

**Gutachten zum
„Discussion Paper - Hospital Capacity Reporting
in Germany During COVID-19“
von Dr. Simon Reif und Sabrina Schubert,
ZEW, Nr. 23-012 | 06/2023**

1. Zusammenfassung

Das Diskussionspapier enthält eine kaum mehr überschaubare Menge an schweren und kritischen Mängeln. Selbst ein einziger kritischer Mangel würde zur Ablehnung im Peer Review-Prozess führen.

Essentiell ist, dass die Schlussfolgerung, nämlich dass es angeblich keine Anhaltspunkte für „strategisches Berichten“ gab, unhaltbar ist. Außerdem widersprechen sich die Autoren hier sogar selbst, denn nach diesem Befund hätten sie die weiteren Analysen gar nicht mehr machen müssen.

Die Kernfrage, ob es „strategisches Berichten“ gab, haben die Autoren ungenügend untersucht. Tatsächlich sind aber die Hinweise auf „strategisches Berichten“ erdrückend.

Jeder Artikel mit falscher oder irreführender Schlussfolgerung muss zurückgezogen werden und darf nicht mehr zitierfähig sein, es sei denn als abschreckendes Beispiel.

Inhaltsverzeichnis

1. Zusammenfassung	1
2. Vorbemerkungen	2
2.1 Sprache.....	2
2.2 Grundsätzliches zu Intensivstationen.....	2
3. Hypothesen und Schlussfolgerungen	2
4. Darstellung der Regeln	3
4.1 Zitate und Referenzen.....	3
4.2 Systematische Darstellung	3
4.3 Unklarheiten zu Regeländerungen.....	4
4.4 „Bestimmte Krankenhäuser“	4
4.5 Erstattungsregel.....	4
4.6 Problem „unfilled“ vs. 0% Freie Betten.....	5
5. Überlegungen zur Auswertung	5
5.1 Zeitfenster	5
5.2 Überlegungen zur Auswirkung auf die Datenstruktur	6
5.3 Mathematische Überlegungen zu Datenstruktur	7
6. Was untersuchten die Autoren?	7
6.1 Zu Section 4.1: Data description.....	8
6.2 Untersuchung des „Ob?“	10
6.3 Zu Section 4.2: Event study analysis, TWFE model.....	10
6.4 Zu Section 4.5 Heterogeneity analysis	12
7. Sonstiges	12
7.1 Der „Inzidenzwert“	12
7.2 Sprachliches.....	13
7.3 Zum Schluss ein besonderes Schmankerl	13
8. Autor des Gutachtens	14

2. Vorbemerkungen

2.1 Sprache

Aus irgendeinem Grund fanden es die Autoren sinnvoll, den Artikel auf Englisch zu verfassen. Das kritisiere ich ausdrücklich nicht.

Da hier aber alle relevanten Seiten, also offensichtlich die Autoren, das ZEW, die interessierte Öffentlichkeit und ich selbst Deutsche sind, ist naheliegend, der Einfachheit halber diese Kommentare auch auf Deutsch zu verfassen. Außerdem halte ich es für undenkbar, den Artikel in dieser Form in einem internationalen wissenschaftlichen Journal zu veröffentlichen.

2.2 Grundsätzliches zu Intensivstationen

Essentiell bei der Betrachtung der Auslastung von Intensivstationen (ITS) sind die Hauptkräfte. Krankenhäuser erreichen heutzutage die meiste Wertschöpfung, vulgo Gewinne, mit mehr oder weniger geplanten Operationen und der entsprechenden Nachsorge auf ITS. Da ITS teuer im Unterhalt, aber auch gewinnträchtig sind, wird das Krankenhausmanagement i.d.R. auf adäquate Auslastung achten.

Der Einfluss von Grippe- oder Coronawellen ist meistens relativ gering. So hatte man das ja auch während der gesamten Pandemie am Verlauf der DIVI-Zahlen erkennen können. Die C19-Fälle machten anscheinend nie mehr als 30% aller belegten ITS-Betten aus. Wobei da durchaus gefragt werden kann, ob solche hohe Raten überhaupt jemals medizinisch geboten waren. Meistens dürfte der Anteil an echten C19-Patienten, also jenen mit akutem Lungenversagen (*acute respiratory distress syndrome, ARDS*) unter 10% gelegen haben.

Eine solche Einordnung wäre dringend geboten, damit niemand argwöhnt, dass die angegebenen Auslastungen im Wesentlichen auf die angebliche C19-Pandemie zurückzuführen seien. Ein allzu wohlfeiles Argument.

3. Hypothesen und Schlussfolgerungen

Die Autoren zieren sich etwas bei einer klaren Zielsetzung. Daher hier diese Zusammenstellung.

Abstract, Zeile 6:

We analyze how hospitals in Germany responded to a scheme that provided financial support if the intensive care unit (ICU) occupancy rate in a county exceeded 75%.

Abstract, Zeile 10:

To analyze hospitals' reactions to this scheme, we employ event study analyses comparing ICU capacity before and after regions became eligible.

In der Einleitung stellen die Autoren zudem klar, dass mit gewissen Manipulationen zu rechnen sei, hier beschönigend als „strategisches Berichten“ (*strategic reporting*) bezeichnet:

Seite 2, dritter Absatz:

Strategic reporting by hospitals in order to increase revenue is nothing unfamiliar to the health economics literature. There is ample evidence that ...

Kurz danach werden die Autoren dann ziemlich eindeutig:

It remains an empirical question whether – and if so – how and to what extent hospitals strategically reported their ICU capacities under the German scheme.

Und erneut kurz danach:

... we investigate whether strategic reporting took place under this scheme ...

Mit anderen Worten: **Primär sollte untersucht werden, ob strategisches Berichten vorkam oder nicht.** Dabei waren sich die Autoren vollkommen bewusst, dass es gerade im Falle der

Intensivbettenbelegung schärfste Kritik in der Öffentlichkeit gab; die Autoren zitieren einige kritische Stimmen im zweiten Absatz Seite 2, darunter auch einen [Bericht des Bunderechnungshofes](#).

