Modellierung und Modellbildung

Description

â?¦am Beispiel unserer Modellierstudie: â??Identifizierung von verschiedenen Faktoren, die mit Covid-19 TodesfÃ¤llen in Europa wÃ¤hrend der ersten Pandemie-Welle assoziiert sindâ??

Eine groÃ?e Gruppe statistischer Verfahren, die der ErklÃ¤rung vergangener und auch der Vorhersage kÃ¼nftiger Daten dienen soll, ist die Modellbildung oder statistische Modellierung. Damit meint man, dass man fÃ¼r einen gegebenen Datensatz mit sehr unterschiedlichen Variablen eine mathematische Struktur findet, die diesen Datensatz mÃ¶glichst gut abbildet, zunÃ¤chst einmal rein formal. Dieses Verfahren kann man nÃ¼tzen, um den Einfluss verschiedener Variablen auf eine Ergebnisvariable zu untersuchen. In der Sprachregelung der Modellbildung ist die Variable, die man erklÃ¤ren will, die abhÃ¤ngige Variable oder das Kriterium oder die Ergebnisvariable (Outcome), und die verschiedenen Variablen, die zur AufklÃ¤rung dieser einen Variable beitragen sollen, sind mehrere unabhÃ¤ngige Variablen oder auch PrÃ¤diktoren.

Ich verwende unsere eben publizierte Modellierstudie [1] als konkretes Beispiel. Sie wurde von mir konzipiert, ich habe erste Analysen gerechnet, dann stieg mein Kollege Rainer J. Klement ein, der als Physiker wesentlich flinker im Umgang mit solchen Modellen ist als ich.

In dieser Studie haben wir die TodesfÃ¤lle der Menschen, die an Covid-19 in der ersten Welle der SARS-CoV2 Pandemie 2020 gestorben sind als abhÃ¤ngige Variable (Outcome, Ergebnisvariable oder Kriteriumsvariable) verwendet. Wir wollten wissen: Welche EinflÃ¼sse auf diese Variable sind besonders wichtig, welche eher vernachlÃ¤ssigbar, und wie viel der Schwankungsbreite kÃ¶nnen wir mit den Variablen, die wir als PrÃ¤diktoren ausgesucht haben, erklÃ¤ren. Technisch spricht man dabei von der AufklÃ¤rung von Varianz dieser Ergebnisvariable. Oder, nochmals anders ausgedrÃ¼ckt: Welche EinflÃ¼sse tragen zu den Schwankungen der Covid-19 TodesfÃ¤lle in den verschiedenen europÃ¤ischen LÃ¤ndern bei?

Dass die TodesfÃ¤lle an und mit SARS-CoV2 nicht einfach durch das Virus allein erklÃ¤rbar sind, zeigt schon die simple Tatsache, dass die auf BevÃ¶lkerung standardisierten Sterberaten in allen europÃ¤ischen LÃ¤ndern sehr unterschiedlich sind. WÃ¤re das Virus allein die Ursache dafÃ¼r, dann mÃ¼ssten auf 100.000 Personen standardisierte Sterbezahlen zu einem Stichtag etwa gleich groÃ? sein. Das sind sie nicht. Das wurde recht schnell klar. Ebenfalls klar geworden ist rasch, dass es eine Vielzahl anderer GrÃ¼nde fÃ¼r die unterschiedlichen Covid-19-Todesraten in einem Land geben muss, auÃ?er dem Virus. Denn, kleines Beispiel: Belgien, das eine Grenze mit Deutschland hat, hatte in der ersten Welle die mit Abstand hÃ¶chsten Covid-19-Todesraten und Deutschland mit Abstand die kleinsten. WÃ¤ren alle TodesfÃ¤lle ausschlieÃ?lich auf das Virus zurÃ¼ckzufÃ¼hren und wÃ¤re die SARS-CoV2 Epidemie die einzige Ursache fÃ¼r TodesfÃ¤lle, dann mÃ¼ssten die TodesfÃ¤lle in beiden LÃ¤ndern â?? immer auf die gleiche Anzahl von Menschen gerechnet â?? gleich hoch sein. Denn ein Virus lÃ¤sst sich nicht von einer Grenze aufhalten, zumal alle â??MaÃ?nahmenâ?? sowieso zu spÃ¤t kamen.

