

COVID-19-Pandemie: Sind harte Lockdowns unwirksam?

Warum die Studie von Bendavid et al. dafür kein Beweis ist

COVID-19 Pandemic: Are Hard Lockdowns Ineffective?

Why the Bendavid et al. Study does not Provide Sufficient Evidence for this Assumption

Anna Glechner¹, Jan M. Stratil², Manuela Bombana³, Ursula Griebler¹, Hajo Zeeb^{4,5}, Gerald Gartlehner^{1,6}

Hintergrund

Während der Corona-Pandemie führten viele Länder Einschränkungen des täglichen Lebens ein, um die Ausbreitung von SARS-CoV-2 (*severe acute respiratory syndrome coronavirus type 2*) einzuschränken. Diese Maßnahmen waren je nach Region und Ausmaß des Infektionsgeschehens unterschiedlich. Eine Gruppe von Wissenschaftlern der Stanford University unter der Leitung von Prof. John Ioannidis kam aufgrund einer Analyse von vorhandenen Daten aus zehn Ländern zu dem Schluss, dass ein harter Lockdown gegenüber weniger restriktiven Einschränkungen keine signifikanten Vorteile bietet, um Infektionszahlen zu senken. Die Publikation der Studie (Bendavid et al.) wird in sozialen Medien als Beleg dafür verwendet, dass ein Lockdown keinen wesentlichen Nutzen bei der Bekämpfung des Infektionsgeschehens bringt.

Methoden

Sechs Wissenschaftler*innen beurteilten unabhängig voneinander die methodische Qualität der Studie. Dafür wurde das international anerkannte Bewertungsinstrument ROBINS-I (*Risk Of Bias in Non-randomized Studies of Interventions*) angewendet, das sich in sieben Domänen gliedert. Konflikte bei der Bewertung wurden durch Konsens gelöst. Das Risiko für Verzerrungen der Resultate wurde anhand eines adaptierten dreistufigen Schemas als niedrig, moderat oder hoch eingestuft.

Ergebnisse

Die Studie von Bendavid et al. weist schwerwiegende methodische Mängel auf. Die Studie wurde in vier von sieben Domänen mit einem hohen Verzerrungspotenzial bewertet. Einer der schwerwiegendsten Mängel betraf Störfaktoren, da die Länder, die verglichen wurden, Maßnahmen in unterschiedlicher Intensität anwendeten und diese je nach Ausgangssituation unterschiedlich wirksam waren. Wesentliche Informationen über die Kriterien für die Auswahl der Länder sowie eine klare Definition der Interventionen fehlten.

Schlussfolgerungen

Die Ergebnisse der Studie haben ein hohes Verzerrungspotenzial und sollten nicht als Grundlage für Entscheidungen dienen.

Schlüsselwörter

Lockdown; SARS-CoV-2; Maßnahmen; Methodenkritik

Background

During the Corona pandemic, many countries implemented restrictions on daily life to contain the spread of severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2). These measures varied by country, depending on the region and the rate of infections. A group of Stanford University researchers led by Prof. John Ioannidis concluded, based on an analysis of existing data from 10 countries, that a hard lockdown, compared with less restrictive measures, did not offer significant benefits in reducing infection rates. The publication of the study (Bendavid et al.) is being used in social media as evidence that lockdowns have no significant benefit in containing the spread of SARS-CoV-2.

Methods

Six researchers independently assessed the methodological quality of the study. The internationally accepted assessment tool ROBINS-I (Risk Of Bias in Non-randomized Studies of Interventions) was applied, which is divided into seven domains. Conflicts in the evaluation were resolved by consensus. Risk of bias in results was graded as low, moderate, or high using an adapted three-tiered scheme.

Results

The study by Bendavid et al. has serious methodological flaws. The study was rated as having high risk of bias in four of seven domains. One of the most serious shortcomings related to confounding factors: the countries that were compared applied measures at different intensities, and the effectiveness of those measures varied depending on the baseline situation. Crucial information on the criteria used to select the countries included in the study, as well as a clear definition of the interventions, were missing.

Conclusions

The results of the study have a high potential for bias and should not be used as a basis for decision-making.

