

Unterschiede der COVID-19-Mortalität zwischen den deutschen Bundesländern

Stefan Huschens*

Version vom 28. Juli 2021**

Zusammenfassung

Zwischen den deutschen Bundesländern bestehen gravierende Unterschiede bezüglich der COVID-19-Mortalität gemessen durch COVID-19-Todesfälle pro 100 000 Einwohner. Würde man die niedrige COVID-19-Mortalität in Schleswig-Holstein auf die gesamte Bundesrepublik übertragen, so würden sich 47 000 anstelle von 92 000 COVID-19-Todesfällen ergeben; übertrüge man die hohe COVID-19-Mortalität in Sachsen auf die gesamte Bundesrepublik, so erhielte man 207 000 anstelle von 92 000 COVID-19-Todesfällen. Diese Spannweite von 47 000 bis 207 000 Todesfällen verdeutlicht, dass eine Ursachenklärung für die regionalen Unterschiede eine relevante wissenschaftliche Aufgabe ist.

Die Unterschiede der COVID-19-Mortalität zwischen den Bundesländern werden aufgezeigt und es werden Hypothesen zur Ursachenerklärung formuliert, die in einer breit und interdisziplinär angelegten empirischen Untersuchung überprüft werden müssten. Zu einer solchen Untersuchung werden statistisch-methodische Vorüberlegungen präsentiert.

Inhaltsverzeichnis

1	COVID-19-Mortalität, -Inzidenz und -Letalitätsquoten	2
1.1	COVID-19-Mortalität	2
1.2	COVID-19-Inzidenz und COVID-19-Letalitätsquote	3
1.3	Zusammenhang zwischen COVID-19-Mortalität, -Inzidenz und -Letalitätsquote	4
2	Mögliche Ursachen unterschiedlicher COVID-19-Inzidenzen und -Letalitätsquoten	6
2.1	Regionale COVID-19-Inzidenzunterschiede	6
2.2	Regionale COVID-19-Letalitätsunterschiede	7
2.3	Regionale COVID-19-Inzidenz- und -Letalitätsunterschiede	8
3	Statistisch-methodische Vorüberlegungen für eine empirische Untersuchung der Ursachen	9

*stefan.huschens@tu-dresden.de, <https://www.stefan-huschens.de>

**Bitte prüfen Sie, bevor Sie sich auf diese Ausarbeitung beziehen, hier <https://www.stefan-huschens.de/statistik/COVID-19/> ob es eine neuere Version gibt.

1 COVID-19-Mortalität, -Inzidenz und -Letalitätsquoten

Die Ausgangsdaten für die folgende Darstellung entstammen dem Lagebericht des Robert Koch-Instituts (RKI) zur Corona-Virus-Krankheit 2019 (COVID-19) vom 26. Juli 2021[RKI]. Alle Zahlenangaben beziehen sich auf diesen Stichtag.

1.1 COVID-19-Mortalität

Die COVID-19-Mortalität M ist gemessen durch die Zahl der COVID-19-Todesfälle pro 100 000 Einwohner. Zur Berechnung der COVID-19-Mortalität wird für die jeweils betrachtete Region die Anzahl der seit Pandemiebeginn kumulierten dem RKI gemeldeten COVID-19-Todesfälle („kumulative Todesfälle“ in der Terminologie des RKI) durch die jeweilige Einwohnerzahl dividiert, mit 100 000 multipliziert und das Ergebnis ganzzahlig gerundet. In der zweiten Spalte von Tabelle 1 ist die COVID-19-Mortalität für die Bundesländer angegeben

Tabelle 1: COVID-19-Mortalität M , COVID-19-Inzidenz I und COVID-19-Letalitätsquote L in % für deutsche Bundesländer am 26. Juli 2021

Bundesland	M	I	L
Baden-Württemberg (BW)	93	4 542	2,06
Bayern (BY)	117	4 968	2,35
Berlin (BE)	98	4 952	1,97
Brandenburg (BB)	151	4 325	3,50
(Hansestadt) Bremen (HB)	72	4 073	1,78
(Hansestadt) Hamburg (HH)	87	4 246	2,05
Hessen (HE)	120	4 672	2,58
Mecklenburg-Vorpommern (MV)	73	2 756	2,65
Niedersachsen (NI)	73	3 301	2,20
Nordrhein-Westfalen (NW)	96	4 587	2,10
Rheinland-Pfalz (RP)	95	3 830	2,49
Saarland (SL)	104	4 248	2,45
Sachsen (SN)	249	7 033	3,54
Sachsen-Anhalt (ST)	157	4 530	3,47
Schleswig-Holstein (SH)	56	2 230	2,52
Thüringen (TH)	205	6 051	3,38
Gesamt	110	4 517	2,44

(Anmerkung 1). Diese variiert von 56 für Schleswig-Holstein bis zu 249 für Sachsen. Für die gesamte Bundesrepublik ist die COVID-19-Mortalität 110 und die Anzahl der COVID-19-Todesfälle 91 527 (Anmerkung 2).