Falls obige Frage, also das „Ob?“ mit ja zu beantworten war, hätte man sekundär die Fragen „Wie?“ und „In Welchem Umfang?“ angehen sollen.

Schlussfolgerung, Abstract, letzter Satz:

We find no evidence of strategic reporting at an economically meaningful and hence empirically detectable scale.

Es ist sehr bemerkenswert, dass die Autoren sich an dieser Stelle bereits fundamental widersprechen. Denn es galt ja „... **whether – and if so – how and to what extent...**“. Das bedeutet ja, dass die näheren Untersuchungen, die weit mehr als die Hälfte des Artikels ausmachen, nur dann erfolgen sollten, wenn das „Ob?“ mit Ja zu beantworten war.

Kritischer Mangel

Es ist also naheliegend, dass hier ein Gefälligkeitsgutachten vorliegt. Offenbar hatten die Autoren den Auftrag, das Vorliegen von „strategischem Berichten“ zu verneinen, andererseits wollten sie mit weiteren komplexen Untersuchungen glänzen. Zahlreiche weitere Argumente bestätigen diesen Verdacht wie auch die Unhaltbarkeit oben zitierter Schlussfolgerung.

Unhaltbarkeit einer Schlussfolgerung erzwingt notwendigerweise die komplette Rücknahme.

4. Darstellung der Regeln

Regeln sind bei allen Studien essentiell, hier aber doppelt, weil ja gerade der Einfluss bestimmter Regeln der Verordnungsgeber auf das Verhalten, hier der Krankenhäuser, der Untersuchungsgegenstand war.

Die Darstellung der Regeln ist in jeder Hinsicht ungenügend und in der Summe als kritisch zu bewerten.

Kritischer Mangel

4.1 Zitate und Referenzen

Es gibt nicht einmal korrekte Referenzen für die Regeln, die i.W. auf Seite 4 dargestellt wurden. Auf jeden Fall müssten die Gesetze, und soweit zutreffend Verordnungen, juristisch korrekt referenziert werden. Außerdem wäre es äußerst wünschenswert, die Originaltexte der Verordnung, wenigstens auszugsweise, im Anhang zu präsentieren; z.B. als PDF oder als „Fotos“. Eine Übersetzung ins Englische halte ich nicht für geboten.

Schwerer Mangel

Die Regeln waren im Wesentlichen im §21 des Krankenhausfinanzierungsgesetz – KHG, darin insbesondere im Absatz (1a).

<https://www.gesetze-im-internet.de/khg/BJNR010090972.html>

4.2 Systematische Darstellung

Für einen solchen Artikel wäre essentiell, diese Regel tabellarisch so darzustellen, dass der Leser ein Gefühl für den Zeitstrahl bekommt. Inakzeptabel ist, dass so essentielle Regeln wie jene in der Fußnote 3, Seite 4, eben nur in einer Fußnote zu finden sind.

Schwerer Mangel

4.3 Unklarheiten zu Regeländerungen

Die Autoren geben in jener Fußnote an, dass die „Condition 2“ (zwischen 17. Dezember 2020 und 15. Januar 2021) völlig entfiel, wenn die „Inzidenz“ in einem Landkreis den Wert von 150 überstieg. Gleiches gilt für die Änderung, die laut Autoren ab dem 15. Januar 2021 gelten sollten.

Ich konnte weder diese noch eine ähnliche ab 15. Januar verifizieren. Diese Unklarheit hängt direkt mit oben erwähntem Mängeln (4.1 und 4.2) zusammen und wird dort bereits als schwerer Mängel klassifiziert.

Da die Autoren vermutlich von der Existenz dieser Regeln ausgingen, basieren die folgenden Überlegungen eben darauf.

4.4 „Bestimmte Krankenhäuser“

§21 (1a) stellt außerdem klar, dass es eine weitere, von den Autoren nicht diskutierte Bedingung gab, nämlich:

... kann die für die Krankenhausplanung zuständige Landesbehörde Krankenhäuser in dem Landkreis oder in der kreisfreien Stadt bestimmen, die Ausgleichszahlungen nach Satz 1 erhalten, wenn diese

- a) einen Zuschlag für die Teilnahme an der umfassenden oder erweiterten Notfallversorgung gemäß § 9 Absatz 1 a Nummer 5 des Krankenhausentgeltgesetzes für das Jahr 2019 oder für das Jahr 2020 vereinbart haben oder*
- b)...*

Diese Regel dürfte durchaus Bedeutung für den vorliegenden Datensatz hatte. Es gab offenbar eine gewissen Menge von Krankhäusern, die durch die genannten Regeln von 7-Tage-Inzidenz sowie einer Belegungsrate von 75% (25% freie ITS Betten) gar nicht betroffen waren. Anderenfalls wären die vielen Werte weit unterhalb der 75% Belegungsrate nämlich nicht erklärbar.

Berücksichtigt man diese Teilmenge nicht, erhöht sich die Variabilität bzw. senkt sich die Trennschärfe. Letztere erscheint aber bei sinnvoller Methodik völlig hinreichend. Die Berücksichtigung dieser Krankenhäuser macht Ergebnis und Schlussfolgerungen dagegen robuster. Ich plädiere für eine robuste Analyse, also den Einschluss der nicht-bestimmten Krankenhäuser, und eine entsprechende Diskussion. Dies fehlte.

Schwerer Mangel

4.5 Erstattungsregel

Auf Seite 3 / 4 schrieben die Autoren:

This reformed scheme still contained the flat payments from the earlier scheme (between €360 and €760 per unfilled bed compared to 2019) ...

Es wurde nichts zu der Spanne von 360 bis 760 € geschrieben wurde. Tatsächlich war die Berechnung noch etwas komplizierter.

Kleiner Mangel.

Das unscheinbare „per unfilled bed compared to 2019“ implizierte aber zwei Dinge:

1. Eine Auslastung knapp oberhalb der 75%-Marke bewirkte die höchsten Ausgleichszahlungen wegen der dann größtmögliche Rate freier Betten.
2. Im Vorfeld galt es, dem Vergleichswert 2019 möglichst nahe zu kommen.

