

Evaluation der Wirksamkeit von Corona-Maßnahmen in Deutschland: Was der RKI-Abschlussbericht sagt – und was er nicht sagt

Technischer Report

von Werner Baumgarten, Oliver Beige, Daniel Haake, Johannes Merkl und Thomas Wieland

Version 1.2. Erweitert am 02.09.2023.

E-Mail-Kontakt Autoren: geowieland@googlemail.com

Inhalt

Nichttechnische Zusammenfassung	3
Einleitung	4
1 Hypothesentesten und Kausalinferenz in der RKI-Studie	5
1.1 Grundlagen: Kanonische Methodenansätze in der empirischen Kausalinferenz	5
1.2 Der Methodenansatz der RKI-Studie	6
2 Modellierung dynamischer Ausbreitung mit Eingriffen: der Modellansatz des RKI	9
2.1 Modellansatz des RKI: die abhängige Variable R_t	10
2.1.1 Grundcharakteristika des Beschleunigungsmaßes R_t	10
2.1.2 Spezielle Charakteristika des Beschleunigungsmaßes R_t	11
2.1.3 Referenzdatum des Beschleunigungsmaßes R_t	13
2.1.4 Die Berechnung des Nowcast und dessen Wirkung auf den R-Wert	15
2.1.5 Nicht repräsentativ erhobene Fallzahlen als Datengrundlage	20
2.1.6 Effekt der Ferien auf die Meldung von Fallzahlen	24
2.1.7 Fazit zur Wahl von R_t als abhängige Variable	26
2.2 Modellansatz des RKI: die unabhängigen Variablen	26
2.2.1 Modellierung des saisonalen Verlaufs	27
2.2.2 Annahmen zum Effekt neuer Varianten	29
2.2.3 (Nicht-)Berücksichtigung von beschränktem Wachstum	34
2.2.4 Replikation des Modells	36
2.2.5 Fazit zur Konstruktion des Modells und deren inhaltliche Implikationen	40
2.3 Modellansatz des RKI: Annahmenverletzungen in der Modellanalyse	40
2.3.1 Umgang mit Multikollinearität bei den Maßnahmenvariablen	40
2.3.2 Fehlende Voraussetzungen für die statistische Inferenz	41
2.3.3 Räumliche Effekte fehlen	46
3 Zur inhaltlichen Interpretation der Ergebnisse	47
3.1 Maßnahmen "wirken" schon vor ihrer Einführung	47
3.2 Einfluss von Maßnahmenbündeln: Sind NPI im Verbund besonders effektiv?	50
3.3 Die "paradoxen Effekte"	52
3.4 Einfluss der Impfung auf den R-Wert	53
4 Fazit	56
5 Literatur	57
Autoren	60

Nichttechnische Zusammenfassung

Am 20.07.2023 veröffentlichte das Robert-Koch-Institut einen Projektabschlussbericht zum StopptCOVID-Projekt, in dem die Effektivität der Corona-Maßnahmen in Deutschland untersucht werden soll. Diese Studie hat den Anspruch, die Wirksamkeit einer großen Menge einzelner pharmazeutischer (Impfung) und nicht-pharmazeutischer Interventionen (z.B. Schulschließungen, Kontaktbeschränkungen) auf das Infektionsgeschehen zu überprüfen.

Hierbei wird ein übliches statistisches Analyseverfahren – ein sog. Regressionsmodell – verwendet, mit dem der zeitliche Verlauf der Corona-Wellen – gemessen anhand des bekannten „R-Wertes“ – zu erklären versucht wird. Die Modellanalyse liefert Ergebnisse, die dahingehend interpretiert werden, dass nahezu jede getroffene Maßnahme den gewünschten Effekt – also eine Minderung des Infektionsgeschehens bzw. eine Senkung des R-Wertes – hatte. Darüber hinaus wird in der Studie festgestellt, dass bei vielen Maßnahmen die gewünschte Wirkung bereits vor ihrem Inkrafttreten eingetreten ist, was mit freiwilligen Verhaltensänderungen – genauer: einer Antizipation der Maßnahmen durch die Bevölkerung – erklärt wird. Weiterhin wird ausdrücklich darauf hingewiesen, dass Maßnahmen nur im Verbund mit anderen Maßnahmen ihre volle Wirkung entfalten.

Da es sich bei der RKI-Untersuchung um eine quantitative Modellierungsstudie handelt, die den Anspruch erhebt, kausale Einflüsse von Maßnahmen auf das Infektionsgeschehen nachzuweisen, ist es notwendig, das zu Grunde gelegte methodische Konzept sowie die aus den Modellergebnissen getroffenen Schlussfolgerungen einer kritischen Prüfung zu unterziehen, wie es beispielsweise in Begutachtungsprozessen bei wissenschaftlichen Fachzeitschriften unternommen wird. Hierbei zeigt sich, dass die RKI-Studie dem wissenschaftlichen Anspruch, derartige Aussagen zu treffen, nicht genügt.

Erstens wird bereits ein Studienkonzept gewählt, mit dem es gar nicht möglich ist, kausale Einflüsse von Interventionen zu überprüfen; hierfür wäre ein anderer Modelltyp aus dem Bereich der sog. Kausalinferenz von Nöten. Vereinfacht ausgedrückt kann der gewählte Modellansatz nur überprüfen, ob das Infektionsgeschehen *nach* Einsetzen einer Maßnahme geringer war als davor oder danach; es kann damit nicht überprüft werden, ob das Infektionsgeschehen auch *aufgrund* dieser Maßnahme gedrosselt wurde. Doch auch schon im ersten Punkt wird der genannte Nachweis nicht erbracht, da die gewünschte Wirkung zum Teil bereits zeitlich vor dem Einsetzen der Maßnahmen erfolgte.

Zweitens gibt es in der Modellanalyse eine Reihe weiterer Probleme, die, unabhängig von der Wahl des Modelltyps, eine Ableitung belastbarer Aussagen erschweren. Hierzu gehört unter anderem, dass für das Modell Annahmen zur höheren Übertragbarkeit der späteren Virusvarianten getroffen werden, die einfach in das Modell „gesetzt“ werden und die gleichzeitig die Modellergebnisse stark beeinflussen. Die Wirkung von Corona-Maßnahmen wird im vorliegenden Modell nicht ergebnisoffen geprüft, sondern zum Teil bereits durch das Modelldesign vorausgesetzt. Weiterhin werden auch einige Gütekriterien für derartige Modellanalysen, wie sie in der Fachliteratur aus den Bereichen Statistik/Ökonometrie behandelt werden, verletzt bzw. ihre Behandlung wird nicht dokumentiert. Hinzu kommen Probleme in der Berechnung des R-Wertes und Verzerrungen in dessen Datengrundlage.

Drittens sind auch wesentliche Schlussfolgerungen der Studie nicht durch deren eigene Modellergebnisse gesichert. Dass die Bevölkerung Maßnahmen „freiwillig vorweggenommen“ habe, ist lediglich eine Behauptung, die sich nicht ausreichend belegen lässt und für die sich sogar Gegenbeispiele finden lassen. Einige Ergebnisse sind widersprüchlich - zum Beispiel ein kontraproduktiver Effekt von Masken in bestimmten Altersgruppen -, ohne dass diese paradoxen Ergebnisse inhaltlich erklärt werden. Auch die Aussage, wonach Maßnahmen erst „im Verbund“ eine ernsthafte Wirkung haben, ist in dieser Pauschalität bereits logisch nicht nachvollziehbar und wird durch die RKI-Studie auch nicht belegt.

Wir empfehlen daher eine Neubearbeitung der Studie auf der Grundlage des bestehenden Datensatzes (Re-Analyse) durch eine unabhängige Instanz. Vor diesem Hintergrund empfehlen wir weiterhin die Veröffentlichung des zugrunde gelegten Datensatzes sowie des Quellcodes der Auswertung.

Einleitung

Der vom Robert-Koch-Institut (RKI) in Zusammenarbeit mit der Universität Bielefeld veröffentlichte Abschlussbericht, *“Wirksamkeit und Wirkung von anti-epidemischen Maßnahmen auf die COVID-19-Pandemie in Deutschland (StopptCOVIDStudie)”*, ([RKI 2023b](#), nachfolgend “RKI-Studie”) analysiert die Wirksamkeit der staatlich getragenen Eingriffe zur Eindämmung der Ausbreitungsdynamik in der Frühphase der COVID-19 Pandemie, und kommt zu verschiedenen affirmativen Ergebnissen, z.B. *“In der Summe aller NPI führten Verschärfungen jeweils zu einer stärkeren Reduktion der COVID-19 Ausbreitung”* (S. 6) oder *“eine hohe Impfquote [hatte] einen stark reduzierenden Effekt auf den R-Wert”* (S. 6). Abschließend konkludieren die Autoren, dass *“[d]ie in unserer Studie betrachteten NPI wesentlich zur Bekämpfung der Pandemie bei[trugen] und in der Zeit bis zur Entwicklung wirksamer Impfstoffe eine starke Überlastung des Gesundheitssystems [verhinderten]”* (S. 6).

Wir analysieren und bewerten die Aussagen der Studie zu den Themen 1. Methodenansatz, 2. Modellansatz (abhängige Variable, unabhängige Variablen), 3. Ergebnisse und Schlussfolgerungen.

Zusammengefasst konstatieren wir, dass die Studie aufgrund gravierender methodischer Mängel überhaupt nicht zu diesen Ergebnissen kommen konnte, dass das zur Erzeugung der Schlussfolgerungen entwickelte Regressionsmodell weit unterhalb den üblichen wissenschaftlichen Standards konstruiert wurde, damit die formalen Ergebnisse nicht aus dem Modellansatz inferierbar sind, und auch zum Teil die Schlussfolgerungen den formalen Ergebnissen direkt widersprechen. An keiner Stelle kann die Studie die affirmativen Schlussfolgerungen empirisch unterstützen.

Im Einzelnen bemängeln wir: 1. Der verwendete Methodenansatz kann die gestellte Frage nicht beantworten, selbst wenn er sauber ausgearbeitet worden wäre, 2. Das formale Modell enthält mehrere disqualifizierende methodische Fehler, die zu Scheineffekten führen bzw. das Ergebnis z.T. direkt steuern, 3. Den Schlussfolgerungen entgegenstehende Ergebnisse werden ohne Beleg wegerklärt, 4. Die methodischen Fehler und Auslassungen in der Diskussion und Interpretation führen dazu, dass die Studie das, was sie nachzuweisen vermeint, nicht nachweisen kann und disqualifiziert sie daher als wissenschaftliche Arbeit.

Es ist hierbei zu erwähnen, dass die RKI-Studie in manchen Punkten eine Verbesserung gegenüber vielen früheren NPI-Studien darstellt; das betrifft in methodischer Hinsicht v.a. die Wahl eines vergleichsweise langen Untersuchungszeitraums sowie die Tatsache, dass zumindest versucht wird, die Saisonalität im Ausbreitungsgeschehen zu berücksichtigen. Auch in der inhaltlichen Diskussion wird deutlich, dass tendenziell eine differenzierte Sicht auf die Maßnahmen angestrebt wird, wobei sich einige Äußerungen auch als Anreiz für eine “Kosten-Nutzen“-Abwägung bestimmter Maßnahmen lesen lassen (das betrifft v.a. das höchst umstrittene Thema der Schulschließungen).

Diese Aspekte sind jedoch untergeordneter Natur, weil die Schwierigkeiten der RKI-Studie in ihrem “Kern”, nämlich dem methodisch-konzeptionellen Aufbau der ganzen Untersuchung und den inhaltlichen Schlussfolgerungen aus den Untersuchungsergebnissen, liegen.

Wir empfehlen daher eine Neubearbeitung der Studie auf der Grundlage des bestehenden Datensatzes (Re-Analyse) durch eine unabhängige Instanz. Im Sinne guter wissenschaftlicher Praxis (insbesondere im Hinblick auf Reproduzierbarkeit) empfehlen wir weiterhin die Veröffentlichung des zugrunde gelegten Datensatzes sowie des Quellcodes der Auswertung. Wir geben in der Besprechung Empfehlungen zur sachgemäßen Durchführung einer solchen retrospektiven Studie aufgrund existierender Best Practices und belegen die Abweichungen der Studie von diesen Best Practices.

1 Hypothesentesten und Kausalinferenz in der RKI-Studie

1.1 Grundlagen: Kanonische Methodenansätze in der empirischen Kausalinferenz

Zur korrekten empirischen Beurteilung von Eingriffen in ein dynamisches System empfehlen sich zwei grundsätzliche Methodenansätze. Zum einen empfiehlt sich zur Begleitung von Maßnahmen die Erstellung eines repräsentativen Panels aus der Bevölkerung, das vor und während der Wirkungsphase kontinuierlich beobachtet wird. Sowohl der Gesetzgeber als auch das Bundesgesundheitsministerium und das RKI haben es während der COVID-19 Pandemie versäumt, eine solche maßnahmenbegleitende kontrollierte Panelstudie (ein sogenanntes [Randomized Controlled Trial](#) oder RCT) zu initiieren oder durchzuführen.¹

Alternativ zu einem RCT kann man auf den Methodenansatz der natürlichen Experimente und der [Difference-in-Differences Estimation](#) ("Diff-in-Diff") zurückgreifen. In einem solchen Ansatz werden aus den gesammelten Echtweltdaten Szenarien gesucht, die eine Varianz in den Eingriffen zwischen zwei Subpopulationen beinhalten, und den Effekt dieser Varianz auf das darauffolgende Geschehen beobachten (Beispiel: Eine Region hat eine bestimmte Maßnahme eingeführt, eine benachbarte Region nicht). Wichtig zur korrekten Umsetzung eines solchen Ansatzes ist es, dass die Eingriffe exogen zum Geschehen vorgenommen werden und dass die Subpopulationen das *ceteris paribus* Kriterium annähernd erfüllen, dass sie sich also in allen anderen Aspekten stark ähneln und damit die Randomisierung der RCTs simulieren.²

Der Methodenansatz der natürlichen Experimente wurde schon in den 1990er Jahren entwickelt und 2021 auch mit dem [Nobelpreis für Wirtschaftswissenschaften](#) (für David Card sowie Joshua Angrist und Guido Imbens) ausgezeichnet. Insbesondere in den Sozialwissenschaften sind natürliche Experimente und verschiedene Diff-in-Diff-Methoden inzwischen zum Goldstandard für die Kausalinferenz aufgestiegen. Auch wurden bereits mehrere Corona-NPI-Studien mit dem Diff-in-Diff-Ansatz durchgeführt und publiziert (z.B. de

¹ Dieser Kritikpunkt wurde auch im Gutachten des Sachverständigenausschusses (2022) vorgetragen, unter: *Optimierungsbedarfe -> Repräsentative Stichproben*.

² In der zunehmend umfangreicheren und anspruchsvollen Methodenliteratur zu Diff-in-Diff-Methoden reicht es inzwischen häufig aus, Varianzen zwischen Eingriffen, z.B. zeitliche Unterschiede, auszunutzen. Auch können Kontrollgruppen synthetisch aus mehreren Subpopulationen kombiniert werden. Siehe dazu z.B. Imbens & Rubin (2015). Eine solche *synthetische Kontrolleinheit* liegt z.B. der NPI-Studie von Mader & Rüttenauer (2022) zu Grunde.

Haas et al. 2022, Funk et al. 2022). Aufgrund der hohen Datendichte und angesichts des fehlenden repräsentativen Panels empfiehlt sich also eine Diff-in-Diff Studie zur empirischen Bewertung der Kausalhypothese.

1.2 Der Methodenansatz der RKI-Studie

Die Studie des RKI erstellt mehrere Kausalhypothesen zur Wirksamkeit nichtmedizinischer und medizinischer Eingriffe zur Bekämpfung von COVID-19, insbesondere in der Reduktion des Ausbreitungsverhaltens von COVID-19 in der Zeitspanne März 2020 bis August 2021, und sieht diese Kausalhypothesen in der Analyse größtenteils als bestätigt an. Damit fällt die Studie in das Gebiet der Kausalinferenz.

Die Kausalinferenz beschäftigt sich mit der Formalisierung von Behauptungen über Ursache-Wirkungs-Zusammenhängen und deren Verifizierung oder Widerlegung via formaler Methoden aus dem Bereich der Hypothesentestung. Die Grundlage hierbei sind quantitative Daten, die Hypothesentestung erfolgt über statistische und ökonometrische Modelle aus dem Methodenfundus der empirischen Forschung (z.B. Wirtschaftswissenschaften, Psychologie, Epidemiologie).

Dabei werden grundsätzlich zwei Hypothesen gegenübergestellt: die Nullhypothese und die alternative oder kausale Hypothese. Die Kausalhypothese formalisiert einen kausalen Zusammenhang zwischen Eingriff und Effekt in einem beschriebenen System, hier die Implementierung von restriktiven Interventionen zur Eindämmung der Ausbreitungsdynamik in der Frühphase der COVID-19 Pandemie.

Die Nullhypothese ist dabei die Ausgangshypothese, die als gültig angenommen wird, bis sie mit einem hohen Grad statistischer Konfidenz widerlegt werden kann.³ Die Nullhypothese besagt generisch, dass der Effekt nicht auf die Einwirkung der hypothetisierten Eingriffe zurückzuführen ist. In dem Fall der RKI-Studie bedeutet das, dass die Variation in der Ausbreitungsdynamik nicht von der Orchestrierung der restriktiven Maßnahmen abhängig ist. Das bezieht sich explizit auf alle beobachteten Auf- und Abschwünge innerhalb des Beobachtungszeitraums. Die Kausalhypothese negiert diese Aussage der Nullhypothese, indem sie den Einfluss einer Reihe von hypothetisierten Faktoren auf die Ausbreitungsdynamik postuliert. Aufgabe einer Kausalinferenz ist somit, die Nullhypothese zu widerlegen und die Kausalhypothese zu bestätigen.

Zur Widerlegung der Nullhypothese gehören grundsätzlich immer zwei Schritte.

- Erstens muss gezeigt werden, dass auf synchronisierten Zeitreihen der Effekt *nach* dem Eingriff aufgetreten ist ("*post hoc*").
- Zweitens muss gezeigt werden, dass der Effekt nur *aufgrund* des Eingriffs aufgetreten ist, also ohne den Eingriff so nicht aufgetreten wäre ("*propter hoc*").⁴

³ Es gibt in der Hypothesentestung keine formale Refutation einer Hypothese. Eine alternative Hypothese gilt als nicht belegt, bis die Nullhypothese (üblicherweise mehrfach) widerlegt werden konnte.

⁴ Die Bezeichnungen sind aus dem berühmten Fehlschluss "*post hoc, ergo propter hoc*" ("danach, also deswegen") abgeleitet. Die zeitliche Abfolge von Ursache und Wirkung ist eine notwendige Bedingung für einen Kausalzusammenhang, aber erst der Nachweis von Ursächlichkeit macht den Schluss auch hinreichend.

Das bedeutet auch, es muss empirisch belegt werden, dass die Ausbreitungsdynamik sich ohne die Interventionen nicht von selbst (oder durch andere Einflüsse) abgeschwächt hätte (Es wurde empirisch häufig nachgewiesen, dass Corona-Wellen unabhängig von der Eskalationsstufe der nicht-pharmazeutischen Interventionen an Dynamik verloren haben bzw. „gebrochen“ wurden, siehe unter Kap. 3.1). Dies wird nur mit Verweis auf existierende Modellrechnungen impliziert (*„Es war allerdings davon auszugehen, dass der exponentielle Anstieg der COVID-19-Fälle ohne die Implementierung von NPI zu einer massiven Überlastung des Gesundheitssystems geführt hätte“*, S. 9) aber nicht empirisch belegt – ein in der COVID-19 Literatur häufiger auftretender Kategorienfehler: Simulationsmodelle gehören zur Theorie, sie können grundsätzlich nur behaupten, aber nicht empirisch belegen. Die Aufgabe der empirischen Validierung ist es, die Theorie an der Realität zu messen, nicht die Realität an der Theorie. Die Abweichung eines realen Verlaufs von einem modellierten Verlauf sagt uns erstmal nichts über den Erfolg des Eingriffs, sondern sie kann genauso gut die Fehlerhaftigkeit der Modellierung offenbaren.

Das ist insbesondere relevant, weil Eingriffe in das Ausbreitungsgeschehen grundsätzlich endogen sind, d.h. sie werden üblicherweise als Reaktion auf anwachsende Fallzahlen und eine steigende Beschleunigung der Ausbreitung angeordnet.⁵ Wir sollten deshalb eine Häufung von Maßnahmen nahe des Ausbreitungsgipfels beobachten. Dass sie aber für den Rückgang der Ausbreitung verantwortlich sind, muss empirisch bewiesen und nicht impliziert werden. An dieser Aufgabe scheitert die Studie – ein Fehlschluss des Typus *petitio principii*: wesentliche Grundvoraussetzungen der Beweisführung werden axiomatisch als wahr angenommen, obwohl sie selbst der empirischen Validierung bedürfen.

Nach einer erläuternden Einleitung und der Aufreihung der hypothetisierten Einflussfaktoren stellt die RKI-Studie auf S. 26/27 ein formales Modell vor. Es handelt sich um ein lineares bzw. generalisiertes lineares Regressionsmodell⁶ mit dem Logarithmus des Reproduktionsfaktors, „R-Wert“ (auch R_t) als abhängige, also zu erklärende Variable. Ob und inwiefern eine kausale Überprüfung des Einflusses von Interventionen in einer solchen Studie möglich ist, hängt entscheidend davon ab, wie die Maßnahmen ins Modell einfließen.

Die RKI-Studie nutzt hierfür (bevölkerungsgewichtete) Dummy-Variablen für einzelne Maßnahmen: *„Die Maßnahmen wurden ausgehend von binären (0/1) Variablen auf der Landkreis-Ebene bevölkerungsgewichtet gemittelt, so dass ein reeller Wert zwischen 0 und 1 entstand, der den Anteil der Bevölkerung eines Bundeslandes angibt, für den diese Maßnahme mit einer bestimmten Stärke verordnet wurde“* (S. 19)⁷. Für diese Variablen schätzt das Modell Koeffizienten, die etwas über den (statistischen) Einfluss der einzelnen Maßnahmen aussagen: Ist beispielsweise ein Koeffizient für Maßnahme X negativ und statistisch signifikant, bedeutet dies – im Modell – eine Reduzierung des R-Wertes. Hier kommt ein zentrales Problem: Dieses Modelldesign untersucht, ob der R-Wert *in Gegenwart*

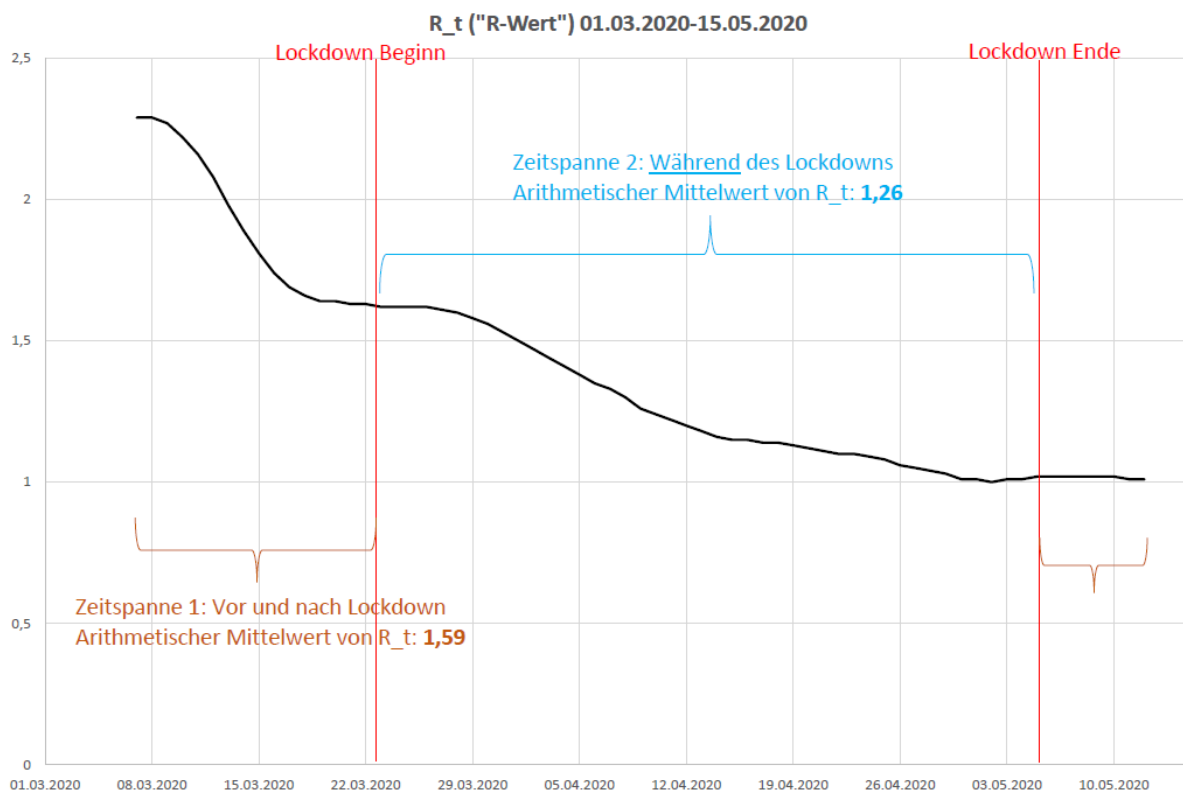
⁵ Ein *endogener* Eingriff erfolgt in Reaktion auf den bisherigen Verlauf und u.U. auf den antizipierten zukünftigen Verlauf eines dynamischen Systems. Ein verlaufsunabhängiger Eingriff heißt *exogen*.

⁶ Hierbei der Hinweis, dass dies nicht bedeutet, dass tatsächlich ein linearer Zusammenhang zwischen den erklärenden Variablen und dem Outcome angenommen wird, denn die abhängige Variable R_t wird logarithmiert (S. 26-27). Im Ergebnis sagen die Modellkoeffizienten etwas über die prozentuale Änderung des R_t -Wertes aus.

⁷ Einzelne Maßnahmen waren zunächst als *Dummy-Variablen* kodiert, d.h. sie nehmen einen Wert von eins an, wenn an einem bestimmten Tag eine bestimmte Maßnahme gegolten hat, und den Wert null, wenn diese Maßnahme an einem bestimmten Tag nicht in Kraft war. Grundlage sind hierbei die Daten der Corona-Datenplattform (<https://www.corona-datenplattform.de/>) – übrigens ein wahrer Datenschatz, der von der Firma infas Geodaten im Auftrag des Wirtschaftsministeriums zusammengestellt wurde.

einer bestimmten Maßnahme *X* *geringer* war als er in Abwesenheit dieser Maßnahme bzw. bei geringerer Maßnahmenstärke war. Es wird *nicht* untersucht, ob der R-Wert nur in Gegenwart dieser Maßnahme *zusätzlich gesunken* ist.

Hier kommt noch ein weiteres Problem hinzu: Da Maßnahmen in der Regel erst ab einem bestimmten Inzidenz-Level eingeführt wurden und der (reale) R-Wert in einer Infektionswelle so oder so mit der Zeit sinkt (siehe unter Kap. 2.2.3), haben wir im absoluten Regelfall ohnehin höhere R-Werte *vor* Inkrafttreten der meisten Maßnahmen als danach. Das kann dazu führen, dass gerade den vollständigen Lockdowns (z.B. Schul-, Kita- und Geschäftsschließungen ab dem 16.12.2020) eine besonders hohe Bedeutung zugeschrieben wird.



Mittelwertunterschiede des R-Wertes von Schweden vor, nach und während des deutschen Lockdowns in der ersten Corona-Welle. Quelle: OWID, eigene Berechnungen.

Die obige Abbildung erläutert das Prinzip, wobei anhand des Beispiels sehr schnell deutlich wird, dass eine zeitliche Koinzidenz nicht notwendigerweise eine Kausalität bedeutet: Es wird der *schwedische* R-Wert von Anfang März bis Mitte Mai 2020 gezeigt, d.h. aus einem Land mit wesentlich geringerer Maßnahmenintensität, in dem auf einen Lockdown vollständig verzichtet wurde. Auch der gezeigte schwedische R-Wert ist *im Mittel* während des *deutschen* Lockdowns deutlich geringer (Arithmetischer Mittelwert von 1,26) als in der Zeitspanne davor und danach (Arithmetischer Mittelwert von 1,59). Dies *kann* natürlich nicht auf einen Lockdown zurückzuführen sein, da Schweden nie einen solchen etabliert hat; es hängt stattdessen (auch) damit zusammen, dass R_t unabhängig von Mitigationsmaßnahmen zwangsläufig sinkt. Das bedeutet wiederum *nicht*, dass staatliche Maßnahmen per se keinerlei Wirkung auf R_t haben können (hierzu siehe auch Kap. 2.2.3), sondern nur, dass

allein von einem solchen Mittelwertunterschied - wie er im RKI-Modell betrachtet wird - noch *keine* kausale Wirkung dieser Maßnahmen abgeleitet werden kann.

Um eine tatsächliche kausale Wirkung zu prüfen, müsste auch der obige zweitgenannte Punkt (signifikant *stärkere Senkung* des R-Wertes als vor Inkrafttreten der Maßnahme) Berücksichtigung finden; mathematisch ausgedrückt wären im Modell Interaktionen von Zeit und Maßnahmen als zusätzliche Variablen (also [Interaktionsterme](#)) von Nöten, was das Grundprinzip der o.g. Difference-in-Differences-Methode zur Feststellung kausaler Einflüsse wäre. Der zu Grunde liegende RKI-Modellansatz *kann* demnach gar keine Maßnahmeneinflüsse überprüfen, weil er de facto keine Kausalanalyse darstellt. Es gab im Kontext der empirischen Prüfung der Wirkung von Corona-Maßnahmen einige Studien, die einen Diff-in-Diff-Ansatz verwendet haben (z.B. de Haas et al. 2022, Funk et al. 2022), im vorliegenden Fall wurde es jedoch unterlassen.