Das also ist die Motivation fÃ¼r diese konkrete Modellierstudie. FÃ¼r alle Modellierungen im Allgemeinen gilt: Wir beobachten Schwankungen, technisch gesprochen Varianz, in einem bestimmten Parameter oder in einer

Â© Prof. Harald Walach

bestimmten Variablen, in diesem Falle in der Zahl der Covid-19 TodesfÃ¤lle pro Land. Wir wollen wissen, wie diese Schwankung zustande kommt. Anders ausgedrÃ¼ckt: Wir wollen wissen, welche mÃ¶glichen Variablen an dieser Schwankung beteiligt sind oder welche Variablen einen Einfluss darauf haben.

Wer das Grundprinzip der Modellierung nicht kennt, kÃ¶nnte sich vorab meinen Methodenblog zum Thema â??Modellierungâ?? ansehen. Dann ist leichter verstÃ¤ndlich, was wir in dieser Modellierstudie gemacht haben, in der wir versucht haben zu verstehen, welche Variablen einen Einfluss auf die Variation der Covid-19 TodesfÃ¤lle in der ersten SARS-CoV2-Welle Anfang 2020 haben.

Die Modellierung von Variablen, die Covid-19 TodesfÃ¤lle in Europa mÃ¶glicherweise beeinflussen

Unsere Kriteriums- oder Zielvariable war also die Rate der Covid-19 TodesfÃ¤lle in jedem europÃ¤ischen Land wÃ¤hrend der ersten Welle der SARS-CoV2-Pandemie. Wir verwendeten die Daten von 43 europÃ¤ischen LÃ¤ndern zum Stichtag 20. August 2020. Dieser Tag war der Tag, an dem die erste Welle so ziemlich Ã¼berall vorÃ¼ber war und wurde einigermaÃ?en pragmatisch gewÃ¤hlt. Wir verwendeten die bis dorthin gezÃ¤hlten TodesfÃ¤lle, die Covid-19 zugeschrieben wurden, aus der Ã¶ffentlich verfÃ¼gbaren Datenbank, wie sie die Webseite â??Our World in Dataâ?? (OWID) anbietet.

Die Verteilung dieser Variablen ist nicht normalverteilt, sondern folgt einer Gamma-Verteilung relativ genau (siehe Abbildung 1, links). FÃ¼hrt man eine logarithmische Transformation dieser Variablen durch, dann folgt sie einigermaÃ?en brauchbar einer Normalverteilung (Abb. 1 rechts).