Ad 1: Diese zentrale Überlegung fehlt im Artikel völlig. Die ist aber essentiell, um den Einfluss der Regelsetzung auf die Datenstruktur zu erklären.

Kritischer Mangel

Zu 2. Der Vergleichswert 2019 dürfte unbeeinflussbar sein. Allerdings änderten die Krankenhäuser gerade im Laufe des Jahres 2020 vielfach ihre ITS-Kapazität. Eine Adjustierung der Kapazität im Herbst 2020 an diese Referenzgröße, egal ob nach oben oder nach unten, war natürlich möglich und dürfte erfolgt sein. Auch dies hätte erwogen werden müssen. Siehe dazu auch die Diskussion im Abschnitt zur **Figure 3a** auf Seite 11. Dieser Aspekt wird dort als schwerer Mangel bewertet.

Zumindest hätten die Autoren angesichts der Bedeutung der Bettenkapazität 2019 als Referenzgröße auf den Gesetzestext verweisen müssen.

Mäßiger Mangel

Dazu machte das KHG Vorgaben, die allerdings m.E. nicht restlos klar waren. Interessant ist ferner, dass das DIVI-Intensivregister erst Anfang April 2020 aufgebaut wurde. Man konnte daher zur Kontrolle nicht auf DIVI-Daten zurückgreifen.

Leichter Mangel

4.6 Problem „unfilled“ vs. 0% Freie Betten

Essentiell ist das „unfilled“ aber im Zusammenhang mit Figure 2a und 2b. Die mit Abstand höchsten und zudem völlig isolierten Säulen zeigen nämlich 0% freie Betten (=100% Auslastung). Dies ist nun aber sehr überraschend, denn 0% freie Betten würde ja bedeuten, dass es keinerlei Sondererstattung gegeben hätte. Es könnte zudem schlicht so sein, dass die größere Wertschöpfung bei 100% Auslastung lag und dass gerade kleinere Häuser bzw. Kreise mit, z.B. 3 ITS-Betten sowieso nur die Wahl zwischen $2/3 = 67%$ Auslastung und $3/3 = 100%$ Auslastung hatten. Ich konnte nirgends eine Erklärung oder Diskussion dieser Beobachtung finden. Eine Erklärung dafür könnte aber von großer Bedeutung sein.

Schwerer Mangel

Außerdem hätte man die zahlreichen Fälle von 100% Auslastung genauer analysieren sollen. Woher kamen die Meldungen? Wie war die ITS-Bettenanzahl, zu welchen Zeiten, wie war die „Inzidenz“ zu diesen Zeitpunkten. Die absolute Anzahl der verfügbaren Betten beeinflusst mathematisch die auf klassifizierten Prozentzahlen basierende Datenstruktur.

Schwerer Mangel.

Ebenso hätte man das andere Ende der Skala, also sehr niedriger bzw. Null-Auslastung diskutieren sollen.

Mäßiger Mangel.

5. Überlegungen zur Auswertung

Angesichts des Zieles, **ob strategisches Berichten vorkam oder nicht**, hätten die Autoren eine Reihe von Überlegungen anstellen müssen. Dies unterblieb aber.

Kritischer Mangel

5.1 Zeitfenster

Siehe Abschnitt 4.3. Derzeit gehe ich von einer Regeländerung zum 17. Dezember 2020 aus.

Sollte es tatsächlich die erwähnte Regeländerung zum 17. Dezember 2020 gegeben haben, sind alle simplen Auswertungen über dieses Datum hinaus **obsolet**. Denn die erwähnte Regeländerung hätte stark und vor allem anders als zuvor die Datenstruktur beeinflussen müssen. Dies gilt für die Table 1 (siehe dazu auch Kapitel 6.1, Seite 8) sowie Figure 1 und Figure 2 des Artikels.

Kritischer Mangel (bedingt, oder aber kritischer Mangel bei Regelbeschreibung)

Zum besseren Verständnis bezeichne ich im Folgenden die Zeit vom 17. November bis 17. Dezember 2020 als **Regelphase A**.

Neben der Regelphase A gab es natürlich auch die **Vor-Regel-Phase**. Dazu ließen sich die Autoren in ihrem Kapitel 4.4 aus. Demnach erschien ein Entwurf der Regeln erstmals am 3. November 2020. Die Autoren schreiben an anderer Stelle, S. 10:

“Counties which already had an incidence above 70 before the scheme was introduced are first treated on November 8th, 2020 ...”

Offensichtlich gab es also eine **Übergangsphase** im November 2020.

Mir vorliegende Daten deuten aber an, dass es, wenn auch sehr wenige, Krankenhäuser gab, die bereits im Oktober 2020 „ahnten“, wie die kommenden Regelungen ungefähr aussehen würden. Zumal die verschiedenen politischen Veranstaltungen in jener Zeit klar machten, wie eng die „Zusammenarbeit“ der federführenden Politik und der Führung der DIVI war.

Man sollte die Monate September und Oktober als **Vor-Regel-Phase** und den 1. bis 17. November als Übergangsphase werten.

Die Autoren hätten solche oder ähnliche Festlegungen treffen müssen.

Kritischer Mangel

Es ist zudem keineswegs so, dass die Autoren sich nicht auch die Vor-Regel-Phase angeschaut hätten, Beweis Abbildung Figure A.1, Seite 26. Dort zeigten sie nämlich die Verteilungen der ITS-Kapazitäten zwischen dem 1. Oktober und 15. November 2020. Warum eigentlich bis zum 15.?

Leichter Mangel

Unter der Voraussetzung, dass die Regeländerungen vom 17. Dezember und 15. Januar so zuträfen, hätten alle Phasen ab dem 17. Dezember 2020 gesondert analysiert werden sollen und keinesfalls mit der Regelphase A vermischt werden dürfen.