Die Studie versucht also nur die erste der Fragen, nämlich nach der zeitlichen Reihenfolge (*post hoc*), zu beantworten, und scheitert dabei, zum Teil nach eigener Erkenntnis, ohne allerdings die Schlussfolgerung anzupassen. Selbst im Studienansatz wird erkennbar, dass die meisten Effekte vor den Maßnahmen eintraten (siehe unter Kap. 3.1). Die zweite Frage nach dem Kausalzusammenhang (*propter hoc*) wird weder angesprochen noch beantwortet. Damit ist die Schlussfolgerung der Kausalität überhaupt nicht aus den Studienergebnissen ableitbar. Diese Omission ist alleine disqualifizierend.

Zur korrekten empirischen Bewertung der Kausalität ist es grundsätzlich erforderlich, zwei Realitäten miteinander zu vergleichen (kanonisch, eine Realität mit Eingriff gegen eine Realität ohne Eingriff, in der Literatur als *Treatment vs Control* bekannt) statt eine Realität gegen eine Modellierung oder eine andere Fiktion. Diese Vorgehensweise kann bestenfalls die Fehlerhaftigkeit der Modellierung bewerten, nicht aber die Kausalität der Eingriffe auf das beobachtete Geschehen. Auch dies ist ein disqualifizierender Methodikfehler.

Wir schließen in diesem Abschnitt mit der Erkenntnis, dass, selbst bei sauberer Ausarbeitung des formalen Modells, dieses nicht in der Lage wäre, die Frage der Kausalität zu beantworten, da es grundsätzlich einen falschen Methodenansatz verfolgt. Selbst wenn das gewählte Modell formal korrekt konkludieren würde, dass sich *nach* den Maßnahmen die Ausbreitungsdynamik reduzierte, könnte es methodisch nicht nachweisen, dass dies *aufgrund* der Maßnahmen geschah.

2 Modellierung dynamischer Ausbreitung mit Eingriffen: der Modellansatz des RKI

Neben der grundlegenden Problematik, dass der gewählte Methodenansatz bestenfalls Aussagen über das *nach*, aber nicht das *aufgrund* machen kann (der sogenannte [post hoc, ergo propter hoc](#) Fehlschluss), sorgen mehrere grundsätzliche methodische Fehler im Modellansatz dafür, dass Erkenntnisse über die zeitliche Sequenz und die Kausalität nicht oder nicht ausreichend empirisch belegt sind, oder aufgrund der Modellspezifikationen reine Scheineffekte sind. Neben der Methodenauswahl empfehlen wir auch, die Modellkonstruktion komplett neu entwickeln zu lassen.

2.1 Modellansatz des RKI: die abhängige Variable R_t

2.1.1 Grundcharakteristika des Beschleunigungsmaßes R_t

Das formale Modell der RKI-Studie verwendet den Logarithmus des Reproduktionsfaktors, "R-Wert" (auch R_t) als abhängige, also zu erklärende Variable. Die Grundlage des R-Wertes bilden die gemeldeten SARS-CoV-2-Infektionsfälle⁸. Statt einem R-Wert nach Meldedatum verwendet die Studie einen auf den imputierten Erkrankungsbeginn zurückgerechneten R-Wert.⁹

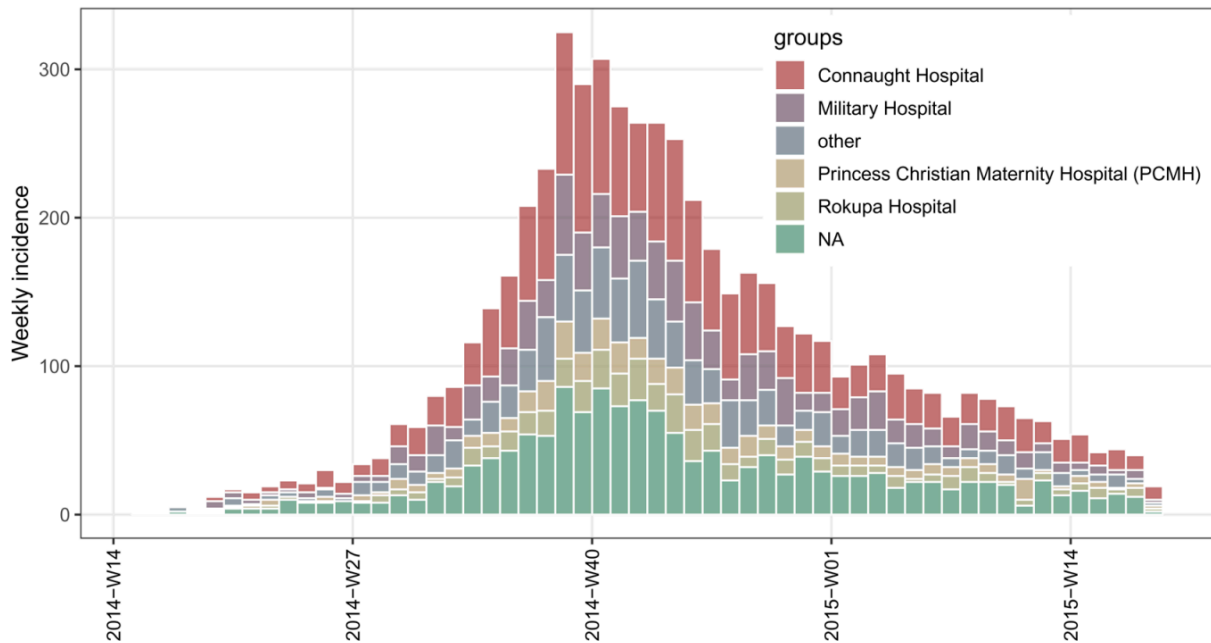
Als Motivation für die Wahl von R_t werden zwei Hauptargumente angeführt. Beide sind methodisch fragwürdig.

Zum ersten wird behauptet, dass "[a]ufgrund der effektiven Mensch-zu-Mensch-Übertragung des SARS-CoV-2 Virus [...], die durchschnittliche Zahl von Folgeinfektionen, die sogenannte Reproduktionszahl (R-Wert) eine zentrale Rolle [spielt]. Ist der R-Wert größer als eins, so führt das zu einem exponentiellen Anstieg der Fallzahlen." (S. 10)

Der zweite Satz enthält einen weitverbreiteten, aber für die Bewertung folgenreichen Semantikfehler. Ein R-Wert über 1 entspräche nur dann einer exponentiellen Ausbreitung, wenn der R-Wert über 1 und *konstant*, also unabhängig von t wäre.

⁸ In der RKI-Studie wird demnach bewusst ein Indikator verwendet, der auf der Meldeinzidenz basiert; dies wird damit begründet, dass die Corona-Maßnahmen darauf abzielten, (infektiöse) Kontakte zu reduzieren (S. 10). Es stimmt zwar, dass die Maßnahmen als *Mittel* zur Kontaktreduzierung etabliert wurden, als eigentliches *Ziel* der Interventionen wurde jedoch stets formuliert, dass hierdurch die Überlastung des Gesundheitssystems durch eine Überzahl an schweren, stationär behandlungsbedürftigen COVID-19-Verläufen verhindert werden sollte. Es stellt sich vor diesem Hintergrund die Frage, warum nicht andere Indikatoren als die Meldeinzidenz bzw. die daraus abgeleitete Reproduktionszahl zugrunde gelegt wurden, die das eigentliche Ziel der COVID-19-Maßnahmen zum Gegenstand haben. Es wäre z.B. möglich gewesen, aus den RKI-Rohdaten die registrierten COVID-19-Todesfälle zu extrahieren. Eine weitere Alternative wären etwa die DIVI-Daten zur Belegung bzw. zu Neueinweisungen von SARS-CoV-2-Infizierten gewesen. Alle genannten Daten sind mindestens auf Bundesländerebene – was die Untersuchungsebene der RKI-Studie ist – verfügbar, meist sogar auf der kleinräumlichen Ebene der Landkreise (oder lassen sich auf diese Ebene aggregieren). Daten zu beispielsweise Todesfällen oder ITS-Belegungen wären also als abhängige Variable mindestens genauso gut begründbar gewesen, zumal sich die RKI-Studie in ihrer Zusammenfassung wiederum explizit auf die Auswirkungen auf das Gesundheitssystem bezieht: "Die in unserer Studie betrachteten NPI trugen wesentlich zur Bekämpfung der Pandemie bei und verhinderten in der Zeit bis zur Entwicklung wirksamer Impfstoffe eine starke Überlastung des Gesundheitssystems." (S. 6). Die Schweizer Studie zur Evaluation der Corona-Maßnahmen, die einen vergleichbaren Ansatz hat, verwendet stattdessen Krankenhauseinweisungen (Funk et al. 2022).

⁹ Wir gehen im Folgenden davon aus, dass zur Rückrechnung der vom RKI täglich gemeldete Nowcast verwendet wurde. Hauptanlass zu dieser Annahme ist, dass einer der mit dem RKI affilierten Studienautoren an der Konstruktion des Nowcasts maßgeblich beteiligt war. Die statistische Imputation des Erkrankungsbeginns war in allen Fällen nötig, die asymptomatisch verliefen oder für die kein Erkrankungsbeginn gemeldet wurde. Dieser Anteil stieg im Verlauf der Pandemie kontinuierlich, was die Nowcast zunehmend erratisch und unzuverlässig machte.



Kanonisches Beispiel für eine epidemische Kurve. Quelle; Kamvar et al. (2019).

Eine Exponentialfunktion $f(t) = e^t$ und deren Ableitungen haben aber keine Extrem- oder Wendepunkte, die man für eine Vorhersage des weiteren Kurvenverlaufs verwenden könnte. Im Gegensatz zur Exponentialfunktion enthält die kanonische epidemische Kurve Extrem- und Wendepunkte, wie in dem Beispiel leicht erkennbar. Ob ein Eingriff in das Ausbreitungsgeschehen diese Extrempunkte verändert, muss aber empirisch nachgewiesen werden.

R_t , die Reproduktionsrate, ist formal ausgedrückt das Verhältnis von Beschleunigung zu Geschwindigkeit des Ansteckungsgeschehens plus 1, und damit ein kalibriertes Maß für die Ausbreitungsbeschleunigung. Dies steht im Gegensatz zu Maßen wie der täglichen Fallrate oder der Inzidenz, die Geschwindigkeitsmaße darstellen. Es ist also zu erwarten, und dies ist leicht empirisch validierbar, dass R_t grundsätzlich vor diesen Maßen sein Maximum erreicht und überschreitet. R_t ist somit ein Frühindikator und damit ein Prädiktor für das weitere Ausbreitungsgeschehen.

Ob dieser Umkehrpunkt R_t max aber durch Eingriffe oder durch natürlichen Verlauf entstand, muss durch die zwei Schritte der Hypothesentestung ermittelt werden. Um die Nullhypothese des natürlichen Verlaufs zu widerlegen, muss sowohl das *post hoc* als auch das *propter hoc* methodisch richtig und mit hoher statistischer Konfidenz etabliert werden. Insbesondere letzteres kann der Methodenansatz aber gar nicht leisten.

2.1.2 Spezielle Charakteristika des Beschleunigungsmaßes R_t

Ein zweite Motivation für die Wahl von R_t als abhängige Variable ist damit auch methodisch nicht ausreichend korrekt dargelegt:

“Eine weitere günstige Eigenschaft des R-Werts als Zielvariable ist, dass er deutlich unabhängiger von der Teststrategie ist als die 7-Tage-Inzidenz, da er als Quotient zweier nur wenige Tage auseinander liegender Werte der 7-Tage-Inzidenz gebildet wird.” (S. 11).¹⁰

Als Beschleunigungsmaß und damit Ableitung der Inzidenz als Geschwindigkeitsmaß ist R_t im Gegenteil zu dieser Behauptung ganz besonders von dem insbesondere in der Frühphase rapiden Anstieg der Testzahlen betroffen. Dieser Amplifikationseffekt wird noch zusätzlich verstärkt, weil R_t der Quotient von Beschleunigung und Geschwindigkeit des Ausbreitungsgeschehens ist. Das ist auch in dem auf Seite 12 der Studie abgebildeten Diagramm sehr gut erkennbar.



Verlauf der 7-Tage-Inzidenz und des R-Wertes im Analysezeitraum in Deutschland. Abbildung 4, S. 12 (RKI 2023b). Der kurzfristige Peak von $R_t > 3$ im März 2020 entstand in einer Phase, in der das Testvolumen deutlich ausgeweitet wurde.

Wenn man, empirisch plausibel, davon ausgeht, dass die 7-Tage-Inzidenz in den frühen Tagen des Beobachtungszeitraums durch rapide ansteigende tägliche Testzahlen verzerrt wurde, ist empirisch nicht separierbar, dass die Anomalie im März 2020 im R-Wert ein Meldeartefakt war, verstärkt durch die Konstruktion von R_t .¹¹ Diese Erkenntnis wird anekdotisch unterstützt dadurch, dass R_t -Spitzen dieser Größenordnung weltweit in allen Ländern einmalig blieben, unabhängig vom lokalen Maßnahmenregime.

¹⁰ Dass Teststrategie und das Testverhalten Auswirkungen auf die 7-Tage-Inzidenz haben, bemerkte bereits der Sachverständigenausschuss zur Bewertung der Corona-Maßnahmen an. Er hat darauf hingewiesen, dass die 7-Tage-Inzidenz ein Wert ist, der von der Teststrategie und damit auch vom Testverhalten abhängig ist: “Dieser Wert ist insofern problematisch, als dass es sich um die 7-Tage-Melderate pro 100.000 Einwohner handelt, die abhängig von Teststrategien ist, nicht alle positiven Testergebnisse (etwa Antigen-Schnelltests) erfasst und die Dunkelziffer an Infektionen nicht berücksichtigt. [...] Eine valide Inzidenz, also die tatsächliche Erkrankungsrate, kann nur mit repräsentativen Zufallsstichproben ermittelt werden bzw. durch Modellrechnungen abgeschätzt werden.” (Sachverständigenausschuss nach § 5 Absatz 9 Infektionsschutzgesetz - Bundesgesundheitsministerium, S. 44f.).

¹¹ Ein kanonischer Parameter für einen Ausbreitungsprozess besteht aus der Differenz von neuen Fällen minus alten Fällen (Beschleunigung), geteilt durch alte Fälle (Geschwindigkeit). R_t addiert 1 dazu, also neue Fälle durch alte Fälle. Bei einer starken Divergenz zwischen alter und neuer Messung wirkt sich diese durch den Nennerterm besonders verzerrend aus.

Die Behauptung der Autoren, dass R_t weniger vom Testregime beeinflusst wird als die Geschwindigkeitsmaße Fallzahlen und Inzidenz, ist damit für diesen relevante Phase des Ausbreitungsgeschehens nicht belegt.

2.1.3 Referenzdatum des Beschleunigungsmaßes R_t

Ein in der COVID-19 Literatur häufig ignoriertes Mandat für die methodisch korrekte empirische Validierung von Kausalhypothesen ist die Notwendigkeit, zwei Zeitreihen miteinander zu synchronisieren: die der viralen Ausbreitung und die der Maßnahmen.

Eine Kausalanalyse muss zur zeitlichen Abfolge demonstrieren, dass: 1. Ansteckungen mit Verzug beobachtet werden, 2. Maßnahmen konzipiert, verordnet und umgesetzt werden, 3. eine Verhaltensänderung in der Bevölkerung eintritt, und 4. diese sich wieder mit Verzug in den laufenden Werten wiederfindet.¹²

Dazu müssen die Ereignisse in den beiden Zeitreihen auf das gleiche Referenzdatum geschrieben werden, um überhaupt Schlüsse über die zeitliche Abfolge von Ausbreitung, Maßnahmen und Wirkung, also das *post hoc*, machen zu können. Nur dann können wir auch Fragen über die Verursachung beantworten.

Das relevante Referenzdatum ist das Ansteckungsdatum. Dies ist aber nur in Ausnahmefällen direkt beobachtbar und wird daher meist rückwirkend aus den beobachtbaren Daten errechnet. Eine kanonische Vorgehensweise ist dabei, das Erkrankungsdatum zu notieren und über den Abzug des Erwartungswertes der Inkubationszeit (bei rigoroseren Studien auch die Verteilung) das Ansteckungsdatum zu imputieren. Schätzer für die Inkubationszeit liegen meist bei 4–5 Tagen.

Damit wird das Erkrankungsdatum zum ersten gemeinhin beobachtbaren Referenzdatum. Dieser Wahl schließt sich auch das RKI an.¹³ Grundsätzlich ist die Wahl methodisch akzeptabel unter drei Voraussetzungen: 1. der Verzug von Ansteckung zu symptomatischer Erkrankung ist statistisch sauber erfasst, 2. das Erkrankungsdatum ist häufig genug beobachtbar und in der Meldkette sauber erfasst und weitergegeben, so dass für nur eine Minderheit der erfassten Fälle das Erkrankungsdatum imputiert werden muss, und 3. die Erkrankung wird dann in der Kausalanalyse auf das Ansteckungsdatum zurückgerechnet und damit mit der zeitlichen Abfolge der Maßnahmen synchronisiert.

Die Korrektur von Verzügen passiert in der Studie des RKI an verschiedenen Orten, aber nur unzureichend und unsystematisch, eine nachvollziehbare Abhandlung der Abfolge der Ereignisse fehlt. Damit ist die Validität der meisten Hypothesen in Bezug auf ihre zeitliche Abfolge fraglich, die oben erstellte Sequenz der Ereignisse wird nirgends etabliert.

Die Wahl der Indizierung des R-Wertes verdeutlicht dieses Vorgehen. Zum theoretischen Hintergrund von R_t in der Schätzung von Ausbreitungsdynamik, siehe z.B. Fraser (2007).

¹² Es ist nicht grundsätzlich abwegig, eine Hypothese zu postulieren, dass die Bevölkerung direkt auf den Anstieg der Fallzahlen oder auf die Ankündigung der Maßnahmen mit einer Verhaltensänderung reagiert, also die Maßnahmen antizipiert. Dies ist aber eine grundsätzlich andere Hypothese als die in der RKI-Studie postulierte. In allen Fällen muss eine Hypothese immer sauber ausgearbeitet werden.

¹³ Andere Meldebehörden, so z.B. das Vereinigte Königreich, verwenden das Datum des ersten positiven Tests.

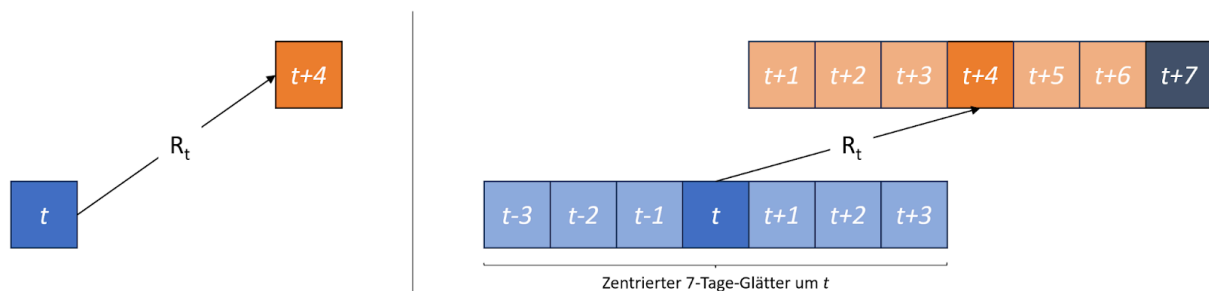
Der R-Wert wurde definiert als 4-Tages-Wachstumsrate der 7-Tage-Inzidenz, die anhand der gemäß IfSG an das Robert Koch-Institut gemeldeten COVID-19-Fälle täglich berechnet wurde. Es ergibt sich also folgende Formel zur Berechnung

$$R_t := \frac{\#neue\ Fälle_{(t-6,t)}}{\#neue\ Fälle_{(t-10,t-4)}}$$

Um die Einschätzung der zeitlichen Effekte auf den R-Wert zu verbessern, wurde anstelle der gemeldeten 7-Tage-Inzidenz die 7-Tage-Inzidenz nach dem Erkrankungsbeginn gewählt.

Definition der Beschleunigungsrate R-Wert/ R_t (S. 11).¹⁴ Alle Textzitate aus der RKI-Studie sind zur besseren Kenntlichmachung invertiert.

Der Reproduktionsfaktor R_t ist ein Schätzer für die Veränderung der Fallzahlen von einem Referenzzeitpunkt auf einen nachfolgenden Zeitpunkt. Kanonisch wird für den Verzug eine Approximation der Inkubationsphase angesetzt, im Modell der Studie wie in vielen anderen COVID19-Modellen 4 Tage. Zusätzlich wird zur Glättung der Wocheneffekte jeweils ein Intervall genommen, im Modell sinnhaft jeweils 7 Tage für Referenz- und Effekterm. Die Studie verwendet als Referenz Erkrankungsbeginn, eine Rückrechnung auf Ansteckungsdatum wird für das Hauptmodell nicht vorgenommen.¹⁵



Korrekte Konstruktion des R-Wertes für Zeitpunkt t in zwei Schritten: 1. Definition von R_t als Quotient von Systemzustand zum Zeitpunkt $t+4$ über Zeitpunkt t . 2. Glättung von Zähler und Nenner über einen zentrierten gleitenden Mittelwert über 7 Tage. Die RKI-Studie verwendet den Beobachtungstag $t+7$ als Referenzdatum, ohne den Verzug zu protokollieren. Korrekt wären t oder unter Umständen $t+2$.

Wo die Definition des R-Wertes irrt, und diese Irrung hat signifikante Konsequenzen auf die Validität der *post hoc* Aussagen, ist bei der Wahl des Referenzdatums t . In der obigen Formel ist t der letzte Tag des Effekterms, damit ist R_t als nachlaufendes Maß definiert. Dies ist inkorrekt und muss korrigiert werden.

Zum ersten bezieht sich das Referenzdatum auf das auslösende und nicht auf das Wirkungsintervall. Zum zweiten muss das 7-Tage-Glättungsintervall um das Referenzdatum t gebildet werden (zentrierter gleitender Mittelwert) und nicht vorlaufend. Der korrekt datierte, in erster Näherung unverzerrte R_t müsste also wie folgt aussehen.

¹⁴ R_t stellt das *Wachstum* der täglichen Fallzahlen dar, also die *Beschleunigung* der kumulativen Fallzahlen.

¹⁵ In Abbildung 32 (S. 48) stellt die Studie ein Zusatzmodell mit einer "rückprojizierten Inzidenz nach Infektionszeitpunkt", mit erkennbar verringertem R^2 (52,5% vs 62,7%) vor. Wir interpretieren die Einlassung dahingehend, dass das Hauptmodell grundsätzlich keine Rückdatierung vom Erkrankungsdatum zum Ansteckungsdatum vornimmt.

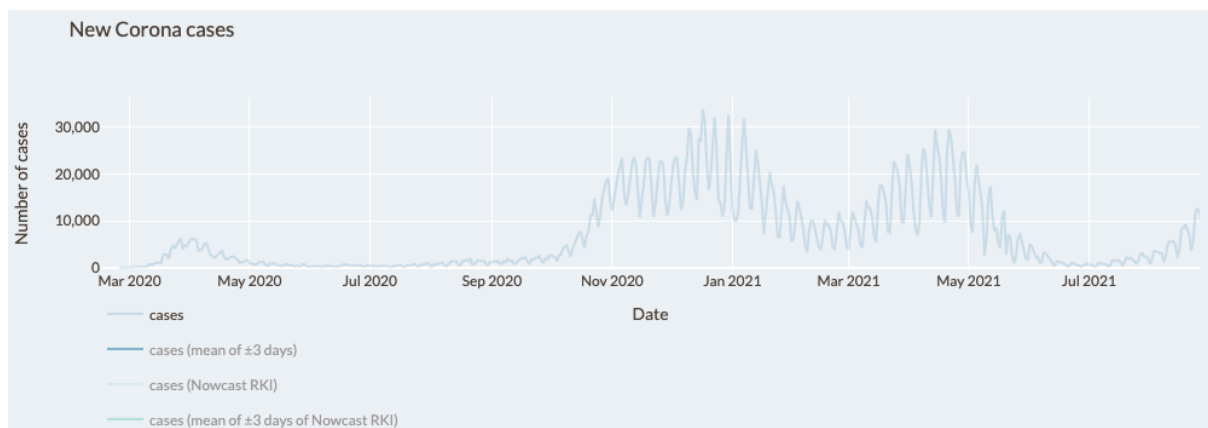
$$R_t = \frac{\#neue\ F\ddot{a}lle(t+1, t+7)}{\#neue\ F\ddot{a}lle(t-3, t+3)}$$

Diese subtile Korrektur verschiebt das Referenzdatum um 7 Tage, von Tag 11 des betrachteten Intervalls auf Tag 4.¹⁶ Ohne Korrektur dieses Verzugs mssen alle Aussagen ber Zeitabfolge invalidiert werden. Wenn man noch die fehlende Korrektur fr die Inkubationsphase mit einbezieht, erhht sich der nicht korrigierte Verzug der Ausbreitungszeitleiste gegenber der Manahmenzeitleiste um geschtzte 11 Tage. Mit einer solchen Diskrepanz in den Zeitreihen sind Aussagen ber Manahmen vor Wirkung (*post hoc*) nicht mehr haltbar. Grundstzlich fehlt eine klare Ausarbeitung und Darstellung, wie Zeitachsen synchronisiert wurden, komplett¹⁷.

Es ist dabei festzustellen, dass die Inkubationszeit im Hauptmodell nicht bercksichtigt wurde (vgl. Aussage auf S. 45). Fr die berprfung einer Manahmenwirkung ist die Verschiebung von R_t um die Inkubationszeit allerdings zwingend notwendig, da hierfr der Infektions- und nicht der Erkrankungsbeginn betrachtet werden muss.

2.1.4 Die Berechnung des Nowcast und dessen Wirkung auf den R-Wert

Die tglich gemeldeten Flle unterlagen mehreren Problemen. Einerseits unterlagen sie einem starken Wochentagseffekt aufgrund der Meldekette von Teststellen ber Gesundheitsmter bis zum RKI, was zu starken Schwankungen fhrte (vgl. Abbildung). Andererseits gab es das Problem, dass hierbei der zeitliche Verlauf des Infektionsgeschehens deutlich versptet dargestellt wurde.

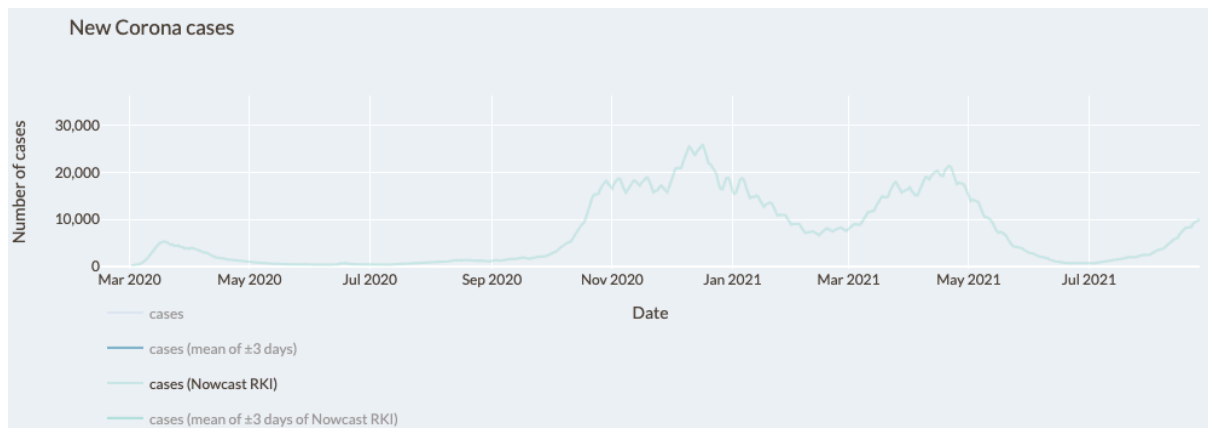


¹⁶ Eine alternative Definition verwendet den Mittelpunkt des ganzen Beobachtungszeitraums, hier $t+2$, als Referenzdatum. Dies hat den Vorteil, dass Geschwindigkeits- und Beschleunigungscharts synchronisiert sind: ein Minimum oder Maximum in den Fallzahlen entspricht einem Durchlauf der 1-er Linie im R_t . Dies wrde immer noch einen nicht protokollierten Verzug von 5 Tagen fr das Referenzdatum des RKI bedeuten.

¹⁷ Weiterhin ist zu erwhnen, dass gewisse Bundeslnder zu bestimmten Tagen aus der Analyse rausgestrichen wurden, wenn dort die Fallzahlen so gering waren, dass sich kein sinnvoller R -Wert berechnen lie (S. 11). Das ist mathematisch nachvollziehbar, da die Berechnung eines Quotienten, was der R -Wert des RKI ist, bei einer null im Nenner (= null Infektionen in der letzten Periode) nicht definiert ist und bei einer null im Zhler (= null Infektionen in der betrachteten Periode) wenig sinnvoll ist, da ein realer R_t -Wert bei einer existierenden Infektionskrankheit nicht bei null liegen kann. Es fhrt aber dazu, dass ausgerechnet Beobachtungen aus der Analyse ausgeschlossen wurden, in denen kaum Infektionen auftraten und gleichzeitig nur wenige oder gar keine Manahmen gegolten haben. Welchen Einfluss dieser Umstand auf die Modellergebnisse hat, ist ungeklrt, aber sollte nicht unbeachtet bleiben.

Visualisierung der neuen Fälle nach Meldedatum. Eigene Darstellung, Datenquelle: RKI.