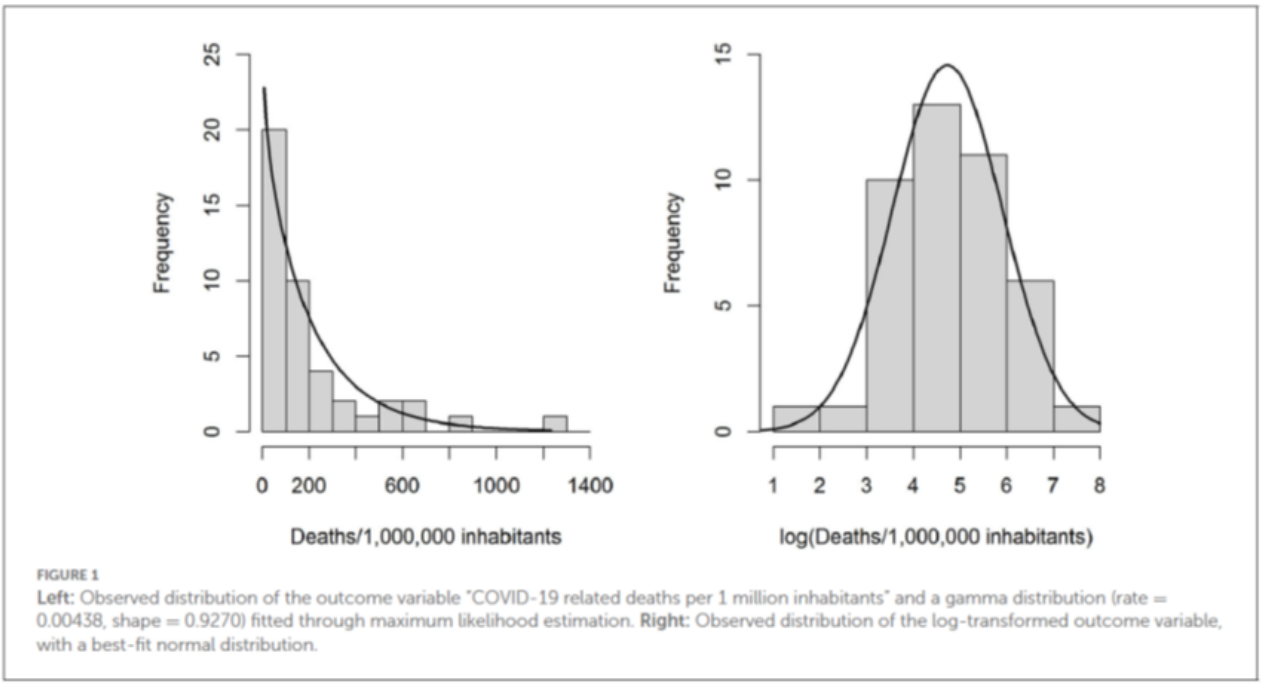

Abbildung 1 â?? Abbildung 1 der Originalpublikation mit der Verteilung der Covid-19 TodesfÃ¤lle pro 100.000 Einwohner in 43 europÃ¤ischen LÃ¤ndern bis zum 20.8.2020. Links die originale Verteilung, die einer Gamma-Verteilung nahekommt, rechts die einigermaÃ?en einer Normalverteilung folgende log-transformierte Verteilung

Â© Prof. Harald Walach

Wir wÃ¤hlten daher einen doppelten analytischen Ansatz, wobei der erste im Protokoll als Hauptanalyse formuliert wurde, der zweite als SensitivitÃ¤tsanalyse.

Protokoll

Protokolle sind Ã¼brigens Projektbeschreibungen, in denen die wichtigsten Auswertungs- und Analyseschritte definiert werden, *bevor* irgendwelche Auswertungen durchgefÃ¼hrt werden. Ich formuliere standardmÃ¤Ã?ig immer Protokolle und stelle sie *vor Beginn* einer Untersuchung Ã¶ffentlich zur VerfÃ¼gung, meistens auf der Open Science Foundation Plattform. Bei klinischen Studien ist dies seit einiger Zeit Standardvorgabe und dort werden Protokolle in entsprechenden Studienregistern in einer Datenbank gefÃ¼hrt (z.B. clinicaltrials.gov), damit man sich am Ende davon Ã¼berzeugen kann, dass sich die Autoren an die Vorgaben hielten. Auch in diesem Fall habe ich ein Protokoll erstellt (https://osf.io/x93np/; dort befindet sich auch ein Update und ein Analyseskript). Mittlerweile haben sich einige Verbesserungen ergeben, die wir in der Publikation erwÃ¤hnt haben.

Analytische Strategie

In diesem Fall diente uns die Modellierung, um zu verstehen, ob und welche Variablen einen Einfluss auf die Todesraten haben. Daher formulierten wir unterschiedliche Modelle mit unterschiedlicher KomplexitÃ¤t. Wir rechneten generalisierte lineare Modelle auf eine gamma-verteilte Kriteriumsvariable. Solche Modelle kann man verstehen wie einfache lineare Modelle, mit dem Unterschied, dass die Variablen nicht mit einem einfachen additiv-linearen Link verbunden sind, sondern als logarithmische Funktion miteinander verknÃ¼pft werden. Weil sie etwas komplexer sind, kann man ihre Ergebnisse auch nicht einfach direkt interpretieren. Zwar kann man auch fÃ¼r sie einen R^2-Ã¤hnlichen Kennwert, den sog. Kullback-Leibler R^2-Wert errechnen, den man als Prozentsatz der aufgeklÃ¤rten Varianz interpretieren kann. Aber damit die Daten leichter interpretierbar sind, haben wir auch ein lineares Modell auf eine log-transformierte Zielvariable gerechnet. Die Ergebnisse sind sehr Ã¤hnlich, was zeigt, dass die Modellbildung erfolgreich war.

Das einfachste Modell bildet die Hypothese ab: AusschlieÃ?lich die Fallzahlen (selbstverstÃ¤ndlich standardisiert auf die Anzahl der Tests) sind fÃ¼r TodesfÃ¤lle verantwortlich. Je mehr FÃ¤lle, umso mehr Tote. Weil diese Variable sehr schief verteilt ist, haben wir sie logarithmiert. (Dies ist Ã¼brigens eine VerÃ¤nderung gegenÃ¼ber meiner ursprÃ¼nglichen Analyse und Beschreibung [2]. Logarithmiert man nÃ¤mlich diese Variable, dann wird das Modell besser und die VarianzaufklÃ¤rung steigt. Waren es mit den rohen Werten nur 7 % VarianzaufklÃ¤rung, sind es mit den logarithmierten Werten 20 % VarianzaufklÃ¤rung).

