



Identifizierung von potentielltem Abrechnungsbetrug bei COVID-19-Schnelltests: Ergebnisse einer Pilotstudie zum Vergleich zwischen konventionellen Methoden und statistischen Verfahren

Autorinnen und Autoren: Michael Bosnjak^{1,2*}, Stefan Dahm², Ronny Kuhnert², Dennis Weihrauch³, Angelika Schaffrath Rosario², Julia Hurraß³, Patrick Schmich², Lothar H. Wieler^{4,2*}

Institutionen: 1 Universität Trier, Abteilung für Psychologische Methodenlehre, Trier 2 Robert Koch-Institut, Abteilung für Epidemiologie und Gesundheitsmonitoring, Berlin 3 Gesundheitsamt Köln, Infektions- und Umwelthygiene, Köln 4 Digital Global Public Health, Hasso-Plattner-Institut, Universität Potsdam *Institution zum Zeitpunkt des Projektstarts

Abstract

Hintergrund: Einige COVID-19-Testzentren haben manipulierte Testanzahlen bei Antigentests/Schnelltests gemeldet. Diese Studie vergleicht statistische Ansätze mit herkömmlichen Betrugserkennungsverfahren. Analysiert wurde das Ausmaß der Übereinstimmung zwischen den herkömmlichen und den statistischen Methoden und inwieweit statistische Ansätze zusätzliche Verdachtsfälle identifizieren können.

Methode: Ausreißeranalyse bei gemeldeten Tests, Modellierung der Positivrate (Poisson-Regression), Prüfung von Verteilungsannahmen bezüglich der ersten (Benford-Gesetz) und der letzten Ziffer der Anzahl gemeldeter Tests. Basis der Analysen waren Abrechnungsdaten (April 2021 bis August 2022) von 907 Testzentren aus einer deutschen Stadt.

Ergebnisse: Die positive Übereinstimmung zwischen den konventionellen und statistischen Ansätzen („Sensitivität“) lag zwischen 8,6% und 24,7%, die negative Übereinstimmung („Spezifität“) zwischen 91,3% und 94,6%. Der Anteil der durch statistische Ansätze zusätzlich identifizierten potenziell betrügerischen Testzentren lag zwischen 7,0% und 8,7%. Die Kombination mindestens zweier statistischer Verfahren ergab eine optimale Detektionsrate von Testzentren mit zuvor unentdecktem Anfangsverdacht.

Schlussfolgerungen: Die statistischen Ansätze waren effektiver und systematischer bei der Identifizierung betrugsverdächtiger Testzentren als die konventionellen Methoden. Testzentren sollten bei zukünftigen Pandemien angehalten werden, Paradata (z. B. Zeitstempel der Testung) abzubilden.

Keywords: COVID-19-Testabrechnungen, Abrechnungsbetrug im Gesundheitswesen, Statistische Betrugsermittlung

1. Einleitung

1.1 Hintergrund und Überblick

Die COVID-19-Pandemie hat die Notwendigkeit schneller und umfassender Antigentestungen zur Eindämmung der Virusausbreitung gezeigt [1].

Die Nachfrage nach umfassenden COVID-19-Testkapazitäten wurde in vielen Ländern durch finanzielle Anreize für bestehende und neu gegründete Testzentren gedeckt. In Deutschland wurde durch eine gesetzliche Regelung („Corona-Testverordnung“, TestV [2]) die Erstattung der Testkosten aus Steuermitteln garantiert („Bürgertests“). Während solche Anreize entscheidend dazu beigetragen haben, schnell umfangreiche Testkapazitäten aufzubauen, hat dies – möglicherweise in

Informationen zu Artikel und Zeitschrift

Eingereicht: 20.12.2023
Akzeptiert: 08.04.2024
Veröffentlicht: 19.06.2024


Artikel peer reviewed

Zitierweise: Bosnjak M, Dahm S, Kuhnert R, Weihrauch D, Schaffrath Rosario A et al. (2024) Identifizierung von potentielltem Abrechnungsbetrug bei COVID-19-Schnelltests: Ergebnisse einer Pilotstudie zum Vergleich zwischen konventionellen Methoden und statistischen Verfahren J Health Monit. 2024;9(2):e 12093. doi: 10.25646/12093

Prof. Dr. Michael Bosnjak
bosnjak@uni-trier.de

Robert Koch-Institut, Berlin
Journal of Health Monitoring
www.rki.de/jhealthmonit

Englische Version des Artikels
www.rki.de/jhealthmonit-en

 Open access



[CC BY 4.0 Lizenzvertrag](https://creativecommons.org/licenses/by/4.0/)
[Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/)

Gesundheitsberichterstattung des Bundes.
Gemeinsam getragen von RKI und Destatis.



Das Robert Koch-Institut ist ein
Bundesinstitut im Geschäftsbereich des
Bundesministeriums für Gesundheit

Kombination mit unzureichenden betrugspräventiven Maßnahmen – auch Abrechnungsbetrug begünstigt, was den Steuerzahlern finanziellen Schaden zugefügt hat. Insbesondere wurde in den Medien berichtet, dass einige Testzentren die Anzahl tatsächlich durchgeführter COVID-19-Tests überhöht angegeben haben und/oder Testtermine in Rechnung gestellt haben, die vereinbart, aber dann abgesagt wurden [3]. Es wurde auch berichtet, dass Testzentren überhaupt nicht existierten (kein tatsächlicher Geschäftsort), aber dennoch ihre gefälschten Rechnungen erstattet bekamen [4]. Wohl auch aufgrund solcher Medienberichte wurde vom Bundesgesundheitsministerium (BMG) beschlossen, verschiedene Maßnahmen zur Identifizierung potenziell betrügerischer COVID-19-Testzentren einzuführen. Die Entwicklung von statistischen Methoden zur Aufdeckung von verdächtigen Testzentren war darunter ein Ansatz.