Die Relevanz der Mortalitätsunterschiede zeigt folgende kontrafaktische Überlegung. Übertrüge man die niedrige COVID-19-Mortalität in Schleswig-Holstein auf die gesamte Bundesrepublik, so resultierten 46 596 anstelle von 91 527 COVID-19-Todesfällen; übertrüge man die hohe COVID-19-Mortalität in Sachsen auf die gesamte Bundesrepublik, so erhielte man 207 184 anstelle von 91 527 COVID-19-Todesfällen (Anmerkung 3). Die Spanne von 46 596 bis 207 184 Todesfällen verdeutlicht die wissenschaftliche und gesellschaftliche Relevanz einer Ursachenerforschung.

1.2 COVID-19-Inzidenz und COVID-19-Letalitätsquote

Die Wahrscheinlichkeit, an einer bestimmten Krankheit zu sterben, kann in zwei Wahrscheinlichkeiten aufgeteilt werden: die Wahrscheinlichkeit, zu erkranken, und die Wahrscheinlichkeit, im Fall einer Erkrankung an dieser Krankheit zu sterben (siehe Anmerkung 4). Entsprechend kann die krankheitsspezifische Mortalität in die krankheitsspezifische Inzidenz – als relative Häufigkeit des Auftretens der Krankheit in der Bevölkerung – und die krankheitsspezifische Letalitätsquote – als Anteil der an der Krankheit Gestorbenen – aufgeteilt werden. Die krankheitsspezifische Letalität und Mortalität sind also zwei spezielle Konzepte der Sterblichkeit, wobei die Bezugsgröße bei der Letalität die Anzahl der Erkrankten und bei der Mortalität die Anzahl der Einwohner ist.

Für die COVID-19-Erkrankungen bzw. COVID-19-Todesfälle sind unterschiedliche Konzepte für Inzidenz und Letalität denkbar, je nachdem, ob man sich auf die Zahl der SARS-COV-2-Infizierten, die Zahl der dem RKI gemeldeten, laborbestätigten SARS-COV-2-Infektionen oder die Zahl der mit Symptomen an COVID-19 Erkrankten bezieht. Die Zahl der Infektionen ist größer (Dunkelziffer) als die Zahl der dem RKI gemeldeten SARS-COV-2-Infektionen, die wiederum größer als die Zahl der symptomatisch an COVID-19 Erkrankten ist. Für epidemiologische Überlegungen interessieren eher Inzidenz- und Letalitätskonzepte, die auf der Anzahl der SARS-COV-2-Infizierten basieren, die allerdings unter anderem wegen des hohen Anteils asymptomatischer Krankheitsverläufe und dadurch unentdeckter Infektionen nur grob geschätzt werden kann. Die folgenden Inzidenz- und Letalitätskonzepte basieren auf den dem RKI gemeldeten, laborbestätigten SARS-COV-2-Infektionen.

Das hier verwendete Inzidenzkonzept bezieht sich auf die dem RKI gemeldeten Zahlen von laborbestätigten SARS-COV-2-Infektionen, den so genannten COVID-19-Fällen, und misst die COVID-19-Inzidenz I durch COVID-19-Fälle pro 100 000 Einwohner. Zur Berechnung der COVID-19-Inzidenz wird für die jeweils betrachtete Region die Anzahl der kumulierten dem RKI seit Pandemiebeginn gemeldeten COVID-19-Fälle („kumulative Fälle“ in der Terminologie des RKI) durch die jeweilige Einwohnerzahl dividiert, mit 100 000 multipliziert und das Ergebnis ganzzahlig gerundet. Durch die Bezugsgröße 100 000 Einwohner und die Ganzzahligkeit sollen die Inzidenzen leichter verstehbar und vergleichbar sein als die relativ kleinen Prozentzahlen, die sich ergeben, wenn man die COVID-19-Fälle pro Einwohner betrachtet. Die Werte der so berechneten COVID-19-Inzidenz für die Bundesländer sind in der dritten Spalte von Tabelle 1 angegeben (Anmerkung 5). Die COVID-19-Inzidenz variiert von 2 230 für Schleswig-Holstein bis 7 033 für Sachsen und hat für die Bundesrepublik Deutschland den Wert 4 517.

Das hier verwendete Letalitätsquote ist eine Meldequote, die sich errechnet, indem die gemeldeten COVID-19-Todesfälle in Relation zu den gemeldeten COVID-19-Fällen gesetzt wird. Die so gebildete COVID-19-Letalitätsquote L für die gesamte Bundesrepublik ist

$$\frac{\text{COVID-19-Todesfälle}}{\text{COVID-19-Fälle}} = \frac{91\,527}{3\,756\,865} \approx 2,44\% .$$

(Anmerkung 6). Die analog gebildeten COVID-19-Letalitätsquoten für die einzelnen Bundesländer sind, gerundet auf zwei Nachkommastellen, in der vierten Spalte von Tabelle 1 zusammengestellt (Anmerkung 7). Sie variieren zwischen 1,78% für Bremen und 3,54% für Sachsen.

1.3 Zusammenhang zwischen COVID-19-Mortalität, -Inzidenz und -Letalitätsquote

Die Begriffe Fälle, Todesfälle, Mortalität, Inzidenz und Letalitätsquote bedeuten im Folgenden COVID-19-Fälle, -Todesfälle, -Mortalität, -Inzidenz und -Letalitätsquote.