Keywords

lockdown; SARS-CoV-2; interventions; methodology critique

¹ Department für Evidenzbasierte Medizin und Evaluation, Cochrane Österreich, Donau-Universität, Krems/Österreich; ² Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, LMU, München; ³ Abteilung Allgemeinmedizin und Versorgungsforschung, Universitätsklinikum, Heidelberg

⁴ Leibniz-Institut für Präventionsforschung und Epidemiologie, Bremen; ⁵ Health Sciences Bremen, Universität Bremen

⁶ Research Triangle Institute (RTI) International, North Carolina, USA

DOI 10.3238/zfa.2021.0103-0107

Hintergrund

Seit Januar 2020 wurden weltweit in vielen Ländern Lockdown-Maßnahmen, also temporäre staatlich verordnete Einschränkungen des öffentlichen Lebens, eingeführt, um die Ausbreitung von SARS-CoV-2 einzudämmen [1]. Diese Maßnahmen umfassen zum Beispiel das Schließen von Schulen, Geschäften, Restaurants oder Kulturbetrieben, die Absage von Massenveranstaltungen oder Ausgangsbeschränkungen. Die Intensität der einzelnen Maßnahmen variierte von Land zu Land und verändert sich laufend, basierend auf wissenschaftlichen Modellrechnungen und politischen Entscheidungen. Lockdown-Maßnahmen werden in der Regel dann eingesetzt, wenn nicht-pharmakologische Interventionen, wie zum Beispiel Teststrategien, Tragen eines Mund-Nasenschutzes, Abstandsregeln, Hygienemaßnahmen oder Reisebeschränkungen, nicht ausreichend wirksam sind, um die Zahl der Neuinfektionen zu senken.

Eine Gruppe von Wissenschaftlern der Stanford University, USA, unter der Leitung von Prof. John Ioannidis widmete sich der Frage, ob die SARS-CoV-2-Pandemie auch durch weniger restriktive Maßnahmen als Lockdowns eingedämmt werden könnte. Die Ergebnisse der Studie wurden im Januar 2021 im *European Journal of Clinical Investigation* von Bendavid et al. publiziert. [2]

Die Autoren verwendeten für die Studie Daten des ersten Lockdowns zu Beginn der Pandemie 2020. Sie versuchten zu ermitteln, wie sich bestimmte Maßnahmen (z.B. das Schließen von Schulen) auf den Verlauf der Pandemie ausgewirkt haben. Berechnet wurden die Veränderungen der täglichen Fallzahlen nach Einführung der Maßnahmen. Dabei wurde auf vorgehende Analysen und Datenbanken [1, 3, 4] zurückgegriffen. Eine offizielle Anweisung zuhause zu bleiben sowie Geschäftsschließungen definierten die Autoren als stark einschränkende Maßnahmen (Lockdowns). Zwei Länder, Schweden und Südkorea, führten diese beiden Maßnahmen zu Beginn der Pandemie nicht durch und wurden in der

Studie als Länder mit „weniger strengen Maßnahmen“ eingestuft. Zusätzlich wurden acht Länder ausgewählt, die zu Beginn der Pandemie 2020 Lockdowns verhängten: Deutschland, England, Frankreich, Italien, Iran, Niederlande, Spanien, und die USA. Die Autoren berechneten, wie gut Lockdowns in diesen acht Ländern wirkten, um den Anstieg der SARS-CoV-2-Fälle einzudämmen. Davon abgezogen wurde der Effekt, der in Schweden und Südkorea beobachtet wurde. Mit dieser Subtraktion berechneten die Autoren, ob Lockdowns mehr oder weniger effektiv waren als weniger restriktive Maßnahmen.

Auf Basis ihrer Ergebnisse schlossen die Autoren, dass Lockdowns im Vergleich zu weniger restriktiven nicht-pharmakologischen Maßnahmen keinen wesentlichen Zusatzeffekt erzielen. In Schweden und Südkorea, die in dieser Studie als Länder mit weniger einschränkenden Maßnahmen definiert wurden, gab es zu Beginn der Pandemie ebenfalls Interventionen, um das Infektionsgeschehen einzudämmen, wie z.B. Empfehlungen für einen Mund-Nasenschutz (in Südkorea für manche Bereiche verpflichtend), Home-Office, Schulschließungen, Verbot von öffentlichen Versammlungen, Einschränkung von Inlands- und internationalem Reiseverkehr, Teststrategien und eine konsequente Kontaktnachverfolgung.