Die vorliegende Arbeit stellt die Ergebnisse einer Pilotstudie des Robert Koch-Instituts vor, in der verschiedene statistische Ansätze zur Identifizierung betrugsverdächtiger Testzentren im Vergleich zu herkömmlichen Verfahren zur Betrugserkennung, die von Gesundheitsämtern standardmäßig verwendet werden, untersucht wurden. Hierfür wurden Daten von einem deutschen Gesundheitsamt aus einer Stadt mit mehr als 900 Testzentren verwendet. Diese Pilotstudie kann als Grundlage für die flächendeckende Anwendung statistischer Ansätze zur Aufdeckung betrügerischer Testzentren gesehen werden und dient gleichzeitig als Beitrag für eine akkuratere Abbildung des Testgeschehens bei zukünftigen Pandemien.

1.2 Forschungsfrage

Im Falle eines Ausbruchsgeschehens spielen die lokalen Gesundheitsämter – in Deutschland sind es etwa 400 – eine Schlüsselrolle bei der Entscheidung über und Umsetzung von geeigneten Maßnahmen. Während der COVID-19-Pandemie hatten sie auch die Aufgabe, Testzentren zu überwachen und Informationen über potenziell betrügerische Aktivitäten an die Strafverfolgungsbehörden weiterzuleiten. Testzentren galten in der Regel als betrugsverdächtig, wenn sie mindestens eines der folgenden Kriterien erfüllten: übermäßig hohe Testzahlen, deutliche Abweichungen von der durchschnittlichen Positivrate (dem Anteil positiver Testergebnisse) oder Hinweise aus der Bevölkerung (Bürgerinnen und Bürger, aktuelle oder ehemalige Mitarbeiterinnen und Mitarbeiter der Testzentren). Darüber hinaus wurden Testzentren als verdächtig betrachtet, wenn sie Gegenstand polizeilicher oder strafrechtlicher Ermittlungen waren. Typische Anlässe für solche Ermittlungen waren über die Meldungen aus der Bevölkerung hinaus Banktransaktionen, die ungewöhnlich hohe Transaktionen auf den Konten der Testzentren verzeichneten.

Diese herkömmlichen Ansätze sind zeitaufwändig, kaum einheitlich umgesetzt und ressourcenintensiv. Es stellt sich die Frage, ob betrugsverdächtige COVID-19-Testzentren anhand der Abrechnungsdaten auch auf der Grundlage von systematisch eingesetzten statistischen Ansätzen identifiziert werden können und diese somit dazu beitragen können, die verfügbaren Ressourcen effizient zu nutzen. Entsprechend wurden in der Pilotstudie zwei Forschungsfragen verfolgt: Erstens, wie hoch ist die Übereinstimmung zwischen herkömmlichen Methoden zur Identifizierung betrugsverdächtiger Testzentren und einer Reihe von statistischen Methoden, die auf die Abrechnungsdaten der COVID-19-Schnelltests angewendet werden? Zweitens, wie hoch ist bei den durch die statistischen Methoden als verdächtig klassifizierten Testzentren der Anteil derer, die bislang nicht aufgefallen waren und bei denen in nachfolgenden gründlichen Ermittlungen der Betrugsverdacht bestätigt werden konnte?

Die Studie vergleicht statistische Methoden mit herkömmlichen Verfahren zur Betrugserkennung.

2. Methode

2.1 Daten

In der Analyse wurden Daten zu COVID-19-Antigentests verwendet, die von 907 Testzentren in einer deutschen Großstadt im Zeitraum vom 8. April 2021 bis zum 28. August 2022 zu Abrechnungszwecken übermittelt wurden. Die Übermittlung fand über ein durch das zugehörige deutsche Landesministerium erstelltes Online-Portal statt, in dem jedes Testzentrum täglich die entsprechenden Daten eintrug. Zu dieser Übermittlung waren die Testzentren gesetzlich verpflichtet. Für jeden Datenpunkt sind folgende Angaben vorhanden: Teststellentyp (Apotheke, Arzt- oder Zahnarztpraxis, privates Testzentrum), Datum des Tests, Anzahl der durchgeführten Tests pro Tag, Anzahl der positiven Tests pro Tag. [Tabelle 1](#) enthält deskriptive Angaben zu den Daten der vorliegenden Pilotstudie.

2.2 Konventionelle und statistische Methoden

Konventionelle Methode: Erkennen von Abrechnungsbetrug durch die Gesundheitsämter

Verschiedene Indikatoren konnten dazu führen, dass ein Testzentrum als betrugsverdächtig klassifiziert wurde: eine überhöhte Anzahl durchgeführter Tests, eine signifikant erniedrigte Positivrate oder Hinweise aus der Bevölkerung. Falls eine daraufhin eingeleitete Vor-Ort-Inspektion den

Tabelle 1: Anzahl der Teststellen, Anzahl der Teststellen unter Betrugsverdacht durch das Gesundheitsamt, Anzahl der Tests und Anteil der positiven Tests nach Teststellentyp. Quelle: Daten zu täglichen COVID-19-Antigentests von 907 Testzentren im Zeitraum vom 8. April 2021 bis 28. August 2022

Teststellentyp	Anzahl der Teststellen	Anzahl der beim Gesundheitsamt unter Betrugsverdacht stehenden Teststellen	Gesamtzahl der Tests	Anzahl der Tests pro Tag und Teststelle			Anteil der positiven Tests (Positivrate in %)
				Mittelwert	Maximum	Median	
Apothek	60	NA	1.657.633	108	3.620	64	2,44
Arzt- oder Zahnarztpraxis	390	4	485.434	21	4.483	5	2,75
Privates Testzentrum	457	89	15.648.440	194	37.373	115	1,95
Gesamt	907	93	17.791.507	150	37.373	79	2,02

NA = nicht verfügbar

Daten zur Anzahl der beim Gesundheitsamt unter Betrugsverdacht stehenden Apotheken liegen nicht vor, da Apotheken unter der Aufsicht einer anderen Abteilung stehen.