Kritischer Mangel

5.2 Überlegungen zur Auswirkung auf die Datenstruktur

Die Regeln galten überhaupt nur für die vom Kreis „bestimmten Krankenhäuser“ (siehe Kapitel 4.4). Die nicht bestimmten waren zunächst außen vor. Die Autoren hätten ihre Anzahl bestimmen sollen, ferner, ob es später Übergänge gab.

Schwerer Mangel

Die „bestimmten Krankenhäuser“ hatten ein Interesse, den Belegungsgrad zwar $\geq 75\%$ zu halten, aber wegen der Vergütung pro freiem Bett auch nicht weit darüber. Dies hätte explizit zum Ausdruck gebracht werden müssen.

Schwerer Mangel

Die Autoren betonen sehr die Inzidenz-Regel (mindestens 70), die sie offenbar für nicht manipulierbar halten; kann sein, muss nicht sein (siehe Kapitel 7.1).

Leichter Mangel

Wichtig ist einerseits, dass im November 2020 bereits viele Kreise über der Marke von 70 lagen. Die anderen Kreise bzw. Krankenhäuser konnten ja absehen, dass diese Marke bald überschritten werden würde und hatten daher ein Interesse daran, sich frühzeitig darauf einzurichten. De facto spielte also die „Inzidenz“ keine oder zumindest keine wesentliche Rolle. Dies hätte diskutiert werden müssen. In diesem Sinne hätten die Autoren gute Gründe gehabt,

diese Nuance gar nicht mehr zu betrachten und statt dessen eine robuste Analyse (also über alle Kreise) wählen sollen; explorativ hätte man dann vielleicht noch die Inzidenz-Frage angehen können. Demnach wäre eine primäre Analyse nach Zeit, also der Vor-Regel-Phase gegen die Regelphase A die vernünftigste gewesen. Eine solche Analyse unterblieb aber.

(Wie schon erwähnt: Kritischer Mangel)

Unter der bereits am 11. Januar 2021 vom RKI geäußerten Annahme, dass es auch „strategisches Berichten“ gab, hätten sich diese Regeln wie folgt auf die Datenstruktur der Regelphase A auswirken sollen:

- Wesentlich weniger Werte für eine Auslastung $<75\%$ als in der Vor-Regel-Phase (Chi-Quadrat-Test, Schätzung mit Disproportionalitätsanalyse).
- Wesentlich andere Verteilung als in der Vor-Regel-Phase (F-Test).
- Eine Ballung bei Werten knapp ≥ 75 geben (Verminderungen des Interquartilebereichs).
- Verschiebung der gesamten Punktwolke nach oben (Erhöhung von Median und Quartilen, Wilcoxon Vorzeichen-Rang-Test).

Man könnte ferner eine Testhierarchie erwägen. Dann wäre der Test auf den kategorischen Daten (Chi-Quadrat) sicher an erster Stelle zu nennen. An zweiter Stelle wohl der F-Test.

Bitte beachten:

Aufgrund mir vorliegender Daten zu den Kapazitätsmeldungen (Anteil $<75\%$ in Klammern) ergeben sich aus Chi-Quadrat-Test bei Yates Korrektur:

September 2020 (66,6%) vs. November 2020 (49,5%), $p < 0,00001$ ($X^2 = 715$)

September 2020 (66,6%) vs. Dezember 2020 (31,7%), $p < 0,00001$ ($X^2 = 2941$)

Oktober 2020 (63,8%) vs. November 2020 (49,5%), $p < 0,00001$ ($X^2 = 503$)

Oktober 2020 (63,8%) vs. Dezember 2020 (31,7%), $p < 0,00001$ ($X^2 = 2530$)

NB: Bei allen X^2 Werten $> 18,44$ gilt $p < 0,00001$

Mithin sind alle Vergleiche der Regelphase(n) mit den Vor-Regel-Phasen höchst signifikant.

Schon damit kann als erwiesen angesehen werden, dass ab November 2020 „strategisches Berichten“ vorlag. Dies widerspricht fundamental den Schlussfolgerungen der Autoren.

5.3 Mathematische Überlegungen zu Datenstruktur

Die Zielvariable war ja ein Prozentwert (ITS-Kapazität), es konnten und kamen tatsächlich auch Werte von 0% bis 100% vor. Prozentzahlen sind nur in den mittleren Bereichen, also etwa zwischen 15% und 85% normalverteilt, in den Schwänzen aber nicht. Die tatsächlichen Daten zeigen zudem eine sehr starke Häufung bei 100%, bisweilen auch bei 0% Auslastung.

Von daher scheiden alle parametrischen Auswertungen von vorneherein aus. Es ist zudem zu vermuten, dass von daher auch die von den Autoren bevorzugte Auswertung TWFE unangemessen gewesen sein dürfte. Explizit schreiben alle Autoren zum TWFE, dass homogene und keine heterogenen Daten vorliegen müssten. Tatsächlich waren die Daten aber sehr heterogen, also die TWFE höchst fragwürdig.

6. Was untersuchten die Autoren?

Laut Aussagen der Autoren (Seite 6) wollten sie zuerst untersuchen, **ob** die Verteilung der ITS-Betten sich um die „Inzidenz“ von 70 änderte und **ob** sich eine Verschiebung in der Verteilung der freien Betten um die entscheidende Schwelle der 75% Belegung zeigte:

First, we use a set of bivariate analyses to show whether the distribution of ICU beds changed around the incidence threshold of 70 and check for heaping in the distribution of free ICU beds at the eligibility threshold of 25%.

Erst danach wollten Sie den TWFE Schätzer nutzen, um die Frage „wie“ zu erklären.

Second, we use two-way-fixed-effects (TWFE) estimations to analyze how the number of available ICU beds changed after the incidence threshold of 70 was crossed in each county.

Demnach hätte eine Testhierarchie vorgelegen, also zuerst das “Ob?”, dies sogar zweifach, dann das “Wie?”.

6.1 Zu Section 4.1: Data description

Kern dieses Kapitels ist die „Table 1: Descriptive statistics“.

