

Der gesellschaftliche Mehrwert verknüpfter Daten: Algorithmen als Entscheidungshilfen bei Antibiotikaverschreibungen

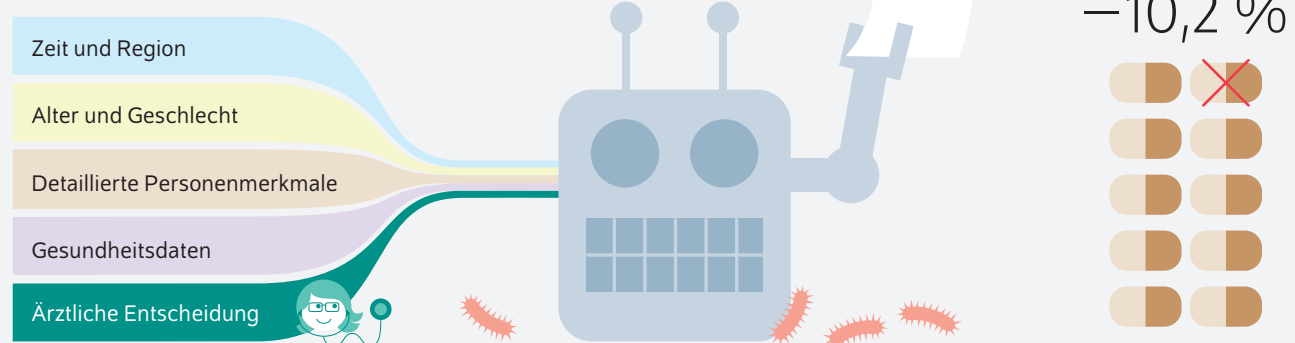
Von Shan Huang, Michael Allan Ribers und Hannes Ullrich

- Es wird die Frage untersucht, welchen gesellschaftlichen Nutzen die Verknüpfung und Auswertung von personalisierten Daten aus der Verwaltung (Administrativdaten) haben kann
- Fehlverschreibungen von Antibiotika verursachen wegen zunehmender Resistenzen erhebliche gesellschaftliche Kosten, die durch datenbasierte Entscheidungshilfen reduziert werden könnten
- Mit dänischen Daten wird analysiert, wie Methoden maschinellen Lernens und die Detailtiefe der verwendeten Daten Fehlverschreibungen reduzieren könnten
- Verschreibungen könnten unter Verwendung aller verfügbarer Daten um bis zu 10,2 Prozent verringert werden ohne die Gesamtzahl behandelter bakterieller Infektionen zu verringern
- Wissenschaftlich evaluierte Anwendungsbeispiele sind auch in anderen Bereichen nötig, um zwischen Nutzen und Risiken verknüpfter Daten abzuwägen

Menschliche Entscheidungen können durch maschinelles Lernen verbessert werden – die Verknüpfung von Daten ist dafür wichtig

Die Verknüpfung verschiedener Daten hilft dem Algorithmus

Unter Verwendung aller verfügbarer Administrativdaten könnte der Algorithmus Antibiotikaverschreibungen um 10,2 Prozent reduzieren



Quellen: Eigene Berechnung auf Basis dänischer Administrativdaten.

© DIW Berlin 2021

ZITAT

„In der Medizin und in vielen anderen Bereichen kann die gesamte Gesellschaft von verknüpften Administrativdaten profitieren. Um nützliche Anwendungsgebiete zu identifizieren, ist eine bessere Forschungsinfrastruktur nötig.“

— Shan Huang —

MEDIATHEK



Audio-Interview mit Hannes Ullrich
www.diw.de/mediathek

Der gesellschaftliche Mehrwert verknüpfter Daten: Algorithmen als Entscheidungshilfen bei Antibiotikaverschreibungen

Von Shan Huang, Michael Allan Ribers und Hannes Ullrich

ABSTRACT

Personendaten aus der Verwaltung (Administrativdaten) haben sich in den letzten Jahrzehnten als wichtige Basis zur Evaluierung von Politikmaßnahmen erwiesen. Durch maschinelles Lernen können basierend auf diesen Daten auch Vorhersagen getroffen werden, die zur Lösung gesellschaftlicher Problemstellungen beitragen. Hierzu gehören beispielsweise das Aufspüren von Steuerbetrug und eine bessere medizinische Versorgung. Welchen Mehrwert hierbei die Verknüpfung weitreichender Administrativdaten bietet, ist bisher noch nicht ausreichend untersucht worden. In diesem Beitrag wird am Beispiel von Antibiotikaverschreibungen für Harnwegsinfekte in Dänemark gezeigt, dass die Verknüpfung von Daten eine Verbesserung von Verschreibungsentscheidungen ermöglichen kann. Um für Deutschland vergleichbare Erkenntnisse zu gewinnen und datenbasierte, politikrelevante Anwendungen unter geltenden Datenschutzstandards zu identifizieren, ist ein Zugang zu verknüpften Administrativdaten für Forschungszwecke nötig. Hierfür müssen fragmentierte Datensilos weiter reduziert und Infrastrukturen mit Blick auf Anwendungsmöglichkeiten aufgebaut werden.