Die Studie fand in den sozialen Medien enorme Beachtung und wird immer wieder als Beleg dafür zitiert, dass weniger restriktive Maßnahmen auch ausgereicht hätten, um die Verbreitung von SARS-CoV-2 einzudämmen. Der *Altmetric Score*, ein Indikator für die Menge an erlangter Aufmerksamkeit der Studie kletterte Mitte Februar auf mehr als 16.000, mit über 26.000 Tweets [5]. Experten kritisierten jedoch, dass die vereinfachten Annahmen in der Studie der komplexen Pandemie-Situation nicht gerecht werden [6]. Die Anfälligkeit der Studie für systematische Fehler ergibt sich primär an zwei Punkten: (i) der Ermittlung der Auswirkung der Maßnahmen auf das Infektionsgeschehen (z.B. welche Auswirkungen hatten Geschäftsschließungen in

Deutschland auf die Zahl der Neuinfektionen?) und (ii) der Methode, die genutzt wurde, um den „Mehreffekt“ von Lockdowns zu ermitteln.

Im Anschluss präsentieren wir eine formale kritische Evaluierung der Methoden der Studie von Bendavid und Kollegen „*Assessing mandatory stay-at-home and business closure effects on the spread of COVID-19*“ [2].

Methoden

Die Qualität von kontrollierten Studien wird in erster Linie von Design, Durchführung und Analyse bestimmt [7]. Gerade die Validität der Ergebnisse von nicht-randomisierten Studien kann durch nicht sorgfältige methodische Planung und Durchführung stark eingeschränkt sein – man spricht vom Verzerrungsrisiko. Unter Verzerrungen versteht man systematische Fehler beim Design, der Durchführung und der Analyse von Studien, welche die Ergebnisse der Studie verzerren [8]. Diese Verzerrungen können sowohl zu einer Unterschätzung als auch einer Überschätzung des tatsächlichen Effektes führen. Ein international anerkanntes Instrument zur Bewertung des Verzerrungsrisikos von nicht-randomisierten Studien ist ROBINS-I (*Risk Of Bias In Non-randomized Studies of Interventions*) [9]. Es bewertet sieben Domänen, die bei nicht-randomisierten Studien zu einer Verzerrung der Studienergebnisse führen können:

1. Sind potenzielle Störfaktoren, die einen Einfluss auf den Endpunkt haben, gleichmäßig zwischen beiden Gruppen verteilt?
2. Wie wurde die Auswahl der Teilnehmer*innen (in der vorliegenden Studie Länder) der Studie durchgeführt?
3. Gab es eine genaue Definition/Klassifikation der Interventionen, die verglichen wurden?
4. Gab es Abweichungen von den geplanten Interventionen?
5. Wie wurde mit fehlenden Daten umgegangen?
6. Waren die Messmethoden der Endpunkte für beide Gruppen gleich?
7. Wie wurden die in der Publikation berichteten Endpunkte ausgewählt?