Wie oben diskutiert, handelt es sich bei dieser Untersuchung um eine Art Längsschnittstudie mit 2 oder besser 3 Phasen: Der **Vor-Regel-Phase** (Monate September und Oktober 2020), einer **Übergangsphase** (1. bis 16. November) und der **Regelphase A** (17. November bis 17. Dezember 2020). Eventuell schlossen sich dann noch weitere Regelphasen an.

Die Fragestellung „Ob?“ kann aber ohne Betrachtung der **Vor-Regel-Phase** gar nicht beantwortet werden. Die Deskription hätte also mindestens 2 oder besser 3 oder sogar mehr Phasen zeigen müssen.

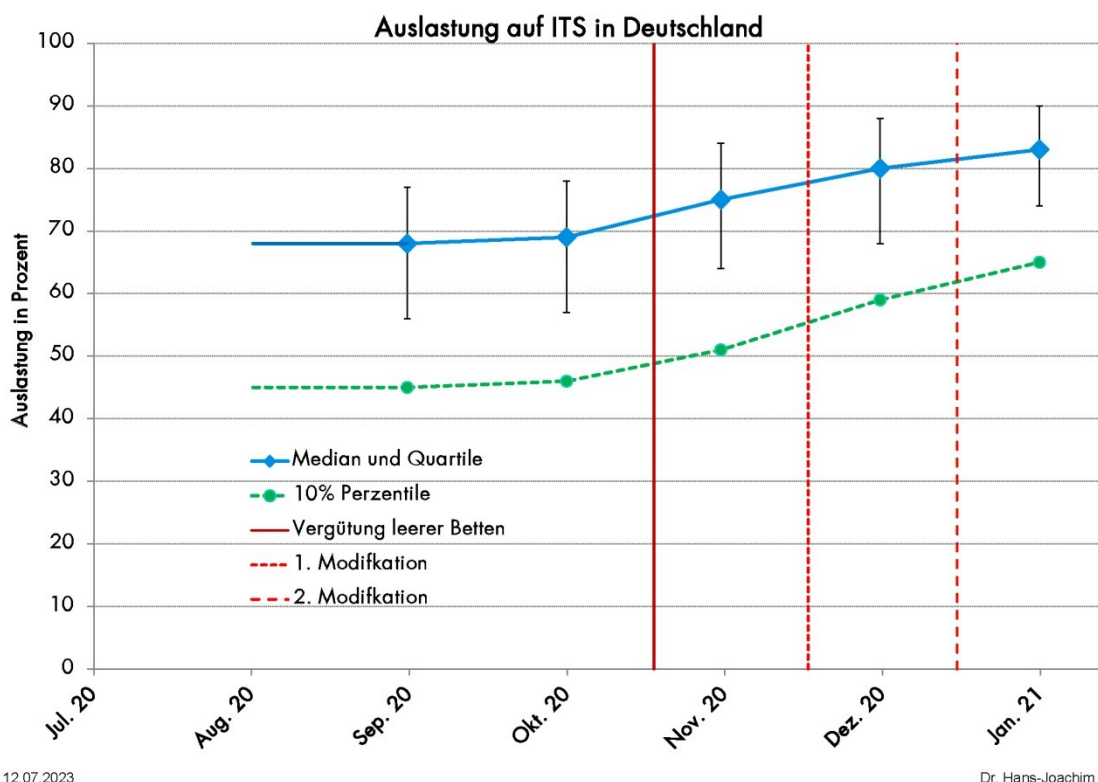
Kritischer Mangel

Dieser Mangel wird sogar noch bedeutsamer, da die Autoren in Figure A1 Verteilungsdaten zu einer Vor-Regel-Phase zeigten. Sie hatten also die Daten, sie wussten um die Bedeutung, aber bereiteten sie nicht numerisch auf.

Des Weiteren sind die gezeigten Parameter völlig unzureichend. Die Autoren hätten mindestens noch Median und Quartile zeigen müssen; da aber gerade die Datenstruktur in Rede stand, besser auch noch die 5er, 10er, 90er und 95er Perzentilen.

Kritischer Mangel

Aus mir zugänglichen Daten habe ich eine Abbildung angefertigt, die eine einfache, aber sinnvolle Datenbeschreibung pro Monat zeigt. Es ist unschwer zu erkennen, dass mit Einführung der Zuzahlungsregeln ab dem 17. November es zu einer erheblichen Verzerrung der Datenstruktur kam.



12.07.2023

Dr. Hans-Joachim Kremer

Abb. 1 **Auslastung auf ITS in Deutschland**

Datenquelle: Raimund Hagemann

Der in „Table 1“ angegebene Wert für die Standardabweichung von „Share free ICU beds“ von 0,13% ist offensichtlich falsch. Vermutlich sollte da 13% stehen.

Mäßiger Mangel

Die „Figure 1“ ist nett zu haben (*nice to have*). Allerdings lässt die Beschreibung offen, ob mit „November“ die Zeit vom 17. Oktober bis zum 17. November (das wäre immerhin eine, wenn auch suboptimale Vor-Regel-Phase) oder die Zeit vom 17. November bis zum 17. Dezember gemeint ist.

Mäßiger Mangel

Die „Figure 2“ ist ebenfalls nett zu haben. Allerdings hätte es eine solche Abbildung auch für die Vor-Regel-Phase sowie die Übergangsphase geben sollen. Außerdem, falls zutreffend, auch für die Regelphasen ab dem 17. Dezember bzw. ab dem 15. Januar 2021.

Kritischer Mangel

In beiden Teilen, also Figure 2a und b, ist die hervorstechendste Säule jene mit 0% freier ITS-Betten bzw. 100% Auslastung. Diese Tatsache hätten die Autoren, wie schon oben ausgeführt, eingehend untersuchen und diskutieren müssen. Dies fehlt komplett.

Kritischer Mangel

Die in Figure 2 sichtbaren Verteilungen zeigen eindeutig, dass keine homogene (oder normale) Verteilung vorliegt. Dies hätten explizit angesprochen werden müssen, da die Voraussetzung für die später vorgenommene TWFE-Analyse demnach vermutlich gar nicht vorlag.

