Dr. rer. nat. Dipl.-Geogr. Thomas Wieland

Natürliche Experimente zur Beurteilung der Effektivität von Covid-19-Eindämmungsmaßnahmen

Methodische Grundlagen, Umsetzung in NPI-Studien und ein Fallbeispiel

Zusammenfassung

Um die Effektivität von nichtpharmazeutischen Interventionen (NPI) zur Eindämmung von Infektionen während der Covid-19-Pandemie zu überprüfen, sind natürliche Experimente hilfreich. Dieser Beitrag stellt das Prinzip natürlicher Experimente und ihrer statistischen Analyse vor und diskutiert im ersten Schritt deren Umsetzung in bisherigen NPI-Studien. Die Aussagekraft vieler Studien zur Wirksamkeit von NPI ist eingeschränkt, weil häufig keine oder nur schwache Kontrollbedingungen vorliegen und weil es in vielen Fällen auch Probleme in der Spezifikation der Analysemodelle gibt. Oft wird implizit davon ausgegangen, dass nur NPI das Infektionsgeschehen entschleunigen können, weil Kontrollvariablen für etwaige andere Ursachen fehlen. Im zweiten Schritt wird die Analyse eines natürlichen Experimentes vorgenommen, und zwar im Hinblick auf die Wirksamkeit von fünf Maßnahmen in der zweiten Corona-Welle. Als Vergleichsregionen fungieren die Schweiz (wo diese Maßnahmen nicht getroffen wurden) und das angrenzende Baden-Württemberg. Es zeigt sich hierbei, dass die Schließung von körpernahen Dienstleistungen eine Senkung der effektiven Reproduktionszahl verursacht, also die Infektionswelle dämpft. Die Schließung von Schulen und Kitas sowie von Hotels ist hingegen nicht effektiv.

Schlüsselwörter

Nichtpharmazeutische Interventionen, Covid-19-Pandemie, Natürliche Experimente, Ökonometrie, Differenz-von-Differenzen-Ansatz

Crossref/doi

http://doi.org/10.24945/MVF.01.24.1866-0533.2580

Nahezu jeder Staat weltweit hat in der Corona-Pandemie Maßnahmen zur Eindämmung von Infektionen (Nichtpharmazeutische Interventionen, NPI) ergriffen. In vielen Ländern wurden umfassende Maßnahmenpakete (Lockdowns) implementiert. Da die meisten dieser Maßnahmen auch negative Effekte haben (z. B. Lerndefizite durch Schulschließungen, Einschränkungen individueller Freiheitsrechte, wirtschaftliche Einbußen), ist die Beurteilung ihrer tatsächlichen Wirksamkeit – also ihres Einflusses auf das gesamte Infektionsgeschehen – von hoher Bedeutung. Die Frage eines kausalen Schlusses ist hier aber keinesfalls trivial. Als „Goldstandard" der Evidenz wird in der Medizin häufig die randomisierte kontrollierte Studie (randomized controlled trial, RCT) angesehen. Der (statistische) Schluss auf kausale Einflüsse von Interventionen wird dabei gewährleistet durch das Vorhandensein einer Kontrollgruppe (Kontrollbedingungen), die mit der Interventionsgruppe vergleichbar ist (Strukturgleichheit), sowie Randomisierung und, in vielen Fällen, auch Verblindung (Kabisch et al. 2011: 663 ff.). Unabhängig davon, dass es bereits in „echten" Experimenten herausfordernd ist, alle genannten Qualitätsmerkmale zu erfüllen, fällt schnell auf, dass sich dieses Prinzip der Evidenzfindung nicht direkt auf die Beurteilung der Effektivität von Pandemie-Eindämmungsmaßnahmen übertragen lässt. Die Wirksamkeit von NPI – die ja politische Eingriffe sind – kann aber mithilfe von sog. natürlichen Experimenten überprüft werden, die mit ökonometrischen Analyseverfahren ausgewertet werden. Dieser Beitrag skizziert zunächst das methodische Prinzip dahinter und diskutiert dann in einem ersten Schritt die Probleme in der Umsetzung bestehender NPI-Studien. Besondere Bedeutung erfahren hierbei einerseits die notwendigen Kontrollbedingungen, um kausale Schlüsse im Hinblick auf die Effektivität von Maßnahmen ziehen zu können, und andererseits die Spezifikation der Analysemodelle in diesen Studien. Im zweiten Schritt werden die Ergebnisse eines natürlichen Experiments präsentiert, und zwar im Hinblick auf fünf NPI, die in der zweiten Corona-Welle in Deutschland implementiert wurden; als Kontrollgruppe dient hierbei die Schweiz, wo auf diese Maßnahmen – darunter Schul- und Kitaschließungen – verzichtet wurde.

Methodische Probleme modellbasierter Studien zur Beurteilung der Effektivität von NPI

Natürliche Experimente und ihre statistische Analyse in der Beurteilung der Effektivität von NPI

>> NPI sind, wie z. B. die Einführung eines Mindestlohns, eine temporäre Mehrwertsteuerermäßigung für einen Wirtschaftszweig oder das Verbot eines bestimmten Unkrautvernichtungsmittels, politische Interventionen. Bei solchen Eingriffen liegt keine Randomisierung vor, da eine bestimmte Maßnahme nicht in zufällig gewählten Teilgebieten eingeführt wird. Auch eine Verblindung ist nicht möglich, da die Angehörigen der Interventionsgruppe wissen, ob die Maßnahme besteht oder nicht. Vorausgesetzt, dass ein bestimmter Eingriff in einem Gebiet vollzogen wird und in einem anderen Gebiet nicht, besteht aber, wie in einem „echten" Experiment, eine Interventions- und eine Kontrollgruppe; man spricht in so einem Fall von einem natürlichen Experiment (oder Quasi-Experiment). Es ist ebenso eine Vorher-Nachher-Untersuchung mit Kontrollgruppe, die einen kausalen Schluss auf die Effektivität einer Intervention erlaubt (Kugler et al. 2014: 114 ff.; Wooldridge 2012: 457). Es bestehen jedoch keine Labor-, sondern Reale-Welt-Bedingungen. Die Varianz der unabhängigen Variablen wird, anders als in einem „echten" Experiment, nicht durch die Forscher:innen gesteuert, sodass auch nicht von einem Feldexperiment gesprochen werden kann, sondern von einer Feldstudie (Eifler/Leitgöb 2019: 210 ff.).