Wegen der Gleichheit

$$\frac{\text{Todesfälle}}{\text{Einwohnerzahl}} = \frac{\text{Fälle}}{\text{Einwohnerzahl}} \times \frac{\text{Todesfälle}}{\text{Fälle}} \quad (1)$$

gilt der Zusammenhang

$$M = I \times L$$

zwischen der Mortalität M , der Inzidenz I und der Letalitätsquote L (Anmerkung 8). Regionale Unterschiede der Mortalität können nur dann auftreten, wenn sich die Inzidenz, die Letalitätsquote oder beide unterscheiden. Die Frage nach den Ursachen regional unterschiedlicher Mortalitäten, kann auf die Fragen nach den Ursachen unterschiedlicher Inzidenzen und unterschiedlicher Letalitätsquoten zurückgeführt werden, wobei gemeinsame Ursachen möglich sind. Der multiplikative Zusammenhang „Mortalität = Inzidenz \times Letalitätsquote“ erklärt auch, warum sich die Bundesländer bei der Mortalität sogar maximal um den Faktor 4,45 ($\approx 249/56$) unterscheiden, obwohl sie sich bei der Inzidenz maximal um den Faktor 3,15 ($\approx 7\,033/2\,230$) und bei der Letalität maximal um den Faktor 1,99 ($\approx 3,54/1,78$) unterscheiden.

In den Tabellen 2 bis 4 sind die Bundesländer jeweils zweifach klassifiziert nach Inzidenz und Mortalität, nach Letalität und Mortalität sowie nach Inzidenz und Letalität. In diesen Tabellen werden für die Bundesländer die Kürzel aus Tabelle 1 verwendet. Der Eintrag **X** markiert Werte für die gesamte Bundesrepublik.

Inzidenz versus Mortalität In Tabelle 2 sind die Bundesländer nach Inzidenz und Mortalität klassifiziert. Die Tabelle verdeutlicht, dass erwartungsgemäß eine höhere Inzidenz

Tabelle 2: Bundesländer klassifiziert nach COVID-19-Inzidenz I und COVID-19-Mortalität M am 26. Juli 2021

	$50 \leq M < 90$	$90 \leq M < 130$	$130 \leq M < 170$	$170 \leq M < 210$	$210 \leq M < 250$
$2000 \leq I < 2500$	SH				
$2500 \leq I < 3000$	MV				
$3000 \leq I < 3500$	NI				
$3500 \leq I < 4000$	RP				
$4000 \leq I < 4500$	HB HH	SL		BB	
$4500 \leq I < 5000$		BE BW BY		ST	
		HE NW			
$5000 \leq I < 5500$					
$5500 \leq I < 6000$					
$6000 \leq I < 6500$	TH				
$6500 \leq I < 7000$					
$7000 \leq I < 7500$	SN				

tendenziell zu einer höheren Mortalität führt und dass die Bundesländer Sachsen und Thüringen bezüglich der Mortalität und der Inzidenz eine hervorgehobene Position einnehmen.

Letalität versus Mortalität In Tabelle 3 sind die Bundesländer nach Letalitätsquote und Mortalität klassifiziert. Diese Tabelle verdeutlicht, dass erwartungsgemäß eine höhere

Tabelle 3: Bundesländer klassifiziert nach COVID-19-Letalitätsquote L in % und COVID-19-Mortalität M am 26. Juli 2021

	$50 \leq M < 90$	$90 \leq M < 130$	$130 \leq M < 170$	$170 \leq M < 210$	$210 \leq M < 250$
$1,6 \leq L < 2,0$	HB	BE			
$2,0 \leq L < 2,4$	HH	BW NW			
$2,4 \leq L < 2,8$	MV NISH	BY HE RP SL X			
$2,8 \leq L < 3,2$					
$3,2 \leq L < 3,6$			BB ST	TH	SN

Letalität tendenziell zu einer höheren Mortalität führt. Es wird die Sonderrolle der vier Bundesländer Brandenburg, Sachsen, Sachsen-Anhalt und Thüringen bzgl. der Letalität deutlich. Da diese vier Bundesländer eine vergleichbar hohe Letalitätsquote haben, gehen die Mortalitätsunterschiede zwischen diesen vier Bundesländer wesentlich auf unterschiedliche Inzidenzen zurück.

Inzidenz versus Letalität In Tabelle 4 sind die Bundesländer nach Inzidenz und Letalitätsquote klassifiziert. Es zeigt sich eine deutlich über dem Durchschnitt liegende Inzidenz in

Tabelle 4: Bundesländer klassifiziert nach COVID-19-Inzidenz I und COVID-19-Letalitätsquote L in % am 26. Juli 2021

	$1,6 \leq L < 2,0$	$2,0 \leq L < 2,4$	$2,4 \leq L < 2,8$	$2,8 \leq L < 3,2$	$3,2 \leq L < 3,6$
$2000 \leq I < 2500$			SH		
$2500 \leq I < 3000$			MV		
$3000 \leq I < 3500$		NI			
$3500 \leq I < 4000$			RP		
$4000 \leq I < 4500$	HB	HH	SL		BB
$4500 \leq I < 5000$	BE	BW BY NW	HE X		ST
$5000 \leq I < 5500$					
$5500 \leq I < 6000$					
$6000 \leq I < 6500$					TH
$6500 \leq I < 7000$					
$7000 \leq I < 7500$					SN

den Bundesländern Sachsen und Thüringen und eine deutlich erhöhte Letalität in den Bundesländern Brandenburg, Sachsen, Sachsen-Anhalt und Thüringen. Die Tabelle legt nahe, dass Inzidenz und Letalität nicht unabhängig sind und es vermutlich gemeinsame Ursachen gibt, die sowohl die Höhe der Inzidenz als auch die Höhe der Letalität beeinflussen. Andererseits weist die Klasse mit einer Inzidenz von 4000 bis unter 4500 eine Spanne der Letalität von 1,78 in Bremen bis 3,50 in Brandenburg auf und die Klasse mit einer Inzidenz von 4500 bis unter 5000 weist eine Spanne der Letalität von 1,97 in Berlin bis 3,47 in Sachsen-Anhalt auf, so dass es auch Faktoren geben muss, welche die Letalität unabhängig von der Inzidenz beeinflussen.