MaÃ?nahmen

Eine weitere interessierende Variable ist der Einfluss der â??MaÃ?nahmenâ??, technisch â??non-pharmaceutical interventionsâ?? (NPIs), im Volksmund gerne â??Lockdownâ?? genannt. In Wirklichkeit bestehen diese â??MaÃ?nahmenâ?? aus einem BÃ¼ndel von mÃ¶glichen Interventionen, von SchulschlieÃ?ungen, Ã¼ber GrenzschlieÃ?ungen, Ausgangssperren, LadenschlieÃ?ungen, Maskenpflicht, SchlieÃ?en von kulturellen RÃ¤umen und Restaurants, bis hin zur strikten Order zu Hause zu bleiben. Die Blavatnik-School of Policiy der UniversitÃ¤t Oxford stellt einen Tracker zur VerfÃ¼gung, der fÃ¼r alle LÃ¤nder dieser Erde die Striktheit dieser MaÃ?nahmen beschreibt in einem auf 100 % genormten Wert (0: keinerlei MaÃ?nahmen; 100: maximal denkbare MaÃ?nahmen). Dieser Wert Ã¤ndert sich und wird wÃ¶chentlich adaptiert. In unserer Analyse taucht diese Variable als Government Response Severity Index (GRSI) auf, und zwar als mittlere Schwere in einem Land Ã¼ber den interessierenden Zeitraum.

Das zweite Modell stellte also die Wirksamkeit der politischen Reaktionen in den Vordergrund.

Â© Prof. Harald Walach

Grippeimpfungen, Vitamin-D-Status, Krankenhausbetten

Das dritte mÃ¶gliche Modell verwendet neben der Fallzahl einige medizinische Variablen, nÃ¤mlich den Prozentsatz der Ã¤lteren BevÃ¶lkerung, der gegen Grippe geimpft ist. Das wurde motiviert durch eine interessante Studie, die einen positiven Zusammenhang zwischen Grippe-Impfung und Covid-19-Todesraten fand [3]. Wir nahmen in dieses Modell als mÃ¶glichen Protektiv-Faktor, der durch verschiedene Studien erhÃ¤rtet war, die Vitamin-D-Versorgung der BevÃ¶lkerung auf, die wir aus unterschiedlichen Studien zusammengestellt hatten, sowie die Anzahl der Krankenhausbetten auf 100.000 Einwohner.

Populationsparameter

Weil vor allem Ã¤ltere Menschen von Covid-19 betroffen sind, formulierten wir ein Modell, in dem neben der Fallzahl und dem Vitamin-D-Status die Lebenserwartung in einem Land und die Anzahl der Ã¤lteren BevÃ¶lkerung als PrÃ¤diktoren verwendet wurden.

Risikofaktoren

Risikofaktoren, wie koronare Herzkrankheit und Diabetes, sowie Rauchen als mÃ¶glicher Risiko- oder gar Protektivfaktor [4, 5] gingen in ein weiteres Modell ein, zusÃ¤tzlich zu Fallzahl und Vitamin-D-Status.

LÃ¤nderspezifische Faktoren

SchlieÃ?lich bildeten wir noch ein Modell, bei dem neben den log-transformierten, standardisierten FÃ¤llen und dem Vitamin-D-Status noch lÃ¤nderspezifische Faktoren wie BevÃ¶lkerungszahl und -dichte, Prozent der Alten, Bruttoinlandsprodukt und Entwicklungsindex eingingen.

Volles Modell und datengesteuertes Modell via LASSO-Analyse

SchlieÃ?lich analysierten wir noch das volle Modell mit allen 13 Variablen und rechneten ein datengetriebenes Modell mit einer relativ neuartigen Analysemethode, die abgekÃ¼rzt LASSO heiÃ?t [6]. LASSO ist eine AbkÃ¼rzung fÃ¼r â??Least absolute shrinkage and selection operatorâ??, eingedeutscht: ein Operator, der die am wenigsten bedeutsamen PrÃ¤diktoren auf null setzt und diejenigen selektiert, die dann an Bedeutung gewinnen.

Das volle Modell ist einfach zu verstehen: Es verwendet einfach alle Variablen. Das LASSO-Modell ist datengesteuert. Hier werden in einem iterativen Verfahren alle Variablen auf null gesetzt, bis auf eine. Von diesen wird zunÃ¤chst diejenige gewÃ¤hlt, die den gewichtigsten Beitrag zur VarianzaufklÃ¤rung leistet. Dann wird dieser iterative Prozess fÃ¼r die nÃ¤chsten Ã¼brig-gebliebenen Variablen wiederholt, usw. Dies ist zunÃ¤chst ein rein exploratives Verfahren, das aber den Vorteil hat, dass es ohne eine multiple Modellbildung auskommt. Auf diese Art und Weise kann man erkennen, welche Variablen wirklich wichtig sind und formuliert dann fÃ¼r diese so ausgewÃ¤hlten Variablen ein Modell.