Unter der Voraussetzung, dass Daten zum Outcome für dieselben Analyseeinheiten für mindestens zwei Zeitpunkte – also Paneldaten – vorliegen, bietet die Ökonometrie verschiedene Verfahren zur statistischen Analyse des Interventionseffektes an (Kugler et al. 2014: 114 ff.). Prominent ist vor allem der Differenz-von-Differenzen-Ansatz (Difference-in-Differences approach, DiD), der auf einem linearen Regressionsmodell basiert. Die Ableitung der Kausalität von Ergebnissen ergibt sich, neben der Tatsache, dass

Kontrollbedingungen bestehen, daraus, dass über die Spezifikation der unabhängigen Variablen die Untersuchungseinheiten der Interventions- bzw. Kontrollgruppe zugeordnet werden. Es ist nicht erforderlich, dass beide Gruppen dasselbe Ausgangsniveau haben; der DiD-Ansatz ist sogar speziell für den Fall gedacht, dass die Interventions- und die Kontrollgruppe vor Beginn der Intervention unterschiedliche Niveaus im interessierenden Outcome haben. Der Ansatz vergleicht, wie sich der Outcome in der Behandlungsgruppe ohne Intervention entwickelt hätte (counterfactual) mit der Entwicklung ihres tatsächlichen Outcomes mit Intervention. Als Anhaltspunkte dienen hierfür die Entwicklung aller Untersuchungseinheiten vor Beginn der Intervention und die Entwicklung der Kontrollgruppe ab dem Zeitpunkt der Intervention. Die Grundannahme ist, dass sich die Interventions- und die Kontrollgruppe ohne Intervention gleich (bzw. nicht statistisch signifikant verschieden) entwickelt hätten (Kugler et al. 2014: 120 ff.; Wooldridge 2012: 454 ff.). Der DiD-Ansatz lässt sich durch feste Effekte (fixed effects) für die Untersuchungseinheiten und/oder die Zeitpunkte erweitern, was eine Verallgemeinerung des Ansatzes erlaubt (Goodman-Bacon 2021: 255 ff.).

Im Hinblick auf die vielzähligen Studien zur Effektivität von Corona-NPI lässt sich zunächst feststellen, dass meist kein kausaler Modellansatz im o. g. Sinne verwendet wird (Ausnahmen z. B. de Haas et al. 2022; Kosfeld et al. 2021; von Bismarck-Osten et al. 2022). Dies ist weniger problematisch, denn dieser Umstand ließe sich durch eine andere Spezifikation der jeweiligen Analysemodelle ausgleichen. Viel schwerwiegender ist allerdings, dass häufig die o. g. Ausgangsbedingungen für ein natürliches Experiment – nämlich die Existenz einer Interventions- *und* einer Kontrollgruppe – gar nicht vorliegen.

Die Mehrheit der NPI-Studien bezieht sich auf die erste Corona-Welle (z. B. Banholzer et al. 2021; Brauner et al. 2021; Flaxman et al. 2020; Haug et al. 2020), was ein Problem hinsichtlich ihrer Aussagekraft mit sich bringt: Im Frühjahr 2020 fielen die Eindämmungsmaßnahmen international sehr ähnlich aus, sodass nur eine geringe Varianz zwischen den verglichenen Ländern besteht. Eine stärkere Ausdifferenzierung erfolgte erst ab der zweiten Corona-Welle (Herbst/Winter 2020/2021), wofür Deutschland und seine Nachbarländer ein gutes Beispiel sind (z. B. Schulschließungen in Deutschland und Österreich, aber nicht in Frankreich und der Schweiz). Das Problem geringer Varianz zwischen den Untersuchungseinheiten trifft auch dann zu, wenn nur die Teilräume eines einzelnen Landes in die Analyse einfließen und innerhalb des Landes nur geringe Unterschiede in der Implementierung von Maßnahmen bestanden. Dieser Fall findet sich sowohl in begutachteten wissenschaftlichen Arbeiten, die ansonsten sicher nicht zu beanstanden sind (z. B. Kosfeld et al. 2021; Khazaei et al. 2023), als auch in öffentlich geförderten NPI-Studien wie der Schweizer Maßnahmenevaluation (Funk et al. 2022) und der „StopptCOVID"-Studie des Robert Koch-Institutes zu den deutschen Maßnahmen (an der Heiden et al. 2023). Die Ableitung kausaler Effekte von Maßnahmen ist nicht möglich, wenn es aufgrund mangelnder Varianz zwischen den Untersuchungseinheiten keine Kontrollbedingungen gibt, d. h. wenn dieselben Maßnahmen (nahezu) zeitgleich im gesamten Untersuchungsgebiet implementiert wurden.

Die Kontrollbedingungen werden auch dadurch geschwächt, dass in vielen Studien einzelne Maßnahmen aggregiert werden, die zwischen den untersuchten Ländern bzw. Regionen eigentlich variieren. Häufig wird z. B. die Schließung von Grund- und weiterführenden Schulen sowie Hochschulen zu einer NPI zusammengeführt und dann für dieses Aggregat ein Effekt geschätzt (z. B. Brauner et al. 2021; Gavenciak et al. 2022; Haug et al. 2020; Sharma et al. 2021). Hier besteht das Problem, dass unterschiedliche Schultypen und Universitäten keinesfalls gleichbehandelt wurden; in vielen Ländern wurden ab der zweiten Corona-Welle zwar Hochschulen auf Fernunterricht umgestellt, Schulen aber offengelassen (z. B. Frankreich, Schweiz). Andere Studien aggregieren sogar die Schließungen von Schulen, Universitäten und Kitas in eine NPI (z. B. Kosfeld et al. 2021), wodurch eine weitere Vergleichsdimension verloren geht. In den genannten Fällen lässt sich z. B. ein (möglicher) Effekt der Schließung eines einzelnen Schultyps nicht mehr von einem (möglichen) Effekt der Schließung anderer Schultypen oder Hochschulen trennen.