Auffällig ist ein tendenzieller Zusammenhang zwischen der relativen geographischen Lage der Bundesländer und deren Positionierung im Koordinatensystem der Tabelle 4. Dies betrifft sowohl die relative West-Ost-Positionierung als auch die relative Nord-Süd-Positionierung

und lässt vermuten, dass es Einflussfaktoren gibt, die im weitesten Sinn mit der geographischen Lage der Länder zusammenhängen.

2 Mögliche Ursachen unterschiedlicher COVID-19-Inzidenzen und -Letalitätsquoten

Die Versuchung ist groß, die Unterschiede der COVID-19-Inzidenzen oder -Letalitätsquoten voreilig einzelnen plausiblen Faktoren zuzuschreiben. Beispielsweise liegen die drei Bundesländer mit den höchsten Inzidenzen (Bayern, Sachsen und Thüringen) nahe an Tschechien, wo es zwischenzeitlich im Vergleich zu Deutschland sehr hohe Inzidenzen gab. Es ist plausibel, dass Faktoren wie Berufspendler und kleiner Grenzverkehr zu Einkaufszwecken, die mit der geographischen Lage zusammenhängen, Mitursachen für erhöhte Inzidenzen sind, aber nur in einer Studie, die methodisch auch andere Erklärungsfaktoren zulässt, kann die Relevanz dieses regionalen Faktors beurteilt werden. Die folgenden Listen möglicher Ursachen regionaler Inzidenz- und Letalitätsunterschiede verdeutlichen die Komplexität der Fragestellung.

2.1 Regionale COVID-19-Inzidenzunterschiede

Die höchsten Inzidenzen gibt es in den Bundesländern Sachsen und Thüringen. Die niedrigsten Inzidenzen gibt es in den Bundesländern Schleswig-Holstein und Mecklenburg-Vorpommern. Mögliche Ursachen unterschiedlicher regionaler Inzidenzen sind z. B. Unterschiede bei:

1. der Altersstruktur der Bevölkerung,
2. der Struktur der Vorerkrankungen in der Bevölkerung,
3. der Bevölkerungsdichte (z. B. Einwohnerzahl pro qkm),
4. der Zusammensetzung aus Stadt- und Landbevölkerung,
5. der Wohnverhältnissen (z. B. Wohnfläche pro Einwohner, Anteil eigenbewohnter Häuser),
6. den gemeinsamen Grenzen mit Nachbarstaaten (Berufspendler, kleiner Grenzverkehr zu Einkaufs- und Naherholungszwecken),
7. der Geographie (Anteil landwirtschaftlicher Flächen, klimatische Unterschiede, Meer-nähe oder -ferne, Berg- oder Flachland),
8. den Reise- und Urlaubsgewohnheiten der Einheimischen,
9. dem Anteil des Tourismus,
10. dem Anteil der Geschäfts- und Tagungsreisenden,
11. der Nutzung öffentlicher Verkehrsmittel,
12. den Formen des Sozialkontaktes (z. B. Händeschütteln, Umarmung bei Begrüßung, Wangenkuss, enges Zusammensitzen in schlecht gelüfteten Räumen, gemeinsames Frühstück am Arbeitsplatz),

13. den sozialen Kollektivritualen (z. B. Form der Messen in Kirchen, Verhalten in Sportstätten und Sportvereinen, Verhalten auf Vereins- und Dorffesten)
14. der regionalüblichen körperlichen Distanz zwischen unbekanntenen Personen (z. B. auf Rolltreppen, in Warteschlangen, in Geschäften, in Gaststätten, auf Märkten, in Biergärten, in Festzelten),
15. den Sprachstilen (gedämpfte Stimme versus lautes Sprechen und Brüllen; unterschiedliche Singtraditionen in Kindergärten, Schulen, Vereinen und auf Betriebs- und Familienfesten),
16. der Reaktion auf staatliche Empfehlungen und Vorschriften, die COVID-19 betreffen (Anteil der Coronaleugner, Esoteriker und Verschwörungstheoretiker),
17. der Akzeptanz medialer Informationen, die COVID-19 betreffen (Anteil der Coronaleugner, Esoteriker und Verschwörungstheoretiker),
18. den Hygienekonzepten und deren Umsetzung in Arztpraxen, Physiotherapiepraxen und Krankenhäusern; Tatoostudios, Sonnenstudios, Fitnessstudios, Saunen und Bädern; Ämtern und Schulen,
19. der Empfänglichkeit für SARS-COV-2-Infektionen (z. B. regionale Impftraditionen, Anmerkung 9),
20. der Organisation der medizinischen Institutionen (z. B. in Gemeinschaftspraxen und Polikliniken gemeinsame Wartezimmer und „Schwestern“ für mehrere Arztpraxen),
21. der Impfneigung (der Bevölkerung, insbesondere des medizinischen Personals und des Pflegepersonals),
22. den hygienischen Traditionen (z. B. Ausstattung öffentlicher Toiletten, rustikale Biergartentradition, Verwendung bzw. Nicht- oder Mehrfachverwendung von Tischwäsche in Restaurants und Gaststätten).