Auswahl der Modelle und das Akaike-Informationskriterium

Diese 8 Modelle ergeben unterschiedlich gute Anpassungen an die Daten. Wie entscheidet man nun? Hier hilft das sog. Akaike-Informationskriterium (AIC), das der japanische Statistiker Akaike vorgeschlagen hat. Es ist eine absolute und fÃ¼r sich genommen wenig aussagekrÃ¤ftige Zahl, in die die AnpassungsgÃ¼te eines Modells und die Anzahl der dafÃ¼r notwendigen Parameter eingeht. Ein Modell mit grÃ¶Ã?erer Parameteranzahl wird â??bestraftâ??, indem das AIC ansteigt. Daher kann man das AIC verwenden, um die relative ModellgÃ¼te zu bestimmen.

Â© Prof. Harald Walach

Aus einer Klasse von Modellen ist dasjenige das Beste, das das vergleichsweise niedrigste AIC aufweist.

Wir haben die Analyseergebnisse und die entsprechenden Modellparameter in Tabellen 1 und 2 der Originalpublikation dargestellt.

Ich gebe hier in Tabelle 1 nur die wichtigsten Ergebnisse wieder.

Modell No und Variablen im Modell	**Regressions-gewicht b**	**Signifikanz von b p-Wert**	**KL-R^2**	**AIC**	**Delta AIC**
Modell 1			*0.2*	*541.3*	*36.2*
FÃ¤lle	0.6	.0008			
Modell 2			*0.26*	*538.5*	*33.4*
FÃ¤lle	0.6	.0005			
GRSI	0.04	.035			
Modell 3			*0.47*	*524.8*	*19.7*
FÃ¤lle	0.73	$2*10^{-5}$			
Vitamin D	-0.3	.4			
Betten	0.025	.7			
Grippeimpfung	0.032	$3*10^{-5}$			
Modell 4			*0.45*	*527.5*	*22.43*
FÃ¤lle	0.73	$3*10^{-6}$			
Vitamin D	-0.47	.11			
Lebenserwartung	0.19	$4 * 10^{-5}$			
Anteil Alter	-0.033	.5			
Modell 5			*0.49*	*524.9*	*19.8*
FÃ¤lle	0.71	$8*10^{-5}$			
Vitamin D	-0.36	.2			
Raucher (%)	-0.007	.8			
CVD TodesfÃ¤lle	-0.999	.02			
Diabetes PrÃ¤valenz	-0.09	.18			
Modell 6			*0.48*	*527.5*	*22.42*
FÃ¤lle	0.96	$6*10^{-6}$			
BevÃ¶lkerungdichte	0.22	.058			
Lebenserwartung	-0.011	.88			
BPI	$2.6*10^{-5}$	.09			
HDI	1.94	.7			
%Alte	0.08	.2			
Modell 7			*0.67*	*524.2*	*19.1*
FÃ¤lle	0.97	$2*10^{-6}$			
GRSI	0.03	.055			
Vitamin D	-0.716	.043			

Â© Prof. Harald Walach

Grippe Impfungsrate	0.02	.0055		
Lebenserwartung	-0.05	.7		
BevÃ¶lkerungsdichte	0.09	.4		
%Raucher	-0.014	.5		
CVD MortalitÃ¤t	-0.04	.9		
Diabetes PrÃ¤valenz	-0.06	.4		
Betten	0.03	.6		
BPI	$4*10^{-8}$	.005		
HDI	-2.5	.6		
%Alte	0.09	.14		
Modell 8			*0.68*	*505.1 â??*
FÃ¤lle	0.84	$9*10^{-7}$		
GRSI	0.02	.07		
Vitamin D	-0.7	.02		
Grippe Impfungsrate	0.02	.0002		
BevÃ¶lkerungsdichte	0.13	.2		
BPI	$3*10^{-5}$	.007		

Tabelle 1 â?? Ergebnis der Modellierung von Covid-19 TodesfÃ¤llen wÃ¤hrend der ersten Welle in Europa: Modelle, Variablen im Modell, Regressionsgewicht b, Signifikanz von b, Kullback-Leibler R^2, Akaike Information Criterion AIC, Unterschied des AIC eines Modells zum besten Modell 8 (Delta > 15 signalisieren, dass das Modell nicht plausibel ist)