Die Kontrollbedingungen werden weiterhin durch die Wahl der räumlichen Analyseebene und des Untersuchungszeitraumes beeinflusst: Häufig wird eine großräumige Analyseebene gewählt, v. a. ganze Länder (z. B. Banholzer et al. 2021; Brauner et al. 2021; Flaxman et al. 2020; Haug et al. 2020; Gavenciak et al. 2022), was regionale Unterschiede im Infektionsgeschehen und in der Ausgestaltung von Pandemiemaßnahmen ausblendet. Beispielsweise wurden einige NPI in Deutschland auf der Ebene von Landkreisen implementiert (z. B. Ausgangssperren); auch die spätere „Bundesnotbremse" war kreisspezifisch. Zwei Studien, die ansonsten Kontrollbedingungen aufweisen, beschränken ihre Daten auf das Jahr 2020 und beenden damit ihre Analyse inmitten der zweiten Corona-Welle (Levelu/Sandkamp 2023; Sharma et al. 2021). Hierbei sind Effekte schwierig abzuleiten, weil nur ein (verhältnismäßig kleiner) Teil der gesamten Interventionsperiode in die Analyse einfließt.

Probleme in der Modellspezifikation bei NPI-Studien

Nahezu alle empirischen NPI-Studien basieren auf Regressionsanalysen mit Maßnahmen als unabhängigen Variablen. Als abhängige Variable im Regressionsmodell fungiert meist die effektive Reproduktionszahl (z. B. an der Heiden et al. 2023; Brauner et al. 2021; Flaxman et al. 2020; Gavenciak et al. 2022; Khazaei et al. 2023; Levelu/Sandkamp 2023; Li et al. 2021; Sharma et al. 2021). In einem solchen Modell wird die Wirkung einer einzelnen erklärenden Variablen isoliert betrachtet, also unter ansonsten konstant gehaltenen Bedingungen (ceteris paribus). Der empirisch geschätzte Einfluss einer Variablen ist daher, sinngemäß, nur „gültig" in Kombination mit den anderen erklärenden Variablen im Modell. Um ihn verzerrungsfrei zu schätzen, müssen deshalb auch andere Variablen ins Modell aufgenommen werden, die zwar nicht von Interesse sind, aber möglicherweise die abhängige Variable beeinflussen. Das ist das Prinzip von Kontrollvariablen, die in solche Modelle eingebunden werden, um ihren Einfluss „herauszurechnen" (Hilmert 2019: 4 ff.; Wooldridge 2012: 97). Dieser Aspekt der Modellspezifikation lässt sich verallgemeinern: Ein Regressionsmodell, in dem Variablen fehlen, die eigentlich die abhängige Variable beeinflussen, ist unterspezifiziert. In der Folge werden auch die Einflüsse der im Modell inkludierten erklärenden Variablen ggf. über- oder

unterschätzt; man spricht hier vom omitted variable bias (Greene 2012: 96 ff.; Wooldridge 2012: 88 ff.).

Ein Blick in die Analysemethodik bisheriger NPI-Literatur zeigt, dass viele der Studien mit der effektiven Reproduktionszahl als abhängiger Variable *ausschließlich* NPI als unabhängige Variablen in ihre Regressionsmodelle integrieren (z. B. Brauner et al. 2021; Flaxman et al. 2020; Li et al. 2021; Sharma et al. 2021). Beispielhaft hierfür seien Gavenciak et al. (2022: 4) mit Bezug auf die Arbeiten von Brauner et al. (2021) und Sharma et al. (2021) zitiert: „Both models assume the time-varying reproduction number R(t) to be a product of R0, the 'natural' reproduction number when no NPIs are in place, multiple terms representing the effects of interventions, and noise terms modelling the influence of other, unobserved, factors on R(t)." In so einem Modell sind demnach keine anderen unabhängigen Variablen vorgesehen, die einen Einfluss auf den zeitlichen Verlauf von Rt haben könnten. Implizit wird somit davon ausgegangen, dass die effektive Reproduktionszahl konstant ist, solange keine Corona-Maßnahmen in Kraft sind.

Der Zeitverlauf von Rt hängt aber, wie in der epidemiologischen Literatur hinreichend dokumentiert wird, nicht nur von Eindämmungsmaßnahmen ab, sondern u. a. auch von folgenden Aspekten:

- Abnahme suszeptibler Bevölkerung: Eine betroffene Person ist nach der eigenen Infektion – zumindest für einen gewissen Zeitraum – immunisiert und kann daher (vorübergehend) weder selbst neu infiziert werden noch eine andere Person infizieren; dasselbe gilt für eine Impfung (Mercer et al. 2011: 984 f.; Nishiura/Chowell 2009: 104)
- Populationsheterogenität: Nicht jedes Individuum hat, auch aufgrund sehr unterschiedlicher Kontaktmuster, dieselbe Wahrscheinlichkeit, sich selbst bzw. andere zu infizieren (Müller 2021: 70 f.; Viboud et al. 2016: 31 ff.)
- Freiwillige Verhaltensänderungen: Individuelle Senkung von Infektionsrisiken unabhängig von Maßnahmen (Mercer et al. 2011: 984 f.; Nishiura/Chowell 2009: 104; Viboud et al. 2016: 31 ff.), was auch in der Corona-Pandemie z. B. bezüglich des Mobilitätsverhaltens nachgewiesen wurde (z. B. Goolsbee/Syverson 2021; Jacobsen/Jacobsen 2020).
- Saisonalität: SARS-CoV-2 folgt, wie andere respiratorische Viruserkrankungen (z. B. Influenza, RSV, andere Coronaviren), in den gemäßigten Klimazonen einem saisonalen Zyklus mit dem Schwerpunkt auf Herbst und Winter (Dbouk/Drikakis 2021; Moriyama et al. 2020: 84 ff.).

Natürlich ist es unmöglich, in einem statistischen Modell sämtliche realen Erklärungstatbestände zu berücksichtigen. Die Modelle der o. g. Studien können aber nicht vollständig spezifiziert sein, was die Interpretation der Ergebnisse bezüglich der NPI-Variablen erschwert (omitted variable bias, s. o.). Denn wenn *nur* Maßnahmen als erklärende Variablen im Modell vorgesehen sind, wird zwingend jede Veränderung im R-Wert der Wirkung dieser Maßnahmen zugeschrieben, auch wenn sie aus anderen Gründen erfolgte. Einzelne spätere NPI-Studien haben ihre Analysemodelle um bestimmte o. g. Teilaspekte erweitert, z. B. durch Modellterme für saisonale Schwankungen (an der Heiden et al. 2023; Gavenciak et al. 2022), die Wirkung der Covid-19-Impfungen auf den R-Wert (an der Heiden et al. 2023) oder (mögliche) freiwillige Verhaltensänderungen (Levelu/Sandkamp 2023); allerdings finden sich auch in diesen Modellen keine Variablen, die, von diesen Teilaspekten abgesehen, eine Veränderung der effektiven Reproduktionszahl herbeiführen könnten. Einzig die im November 2023 erschienene Arbeit von Khazaei et al. (2023) berücksichtigt freiwillige Verhaltensänderungen als eine Art „eigene Maßnahme" und darüber hinaus die sinkende (Wieder-)Ansteckungswahrscheinlichkeit durch vorherige Infektionen und Impfungen in der Modellgleichung.