Sechs Wissenschaftler*innen mit Erfahrung in der kritischen Evaluie-

ROBINS-I-Domäne	Risiko für Verzerrung	Begründung der Bewertung
Verzerrung aufgrund einer Ungleichverteilung von Störfaktoren am Beginn der Studie	Hoch	<ul style="list-style-type: none"> Keine Adjustierung für Störfaktoren am Beginn der Studie, z.B. unterschiedliche Intensität und Wirksamkeit von nicht-pharmakologischen Interventionen Die Wirksamkeit von nicht-pharmakologischen Interventionen bestimmt, ob ein Lockdown eingeführt wird und ist gleichzeitig auch ein prognostischer Faktor für den Endpunkt (Anstieg der SARS-CoV-2 Neuinfektionen).
Verzerrung aufgrund der Auswahl der eingeschlossenen Länder	Hoch	<ul style="list-style-type: none"> Risiko für Verzerrungen durch (systematisches) Auslassen von Ländern mit hoher Wirksamkeit der restriktiven Maßnahmen Risiko für Verzerrungen bei dem Ansatz, Effekte zwischen Länder zu übertragen, durch Unterschiedlichkeit der Länder
Verzerrung bei der Klassifikation der Interventionen	Hoch	<ul style="list-style-type: none"> Unzureichende Definitionen von restriktiven und weniger restriktiven nicht-pharmakologischen Interventionen Überlappungen zwischen den Interventionen Intensität, Strenge (z.B. Strafen bei Nicht-Umsetzung) und Dauer der Interventionen, die zwischen den Ländern variieren, wurden nicht miteinbezogen.
Verzerrung aufgrund von Abweichungen bei den Interventionen	Unzureichende Information	<ul style="list-style-type: none"> Unzureichende Berichterstattung der angewandten Methoden
Verzerrung aufgrund von fehlenden Daten	Moderat	<ul style="list-style-type: none"> Es wird nicht berichtet, ob und welche Daten fehlen (Endpunktdaten, Indikatoren, Störfaktoren, etc.) und wie mit fehlenden Daten umgegangen wurde. Auf der Webseite einer der Datenquellen (<i>Oxford Covid-19 Government Response Trackers</i>) wird erläutert, dass bestimmte Indikatoren an manchen Tagen fehlen – dies bleibt in der Studie unerwähnt.
Verzerrung beim Messen des Endpunkts	Hoch	<ul style="list-style-type: none"> Vorgehen beim Testen ist sowohl innerhalb von Ländern als auch zwischen den Ländern äußerst variabel (Teststrategien, Anzahl durchgeführter Tests pro Zeiteinheit und Unterschiede bei den verwendeten Tests); keine Berücksichtigung dieser Unterschiede im Modell. Zeitlicher Verzug zwischen Beginn einer Strategie und Auswirkungen auf die Infektionsrate wird nicht berücksichtigt bzw. dargestellt. Auswahl des Startpunkts für die Ermittlung der Infektionsrate ist nicht transparent.
Verzerrung bei der Auswahl der berichteten Endpunkte	Unzureichende Information	<ul style="list-style-type: none"> Es wurde kein Studienprotokoll publiziert
Gesamtbewertung	Hoch	<ul style="list-style-type: none"> Störfaktoren im Sinne von wesentlichen Unterschieden zwischen den Ländern, die die Infektionsraten beeinflussen, wurden nicht berücksichtigt. Die Einteilung in Länder mit restriktiveren Interventionen (Lockdowns) und weniger restriktiven Interventionen ist willkürlich und nicht nachvollziehbar. Unterschiede zwischen Ländern in Bezug auf Teststrategien und Anzahl der durchgeführten Tests wurden nicht berücksichtigt. Fehlende Transparenz bei wesentlichen methodischen Entscheidungen

Tabelle 1 Zusammenfassung der Bewertung der Studie von Bendavid et al. mit ROBINS-I

Die Bewertung von nicht-randomisierten Studien führte unabhängig voneinander eine Bewertung der Studie von Bendavid et al. mit ROBINS-I durch. Unterschiede bei den Bewertungen der einzelnen Domänen und bei der Gesamtbewertung wurden bei einer Videokonferenz durch Konsens gelöst. Die von ROBINS-I vorgeschlagene 4-stufige Bewertung des Verzerrungsrisikos, wurde auf ein 3-stufiges Schema (niedriges, moderates, hohes Verzerrungsrisiko) modifiziert.

Ergebnisse

Die Studie von Bendavid et al. wurde in vier von sieben Domänen mit hohem Verzerrungsrisiko bewertet. Insgesamt ergab sich daraus ein hohes Verzerrungsrisiko als Gesamtbewertung für den von der Studie untersuchten Endpunkt „Anstieg der SARS-CoV-2 Neuinfektionen“. Im Folgenden fassen wir die wesentlichsten Punkte bei der Bewertung der einzelnen Domänen von ROBINS-I kurz zusammen. Tabelle 1 bietet einen Überblick über das Risiko für Verzerrung

über die sieben bewerteten Domänen hinweg und eine entsprechende Begründung der jeweiligen Bewertung.