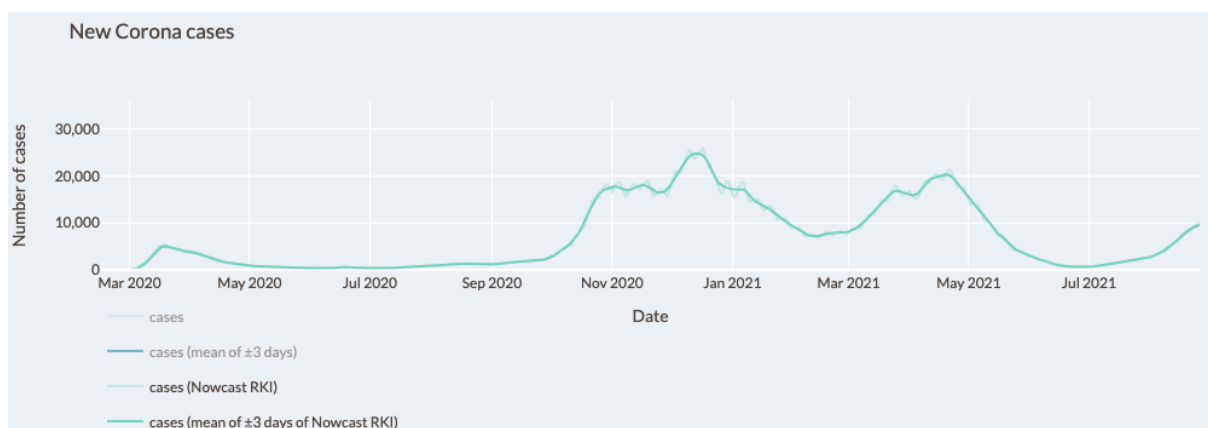
Bei den gemeldeten Fällen war bei einigen von ihnen ein Erkrankungsbeginn gemeldet worden, bei anderen hingegen nicht. Das konnte vorkommen, wenn diese Information entweder nicht erhoben wurde oder durch einen asymptomatischen Verlauf nicht bekannt war. Aus diesen Gründen entwickelte das RKI den sog. Nowcast. Dabei handelt es sich um eine Imputierung des Erkrankungsbeginns von den gemeldeten Fällen mit unbekanntem Erkrankungsbeginn, so dass dann für alle gemeldeten Fälle ein entweder gemeldeter oder zumindest approximierter Erkrankungsbeginn je Fall vorlag. Dadurch kann eine Visualisierung der gemeldeten Fälle nach (imputiertem) Erkrankungsbeginn erfolgen.



Visualisierung des Nowcast des RKI.

Ein entscheidender Parameter für die Approximation des Erkrankungsbeginns war das Datum des Eingangs der Fallmeldung am RKI ([RKI 2020](#)). Das unterlag aber weiterhin dem Wochentageeffekt, so dass auch der Nowcast innerhalb der Woche deutlich oszillierte, wie in der Abbildung zu sehen, wenn auch nicht mehr so stark wie die vom RKI täglich gemeldeten Fälle.

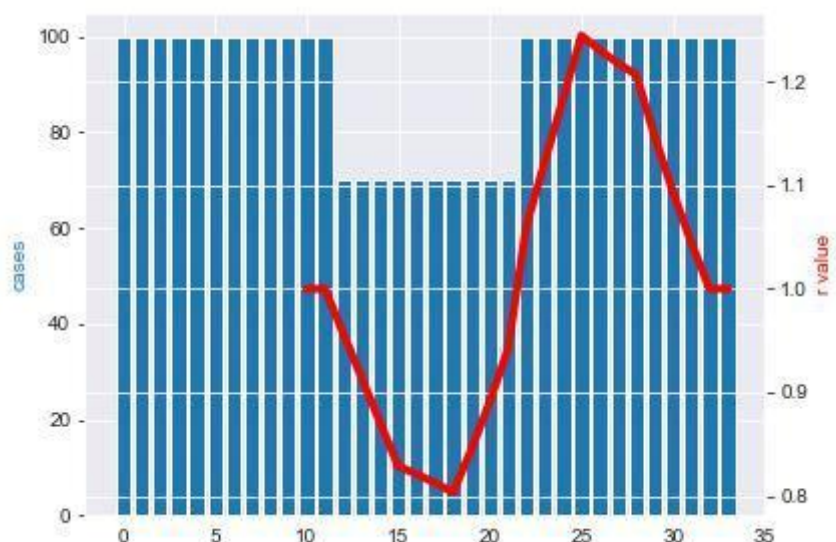
Besser wäre es daher gewesen, dem Wochentageeffekt durch eine Glättung in Form eines 7-Tage-Mittelwertes des Nowcast entgegenzuwirken.



Visualisierung des Nowcast des RKI mit überlagertem 7-Tage-Mittelwert zur Glättung.

Diese Wochentagsschwankungen haben selbstverständlich auch eine Auswirkung auf den berechneten R-Wert. Das hat auch das RKI (teilweise) erkannt. Zunächst hatten sie für die Berechnung des R-Wertes die letzten 4 Tage mit den 4 Tagen davor ins Verhältnis gesetzt. Da hierbei an den verschiedenen Tagen aber immer unterschiedliche Wochentage in die Berechnung einfließen und das aufgrund der Wochentagsschwankung zu starken Schwankungen im R-Wert führte, stellte das RKI auf den 7-Tage-R-Wert um. Dabei wurden 7-Tage-Zeitreihen betrachtet, wobei der Nenner um die Generationszeit von 4 Tagen verschoben war (siehe Erläuterung unter 2.1.3). Das hatte den Vorteil, dass sowohl im Zähler als auch im Nenner nun jeweils 7 Tage berücksichtigt wurden und damit dem Wochentageffekt entgegengewirkt wurde.

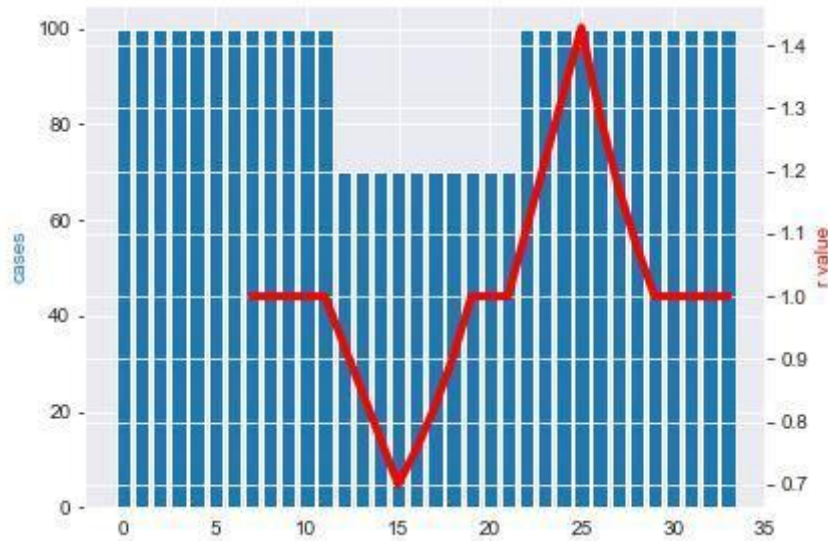
Allerdings führte das dazu, dass sich betrachtete Tage überlappten und sowohl im Zähler als auch im Nenner vorhanden waren. Das hat nun eine Wirkung auf die Berechnung des R-Wertes.



Wirkung eines angenommenen abrupten Abfallens der Fallzahlen auf den 7-Tage-R-Wert.

Nehmen wir an, wir hätten eine Zeit lang täglich konstant 100 Fälle pro Tag. Abrupt gibt es nun nur noch 70 Fälle pro Tag. Es wäre davon auszugehen, dass mit etwas Zeitverzug ein R-Wert von 0,7 erreicht wird. Wie aus der Abbildung zu erkennen ist, wird aufgrund der sich im Zähler und Nenner überlappenden Zeitreihen in diesem Beispiel jedoch minimal ein Wert von ca. 0,8 erreicht. Die Wahl der Berechnungsmethode des 7-Tage-R-Wertes ist daher als problematisch zu erachten.

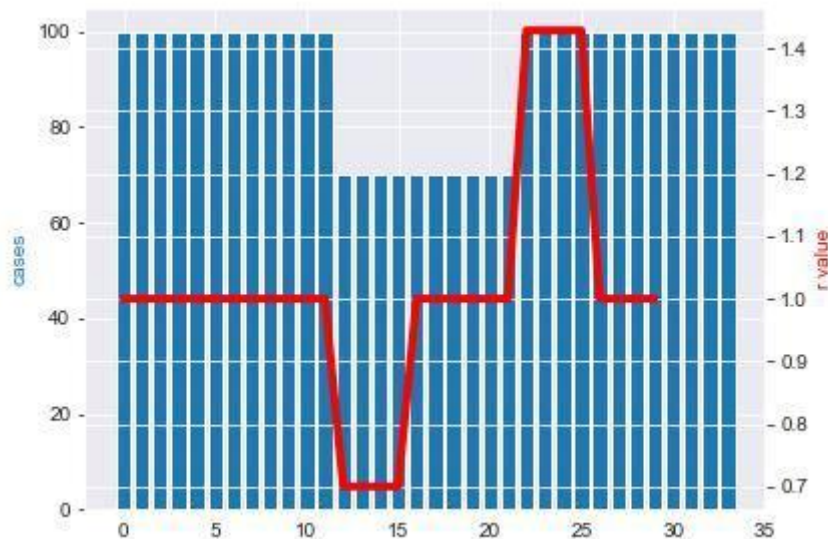
Zur korrekten Analyse der Wirkung von Maßnahmen wäre eine repräsentative Datenerhebung notwendig gewesen. Diese Daten gibt es jedoch nicht. Mit den vorliegenden Daten wäre es besser gewesen, den 7-Tage-Mittelwert des Nowcast zu nehmen und anhand diesem den R-Wert mit der ursprünglichen Berechnungsmethode (letzte 4 Tage im Verhältnis zu den 4 Tagen davor) zu berechnen.



Wirkung eines abrupten Abfallens der Fallzahlen auf den 4-Tage-R-Wert.

Dieser R-Wert hätte weder das Problem des Wochentageffekts aufgewiesen, noch das Problem der überlappenden Zeitreihen im Zähler und Nenner, weshalb auch der zu erwartende Wert von 0,7 erreicht wird. Die in Abschnitt 2.1.3 genannte zeitliche Verschiebung ist auch bei dieser Berechnungsmethode zu berücksichtigen, wobei die zeitliche Verschiebung entsprechend geringer ausfallen würde.

Eine weitere Möglichkeit wäre es, die Fallzahlen des aktuellen Tages direkt mit den Fallzahlen vor 4 Tagen ins Verhältnis zu setzen und so R_t zu berechnen.



Alternative Berechnung von R_t mittels der Berechnungsmethode: neue Fälle / neue Fälle vor vier Tagen

Aufgrund dessen, dass hierbei zwei Tage miteinander verglichen werden und nicht zwei Zeiträume, schlägt sich die abrupte Änderung in den Fallzahlen auch direkt im R_t nieder. Da hier keine Zeiträume sondern explizite Tage miteinander verglichen werden, ist diese

Methode nur geeignet, wenn die Fallzahlen sauber erhoben sind und damit keine unnatürliche Oszillation aufweist bzw. eine Glättung bspw. durch einen 7-Tage-Mittelwert durchgeführt wurde.

Der Nachteil an dieser Berechnungsmethode ist, dass auch die nachfolgenden drei Tage einen Wert von 0,7 aufweisen, obwohl es zu keiner weiteren Änderung in den Fallzahlen gekommen ist. Dabei zeigt sich das Problem der Berechnung einer Beschleunigung. Es handelt sich dabei um die Beschleunigung über den beobachteten Zeitraum hinweg, in diesem Fall über die Generationszeit von vier Tagen.

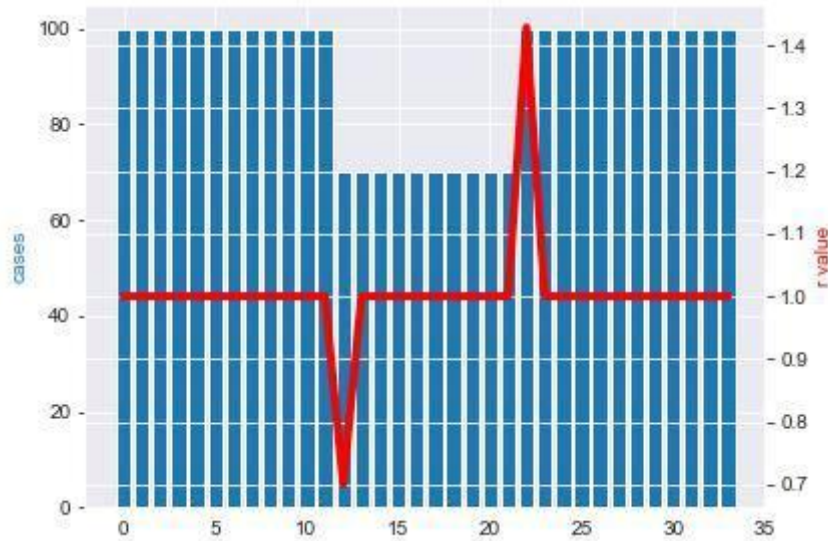
Um diesem Effekt gerecht zu werden, kann der R-Wert über den Zeitraum der Generationszeit zentriert werden.



Alternative Berechnung von R_t mittels der Berechnungsmethode: neue Fälle / neue Fälle vor vier Tagen mit Zentrierung über den beobachteten Zeitraum (Generationszeit von vier Tagen)

Die zuvor genannten Probleme der Berechnungsmethode einer Beschleunigung können für die korrekte Beurteilung eines möglichen Maßnahmeneffekts problematisch sein. Es wäre daher wünschenswert, den Zeitraum für die Berechnung der Beschleunigung möglichst kurz zu halten.

Hierbei kann dann von der Betrachtung eines R-Wertes über die Generationszeit abgewichen werden. Das kürzeste Zeitfenster wäre gegeben, wenn die Beschleunigung anhand direkt aufeinander folgender Tage gemessen werden würde. So kann den zuvor genannten Problemen entgegengewirkt werden.



Alternative Berechnung eines Beschleunigungsmaßes mittels der Berechnungsmethode: neue Fälle / neue Fälle einen Tag zuvor

Hierbei gilt umso mehr die Prämisse, dass dazu eine möglichst saubere Datengrundlage benötigt wird, ohne unnatürliche Oszillation bzw. einer Glättung eines solchen Effektes bspw. mittels eines 7-Tage-Mittelwertes.

2.1.5 Nicht repräsentativ erhobene Fallzahlen als Datengrundlage

Die Teststrategie und das Testverhalten haben Auswirkungen auf die Zahl der gemeldeten Fälle. Aufbauend auf den gemeldeten Fällen wurde der Nowcast als auch die 7-Tage-Inzidenz berechnet. In Bezug auf die 7-Tage-Inzidenz merkte bereits der Sachverständigenausschuss zur Bewertung der Corona-Maßnahmen diese Problematik an. Er hat darauf hingewiesen, dass die 7-Tage-Inzidenz ein Wert ist, der von der Teststrategie und damit auch vom Testverhalten abhängig ist:

“Dieser Wert ist insofern problematisch, als dass es sich um die 7-Tage-Melderate pro 100.000 Einwohner handelt, die abhängig von Teststrategien ist, nicht alle positiven Testergebnisse (etwa Antigen-Schnelltests) erfasst und die Dunkelziffer an Infektionen nicht berücksichtigt. [...] Eine valide Inzidenz, also die tatsächliche Erkrankungsrate, kann nur mit repräsentativen Zufallsstichproben ermittelt werden bzw. durch Modellrechnungen abgeschätzt werden” (Sachverständigenausschuss nach § 5 Absatz 9 Infektionsschutzgesetz - Bundesgesundheitsministerium 2022).

Auch das RKI merkte im Wochenbericht 5.5.2022 zur Inzidenz allgemeingültig an:

“Die Erfassung der durchgeführten Teste sowie die Ermittlung des Anteils der positiven Teste ermöglichen eine Einschätzung zur Wirksamkeit der Teststrategie. Bei der Interpretation der Daten müssen Veränderungen in der Teststrategie sowie im Testverhalten der Bevölkerung berücksichtigt werden.”

Die Aussagen des Sachverständigenausschusses und des RKI implizieren eine enorme Verzerrung der Datenlage der gemeldeten Fälle und damit auch der 7-Tage-Inzidenz und des Nowcast. Denn je mehr Menschen man testet, desto mehr Corona-Fälle findet und meldet man, auch wenn diese Abhängigkeit nicht zwangsläufig linear ist. Es wird damit aber nicht der *Anteil der positiven Bevölkerung* dargestellt.

Die Kette zur Bestimmung eines Corona-Testergebnisses bestand aus: Verpflichtung zum Test oder Krankheitsgefühl, Schnelltest, Arztgang, PCR-Test. Das Testverhalten ist personenbezogen, betrifft also die Frage, ob man bei einem Krankheitsgefühl einen Schnelltest macht und auch wirklich zum Arzt geht. Die Teststrategie ist dagegen politischer/angeordneter Natur und wirkt sich auf den Anteil der Tests aufgrund von Verpflichtungen aus. Beides: Die Teststrategie und das Testverhalten waren beeinflussbar und beides wirkte sich somit signifikant auf die gemeldeten Fälle aus. Es ist bekannt durch die unzähligen Verordnungen aller Bundesländer, dass die Strategien in der Testpolitik sich vielfach geändert haben.

Im Fazit bedeutet das:

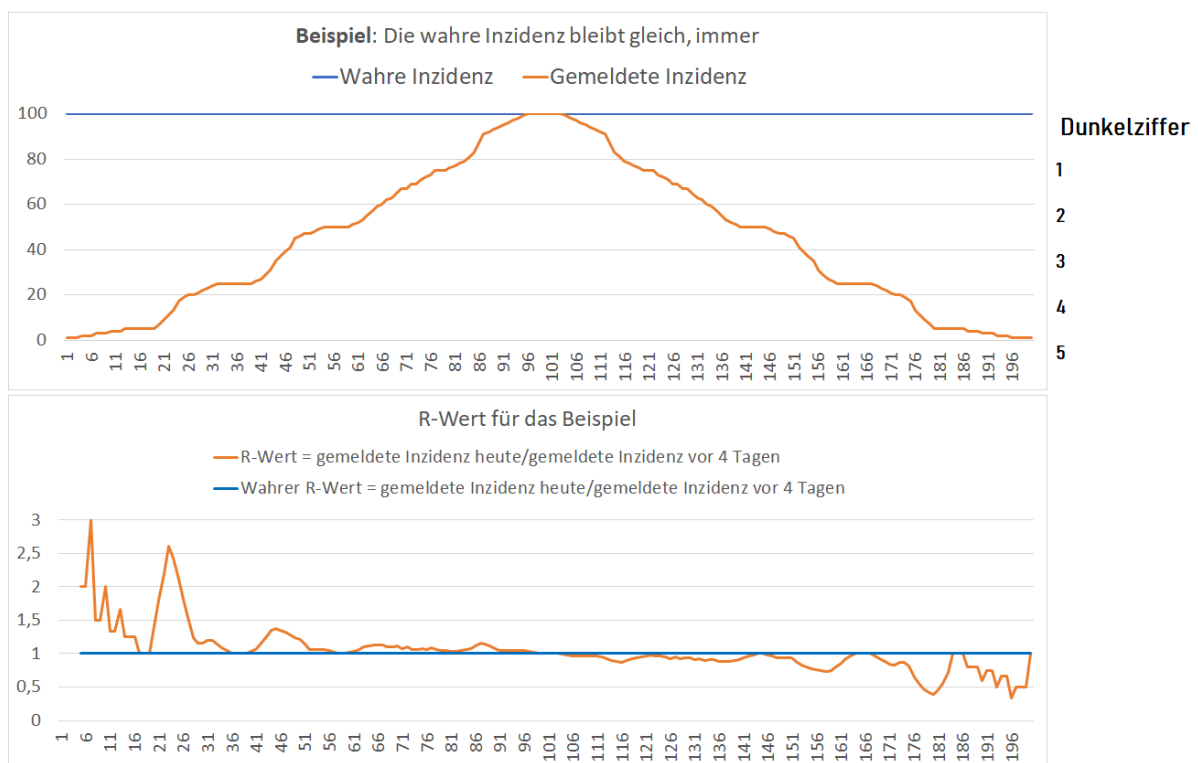
So, wie die 7-Tage-Inzidenz definiert wurde, entspricht sie einer 7-Tage-Fallzahl pro 100.000 Einwohner, aber sie stellt keine Inzidenz im medizinischen Sinne - also eine Fallrate pro 100.000 *getesteter* Einwohner - dar. Das ist elementar, um die Kritik zu begreifen, warum die Teststrategie und das Testverhalten so enorm wichtig für die gemeldeten Fälle und damit auch für die 7-Tage-Inzidenz und den Nowcast sind. Alternativ wäre eine dauerhaft erhobene repräsentative Zufallsstichprobe zielführend gewesen. Es gab, wie der Sachverständigenausschuss auch beschreibt, eine hohe Dunkelziffer an Fällen. Die Lage änderte sich dann eben nach dem starken Ausbau der Testkapazitäten zu Beginn der Pandemie, der flächendeckenden Einführung von Schnelltests und nochmals änderte sich die Lage mit der verpflichtenden Schnelltestung in vielen Bereichen durch Verpflichtungen zu Tests in: Medizinischen Einrichtungen, Heimen, Schulen, 3G auf Arbeit, Zugang zu ÖPNV, zu Gastronomie, zu Kinos usw. Das alles trieb die 7-Tage-Inzidenz in den jeweils durch Verpflichtungen betroffenen Altersgruppen enorm hoch, denn überall wurde nach dem Prinzip: Anzahl von gemeldeten Fällen (7-Tage-Inzidenz) aber nicht Anteil an Bevölkerung (korrekte Inzidenz) gemessen. Gleichzeitig gab es aber auch eine Komponente in der Teststrategie, die die gemeldeten Fälle sinken lassen konnte: Testpflichtbefreiungen gab es für z.B. frisch Geimpfte oder vollständig gegen Corona geimpfte Personen.

Die 7-Tage-Inzidenz wie auch der Nowcast beruhen auf den gemeldeten Fallzahlen, welche aufgrund der o.g. Probleme großen Unsicherheiten unterliegen. R_t beruht wiederum auf dem Nowcast, weshalb sich diese Unsicherheiten gerade zu Zeiten von Änderungen im Test- oder Meldeverhalten sich eben auch auf diesen niederschlagen. Eine pandemiebegleitende, repräsentative Datenerhebung wäre daher unerlässlich gewesen, um zuverlässige Aussagen über das Infektionsgeschehen treffen zu können.

Welchen Effekt ein sich veränderndes Meldegeschehen bei einer hohen Dunkelziffer auf den R_t -Wert (bzw. im Folgenden R-Wert) haben kann, sollen die beiden folgenden Beispiele verdeutlichen.

Als erstes wird ein Beispiel angenommen, in dem von einer Bevölkerung die gesamte Zeit über immer die gleiche Menge Menschen infiziert ist. Dann wären die "wahre" Fallzahl oder auch die "wahre" Inzidenz jeden Tag identisch. Der "wahre R-Wert" wäre für so ein Beispiel von Anfang bis Ende immer gleich groß, nämlich "1", da gestern so viele Fälle waren wie vorgestern und davor und davor. Niemals ändert sich die Anzahl der gemeldeten Fälle von einem Tag zum anderen.

Wenn man aber gar nicht alle Menschen testet (bzw. einen repräsentativen Teil der Bevölkerung), man also eine Dunkelziffer hat, und die Vorschriften zum Testen sich auch noch sprunghaft ändern, die Testpflichten erst also schrittweise aufgebaut oder wieder abgebaut werden, dann sieht man, dass der R-Wert schwankt und zwar je nach Veränderung der Testvorschriften - dabei müsste er eigentlich immer gleich groß bleiben und nämlich den Wert "1" (keine Änderung) haben, also eine gleich bleibende Anzahl von Fällen aufzeigen. Genau das passiert aber nicht, wenn man Teststrategien ändert. Es ändern sich dann immer die Bedingungen und der R-Wert wird unmöglich korrekt zu interpretieren sein.

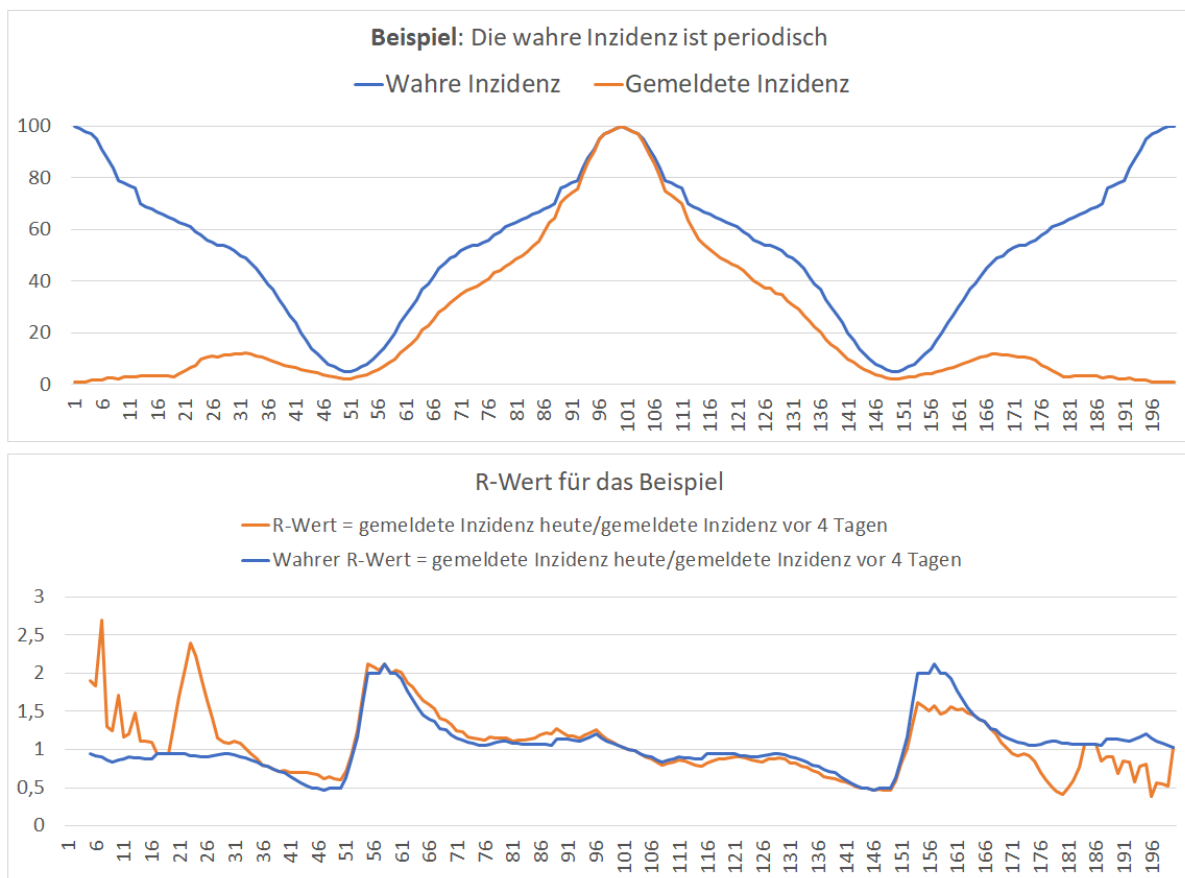


Beispiel 1 zur Auswirkung der Teststrategie auf die gemeldete Inzidenz. In einer Bevölkerung, bei der jeden Tag die gleiche Menge Menschen infiziert ist, wäre die Inzidenz in "Wahrheit" immer gleich und der R-Wert immer gleich 1. Durch eine sich verändernde Teststrategie, bei der bis Tag 100 die Testpflichten Schritt für Schritt auf bis zu 100% der Bevölkerung angehoben und danach symmetrisch bis Tag 200 wieder abgesenkt werden, sieht man Scheinänderungen im R-Wert.

Leider fehlt in der Studie des RKI eine Abschätzung darüber, wie sich die ungenau bestimmte Anzahl von Infektionen auf das Endergebnis des R_t -wertes auswirkt.

Im zweiten Beispiel verhält sich die Inzidenz periodisch, z.B. weil es ein saisonal auftretendes Virus gibt. Ebenso wie im ersten Beispiel erhöht man 100 Tage lang

schrittweise die Testpflicht bis man 100% der Bevölkerung testet und senkt danach die Testpflicht wieder so ab, wie man sie aufgebaut hatte.



Beispiel 2 zur Auswirkung der Teststrategie auf die gemeldete Inzidenz. Die Inzidenz schwankt periodisch. Die Teststrategie sieht vor, dass steigend bis zum Tag 100 immer mehr Menschen getestet werden, bis 100% der Bevölkerung getestet wird. Dann entspricht die "gemeldete Inzidenz" der "wahren Inzidenz" und der "gemeldete R-Wert" dem "wahren R-Wert". Danach wird die Testpflicht schrittweise spiegelverkehrt abgebaut, wie sie aufgebaut wurde. Die Dunkelziffer der Inzidenz sinkt bis zum 100. Tag und steigt danach wieder an. Es wird deutlich, dass in Zeiten hoher Dunkelziffern, also etwa zu Beginn bis Tag 34 und dann ab Tag 170 der "gemeldete R-Wert" ein steigendes bzw. fallendes Infektionsgeschehen andeutet, obwohl das in Wahrheit nicht der Fall ist. Man hat schlicht und einfach einerseits zu wenig getestet und andererseits zu oft die Maßnahmen verändert, um das wahre Infektionsgeschehen zu sehen.

Die "wahre Inzidenz" zeigt nun ein periodisch fallendes und steigendes Muster an, während die "gemeldete Inzidenz" das zwar auch tut, aber sie zeigt nicht immer in die gleiche Richtung wie die "wahre Inzidenz". Der Grund liegt darin, dass man ja die Testkapazität erst richtig aufbauen muss, um auf die wahre Inzidenzkurve zu kommen. Im R-Wert spiegelt sich das auch wider. Es gibt z.B. im Bereich 1 bis 34 Tage eine Zeit, in der der "gemeldete R-Wert" positiv ist, während der "wahre R-Wert" negativ ist. Hier hätte die Bevölkerung gesagt bekommen, dass die Infektionen steigen, obwohl sie in Wahrheit gefallen sind. Man hat halt nicht alle Menschen bzw. einen repräsentativen Teil der Bevölkerung getestet, um den wahren Verlauf der Infektionen zu sehen. Den umgekehrten Effekt sieht man dann ab etwa 170 Tagen, wo die "wahre Inzidenz" steigt, während die aufgrund zurückgenommener Testvorschriften die "gemeldete Inzidenz" sinkt.