Ein weiterer Aspekt im Hinblick auf die Berücksichtigung von Kontrollvariablen ist, dass viele Studien nur einen Teil der implementierten Corona-Maßnahmen in die Regressionsmodelle aufnehmen. Beispielsweise berücksichtigen die o.g. Studien, die die Schließung von Schulen und Universitäten zu einer NPI zusammenfassen, gar nicht die Schließung von Kindertageseinrichtungen oder weisen zumindest nicht darauf hin, ob diese Maßnahme dem genannten Aggregat zugerechnet wird (Brauner et al. 2021; Gavenciak et al. 2022; Sharma et al. 2021). Khazaei et al. (2023) treffen z. B. hingegen eine bewusste Auswahl von acht NPI (u. a. Schulschließungen, Ausgangssperre, Personenbeschränkungen für Veranstaltungen). Völlig unabhängig davon, ob diese Auslassungen einer bewussten Auswahl oder mangelnden Daten geschuldet sind, besteht hierbei ein ähnliches Problem wie bei anderen denkbaren Kontrollvariablen: Der im Modell gefundene Effekt einer spezifischen Maßnahme kann über- oder unterschätzt werden, weil nicht alle anderen Maßnahmen im Modell berücksichtigt werden. Es ist z. B. möglich, dass ein gefundener Effekt in Wirklichkeit auf andere Maßnahmen zurückzuführen ist, die im Modell nicht auftauchen.

Fallbeispiel: Baden-Württemberg und Schweiz in der zweiten Corona-Welle

In der zweiten Corona-Welle hatte sich die Intensität staatlicher Eindämmungsmaßnahmen bereits deutlich zwischen den Ländern ausdifferenziert, sodass Wirksamkeitsanalysen anhand natürlicher Experimente vielerorts möglich sind. Das zeigt sich u. a. am direkten Vergleich Deutschlands mit der benachbarten Schweiz: In beiden Ländern wurden zwar (zeitversetzt) Maßnahmen unterschiedlicher Intensität etabliert, jedoch verzichtete die Schweiz, im Gegensatz zu Deutschland, auf die Schließung von Kinderbetreuungseinrichtungen, Grund- und Sekundarschulen, Hotels und körpernahen Dienstleistungen (z. B. Friseuren). Dieser Unterschied bildet eine gute Grundlage für ein natürliches Experiment, da in Bezug auf diese Interventionen No-Treatment-Kontrollbedingungen vorherrschen. Daher wird hier der Einfluss dieser fünf NPI analysiert, und zwar in Bezug auf die Schweiz und das angrenzende Baden-Württemberg. Die räumliche Analyseebene ist die NUTS3-Ebene, d. h. Stadt- und Landkreise (Baden-Württemberg) bzw. Kantone (Schweiz).

Die empirische Analyse baut auf den bisherigen NPI-Studien (s. o.) auf: Es wird ein Regressionsmodell mit der effektiven Reproduktionszahl als abhängige Variable und den Corona-Maßnahmen als unabhängige Variablen (in Form von Dummy-Variablen) aufgebaut (siehe Abb. 1). Die Modellspezifikation wird allerdings in zweierlei Hinsicht modifiziert: Erstens erfolgt die Analyse des Interventionseffektes entsprechend des Differenz-in-Differenzen-Ansatzes (DiD). In Anlehnung an die Formalisierung in Wooldridge

(2012: 457 f.) wird die Interventionsgruppe für eine spezifische Maßnahme (z. B. Schließung von Grundschulen) und der Interventionszeitraum (z. B. Tage, an denen die Grundschulen in der Interventionsgruppe nicht aufgrund von Ferien, sondern als Corona-Maßnahme geschlossen waren) mit je einer Dummy-Variable markiert und zusätzlich die Interaktion dieser beiden Variablen ins Modell aufgenommen. Der Koeffizient dieses Interaktionsterms ist der DiD-Schätzer und gibt den Interventionseffekt der spezifischen Maßnahme an (prozentuale Veränderung des R-Wertes).

Zweitens werden dem Modell Kontrollvariablen hinzugefügt, die, von Pandemiemaßnahmen unabhängig, den Verlauf von Rt beeinflussen können. Analog zu dem Modell von Levelu/Sandkamp (2023: 9 f.) wird der R-Wert zehn Tage zuvor ins Modell aufgenommen, da angenommen wird, dass Menschen möglicherweise auf erhöhte Infektionszahlen reagieren und individuelle Risikomeidung betreiben. Weiterhin werden kumulierte Zahlen der Infektionen und der Geimpften sowie deren Interaktion mit der Zeit ins Modell aufgenommen, um der (möglichen) Entschleunigung der Infektionswelle durch (temporäre) Immunität Rechnung zu tragen (s. o.). Für jede der fünf NPI wurde ein Modell geschätzt. Alle anderen Corona-