Durch die Auswertung verschiedener personenbezogener Daten aus der Verwaltung (Administrativdaten, Kasten 1) kann die Forschung Erkenntnisse gewinnen, die dem Gemeinwohl dienen. Der wachsende Zugang zu verknüpfbaren Administrativdaten hat in den letzten zwei Jahrzehnten zahlreiche hochwertige Forschungsbeiträge hervorgebracht.¹ Dadurch kommen häufig wertvolle wissenschaftliche und politikrelevante Erkenntnisse aus den Ländern, die bei der Bereitstellung von Administrativdaten Vorreiter sind und auch stärker evidenzbasiert politische Entscheidungen treffen.

Hierzu gehörten von Beginn an die skandinavischen Länder. Zugang zu vereinzelt Administrativdatensätzen besteht für die Forschung auch in weiteren Ländern. Für Analysen zur sozialen Mobilität anhand von Steuerdaten sind die USA prominente Beispiele.² In der SARS-Cov-2 Pandemie haben Politik und Öffentlichkeit in Israel und den skandinavischen Ländern regelmäßige Berichte zu Infektionszahlen nach etwa Berufsgruppen, Alter und sozioökonomischen Eigenschaften zur Hand. In Dänemark konnte man so bereits im Sommer 2020 die Übertragungsrisiken in Haushalten und Schulen quantifizieren und die Rolle von Alter und individueller Virenlast für Übertragungsrisiken in Haushalten beschreiben.³

Während die Nutzung einzelner Administrativdatenquellen bereits in vielen Fällen in Deutschland möglich ist, sind Verknüpfungen zwischen verschiedenen Datenquellen oft nicht möglich. Die Forderung nach einer umfanglicheren, gemeinwohlorientierten Nutzung personenbezogener Daten ist daher Kernthema aktueller Debatten. Ein Hauptziel politischer Initiativen ist die Förderung des Austausches und der

¹ Vgl. Liran Einav und Jonathan Levin (2014): Economics in the age of big data. *Science*, 346, 1243089.

² Vgl. Raj Chetty et al. (2014): Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *Quarterly Journal of Economics*, 129(4), 1553–1623; Raj Chetty et al. (2014): Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility. *American Economic Review Papers and Proceedings*, 104(5), 141–147.

³ Vgl. Frederik Plesner Lyngse et al. (2020): COVID-19 Transmission Within Danish Households: A Nationwide Study from Lockdown to Reopening. *medRxiv* 2020.09.09.20191239; Frederik Plesner Lyngse et al. (2021): Association between SARS-CoV-2 Transmission Risk, Viral Load, and Age: A Nationwide Study in Danish Households. *medRxiv* 2021.02.28.21252608.

Kasten 1

Erhebung und Nutzung personenbezogener Daten

In einer weitgehend digitalisierten Verwaltung werden personenbezogene Mikrodaten aus administrativen Zwecken als Nebenprodukt generiert. Die Nutzung solcher Administrativdaten für statistische Analysen bringt Vorteile mit sich, da häufig die gesamte Bevölkerung einbezogen werden kann. Gleichzeitig sind mit dieser Datenart aber auch Einschränkungen verbunden, da der Grund der Datenerhebung (Administration) nicht dem Ziel der Nutzung – zum Beispiel Forschung oder Politikevaluation – entspricht. Dennoch hat sich international gezeigt, dass Administrativdaten die Grundlage für vielfältige, gesellschaftlich relevante Erkenntnisse sein können. Grundvoraussetzungen dafür stellen eine präzise Zusammenführung aus unterschiedlichen Datenquellen, die sichere Zugangsgewährung zum Schutz vor Datendiebstahl und ein verantwortungsvoller, überprüfbarer Umgang mit den Daten dar. Dänemark bietet sich als Datenbasis für die vorliegende Studie an, weil dort seit geraumer Zeit und unter Einhaltung strenger Datenschutzregelungen Personendaten zu Forschungszwecken verknüpft werden dürfen.

Eine umfassendere Datenbasis der Bevölkerung kann mit großem gesellschaftlichem Nutzen einhergehen, aber auch den sozialen Zusammenhalt untergraben, wenn unbegrenzt persönliche Informationen gesammelt werden.¹ In demokratischen Staaten, deren Regierung und Verwaltung der Kontrolle der Bürger unterliegen, können durch die Auswertung solcher Daten effiziente Politik-

maßnahmen identifiziert werden. Zudem können systematische Fehlentscheidungen aufgedeckt werden und so insgesamt mehr Fairness und Transparenz im politischen Handeln erreicht werden. Entgleitet den Bürgern jedoch die Kontrolle über persönliche Daten, sind auch Szenarien einer digitalen Dystopie vorstellbar. Daher wird der öffentliche Diskurs dazu, für welche Zwecke und mit welcher Wirkung Daten verwendet werden dürfen, ein fortwährender Teil der politischen Diskussion bleiben.