2.2 Regionale COVID-19-Letalitätsunterschiede

Die mit Abstand höchsten Letalitätsquoten gibt es in den Bundesländern Brandenburg, Sachsen, Sachsen-Anhalt und Thüringen. Die niedrigsten Letalitätsquoten gibt es in Bremen, Berlin und Baden-Württemberg. Mögliche Ursachen unterschiedlicher regionaler Letalitätsquoten sind z. B. Unterschiede bei:

1. der Altersstruktur der Infizierten,
2. der Struktur der Vorerkrankungen bei den Infizierten,
3. der Empfänglichkeit für COVID-19-Erkrankungen im Fall einer SARS-COV-2-Infektion (z. B. bereits durchgemachte Infektionen mit anderen Viren; Anmerkung 10),
4. dem Anteil asymptomatischer und schwach symptomatischer Fälle an den gemeldeten Infektionen als Folge unterschiedlicher Teststrategien,

5. den Behandlungs-, Diagnose- und Pflegemethoden (z. B. Arztpraxen versus Polikliniken, Qualität der Pflegeheime, Einweisung in Krankenhäuser),
6. der Ausstattung, Struktur und Qualität des Gesundheitssystems,
7. der Auslastung oder Überlastung des Gesundheitssystems, insbesondere auf den Höhepunkten der Pandemiewellen,
8. der Feststellung der Todesursache bei COVID-19-Erkrankten (Zweifelsfälle bei Komorbidität),
9. der Feststellung der Todesursache bei nicht offensichtlich an COVID-19-Erkrankten (Obduktionen, nachträgliche Testung),
10. der rechtzeitigen Inanspruchnahme des Gesundheitssystems im Erkrankungsfall (Anteil der Coronaleugner, Esoteriker und Verschwörungstheoretiker),
11. der Impfbeteiligung (Anteil der Impfgegner, Coronaleugner, Esoteriker und Verschwörungstheoretiker).

2.3 Regionale COVID-19-Inzidenz- und -Letalitätsunterschiede

Es gibt denkbare Ursachen, die gleichzeitig die Inzidenz und die Letalität beeinflussen.

Dies ist z. B. eine unterschiedliche regionale Verteilung von Virusvarianten, die unterschiedliche Eigenschaften sowohl bezüglich der Ansteckung als auch bezüglich der Schwere des Krankheitsverlaufs besitzen können. Dies könnte eine Ursache sowohl für unterschiedliche Inzidenzen als auch für unterschiedliche Letalitätsquoten sein.

Im Schnittpunkt der beiden obigen Listen denkbarer Ursachen für unterschiedliche Inzidenzen einerseits und Letalitäten andererseits befinden sich drei Ursachenkomplexe.

- Die Altersstruktur und die Struktur der Vorerkrankungen wirken sich sowohl auf die Infektionen als auch auf die Letalität aus. Vgl. dazu die Nr. 1 und 2 in der Liste aus Abschnitt 2.1 mit denselben Nummern in der Liste aus Abschnitt 2.2.
- Der „Anteil der Coronaleugner, Esoteriker und Verschwörungstheoretiker“ kann sich sowohl auf die Infektionen als auch auf die Letalität auswirken. Vgl. dazu die Nr. 16, 17, 18 und 21 in der Liste aus Abschnitt 2.1 mit den Nr. 5, 8, 9 und 10 in der Liste aus Abschnitt 2.2.
- Ein dritter Schnittpunkt findet sich bei Strukturunterschieden des Gesundheitssystems. Vgl. dazu die Nr. 18 und 20 in der Liste aus Abschnitt 2.1 mit den Nr. 5, 8, 9, 10 und 11 in der Liste aus Abschnitt 2.2.

Für die Erklärung unterschiedlicher aktueller Inzidenz- oder Letalitätsentwicklungen oder beim Vergleich unterschiedlicher Zeitperioden spielen weitere Faktoren eine Rolle, wie Unterschiede bei

1. dem Anteil der bereits gegen COVID-19 Geimpften,
2. dem Anteil der bereits in der Vergangenheit SARS-COV-2-Infizierten.

3 Statistisch-methodische Vorüberlegungen für eine empirische Untersuchung der Ursachen

Mögliche Teiluntersuchungen Eine empirische Untersuchung zur Identifizierung eines Ursachenbündels, das die regionalen Unterschiede der Inzidenzen und Letalitätsquoten und damit auch der Mortalitäten erklären kann, könnte zunächst in zwei Teiluntersuchungen aufgeteilt werden. In jeweils einer Untersuchung könnten Ursachen unterschiedlicher regionaler Inzidenzen und Ursachen unterschiedlicher regionaler Letalitätsquoten identifiziert werden. Während bei der ersten Untersuchung sozio-kulturelle Faktoren und Verhaltensfaktoren dominieren, stehen bei der zweiten Untersuchung medizinische und gesundheitsökonomische Aspekte im Vordergrund. Im Prinzip könnten die beiden Teiluntersuchungen arbeitsteilig von zwei verschiedenen Wissenschaftlergruppen mit unterschiedlichen Forschungsschwerpunkten durchgeführt werden.