Betrugsverdacht nicht ausräumen konnte, galt das Testzentrum als verdächtig. Auch Testzentren, die Gegenstand polizeilicher oder strafrechtlicher Ermittlungen waren, wurden als verdächtig betrachtet. All diese herkömmlichen Verfahren wurden sporadisch umgesetzt und konnten nicht systematisch eingesetzt werden. Zusätzlich waren der Umfang und die Intensität der Prüfung stark von den vorhandenen personellen und finanziellen Ressourcen abhängig.

Betrugsverdacht aufgrund von statistischen Auffälligkeiten

Ein Testzentrum wurde als statistisch auffällig und damit betrugsverdächtig klassifiziert, wenn mindestens einer der folgenden vier Indikatoren auffällig war: eine im Vergleich zur durchschnittlichen Anzahl von Tests pro Tag überhöhte Anzahl von abgerechneten Tests (unterteilt nach Teststellentyp: Apotheke, Arzt- oder Zahnarztpraxis, privates Testzentrum); eine niedrige Positivrate eines Testzentrums (nach Teststellentyp und analysiert mit einer Poisson-Regression); oder Abweichungen der gemeldeten Anzahlen vom Benford'schen Gesetz oder von der Annahme gleichverteilter Endziffern. Im Gegensatz zum herkömmlichen Ansatz repräsentieren diese statistischen Methoden einen systematischen Workflow, bei dem eine Reihe von statistischen Prüfverfahren auf alle Testzentren angewendet werden. Die statistischen Methoden werden unten näher beschrieben.

2.3 Ergebniskriterien

Zum Vergleich zwischen konventionellen und statistischen Methoden wurden drei Ergebniskriterien herangezogen. Zunächst wurde der Prozentsatz mit positiver Übereinstimmung (Formel 1) betrachtet (entsprechend dem Konzept der Sensitivität in Studien zur Güte diagnostischer Tests [5]):

$$\text{positive Übereinstimmung} = \frac{\text{Anzahl (statistisch auffällig und Betrugsverdacht beim Gesundheitsamt)}}{\text{Anzahl (Betrugsverdacht beim Gesundheitsamt)}} \cdot 100 \quad (1)$$

Als zweites Ergebniskriterium (Formel 2) diente der Prozentsatz der negativen Übereinstimmung (entsprechend dem Konzept der Spezifität in Studien zur Güte diagnostischer Tests [5]):

$$\text{negative Übereinstimmung} = \frac{\text{Anzahl (nicht statistisch auffällig und kein Betrugsverdacht beim Gesundheitsamt)}}{\text{Anzahl (kein Betrugsverdacht beim Gesundheitsamt)}} \cdot 100 \quad (2)$$

Als drittes Ergebniskriterium wurde der Anteil der durch die neue Methode zusätzlich identifizierten, potenziell betrügerischen Testzentren berechnet (Formel 3), die von konventionellen Methoden nicht entdeckt worden waren (entsprechend 1 minus Spezifität bzw. 100 minus negative Übereinstimmung):

$$\text{inkrementell identifizierte, potenziell betrügerische Testzentren} = \frac{100 - \text{negative Übereinstimmung}}{100 - \text{negative Übereinstimmung}} \quad (3)$$

2.4 Verwendete statistische Methoden zur Erkennung auffälliger Teststellen

Die im Folgenden beschriebenen statistischen Verfahren wurden auf die Abrechnungsdaten angewandt. Eine umfassende Beschreibung findet sich in [6].

Identifikation von Ausreißern anhand einer hohen Anzahl von abgerechneten Tests

Testzentren mit einer ungewöhnlich hohen durchschnittlichen Anzahl abgerechneter Tests pro Tag im Vergleich zu anderen Testzentren derselben Kategorie (Apotheke, Arzt- oder Zahnarztpraxis, privates Testzentrum) wurden als auffällig eingestuft. Die Grenze wurde für jeden Teststellentyp auf das jeweilige 90. Perzentil festgelegt. Das Histogramm in [Abbildung 1](#) visualisiert Ausreißer (dunkelblau markiert), d. h. auffällige Testzentren.

Poisson-Regression zur Identifikation von Ausreißern anhand auffällig niedriger Positivraten

Bei diesem Ansatz wird ermittelt, inwieweit einzelne Testzentren eine zu geringe Positivrate berichteten. Die Geeignet-

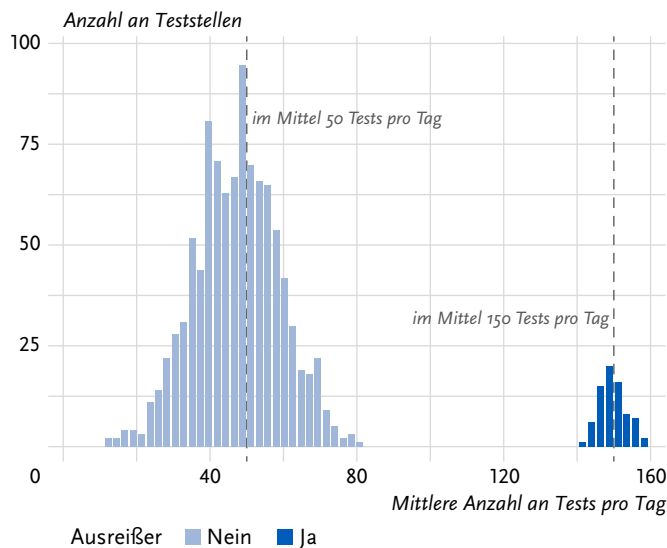


Abbildung 1: Histogramm zweier simulierter Verteilungen: Durchschnittliche Anzahl abgerechneter Tests pro Tag mit Identifizierung von Ausreißern. Quelle: Eigene Darstellung

heit der niedrigen Positivrate der Testzentren als Indikator für einen Anfangsverdacht kann darauf zurückgeführt werden, dass eine positive Testung Auflagen für die betroffene Person nach sich zog. Förderlich für einen Betrug wäre deshalb eine negative Testung gewesen, da sie keinerlei weitere Maßnahmen nach sich zog.