FÃ¤lle: Anzahl der FÃ¤lle auf 100.000 Tests standardisiert (Variable log-transformiert); GRSI: Government Response Index (HÃ¤rte der MaÃ?nahmen), Vitamin D (ausreichend oder nicht), Betten: Anzahl Krankenhausbetten fÃ¼r 100.000 Bewohner; Grippeimpfung: Durchimpfungsrate mit Grippeimpfung bei der alten BevÃ¶lkerung (i.d.R. Ã¼ber 65 Jahre), CVD: kardiovaskulÃ¤re Erkrankung (log-transformiert); Anteil Alter: % der Ã¼ber 70-JÃ¤hrigen in einer Population; BPI: Bruttoinlandsprodukt; HDI: Human Development Index

Die lineare Regression auf die log-transformierte Kriteriumsvariable erbrachte wie gesagt fast die gleichen Ergebnisse, daher gehe ich darauf nicht ein. Wer mag, kann sich die Originaltabelle ansehen. AuÃ?erdem sei noch erwÃ¤hnt: Wir haben fehlende FÃ¤lle durch einen Interpolationsalgorithmus ersetzt, weil wir nicht fÃ¼r alle Variablen fÃ¼r jedes Land Daten hatten und weil bei fehlenden Werten ein Fall von der Analyse ausgeschlossen wird, was die statistische MÃ¤chtigkeit reduziert. Um zu kontrollieren, ob diese Prozedur einen Einfluss hatte, haben wir die Analyse fÃ¼r die Daten angepasst, fÃ¼r die alle FÃ¤lle vorlagen und erhielten ziemlich die gleichen Ergebnisse; auch das stelle ich hier nicht extra dar. Die Tabelle befindet sich in der Originalpublikation.

Widmen wir uns nun den Modellen und was sie uns sagen. Beginnen wir mit dem schlechtesten und einfachsten, Modell 1, bei dem nur die standardisierte Fallzahl als PrÃ¤diktor verwendet wird. Dieses Modell testet die Hypothese, ob die Anzahl der FÃ¤lle als PrÃ¤diktor ausreicht. Offensichtlich ist das nicht der Fall. Sowohl die Modellpassung, erfasst durch das Akaike Informationskriterium, als auch die VarianzaufklÃ¤rung sind am schlechtesten bei diesem Modell. Die Differenz des AIC zum besten Modell ist am grÃ¶Ã?ten, nÃ¤mlich 36.2.

Â© Prof. Harald Walach

Man geht in der Regel davon aus, dass AIC-Differenzen unter 5 in etwa gleichwertige Modelle signalisieren und Ã¼ber 15 ein Modell als unplausibel ausscheiden lassen.

Das beste Modell ist eindeutig das durch die LASSO-Regression gefundene datengestÃ¼tzte Modell Nummer 8. Es kann mit nur 6 Variablen mehr Varianz aufklÃ¤ren, nÃ¤mlich 68 %, als das Modell, das alle Variablen nutzt, nÃ¤mlich Modell 7. Dieses hat mit einer AIC-Differenz von 19.1 Punkten auch einen wesentlich schlechteren Wert. Das Gleiche gilt fÃ¼r alle anderen Modelle, die ich deshalb gar nicht erst diskutiere.

Die Anzahl der FÃ¤lle geht auch im besten Modell Nummer 8 als signifikanter und stÃ¤rkster PrÃ¤diktor ein. Der zweitstÃ¤rkste PrÃ¤diktor ist die Versorgung der BevÃ¶lkerung mit Vitamin D. Dieser PrÃ¤diktor ist negativ, also protektiv und signifikant. Die Daten zur Vitamin-D-Versorgung sind an sich schlecht, weil es wenig gute populationsgestÃ¼tzte Studien fÃ¼r alle europÃ¤ischen LÃ¤nder gibt. Daher haben wir die Variable dichotomisiert in ausreichend oder nicht-ausreichend. Das ist natÃ¼rlich extrem grobkÃ¶rnig. Umso erstaunter waren wir, dass wir trotzdem einen relativ starken Effekt gefunden haben. Wir haben die Analyse mit der geografischen Breite nachvollzogen und sahen, dass diese allein den Effekt nicht erklÃ¤rt. Der PrÃ¤diktor, der als Drittes kommt, gemessen an der GrÃ¶Ã?e des Regressionsgewichtes, ist die BevÃ¶lkerungsdichte: je grÃ¶Ã?er, umso mehr FÃ¤lle. Allerdings ist dieser PrÃ¤diktor nicht signifikant, spielt also trotz seiner GrÃ¶Ã?e eine untergeordnete Rolle.