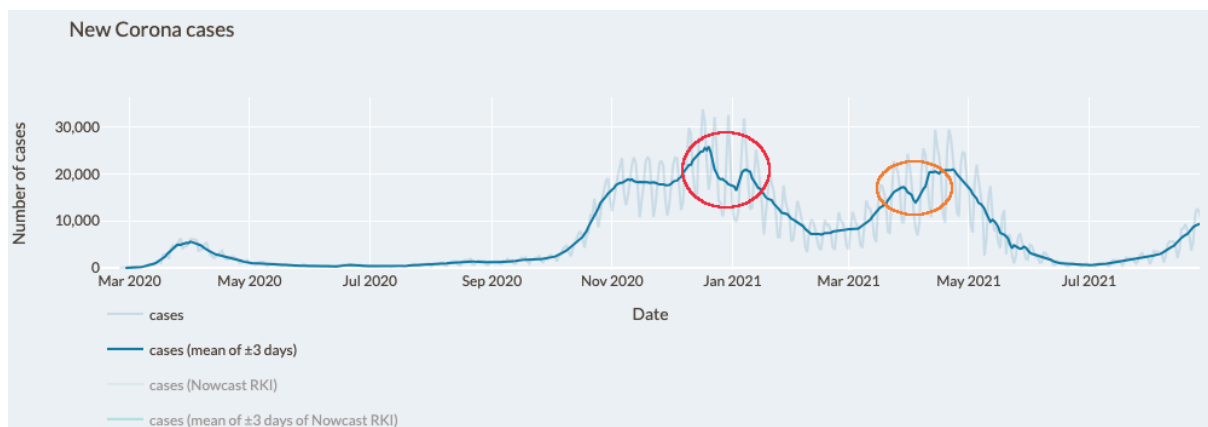
Ein besonderes Augenmerk bekommen bei R-Werten also die Zeiten, in denen die Teststrategien geändert wurden bzw. Testkapazitäten erhöht wurden. Wann immer die Vorschriften hin zu mehr Testpflicht verändert wurden, hat sich das direkt auf den gemeldeten R-Wert ausgewirkt.

Man sieht anhand der beiden Beispiele, dass die Änderung der Teststrategie sich auf die Dunkelziffer der gemeldeten Fallzahlen auswirkt, was sich dann auf die Inzidenz und auf den gemeldeten R-Wert auswirkt - und das führt zu paradoxen Effekten, wenn man auf Basis eines so bestimmten R-Wertes Corona-Maßnahmen anordnet.

Die Hauptkritik hier ist also, dass das RKI hinsichtlich dieser Problematik keine Unsicherheitsanalyse durchführt, die angibt, wie stark die Änderung der Teststrategie und Dunkelziffer sich auf den R-Wert auswirken, bzw. es keine Korrektur solcher Effekte vornimmt.

2.1.6 Effekt der Ferien auf die Meldung von Fallzahlen

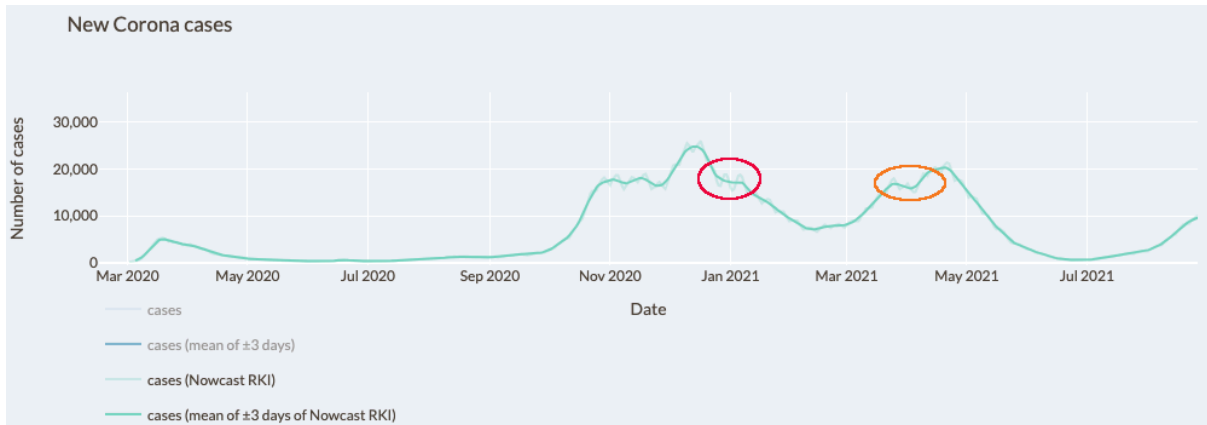
Ferien hatten einen Effekt auf die *gemeldeten* Fallzahlen, wie der Abbildung entnommen werden kann.



Effekt der Ferien auf die gemeldeten Fälle (rot: Weihnachtsferien, orange: Osterferien).

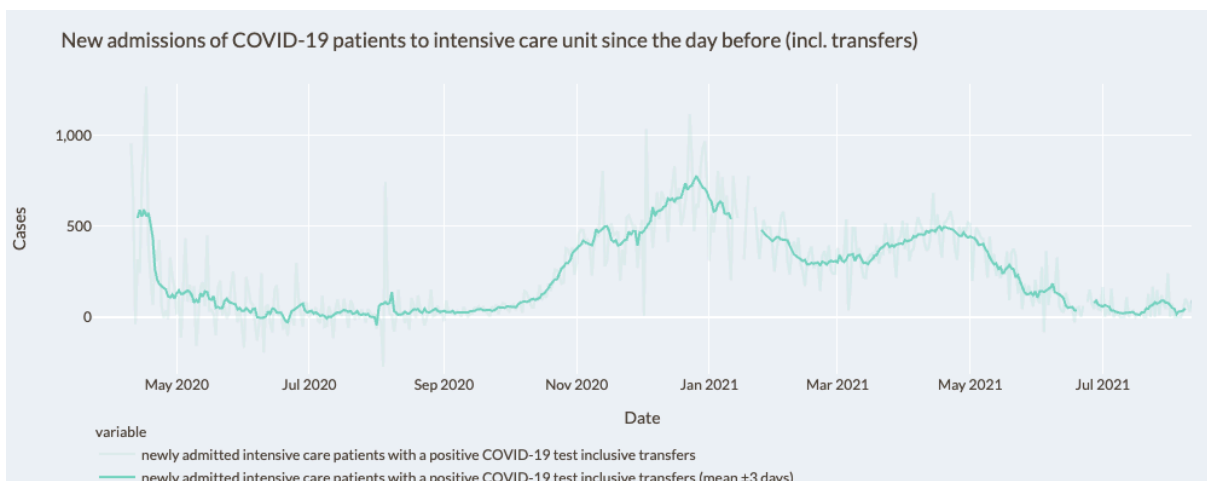
Dies lag u.a. daran, dass einige Meldungen durch die Gesundheitsämter erst nach den Ferien an das RKI übermittelt wurden, weshalb während der Ferien eine zu geringe und nach den Ferien eine zu hohe Meldung erfolgte.

Durch den Nowcast des RKI, wobei der Erkrankungsbeginn approximiert wird, kann dieser Effekt geglättet werden.



Effekt der Ferien auf den Nowcast (rot: Weihnachtsferien, orange: Osterferien).

Man kann beobachten, dass durch die Approximation der Effekt verringert werden konnte. Nun kann man anhand der Daten des Nowcast zu der Schlussfolgerung kommen, dass die Ferien einen bremsenden und nach den Ferien einen beschleunigenden Effekt hatten. Hierbei muss sich jedoch vor Augen geführt werden, dass der wichtigste Parameter für die Approximation des Erkrankungsbeginns das Datum des Eingangs der Fallmeldung am RKI ist ([RKI 2020](#)), welches in vielen Fällen verfälschenderweise nach den Ferien lag. Das führt zu einer Approximation, die die Werte für die Ferien weiterhin zu niedrig und nach den Ferien zu hoch annimmt, auch wenn der Effekt verringert werden konnte. Eine weitere Problematik ist, dass während der Ferien auch weniger Fälle festgestellt werden konnten, sei es durch nicht mehr durchgeführte Schnelltests in Schulen und damit einem höheren Dunkelfeld bei Schülern, als auch ausbleibende Testungen und Arztbesuche während des Urlaubs.



Tägliche Neuaufnahmen auf Intensivstationen (inkl. Verlegungen).

Schaut man sich zum Vergleich die täglichen Neuaufnahmen auf Intensivstationen an, stellt man fest, dass dieser zuvor noch im Nowcast zu sehende Ferieneffekt nun nicht mehr sichtbar ist. Das ist zumindest ein Indiz dafür, dass der im Nowcast zu sehende Effekt nicht auf ein sich änderndes Infektionsgeschehen während der Ferien beruhen könnte, sondern vielmehr auf das Test- und Meldeverhalten während der Ferien und das eigentliche Infektionsgeschehen davon abwich.

Da R_t anhand des Nowcast berechnet wurde und dies die Grundlage für die Analyse in der Studie war, führt dies zu einer fehlerhaften Analyse in Bezug auf die Ferienzeit. Die Daten suggerieren aufgrund der zuvor genannten Punkte zwangsläufig einen Abfall während und einen Anstieg nach den Ferien, auch wenn das Infektionsgeschehen ein anderes gewesen sein sollte, worauf die Daten von den Intensivstationen hindeuten. Mit einer repräsentativen Datenerhebung hätte man dies feststellen können.

2.1.7 Fazit zur Wahl von R_t als abhängige Variable

Grundsätzlich ist es nachvollziehbar, dass R_t als abhängige, also zu erklärende Variable aufgrund der Eigenschaft als Beschleunigungsmaß gewählt wurde. Als äußerst problematisch anzusehen ist dabei allerdings, dass teilweise von falschen Annahmen ausgegangen wurde, bspw., dass das Testverhalten fast keinen Einfluss auf R_t habe.

Generell wurden Probleme in der Datengrundlage ignoriert, eine Korrektur, um diesen Effekten entgegenzuwirken, fand nicht statt. So wurde bspw. der Effekt des Test- und Meldeverhaltens während der Ferien nicht gewürdigt, welches sich durch die Berechnungsmethode auch auf den Nowcast auswirkt und damit auch auf R_t . Wenn der berechnete R_t während der Ferien auf das Test- und Meldeverhalten zurückzuführen ist, aber das eigentliche Infektionsgeschehen von dem berechneten R_t abweicht, kommt es durch das Modell zwangsläufig zu Fehlinterpretationen, dass zu Beginn der Ferien ein Abfallen und nach den Ferien ein Anstieg im R_t zu verzeichnen ist. Eben auch, wenn es diesen Abfall und anschließenden Anstieg im realen Infektionsgeschehen nicht geben würde.

Ebenfalls wurde nicht berücksichtigt, dass R_t aufgrund der Berechnungsmethode ein nachlaufender Wert ist. Eine notwendige Zuordnung von R_t auf den korrekten Tag des vollständigen Wirkens erfolgte nicht. Auch erfolgte keine Berücksichtigung der Inkubationszeit im Hauptmodell. Beides wäre für die korrekte Einschätzung, zu welchem Zeitpunkt eine Änderung in der Ausbreitungsgeschwindigkeit bzgl. Infektionen zu verzeichnen war, erforderlich gewesen.

2.2 Modellansatz des RKI: die unabhängigen Variablen

Die rechte Seite des Modells (siehe nachfolgende Abbildung) beinhaltet die unabhängigen, also erklärenden Variablen. Sie enthält zwei Terme (untere Zeile) für pharmazeutische und nicht-pharmazeutische Eingriffe, die die von einem saisonal oszillierenden Grundmodell (Sinus- und Cosinus-Terme) mit konstant angenommenen Abweichungen für regionale Varianzen in Amplitude, aber nicht in Frequenz, einem festen Effekt für die Bundesländer¹⁸, θ_b , und für die im Beobachtungszeitraum auftretenden Varianten Alpha und Delta (die Terme v_α und v_δ) auftretenden Abweichungen abfangen sollen. Im RKI-Modell ist der Verlauf von R_t über die Zeit demnach abhängig von 1. Saisonalität, 2. der Virusvariante und 3. den Corona-Maßnahmen sowie 4. der Impfung.

¹⁸ In derartigen Modellen mit Paneldaten (Mehrere Untersuchungsobjekte über die Zeit) werden häufig feste Effekte (engl. *fixed effects*) für die Untersuchungsobjekte (hier: Bundesländer) ins Modell integriert, um die unbeobachtete Heterogenität zwischen ihnen zu berücksichtigen. Dies ist aber *nicht* zu verwechseln mit den in der RKI-Studie etablierten "fixierten Effekten" für die Varianten (siehe unter Kap. 2.2.2).

Positiv zu vermerken ist, dass - im Vergleich zu vielen früheren Corona-NPI-Studien - ein vergleichsweise langer Beobachtungszeitraum zu Grunde gelegt wurde (März 2020 bis August 2021) und somit Phasen vor, während und nach den Lockdowns berücksichtigt werden¹⁹. Ansonsten deckt sich das Modelldesign grundsätzlich mit dem vieler anderer NPI-Studien (z.B. Sharma et al. 2021), hat aber, abgesehen vom fehlenden Kausalansatz (siehe unter Kap. 1.2), auch dieselben Probleme.

$$\begin{aligned}
 \mathbb{E}(\log(R_t)) = & \underbrace{0.3v_\alpha + 0.6v_\delta}_{\text{Definierter ("fixierter") Effekt der Alpha- und Delta-Variante --> Erhöhter R-Wert (30 bzw. 60 % höher)}} + \underbrace{\theta_b}_{\text{Feste Effekte für jedes Bundesland --> Niveau-Unterschiede zwischen Bundesländern}} + \underbrace{\beta_1 \sin\left(2\pi \frac{t}{365}\right) + \beta_2 \cos\left(2\pi \frac{t}{365}\right)}_{\text{Saisonale Schwankungen des R-Wertes}} \\
 & + \underbrace{\beta_3 f(\text{Impf}_{t+t})}_{\text{Effekte der Impfung auf R-Wert}} + \underbrace{\sum_{k \in K} \gamma_k m_{t+\tau}^k}_{\text{Effekte von K Maßnahmen auf R-Wert}}
 \end{aligned}$$

Darstellung des Hauptmodells, farbliche Abgrenzung nach Termen, S. 27 (RKI 2023a), verändert.²⁰

Man bemerke zunächst, dass im Modell kein autoregressiver Term R_{t-1} existiert, auch wenn das eine häufige Formulierung in diesen Umständen ist. Der beste Prädiktor von R_t ist normalerweise der Vorgängerwert R_{t-1} oder eine lineare Kombination von Vorgängerwerten. Der Vorteil eines Modells mit autoregressivem Term wäre, dass nur noch der Anteil der abhängigen Variable R_t modelliert wird, der nicht durch ihre Eigendynamik erklärt wird. Man spricht hier von Autoregressionmodellen, die in diesem Anwendungsbereich best-practice-Verfahren sind. Es wäre an dieser Stelle mindestens notwendig gewesen, diese Omission zu begründen.

2.2.1 Modellierung des saisonalen Verlaufs

Anders als viele frühere NPI-Studien (z.B. Sharma et al. 2021) berücksichtigt das RKI-Modell einen saisonalen Effekt. Dieser saisonale Effekt fließt in das Modell über zwei periodische Terme ein, die auf Sinus/Cosinus-Transformationen der Zeit beruhen (S. 26/27). Eine solche Berücksichtigung von Saisonalität wurde bereits in anderen Studien zu Corona-Maßnahmen angewendet, allerdings nicht bezogen auf den R-Wert, sondern auf Infektionskurven (z.B. Berlemann & Haustein 2023). Hier besteht ein entscheidender Unterschied: Infektionskurven vieler Atemwegserkrankungen verlaufen erkennbar periodisch

¹⁹ Einige NPI-Studien hatten einen völlig inadäquaten Beobachtungszeitraum, der eine Aussage zur Wirkung von Corona-Maßnahmen schon alleine aus diesem Grund unmöglich macht; so endete in einigen Studien der Beobachtungszeitraum beispielsweise mitten in der Corona-Winterwelle 2020/21 (z.B. Levelu & Sandkamp 2022, Sharma et al. 2021), wodurch per Definition die (möglichen) Auswirkungen von Lockerungen der Lockdown-Maßnahmen ausgeblendet werden bzw. überhaupt kein Kontrolldesign gegeben ist.

²⁰ Das RKI gibt an, dass die Fixierung des Wertes bspw. zur Delta-Variante in Höhe von 0,6 zu einer Erhöhung von 60% führt. In Wirklichkeit ist die Erhöhung noch höher. Vgl. hierzu Berechnung unter 2.2.2, wonach das Setzen des Wertes auf 0,6 zu einer Erhöhung von über 82% führt.

bzw. saisonal, und zwar, bei SARS-CoV-2, mit einem Peak im (frühen) Winter, hier dargestellt anhand einer RKI-Grafik mit gemeldeten Fällen nach Meldedatum bzw. Erkrankungsbeginn:

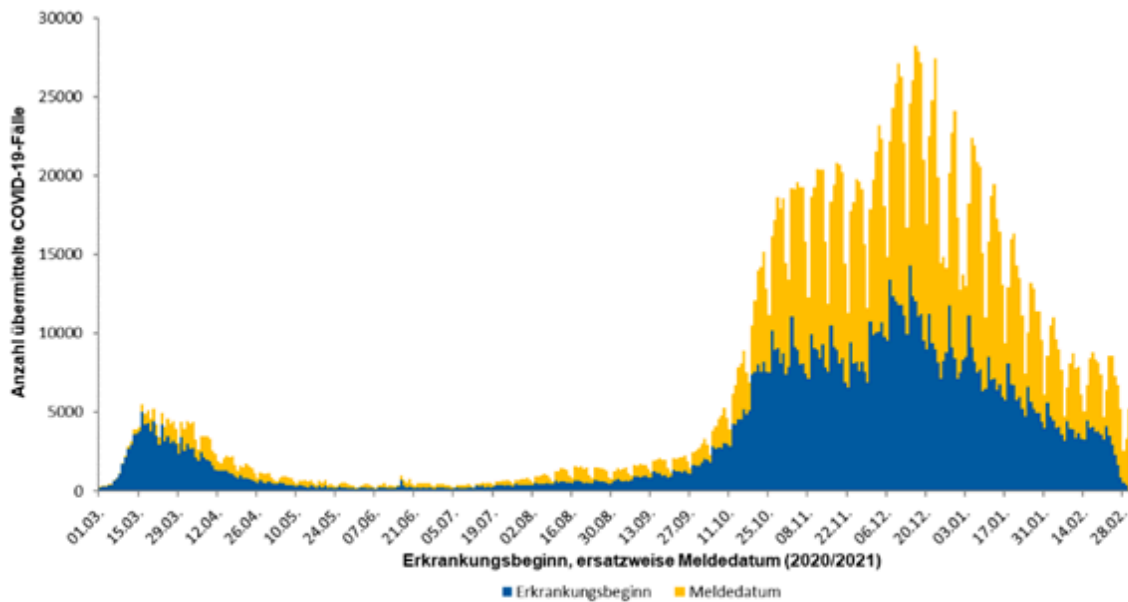
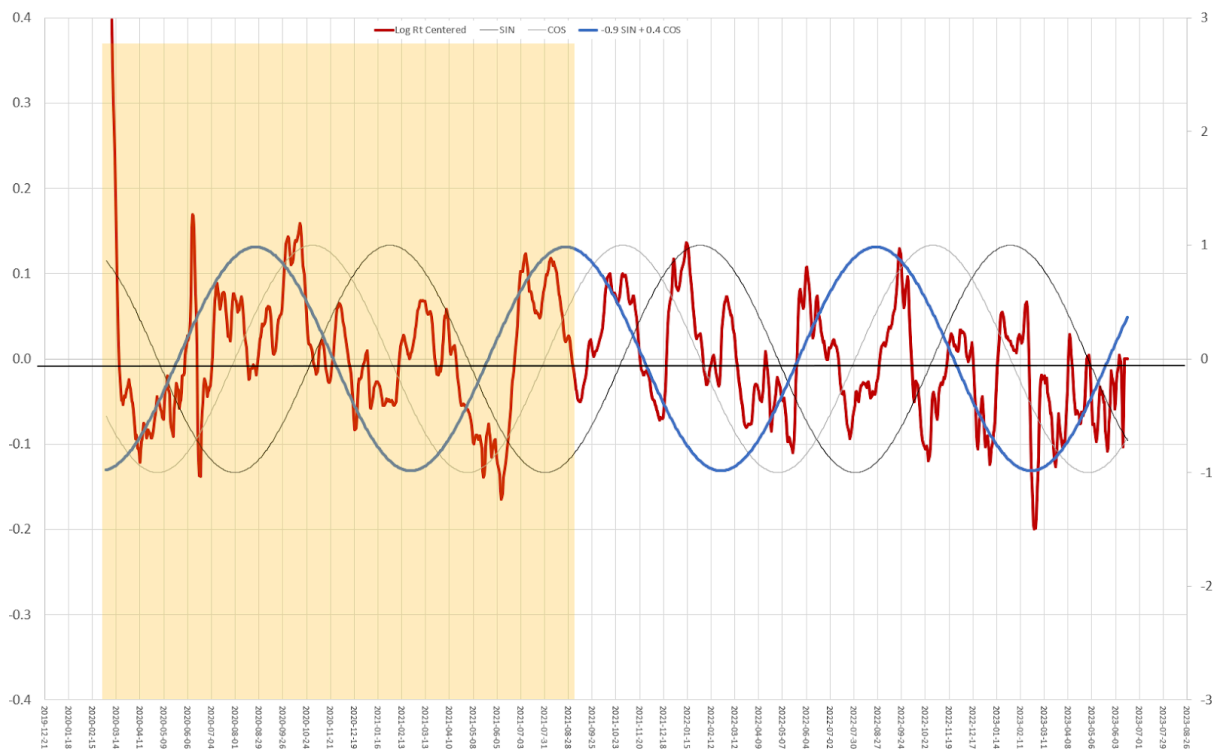


Abbildung 2: Anzahl der an das RKI übermittelten COVID-19-Fälle nach Erkrankungsbeginn, ersatzweise nach Meldedatum. Dargestellt werden nur Fälle mit Erkrankungsbeginn oder Meldedatum seit dem 01.03.2020 (03.03.2021, 0:00 Uhr).

Anzahl der übermittelten Fälle nach Erkrankungsbeginn (blau) bzw nach Meldedatum (gelb). Quelle: RKI (2021a), S. 4.

Es stellt sich nun die Frage, ob eine Sinus/Cosinus-Oszillation auch tatsächlich den “natürlichen” saisonalen Verlauf des R-Wertes darstellt. Im Modell der RKI-Studie hängt die Ausbreitungsdynamik fast vollständig von den Oszillationstermen Sinus und Cosinus ab. Angenommen, der Startpunkt dieser Oszillation ist der 1. Januar 2020²¹, ergibt sich ein Grundmodell ohne Interventionen aus einer Kombination der Sinus- und Cosinuskurven. (Beobachtungszeitraum in gelb, andere Faktoren sind noch nicht berücksichtigt).

²¹ Es wird nicht spezifiziert, ob der 1. Januar oder der 1. März 2020 als Startdatum gewählt wurde. Wir verwenden hier 1. Januar. Alle Aussagen treffen auf beide Möglichkeiten zu.



Auf Amplitude 1 normalisierte Sinus- (schwarz) und Cosinuskurven (grau) sowie eine beispielhafte Linearkombination mit Koeffizienten $-0.9 \sin + 0.4 \cos$ (blau), im Vergleich zum zentrierten $\text{Log } R_t$ (rot). Der gelbe Hintergrund stellt den Beobachtungszeitraum dar. Quelle: eigene Darstellung, Datengrundlage: [RKI Nowcast](#).

Es ist sowohl an dieser Darstellung als auch am Vergleich mit den Infektionskurven erkennbar, dass die Sinus/Cosinus-Transformation der Zeit keine hinreichende Abbildungskraft für den zeitlichen Verlauf des R -Wertes - so wie ihn das RKI immer berechnet hat - hat. Das zeigt sich schon daran, dass in diesem Grundmodell Amplitude und Frequenz gleich bleiben (eine Veränderung der Amplitude wird erst durch die Berücksichtigung der neuen Virusvarianten hergestellt, siehe Kap. 2.2.2) - somit lässt sich beispielsweise das Maximum des R -Wertes zu Beginn der ersten Welle (März 2020) nicht darstellen. Danach ist der R -Wert nie wieder annähernd auf diese Höhe gestiegen, was einer sinusoidalen Oszillation widerspricht. Es ist zwar korrekt, dass R_t oszilliert, jedoch ist nicht geklärt, ob diese Oszillation saisonal ist, zufälligen Schwankungen unterliegt und/oder von systematischen Verzerrungen - zum Beispiel Testeffekten (siehe Kap. 2.1) - beeinflusst ist²².

2.2.2 Annahmen zum Effekt neuer Varianten

Um den Unterschieden der Infektiosität der verschiedenen SARS-CoV-2-Varianten Rechnung zu tragen, enthält die Modellgleichung zwei Terme, die für höhere R_t -Werte bei den nach dem "Wildtyp" (erste Welle) einsetzenden Varianten sorgen. Die zugehörigen Koeffizienten für v_α und v_δ sind, entgegen allen wissenschaftlichen Gepflogenheiten, bereits

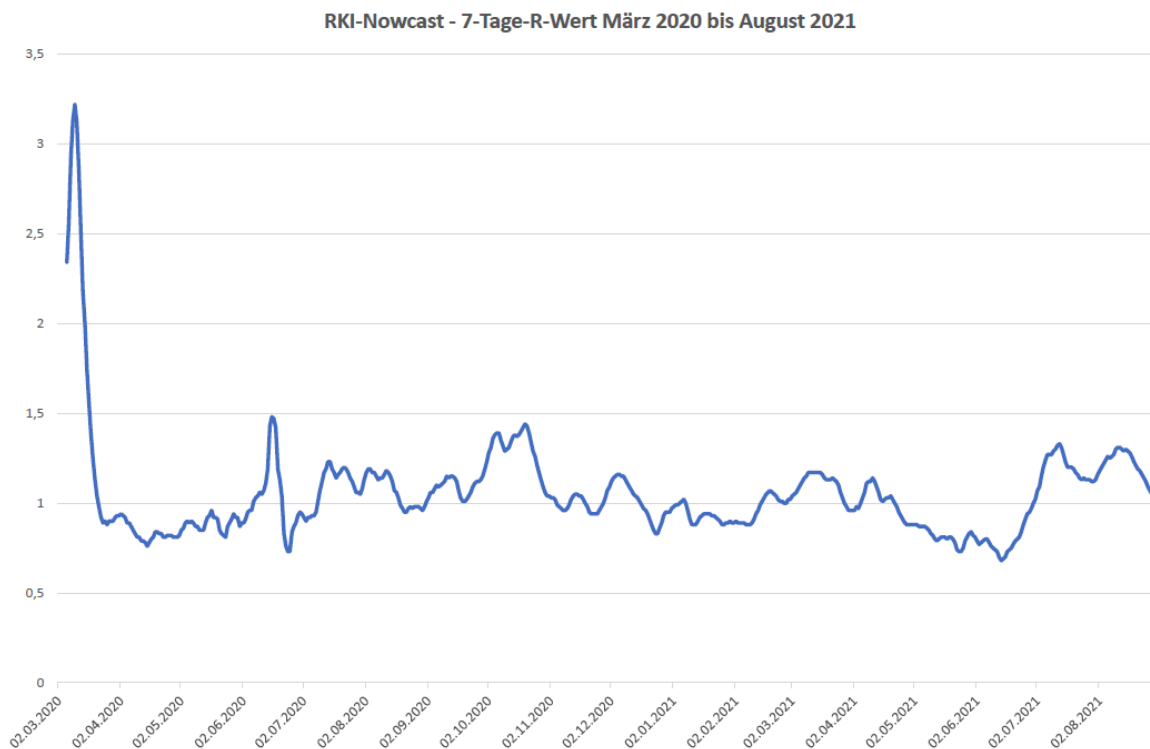
²² Hier kommt noch ein räumlicher Aspekt hinzu: Bundesländer sind bereits recht grobe Aggregate. Streng genommen sind Wellen aber kleinräumig-regional (Landkreise, Gemeinden, Wohnviertel etc.) und werden über die räumliche Aggregation erst "künstlich" zu einer Welle zusammengefasst. Wenn nun in einem großen Bundesland - wie etwa Baden-Württemberg oder Niedersachsen - sukzessive Ausbrüche in den einzelnen Landkreisen auftreten, sorgt auch dies für Oszillationen im R -Wert für das gesamte Bundesland.

mit $\beta_3 = 0,3$ und $\beta_4 = 0,6$ fixiert (S. 27). Diese Parameter bedeuten, dass für die Alpha-Variante ein 30 % höherer und für die Delta-Variante ein 60 % höherer R-Wert angenommen wird.

Hierbei sind *erstens* bereits die Prozentwerte in Frage zu stellen, da diese weder anhand theoretischer Überlegungen noch auf empirischer Basis irgendwie begründet werden; woraus diese Koeffizienten entstehen, bleibt völlig offen. Es gibt keinen Literaturverweis hierzu²³. Es wird also *angenommen*, dass die R-Werte bei der Alpha- und Delta-Variante wesentlich höher *wären*.

Zweitens geht aus der Modelldarstellung in der RKI-Studie nicht hervor, wie die Variantenvariablen v_α und v_δ überhaupt operationalisiert wurden. Es könnte sich hierbei um zeitabhängige Anteile dieser Varianten (z.B. in %) handeln, aber auch um Dummy-Variablen, die jeweils die dominante Variante anzeigen. Eine höhere Aussagekraft hätten Anteile, und da das RKI diese Anteile selbst veröffentlicht, gehen wir an dieser Stelle davon aus, dass eben diese Anteile verwendet wurden.