Verzerrung aufgrund einer Ungleichverteilung von Störfaktoren am Beginn der Studie

Unterschiede zwischen Ländern können als Störfaktoren agieren, wenn sie sowohl Einfluss auf die Einführung eines Lockdowns als auch Auswirkung auf die Infektionsrate haben. Solche potenziellen Störfaktoren wur-

den in der Studie nicht statistisch berücksichtigt. Zum Beispiel führte Südkorea eine sehr restriktive Kontaktnachverfolgung unter Nutzung von persönlichen Bewegungsprofilen von Mobiltelefonaten durch (ein Vorgehen, das in europäischen Ländern nicht legal wäre). Dies führt zu besserer Kontaktnachverfolgung, senkt die Infektionsraten und beeinflusst möglicherweise das Verhalten der überwachten Personen. Wenn solche Unterschiede zwischen Ländern statistisch nicht berücksichtigt werden, kann es dazu führen, dass nicht-pharmakologische Interventionen wie das Tragen eines Mund-Nasenschutzes oder Handhygiene scheinbar besser wirken als es tatsächlich der Fall ist. Andere Faktoren, die als Störfaktoren agieren können, aber nicht berücksichtigt wurden, sind die generelle Bereitschaft der Bevölkerung Regierungsmaßnahmen zu folgen, Intensität der Teststrategien, Effizienz der Kontaktnachverfolgung, Altersverteilung der Bevölkerung oder Bevölkerungsdichte.

Das Risiko für Verzerrungen des Ergebnisses aufgrund von Ungleichverteilung von Störfaktoren am Beginn der Studie wurde als hoch bewertet.

Verzerrung aufgrund der Auswahl der eingeschlossenen Länder

In der Publikation wird nicht schlüssig erklärt, warum jene zehn Länder (Deutschland, England, Frankreich, Italien, Iran, Niederlande, Schweden, Spanien, Südkorea und die USA) für die Studie ausgewählt wurden und andere Länder nicht. Wenn die Auswahl nicht zufällig oder nicht vollständig war, besteht das Risiko einer Verzerrung. Auffällig ist beispielsweise, dass Länder wie China, Australien oder Neuseeland, in denen sehr restriktive Maßnahmen zu einer (fast) vollständigen Durchbrechung des Pandemiegeschehens geführt hatten, nicht bei der Auswertung berücksichtigt wurden.

Ein zweites Risiko für Verzerrungen ergibt sich, wenn angenommen wird, dass Maßnahmen in verschiedenen Ländern gleich wirken würden, was methodisch durch die Ermittlung des Effekts durch Subtraktion suggeriert wird. Beispielsweise könnten Maßnahmen durch verschiedene soziokulturelle Praktiken

in einzelnen Ländern eine deutlich unterschiedliche Wirksamkeit entfalten (z.B. eine gesetzliche Pflicht zum Tragen von Alltagsmasken wird weniger effektiv sein in Ländern, bei denen die Bevölkerung schon vor der Pandemie häufig oder regelmäßig ei-



Dr. med. Anna Glechner ...

... ist Allgemeinmedizinerin und seit 2011 wissenschaftliche Mitarbeiterin bei Cochrane Österreich am Department für Evidenzbasierte Medizin und Evaluation der Donau-Universität Krems. Seit 2016 leitet sie das Evidenzbasierte Informationszentrum für Ärztinnen und beantwortet Anfragen von Mediziner*innen und fasst dafür gemeinsam mit ihrem Team die besten Ergebnisse aus internationalen Studien übersichtlich und in leicht verständlicher Form zu Rapid Reviews zusammen.

Foto: Andrea Reischer

ne Alltagsmaske getragen hat). Wenn für derartige Faktoren nicht statistisch ausgeglichen wurde, besteht die Gefahr, dass Gleiches nicht mit Gleichem verglichen wurde und verzerrte Ergebnisse entstehen.