Methodik Angezeigt sind Regressionsmodelle mit der Inzidenz bzw. der Letalität als der durch die Regression erklärten Variablen und den hypothetischen Ursachen als den erklärenden Variablen. Eine erklärte Variable heißt auch Regressand, abhängige Variable oder Response-Variable; die erklärenden Variablen heißen auch Regressoren, unabhängige Variablen oder explanatorische Variablen. Bei der Spezifikation eines Modells ist zu berücksichtigen, dass die erklärten Variablen I und L jeweils nur Werte in einem beschränkten Intervall annehmen. Eine multiple lineare Regression der Form

$$Y = \alpha + \sum_{k=1}^m \beta_k X_k + U$$

mit Y als der erklärten Variablen ist nur approximativ und mit beschränkten Interpretationsmöglichkeiten für die erklärten Variablen I und L anwendbar, da diese nur Werte in den Intervallen $(0,100\,000)$ und $(0,1)$ annehmen. Mit einer als Ergebnis eines statistischen Schätzverfahrens numerisch spezifizierten Parameterstruktur (a, b_1, \dots, b_m) für die Regressionsparameter $(\alpha, \beta_1, \dots, \beta_m)$ resultiert die geschätzte Gleichung

$$\hat{Y} = a + \sum_{k=1}^m b_k X_k .$$

Für bestimmte Werte der erklärenden Variablen X_1, \dots, X_m resultieren auch nicht interpretierbare Werte von \hat{Y} außerhalb dieser Intervalle. Daher scheidet der einfachste Fall einer multiplen linearen Regression aus. Als Ausweg können geeignete nichtlineare Regressionsmodelle formuliert oder geeignete Variablentransformationen der erklärten Variablen durchgeführt werden.

Beispielsweise kann die Inzidenz durch die Transformation $J = I/100\,000$ zunächst auf das Einheitsintervall beschränkt werden. Inhaltlich ist dann J die Inzidenzquote als Anzahl der COVID-19-Fälle pro Einwohner. Durch eine Probit- oder Logit-Transformation der Variablen J und L können transformierte Variablen $g(J)$ und $g(L)$ gebildet werden, die jeweils durch ein multiples lineares Regressionsmodell erklärt werden können. Es resultieren die Modellgleichungen

$$g(J) = \alpha + \sum_{k=1}^m \beta_k X_k + U \quad \text{und} \quad g(L) = \alpha' + \sum_{k=1}^{m'} \beta'_k X'_k + U'$$

für die transformierten Variablen $g(J)$ und $g(L)$. Im Fall einer Probit-Transformation gilt $g(x) = \Phi^{-1}(x)$ für $x \in (0,1)$, wobei Φ^{-1} die Umkehrfunktion der Verteilungsfunktion Φ der Standardnormalverteilung bezeichnet; im Fall einer Logit-Transformation gilt $g(x) = \ln(x/(1-x))$ für $x \in (0,1)$. Der Vorteil dieser Modellierung ist, dass die transformierten Variablen $g(J)$ und $g(L)$ als erklärte Variablen in gewöhnlichen multiplen linearen Regressionen verwendet werden können, deren Parameter mit Standardsoftware geschätzt werden können. Mit geschätzten Parameterwerten a, b_1, \dots, b_m für die erste und a', b'_1, \dots, b'_m für die zweite Modellgleichung erhält man die geschätzten Modellgleichungen

$$\hat{I} = h \left(a + \sum_{k=1}^m b_k X_k \right) \times 100\,000 \quad \text{und} \quad \hat{L} = h \left(a' + \sum_{k=1}^m b'_k X'_k \right).$$

Dabei ist h die Umkehrfunktion von g , es gilt also $h(x) = \Phi(x)$ im Fall einer Probit-Transformation und $h(x) = 1/(1+e^{-x})$ im Fall einer Logit-Transformation.

Die einfachste Form der Zusammenführung der beiden Regressionen für I und L besteht darin, die beiden getrennt geschätzten Modellgleichungen zur Gleichung

$$\hat{M} = \hat{I} \times \hat{L} = h \left(a + \sum_{k=1}^m b_k X_k \right) \times 100\,000 \times h \left(a' + \sum_{j=1}^{m'} b'_j X'_j \right)$$

zu kombinieren.

Falls die beiden Teiluntersuchungen zur Erklärung der Inzidenz I und der Letalität L mit konsistenten Datenstrukturen konzipiert werden, kann auch das zusammengesetzte Modell

$$M = h \left(\alpha + \sum_{k=1}^m \beta_k X_k + U \right) \times 100\,000 \times h \left(\alpha' + \sum_{k=1}^{m'} \beta'_k X'_k + U' \right)$$

für die Mortalität gebildet werden. Dieses Modell ist in den zu schätzenden Regressionsparametern nicht linear, so dass die Standardtheorie linearer Regressionen nicht angewendet werden kann. Schätzwerte für die Parameter können numerisch z. B. durch Maximum-Likelihood-Schätzung bestimmt werden. Dabei sind simultan insgesamt $1 + m + 1 + m' + 3 = 5 + m + m'$ Parameter zu schätzen, da die Kovarianzmatrix der beiden Residualvariablen U und U' durch drei Parameter (zwei Varianzen und eine Kovarianz) bestimmt ist. Bei diesem Vorgehen sind zwar gewisse Effizienzgewinne im Vergleich zur Zusammenführung von zwei Teilschätzungen zu erwarten. Diese Effizienzgewinne sind aber schwierig quantifizierbar und werden mit erheblichen numerischen Komplikationen und dem Verlust von angebbaren Schätzgleichungen für die Parameter erkauft. Es ist auch denkbar, dass gemeinsame erklärende Variablen in beiden Modellteilen Identifizierbarkeitsprobleme erzeugen.