Drittens fällt bei der Betrachtung der R_t -Zeitreihen des RKI auf, dass der berechnete R_t -Wert zu keinem Zeitpunkt der Pandemie jemals wieder den Höchstwert von Anfang März 2020 (erste Welle) erreichte, und zwar weder vor noch nach noch während der Lockdowns.



²³ Möglicherweise war hier der Hintergrund, dass es Studien gibt, die für spätere SARS-CoV-2-Varianten einen höheren R_0 -Wert (also die Basisreproduktionszahl, *nicht* die effektive Reproduktionszahl!) ermittelt oder modelliert haben. Auf diese Studien wird aber kein Bezug genommen bzw. sie werden im Text bei der Herleitung des Modells nicht zitiert. Abgesehen davon ist die Schätzung der Basisreproduktionszahl sehr fehleranfällig und diese wird oftmals zunächst überschätzt (siehe z.B. Mercer et al. 2011, Prada et al. 2022).

Quelle: eigene Darstellung. Datengrundlage: RKI (2023).

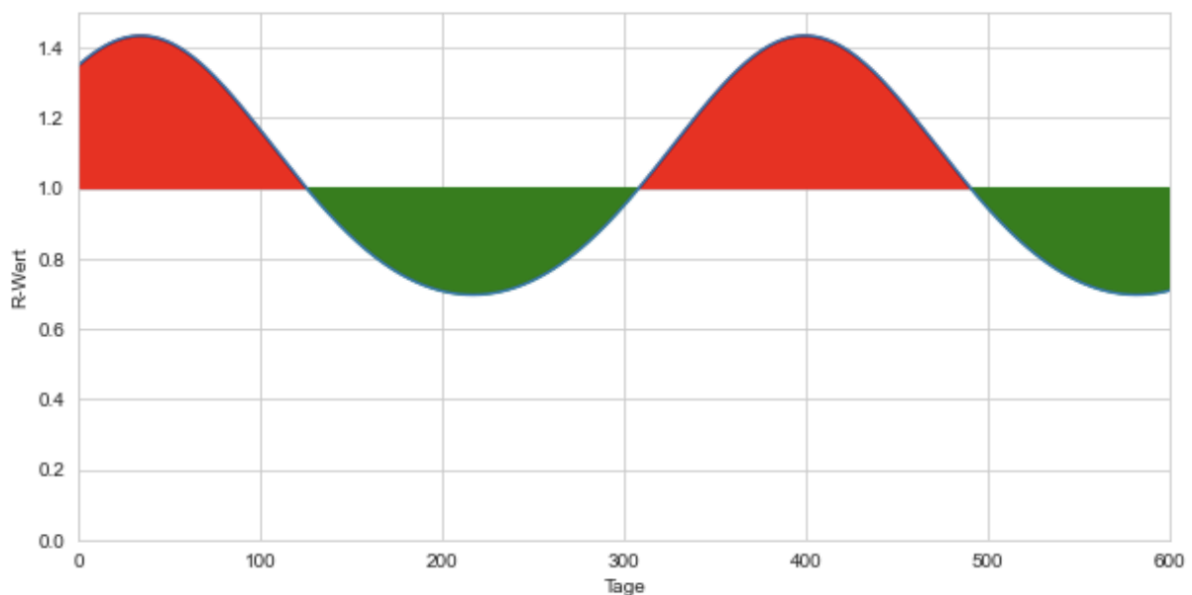
Viel fragwürdiger ist generell, *viertens*, warum für die Abbildung einer höheren Übertragungswahrscheinlichkeit im Modell überhaupt Parameter *definiert* (*“fixiert”*) werden, anstatt sie empirisch zu ermitteln. Üblicherweise besteht ein solches Modell auf der rechten Seite der Gleichung aus unabhängigen (erklärenden) Variablen, deren Parameter empirisch geschätzt werden. So wäre es möglich, diesen (möglichen) Effekt zu kontrollieren *und* dabei empirisch zu quantifizieren - sofern er denn besteht. Dieses Vorgehen wäre dringend erforderlich, unabhängig davon, ob Anteilswerte oder Dummy-Variablen verwendet wurden. Vermutlich wurden in der RKI-Studie Anteile von Varianten ins Modell aufgenommen, was, für sich genommen, nicht zu beanstanden wäre - wenn deren Einfluss über das Modell empirisch ermittelt und nicht *“gesetzt”* werden würde. Durch diese Modellformulierung werden Theorie (a priori-Annahme bezüglich eines höheren R-Wertes) und Empirie (stochastische Schätzung der sonstigen Modellkoeffizienten) in inakzeptabler Weise vermischt.

Egal wie das Auftreten neuer Varianten im Modell operationalisiert wird: In allen Fällen würde man die Parameter für die (angenommene) Erhöhung von R_t durch neue Varianten empirisch ermitteln und nicht einfach *setzen*; d.h. man sollte die *“fixierten”* Parameter eigentlich durch stochastische ersetzen. *Wenn* die später aufgetretenen Virusvarianten, unter ansonsten konstant gehaltenen Bedingungen, tatsächlich für einen wesentlich höheren R_t -Wert gesorgt haben sollten, müsste dies im Modell auch empirisch nachweisbar sein. Dieser unmotivierte Schritt alleine sollte zur Abwertung des Modells führen. Es ist Aufgabe des Modells, nicht des Modellierers, diese Koeffizienten zu finden und statistisch zu qualifizieren.

Diese (hohen) Werte haben im Modell außerdem einen entscheidenden Effekt und beeinflussen die empirischen Ergebnisse der anderen Modellterme: Die a priori-*Definition* höherer R_t -Werte für spätere Varianten bedeutet nämlich, dass die „eigentlich zu erwartenden“ R-Werte in der zweiten und dritten Welle wesentlich höher gewesen *sein müssten* als in der ersten. Da sie das aber nachweislich nicht waren (s.o.), *müssen* – gemäß Modell (!) – zwingend andere Umstände dafür gesorgt haben, diese R_t -Werte zu drosseln. Da aber abgesehen von den Saisonalitätstermen und den Maßnahmenvariablen im Modell keine weiteren erklärenden Variablen auftauchen, ist implizit bereits definiert, dass Maßnahmen zur Senkung des R-Wertes beigetragen haben *müssen*. Die Definition der höheren R-Werte für spätere Varianten steuert also das Modellergebnis mit. Um die Größe dieses problematischen Effektes zu quantifizieren, müssten mehrere Modellvarianten ohne die Variantenterme bzw. mit modifizierten Prozentzahlen geschätzt werden.

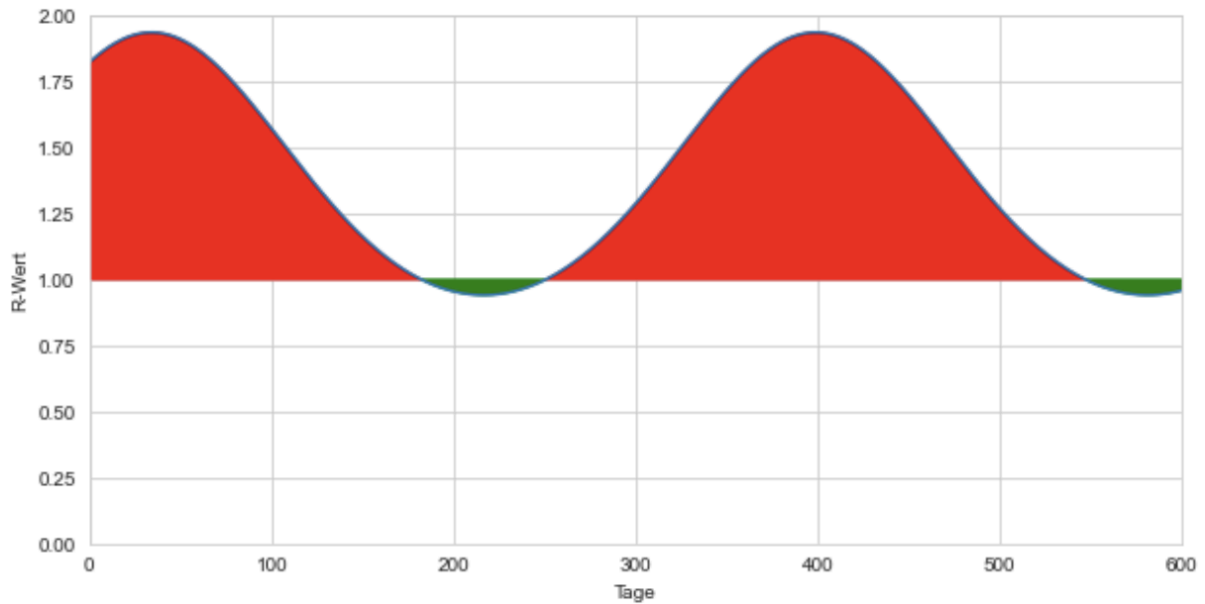
Die Wirkung auf R_t durch das Setzen der Werte für die Parameter der Virusvarianten lässt sich auch direkt berechnen. Mittels der Berechnung $e^{0,3} = 1,35$ stellt man fest, dass das Setzen des Wertes auf 0,3 für den Parameter für die Alpha-Variante im Modell zu einer pauschalen Erhöhung des angenommenen Verlaufs des R-Wertes um 35% führt, im Gegensatz zu der Behauptung in der Studie, dass dies zu einer Erhöhung von R_t von 30% führt. Noch deutlicher ist die Abweichung für die Delta-Variante. Die Berechnung $e^{0,6} = 1,82$ zeigt, dass damit der angenommene Verlauf des R-Wertes für die Delta-Variante pauschal um 82% angehoben wird und nicht, wie in der Studie dargestellt, um 60%!

Dies hat Auswirkungen auf den angenommenen Verlauf der R-Werte für die verschiedenen Virusvarianten. Laut Gleichungssystem des Modells in der Studie wird die Saisonalität durch eine Kombination eines faktorisierten Sinus- und eines faktorisierten Kosinusterm dargestellt (vgl. Abschnitt 2.2). Die Addition dieser Terme (und weiterer Terme) ergeben den natürlichen Logarithmus des R-Wertes. Nehmen wir beispielhaft an, dass der Wert für den Sinusterm bei 0,2 liegt und für den Kosinusterm bei 0,3. Mittels der Berechnung $e^{0,2*\sin(x) + 0,3*\cos(x)}$ erhalten wir den beispielhaften Verlauf des Modells für den Wildtyp ohne Betrachtung weiterer Terme. Dabei ergibt sich diese Grafik, wobei zu sehen ist, dass sich die Phasen, in denen die Fallzahlen steigen (roter Bereich, R-Wert > 1) und in denen sie fallen (grüner Bereich, R-Wert < 1), abwechseln und eine ähnliche Fläche besitzen.



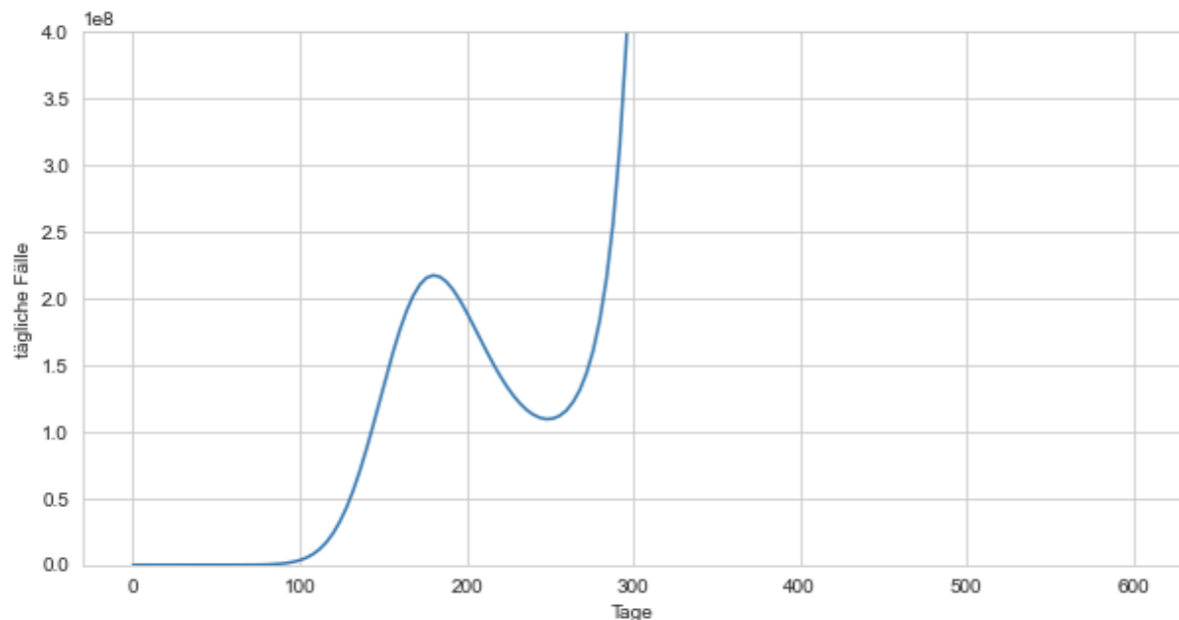
Die Werte wurden in dem Beispiel so gewählt, da hiermit ein Kurvenverlauf entsteht, bei dem sich der R-Wert im Bereich 0,7 bis 1,43 bewegt und damit in etwa in dem Bereich, in dem sich der R-Wert während der Phase des Wildtyps bewegte (abgesehen von der Anfangsphase, in der aber auch Testkapazitäten aufgebaut wurden). Ob Maßnahmen einen höheren R-Wert als 1,43 verhindert haben, spielt für die nachfolgende Betrachtung des Problems der Festlegung der Werte für die Virusvarianten keine Rolle. Vielmehr würden die nachfolgend aufgezeigten Probleme noch verstärkt werden. Entscheidender für die nachfolgende Betrachtung ist ohnehin der untere Wert. Auch die korrekte Einordnung auf der Zeitachse ist für die Betrachtung der Probleme erst einmal nicht bedeutsam.

Um nun den Effekt des Setzens des Wertes auf 0,3 für die Alpha-Variante äquivalent zur vorherigen Betrachtung darzustellen, muss die Rechnung auf $e^{0,2*\sin(x) + 0,3*\cos(x) + 0,3}$ angepasst werden.



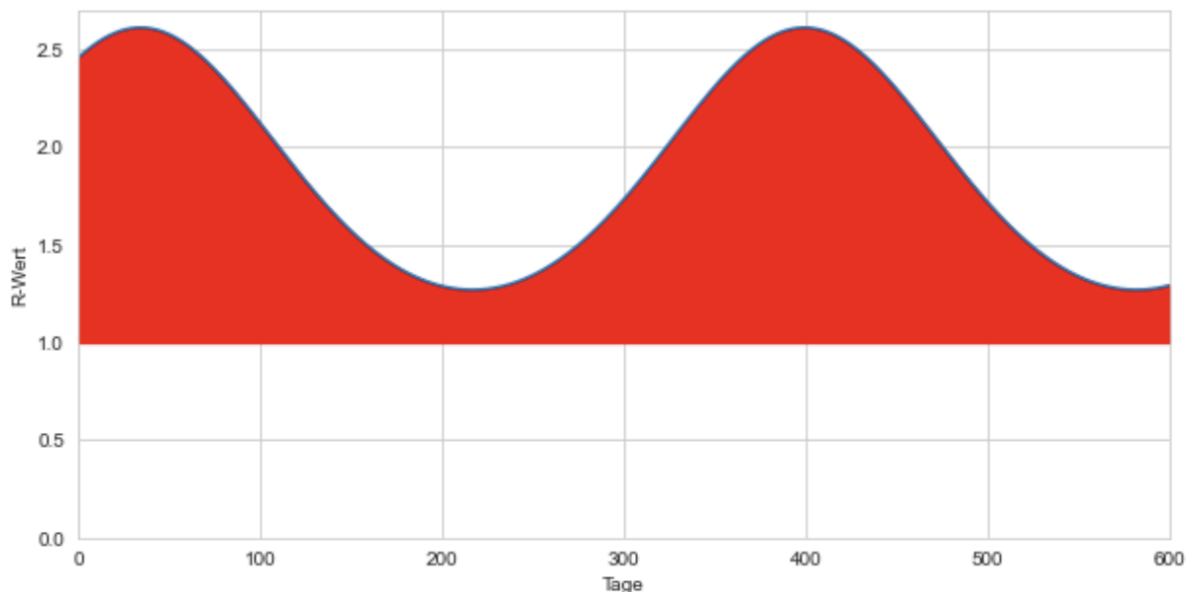
Daraus ergibt sich, dass nun nicht nur der Hochpunkt des Verlaufs für den R-Wert auf 1,94 steigen, sondern auch der Tiefpunkt auf 0,94 angehoben werden würde. Das führt dazu, dass es nur noch eine sehr kurze Phase gibt, in der Fallzahlen überhaupt noch fallen könnten und dann auch nur in einem sehr geringen Maße. Die Fläche für den Anstieg der Fallzahlen ist deutlich größer als die Fläche für das Sinken.

Anhand dieses Verlaufs für den R-Wert lässt sich eine beispielhafte Berechnung für einen sich daraus ergebenden Verlauf der täglichen Fälle durchführen.



Hierbei wird nochmal deutlich, dass es nur eine kurze Phase des Sinkens der Fallzahlen geben würde, die Fallzahlen dann von einem hohen Plateau aus wieder beginnen zu steigen. Es würde also einen Anstieg bis in die Unendlichkeit geben, denn im Modell ist nicht berücksichtigt, wieviele Menschen sich bereits infiziert haben und wieviele sich zum Zeitpunkt t überhaupt infizieren könnten.

Durch Anpassung der Rechnung auf $e^{0,2*\sin(x) + 0,3*\cos(x) + 0,6}$ können wir nun die Wirkung betrachten, die damit durch das Modell für die Delta-Phase angenommen wird.



Hierbei ergibt sich die zuvor schon dargelegte pauschale Erhöhung des R-Wertes um 82% (entgegen der Beschreibung in der Studie, wonach sich der R-Wert um 60% erhöhen soll). Dadurch steigt der Hochpunkt der Verlaufskurve für den R-Wert auf 2,61 und der Tiefpunkt auf 1,27. Ein Sinken der Fallzahlen wäre demnach nicht mal mehr kurzfristig möglich.

Daraus lässt sich erahnen, dass damit das Modell einen Kurvenverlauf ohne Maßnahmen annimmt, welcher so wohl in Realität nicht beobachtet worden wäre, da er zu hoch angenommen wird. Diese Diskrepanz muss das Modell nun durch ihr zur Verfügung stehende Parameter erklären. Da hierfür nur Maßnahmen und Impfungen im Modell vorgesehen sind, führt das automatisch dazu, dass den Maßnahmen und Impfungen eine positive Wirkung bescheinigt wird, auch wenn diese nur auf der Fehlspezifikation des Modells mit der Folge der zu hohen Annahme eines Verlaufs ohne Maßnahmen zurückzuführen wäre (vgl. hierzu Abschnitt 2.2.4).

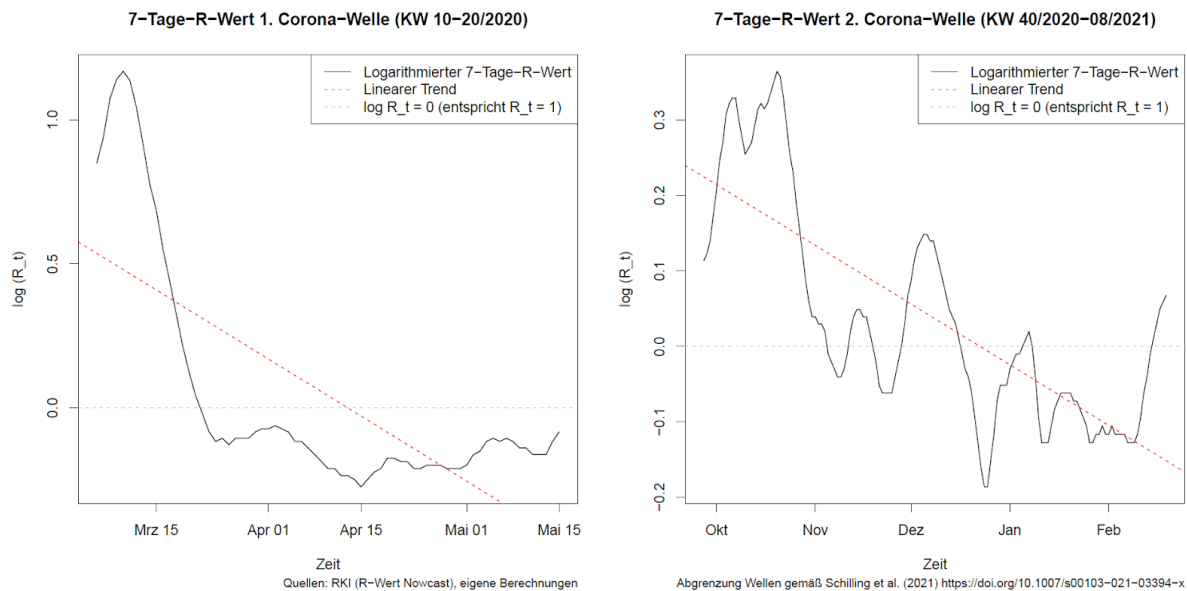
2.2.3 (Nicht-)Berücksichtigung von beschränktem Wachstum

Der R-Wert ist, vereinfacht gesagt, ein Maß für die Ausbreitungsbeschleunigung einer Infektionskrankheit, genau genommen die Beschleunigung über die Geschwindigkeit der Ausbreitung (siehe unter Kap. 2.1). R_t sinkt auch unabhängig von jedweden Interventionen oder der Jahreszeit *während* einer Infektionswelle; die epidemiologische Literatur differenziert deswegen im Hinblick auf die Absenkung von R_t über die Zeit ausdrücklich zwischen *intrinsischen* Faktoren (Abnahme der suszeptiblen Bevölkerung durch fortschreitende Ausbreitung²⁴) und *extrinsischen* Faktoren (individuelle Schutzmaßnahmen

²⁴ Je mehr Individuen sich innerhalb einer Welle mit diesem Erreger infizieren, desto geringer ist die verbleibende Zahl von suszeptiblen Individuen, also jenen, die (noch) nicht infiziert wurden. Bereits deshalb verlangsamt sich die Ausbreitung im Verlauf einer Infektionswelle.

sowie, falls sie etabliert werden und wirksam sind, nicht-pharmazeutische Interventionen, Impfungen etc.) (siehe z.B. Mercer et al. 2011, Nishiura & Chowell 2009).

Die folgende Abbildung zeigt den zeitlichen Trend des (logarithmierten) 7-Tages-R-Wertes aus dem RKI-Nowcast für die ersten beiden Corona-Wellen (die Definition des Zeitraums beider Wellen entspricht hierbei der des RKI):



Verlauf des 7-Tages-R-Wertes in Deutschland in der 1. und 2. Corona-Welle. Quelle: eigene Darstellung. Datengrundlage: RKI (2023).

Weiterhin infizieren sich, egal ob staatliche Maßnahmen ergriffen werden oder nicht, nie alle Individuen der Population in derselben Welle; die Welle erreicht ihren Wendepunkt lange bevor sich annähernd alle Einwohner der betroffenen Region angesteckt hätten. Dies ist auch auf Bevölkerungsheterogenität zurückzuführen, d.h. nicht jedes Individuum hat, alle anderen Bedingungen konstant gelassen, dieselbe Wahrscheinlichkeit, überhaupt mit einem Erreger in Kontakt zu kommen (hierzu siehe z.B. Chowell et al. 2015, Thomas et al. 2020; zur besonderen Bedeutung von "super-spreaders" siehe z.B. Stein 2011). Bereits das reguläre Sozialverhalten von Menschen - völlig unabhängig davon, ob eine Gefahrensituation besteht oder vermutet wird - bremst also schon die Ausbreitung, weil ein bestimmter (großer) Anteil der Population ohnehin nicht infiziert wird.

Es handelt sich daher um *gesättigtes* bzw. *beschränktes Wachstum*, das gut mit einer Sigmoid-Funktion (z.B. logistische Funktion, Gompertz-Funktion) approximiert werden kann, und nicht um exponentielles Wachstum. Diese Beschreibung betrifft Infektionskrankheiten im Allgemeinen und stellt eine epidemiologische Grundlage dar, die auch in die gängigen epidemiologischen Modelle implementiert ist (siehe z.B. Li 2018, Nishiura & Chowell 2009).

Exponentielles Wachstum würde bedeuten, dass der Wachstumsfaktor R_t über eins *und* konstant (!) ist bzw. dass eine konstante, positive Wachstumsrate vorliegt²⁵.

Dass auch (Corona-)Infektionswellen sich ohne Interventionen bzw. unabhängig vom Eskalationsniveau dieser Interventionen abschwächen, ist viele Male empirisch nachgewiesen worden (siehe unter Kap. 3.1). Auch im Evaluationsbericht des Sachverständigenausschusses zu den Corona-Maßnahmen wird darauf eingegangen (Sachverständigenausschuss nach § 5 Absatz 9 Infektionsschutzgesetz - Bundesgesundheitsministerium 2022, Kap. 6.1.1.1).

Der Terminus eines exponentiellen Wachstums der Fallzahlen findet sich an mehreren Stellen in der RKI-Studie, obwohl die Ausbreitung einer Infektionskrankheit, wie bereits dargelegt, nicht exponentiell verläuft. Im RKI-Regressionsmodell gibt es ein stetiges Sinken und Steigen des R-Wertes, eine sinusoidale Oszillation; diese führt, unter ansonsten gleichen Bedingungen, dazu, dass der R-Wert immer wieder auf ein Minimum sinkt und dann wieder auf das vorherige Maximum steigt - Amplitude und Frequenz bleiben gleich (siehe unter Kap. 2.2.1). Durch die Berücksichtigung definierter Varianteneffekte wird sogar davon ausgegangen, dass der R-Wert über die Zeit steigt (siehe Kap. 2.2.2). Die selbstlimitierende Eigenschaft - also dass R_t auch innerhalb einer saisonalen Welle sinkt (durch Abnahme suszeptibler Bevölkerung und Bevölkerungsheterogenität, s.o.) - ist im RKI-Modellansatz nicht implementiert. Vereinfacht ausgedrückt: Von den saisonalen Schwankungen abgesehen *kann* der R-Wert im zu Grunde gelegten Modell nicht sinken - außer durch Corona-Maßnahmen und die Corona-Impfung.

2.2.4 Replikation des Modells

Das Grundprinzip des Modells aus der RKI-Studie lässt sich veranschaulichen, indem man Teile des Modells repliziert, und zwar die Terme der Modellgleichung, die sich auf den Verlauf des R-Wertes ohne den Einfluss von Maßnahmen und Impfung beziehen. Wir haben hierfür im ersten Schritt die Saisonalitätsterme (Sinus/Cosinus; siehe Kap. 2.2.1) und im zweiten Schritt die *vordefinierten* Variantenanteile (siehe Kap. 2.2.2) eingeführt. Diese Terme sind in der folgenden Darstellung rot markiert.

²⁵ Der Hintergrund der häufig anzutreffenden, aber unrichtigen Formulierung des "exponentiellen Wachstums" in der Corona-Pandemie ist vermutlich, dass in der Epidemiologie Exponentialfunktionen zur *Approximation* der *frühen* Phase einer epidemischen Welle (z.B. die ersten Wochen) verwendet werden; dass der Gesamtwachstumsprozess *nicht* exponentiell verläuft, wird dabei aber stets herausgestellt (hierzu siehe z.B. Chowell et al. 2015). Manchmal wird die Frühphase eines solchen Ausbruchs auch als "exponentielle Phase" bezeichnet, was in mathematischer Hinsicht irreführend ist, da auch bereits das Wachstum in der Frühphase einer Infektionswelle nicht "exponentiell" ist, sondern sich nur passabel dadurch *darstellen* lässt.

Unterschiede zwischen
Bundesländern

Erhöhter R-Wert
durch neue
Varianten

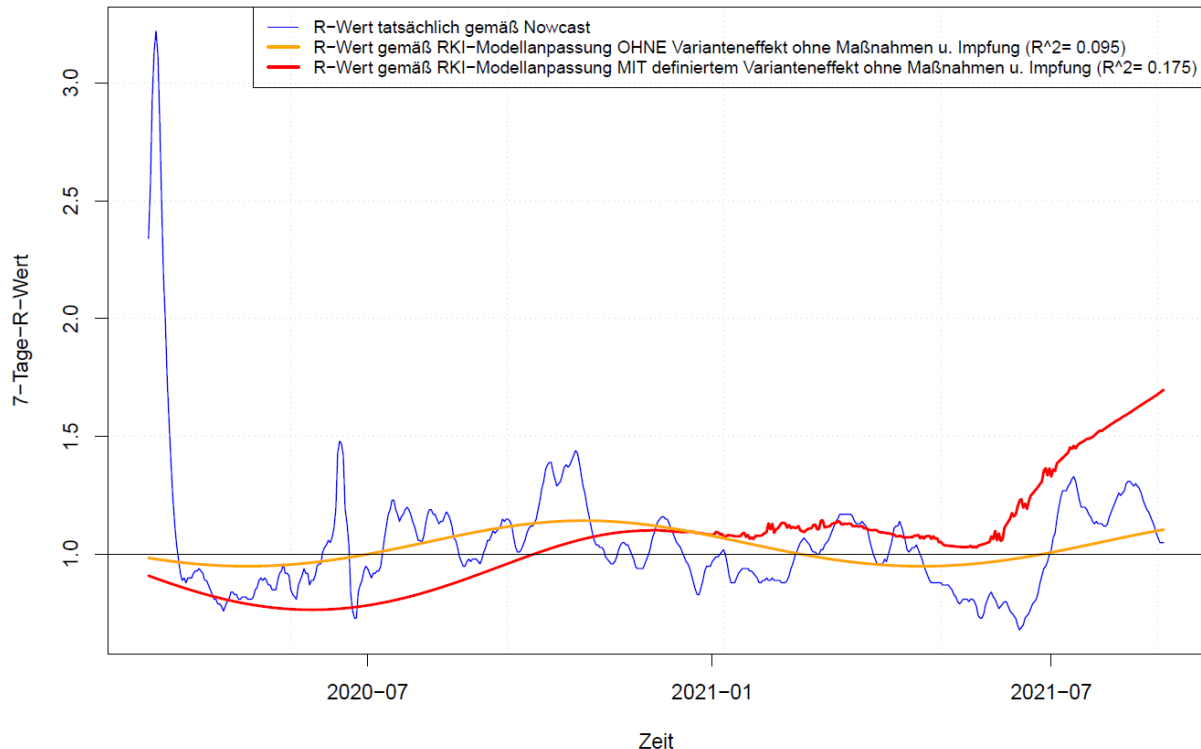
Saisonale Schwankungen R-Wert

$$\mathbb{E}(\log(R_t)) = 0.3v_\alpha + 0.6v_\delta + \theta_b + \beta_1 \sin\left(2\pi \frac{t}{365}\right) + \beta_2 \cos\left(2\pi \frac{t}{365}\right) + \beta_3 f(\text{Impf}_{t+l}) + \sum_{k \in K} \gamma_k m_{t+\tau}^k$$

Effekte von Maßnahmen und
Impfung auf R-Wert

Als abhängige Variable fungiert hierbei der 7-Tage-R-Wert auf nationaler Ebene (Deutschland); daher fallen die festen Effekte für die Bundesländer (θ_b) weg, da keine regionale Differenzierung erfolgt. Entsprechend des RKI-Modells werden die Varianteneffekte (Koeffizienten der Terme v_α und v_δ) mit 0,3 bzw. 0,6 vordefiniert²⁶. Die Schätzung der Parameter β_1 und β_2 erfolgt über das Regressionsmodell.