Kritischer Mangel

6.2 Untersuchung des „Ob?“

Die Autoren widmen dieser zentralen Frage kein eigenes Kapitel. Folgende Kernaussagen sind in Section 4.1 versteckt:

If hospitals strategically report their ICU capacity to be eligible for the emergency financing scheme, there should be heaping in the distribution of free ICU beds below the threshold of 25%. To get a first impression whether such a pattern occurs, we look at heaping in the univariate distribution of free ICU beds.

Es folgt dann die bereits besprochene Figure 2. Die Autoren schreiben lediglich:

“There is, however, no clear heaping below the 25% eligibility threshold.”

NB: Unterhalb 25% freie ITS-Betten entspricht oberhalb 75% belegte Betten.

Eine Aussage, ob eine Häufung („heaping“) unterhalb 25% seit der Einführung der Regel vorliegt, ist ohne Betrachtung der Vor-Regel-Phase logisch unmöglich. Dies unterließen die Autoren. So gesehen ist obiger Kernsatz wie auch die eingangs zitierte Schlussfolgerung unhaltbar.

Kritischer Mangel

6.3 Zu Section 4.2: Event study analysis, TWFE model

Ohne nochmals auf die Kernfrage des „Ob?“ einzugehen, widmen sich die Autoren dieser Auswertung. Dabei hätten sie nach den eigenen Aussagen diese Fragestellung nur dann angehen sollen, wenn das „Ob?“ zuvor bejaht wurde. Dies stellt einen logischen Bruch dar.

(Wie schon erwähnt: Kritischer Mangel)

Die Autoren untersuchen dann die Frage, wie stark sich die Kapazität der freien ITS-Betten nach der Erreichung der „Inzidenz“ von 70 änderte. Das mag eine interessante Untersuchung sein, allerdings ist tatsächlich schon die Frage, ob das Datum 17. November 2020 nicht viel entscheidender als die „Inzidenz“-Schwelle war. Die ergibt sich aus den Überlegungen: A) Viele Kreise erreichten die Schwelle schon vor diesem Stichtag, als die Zuzahlungsregel noch gar nicht griff. B) Krankenhäuser unterhalb jener Schwelle, zumal solche kurz vor Erreichen dieser Schwelle, dürften gut daran getan haben, sich in Sachen ITS-Kapazität darauf vorzubereiten. Mindestens hätte man also diesen Stichtag ebenfalls untersuchen müssen.

(Wie schon erwähnt: Kritischer Mangel)

In diesem Zusammenhang sollte die Aussage in **Section 4.5.5 „Timing of fulfilling incidence condition“**, Seite 19 beachtet werden:

For our main specification, we assume that as soon as the incidence threshold is reached, hospitals might start to report strategically, but it is also possible that they only adjust reporting when the seven days in a row of high incidences are reached.

Mindestens muss dieser Satz sprachlich bemängelt werden, denn eigentlich beschreibt der erste Teil („as soon as“) das Gleiche wie der zweite Teil („when“). Unter der Überschrift „Timing“ hätte hier also der Stichtag 17. November 2020 gehört. So aber wurde dieser Stichtag gar nicht weiter untersucht.

(Wie schon erwähnt: Kritischer Mangel)

Meiner Erfahrung nach sind Sätze wie der vorhin zitierte ein starker Hinweis darauf, dass den Autoren bewusst war, dass hier eine Unlogik vorkam und sie daher diesen Satz verschwebelten. Warum wohl ging man bei dieser Überschrift nicht wirklich auf das „Timing“ ein?

Die Autoren nutzen für diese Analyse ab Section 4.2 das „**two-way fixed-effects**“ (TWFE) Modell. Derartige Modelle wurden offenbar von Ökonomen vielfach benutzt, um den Effekt

von Maßnahmen (z.B. Maskenpflicht, Schulschließungen usw.) während der „Pandemie“ zu bewerten. Auch von theoretischen Überlegungen her ist das TWFE-Modell sehr kritisch zu betrachten [1].

1. Gauthier G. On the use of two-way fixed effects models for policy evaluation during pandemics arXiv preprint arXiv:2106.10949, 2021 - arxiv.org

Die Autoren verweisen nicht auf diesen Artikel, obwohl er Ihnen hätte bekannt sein können.

Schwerer Mangel

Allerdings fordert auch alle einschlägige theoretische Literatur zu TWFE [2] grundsätzlich:

- Die Effekte müssen homogen über Einheiten und Zeitphasen sein.

2. A Guide on Data Analysis. Section 25. Two-way Fixed effects.

https://bookdown.org/mike/data_analysis/two-way-fixed-effects.html

Demnach ist Heterogenität ein Kernproblem der TWFE, das eingehend untersucht und diskutiert werden muss. Gerade die Häufungen von 100% Belegung (bzw. 0% freie ITS-Kapazität, siehe oben in Abschnitt 6.1) zeigen, dass die Datenstruktur sehr heterogen war, und dies in möglicherweise relevantem Ausmaß. Außerdem wird aus Abb. 1 ersichtlich, dass die Heterogenität nach Einführung der Regel deutlich zunimmt. Dies hätte eingehender diskutiert werden müssen.

Schwerer Mangel

Die Beschreibung der **TWFE**-Variablen erscheint unzureichend. Es gibt da ein „beds or share“, was Beliebigkeit suggeriert, in Wahrheit aber korrekt ist, denn man betrachtete beide Variablen; vielleicht sprachlich klarer trennen. Welches genau sind die „two ways“, welches sind die „fixed effects“?

Mäßiger Mangel

Die **Figure 3** zeigt sowohl in (a) als auch in (b) den Datenpunkt -1 ohne Fehlerbalken, der Wert ist exakt 0. Ich habe dazu keine Erklärungen gefunden. Offensichtlich nahm man diesen Tag als Bezugspunkt. Von da aus konnten die, vermutlichen Differenzen* nach oben oder unten schwanken. Wieso definierte man als Anker nicht den Tag 0? Zumal dieser Tag sogar mit einer Senkrechten hervorgehoben wurde. Gerade in (b) überrascht, dass der höchste Punkt der Kurve an Tag 0 liegt. Dies hätte ebenfalls diskutiert werden sollen.