Die Anzahl der positiven Tests pro Tag und Testzentrum wurde durch eine Poisson-Regression mit zufälligen Effekten modelliert. Die Poisson-Regression ist eine Regressionsanalyse, die zur Modellierung von Häufigkeiten genutzt wird [7,8]. Sie kann auch verwendet werden, um die Positivrate zu modellieren, d. h. den Anteil der positiven Tests pro Tag und Testzentrum, indem die jeweilige Gesamtzahl der Tests als sogenanntes Offset ins Modell aufgenommen wird. Zusätzlich zu diesem Offset enthielt das Modell die Teststellen-ID ($j = 1, \dots, 907$) und einen Kalenderwochen-Indikator ($k = 1, \dots, 73$) als zufällige Effekte (als random intercepts, also als zufällige Achsenabschnitte im Modell) sowie den Teststellentyp (Apotheke, Arzt- oder Zahnarztpraxis, privates Testzentrum) als festen Prädiktor. Die Kalenderwoche wurde ins Modell genommen, um Veränderungen der Positivrate im Laufe der Zeit, wie sie beispielsweise durch eine veränderte Inzidenz von COVID-19-Infektionen verursacht werden, zu berücksichtigen. Eine ausführliche Darstellung des Modells findet sich in [8].

Die Variabilität zwischen den Teststellen wurde über einen zufälligen Effekt modelliert, der für jede Teststelle die Abweichung vom allgemeinen Mittel beschreibt. Dabei weist ein niedriger Schätzwert für eine Teststelle auf eine niedrige Positivrate in dieser Teststelle hin. Die von einer Teststelle gemeldeten Daten wurden als auffällig betrachtet, wenn der Schätzwert für eine Teststelle signifikant niedrig war (siehe [6]). Die Analyse wurde mit R, Version 4.2.2, und dem R-Paket lme4, Version 1.1–31, durchgeführt.

Abweichungen vom Benford'schen Gesetz

Das Benford'sche Gesetz [9] (siehe auch [10,11]) ist eine Beobachtung über die Häufigkeitsverteilung der führenden Ziffern in realen numerischen Datensätzen. Es besagt, dass die erste Ziffer in natürlich vorkommenden Datensätzen mit höherer Wahrscheinlichkeit eine kleine und keine große Ziffer ist. Genauer gesagt beträgt die Wahrscheinlichkeit, dass die führende Ziffer eine 1 ist, etwa 30%, während die Wahrscheinlichkeit, dass die führende Ziffer eine 9 ist, nur etwa 5% beträgt. [Abbildung 2](#) zeigt die nach der Benford-Verteilung erwarteten Wahrscheinlichkeiten für die einzelnen Ziffern als erste Ziffer.

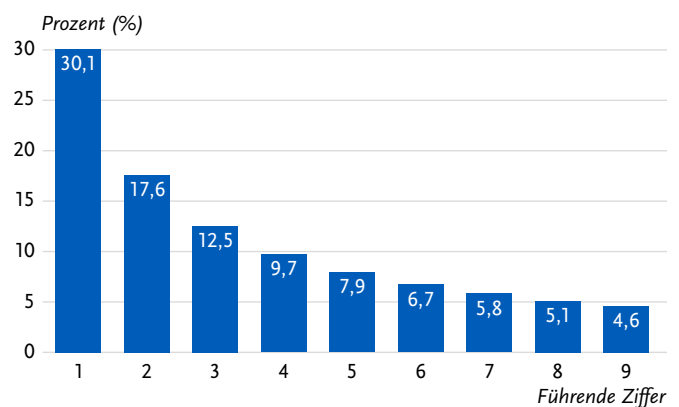


Abbildung 2: Erwartete Verteilung der führenden Ziffern gemäß dem Benford'schen Gesetz. Quelle: Eigene Darstellung nach [9]

Dieses Gesetz ist nach dem amerikanischen Physiker Frank Benford benannt, der es erstmals 1938 beschrieben hat. Es wurde festgestellt, dass es für eine Vielzahl von Datensätzen gilt, einschließlich Finanzdaten (z. B. [12]), Umfragedaten [13] und physikalischen Messungen (z. B. [14,15]). Forscherinnen und Forscher haben Benfords Gesetz auf COVID-19-Daten in verschiedenen Kontexten angewendet, darunter die Analyse der Verteilung erster Ziffern in gemeldeten Fall- und Todeszahlen (z. B. [16,17]).

In die Analyse zur Überprüfung des Benford'schen Gesetzes wurden nur Teststellen einbezogen, die an mindestens 30 Tagen Tests durchgeführt hatten, da nur dann die Verteilung der ersten Ziffer beurteilt werden kann. Für jede dieser Teststellen wurde ein Chi-Quadrat-Testwert zur Abweichung vom Benford'schen Gesetz berechnet. Die 10% Teststellen mit den größten Chi-Quadrat-Testwerten wurden als statistisch auffällig klassifiziert. Teststellen, die weniger als 30 Tage in Betrieb waren, wurden als unauffällig klassifiziert. Die Analyse wurde mit dem R-Paket BenfordTests, Version 1.2.0, durchgeführt. Der entsprechende Analysecode sowie die Beschreibung der Voraussetzungen bei der Anwendung des Benford'schen Gesetzes finden sich in [6].

Abweichungen von der Annahme der Gleichverteilung der letzten Ziffer

Diese Methode untersucht nicht die erste Ziffer, wie es beim Benford'schen Gesetz der Fall ist, sondern die Verteilung der letzten Ziffer. Die Annahme ist, dass die letzte Ziffer von wahrheitsgetreuen Anzahlen vollständig zufällig verteilt ist, sodass jede der zehn Ziffern (0 bis 9) mit einer Wahrscheinlichkeit von 10% in den Daten vorhanden sein sollte. Gemeldete Testzahlen, die unverhältnismäßig oft als letzte Ziffer die 0 oder 5 aufweisen, wurden wahrscheinlich gerundet und somit manipuliert. Wie bei der Überprüfung des Benford'schen Gesetzes wurden auch hier nur Teststellen in die Prüfung einbezogen, die an mindestens 30 Tagen Tests gemeldet hatten. Als zusätzliche Voraussetzung wurden nur Tage mit mehr als neun durchgeführten Tests gezählt. Die übrigen Teststellen wurden als unauffällig klassifiziert. Die 10% der Teststellen mit den größten Chi-Quadrat-Testwerten wurden als statistisch auffällig festgelegt. Die Voraussetzungen für die Anwendung der Methode zur Überprüfung der letzten Ziffer sind in [6] beschrieben.