²⁶ Es ist aus der RKI-Studie nicht ersichtlich, woher die Variantenanteile im Modell (Terme v_α und v_δ) stammen. Die diesbezüglichen Daten für die hier aufgezeigte Teilreplikation des Modells stammen vom European Centre for Disease Prevention and Control (ECDC); diese lagen als wöchentliche Anteile der Alpha- und Delta-Variante vor, die hier mit einem zentrierten 7-Tage-Mittelwert geglättet wurden.



Quellen: European Centre for Disease Prevention and Control (Anteile Varianten), RKI (R-Wert Nowcast), eigene Berechnungen

Replikation des RKI-Modells ohne Maßnahmen und Impfung (gelb = nur Oszillation, rot = Oszillation mit Varianteneffekten). Quelle: Eigene Darstellung.

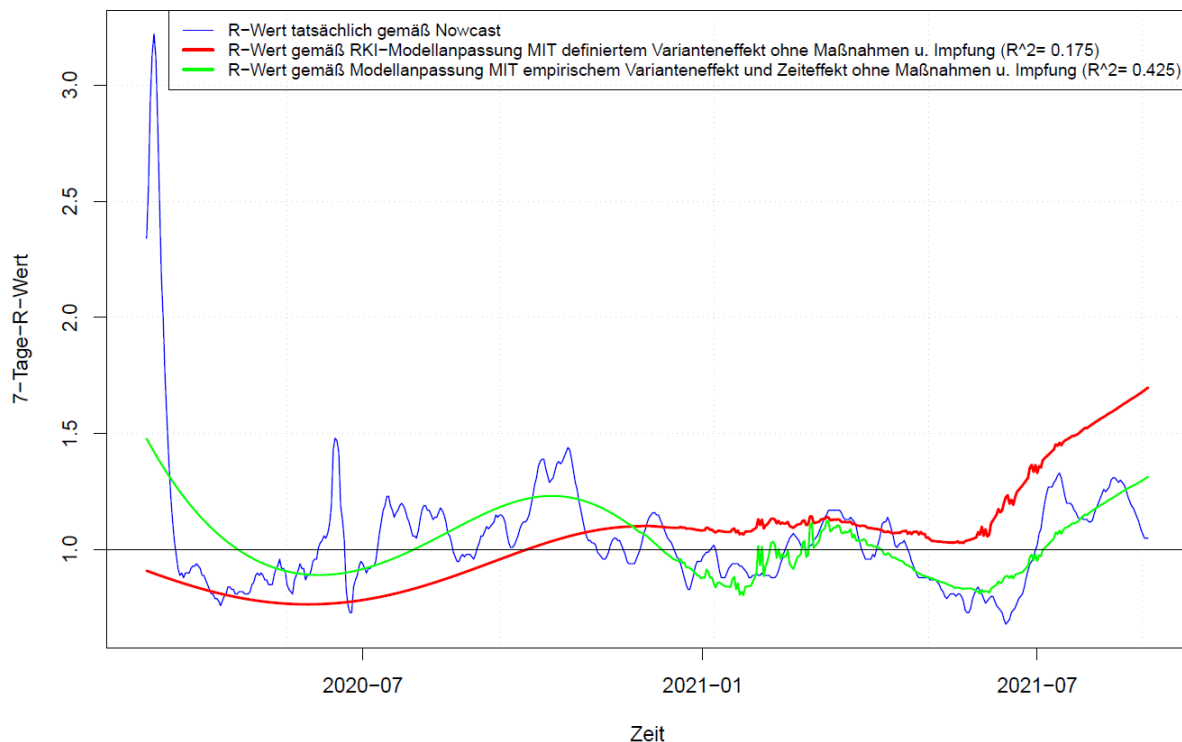
Ein Modell, das ausschließlich die sinusoidale Oszillation des R-Wertes berücksichtigt (gelbe Linie), erreicht eine Varianzaufklärung von ca. 9,5 % ($R^2 = 0,095$). Diese wird zwar durch die Hinzunahme der Varianteneffekte insgesamt etwas auf 17,5 % verbessert ($R^2 = 0,175$), jedoch zeigt sich bereits durch eine bloße Inaugenscheinnahme eine drastische Abweichung des im Modell (ohne Einflüsse von Maßnahmen und Impfung) *angenommenen* R-Wertes vom realen R-Wert. Ab Januar 2021 ist der *modellierte* R-Wert (rote Linie) durchgängig über 1 und fast immer deutlich oberhalb des *echten* R-Wertes (blaue Linie). Dieses Modell ist natürlich noch nicht vollständig spezifiziert: im RKI-Modell werden nun noch weitere Variablen aufgenommen, nämlich die Corona-Maßnahmen und die Corona-Impfung. Diese Variablen erhöhen dann die Varianzaufklärung, sinngemäß und allgemeinverständlich ausgedrückt: Sie "drücken" die rote Linie wieder näher an die blaue Linie (also den Echtwert) heran. Andere Gründe für den Unterschied zwischen dem *angenommenen* (roten) und dem *tatsächlichen* Verlauf des R-Wertes (blaue Linie) sieht das Modell nicht vor, da - außer Maßnahmen und Impfung - keine weiteren erklärenden Variablen in das Modell aufgenommen wurden.

Es wird also an diesem Beispiel plastisch erläutert, dass im RKI-Modell eine Senkung des R-Wertes *ausschließlich* durch Containment-Maßnahmen und Impfung vorgesehen ist. Da *nur* diese Variablen im Modell vorgesehen sind, ist gewissermaßen „vorprogrammiert“, dass im gegebenen Modelldesign Maßnahmen und Impfung eine Wirkung im Sinne einer Senkung des R-Wertes haben *müssen*.

Im folgenden Beispiel wurde eine leicht modifizierte Modellspezifikation und -schätzung durchgeführt: Die Koeffizienten β_3 und β_4 der Terme v_α und v_δ (also der Einfluss der Anteile

neuer Virusvarianten) wurden *nicht* vordefiniert, sondern, wie dies üblicherweise in Regressionsmodellen getan wird, aus den Daten geschätzt. Außerdem wurde ein Zeitterm (t) als unabhängige Variable hinzugefügt, um dem Absinken des R-Wertes innerhalb einer Welle (siehe unter Kap. 2.2.3) Rechnung zu tragen (in diesem Fall ein simpler Tageszähler):

$$E(\log(R_t)) = \beta_1 \sin\left(2\pi \frac{t}{365}\right) + \beta_2 \cos\left(2\pi \frac{t}{365}\right) + \beta_3 v_\alpha + \beta_4 v_\delta + \beta_5 t$$



Quellen: European Centre for Disease Prevention and Control (Anteile Varianten), RKI (R-Wert Nowcast), eigene Berechnungen

Replikation des RKI-Modells ohne Maßnahmen und Impfung (rot = Oszillation mit Varianteneffekten) und alternative Spezifikation mit geschätzten Parametern für den Varianteneffekt und Zeitterm (grün). Quelle: Eigene Darstellung.

Es ist bereits durch Inaugenscheinnahme zu erkennen, dass die alternative Modellspezifikation (grüne Linie) eine bessere Abbildung des realen R-Wertes (blaue Linie) darstellt als die RKI-Modellspezifikation (rote Linie). Die Varianzaufklärung ist mit 42,5 % ($R^2 = 0,425$) wesentlich höher als im Ausgangsmodell; auch das Maximum von R_t zu Beginn der ersten Corona-Welle wird besser (wenn auch nicht annähernd zufriedenstellend) abgebildet²⁷. Diese alternative Spezifikation könnte also ein erster Schritt zur Verbesserung des RKI-Modells sein, wobei klargestellt sein muss, dass alleine die Berücksichtigung der Zeit als erklärende Variable eine starke Vereinfachung darstellt, da hierbei die kausalen Mechanismen hinter dieser Senkung nicht implementiert werden. Sinnvoll wäre z.B. die Berücksichtigung bereits stattgefundenener Infektionen, um dem Umstand einer - kurzfristigen - Immunisierung durch Infektion²⁸ Rechnung zu tragen.

²⁷ Beide Modelle können den kurzfristig sehr hohen R-Wert im Juni 2020 nicht adäquat abbilden. Hierbei handelt es sich um ein einzelnes entdecktes Ausbruchsgeschehen in einer Fleischfabrik in Rheda-Wiedenbrück, die zu dieser Zeit in Medien intensiv diskutiert wurde.

²⁸ Wie lange eine durchgemachte Infektion vor einer erneuten Ansteckung bzw. einer Weitergabe des Erregers schützt, ist umstritten und natürlich auch von den jeweiligen Varianten abhängig. Wir können und wollen hierzu

2.2.5 Fazit zur Konstruktion des Modells und deren inhaltliche Implikationen

Der R-Wert hängt im RKI-Modellansatz ausschließlich von Saisonalität, Varianten, Maßnahmen und Impfungen ab. Es gibt keine weiteren Variablen im Modell, die eine Senkung des R-Wertes, z.B. aufgrund der o.g. Selbstlimitierung oder anderer Faktoren, herbeiführen könnten - außer Maßnahmen und Impfungen.

Insbesondere die - vom Modellierungsteam *vor-definierten* - Effekte der neuen Varianten sorgen dafür, dass der R-Wert im Modell erst "künstlich" höher gemacht wird, als er es in der Realität war. Dann vergleicht das Modell diesen künstlich erhöhten R-Wert mit dem tatsächlichen R-Wert. Da es außer Maßnahmen und Impfungen keine anderen Erklärungsvariablen im Modell gibt, auf die diese Abweichung zurückzuführen sein könnte, ist ein drosselnder Effekt von Maßnahmen und Impfungen bereits vordefiniert. Vereinfacht ausgedrückt: Die Abweichungen des echten R_t -Verlaufs von der angenommenen Modellkurve werden also vollständig Maßnahmen und Impfung zuerkannt, da im Modell überhaupt gar keine anderen Gründe hierfür vorgesehen sind.

Vor dem Hintergrund ist festzuhalten, dass - unabhängig von etwaigen Annahmenverletzungen in der Modellanalyse (siehe unter Kap. 2.3) - bereits die Spezifikation des Modells eine Verzerrung beinhaltet, da im Modell *nur* Maßnahmen und Impfungen das Infektionsgeschehen entschleunigen können. Die angenommene bzw. erwünschte Wirkung von Maßnahmen und Impfungen wird hierdurch bereits stillschweigend vorausgesetzt.

2.3 Modellansatz des RKI: Annahmenverletzungen in der Modellanalyse

2.3.1 Umgang mit Multikollinearität bei den Maßnahmenvariablen

In der Corona-Pandemie wurden unzählige Maßnahmen über die jeweiligen Verordnungen der Bundesländer etabliert, die sich in weiten Teilen zeitlich überschneiden haben (z.B. zeitlich weitgehend parallele Schließung von Schulen und Einzelhandel des nicht-täglichen Bedarfs). Sobald es Koinzidenzen zwischen dem Einsetzen (und Lockern) verschiedener Maßnahmen gibt, drückt sich das in Korrelationen derjenigen unabhängigen Variablen aus, die eben diese Maßnahmen beschreiben; es liegt sog. [Multikollinearität](#) vor, ein in der Ökonometrie häufig auftretendes und viel diskutiertes Problem (Prinzip: die abhängige Variable Y wird statistisch in hohem Maße von den unabhängigen Variablen X_1 und X_2 erklärt, allerdings hängen X_1 und X_2 selbst stark voneinander ab). Wenn Multikollinearität in

keine Aussage treffen. Das RKI ist diesbezüglich zeitweise ("Genesenenstatus" von ungeimpften Personen bei früherer Infektion mit der Delta-Variante oder einer früheren Virusvariante) von 90 Tagen ausgegangen (https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Genesenennachweis-old.html). Da auch das RKI stets von einer *vorübergehenden* Immunisierung durch Infektion ausgegangen ist, verwundert es, dass im RKI-Modell ausschließlich die Impfung als Immunisierungsgrund auftaucht. In einem zeitbezogenen Modell hätte genauso die "natürliche" zeitweise Immunisierung berücksichtigt werden müssen. Ein Zeitterm wäre hier aber nur eine stark vereinfachte Herangehensweise.

einem Modell in hohem Maße vorliegt, ist die inferenzstatistische Einordnung der erklärenden Variablen (Signifikanztest, Konfidenzintervalle) nicht mehr korrekt.

Hierbei ist es zunächst erfreulich, dass dieses Problem in der RKI-Studie auch aufgegriffen wird (S. 19-21). Es werden einerseits derartige Abhängigkeiten mit Hilfe von Korrelationsmatrizen dokumentiert, andererseits werden, aufgrund als zu hoch empfundener Korrelationen, einzelne Bereiche (z.B. Maßnahmen für Grund- und weiterführende Schulen) zusammengefasst: *“Da manche der NPI zeitgleich in Kraft traten, verstärkt oder gelockert wurden, gibt es im Aktivierungsprofil der Maßnahmen teilweise starke positive Korrelationen [...] Diese drei Paare von Bereichen [Schulen, öffentliche Veranstaltungen, Sport] wurden daher zunächst wie in Tabelle 1 beschrieben zusammengefasst [...] Diese stark korrelierten Bereiche [u.a. Sport/Kultur/Beherbergung/Gastronomie] konnten im Rahmen von statistischen Analysen nicht gleichzeitig in einem gemeinsamen Modell analysiert werden, da dies zu instabilen Schätzwerten führte. Daher konnten bei stark korrelierenden NPI nur kombinierte Variablen in das Modell eingeschlossen werden”* (S. 19-20). Schon diese Auseinandersetzung mit dem Thema Multikollinearität wurde in vielen früheren NPI-Studien versäumt.

Es ist jedoch sehr unwahrscheinlich, dass das Multikollinearitätsproblem damit annähernd gelöst worden wäre: In den Abbildungen 22 und 23 (S. 35/36) werden die Modellergebnisse zunächst gesamt und dann getrennt nach Altersgruppen dargestellt, wobei sie jeweils aufsteigend nach den Modellkoeffizienten, sprich: absteigend nach ihrer im Modell ermittelten reduzierenden Wirkung auf den R-Wert, geordnet sind. Hierbei finden sich etliche Variablen, die mindestens mittlere positive Korrelationen aufweisen müssen, was weitgehend schon aus der Korrelationsmatrix zu den “Aktivierungsprofilen” der Maßnahmen in Abb. 6 (S. 20) deutlich wird. Zudem bezieht sich diese Voranalyse (und die darauf aufbauende Teilzusammenfassung von Maßnahmenvariablen) nur auf einzelne Prädiktoren, nicht auf deren Linearkombinationen (Prinzip: die unabhängige Variable X_1 wird in hohem Maße von den anderen unabhängigen Variablen X_2 und X_3 erklärt).

Es wäre daher zumindest vonnöten gewesen, diese Effekte zu überprüfen. Hierfür sind z.B. Varianzinflationsfaktoren (VIF), die auf Hilfsregressionen für die unabhängigen Variablen basieren, oder andere Diagnosemaße denkbar. Eine andere (und häufig verwendete) Möglichkeit im Sinne sparsamerer Regressionsmodelle wäre es gewesen, die Anzahl der erklärenden Variablen deutlich zu reduzieren²⁹. Im letztgenannten Fall könnte das etwa bedeuten, die Analyse der Maßnahmenwirkungen auf eine Auswahl besonders empfindlicher Eingriffe zu reduzieren (z.B. Ausgangssperre, vollständige Schul- und Gastronomieschließungen) und deren Wirkung zu untersuchen. Im vorliegenden Modellansatz der RKI-Studie wird jedoch die Maßgabe für Regressionsmodelle, dass die unabhängigen Variablen keine hohen Korrelationen aufweisen dürfen, definitiv verletzt.

2.3.2 Fehlende Voraussetzungen für die statistische Inferenz

Die grundlegende Problematik der inferenzstatistischen Vorgehensweise dieser Studie des RKI besteht darin, dass die in das Regressionsmodell integrierten theoretischen

²⁹ Einen Überblick zur Diagnose und Eindämmung von Multikollinearität bietet z.B. das Ökonometrie-Standardwerk von Greene (2003).

Beziehungen bereits *a priori* als wahr angenommen werden. Dabei werden die grundlegenden Informationen als gesichertes Wissen behandelt und nicht als vorläufige Zusammenhänge, die anhand der Daten getestet werden müssen (*petitio principii* Fehlschluss). Dies manifestiert sich beispielsweise in der folgenden Aussage: „...zeigt diese Analyse als wichtiges Ergebnis, dass die Verordnungen bereits vor ihrem Inkrafttreten einen Effekt auf die Ausbreitung der COVID-19-Epidemie in Deutschland ausgeübt haben“ (S. 29).

Aus dieser Perspektive wird die empirische Modellierung als reines „curve fitting“ betrachtet, das in erster Linie von der Anpassungsgüte geleitet wird. Konkret wird bei dieser Methode eine Familie von Wahrscheinlichkeitsverteilungen (hier: Log-Normalverteilung) gewählt und das darauf basierende statistische Modell so geschätzt, dass es am besten zu den vorliegenden Daten passt. Dies wird auf den Seiten 29-33 der Studie unmittelbar deutlich, da die Selektion des konkreten Modells mittels der Verzugstage lediglich auf dem AIC-Informationskriterium basiert, welches die Anpassungsgüte unter Berücksichtigung der Komplexität optimiert (Lütkepohl & Krätzig 2004, S. 33).

Das Endergebnis des „curve fittings“ ist typischerweise ein geschätztes Modell, das sowohl statistisch als auch inhaltlich falsch spezifiziert ist: Die eigentlich interessierenden Phänomene des zugrundeliegenden stochastischen Mechanismus, der die vorliegenden Daten erzeugt hat, werden nicht ausreichend aufgeklärt. Für induktive Schlussfolgerungen wäre vielmehr ein statistisches Modell notwendig, das die systematischen Informationen aus den Daten extrahiert, wodurch die Residuen der Regression keine wesentlichen systematischen Informationen beinhalten. Der entscheidende Nachteil des reinen „curve fittings“ besteht diesbezüglich darin, dass die Residuen anhand einer mehr oder weniger willkürlichen Distanzfunktion minimiert werden, ohne näher auf ihre Systematik einzugehen.

In der Literatur zur mathematischen Statistik ist bestens dokumentiert, dass die Höhe der Fehlervarianz für die Robustheit der Inferenz von untergeordneter Bedeutung ist, solange die Stichprobengröße angemessen ist und sich die Korrelationen zwischen den erklärenden Variablen in einem gewissen Rahmen halten.³⁰ Hingegen führen systematische Fehler grundsätzlich zu falschen inferenzstatistischen Schlüssen, die auch nicht durch große Stichproben geheilt werden können. Das Modell ist in diesem Fall nicht in der Lage, zwischen dem Einfluss der erklärenden Variablen und sonstigen, nicht modellierten Effekten zu unterscheiden. Dass sich die vorliegende Studie des RKI mit einigen paradoxen Effekten zu beschäftigen hat (S. 33), passt nur zu gut in dieses Bild. Auch die Höhe des geschätzten Einflusses verschiedener Maßnahmen – besonders mit Blick auf den auf Seite 34 beschriebenen Impfeffekt – ist vor diesem Hintergrund kritisch zu hinterfragen.

Die grundlegende Systematik eines statistischen Modells lässt sich grob in die folgenden Konzepte der Wahrscheinlichkeitstheorie einteilen, auf die im Folgenden bezüglich der Studie näher eingegangen wird:

1. Eine angemessene Wahrscheinlichkeitsverteilung für den R-Wert.
2. Die Homogenität dieser Wahrscheinlichkeitsverteilung im gesamten Zeitraum (sind z. B. Momente dieser Verteilung abhängig von der Zeit oder von anderen Variablen?).
3. Die Abhängigkeiten zwischen den Ausprägungen des R-Wertes in verschiedenen Zeitpunkten.

³⁰ Vgl. u. a. Mittelhammer (2013) S. 444 - 446.

Diese Merkmale sind grundsätzlich aus den vorhandenen Daten abzuleiten und in das statistische Modell zu integrieren. So spiegelt sich beispielsweise die Wahrscheinlichkeitsverteilung im Histogramm der Daten und den zugehörigen deskriptiven Statistiken wider, während sich [zeitliche Abhängigkeiten](#) in der (partiellen) Autokorrelationsfunktion manifestieren. Derartige Methoden zur Gewinnung von systematischen Informationen aus den Daten sind in dieser Studie nicht vorzufinden.

Bezüglich der Wahrscheinlichkeitsverteilung wurde in der Studie eine Normalverteilung des logarithmierten R-Wertes gewählt, die nicht durch die herangezogenen Daten visualisiert wurde. Im Allgemeinen ist die Inferenz auf der Basis von statistischen Tests und Konfidenzintervallen robust gegenüber Abweichungen von dieser Annahme, insbesondere dann, wenn die Verteilung einen hohen Grad an Symmetrie aufweist. Während diese Annahme gemäß der Grafik zum R-Wert im weiteren Verlauf durchaus plausibel erscheint, stellt die Anfangsphase eine gravierende Abweichung mit großen Ausreißern dar, welche die (Rechts-)Schiefe der Verteilung erhöht.

Bei einem Regressionsmodell kommt hinzu, dass die Sensibilität der Inferenz in Bezug auf Abweichungen des logarithmierten R-Wertes von der Normalverteilung umso stärker ist, je stärker die erklärenden Variablen von einer Normalverteilung abweichen (Ali & Sharma 1996). In Anbetracht von Tabelle 1 (S. 13ff.) ist deutlich erkennbar, dass es sich bei den erklärenden Variablen nahezu ausschließlich um diskrete Variablen mit wenigen Ausprägungen (Dummy-Variablen) handelt, die sich nicht durch eine Normalverteilung beschreiben lassen; die Bevölkerungsgewichtung ändert daran nichts, da in den meisten Fällen für alle Bundesländer dieselben Maßnahmen galten.

Zwar werden in Abbildung 28 (S. 43) Sensitivitätsanalysen in Bezug auf verschiedene Zeiträume gezeigt, aber es erfolgt keine Darstellung der Ergebnisse unter Ausschluss des ersten Zeitraumes. Dieser ist zwar relativ kurz, jedoch sind die Ergebnisse teilweise sehr widersprüchlich, wie man am Beispiel der erklärenden Variable „Nachtleben“ gut sehen kann. Die Gefahr, dass die Einbeziehung des ersten Zeitraumes in das Hauptmodell zumindest zu einer leicht verzerrten Abbildung der Realität im Gesamtzeitraum führt, wird indirekt in Abbildung 31 (S. 47) anhand des Vergleichs mit einem Gamma-Modell bestätigt. Relativ zur Gamma-Verteilung kann die Log-Normalverteilung Ausreißer nach oben aufgrund ihrer größeren Rechtsschiefe sowie ihres höheren Variationskoeffizienten bei gleichen Erwartungswerten und Varianzen besser abbilden, was auf einen möglichen Einfluss des ersten Zeitraumes hindeutet.

Bei einem Regressionsmodell ist auch die Betrachtung der auf die Werte der erklärenden Variablen bedingten Wahrscheinlichkeitsverteilung von essenzieller Bedeutung. Konkret ist zu fordern, dass der auf die erklärenden Variablen und bundeslandspezifischen Konstanten bedingte Erwartungswert der Fehlerterme seinem unbedingten Erwartungswert entspricht. Dies impliziert den leichter verständlichen Sachverhalt, dass der Fehler mit den erklärenden Variablen und den bundeslandspezifischen Konstanten – im Sinne einer strikten [Exogenität](#) über alle Zeitverzögerungen hinweg – unkorreliert ist. Insbesondere wird dadurch ausgeschlossen, dass der heutige Fehlerterm eine Auswirkung auf die Ausprägungen der erklärenden Variablen in der Zukunft hat. Eine Verletzung dieser Bedingung hat weitreichende Folgen, da die Modellparameter generell verzerrt sind, sodass der Einfluss der erklärenden Variablen durch sonstige Einflüsse überlagert wird.

Übertragen auf die vorliegende Studie bedeutet dies, dass die Maßnahmen nicht auf Ereignisse in der Vergangenheit bezüglich des R-Werts reagieren dürfen. Da der R-Wert jedoch gemäß den Ergebnissen der vorliegenden Studie einen zeitlichen Vorlauf zu den Maßnahmen aufweist, kann diese Annahme kaum aufrechterhalten werden. Zudem existiert auch keine direkte Erweiterung des zugrundeliegenden Modells, um diesen Einfluss zu kontrollieren. Vielmehr würde dies eine simultane Modellierung des R-Werts und der Maßnahmen erfordern. Daher ist auch die Begründung, wonach „*die Verordnungen bereits vor ihrem Inkrafttreten einen Effekt auf die Ausbreitung der COVID-19-Epidemie in Deutschland ausgeübt haben*“ (S. 29) bereits aus diesem Grund äußerst heikel und ohne eingehende Analyse der wechselseitigen Beziehungen zwischen dem R-Wert und den Maßnahmen in dieser Form nicht haltbar.

Außerdem wird die Annahme der Exogenität durch die Vernachlässigung von wesentlichen Variablen im Modell verletzt. Die vorliegenden komplexen Umweltbedingungen erlauben es sicherlich nicht, sämtliche Einflussfaktoren des R-Wertes zu beobachten und zu messen und in einem Modell zu berücksichtigen. Indes ist es dringend erforderlich, mögliche Effekte von vernachlässigten Variablen zu analysieren, welche die Koeffizienten der im Modell befindlichen erklärenden Variablen beeinflussen können und folglich deren partielle Effekte (“*ceteris paribus*”) verzerren. In den einzelnen Absätzen des vorliegenden Kommentars lassen sich einige bedeutende Quellen solcher Verzerrungen identifizieren:

1. durch fehlende zeitliche Synchronisation des R-Wertes (Abschnitt 2.1.3)
2. durch fehlende Abbildung des Testverhaltens (Abschnitt 2.1.5)
3. durch zeitliche Verzögerungen von Meldungen (Abschnitt 2.1.6)
4. durch vordefinierte Parameter für die jeweiligen Virusvarianten (Abschnitt 2.2.2)
5. durch weggelassene Variablen zur räumlichen Abhängigkeit (Abschnitt 2.3.3)

Beispielsweise wird aus der Diskussion zum Testverhalten (Abschnitt 2.1.5) klar, dass die Wirkungen der Impfung aufgrund der Befreiung von der Testpflicht nicht unabhängig davon betrachtet werden dürfen. In Abschnitt 3.4 wird diesbezüglich die Richtung der Verzerrung des Parameters für die Impfung dargestellt. Neben einer solchen Analyse der Verzerrung stellen Proxy-Variablen oder bei deren Nichtverfügbarkeit die Schätzung mittels der Methode von Instrumentenvariablen mögliche Lösungsansätze dar.

Es ist somit davon auszugehen, dass die dargestellten Effekte der Maßnahmen aufgrund der fehlenden zeitlichen Exogenität und der vernachlässigten Variablen merklich von den tatsächlichen Effekten abweichen. Die oben genannten paradoxen Effekte einiger Maßnahmen (S. 33) und der hohe Impfeffekt mit einem Faktor von 0.66 (S. 34) können durchaus zu einem großen Teil aus der Verletzung der Annahme der Exogenität resultieren.