Schwerer Mangel.

* Falls es da nicht um Differenzen ging, wäre wohl die Y-Achsbeschriftung falsch. Allerdings könnte man bei den „shares“ auch argwöhnen, dass Quotienten sowie eine logarithmische Y-Achse sinnvoller wären.

Die **Figure 3a** zeigt, dass sich die absolute Anzahl der ITS-Betten über den analysierten Zeitraum insgesamt nur marginal änderte, im Mittel nämlich nur von etwa 0,7 bis -0,5. Hätte es damals wirklich eine schlimme C-19 oder Grippe-Welle gegeben, hätte man doch im Trend eine Zunahme der Bettenanzahl erwarten sollen. Dies war aber nicht der Fall, sondern im Gegenteil. Dies führt zum Schluss, dass entweder die C-19 oder Grippe-Welle keinen Einfluss auf die Belegung hatte, was den medialen und politischen Interpretationen diametral widerspricht, oder dass eben doch „strategisches Berichten“ vorlag. Trotz möglicherweise fehlender Signifikanz hätten die Autoren dies diskutieren müssen, und zwar wegen der Abweichung vom offiziell zu erwartenden Ergebnis.

Schwerer Mangel

Die Annahme des strategischen Berichtens wird durch **Figure 3b** bestärkt, denn 1. ist der Verlauf völlig anders als der von Figure 3a, 2. ab Tag 4 sinken die Anteile der freien Kapazitäten signifikant unter den Wert von Tag -1. Da die Kapazität den Quotienten aus „total reported ICU beds“ und der aktuell berichteten Gesamtkapazität darstellt, können die

unterschiedlichen Kurven nur durch Änderungen in Zähler **und** Nenner erklärt werden, mithin erneut ein starker Hinweis auf „strategisches Berichten“. Dieser Effekt könnte sogar noch stärker werden, wenn man tatsächlich den Tag 0 als Bezugspunkt genommen hätte.

Diese Beobachtungen hätten die Autoren selbst anstellen und deuten müssen.

Kritischer Mangel

Wenn sich aber strategisches Berichten sowohl im Nenner als auch im Zähler manifestierte, und eben diese zeigen die Figure 3a und 3b zusammengenommen, hätte man noch andere Variablen und Modelle zum Beschreiben des Effektes verwenden sollen.

Schwerer Mangel

Schließlich noch ein Detail: Es gab einen beinahe signifikanten Abfall bei der Anzahl der Betten (Figure 3a) zwischen Tag -6 und -5 im Vergleich zu Tag -1. Auch hier deutet sich erneut strategisches Berichten an, denn bei zunehmender Inzidenz hätte man ja eher eine Zunahme von belegten Betten erwarten sollen. Dies wurde nicht diskutiert.

Mäßiger Mangel

In **Figure 4** fällt auf, dass nun kein offensichtlicher Bezugspunkt für die Schätzungen genommen wurde, sondern lediglich die X-Achse um den Tag 0 gelegt wurde. Dafür werden die Fehlerbalken deutlich größer als in Figure 3a oder 3b, vor allem nach dem Tag 0. Diese Umstände hätten erwähnt werden sollen.

Mäßiger Mangel

6.4 Zu Section 4.5 Heterogeneity analysis

Auf Seite 18 schreiben die Autoren:

The group of counties which are treated in our event study analyses is very heterogeneous.

Demnach stellen die Autoren ihr eigenes Analysekonzept, also die TWFE, massiv in Frage.

In **Section 4.5.2** untersuchen die Autoren „Pandemic waves“. Dazu gibt es im Anhang die Figure A.5 und A. 6. Erneut wird nach Tag des Erreichens der Inzidenzschwelle von 70 analysiert. Es verwundert, dass es in der „dritten Welle“, definiert ab 1. März 2021, überhaupt noch viele Kreise unterhalb dieser Schwelle gab. Diese Befürchtung wird durch die wesentlich größeren Konfidenzintervalle in Figure A.6 bestätigt. Mindestens hätten die Autoren die Anzahl der untersuchten Kreise angeben müssen.

Schwerer Mangel

Hier steht überhaupt das Konzept, das Überschreiten der Inzidenzschwelle primär zu analysieren, in Frage. War die Beobachtung der Inzidenzschwelle überhaupt noch relevant für die Krankenhäuser, wenn doch schon weit höhere Inzidenzen vorgekommen waren und ein nachhaltiges Unterschreiten dieser Schwelle unwahrscheinlich erschien? Dies wurde, so weit erkennbar, nicht diskutiert.

Schwerer Mangel

7. Sonstiges

7.1 Der „Inzidenzwert“

Der Artikel geht davon aus, dass Krankenhäuser keinen Einfluss auf den Inzidenzwert gehabt hätten. Das muss so aber gar nicht stimmen. Denn der „Inzidenzwert“ hing ja von der Zahl der Testungen sowie, jedenfalls beim damals ausschließlich genutzten rt-PCR-Test, vom CT-Wert ab (höhere CT-Werte erhöhen die Rate der falsch-positiven Befunde). Da aber Krankenhäuser auf jeden Fall alle Patienten und zusätzlich zum Teil auch noch Personal testeten und zudem Einfluss auf ihr eigenes Labor (→ CT-Wert) ausüben konnten, lagen durchaus Möglichkeiten zur

Beeinflussung vor. Zudem gab der Gesetzgeber mit der [Anpassung des Bundesgesetzblattes](#) 2020 Teil I Nr. 33, ausgegeben am 8. Juli 2020 in Ziffer 2 den für Kliniken umsetzbaren Anreiz, „abweichend von § 21 Absatz 6 Satz 1 des Krankenhausfinanzierungsgesetzes [den] Zuschlag [von] 100 Euro [zu gewähren], wenn voll- oder teilstationär behandelte Patientinnen oder Patienten mit dem Coronavirus SARS-CoV-2 infiziert sind.“

Dies sollte hier bzw. im Artikel nur erwähnt werden, da eine ernsthafte Auseinandersetzung mit dieser Problematik den Rahmen dieses Artikels sprengen würde. Ich gestehe zu, dass der Umfang derartiger Manipulation möglicherweise gar nicht solide abzuschätzen sein dürfte.