2.5 Prüfung der inkrementellen Validität der statistischen Methoden

Um den Zuwachs an Validität (inkrementelle Validität) durch die Anwendung statistischer Ansätze im Vergleich zu den konventionellen Methoden zu bewerten, wurden alle statistisch auffälligen Testzentren, für die beim Gesundheitsamt noch kein Betrugsverdacht vorgelegen hatte, einer nachgelagerten gründlichen Überprüfung unterzogen. Basierend auf den Daten aus zurückliegenden und in einigen Fällen neu durchgeführten Vor-Ort-Inspektionen der statistisch auffälligen Testzentren wurden verschiedene Faktoren als mögliche Erklärungen für Ausreißer bei den Positivraten oder den täglichen Testzahlen berücksichtigt. Dazu gehörten die Öffnungszeiten, die Anzahl der Testkabinen, die Anzahl der Beschäftigten und der Standort. Wenn möglich, wurden auch mögliche Ursachen für Anomalien im Benford'schen Gesetz oder der Methode der letzten Ziffer (z. B. Testzentren mit einer konstanten Anzahl von Terminen pro Tag) untersucht. Informationen aus vorliegenden Beschwerden und von ermittelnden Behörden wurden genutzt, um den Verdacht auf Betrug zu erhärten oder zu widerlegen. Das Ergebnis dieser Untersuchung diente als dichotome Ergebnisvariable, aus der die Prognosegüte bzw. Vorhersagevalidität [18] berechnet wurde. Die Vorhersagevalidität ergibt sich als der Anteil inkrementell identifizierter Teststellen mit bestätigtem Betrugsverdacht. Dieser Anteil berechnet sich aus der Anzahl der Testzentren, die zunächst unverdächtig beim Gesundheitsamt aber statistisch auffällig waren und bei denen sich zusätzlich der Betrugsverdacht erhärtet hat, und der Anzahl der Testzentren, die insgesamt statistisch auffällig waren:

$$\text{Inkrementelle Validität} = \frac{\text{Anzahl (neu auffällig und Betrugsverdacht erhärtet)}}{\text{Anzahl (statistisch auffällig insgesamt)}} \cdot 100 \quad (4)$$

3. Ergebnisse

Im Folgenden werden die Ergebnisse der statistischen Prüfung der Abrechnungsdaten zusammengefasst. Eine detaillierte Darstellung findet sich in [6].

3.1 Hohe Anzahl an abgerechneten Tests pro Tag

Mit der ersten statistischen Methode wurden übermäßig hohe abgerechnete Testzahlen als auffällig eingestuft. Dabei wurde als Grenze das 90. Perzentil in Bezug auf die durchschnittliche Anzahl der pro Tag durchgeführten Tests innerhalb des Teststellentyps verwendet. Insgesamt wurden 91 Teststellen (6 Apotheken, 39 ärztliche oder zahnärztliche Praxen und 46 private Testzentren) als auffällig eingestuft. [Tabelle 2](#) fasst diese Zahlen im Vergleich zum konventionellen Ansatz zusammen. Der Prozentsatz der positiven Übereinstimmung zwischen den konventionellen und statistischen Methoden beträgt 24,7% (23/93), der Prozentsatz der negativen Übereinstimmung 91,6% (746/814) und der Anteil der zusätzlich identifizierten potenziell betrügerischen Testzentren, die von der statistischen Methode als auffällig identifiziert wurden und bei konventionellen Ansätzen unentdeckt blieben, beträgt 8,4% (68/814).

Manche COVID-19-Testzentren haben in betrügerischer Weise die Anzahl durchgeführter Tests falsch angegeben, um höhere Erstattungen geltend zu machen.

3.2 Auffällig niedrige Positivraten in der Poisson-Regressionsanalyse

Im Berichtszeitraum (8. April 2021 bis 28. August 2022) wurden COVID-19-Tests an 508 Tagen in 907 Teststellen durchgeführt; nicht alle Teststellen waren während des gesamten Zeitraums in Betrieb. An diesen Tagen meldeten die Teststellen insgesamt $N = 118.892$ positive Tests, deren Positivrate durch eine Poisson-Regression modelliert wurde. Dabei wurden 88 (9,7%) der 907 Teststellen als statistisch auffällig im Hinblick auf ihre relativ niedrigen Raten positiver Tests eingestuft ([Tabelle 2](#)). Die positive Übereinstimmung zwischen der konventionellen Methode und der Klassifizierung durch Poisson-Regression beträgt 18,3% (17/93), die negative Übereinstimmung 91,3% (743/814), und der Anteil der durch die Poisson-Regression zusätzlich identifizierten potenziell betrügerischen Teststellen beträgt 8,7% (71/814).