Modellkonzept

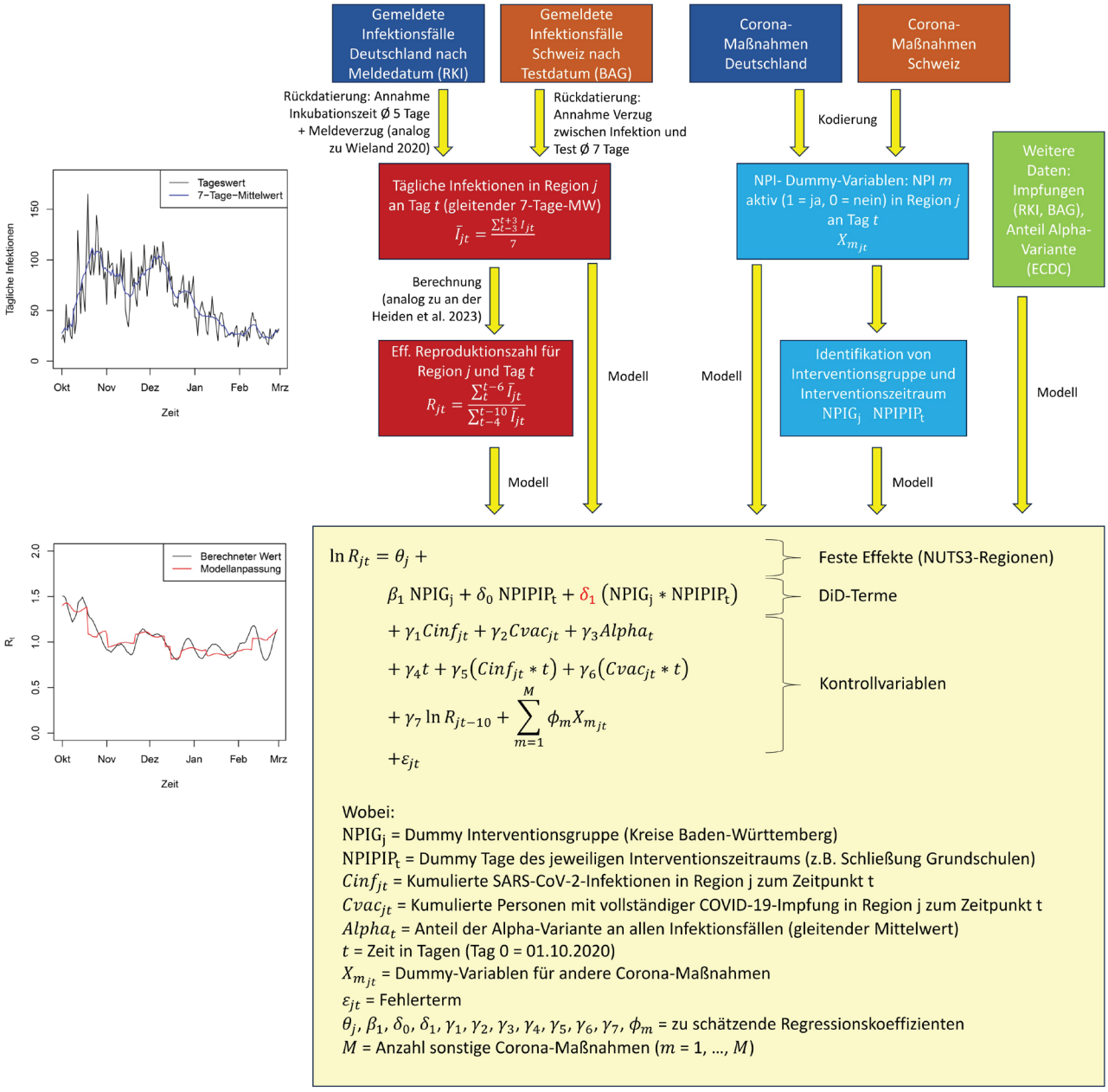


Abb. 1: Modellkonzept.

Maßnahmen gehen hierbei ebenso als Kontrollvariablen ins Modell ein, um den tatsächlichen Effekt der interessierenden Intervention isolieren zu können (Abb. 1.).

Tabelle 1 zeigt die Modellschätzungen und die zugehörigen Modellgütestatistiken. Abbildung 2 zeigt die DiD-Schätzer aus den fünf Modellen und die zugehörigen Konfidenzintervalle, die auf robusten Standardfehlern basieren. Ein signifikanter DiD-Schätzer unterhalb von null signalisiert die Effektivität einer Maßnahme in Form einer durchschnittlichen Senkung der effektiven Reproduktionszahl. Hierbei zeigt sich, dass die Schließung körpernaher Dienstleistungen eine durchschnittliche Senkung der effektiven Reproduktionszahl von ca. 1,7% verursacht. Diese Maßnahme, die in Baden-Württemberg (und ganz Deutschland) in der zweiten Corona-Welle getroffen wurde, in der Schweiz jedoch nicht, ist also als wirksam zu beurteilen. Für die anderen vier untersuchten NPI kann ein solcher Effekt nicht gefunden werden. Die DiD-Schätzer für die Schließung von Schulen, Kindertageseinrichtungen und Hotels sind sogar positiv, was einen kontraproduktiven Effekt andeutet. Das kann nicht darauf zurückzuführen sein, dass das Modelldesign eine Ablehnung der Nullhypothese im Sinne statistischer Signifikanz erschwert, denn einerseits sind auch diese Punktschätzer signifikant (nur eben mit einem anderen als dem erwarteten Vorzeichen), und andererseits ist der theoretisch erwartete Effekt ja für eine andere untersuchte NPI – Schließung von körpernahen Dienstleistungen – nachgewiesen worden (Abb. 1, Tab. 1).