Untersuchungseinheiten und Datenbasis Als Untersuchungseinheiten einer ersten Studie bieten sich die Bundesländer an, wobei dann allerdings lediglich 16 Beobachtungen vorliegen. Dadurch werden im Zusammenhang mit einer größeren Anzahl von Regressoren spezielle methodische Probleme (Anmerkung 11) bei der Schätzung und der Interpretation aufgeworfen, die zu statistischen Methodenkomplexen (Anmerkung 12) führen, die mit elementaren Statistikkenntnissen nicht zu bewältigen sind.

Eine genauere Studie kann sich als Untersuchungseinheiten auf die 401 Kreise (Landkreise, Stadtkreise und kreisfreie Städte) oder die 412 Kreise in der erweiterten Definition

(Anmerkung 13) des RKI stützen, da für diese Einheiten durch das RKI Inzidenz- und Letalitätsdaten vorliegen. Schwierig und aufwendig ist allerdings die Identifikation von Variablen und Indikatoren, welche die oben aufgeführten Ursachenvariablen erfassen und für die zugleich auf Kreisebene Daten vorliegen. Bei einem Regressionsansatz mit Kreisdaten sollte die Zugehörigkeit zu Bundesländern oder eventuell anderen Großregionen durch geeignete Indikatorvariablen (Dummy-Variablen) erfasst werden.

Eine weitere Ausweitung der Datenbasis bestünde in einem Übergang zu Panel-Daten, wobei die Länder- oder Kreisdaten aus aufeinanderfolgenden Zeitpunkten (Tagen, Wochen oder Monaten) verwendet werden. Dieser Ansatz verbreitert zwar die Datenbasis, verursacht aber erhebliche Modellierungsprobleme dadurch, dass die periodischen und singulären Strukturen (Wochenzyklus, Jahreszeitenzyklus; Urlaubszeiten, Großveranstaltungen usw.) der Meldedaten und der Infektionen sowie die Struktur der bisherigen Pandemiewellen (drei große Wellen mit zwei bis drei kleineren Unterwellen) modelliert und bei der statistischen Schätzung berücksichtigt werden müssen. Grob fehlerhaft wäre es, die zeitlichen Daten wie stochastisch unabhängige und identisch verteilte Stichprobenbeobachtungen zu behandeln und dann statistische Standardverfahren anzuwenden.

Bei einer Nutzung zeitlicher Daten, wäre zu erwägen als erklärte Variablen nicht die Niveaus von Inzidenz oder Letalität zu verwenden, sondern deren absolute oder relative Änderungen. Bei einer Erklärung der Änderungen ist die Frage der Variablentransformation neu zu überdenken. Außerdem sind die treibenden Faktoren für Änderungen in der Regel nicht Niveauvariablen, sondern Veränderungsvariablen.

Zeitliche Stabilität der Unterschiede Denkbar ist auch, die Bundesländer bezüglich der Mortalität in den bisherigen drei größeren Infektionswellen zu vergleichen, indem kumulierte Fallzahlen und Todesfallzahlen nicht für den gesamten Zeitraum seit Pandemiebeginn, sondern für einzelne Zeiträume betrachtet werden, die den Infektionswellen zugeordnet werden können. Um zu untersuchen, wie zeitstabil die Mortalitätsunterschiede zwischen den Bundesländern sind, wäre es auch denkbar, die beiden Zeiträume vor und nach der zweiten Kalenderwoche des Jahres 2021 zu vergleichen, auf die jeweils etwa die Hälfte der bisherigen COVID-19-Todesfälle entfällt. Beim Vergleich von Zeiträumen müssen Faktoren wie der Anteil der bereits gegen COVID-19 geimpften Personen und der Anteil derjenigen Personen, die mit SARS-COV-2 infiziert waren, berücksichtigt werden.

Anmerkungen

1. Die zweite Spalte von Tabelle 1 enthält die Zahlen der letzten Spalte von Tabelle 1 aus [RKI], die mit „Todesfälle kumulativ: Fälle/100.000 EW“ überschrieben ist.
2. Die Gesamtzahl der Todesfälle für die Bundesrepublik Deutschland findet sich in der siebten Spalte und letzten Zeile von Tabelle 1 in [RKI].
3. Bei einer angenommenen Covid-19-Mortalität von 56 – das ist die COVID-19-Mortalität in Schleswig-Holstein – anstelle von 110 erhielte man für die gesamte Bundesrepublik

$$91\,527 \times \frac{56}{110} \approx 46\,596$$

Covid-19-Todesfälle, wobei das Ergebnis ganzzahlig gerundet ist. Bei einer angenommenen Covid-19-Mortalität von 249 – das ist die COVID-19-Mortalität in Sachsen – anstelle von 110 erhielte man für die gesamte Bundesrepublik

$$91\,527 \times \frac{249}{110} \approx 207\,184$$

Covid-19-Todesfälle, wobei das Ergebnis ganzzahlig gerundet ist.