Immerhin stellt dieser Umstand ebenfalls das Analysekonzept mit Fokus auf Überschreiten der „Inzidenzschwelle“ in Frage.

Mäßiger Mangel

7.2 Sprachliches

Im Deutschen gibt es keine so strengen Regeln zu Zeiten wie im Englischen. Gerne nutzen Deutsche selbst bei Ergebnissen, also eindeutig abgeschlossenen Handlungen oder Beobachtungen den Präsens, vermutlich um etwas Dramatik in den Text zu bringen.

Das geht im Englischen aber gar nicht. In Einleitung (weitestgehend „simple past“, vielleicht auch mal „past perfect“), Methoden (100%) und Ergebnisse (100%) nutzt man das „simple past“, also z.B. „we performed“ oder „were performed“. In der Diskussion wird man i.d.R. ebenfalls hohe Prozentsätze an „simple past“ bekommen, bisweilen mal einen Ausblick in Präsens oder Futur.

Die Autoren sollten sich an diese Regeln halten, die zumindest für bio-medizinische Dokumente gelten. Der andauernde Gebrauch von „simple present“ bei Ergebnissen, die vollständig abgeschlossen sind, verwirrt und behindert den Lesefluss.

7.3 Zum Schluss ein besonderes Schmankerl

Hier möchte ich nochmals auf die Schlussfolgerung eingehen:

We find no evidence of strategic reporting at an economically meaningful and hence empirically detectable scale.

Und zwar zunächst auf das

economically meaningful

Gewiss, des ZEW führt Wirtschaftsforschung im Namen. Allerdings betrachten die Autoren in den Auswertungen nirgends auch nur annähernd ökonomische Fragestellungen. Lediglich in der Einleitung wird allgemein der [Bericht des Bunderechnungshofes](#) zitiert, der aber gerade diesen Regeln anhand der Zahlungen hohe wirtschaftliche Bedeutung beimaß; also auch hier das Gegenteil dessen, was die Schlussfolgerung behauptet. Tatsächlich schlussfolgerte d der [Bericht des Bunderechnungshofes](#):

Das derzeitige System der Ausgleichszahlungen hat unerwünschte Mitnahmeeffekte eröffnet.

...

Die Ausgleichszahlungen nach § 21 KHG ermöglichten vielen Krankenhäusern im vergangenen Jahr eine massive Überkompensation aus Steuermitteln: Bei sinkender Bettenauslastung um knapp 8 Prozentpunkte wuchsen die Zahlungen der Krankenkassen für Krankenhausbehandlungen im Jahr 2020 gegenüber 2019 um 1,7 %. Hinzu traten die Ausgleichszahlungen des Bundes, die allein im Jahr 2020 10,2 Mrd. Euro betragen.

Der Bundesrechnungshof hat beanstandet, dass das BMG bis heute nicht in der Lage ist, die Zahl der tatsächlich aufgestellten sowie die der zusätzlich angeschafften Intensivbetten verlässlich zu ermitteln.

...

Ein solcher Kapazitätszuwachs ist aus den vorliegenden Statistiken und Datensammlungen indes nicht abzulesen. Dies gilt auch für die im DIVI-Intensivregister abrufbaren Zahlen.

Vereinfacht gesagt: Es wurden 10 Milliarden € aufgewendet, um Krankenhäuser zu unterstützen und Intensivbetten auszubauen, aber ein Kapazitätszuwachs war nicht gegeben. Von daher ist weder die Aussage „no evidence“ noch „on a economically meaningful ... scale“ sachlich haltbar. Zumal von den Autoren letzteres nicht untersucht wurde.

Schlussfolgerungen von Texten mit wissenschaftlichem Anspruch dürfen sich nur auf die eigenen Untersuchungen stützen. Diese Attribute sind sachlich völlig unberechtigt. Die stehen zudem nicht irgendwo im Prosatext, sondern im wichtigsten Satz des Artikels.

Kritischer Mangel

Aber auch das

and hence

ist nicht weniger fragwürdig. Denn es war ja anders herum. Die Autoren bemühten sich, das „Wie“ des befürchteten strategischen Berichtens anhand der TWFE-Analysen zu untersuchen. Grundsätzlich und abgesehen von obiger Kritik könnte man dies als Untersuchung auf einer „*empirically detectable scale*“ ansehen, hier also freie Intensivbetten und ITS-Kapazität. Erst in zweiter Linie hätte man den wirtschaftlichen Einfluss betrachten können, was wie gesagt unterblieb. Auch hier gilt, in Prosa einigermaßen verzeihlich, aber nicht in der Schlussfolgerung.

Schwerer Mangel

Diese seltsamen Formulierungen in den Schlussfolgerungen geben Anlass zur Überlegungen, wie diese überhaupt zu Stande kommen konnten. Ich kann mir es nur so vorstellen: Eine besonders mächtige Person sollte das finale Plazet geben und hat bei dieser Gelegenheit diese Ergänzungen eingefordert. Offensichtlich hatte die Person den Artikel nicht gelesen oder verstanden, sich aber gedacht, ZEW hat etwas mit Ökonomie zu tun und daher diese Attribute eingesetzt. Ebenso wie die Logikverdrehung mit „*and hence*“.

Offenbar ist diese Person derart mächtig, dass die Autoren sich nicht trauten, diese Korrektur nochmals zu redigieren.

8. Autor des Gutachtens

Dr. Hans-Joachim Kremer
Alemannenstraße 101
79117 Freiburg
Festnetz: 0761 6966 480
www.medical-writing-service.com

Mein Dank geht an Raimund Hagemann für die Bereitstellung von ITS-Daten sowie für redaktionelle Arbeiten.