Ein Blick auf die Kontrollvariablen zeigt, dass diese im über-

Modellergebnisse					
	Koeffizienten (Standardfehler)				
Unabhängige Variablen	**Modell 1: Schließung Kitas/ Kindergärten**	**Modell 2: Schließung Grundschulen**	**Modell 3: Schließung Weiterführende Schulen**	**Modell 4: Schließung Hotels**	**Modell 5: Schließung körpernahe Dienstleistungen**
DiD-Terme:					
Interventionsgruppe	0,568*** (0,034)	0,568*** (0,034)	0,557*** (0,034)	0,585*** (0,033)	0,554*** (0,032)
Interventionszeitraum	-0,048*** (0,008)	-0,080*** (0,007)	-0,110*** (0,008)	-0,053 (0,038)	-0,078*** (0,006)
Interventionsgruppe x Interventionszeitraum	0,038* (0,017)	0,072*** (0,010)	0,089*** (0,010)	0,081** (0,030)	-0,017* (0,008)
Kontrollvariablen:					
Infizierte kumuliert	-0,00001*** (0,00000)	-0,00001*** (0,00000)	-0,00001*** (0,00000)	-0,00000*** (0,00000)	-0,00001*** (0,00000)
Geimpfte kumuliert	-0,00001*** (0,00000)	-0,00001*** (0,00000)	-0,00002*** (0,00000)	-0,00002*** (0,00000)	-0,00002*** (0,00000)
Anteil Alpha-Variante	-0,001** (0,0003)	-0,001** (0,0003)	-0,0004 (0,0003)	-0,0004 (0,0003)	0,001** (0,0003)
Zeit [Tage]	0,002*** (0,0002)	0,003*** (0,0002)	0,003*** (0,0002)	0,002*** (0,0002)	0,002*** (0,0002)
Infizierte kumuliert x Zeit [Tage]	0,00000*** (0,000)	0,00000*** (0,000)	0,00000*** (0,000)	0,00000*** (0,000)	0,00000*** (0,000)
Geimpfte kumuliert x Zeit [Tage]	0,00000*** (0,000)	0,00000*** (0,00000)	0,00000*** (0,000)	0,00000*** (0,00000)	0,00000*** (0,000)
ln R_{jt-10}	-0,180*** (0,010)	-0,183*** (0,010)	-0,173*** (0,011)	-0,180*** (0,011)	-0,174*** (0,010)
Andere NPI (Anzahl)	35	35	36	35	35
Feste Effekte (Region)	ja	ja	ja	ja	ja
Beobachtungen	10.570	10.570	10.570	10.570	10.570
R^2	0,626	0,630	0,631	0,625	0,633
Adj. R^2	0,622	0,626	0,627	0,621	0,630
Resid.-Std.-fehler	0,124 (df = 10459)	0,123 (df = 10459)	0,123 (df = 10458)	0,124 (df = 10459)	0,123 (df = 10459)
F-Statistik	157,708*** (df = 111; 10459)	160,284*** (df = 111; 10459)	159,628*** (df = 112; 10458)	157,133*** (df = 111; 10459)	162,849*** (df = 111; 10459)
Legende: *p < 0,05, **p < 0,01, ***p < 0,001; Robuste Standardfehler; Interventionsgruppe = Kreise Baden-Württemberg. Abhängige Variable: ln R_{jt}					

Tab. 1: Modellergebnis.

wiegenden Fall einen statistisch signifikanten Einfluss auf die abhängige Variable haben. Es zeigt sich z. B., dass der R-Wert mit steigenden Zahlen sowohl von Infizierten als auch von Geimpften signifikant sinkt, was den epidemiologischen Grundlagen in Bezug auf die effektive Reproduktionszahl entspricht (siehe z. B. Mercer et al. 2011: 984 f.; Nishiura/Chowell 2009: 104). Dass zusätzlich auch der zehn Tage frühere R-Wert einen signifikant senkenden Einfluss hat, kann darauf hinweisen, dass es freiwillige Verhaltensänderungen gab in dem Sinne, dass Menschen auf erhöhte Infektionszahlen reagiert und Infektionsrisiken gemieden haben. Inwiefern dies tatsächlich passiert ist und welche Auswirkungen diese Verhaltensänderungen auf die Infektionswelle hatten, kann aber anhand von diesem statistischen Ergebnis nicht festgemacht werden. Überhaupt sollten die Kontrollvariablen inhaltlich nicht überinterpretiert werden. Es bleibt aber festzuhalten, dass sie in ein solches Modell zur Analyse der Effektivität von NPI notwendigerweise integriert werden müssen; ohne sie wäre das Modell unterspezifiziert.

Fazit

Die Beurteilung der Effektivität von Pandemiemaßnahmen ist, anders als die Wirkung eines Arzneimittels oder einer Therapie, keine medizinische Fragestellung. NPI sind politische Interventionen, deren Wirksamkeit sich genauso überprüfen lässt wie die anderer staatlicher Eingriffe. Voraussetzung hierfür ist, dass die Ausgangsbedingungen eines natürlichen Experimentes bestehen, also dass es eine Interventions- und eine Kontrollgruppe gibt, die verglichen werden können, und dass Paneldaten für die Untersuchungseinheiten vorliegen. Sind diese Voraussetzungen erfüllt, bietet die Ökonometrie mit dem Differenz-von-Differenzen-Ansatz eine etablierte Analysemethodik an, die etwaige Ausgangsunterschiede zwischen den Untersuchungseinheiten der beiden Gruppen abbilden und zugleich andere Einflüsse in Form von Kontrollvariablen berücksichtigen kann.

Die Literatur zu Corona-Maßnahmen – wobei hier lediglich repräsentative Beispiele genannt werden können, und keine vollständige Literaturliste dargeboten werden kann – zeigt, dass viele Studien diese Anforderungen nicht vollständig erfüllen und dass deren Ergebnisse unter Berücksichtigung dieser Limitationen interpretiert werden müssen. Keinesfalls kann auf die kausale Wirkung einer Maßnahme geschlossen werden, wenn die Studien, die diesen Effekt gefunden haben, keine Kontrollbedingungen aufweisen, also keine Vergleichsregion(en) ohne diese Maßnahme in die Analyse einbezogen wurden. Auch muss zumindest die Effektstärke bei Maßnahmen hinterfragt werden, wenn Studien nur NPI als erklärende Variablen berücksichtigen bzw. wichtige Kontrollvariablen fehlen. Diese Problematik betrifft nicht die beteiligten Forscher:innen, denn die haben ihre Methodik und deren Limitationen dokumentiert, sondern vor allem die Rezeption solcher Studien in den Medien und der Politik. In jedem Fall muss die Methodik einer NPI-Studie im Hinblick auf Kontrollbedingungen und Modellspezifikation kritisch beleuchtet werden, um folgende Frage zweifelsfrei beantworten zu können: Kann die Studie eine kausale Aussage zur Wirkung einer Maßnahme treffen? Ein weiterer Aspekt, der hier nicht behandelt