4. Wenn für eine bestimmte Krankheit mit K das Ereignis bezeichnet ist, zu erkranken, mit S das Ereignis bezeichnet ist, an der Krankheit sterben, und somit $K \cap S$ das Ereignis ist, zu erkranken und an dieser Krankheit zu sterben, dann besteht für die entsprechenden Wahrscheinlichkeiten der Zusammenhang

$$P(K \cap S) = P(K)P(S|K) .$$

Dabei ist die bedingte Wahrscheinlichkeit $P(S|K)$ die Wahrscheinlichkeit zu sterben, falls man erkrankt ist. Die Wahrscheinlichkeit, an der Krankheit zu sterben ist also das Produkt aus einer unbedingten und einer bedingten Wahrscheinlichkeit. Diese Faktorisierung entspricht dem Modell eines zweistufigen Zufallsexperimentes, bei dem sich auf der ersten Stufe aus einem Bernoulli-Experiment ergibt, ob es zu einer Erkrankung kommt oder nicht, und bei dem im Fall einer Erkrankung ein zweites Bernoulli-Experiment darüber entscheidet, ob es zu einem Todesfall kommt oder nicht.

5. Die dritte Spalte von Tabelle 1 enthält die Zahlen der vierten Spalte in Tabelle 1 aus [RKI], die mit „Fälle kumulativ: Fälle/100.000 EW“ überschrieben ist.
6. Die Gesamtzahl der Fälle und der Todesfälle für die Bundesrepublik Deutschland findet sich in der letzten Zeile von Tabelle 1 in [RKI].
7. Die Zahlen für die Letalitätquote L in der vierten Spalte von Tabelle 1 erhält man mit Tabelle 1 aus [RKI], indem man die Anzahl der Todesfälle aus der siebten Spalte durch die Anzahl der Fällen aus der zweiten Spalte dividiert. Alternativ, aber durch Rundungsfehler etwas ungenauer, erhält man die Letalitätquote, indem man die Todesfälle pro 100 000 Einwohner aus der letzten Spalte durch die Fälle pro 100 000 Einwohner aus der vierten Spalte dividiert.
8. Wegen Gleichung (1) gilt auch

$$\frac{\text{Todesfälle}}{\text{Einwohnerzahl}} \times 100\,000 = \frac{\text{Fälle}}{\text{Einwohnerzahl}} \times 100\,000 \times \frac{\text{Todesfälle}}{\text{Fälle}}$$

und somit der Zusammenhang

$$M = I \times L$$

zwischen der Mortalität

$$M = \frac{\text{Todesfälle}}{\text{Einwohnerzahl}} \times 100\,000 ,$$

der Inzidenz

$$I = \frac{\text{Fälle}}{\text{Einwohnerzahl}} \times 100\,000$$

und der Letalitätquote

$$L = \frac{\text{Todesfälle}}{\text{Fälle}} .$$

9. Als in der ersten Pandemiewelle die neuen Bundesländer schwächer betroffen waren, wurde die These aufgestellt, dies könne mit den in der DDR üblichen Tuberkuloseimpfungen zusammenhängen.
10. Es gab die These, dass Vorinfektionen mit den verschiedenen bereits seit langer Zeit endemischen Coronaviren eine Rolle durch eine teilweise Immunität spielen könnten.
11. Die Multikollinearität von erklärenden Variablen führt zur Nichtidentifizierbarkeit einzelner Parameter bei gleichzeitiger Identifizierbarkeit von linearen Teilräumen des Parameterraumes. Die Relevanz der Multikollinearität ist unterschiedlich einzuschätzen je nachdem, ob die Schätzung einzelner Parameter oder die Prognose der erklärten Variable durch Linearkombinationen erklärender Variablen im Vordergrund steht.
12. In Frage kommen z. B. die Hauptkomponentenregression, Lasso-Verfahren und die Indexbildung mit Hilfe der Hauptkomponentenanalyse.
13. Vom 25. März 2020 bis zum 16. Juni 2020 wurde die Stadt Berlin vom RKI in den täglichen Lageberichten als ein Kreis ausgewiesen. Seit dem 17. Juni 2020 weist das RKI die 12 Stadtbezirke von Berlin als eigene Kreise aus, so dass seitdem in den täglichen Lageberichten die 7-Tage-Inzidenzen für 412 Kreise vorliegen.

Literatur

[RKI] Robert Koch-Institut, Täglicher Lagebericht des RKI zur Coronavirus-Krankheit 2019 (COVID-19): 26.07.2021 – Aktualisierter Stand für Deutschland, https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Jul_2021/2021-07-26-de.pdf

Tabellenverzeichnis

1	COVID-19-Mortalität M , COVID-19-Inzidenz I und COVID-19-Letalitätsquote L in % für deutsche Bundesländer am 26. Juli 2021	2
2	Bundesländer klassifiziert nach COVID-19-Inzidenz I und COVID-19-Mortalität M am 26. Juli 2021	4
3	Bundesländer klassifiziert nach COVID-19-Letalitätsquote L in % und COVID-19-Mortalität M am 26. Juli 2021	5
4	Bundesländer klassifiziert nach COVID-19-Inzidenz I und COVID-19-Letalitätsquote L in % am 26. Juli 2021	5