Im Folgenden werden die beiden anderen Punkte – Homogenität und zeitliche Abhängigkeiten – gemeinsam betrachtet. In diesem Kontext geht die Modellgleichung von einer zeitkonstanten Varianz aus, während mögliche Autokorrelationen weder durch dynamische ARMA-Terme noch durch eine besondere Fehlerspezifikation berücksichtigt wurden (S. 26). Dies ist insbesondere deshalb überraschend, da gemäß der Definition des R-Wertes (S. 11) eine Überlappung von 2 Tagen stattfindet, die für diesen Zeitraum auch eine entsprechende Autokorrelation einführt. [*Heteroskedastizität*](#) (die Varianz des Fehlerterms hängt von der Zeit und/oder den Ausprägungen der erklärenden Variablen ab)

und Autokorrelationen führen zwar im Gegensatz zu fehlender Exogenität zu keiner Verzerrung der zu schätzenden Parameter, aber sie verzerren deren Varianz. Dies gilt auch für die entsprechenden Eigenschaften des Fehlerterms über die Bundesländer hinweg (Abschnitt 2.3.3). Da die gewöhnlichen Standardfehler der Regressionsparameter direkt auf diesen Varianzen basieren, können sie nicht mehr für die übliche Konstruktion von Konfidenzintervallen auf der Basis von t-Statistiken herangezogen werden. Zum anderen verzerrt dies auch das Bestimmtheitsmaß R-Quadrat, das insbesondere bei positiver Autokorrelation der Fehler und moderater Schwankung der erklärenden Variablen automatisch in Richtung des Wertes 1 getrieben wird (Granger & Newbold 1974). In der Tat ist der Wert des Bestimmtheitsmaßes R-Quadrat von 62,7 % mit Blick auf die Abbildung 28 (S. 43) bemerkenswert hoch, wenn man bedenkt, dass keine dynamischen ARMA-Terme in das Modell integriert wurden und der Umfang der erklärenden Variablen für eine breite Abbildung des Infektionsgeschehens zu gering erscheint.³¹

Aus der Grafik des R-Wertes ist direkt ersichtlich, dass dieser irregulären Zyklen folgt, welche eine typische Ausprägung positiver Autokorrelationen kennzeichnen. Diese Irregularität der Zyklen erlaubt auch keine rein deterministische Modellierung der Periodizität mittels Sinus- und Cosinuswellen, die in dieser Studie implementiert wurde. Teilt man die Grafik zum R-Wert in verschiedene zeitliche Fenster auf, wird auch die fehlende Homogenität der Varianz im Zeitverlauf deutlich. Eine mögliche Problemlösung ist die Verwendung von robusten Standardfehlern. Im vorliegenden Fall mit einer in Relation zu den Bundesländern großen Zahl an zeitlichen Beobachtungen eignet sich der Standardfehler von Driscoll und Kraay (1998), um die Verzerrung der Varianz durch Heteroskedastizität und Autokorrelation zu beheben. Da dieser auf asymptotischen Resultaten basiert, die in praktischen Anwendungen möglicherweise nicht gelten, findet häufig ein Vergleich mit anderen Methoden wie dem simulationsbasierten Bootstrapping für Panel-Daten statt.

Die direkt aus den Grafiken erkennbaren und zum Teil durch die Konstruktion des R-Wertes begründeten positiven Autokorrelationen führen dazu,³² dass die gewöhnlich verwendeten Konfidenzintervalle zu eng sind.³³ Bei der Betrachtung von Abbildung 22 sieht man beispielsweise an den Stufen 3, 4 & 6 die knappe statistische Signifikanz dieser Maßnahmen in der Altersgruppe 0-17.

Diese Erkenntnis resultiert daraus, dass ein Konfidenzintervall aus allen möglichen Nullhypothesen auf dem entsprechenden Konfidenzniveau besteht, die nicht verworfen werden. Da diese Konfidenzintervalle die Nullhypothese $H_0 = 0$ nicht enthalten, wird der Effekt der Maßnahmen statistisch bestätigt. Auf Basis der vorliegenden Daten ist hingegen davon auszugehen, dass diese Intervalle bei korrekter Ermittlung breiter sind und so möglicherweise auch den Wert Null enthalten. Dies würde wiederum gegen die in der Grafik

³¹ So ist es im Kontext der Regressionsanalyse von Zeitreihen in der Praxis üblich, vergangene Werte der zu erklärenden Variablen (AR-Terme) in das Modell zu integrieren, um historische Einflüsse auf den heutigen Wert der zu erklärenden Variable zu beschreiben, die mit anderen Variablen meist nicht angemessen abgebildet werden können.

³² Der entsprechende Einfluss der Heteroskedastizität hängt von deren konkreter Form ab und kann ohne die Analyse der Daten an dieser Stelle nicht beurteilt werden.

³³ Anhand der vorliegenden Studie des RKI ist nicht erkennbar, dass für die Kalkulation der Konfidenzintervalle besondere Methoden verwendet wurden. Wie im Haupttext bereits erläutert, spricht zumindest die Spezifikation einer zeitkonstanten Varianz im Modell (S. 26) gegen eine Berücksichtigung dieser speziellen Sachverhalte. Daher wird im Folgenden das Problem erklärt, das mit der Anwendung von gewöhnlichen Konfidenzintervallen verbunden ist.

suggestierte statistische Signifikanz der genannten Maßnahmen sprechen. Folglich ist bei der Interpretation der dargestellten Ergebnisse äußerste Zurückhaltung geboten.

Zusammenfassend ist zu konstatieren, dass notwendige Annahmen der Inferenzstatistik bei der Konstruktion des Regressionsmodells nicht berücksichtigt wurden und somit verletzt sind.³⁴ Besonders schwer wiegen die fehlende Exogenität der erklärenden Variablen und die Autokorrelationen in den Fehlertermen. Während die Autokorrelationen in angemessener Weise in das verwendete Modell integriert werden können, erfordert die fehlende Exogenität – mit Verweis auf die obigen Ausführungen und den Abschnitt 1.2 zum Methodenansatz – dringend eine Erweiterung des bestehenden Modellrahmens. Zwar kann das Verfahren der Differences-in-Differences als Spezialfall eines allgemeinen Modells mit festen Effekten betrachtet werden,³⁵ jedoch darf man sich dann im Untersuchungsdesign nicht darauf beschränken, die Korrelationen zwischen den Maßnahmen und bundeslandspezifischen Einflüssen zu kontrollieren.

Auch an dieser Stelle ist nochmals zu betonen, dass die zu Grunde gelegten Daten sowie der Quellcode der Auswertung zur Gewährleistung der Nachvollziehbarkeit der Resultate veröffentlicht werden sollten.

2.3.3 Räumliche Effekte fehlen

Es ist zunächst durchaus zu befürworten, dass die RKI-Studie in ihrem Modell einen räumlichen Ansatz gewählt hat, und zwar in dem Sinne, dass auf subnationaler Ebene (Bundesländer) gearbeitet wird. Allerdings ist der Modellansatz ein Paneldatenmodell ohne räumliche Komponente(n). Eine (oder mehrere) räumliche Modellkomponenten würden dem Umstand Rechnung tragen, dass die unabhängige Variable und/oder der Fehlerterm und/oder auch einzelne unabhängige/erklärende Variablen möglicherweise räumlich autokorreliert sind.

Räumliche Autokorrelation ist ein Begriff auf der Geostatistik bzw. der räumlichen Ökonometrie und bedeutet sinngemäß, dass Beobachtungen in einer Region nicht unabhängig sind von Beobachtungen in benachbarten bzw. nahe liegenden Regionen (es ist gewissermaßen das räumliche Gegenstück zur zeitlichen Autokorrelation, s.o.). Die Berücksichtigung räumlicher Komponenten ist dabei einerseits ein Informationsgewinn, da auch Spillover-Effekte berücksichtigt werden; andererseits werden bei der Nutzung räumlich autokorrelierter Daten in einem nicht-räumlichen Modell auch Modellannahmen verletzt (siehe hierzu z.B. Eckey et al. 2006).

³⁴ Es sei hier angemerkt, dass die spezifischen Charakteristika von Zeitreihen bei der Beurteilung der notwendigen Modellannahmen Eingang finden müssen. Die ökonometrischen Werke von Hayashi (2000) und Greene (2003) bieten hierzu einen hervorragenden Überblick.

³⁵ Siehe hierzu Wooldridge (2001) S. 266 - 267.

Die Ausbreitung von Infektionskrankheiten – ein räumlich-zeitlicher Prozess – ist sogar ein außergewöhnlich gutes Beispiel für räumliche Autokorrelation³⁶. Dass bei SARS-CoV-2 (ebenso wie bei anderen infektiösen Atemwegserkrankungen) die beschreibenden Variablen (z.B. Infektionsraten, Todesfälle, Wachstumsraten) hochgradig räumlich autokorreliert sind, wurde unzählige Male empirisch nachgewiesen (z.B. Bourdin et al. 2021, Saffary et al. 2020, Wang et al. 2021, Wieland 2020).

Um diese Effekte in Regressionsmodellen zu berücksichtigen, bieten sich verschiedene Formen von räumlichen Paneldatenmodellen an, die die unterschiedlichen Formen von spatial lags inkludieren (Überblick siehe Elhorst 2014), was im Modell der RKI-Studie jedoch unterblieben ist. Somit wird zwar – durch die festen Effekte für Bundesländer – ansatzweise räumliche Heterogenität abgebildet, nicht aber raumbezogene Zusammenhänge.

3 Zur inhaltlichen Interpretation der Ergebnisse

Unabhängig vom fehlenden Kausalansatz und den weiteren methodischen Problemen, die in der RKI-Studie auftreten, lässt sich feststellen, dass auch Teile der inhaltlichen Schlussfolgerungen nicht mit den formalen Modellergebnissen kongruent sind.

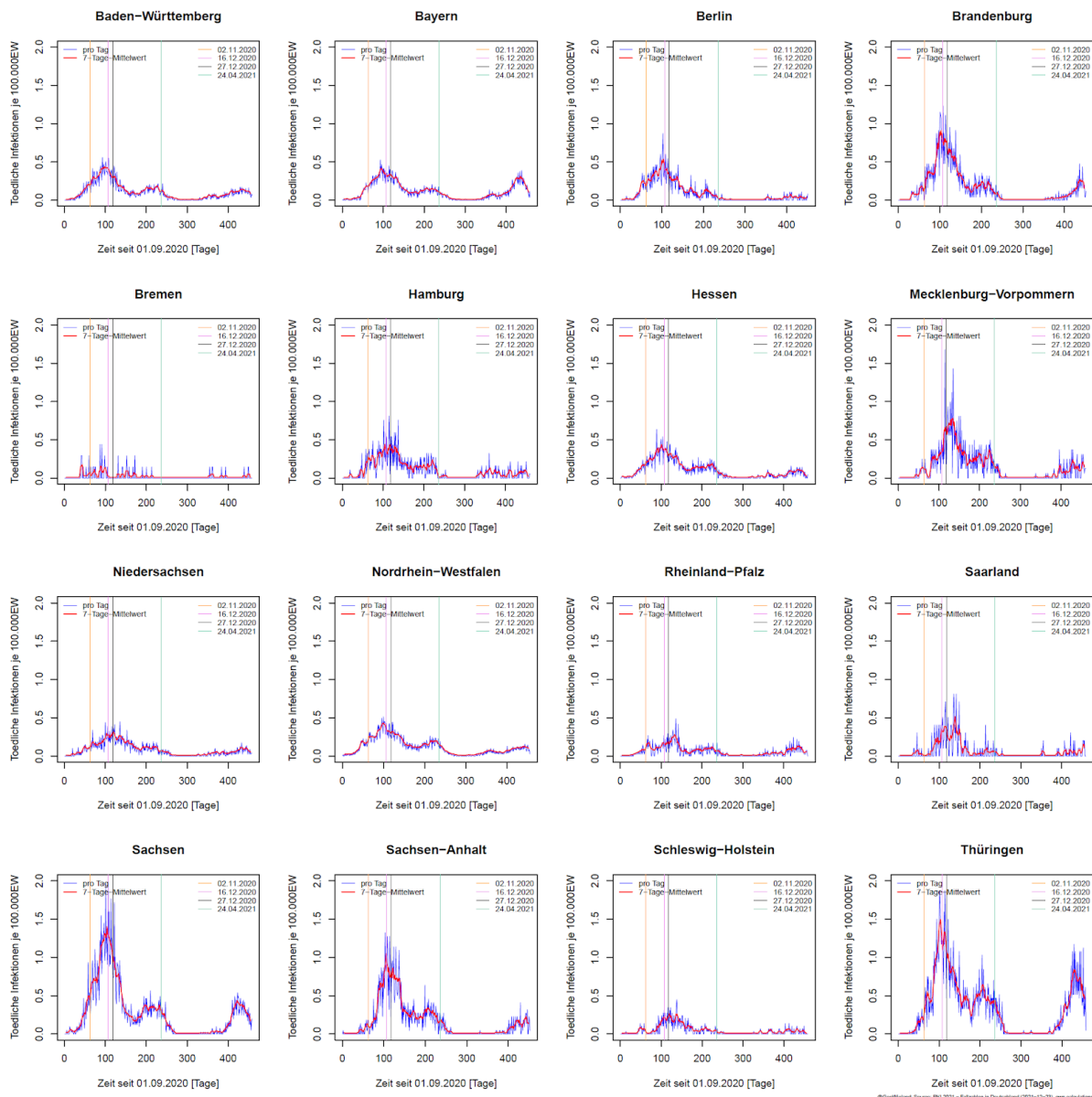
3.1 Maßnahmen “wirken” schon vor ihrer Einführung

In der RKI-Studie wird der zeitliche Verzug zwischen dem Inkrafttreten einer bestimmten Maßnahme und deren (angenommener) Wirkung auf den R-Wert untersucht. Hierfür werden Modellvarianten mit unterschiedlichen Verzügen durchgespielt und dabei festgestellt, *“dass relativ unabhängig von dem gewählten Zeitraum ein angenommener Beginn der Wirkung auf den R-Wert bereits einen oder zwei Tag vor Inkrafttreten einer Verordnung („Verzug“ von -1 Tag oder -2 Tagen) die beste Nachbildung des beobachteten Verlaufs der R-Werte in den Altersgruppen und insgesamt erlaubt [...] Da dem von uns betrachteten R-Wert der Erkrankungsbeginn der COVID-19 Fälle zugrunde liegt und die Infektion jeweils noch vor dem Erkrankungsbeginn erfolgte, zeigt diese Analyse als wichtiges Ergebnis, dass die Verordnungen bereits vor ihrem Inkrafttreten einen Effekt auf die Ausbreitung der COVID-19-Epidemie in Deutschland ausgeübt haben”* (S. 29).

Zunächst bedeutet das, dass spürbare Abschwächungen des Infektionsgeschehens nachweisbar sind, *bevor* bestimmte Maßnahmen bzw. Maßnahmenpakete in Kraft getreten sind, d.h. die Wirkung vor der (angenommenen) Ursache einsetzte. Kausalanalytisch bedeutet dieses Ergebnis, dass die Nullhypothese nicht widerlegt werden kann, da der Eingriff nicht vor dem Effekt kam (siehe Kap. 1.2).

³⁶ Einerseits ähnelt sich das Infektionsgeschehen benachbarter bzw. nahe liegender Regionen; dies ist auch darauf zurückzuführen, dass zwischen Regionen *räumliche Interaktionen* zwischen Menschen stattfinden, die zu Infektionen führen können (*Spillover*-Effekt, Beispiel: Pendlerbewegungen zwischen Landkreis A und Stadt B, die dann zu Ansteckungen z.B. im Großraumbüro führen können). Dieser Umstand führt zu einem *spatial lag* in der abhängigen Variable. Andererseits sind auch viele Umstände, die sich auf den Verlauf von Infektionswellen und Krankheitsverläufen auswirken können, räumlich autokorreliert, z.B. klimatische Ausgangsbedingungen (z.B. Temperatur, Luftfeuchte) oder gesundheitliche Faktoren (z.B. Luftverschmutzung). In diesen Fällen bestehen spatial lags bei den erklärenden Variablen. Auch ein *spatial lag* im Fehlerterm ist möglich, der - bei Nichtbeachtung - für eine Verzerrung der geschätzten Modellkoeffizienten sorgt.

Insofern bestätigt dieses Ergebnis der RKI-Studie eine Reihe früherer Studien aus dem In- und Ausland, die die Notwendigkeit bzw. Effektivität bestimmter Maßnahmen oder Maßnahmenpakete (z.B. vollständiger Lockdown) auf dieser Grundlage in Frage gestellt haben (z.B. Berlemann & Haustein 2023, Boudou et al. 2021, Gibson 2022, Küchenhoff et al. 2021, Santamaria & Hortal 2021, Wieland 2020, Wood 2022). Dass Infektionswellen räumlich-zeitlich versetzt ablaufen und teilweise lange vor, teilweise lange nach Inkrafttreten von Maßnahmen(-paketen) brechen, zeigt auch die folgende Analyse:



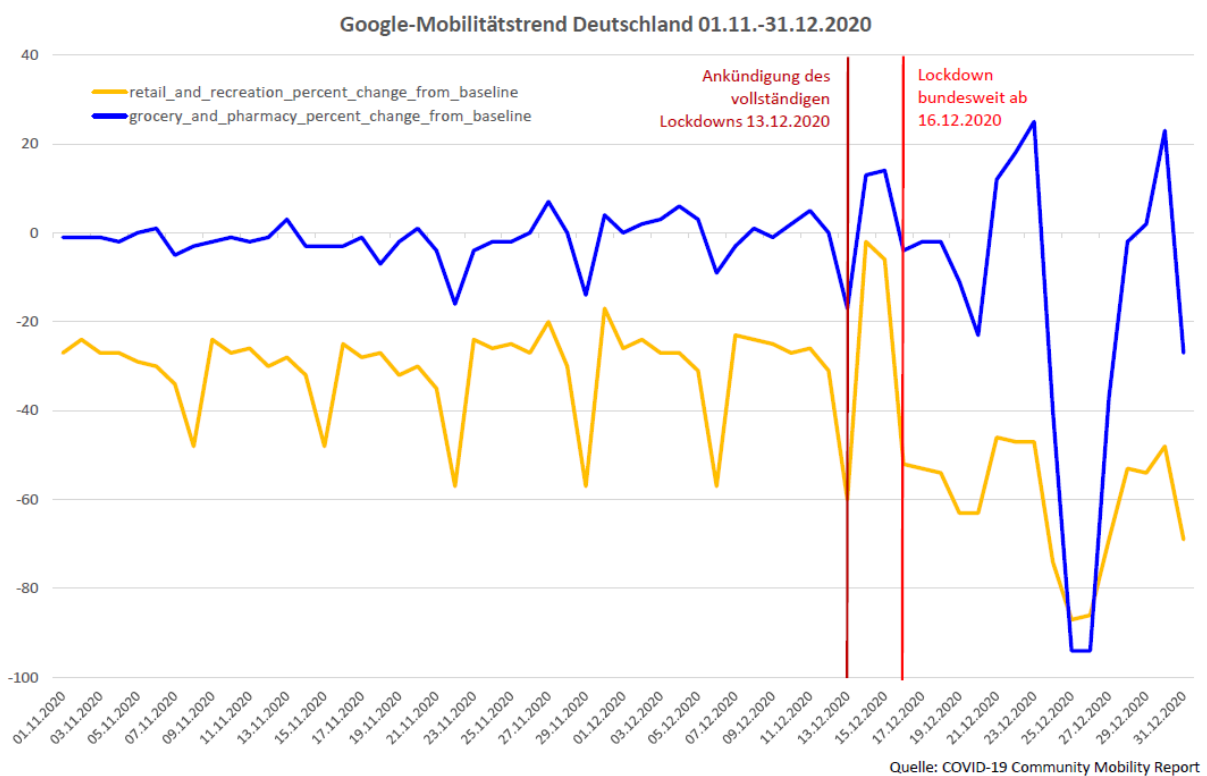
COVID-19-Todesfälle nach imputiertem Infektionsdatum je 100.000 Einwohnern nach Bundesländern. Die vertikalen Striche kennzeichnen das Inkrafttreten verschiedener Maßnahmen(-pakete) in der zweiten Corona-Welle (Winter 2020/21). Quelle: Eigene Darstellung.

Anstatt jedoch vor diesem Hintergrund die kritische Frage nach der tatsächlichen Effektivität von einzelnen Maßnahmen oder gar kompletten Lockdowns zu stellen, kehrt die RKI-Studie die Interpretation dieses Ergebnisses als Beleg für deren Wirksamkeit um. Die Argumentation geht dahin, dass die Bevölkerung jene Maßnahmen antizipiert und "freiwillig"

ihr Verhalten geändert habe: *“Der Effekt der NPI zeigte sich dabei bereits kurz vor dem Inkrafttreten der jeweiligen Verordnungen. Die naheliegendste Erklärung dafür ist, dass Verhaltensanpassungen in der Bevölkerung bereits vor dem Inkrafttreten der Einschränkungen erfolgten. Dies weist auf die besondere Rolle der medialen Berichterstattung und der öffentlichen Diskussionen, insbesondere im Zusammenhang mit den Ministerpräsidentenkonferenzen hin”* (S. 49).

Hierzu muss man *erstens* feststellen, dass dies lediglich *eine* mögliche Interpretation ist, und ob es *“die naheliegendste”* ist, wie es in der Zusammenfassung der RKI-Studie geschrieben wird (S. 5), sollte zumindest ein Diskussionsgegenstand sein. Hier spielt u.a. die Tatsache eine Rolle, dass der R-Wert im Laufe einer Welle zwangsläufig absinkt (siehe Kap. 2.2.3).

Zweitens, darauf aufbauend, wird durch die RKI-Studie kein empirischer Beweis für diese Aussage geleistet. Dabei wäre es - wenn auch nur eingeschränkt - durchaus möglich, Verhaltensänderungen empirisch zu prüfen, etwa anhand von Mobilitätsdaten³⁷. Was die Frequentierung von Einzelhandelsbetrieben, die, mit Ausnahme von Gütern des täglichen Bedarfs, im Zuge der Lockdowns geschlossen wurden, betrifft, zeigen die Google-Mobilitätsdaten sogar den umgekehrten Effekt:



Quelle: Eigene Darstellung. Datengrundlage: COVID-19 Community Mobility Report.

Der zweite vollständige Lockdown trat am 16.12.2020 in Kraft und wurde am 13.12.2020 offiziell angekündigt. In den beiden Tagen danach kam es sogar zur übermäßig *starken*

³⁷ Für die ersten Monate der Pandemiezeit gibt es mehrere Studien, die anhand von Mobilitätsdaten (z.B. Google, Mobilfunkanbieter) Veränderungen im Mobilitätsverhalten der Bevölkerung empirisch analysiert haben, auch im Hinblick auf Unterschiede zwischen Regionen mit und ohne vollständigen Lockdown (z.B. Goolsbee & Syverson 2021, Jacobsen & Jacobsen 2020). Auf diesem Wege ist es möglich, zumindest einen *“freiwilligen”* Rückgang bei der Mobilität (nicht notwendigerweise von physischen Kontakten) zu belegen.

Frequentierung von “nicht essentiellen” Einzelhandelsbetrieben (in der Grafik unter “retail and recreation” zusammengefasst) sowie von Anbietern von Gütern des täglichen Bedarfs (in der Grafik: “grocery and pharmacy”). Am 14. und 15.12.2020 stieg der Mobilitätswert bei den erstgenannten Anbietern sogar erstmalig seit Monaten fast wieder auf die Höhe des Referenzwertes (oranger Graph ist kurzzeitig nur knapp unter null), um dann erst ab Beginn des Lockdowns (16.12.2020) auf das Sonntagsniveau (siehe Minimalwerte alle 7 Tage im orangenen Graphen) abzusinken.

Zumindest in diesem Fall hat die Ankündigung des Lockdowns also sogar einen mobilitäts- und kontaktinduzierenden Effekt gehabt, da hierdurch Weihnachtseinkäufe zwangsweise vorgezogen und die Innenstädte und Shopping-Center sozusagen “künstlich gefüllt” wurden (FAZ 14.12.2020). Es handelt sich dabei tatsächlich um eine Art Antizipationseffekt in der Bevölkerung, jedoch in der gegenteiligen Weise, wie es die Interpretation der RKI-Studie vorsieht.

Drittens muss an dieser Stelle auch an die künstliche Senkung des Verzugs durch die Berechnung des R-Wertes (siehe unter Kap. 2.1.3) gedacht werden. Bei einer anderen Berechnungsweise würde sich ein deutlich größerer Verzug zwischen dem Einsetzen der Maßnahmen und spürbaren Senkungen des R-Wertes ergeben. Insbesondere wenn man bedenkt, dass viele Maßnahmen nur vergleichsweise kurz vor ihrer Einführung angekündigt wurden, wird das Argument der Antizipation dieser Maßnahmen mit steigender Zeitspanne immer unrealistischer³⁸.

3.2 Einfluss von Maßnahmenbündeln: Sind NPI im Verbund besonders effektiv?

In der inhaltlichen Interpretation der Modellergebnisse argumentiert die RKI-Studie explizit damit, dass die besondere Stärke der Maßnahmen ihre Kombination sei bzw., dass erst Maßnahmenbündel effektiv seien: *“Erst die Kombination von NPI in ganz verschiedenen Bereichen konnte diese Kontrolle bewirken. Einzelne NPI konnten nur im Kontext der parallel bestehenden anderen Maßnahmen bewertet werden. Der Effekt einer einzelnen Maßnahme darf nicht so interpretiert werden, dass dieser Effekt auch gewirkt hätte, wenn keine andere Maßnahme ergriffen worden wäre. Insbesondere wurden die Maßnahmen in unserem Modell als unabhängig angenommen, die realen Abhängigkeiten konnten also nicht abgebildet werden”* (S. 49).

An anderer Stelle wird dies anhand von Beispielen begründet: *“Hierdurch zeigt sich, dass die Effektivität der Maßnahmen in den einzelnen Bereichen zur Kontrolle der COVID-19 Pandemie in Deutschland in unserem Modell nicht entkoppelt werden kann [...] Dies liegt zu einem wesentlichen Teil in der Natur dieser Einschränkungen: wenn etwa im öffentlichen*

³⁸ Es fällt außerdem auf, dass in der RKI-Studie eines der zentralen Ergebnisse - Maßnahmen wirken vor ihrer Einführung - durch freiwillige Verhaltensänderungen erklärt wird, gleichzeitig aber darauf hingewiesen wird, dass die tatsächliche Durchführung bestimmter Maßnahmen gar nicht wirklich überprüft werden kann und Verstöße dagegen möglich seien: *“Manche NPI, etwa Restriktionen in Bezug auf private Treffen, konnten leicht umgangen werden, während andere NPI wie die Schließung von Schulen oder Restaurants unumgänglich waren”* (S. 50). In einem Tagesschau-Bericht werden die Autorinnen und Autoren der Studie sogar dahingehend zitiert, dass *“davon auszugehen sei, dass solche Auflagen vielfach missachtet wurden”* (Tagesschau.de 2023). Das Argument freiwilliger Verhaltensänderungen im Sinne einer Antizipation der kontaktreduzierenden Maßnahmen wird durch diese Äußerungen nicht nachvollziehbarer.

Raum die Versammlung von nur höchstens fünf Personen erlaubt ist, so kann das nur einen wesentlichen Beitrag zur Kontrolle der Epidemie liefern, wenn auch öffentliche Veranstaltungen praktisch nicht mehr oder nur sehr eingeschränkt stattfinden und wenn sich Menschen auch in den Bereichen Sport, Kultur, Beherbergung und Gastronomie kaum noch treffen – denn ansonsten wird die Infektion, die in einem Bereich vermieden werden konnte, in einem anderen Bereich dann doch erworben” (S. 38).

Es wird also ausgesagt, dass eine spezifische Maßnahme – Beispiel: Versammlungsbeschränkung im öffentlichen Raum – *nur dann* eine ernsthafte Wirkung habe, wenn andere Maßnahmen – z.B. Einschränkung von Veranstaltungen oder Gastronomie – parallel bestehen.

Erstens: Dass Maßnahmen ihre Wirkung erst im Verbund entfalten würden, mag zunächst einleuchtend oder zumindest interessant klingen, ist aber lediglich eine Behauptung, die bei genauerer Betrachtung schon aus logischen Erwägungen heraus nicht (oder zumindest nicht pauschal) haltbar ist. Dies lässt sich einfach am o.g. Textbeispiel aus der RKI-Studie illustrieren: *Wenn* etwa die Versammlungsbeschränkung im öffentlichen Raum eine Wirkung auf das Infektionsgeschehen hat, dann hat sie diesen Effekt, völlig unabhängig davon, ob es noch irgendeine andere Beschränkung gibt; sie würde, eine reale Wirkung vorausgesetzt, den R-Wert um X % reduzieren. Wenn nun beispielsweise zusätzlich Restaurants geschlossen werden, ergibt sich daraus ein weiterer möglicher Effekt auf das Infektionsgeschehen, der für sich allein steht und den R-Wert abermals um Y % senkt; der tatsächliche Effekt der Versammlungsbeschränkung – sofern er denn besteht – ändert sich dadurch nicht³⁹.

Zweitens ist diese Aussage gerade *nicht* durch die formalen Modellergebnisse gesichert. Denn der Modellansatz sieht, wie korrekterweise geschrieben wird, die Maßnahmen als unabhängig voneinander an. In Regressionsmodellen ist ein wesentliches Prinzip, das – sofern das Modell korrekt spezifiziert ist – die Wirkung einzelner Variablen isoliert bzw. unter der *ceteris-paribus*-Klausel betrachtet wird, also wenn alle anderen Bedingungen konstant gehalten werden. Gleichzeitig verändert jede Modifikation von Variablen im Modell auch die Wirkungsweise der anderen Variablen, da diese i.d.R. nicht vollkommen stochastisch unabhängig voneinander sind (Greene 2003; zu Multikollinearität siehe unter Kap. 2.3.1).

Dadurch wird in einem solchen Modell aber gerade kein *gemeinsamer* Effekt untersucht, und es wird erst recht nicht *bewiesen*, dass Maßnahmen in Kombination besonders wirksam seien, sondern im Modell wird der Effekt einer spezifischen Maßnahme betrachtet, dann der Effekt der nächsten spezifischen Maßnahme usw. Die Maßnahmeneffekte im Modell sind *additiver* Natur, d.h. ihre (angenommene) Wirkung läuft parallel, aber nicht im Verbund.

Die o.g. Interpretation der NPI-Wirkungen in der RKI-Studie zielen aber darauf ab, bzw. es wird sogar betont, dass es Wechselwirkungen also *Interaktionen*, zwischen den Maßnahmen gibt; das wird u.a. daran deutlich, dass einer bestimmten Maßnahme nur dann eine ernsthafte Wirkung zugesprochen wird, wenn parallel andere in Kraft sind.

³⁹ Es gäbe nur den möglichen Fall, dass von Ausweicheffekten ausgegangen wird (z.B. *mehr* Restaurantbesuche wegen der Versammlungsbeschränkung im öffentlichen Raum); dass dieser Fall gemeint wäre, lässt sich aber aus der Aussage, dass eine “[...] Infektion, die in einem Bereich vermieden werden konnte, in einem anderen Bereich dann doch erworben [wird]” (S. 38), nicht schließen.