Die Rate der Durchimpfung der Ã¤lteren BevÃ¶lkerung mit Grippeimpfung ist ein hochsignifikanter PrÃ¤diktor, auch wenn der Einfluss eher klein ist: je mehr grippegeimpfte Alte in einem Land, umso mehr Covid-19-TodesfÃ¤lle.

Wie lÃ¤sst sich das verstehen? Zum einen ist denkbar, dass es noch immer Hintergrundvariablen gibt, die wir nicht erfasst haben, die diesen Einfluss treiben. Diejenigen, die wir erfasst haben, spielen darin jedenfalls keine Rolle. Denn die Variable ist selbst im vollen Modell, in das alle Variablen eingehen, ein signifikanter PrÃ¤diktor und in allen anderen Modellen, in denen wir sie eingebracht haben. Sie hÃ¤ngt univariat nur mit der Lebenserwartung zusammen (r = .56) und mit der kardiovaskulÃ¤ren MortalitÃ¤t negativ (r = -.57), also je mehr Grippeimpfungen, umso weniger kardiovaskulÃ¤re MortalitÃ¤t in einem Land. Aber: je mehr Grippeimpfungen, umso mehr Covid-19 MortalitÃ¤t. Wir haben anscheinend die Wahl zwischen Pest und Cholera sozusagen. Gegen alles gleichzeitig scheinen wir uns nicht schÃ¼tzen zu kÃ¶nnen. Zum einen ist da das PhÃ¤nomen, dass Erreger offenbar eine Art symbiotisches Ã?kosystem bilden. Vertreibt man einen, kommen eben andere. Zum anderen kÃ¶nnte es sein, dass die Impfung gegen Grippe das Immunsystem eine Weile beschÃ¤ftigt. [7] In dieser Zeit hat es vielleicht ein anderer Erreger leichter, das Immunsystem zu Ã¼berwinden. Eine sehr gute randomisierte Studie an Kindern zeigte: Gegen Grippe geimpfte Kinder hatten zwar deutlich weniger Risiko, an Grippe zu erkranken, aber ein vierfach gesteigertes Risiko, an anderen Atemwegsinfekten zu erkranken. [8]

Das Bruttoinlandsprodukt ist in diesem Modell ein typisches Beispiel fÃ¼r einen hochsignifikanten, aber sachlich wenig wichtigen PrÃ¤diktor, denn sein numerischer Wert ist extrem klein, nÃ¤mlich 0,00003: je mehr ein Land erwirtschaftet, umso grÃ¶Ã?er die Covid-19 MortalitÃ¤t in diesem Land.

Am interessantesten aus konzeptueller Sicht ist fÃ¼r mich die Tatsache, dass der Government Response Index GRSI, der die HÃ¤rte der MaÃ?nahmen abbildet, ein zwar schwacher, aber *positiver* PrÃ¤diktor ist. D.h.: je stÃ¤rker die MaÃ?nahmen, umso mehr TodesfÃ¤lle. HÃ¤tten die MaÃ?nahmen in der ersten Welle irgendetwas Positives bewirkt, hÃ¤tten wir hier ein negatives Vorzeichen erwartet. Genau das ist nicht der Fall. Zwar ist der PrÃ¤diktor auch nicht gerade sonderlich signifikant, spielt also auch keine groÃ?e Rolle, aber wenn schon, dann eine unrÃ¼hmliche. Wir hatten schon in einer anderen Publikation gezeigt, dass die unselige Modellierstudie, die angeblich belegt, dass Deutschlands Lockdown nÃ¶tig war [9], eine falsche Datengrundlage verwendete [10]. Die Datenbasis fÃ¼r die Wirksamkeit von Lockdowns und MaÃ?nahmen ist aus meiner Sicht schlecht. Diese neue Analyse erhÃ¤rtet diesen Befund.