Tabelle 2: Anzahl der Teststellen (n = 907) nach Teststellentyp, mit und ohne Betrugsverdacht nach konventionellem Ansatz und Auffälligkeit in den vier statistischen Methoden. Quelle: Daten zu täglichen COVID-19-Antigentests von 907 Testzentren im Zeitraum vom 8. April 2021 bis 28. August 2022

Statistische Methode	Teststellentyp	Beim Gesundheitsamt unter Betrugsverdacht stehende Teststellen			Kein Betrugsverdacht durch das Gesundheitsamt		
		Statistisch auffällig	Statistisch unauffällig	Gesamt	Statistisch auffällig	Statistisch unauffällig	Gesamt
Hohe Testanzahl	Apotheke*	0	0	0	6	54	60
	Arzt- oder Zahnarztpraxis	4	0	4	35	351	386
	Privates Testzentrum	19	70	89	27	341	368
	Gesamt	23	70	93	68	746	814
Niedrige Positivrate	Apotheke*	0	0	0	11	49	60
	Arzt- oder Zahnarztpraxis	1	3	4	15	371	386
	Privates Testzentrum	16	73	89	45	323	368
	Gesamt	17	76	93	71	743	814
Benford'sches Gesetz	Apotheke*	0	0	0	5	55	60
	Arzt- oder Zahnarztpraxis	0	4	4	8	378	386
	Privates Testzentrum	10	79	89	44	324	368
	Gesamt	10	83	93	57	757	814
Letzte Ziffer	Apotheke*	0	0	0	1	59	60
	Arzt- oder Zahnarztpraxis	1	3	4	14	372	386
	Privates Testzentrum	7	82	89	29	339	368
	Gesamt	8	85	93	44	770	814

*Die Daten zur Anzahl der beim Gesundheitsamt unter Betrugsverdacht stehenden Apotheken liegen nicht vor, da Apotheken unter der Aufsicht einer anderen Abteilung des Gesundheitsamtes standen.

3.3 Abweichungen vom Benford'schen Gesetz

665 der 907 Teststellen hatten für mindestens 30 Tage Tests gemeldet und konnten somit in die Analysen zur Überprüfung der Verteilung der ersten Ziffer anhand des Benford'schen Gesetzes einbezogen werden. Teststellen mit Meldungen für weniger als 30 Tage galten als unauffällig. Insgesamt wurden 67 (7,4%) aller 907 Teststellen als auffällig klassifiziert. Der Prozentsatz der positiven Übereinstimmung zwischen den herkömmlichen Methoden und der Überprüfung des Benford'schen Gesetzes beträgt 10,8% (10/93) (Tabelle 2), der Prozentsatz der negativen Übereinstimmung beträgt 93,0% (757/814), und der Anteil der durch die Überprüfung des Benford'schen Gesetzes zusätzlich identifizierten potenziell betrügerischen Testzentren beträgt 7,0% (57/814).

3.4 Abweichungen von der Annahme der Gleichverteilung der letzten Ziffer

Auch bei der Überprüfung der Gleichverteilung der letzten Ziffer konnten nur 512 Teststellen mit einer ausreichenden Anzahl an Tageswerten untersucht werden, die übrigen galten als unauffällig. 52 (5,7%) aller 907 Teststellen wurden als statistisch auffällig klassifiziert. Der Prozentsatz der positiven Übereinstimmung zwischen dem konventionellen Ansatz und der Endziffermethode beträgt 8,6% (8/93), der Prozentsatz der negativen Übereinstimmung 94,6% (770/814). Der

Anteil der durch die Endziffermethode zusätzlich identifizierten potenziell betrügerischen Teststellen, die mit dem konventionellen Ansatz unentdeckt geblieben waren, beträgt 5,4% (44/814).

3.5 Inkrementelle Validität der statistischen Ansätze zur Betrugserkennung

In den vorherigen Abschnitten wurde der inkrementelle Beitrag jeder statistischen Methode im Vergleich zu konventionellen Ansätzen zur Erkennung von möglichem Abrechnungsbetrug der Teststellen untersucht. Als nächstes lag der Fokus auf der Prognosegüte bzw. Vorhersagevalidität der statistischen Methoden. Genauer gesagt, welcher Anteil der als statistisch auffällig identifizierten Teststellen konnte nach einer erneuten Untersuchung durch das Gesundheitsamt als Verdachtsfall bestätigt werden? Tabelle 3 zeigt die Ergebnisse, stratifiziert nach den einzelnen statistischen Methoden und nach verschiedenen Kombinationen der Methoden.

Unter den einzelnen statistischen Ansätzen hat die Prüfung auf eine hohe Zahl von gemeldeten Tests die höchste Anzahl potenziell betrügerischer Teststellen als auffällig klassifiziert (91 Teststellen). Die Prognosegüte, also der inkrementelle Anteil (in Prozent) an Teststellen mit bestätigtem Betrugsverdacht, war am höchsten für die Prüfung der letzten Ziffer (Betrugsverdacht wurde in 34,6% der Fälle bestätigt), gefolgt von der Prüfung auf eine niedrige Positivrate

Tabelle 3: Inkrementelle prädiktive Validität der statistischen Methoden bei der Erkennung von Abrechnungsbetrug durch Teststellen. Quelle: Daten zu täglichen COVID-19-Antigentests von 907 Testzentren im Zeitraum vom 8. April 2021 bis 28. August 2022

	Statistische Methoden	Statistisch auffällige Teststellen N (in %)			
		Bereits beim Gesundheitsamt unter Betrugsverdacht stehende Teststellen	Nach zusätzlicher Prüfung durch das Gesundheitsamt bestätigter Verdacht		Gesamt
			Ja	Nein	
Statistische Methoden einzeln	Hohe Testanzahl	23 (25,3)	23 (25,3)	45 (49,5)	91 (100,0)
	Niedrige Positivrate	17 (19,3)	27 (30,7)	44 (50,0)	88 (100,0)
	Benford'sches Gesetz	10 (14,9)	12 (17,9)	45 (67,2)	67 (100,0)
	Letzte Ziffer	8 (15,4)	18 (34,6)	26 (50,0)	52 (100,0)
Statistische Methoden kombiniert	Auffällig in mindestens einer Methode	47 (20,9)	49 (21,8)	129 (57,3)	225 (100,0)
	Auffällig in mindestens zwei Methoden	9 (14,8)	25 (41,0)	27 (44,3)	61 (100,0)

(30,7%). Da die Prüfung auf eine niedrige Positivrate im Gegensatz zur Prüfung der letzten Ziffer auf alle Teststellen angewendet werden konnte, erzielte sie jedoch die höchste absolute Anzahl zusätzlich identifizierter Teststellen.