Um diese Interaktionen zu untersuchen, die nicht auf einen additiven, sondern auf einen *multiplikativen* Effekt abzielen, gäbe es in der Anwendung derartiger Regressionsmodelle durchaus Möglichkeiten: In solchen Fällen wird üblicherweise mit [Interaktionstermen](#) gearbeitet, die aus dem Produkt zweier (oder mehrerer) erklärender Variablen gebildet und für die dann separate Koeffizienten geschätzt werden (Beispiel: eine Dummy-Variable für Versammlungsbeschränkungen, eine für die Schließung von Restaurants und eine für Versammlungsbeschränkungen *und* Restaurantschließungen). Auf Interaktionseffekte wurde aber im RKI-Modell verzichtet.

Derartige Wechselwirkungen - Maßnahmen wirken nur in Verbindung mit anderen - lassen sich also in der Studie weder theoretisch sauber begründen, noch empirisch nachweisen.

3.3 Die “paradoxen Effekte”

In den Abbildungen 22 und 23 (S. 35-36) der RKI-Studie werden die Modellergebnisse grafisch dargestellt, auch differenziert nach Altersgruppen, für die die unabhängige Variable *RT separat* berechnet wurde. Eine vertikale Referenzlinie differenziert zwischen negativen Effekten (d.h. Koeffizienten sind negativ bzw. das Modellergebnis ist, dass diese unabhängige Variable den R-Wert senkt) und positiven Effekten (d.h. Koeffizienten sind positiv bzw. diese unabhängige Variable erhöht den R-Wert). Hierbei fällt auf, dass einzelne Maßnahmen rechts von der vertikalen Referenzlinie ($R_t = 1$) zu finden sind, also - gemäß Modellanalyse in der RKI-Studie - den R-Wert erhöhen und demnach, inhaltlich interpretiert, kontraproduktiv sind, weil sie das Infektionsgeschehen sogar beschleunigen, anstatt dessen Geschwindigkeit zu drosseln. In der RKI-Studie findet sich zu diesen Modellergebnissen nur ein einziger kurzer Verweis in der Ergebnisdiskussion: *“Unter den anderen Maßnahmen zeigte nur die erste Eskalationsstufe der Einschränkungen bei Masken und bei den Beschränkungen am Arbeitsplatz einen paradoxen Effekt (der den R-Wert signifikant erhöht). Bei Kindern zeigten darüber hinaus die ersten Stufen der Einschränkungen in privaten Räumen und im Groß- und Einzelhandel paradoxe Effekte”* (S. 33).

Hierzu ist zunächst anzumerken, dass es natürlich prinzipiell denkbar ist, dass eine Corona-Maßnahme nicht nur keinen Einfluss haben, sondern sogar kontraproduktiv sein kann⁴⁰; dies wird vermutlich kaum bestritten, ist aber an dieser Stelle keine relevante Frage. Viel wichtiger ist, dass diese “paradoxen Effekte” ein Modellergebnis der RKI-Studie bedeuten, das den Erwartungen, wonach Maßnahmen eine mindernde Wirkung auf den R-Wert haben, zumindest in diesen Fällen widerspricht; kurz gesagt: in diesen genannten Fällen ist, ausgehend vom gegenwärtigen Untersuchungsdesign, die Hypothese einer infektionsreduzierenden Wirkung zu verwerfen. Es ist in wissenschaftlichen Arbeiten unbedingt zu erwarten, dass solche Ergebnisse nicht nur *dargestellt* bzw. kurz erwähnt, sondern auch inhaltlich interpretiert werden und/oder dass mögliche methodische Probleme, die für inhaltliche Fehlschlüsse verantwortlich sein können, mindestens benannt werden.

Zum erstgenannten Punkt - inhaltliche Interpretation - wäre es zu erwarten, dass sich die Autorinnen und Autoren der Studie damit auseinandersetzen, *warum* eben diese

⁴⁰ Ein *plausibles Beispiel* hierfür könnten etwa Kontaktbeschränkungen im öffentlichen Raum (z.B. Park) sein, die Menschen dazu animiert, sich stattdessen (heimlich) in Innenräumen (z.B. zuhause) zu treffen, wo das Infektionsrisiko (z.B. aufgrund schlechterer Belüftung) höher ist.

Maßnahmen sogar den gegenteiligen Effekt haben sollen. Hierzu Beispiele aus Abbildung 23 (S. 36): Warum sollten Maßnahmen im Einzelhandel unterschiedlicher Eskalationsstufe bei 18-59-Jährigen das Infektionsgeschehen drosseln und bei Kindern und Jugendlichen sowie in der Altersgruppe 60+ das Gegenteil bewirken? Es geht aus diesen Ergebnissen auch hervor, dass Masken im ÖPNV und im Einzelhandel (Stufe 2) bei Personen ab 60 Jahren das Infektionsgeschehen entschleunigen, aber in den anderen Altersgruppen deutlich verstärken. Warum sind Maßnahmen am Arbeitsplatz (Stufe 2) in der Gruppe von 18 bis 59 Jahren - also der mehrheitlich erwerbstätigen Bevölkerung - kontraproduktiv, während sie in der Arbeitsgruppe unter 18 Jahren den erwarteten Effekt zeigen? Warum soll die Maskenpflicht in Schulen keinen nachweisbaren Effekt in der Altersgruppe U18 gehabt haben, aber einen deutlich entschleunigenden Effekt in der Altersgruppe 60+? Diese Ergebnisse sind bereits logisch widersprüchlich, und in der RKI-Studie wird darauf überhaupt nicht eingegangen.

Der zweitgenannte Punkt betrifft die Möglichkeit, dass "falsche" Ergebnisse durch methodische Fehler im Untersuchungskonzept auftreten. Hier können verschiedene Faktoren im empirischen bzw. Modelldesign zu Grunde liegen, die zu verzerrten Koeffizienten führen (z.B. [ausgelassene Variablen](#), serielle und/oder räumliche [Autokorrelation](#); siehe Kap. 2.2 und 2.3). Weiterhin ist es wahrscheinlich, dass die Steuerung der Modellergebnisse durch die "fixierten" Variantenparameter oder die Sinus/Cosinus-Oszillationen einen Teil dazu beitragen (siehe unter Kap. 2.2.2). Es ist allein aufgrund dieser Teilergebnisse die Frage zu stellen, wie belastbar die anderen Ergebnisse sind, wenn bei dieser Modellspezifikation derart "paradoxe" Effekte überhaupt auftreten. Diese Aspekte würden sich nur nachprüfen lassen, wenn die Daten bzw. der zur Auswertung verwendete Quellcode veröffentlicht werden.

3.4 Einfluss der Impfung auf den R-Wert

Ein Ergebnis, das besondere Aufmerksamkeit verdient, ist das affirmative Ergebnis zur Reduzierung von R_t durch Impfung: *"Unsere Ergebnisse zeigen, dass eine hohe Impfquote einen stark reduzierenden Effekt auf den R-Wert hatte. Dies führte insbesondere in der älteren Bevölkerung zu einer deutlich schwächeren dritten COVID-19 Welle"* (S. 6). Das ist insbesondere auffällig, weil die empirische Wissenschaft spätestens ab der Omikron-Welle ab Winter 2021/22 von der Hypothese, dass diese Impfung einen Einfluss auf die Ausbreitung hätte, abgerückt ist und sich inzwischen auf den Einfluss auf die Krankheitsschwere fokussiert.⁴¹ Ein solches Ergebnis, wenn empirisch replizierbar, wäre ein bemerkenswerter Beitrag zur fachlichen Diskussion. Dieses muss aber wissenschaftlichen Standards entsprechen.

Auch bei diesem Ergebnis müssen wir erstmal davon ausgehen, dass es ein nichtreplizierbarer Scheineffekt ist, der durch die Konstruktion des Modells entstand.

⁴¹ Dass die COVID-19-Schutzimpfungen – je nach Impfschema – bereits nach kurzer Zeit die Ansteckungswahrscheinlichkeit nicht mehr senken, ist u.a. in einer umfangreichen Studie aus Schweden gezeigt worden (Nordström et al. 2022), die auf Daten aus dem dortigen Impf- und Gesundheitsregister basiert. Würde sich die abhängige Variable im Modell der RKI-Studie auf schwere Krankheitsverläufe beziehen (z.B. Krankenhaus/ITS-Einweisungen, Todesfälle), wäre ein signifikanter Impfeffekt plausibel und wurde auch schon in kausalanalytischen NPI-Studien festgestellt (z.B. Mader & Rüttenauer 2022); allerdings bezieht sich das RKI-Modell bewusst allein auf die Ausbreitung von Infektionen, nicht auf schwere Verläufe.

Aufgrund des fehlenden Autokorrelationsterms sind alle Effekte erstmal Abweichungen der theoretisierten Grundoszillation und nicht Abweichungen von einem empirisch konstruierten Counterfactual einer ungedämmten Ausbreitung.

Da der Effekt der SARS-CoV-2 Varianten auf den R-Wert nicht ohne Weiteres von den Effekten der Impfung, den erweiterten Textkapazitäten ab dem Sommer 2020 und auch der im Verlauf der Pandemie weniger gut an die realen Maßnahmen angepasste Dokumentation abgegrenzt werden kann, betrachten wir schließlich hauptsächlich ein Modell mit fixiertem Effekt der Varianten. Hierbei wurde festgelegt, dass die Alpha-Variante zu einem 30% höheren R-Wert und die Delta-Variante zu einem 60% höheren R-Wert im Vergleich zum SARS-CoV-2 Wildtyp führt.

$$\log(R_t) \sim \mathcal{N}(\mathbb{E}(\log(R_t)), \sigma^2)$$

$$\mathbb{E}(\log(R_t)) = 0.3v_\alpha + 0.6v_\delta + \theta_b + \beta_1 \sin\left(2\pi \frac{t}{365}\right) + \beta_2 \cos\left(2\pi \frac{t}{365}\right)$$

$$+ \beta_3 f(\text{Impf}_{t+\tau}) + \sum_{k \in K} \gamma_k m_{t+\tau}^k$$

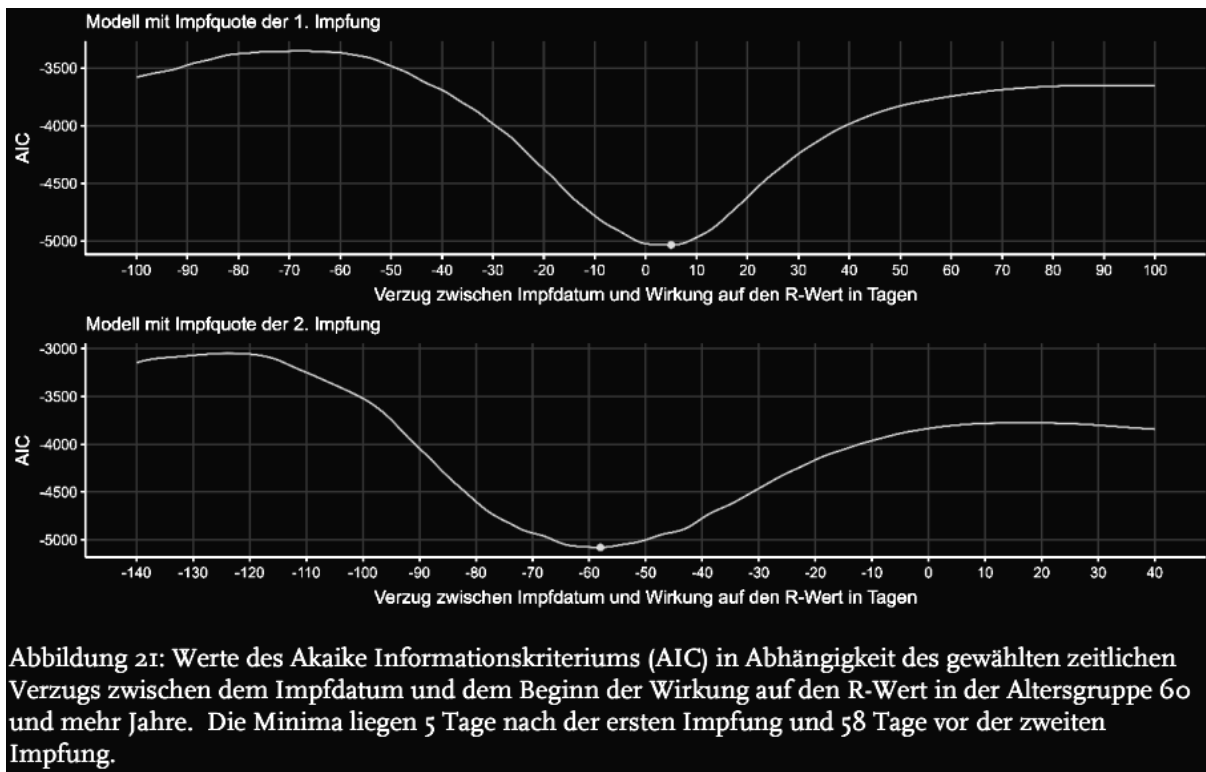
Das Hauptmodell mit händisch fixierten Parametern für die Varianten Alpha (0.3) und Delta (0.6) (RKI 2023b, S. 27).

Eine Besonderheit des Modells, die auch schon oben beschrieben wurde, ist die manuelle Fixierung der Koeffizienten für die Alpha- und Delta Varianten (siehe Kap. 2.2.2). Damit wird per fiat angenommen, dass die durch die Sinus- und Cosinus-Terme erzeugte postulierte Grundoszillation während der Hauptausbreitungszeit noch verstärkt wird. Beide – insbesondere Alpha – fallen aber in die Hauptimpfzeit im Jahr 2021.

Es ist also zu vermuten, dass der hypothetisierte Impfeffekt dadurch entsteht, dass die Divergenz zwischen modelliertem angestiegenen R-Wert und empirisch stetig im Band $R_t \sim 0,7-1.5$ verlaufenden R-Wert als “Impfeffekt” interpretiert wird. Das wäre dann sicher nicht replizierbar. Aufgrund der fehlenden Informationen können wir das nicht weiter validieren. Wir müssen aber davon ausgehen, dass dieser Effekt sich auch nicht in einem sorgfältig aufgesetzten Modell – mit empirisch ermittelten Parametern für den Varianteneffekt – replizieren lässt.

Indikatoren, die diese Einschätzung unterstützen, sind die Betrachtung der deutschen Nachbarländer oder auch der gesamten europäischen Staaten, die ausreichend Varianz in Impf- und Ausbreitungsdynamik anbieten sollten. Eine Korrelation, die die Nullhypothese eindeutig widerlegen kann, können wir auf dieser Ebene auf den ersten Blick nicht erkennen, aber auch diese Hypothese bedarf sorgfältiger empirischer Validierung. Unsere Kommentare zu den Best Practices in den vorhergehenden Abschnitten gelten auch hier.

Ein anderer Indikator ist die Ausbreitungsdynamik nach dem Beobachtungszeitraum, die auch die sehr dynamische Omikron-Dominanz beinhaltet.



Wirkungsverzug der Impfungen nach AIC-Kriterium, für erste und zweite Impfung. Quelle: RKI 2023b, S.33.

Angemerkt sei, dass gerade bei der Betrachtung der Impfeffektivität die Art der Datenerhebung zu berücksichtigen ist. Denn sowohl nicht mehr verpflichtend durchzuführende Schnelltests wie in Schulen, aber auch die Möglichkeit, dass es mit frischer Impfung eher zu asymptomatischen Verläufen kommen und damit Infektionen unentdeckt bleiben könnten, erhöht das Dunkelfeld in diesen Bereichen. Das würde zu einer Verzerrung in den Daten führen, die wiederum die Grundlage der Analyse darstellt.

Auf diese Weise lässt sich auch die Verzerrung des Parameters für die Impfung durch die Teststrategie ableiten: Die vernachlässigte Variable "Anzahl an Tests" ist mit der Anzahl der Geimpften negativ korreliert ("je mehr Geimpfte, desto weniger Tests"). Zudem hat diese vernachlässigte Variable einen nicht modellierten, positiven Koeffizienten ("je weniger Tests, desto geringer ist der R-Wert"). Dies führt zu einer negativen Verzerrung des negativen Parameters für die Impfung. Die unberücksichtigte Teststrategie führt dazu, dass der Effekt der Impfung zu hoch eingeschätzt wird.

Hinzu kommt, dass in der Studie (S. 33) angemerkt wird, dass sich der Impfeffekt etwa 5 Tage nach der ersten Impfung (und 58 Tage vor der zweiten Impfung) im R-Wert zu zeigen begann. Jedoch weist der verwendete R-Wert gegenüber den Maßnahmen nach Abschnitt 2.1.3 einen mittleren Verzugs von ca. 11 Tagen auf. Selbst wenn man berücksichtigt, dass der Impfeffekt ab dem Zeitpunkt t kontinuierlich in den herangezogenen R-Wert einfließt, müsste bei zusätzlicher Betrachtung der Inkubationszeit von mindestens 4 Tagen der Impfeffekt entweder zeitgleich oder vor der Impfung auftreten. Ein derartiger Zusammenhang lässt sich logisch nicht nachvollziehen, eine empirische Unterstützung fehlt.

4 Fazit

In der rigorosen Analyse des Abschlussberichts zur Corona-Pandemie des Robert-Koch-Instituts (RKI 2023b) finden wir:

- Mit dem eingesetzten methodischen Ansatz kann keine vollständige Kausalanalyse durchgeführt werden. Bestenfalls kann ein Nachweis der zeitlichen Abfolge (*post hoc*) geliefert werden, jedoch nicht der Verursachung (*propter hoc*). Der Modellansatz in der RKI-Studie ist somit, völlig unabhängig von anderen Defiziten in der Modellbildung, bereits in formaler Hinsicht nicht in der Lage, den Einfluss von Corona-Maßnahmen auf das Infektionsgeschehen nachzuweisen.
- Aufgrund von gravierenden methodischen Problemen im Modellansatz muss davon ausgegangen werden, dass postulierte Effekte mehrheitlich nicht replizierbare Scheineffekte sind, die durch die Modellkonzeption entstanden sind.
 - Dies betrifft unter anderem die Aufnahme von Modellparametern, die weder theoretisch noch empirisch begründet, sondern von den Modellieren händisch eingefügt werden. Diese Parameter beeinflussen aber das Modellergebnis in entscheidender Weise.
 - Die gesamte Spezifikation des Modells ist so aufgebaut, dass ein senkender Einfluss von Corona-Maßnahmen und der Impfung bereits stillschweigend vorausgesetzt wird.
 - Zudem sind mehrere Annahmen bzw. Gütekriterien von Regressionsmodellen, wie sie in der Studie verwendet werden, nicht beachtet bzw. ihre Beachtung zumindest nicht dokumentiert worden.
- Aufgrund der Nutzung von nicht repräsentativ erhobenen Daten kommt es durch das vorhandene Test- und Meldeverhalten, welche sich in den gemeldeten Fallzahlen und damit im Nowcast und dem verwendeten R_t ("R-Wert") widerspiegeln, ohne Korrektur der Effekte zwangsläufig zu Fehlinterpretationen.
- Aufgrund von unzureichender Synchronisation der Zeitreihen von Ausbreitung und Maßnahmen sind Aussagen zur zeitlichen Abfolge von Effekten nach Eingriffen fragwürdig. Hier spielt, unter anderem, bereits die problematische Berechnung der abhängigen Variable im Modell - der "R-Wert"- eine wichtige Rolle.
- Unabhängig von methodischen Problemen sind auch die inhaltlichen Schlussfolgerungen aus den formalen Modellergebnissen nicht (oder zumindest nicht annähernd vollständig) durch diese gedeckt.
 - Insbesondere wird die Erkenntnis, dass bei mehreren Maßnahmen die Wirkung vor dem Eingriff eintrat, nicht aufgearbeitet, sondern mit der unbelegten Behauptung, dass die Bevölkerung die Maßnahmen antizipiert habe, wegeargumentiert; ein Nachweis hierfür wird nicht erbracht - und es lassen sich sogar Gegenbeispiele für diese Aussage finden. Damit muss auch der Versuch, die zeitliche Abfolge (*post hoc*) statistisch sauber zu etablieren, als gescheitert angesehen werden.
 - Auch ist die Schlussfolgerung, dass Corona-Maßnahmen erst im Verbund mit anderen Maßnahmen eine relevante Wirkung entfalten, theoretisch nicht adäquat begründet und empirisch nicht nachgewiesen, da das Studiendesign nicht wechselwirkende Effekte annimmt und solche Rückschlüsse

konstruktionsbedingt gar nicht erlaubt. Ähnlich gilt das auch für die postulierte Impfwirkung.

- Die Motivation der konzeptionellen Entscheidungen und Dokumentation der Methoden und Ergebnisse ist durchgehend ungenügend. Ergebnisse und Schlussfolgerungen sind nicht nachvollziehbar.

Aufgrund dieser substantiellen Mängel stellen wir fest, dass bei der Studiendurchführung grundlegende wissenschaftliche und fachliche Prinzipien - insbesondere in der Modellbildung - nicht eingehalten wurden und sich somit keine belastbaren Aussagen zur Wirkung von Corona-Maßnahmen ableiten lassen. Wir empfehlen daher eine Neubearbeitung der Studie auf der Basis der bestehenden Daten (Re-Analyse) durch eine unabhängige Instanz. Darüber hinaus sollten im Sinne guter wissenschaftlicher Praxis (insb. im Hinblick auf Reproduzierbarkeit) die zu Grunde gelegten Daten sowie der Quellcode der Auswertung veröffentlicht werden.

5 Literatur

Ali MM, Sharma SC (1996): Robustness to nonnormality of regression F-tests. Journal of econometrics. Vol. 71. No. 1-2.

Berlemann M, Haustein E (2023) Right and yet wrong: a politico-economic perspective on Germany's early COVID-19 policy. Regional Studies..

Boudou M, ÓhAiseadha C, Garvey P, O'Dwyer J, Hynds P (2021) Breakpoint modelling of temporal associations between non-pharmaceutical interventions and symptomatic COVID-19 incidence in the Republic of Ireland. PLoS ONE 16(7): e0255254..

Bourdin S, Jeanne L, Nadou F. & Noiret, G. (2021). Does lockdown work? A spatial analysis of the spread and concentration of Covid-19 in Italy. Regional Studies 55 (7): 1182-1193.

Chowell G, Viboud C, Hyman JM et al. (2015) The Western Africa ebola virus disease epidemic exhibits both global exponential and local polynomial growth rates. PLoS Currents 21(7): ecurrents.outbreaks.8b55f4bad99ac5c5db3663e916803261.

Driscoll JC, Kraay AC (1998) Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. The Review of Economics and Statistics 80 (4): 549-560.

de Haas S, Götz G, Heim S (2022) Measuring the effect of COVID-19-related night curfews in a bundled intervention within Germany. Nature Scientific Reports 12: 19732.

Eckey HF, Kosfeld R, Türck M (2006) Räumliche Ökonometrie. WiSt 35 (10): 548-554.

Elhorst JP (2014) Spatial Econometrics. From Cross-Sectional Data to Spatial Panels. Berlin [u.a.]: Springer.

Frankfurter Allgemeine Zeitung (2020) Der Sturm vor der Ruhe. Artikel vom 14.12.2020.

Fraser C (2007) Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic. PLoS ONE 2(8), e758.

Funk M, Wüthrich F, de Luze R, Jenal L, Rutz S, Schelker M (2022) Wirksamkeit von Corona-Massnahmen in der Schweiz - Empirische Analyse basierend auf Daten der zweiten Welle. Grundlagen für die Wirtschaftspolitik Nr. 36. Staatssekretariat für Wirtschaft SECO, Bern, Schweiz.

Gibson J (2022) Hard, not early: putting the New Zealand Covid-19 response in context. New Zealand Economic Papers 56(1): 1-8.

Goolsbee A, Syverson C (2021) Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020. J Public Econ 193:104311.

Granger CWJ, Newbold P (1974) Spurious regressions in econometrics. Journal of econometrics. Vol. 2. No. 2.

Greene WH (2003) Econometric Analysis. New York: Pearson.

Hayashi F (2000) Econometrics. Princeton University Press.

Imbens G, Rubin DW (2015) Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction, Cambridge University Press: New York.

Jacobsen GD, Jacobsen KH (2020) Statewide COVID-19 stay-at-home orders and population mobility in the United States. World Med Health Policy 1:1.

Kamvar ZN, Cai J, Pulliam JRC et al. (2019) Epidemic curves made easy using the R package incidence. F1000 Research 8: 139.

Küchenhoff H, Günther F, Höhle M, Bender A (2021) Analysis of the early COVID-19 epidemic curve in Germany by regression models with change points. Epidemiology & Infection 149: E68.

Levelu A, Sandkamp A (2022) A lockdown a day keeps the doctor away: The effectiveness of non-pharmaceutical interventions during the Covid-19 pandemic. Kiel Working Paper 2221.

Li MY (2018) An Introduction to Mathematical Modeling of Infectious Diseases. Springer: Edmonton.

Lütkepohl H, Krätzig M (2004) Applied Time Series Econometrics. Cambridge University Press.

Mader S, Rüttenauer T (2022) The Effects of Non-pharmaceutical Interventions on COVID-19 Mortality: A Generalized Synthetic Control Approach Across 169 Countries. Frontiers in Public Health 10: 820642.

Mercer GN, Glass K, Becker NG (2011) Effective reproduction numbers are commonly overestimated early in a disease outbreak. Statistics in Medicine 30(9): 984-994.

Mittelhammer RC (2013) Mathematical Statistics for Economics and Business. Springer: New York.

Nishiura H, Chowell G (2009) The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends. In: Chowell G, Hyman JM, Bettencourt LMA, Castillo-Chavez C (Hrsg.) Mathematical and Statistical Estimation Approaches in Epidemiology. Springer: Heidelberg [u.a.]. S. 103-121.

Nordström P, Ballin M, Nordström A (2022) Risk of infection, hospitalisation, and death up to 9 months after a second dose of COVID-19 vaccine: a retrospective, total population cohort study in Sweden. Lancet 399(10327):814-823.

Prada JP, Maag LE, Siegmund L et al. (2022) Estimation of R0 for the spread of SARS-CoV-2 in Germany from excess mortality. Nature Scientific Reports 12: 17221.

RKI (2020) Schätzung der aktuellen Entwicklung der SARS-CoV-2- Epidemie in Deutschland – Nowcasting. Epidemiologisches Bulletin. 17/2020.

RKI (2021a) Täglicher Lagebericht des RKI zur Coronavirus-Krankheit-2019 (COVID-19) vom 03.03.2021.

RKI (2021b) Täglicher Lagebericht des RKI zur Coronavirus-Krankheit-2019 (COVID-19) vom 28.01.2021.

RKI (2022a) Wöchentlicher Lagebericht des RKI zur Coronavirus-Krankheit-2019 (COVID-19) 05.05.2022.

RKI (2023a) Nowcasting und R-Schätzung: Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland.

RKI (2023b) Abschlussbericht: Wirksamkeit und Wirkung von anti-epidemischen Maßnahmen auf die COVID-19-Pandemie in Deutschland (StopptCOVIDStudie).

Sachverständigenausschuss nach § 5 Absatz 9 Infektionsschutzgesetz - Bundesgesundheitsministerium (2022) Evaluation der Rechtsgrundlagen und Maßnahmen der Pandemiepolitik, Bericht des Sachverständigenausschusses nach § 5 ABS. 9 IFSG.

Saffary T, Adegboye OA, Gayawan E, Elfaki F, Kuddus MA, Saffary R (2020) Analysis of COVID-19 Cases' Spatial Dependence in US Counties Reveals Health Inequalities. Frontiers in Public Health 8: 579190.

Sharma M, Mindermann S, Rogers-Smith C et al. (2021) Understanding the effectiveness of government interventions against the resurgence of COVID-19 in Europe. Nature Communications 12: 5820.

Santamaría L, Hortal J (2020) Chasing the ghost of infection past: Identifying thresholds of change during the COVID-19 infection in Spain. Epidemiology & Infection 148: E282..

Stein RA (2011) Super-spreaders in infectious diseases. International Journal of Infectious Diseases 15(8): e510-e513.

Tagesschau.de ([2023](#)): Corona-Schutzmaßnahmen waren laut RKI wirksam. Online-Artikel vom 20.07.2023.

Thomas LJ, Huang P, Yin F. et al. ([2020](#)) Spatial heterogeneity can lead to substantial local variations in COVID-19 timing and severity. Proceedings of the National Academy of Sciences of the United States of America 117(39): 24180-24187.

Wang Q, Dong W, Yang K, Ren Z, Huang D, Zhang P, Wang J ([2021](#)) Temporal and spatial analysis of COVID-19 transmission in China and its influencing factors. International Journal of Infectious Diseases 105: 675-685.

Wieland T ([2020](#)) Flatten the Curve! Modeling SARS-CoV-2/COVID-19 Growth in Germany at the County Level. REGION 7(2): 43-83.

Wood SN ([2022](#)) Inferring UK COVID-19 fatal infection trajectories from daily mortality data: Were infections already in decline before the UK lockdowns? Biometrics 78(3): 1127-1140.

Wooldridge J (2001) Econometric Analysis of Cross Section and Panel Data. The MIT Press: Cambridge, Massachusetts.

Autoren

Dr. Werner Baumgarten, Dipl.-Phys.; Fachliche Schwerpunkte: Biophysik, Nichtlinearität und Strukturbildung.

Dr. Oliver Beige, Ph.D. (UC Berkeley); Fachliche Schwerpunkte: Dynamische Ausbreitungsprozesse in Netzwerken, Ökonometrie, Kausalinferenz.

Daniel Haake, M.Sc. in Data Science, Senior Data Scientist; Fachliche Schwerpunkte: Prognosemodelle, Datenanalyse, Datenqualität.

Dr. Johannes Merkl, M.Sc. Betriebswirtschaft, Fachliche Schwerpunkte: Ökonometrie, Risikoquantifizierung, Inferenzstatistik, Zeitreihen.

Dr. Thomas Wieland, Dipl.-Geogr.; Fachliche Schwerpunkte: Wirtschaftsgeographie, medizinische Geographie, Ökonometrie (insb. räumliche Modelle), GIS.

Version 1.2. Erweitert am 02.09.2023.

E-Mail-Kontakt Autoren: geowieland@googlemail.com