Â© Prof. Harald Walach

Zusammenfassung unserer Covid-19 MortalitÃ¤ts-Modellierung

Wir wissen jetzt also: in der ersten Phase der SARS-CoV2-Ausbreitung waren die Fallzahlen, standardisiert auf die Anzahl der Tests, ein wichtiger PrÃ¤diktor, der etwa 20 % der Varianz erklÃ¤rt. Allerdings kann das beste Modell mit 6 Variablen insgesamt 68 % der Varianz erklÃ¤ren. Die HÃ¤rte der MaÃ?nahmen spielt eine Rolle: je hÃ¤rter die MaÃ?nahmen, umso mehr TodesfÃ¤lle; je dichter die BevÃ¶lkerung zusammenwohnt, umso mehr TodesfÃ¤lle; je besser die Versorgung mit Vitamin D, umso weniger TodesfÃ¤lle; je hÃ¶her die Durchimpfungsrate mit Grippeimpfung bei den Ã?lteren, umso mehr TodesfÃ¤lle; je hÃ¶her das BIP, umso mehr TodesfÃ¤lle. Dabei bleiben 32 % der Varianz ungeklÃ¤rt. Wir wissen, dass die berÃ¼hmten Vorerkrankungsvariablen (PrÃ¤valenz von Diabetes, von kardiovaskulÃ¤ren Erkrankungen und von Rauchen) dabei keine groÃ?e Rolle spielen; denn Modelle, in die wir diese eingebracht haben, sind weniger gut in der AufklÃ¤rung der Varianz. Aber es kÃ¶nnte natÃ¼rlich sein, dass wir andere wichtige Variablen Ã¼bersehen haben: wie glÃ¼cklich ein Land im Durchschnitt ist, wie hoch das Angstniveau ist oder die Depressionsrate. Diese haben wir nicht Ã¼berprÃ¼ft. Das kÃ¶nnen andere gerne noch tun. Unser Datensatz steht zur VerfÃ¼gung. Aber ich finde: 68 % VarianzaufklÃ¤rung ist schon ganz schÃ¶n gut. Und wenn man diese Information ernst nehmen wÃ¼rde und zum Beispiel eine flÃ¤chendeckende Kampagne zur ausreichenden Vitamin-D-Versorgung der BevÃ¶lkerung durchfÃ¼hren wÃ¼rde, oder jedem Haushalt entsprechendes Material zur VerfÃ¼gung stellen wÃ¼rde, dann wÃ¤re das garantiert effektiver als Maskenpflicht in ZÃ¼gen und Schulen. Soviel kann man sagen. Denn die Maskenpflicht ist Teil des GRSI und der spielt eher eine unrÃ¼hmliche Rolle.

Quellen und Literatur

1. Klement RJ, Walach H. Identifying factors associated with Covid-19 related deaths during the first wave of the pandemic in Europe. Frontiers in Public Health. 2022;6th July 2022. doi: https://doi.org/10.3389/fpubh.2022.922230.
2. Klement RJ, Walach H. Low Vitamin D Status and Influenza Vaccination Rates are Positive Predictors of Early Covid-19 Related Deaths in Europe â?? A Modeling Approach. Zenodo. 2021. doi: https://doi.org/10.5281/zenodo.4680691.
3. EBMPHET Consortium. COVID-19 Severity in Europe and the USA: Could the Seasonal Influenza Vaccination Play a Role? SSRN. (7/6/2020). doi: https://doi.org/10.2139/ssrn.3621446
4. Patanavanich R, Glantz SA. Smoking is Associated with COVID-19 Progression: A Meta-Analysis. medRxiv. 2020:2020.04.13.20063669. doi: https://doi.org/10.1101/2020.04.13.20063669.
5. Farsalinos K, Eliopoulos E, Leonidas DD, Papadopoulos GE, Tzartos S, Poulas K. Nicotinic Cholinergic System and COVID-19: In Silico Identification of an Interaction between SARS-CoV-2 and Nicotinic Receptors with Potential Therapeutic Targeting Implications. International Journal of Molecular Sciences. 2020;21(16):5807. PubMed PMID: doi: https://doi.org/10.3390/ijms21165807.
6. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society B. 1996;58:267-88.
7. Cowling BJ, Nishiura H. Virus Interference and Estimates of Influenza Vaccine Effectiveness from Test-Negative Studies. Epidemiology. 2012;23(6):930-1. doi: https://doi.org/10.1097/EDE.0b013e31826b300e. PubMed PMID: 00001648-201211000-00030.
8. Cowling BJ, Fang VJ, Nishiura H, Chan K-H, Ng S, Ip DKM, et al. Increased risk of noninfluenza respiratory virus infections associated with receipt of inactivated influenza vaccine. Clin Infect Dis. 2012;54(12):1778-83. Epub 03/15. doi: https://doi.org/10.1093/cid/cis307. PubMed PMID: 22423139.
9. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science. 2020;369(6500):eabb9789. doi: https://doi.org/10.1126/science.abb9789.

Â© Prof. Harald Walach

10. Kuhbandner C, Homburg S, Walach H, Hockertz S. Was Germanyâ??s Lockdown in Spring 2020 Necessary? How bad data quality can turn a simulation into a dissimulation that shapes the future. Futures. 2022;135:102879. doi: https://doi.org/10.1016/j.futures.2021.102879.

Date Created
27.07.2022

Â© Prof. Harald Walach