Von den durch statistische Methoden als auffällig identifizierten Testzentren wurden 41 % durch nachfolgende Ermittlungen als potentiell betrügerisch bestätigt. Diese Testzentren waren zuvor unverdächtig.

Die Kombination von Methoden mit dem ODER-Operator (statistische Auffälligkeit durch mindestens eine Methode) führte zu der höchsten Anzahl (N=49) von Teststellen mit bestätigtem Betrugsverdacht. Allerdings war der Anteil der neu entdeckten Teststellen mit bestätigtem Betrugsverdacht im Vergleich zu den einzelnen Methoden niedrig (21,8%). Dagegen scheint es bei begrenzten Testressourcen sinnvoll, die Untersuchung mit Teststellen zu beginnen, die in mindestens zwei Methoden statistisch auffällig sind (hier N=61). Bei diesen Teststellen ergab die Nachprüfung durch das Gesundheitsamt eine sehr hohe Trefferquote von 41,0%.

4. Diskussion

Diese Arbeit stellt die Ergebnisse einer Pilotstudie vor, in der die Eignung verschiedener statistischer Methoden zur Identifizierung von potentiell Abrechnungsbetrug durch COVID-19-Teststellen im Vergleich zu den herkömmlichen Verfahren der Gesundheitsämter beurteilt wurde. Hierzu wurden Abrechnungsdaten aus Teststellen in einer deutschen Großstadt ausgewertet.

Die statistischen Ansätze erwiesen sich im Vergleich zu konventionellen Methoden als systematischer und bieten deutliche Vorteile in Bezug auf den Umfang der Daten-

nutzung, Fairness aufgrund der Berücksichtigung aller Teststellen bei den Prüfungen, geringere Fehleranfälligkeit aufgrund zusätzlich aufgeklärter Betrugsfälle sowie Ressourceneffizienz und leichte Skalierbarkeit auf deutlich größere Datenbestände.

Zwei Forschungsfragen wurden empirisch untersucht: Erstens, wie hoch ist die Übereinstimmung zwischen konventionellen Methoden und den statistischen Ansätzen zur Identifizierung von potenziellem Abrechnungsbetrug durch Teststellen? Zur Beantwortung dieser Forschungsfrage wurden – für jeden statistischen Ansatz – drei Kennzahlen zur Bewertung geschätzt: der Prozentsatz positiver Übereinstimmung zwischen der konventionellen und der statistischen Methode (analog zum Konzept der Sensitivität in Studien zur Güte diagnostischer Tests), der Prozentsatz negativer Übereinstimmung (analog zum Konzept der Spezifität in Studien zur Güte diagnostischer Tests) und der Anteil, der durch die statistische Methode zusätzlich identifizierten, potenziell betrügerischen Testzentren, die mit konventionellen Methoden nicht erkannt worden waren (entsprechend 1 – negative Übereinstimmung). Der Prozentsatz positiver Übereinstimmung war bei der Prüfung hoher Anzahlen von gemeldeten Tests am höchsten. Bei der negativen Übereinstimmung zeigte die Prüfung der letzten Ziffer den höchsten Anteil. Die Prüfung der letzten Ziffer hatte auch, gefolgt von der Prüfung auf eine niedrige Positivrate, den höchsten Anteil an zusätzlich identifizierten, potenziell betrügerischen Testzentren.

Die zweite Forschungsfrage betraf die prognostische Validität der statistischen Ansätze: Wie hoch ist der Anteil der von den statistischen Ansätzen als auffällig identifizierten Testzentren, bei denen sich der Betrugsverdacht nach einer gründlichen Untersuchung durch das Gesundheitsamt erhärtete? Abhängig von den einzelnen Methoden und ihren Kombinationen haben wir den Anteil der zusätzlich identifizierten Teststellen, die zuvor von konventionellen Ansätzen als unauffällig eingestuft worden waren und bei denen sich

der Betrugsverdacht bestätigte, auf Werte zwischen 17,9% und 41,0% aller statistisch auffälligen Teststellen geschätzt, was den zusätzlichen Nutzen der statistischen Ansätze bestätigt.

Auf Grundlage dieser Ergebnisse empfehlen wir, alle in dieser Studie beschriebenen statistischen Ansätze parallel zu nutzen, um eine breite Grundlage zur Betrugsbekämpfung zu schaffen. Bei begrenzten Ressourcen sollten anschließend diejenigen Teststellen überprüft werden, die von mindestens zwei der vier Methoden als auffällig klassifiziert wurden. Diese Methodenkombination hatte die höchste inkrementelle prognostische Validität.

Die Verpflichtung der Testzentren zur Nutzung von Software, die detaillierte Daten zu den Tests erhebt (z. B. Zeitstempel und Testergebnis), würde die Betrugserkennung weiter verbessern.

Die Pilotstudie weist zwei Limitationen auf. Erstens waren die Voraussetzungen für die Durchführung einiger statistischer Tests nicht immer erfüllt. So ist beispielsweise die Voraussetzung für die Anwendung des Benford'schen Gesetzes, dass die zugrunde liegenden Daten eine ausreichend große Varianz aufweisen. Diese Annahme ist bei Teststellen, die täglich eine ähnliche Anzahl von Tests melden, verletzt. Zweitens ist der Nutzen der Abrechnungsdaten von Teststellen auf Tagesbasis begrenzt. In einer idealen Welt stünden Paradata (auch Begleitdaten genannt) zur Verfügung, die es ermöglicht hätten, jeden einzelnen Testvorgang zu rekonstruieren und dessen Glaubwürdigkeit zu bewerten. Zum Beispiel wäre mit Zeitstempeln der Tests eine Bewertung möglich gewesen, ob der zeitliche Rahmen der Tests plausibel war oder ob einige Tests außerhalb der regulären Betriebszeiten durchgeführt wurden. Für zukünftige Pandemien wird dringend empfohlen, von Teststellen zu verlangen, geeignete Software zu verwenden, welche die Speicherung solcher Paradata ermöglicht und diese zu Prüfzwecken laufend übermittelt.