Ein geringeres Risiko für systematische Verzerrungen würde bestehen, wenn beispielsweise eine Vollerhebung innerhalb einer bestimmten Gruppierung erfolgt wäre (z.B. alle OECD-Länder). Ein alternativer Ansatz wäre gewesen, ausgehend von den Fällen Südkorea und Schweden, möglichst gute Vergleichspartner zu finden, die sich in möglichst wenigen Faktoren unterscheiden, außer, dass letztere restriktivere Maßnahmen eingeführt haben (z.B. Norwegen oder Dänemark für das Fallbeispiel Schweden). In beiden Fällen wäre es zudem wichtig, statistisch auf andere Einflussfaktoren auszugleichen, die Unterschiede zwischen der Wirksamkeit der Maßnah-

men in einzelnen Ländern erklären könnten.

Das Risiko für Verzerrungen des Ergebnisses aufgrund der Auswahl der eingeschlossenen Länder wurde als hoch bewertet.

Verzerrung bei der Klassifikation der Interventionen

Die Basis der Studie ist eine Einteilung von Ländern in jene mit restriktiveren nicht-pharmakologischen Interventionen, d.h. Lockdowns (Deutschland, England, Frankreich, Italien, Iran, Niederlande, Spanien) und jene mit weniger restriktiven nicht-pharmakologischen Interventionen (Schweden und Südkorea). Die Veränderung der SARS-CoV-2-Infektionsraten der beiden Gruppen von Ländern wurde anschließend verglichen. Eine genaue Definition der Interventionskategorien fehlt in der Publikation. Es werden lediglich zwei Maßnahmen als Beispiele für restriktivere Interventionen erwähnt: die „Anweisung zuhause zu bleiben“ und „Geschäftsschließungen“. Welche Kriterien die Autoren festlegten, um ein Land in der Gruppe mit Lockdown oder in der Vergleichsgruppe mit weniger restriktiven Maßnahmen einzureihen, ist im Artikel nicht beschrieben. Des Weiteren wurden die Intensität der Umsetzung bzw. Sanktionierung und die Dauer der Interventionen, die zwischen den Ländern variierte, nicht in die Analysen miteinbezogen. Die von der Studie zitierte Datenquelle „*Coronavirus Government Response Tracker*“ der Universität Oxford [1] weist mittlerweile Daten von 180 Ländern weltweit auf mit differenzierten Daten über Maßnahmen in den einzelnen Ländern. Durch die unklare Definition der Interventionskategorien ist fraglich, ob die Vergleiche wirklich valide sind. So wurde z.B. Südkorea in die Gruppe der weniger restriktiven Länder eingeordnet, obwohl ab Ende März 2020 auch in Südkorea die Anweisung galt, nur für sportliche Aktivitäten, Lebensmitteleinkäufe und notwendige Fahrten das Haus zu verlassen [1].

Das Risiko für Verzerrungen des Ergebnisses aufgrund der Klassifikation der Interventionen wurde als hoch bewertet.

Verzerrung aufgrund von Abweichungen bei den Interventionen

Die Studie bietet nicht genug Information, um diese Domäne beurteilen zu können.

Verzerrung aufgrund von fehlenden Daten

In der vorliegenden Studie wird an keiner Stelle erwähnt, für welche Daten fehlende Werte vorliegen, es gibt also keinerlei Informationen darüber, ob relevante Daten wie Endpunktdaten, Indikatoren und Störfaktoren vollständig sind, bzw. wie viele Daten pro Variable fehlen. Es wird im Methodenteil auch nicht beschrieben, wie mit fehlenden Daten im statistischen Modell umgegangen wurde. Auf Seiten des *Oxford Covid-19 Government Response Trackers* (OxCGRT) wird erläutert, dass bestimmte Indikatoren an manchen Tagen fehlen [1]. Das Ausmaß der fehlenden Daten und die Auswirkung auf die Ergebnisse der Studie ist schwer einschätzbar.

Das Risiko für Verzerrungen des Ergebnisses aufgrund von fehlenden Daten wurde als moderat bewertet.