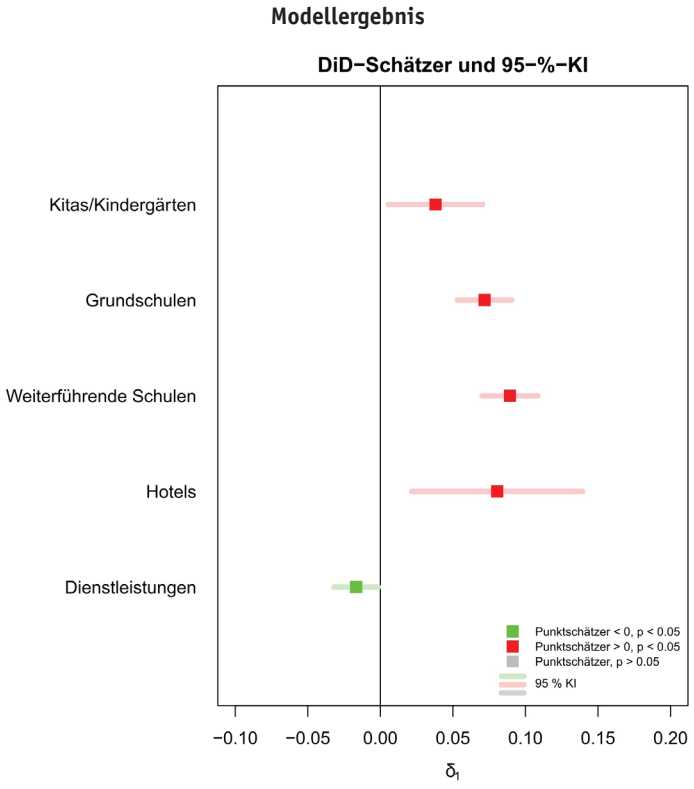


Abb. 2: Modellergebnis.

wird, ist die Effizienz bzw. Verhältnismäßigkeit einer Maßnahme, was Gegenstand politischer bzw. gesellschaftlicher Diskussionen sein sollte.

Das Fallbeispiel zeigt, dass die Schließungen von Schulen, Kinderbetreuungseinrichtungen und Hotels im direkten Vergleich zweier benachbarter Regionen mit „echter" Kontrollgruppe keine dämpfende Wirkung auf die effektive Reproduktionszahl haben, also ineffektiv im Sinne der Methodik bisheriger NPI-Studien sind.

Dies offenbart einerseits die Wichtigkeit natürlicher Experimente und ihrer adäquaten statistischen Analyse und andererseits die Notwendigkeit von Kontrollvariablen im Modell, denn diese haben nachweisbare Einflüsse auf das Infektionsgeschehen. Ohne sie wäre das Modell fehlspezifiziert und die Effekte der untersuchten Maßnahmen würden ggf. falsch eingeschätzt werden. Der Befund ähnelt tendenziell dem in den (wenigen) Studien, die Corona-Maßnahmen mit „echten" Kontrollbedingungen und einem DiD-Ansatz untersucht haben; so wurde z. B. kein Effekt von nächtlichen Ausgangssperren (de Haas et al. 2022) und von Schulschließungen gefunden (von Bismarck-Osten et al. 2022).

Grenzen des Verfahrens

Natürlich zeigen sich hier auch die Grenzen des Verfahrens: Es war nur für fünf Maßnahmen möglich, eine kausale Analyse im Sinne eines natürlichen Experimentes durchzuführen, da nur hier „echte" Kontrollbedingungen vorlagen. Eine Aussage zu anderen NPI kann auf dieser Grundlage nicht getroffen werden.

Weiterhin muss darauf hingewiesen werden, dass ein natür-

liches Experiment wie dieses immer nur Aussagen im Hinblick auf einen durchschnittlichen Interventionseffekt, also eine Wirkung auf Populationsebene, machen kann. Die Ergebnisse spiegeln kein individuelles Risiko wider. Dass z. B. Schulschließungen keinen infektionsreduzierenden Effekt in der Gesellschaft zeigen, bedeutet natürlich nicht, dass eine Person sich nicht in einer Schule mit SARS-CoV-2 infizieren kann oder dass keinerlei Infektionen in Schulen stattgefunden hätten. Nur sind die Infektionen, die (möglicherweise) durch Schulschließungen verhindert wurden, so wenige, dass sich ihr aggregierter Effekt nicht gesamtgesellschaftlich im Infektionsgeschehen zeigt. Es sind auch Ausweicheffekte denkbar, d. h., dass Infektionen, die (möglicherweise) durch Schulschließungen verhindert wurden, sogar dadurch überkompensiert wurden, dass sich Schüler:innen und Lehrer:innen stattdessen an einem anderen Ort aufgehalten haben, an dem das Infektionsrisiko (möglicherweise) viel höher ist. <<

Literatur

an der Heiden, M./Hicketier, A./Bremer, V./Bozorgmehr, K./Costa, D./Rohleder, S. (2023): Wirksamkeit und Wirkung von anti-epidemischen Maßnahmen auf die COVID-19-Pandemie in Deutschland (StopptCOVID-Studie). Berlin: Robert Koch-Institut.

Banholzer, N./van Weenen, E./Lison, A./Cenedese, A./Seeliger, A./Kratzwald, B./Tschernutter, D./Puig Salles, J./Bottrighi, P./Lehtinen, S./Feuerriegel, S./Vach, W. (2021): Estimating the effects of non-pharmaceutical interventions on the number of new infections with COVID-19 during the first epidemic wave. In: PLoS ONE, 2021, 16, 6: e0252827.

Brauner, J. M./Mindermann, S./Sharma, M./Johnston, D./Salvatier, J./Gavenciak, T./Stephenson, A. B./Leech, G./Altman, G./Mikulik, V./Norman, A. J./Teperowski Monrad, J./Besiroglu, T./Ge, H./Hartwick, M. A./Whye Teh, Y./Chindelevitch, L./Gal, Y./Kulveit, J. (2021): Inferring the effectiveness of government interventions against COVID-19. In: Science, 2021, 371, 6531: eabd9338.

Dbouk, T./Drikakis, D. (2021): Fluid dynamics and epidemiology: Seasonality and transmission dynamics. In: Physics of Fluids 2021, 33, 2: 021901.

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

Eifler, S./Leitgöb, H. (2019): Experiment. In: Baur, N./Blasius, J (Hrsg.) (2019): Handbuch Methoden der empirischen Sozialforschung, Wiesbaden: Springer, 203-218.

Flaxman, S./Mishra, S./Gandy, A./Unwin, H. J. T./Mellan, T. A./Coupland, H./Whittaker, C./Zhu, H./Berah, T./Eaton, J. W./Monod, M./Imperial College COVID-19 Response Team/Ghani, A. C./Donnelly, C. A./Riley, S./Vollmer, M. A. C./Ferguson, N. M./Okell, L. C./Bhatt, S. (2020): Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature, 2020, 584: 257-261.

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. Bern: Staatssekretariat für Wirtschaft SECO.

Gavenciak, T./Teperowski Monrad, J./Leech, G./Sharma, M./Mindermann, S./Bhatt, S./Brauner, J./Kulveit, J. (2022): Seasonal variation in SARS-CoV-2 transmission in temperate climates. In: PLoS Computational Biology, 2022, 18, 8: e1010435.