Datenschutz und Ethik

Die bereitgestellten Daten unterliegen der strikten Einhaltung der datenschutzrechtlichen Bestimmungen der EU-Datenschutzgrundverordnung (DSGVO) und des Bundesdatenschutzgesetzes (BDSG).

Datenverfügbarkeit

Die Daten der COVID-19-Antigentests sind nicht öffentlich verfügbar und wurden nur für den Zweck der Pilotstudie vom Projektpartner [7] zur Verfügung gestellt. Zur Reproduktion der Analyseskripte ist ein künstlich generierter Pseudo-Datensatz unter [6] verfügbar.

Förderungshinweis

Keine Förderung.

Beiträge der Autorinnen und Autoren

Maßgebliche Beiträge zur Konzeption der Arbeit: MB, AS, SD, RK, PS, LW; zum Design der Arbeit: MB, AS, SD, RK; zur Erhebung der verwendeten

Daten: DW, JH; zur Analyse der verwendeten Daten: SD, RK; zur Interpretation der verwendeten Daten: MB, AS, SD, RK, DW; Ausarbeitung des Manuskripts: MB, AS, SD, RK; kritische Überarbeitung bedeutender Inhalte: LW, AS, MB, DW, JH, PS; Finale Version des Manuskripts gelesen und der Veröffentlichung zugestimmt: MB, AS, SD, RK, PS, LW, DW, JH.

Interessenkonflikt

Die Autorinnen und Autoren geben an, dass kein Interessenkonflikt besteht.

Danksagung

Stellvertretend für die vielen Kolleginnen und Kollegen, die im Gesundheitsamt an der Studie beteiligt waren, danken wir Roshanak Golmohammadi und Steffen Bujok für die Koordination des Teams, das über tausend Vor-Ort-Inspektionen durchgeführt hat. Ebenso gebührt der Dank dem gesamten Team Bürgertests für ihre komplexen und detaillierten Untersuchungen von statistisch auffälligen Testzentren.

Hinweis

Die ausführliche technische Beschreibung der Methoden, die detaillierte Beschreibung der Ergebnisse sowie Programmsyntax und Pseudo-Datensatz wurden als elektronischer Anhang unter Zenodo veröffentlicht [6].

Literatur

- Lopera TJ, Alzate-Angel JC, Diaz FJ, Rugeles MT, Aguilar-Jimenez W. The Usefulness of Antigen Testing in Predicting Contagiousness in COVID-19. *Microbiol Spectr*. 2022;10(2):e0196221. Epub 20220329. doi: 10.1128/spectrum.01962-21.
- Bundesministerium der Justiz und für Verbraucherschutz (Hrsg). Verordnung zum Anspruch auf Testung in Bezug auf einen direkten Erregernachweis des Coronavirus SARS-CoV-2 (Coronavirus-Testverordnung – TestV). *Bundesanzeiger BAnz AT 27.01.2021 V22021* [cited 08.04.2024]. Available from: <https://www.bundesanzeiger.de/pub/publication/rjv0GJUnP0bov4hKuas/content/rjv0GJUnP0bov4hKuas/BAnz%20AT%2027.01.2021%20V2.pdf?inline>.
- Grill M, Ott K. Betrug mit Corona-Schnelltests: Ermittler gehen von Schaden in Milliardenhöhe aus. 2022 [cited 08.04.2024]. Available from: <https://www.sueddeutsche.de/politik/schnelltests-corona-betrug-1.5590484?reduced=true>.
- Südwestrundfunk. Jugendstrafe für millionenschweren Corona-Testzentrum-Betrug in Freiburg. 2022 [cited 08.04.2024]. Available from: <https://www.swr.de/swraktuell/baden-wuerttemberg/suedbaden/betrugsprozess-corona-testzentrum-100.html>.
- Bland M. An introduction in medical statistics. Oxford: Oxford University Press Inc; 2009.
- Bosnjak M, Dahm S, Kuhnert R, Weihsrauch D, Schaffrath Rosario A, Hurraß J, et al. Appendix – Potential COVID-19 test fraud detection: Findings from a pilot study comparing conventional and statistical approaches. [Dataset] Zenodo; 2024 [cited 08.04.2024]. Available from: <https://zenodo.org/records/12168750>.
- Snijders TAB, Bosker RJ. Multilevel Analysis. London: SAGE Publications Ltd; 2012.
- Vittinghoff E, Glidden D, Shiboski S, McCulloch C. Regression Methods in Biostatistics. Second Edition. New York: Springer; 2012.
- Benford F. The law of anomalous numbers. *Proceedings of the American Philosophical Society*. 1938;78(4):551–72.
- Berger A, Hill TP. The mathematics of Benford's law: a primer. *Stat. Methods Appl*. 2020;30(3):779–95. doi: 10.1007/s10260-020-00532-8.
- Hill T. A Statistical Derivation of the Significant-Digit Law. *Statistical Science* 10(4) 1995;10(4). doi: 10.1214/ss/1177009869.
- Nigrini MJ. Benford's Law: Applications for forensic accounting, auditing, and fraud detection. Hoboken: John Wiley & Sons; 2012.
- Kaiser M. Benford's law as an indicator of survey reliability – can we trust our data? *Journal of Economic Surveys*. 2019;33(5):1602–18.
- Alexopoulos T, Leontsinis S. Benford's Law in Astronomy. *Astron. Astrophys*. 2014;35(4):639–48.
- Pain JC. Benford's law and complex atomic spectra. *Physical Review*. 2008;E77(1):012102.

- 16 Kolas P. Applying Benford's law to COVID-19 data: the case of the European Union. *J Public Health (Oxf)*. 2022;44(2):e221-e6. doi: 10.1093/pubmed/fdac005.
- 17 Silva L, Figueiredo Filho D. Using Benford's law to assess the quality of COVID-19 register data in Brazil. *J Public Health (Oxf)*. 2021;43(1):107-10. doi: 10.1093/pubmed/fdaa193.
- 18 Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52(4):281-302. doi: 10.1037/h0040957.