Bias beim Messen des Endpunkts

Für die Ermittlung des Endpunkts „Anstieg der SARS-CoV-Neuinfektionen“, ist eine Registrierung neuer Infektionen notwendig. Das Vorgehen beim Testen wird sowohl innerhalb der Länder als auch zwischen den Ländern als äußerst variabel eingeschätzt, z.B. in Bezug auf Teststrategien, Anzahl durchgeführter Tests pro Zeiteinheit und Unterschiede bei den verwendeten Tests. Es ist zudem unklar, wie mit dem zeitlichen Verzug zwischen einsetzender Strategie und dem Auftreten von Infektionen umgegangen wurde. Die Ermittlung der Wachstumsrate hängt stark vom gewählten Startzeitpunkt ab, von dem an die Fallzahlen berechnet werden. Eine differenzierte Begründung und Erläuterung des gewählten Startzeitpunkts per Land oder subnationaler Region konnten wir nicht erkennen.

Das Risiko für Verzerrungen des Ergebnisses aufgrund von Ungleichheiten beim Messen des Endpunkts wurde als hoch bewertet.

Verzerrung bei der Auswahl der berichteten Endpunkte

Es wurde kein Studienprotokoll veröffentlicht, in dem a priori festgelegte Endpunkte dokumentiert werden. Die Studie bietet nicht genug Information, um diese Domäne beurteilen zu können.

Diskussion

Die formale methodische Bewertung der Studie von Bendavid et al. [2] zeigt schwerwiegende methodische Probleme auf. Die Bewertung mit hohem Risiko für Verzerrungen bedeutet, dass diese Studie keine belastbare Evidenz für oder gegen die Wirksamkeit von restriktiven nicht-pharmakologischen Maßnahmen liefert. Sie sollte nicht als Grundlage für Entscheidungen verwendet werden. Unseres Wissens ist unsere Untersuchung die erste formale methodische Bewertung dieser Studie. Informelle Bewertungen durch *PubPeer*, einem Online Journal Club, zeigen jedoch ebenfalls massive methodische Bedenken auf [10].

Zusammenfassend sind bei einem Vergleich von mehr oder weniger restriktiven Maßnahmen viele Faktoren zu berücksichtigen, um die Komplexität der Maßnahmen und den Einfluss auf die Infektionsrate in den unterschiedlichen Ländern zu erfassen. Eine solche Studie stellt eine große Herausforderung dar, und ist möglicherweise nicht valide durchführbar.

Interessenkonflikte:

Keine angegeben.

Literatur

1. Coronavirus government response tracker: University of Oxford. www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker
2. Bendavid E, Oh C, Bhattacharya J, Ioannidis JPA. Assessing mandatory

stay-at-home and business closure effects on the spread of COVID-19. *Eur J Clin Invest* 2021; e13484

3. Hsiang S, Allen D, Annan-Phan S, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* 2020; 584: 262–7
4. Number of coronavirus (COVID-19) cases in Sweden by region. www.statista.com/statistics/1103949/number-of-coronavirus-covid-19-cases-in-sweden-by-region/
5. Assessing mandatory stay-at-home and business closure effects on the spread of COVID-19. Overview of attention for article published in *European Journal of Clinical Investigation*, February 2021: Altmetric. www.altmetric.com/details/97143657?src=bookmarklet
6. Bartens W. „Da gibt es doch diese Studie aus Stanford ...“. *Süddeutsche Zeitung* 30.1.2021
7. Buchberger B, von Elm E, Gartlehner G, et al. Bewertung des Risikos für Bias in kontrollierten Studien. *Bundesgesundheitsbl* 2014; 57: 1432–8
8. Fletcher RH, Fletcher SW, Fletcher GS. *Clinical epidemiology: the essentials*. Alphen aan den Rijn, Walters Kluwer, Lippincott Williams & Wilkins, 2014
9. Sterne JAC, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016; 355: i4919
10. Comments on the study of Bendavid et al: assessing mandatory stay at home business closure effects on the spread of COVID-19. <https://pubpeer.com/publications/3D81CAC483C2021C00E27C8826DF71> (January 2021)

Korrespondenzadresse

Dr. med. Anna Glechner
Cochrane Österreich,
Department für Evidenzbasierte
Medizin und Evaluation
Donau-Universität Krems
Dr.-Karl-Dorrek-Str. 30
A-3500 Krems an der Donau
anna.glechner@donau-uni.ac.at