Goodman-Bacon, A. (2021): Difference-in-differences with variation in treatment timing. In: Journal of Econometrics 2021, 225, 254-277.

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

Greene, W. H. (2012): Econometric Analysis. Edinburgh Gate: Pearson.

Haug, N./Geyrhofer, L./Londei, A./Dervic, E./Desvars-Larrive, A./Loreto, V./Pinior, B./Thurner, S./Klimek, P. (2020): Ranking the effectiveness of worldwide COVID-19 government interventions. In: Nature Human Behavior, 2020, 4: 1303-1312.

Hillmert, S. (2019): Auswahl und Kontrolle von Drittvariablen bei der Bestimmung kausaler Zusammenhänge – illustriert mit Directed Acyclic Graphs (DAGs). Universität Tübingen: Institut für Soziologie.

Jacobsen, G. D./Jacobsen, K. H. (2020): Statewide COVID-19 Stay-at-Home Orders and Population Mobility in the United States. In: World Medical and Health Policy, 2020, 12, 4: 347-356.

Kabisch, M./Ruckes, C./Seibert-Grafe, M./Blettner, M. (2011): Randomized controlled trials: part 17 of a series on evaluation of scientific publications. In: Deutsches Ärzteblatt International 2011, 108, 39: 663-688.

Khazaei, Y./Küchenhoff, H./Hoffmann, S./Syliqi, D./Rehms, R. (2023): Using a Bayesian hierarchical approach to study the association between non-pharmaceutical interventions and the spread of Covid-19 in Germany. In: Scientific Reports, 2023, 13: 18900.

Kosfeld, R./Mitze, T./Rode, J./Wälde, K. (2021): The Covid-19 containment effects of public health measures: A spatial difference-in-differences approach. In: Journal of Regional Science, 2021, 61, 4: 799-825.

Kugler, F./Schwerdt, G./Wößmann, L. (2014): Ökonometrische Methoden zur Evaluierung kausaler Effekte der Wirtschaftspolitik. In: Perspektiven der Wirtschaftspolitik, 2014, 15, 2: 105-132.

Levelu, A./Sandkamp, A. (2023): A Lockdown a Day Keeps the Doctor Away: The Global Effectiveness of Non-Pharmaceutical Interventions in Mitigating the Covid-19 Pandemic. CESifo Working Papers, 10023.

Li, Y./Campbell, H./Kulkarni, D./Harpur, A./Nundy, M./Wang, X./Nair, H. (2021): The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. In: Lancet Infectious Diseases 2021, 21, 2: 193-202.

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

Moriyama, M./Hugentobler, W. J./Iwasaki, A. (2020): Seasonality of Respiratory Viral Infections. In: Annual Review of Virology 2020, 7, 1: 83-101.

Müller, B. (2021): Zur Modellierung der Corona-Pandemie – eine Streitschrift. In: Monitor Versorgungsforschung, 2021, 6: 68-79.

Nishiura, H./Chowell, G. (2009): The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends. In: Chowell, G./Hyman, J. M./Bettencourt, L. M. A. (Hrsg.): Mathematical and Statistical Estimation Approaches in Epidemiology, 103-121.

Sharma, M., Mindermann, S./Rogers-Smith, C./Leech, G./Snodin, B./Ahuja, J./Sandbrink, J. B./Teperowski Monrad, J./Altman, G./Dhaliwal, G./Finnveden, L./Norman, A. J./Oehm, S. B./Sandkühler, J. F./Aitchison, L./Gavenciak, T./Mellan, T./Kulveit, J./Chindelevitch, L./Flaxman, S./Gal, Y./Mishra, S./Bhatt, S./Brauner, J. M. (2021): Understanding the effectiveness of government interventions against the resurgence of COVID-19 in Europe. In: Nature Communications, 2021, 12: 5820.

Viboud, C./Simonsen, L./Chowell, G. (2016): A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. In: Epidemics 2016, 15: 27-37.

Wieland, T. (2020): A phenomenological approach to assessing the effectiveness of COVID-19 related nonpharmaceutical interventions in Germany. In: Safety Science, 2020, 131: 104924.

Wooldridge, J. M. (2012): Introductory Econometrics. A Modern Approach. South-Western: Mason.

Natural experiments to assess the effectiveness of Covid-19 containment measures – methodological principles, implementation in NPI studies and a case study

To assess the effectiveness of non-pharmaceutical interventions (NPIs) in controlling infections during the Covid-19 pandemic, natural experiments are useful. This article introduces the principle of natural experiments and their statistical analysis and, in the first step, discusses their implementation in previous NPI studies. The validity of many studies on the effectiveness of NPIs is limited because, on the one hand, there are often no or only weak control conditions and, on the other hand, because there are problems in the specification of the analysis models. It is often implicitly assumed that only NPIs can slow down the infection process because there are no control variables for any other causes. In the second step, a natural experiment is analyzed about the effectiveness of five measures in the second Corona wave. Switzerland (where these measures were not taken) and neighboring Baden-Württemberg serve as comparison regions. It turns out that the closure of close-contact services causes a reduction in the effective reproduction number, i.e., dampens the wave of infections. However, closing schools and daycare facilities as well as hotels is not effective.

Keywords

Non-pharmaceutical interventions, Covid-19, Pandemic, Natural experiments, Econometrics, Difference-in-differences approach

Autorenerklärung

Der Autor erklärt, dass keine Interessenkonflikte vorliegen.

Zitationshinweis

Wieland, T.: „Natürliche Experimente zur Beurteilung der Effektivität von Covid-19-Eindämmungsmaßnahmen“, in: „Monitor Versorgungsforschung“ (01/24), S. 55–62. http://doi.org/10.24945/MVF.01.24.1866-0533.2580

Dr. rer. nat. Dipl.-Geogr. Thomas Wieland ORCID: 0000-0001-5168-9846
hat Geographie studiert und danach in diesem Fach promoviert. Er hat sowohl in der universitären Forschung und Lehre als auch in der Privatwirtschaft gearbeitet, wo er aktuell wieder tätig ist. Seine inhaltlichen Schwerpunkte sind Wirtschaftsgeographie, medizinische Geographie, Statistik und ökonometrische Modelle sowie Geographische Informationssysteme (GIS).
Kontakt: geowieland@googlemail.com