Auch digitalisierte Firmen nutzen eine Fülle an Daten, um unter anderem ihre Ziele der Gewinnmaximierung und Unternehmenswertbildung zu erreichen. Dadurch wird zwar Wertschöpfung generiert, zum Beispiel, weil innovative Produkte und Dienstleistungen angeboten werden können. Die Datensammlung und -nutzung privater Akteure ist allerdings oft intransparent. Zum Beispiel hat die Akquise von Fitbit durch Google für Aufmerksamkeit gesorgt. Die von Fitbit vertriebene Smart Watch erfasst Gesundheitswerte, die Google nach dem Kauf mit reichhaltigen persönlichen Nutzerdaten verknüpfen kann. Diese Verknüpfungen ermöglichen Vorhersagen über den Gesundheitszustand aller Google-NutzerInnen und haben daher einen großen Wert, zum Beispiel für Versicherungsanbieter. Bei Diskussionen um die Nutzung und Verknüpfung von Daten ist es daher wichtig, die genauen Ziele und Rahmenbedingungen der Auswertungen öffentlich zu diskutieren und gegebenenfalls zu reglementieren.

1 Vgl. Jean Tirole (2021): Digital dystopia. American Economic Review, im Erscheinen.

Verknüpfung von Daten. Hierzu gehören Vorhaben wie die europäische Digitalstrategie, die europäische Cloud Gaia-X,⁴ aber auch Bemühungen zum Aufbau digitaler Infrastruktur und nationaler Forschungsdatenplattformen. Die im Januar 2021 beschlossene Datenstrategie der Bundesregierung sieht beispielsweise vor, die Nutzung von Daten durch Treuhänder und die Ansiedlung von „Chief Data Scientists“ in den öffentlichen Verwaltungen zu fördern. Bisher hat sich jedoch noch nicht herauskristallisiert, welche Anwendungen politische Dateninitiativen in der Praxis hervorbringen werden und welche Relevanz dabei die Verknüpfung von Daten hat. Anwendungsbeispiele und deren Evaluation sind nötig, um in konkreten Fällen Nutzen und Risiken abzuwägen.⁵

Für politische Entscheidungen nehmen Kausalanalysen eine bedeutende Stellung ein. Administrativdaten ermöglichen eine Vielzahl solcher Analysen, zum Beispiel zur Wirkung

von bestimmten Politikmaßnahmen auf finanzielle Vorsorgeentscheidungen, den Ursachen für steigende Lohnungleichheit in Deutschland, Ausweichverhalten nach Einführung einer Vermögenssteuer oder die Rolle der Elternschaft für Geschlechterunterschiede auf dem Arbeitsmarkt.⁶

Weniger Beachtung haben bisher gesellschaftliche Problemstellungen erhalten, bei denen reine Vorhersagen im Mittelpunkt stehen. Auch für diese Art von Entscheidungsproblemen können verknüpfte Administrativdaten von hoher Relevanz sein.⁷ Im medizinischen Bereich können beispielsweise datengetriebene Verfahren dazu beitragen, dass Behandlungsentscheidungen – wie etwa für oder gegen eine

6 Vgl. Raj Chetty et al. (2014): Active vs. Passive Decisions and Crowd-out in Retirement Savings Accounts: Evidence from Denmark. Quarterly Journal of Economics. 129(3), 1141–1219; David Card, Jörg Heining, and Patrick Kline (2013): Workplace heterogeneity and the rise of West German wage inequality. Quarterly Journal of Economics. 128(3), 967–1015; David Seim (2019): Behavioral Responses to Wealth Taxes: Evidence from Sweden. American Economic Journal: Economic Policy, 9(4), 395–421; Henrik Kleven, Camille Landais und Jakob Egholt Søgaard (2019): Children and Gender Inequality: Evidence from Denmark. American Economic Journal: Applied Economics, 11(4), 181–209.

7 Vgl. Jon Kleinberg et al. (2015): Prediction policy problems, American Economic Review Papers & Proceedings, 105(5), 491–95; Ajay Agrawal, Joshua Gans und Avi Goldfarb (2018): Prediction machines: the simple economics of artificial intelligence. Harvard Business Press; Michael A. Ribers und Hannes Ullrich (2019a): Battling Antibiotic Resistance: Can Machine Learning Improve Prescribing? DIW Diskussionspapier Nr. 1803 (online verfügbar).

4 Die europäische Digitalstrategie beinhaltet mehrere Pakete zur Digitalisierung in Europa, der Nutzung von Künstlicher Intelligenz und eine europäische Datenstrategie, vgl. European Commission – The European Digital Strategy (online verfügbar, abgerufen am 12. März 2021; das gilt auch für alle anderen Onlinequellen des Berichts sofern nicht anders vermerkt). Gaia-X ist eine sich im Aufbau befindende europäische Dateninfrastruktur (online verfügbar).

5 Vgl. Corinna Budras (2021): Frontalangriff auf die Datenstrategie. FAZ.net vom 11. Februar 2021 (online verfügbar); Barbara Gillmann (2021): Kooperation von Wirtschaft und Wissenschaft: Forschungsdaten bleiben oft ungenutzt. Handelsblatt vom 15. Februar 2021 (online verfügbar).

Verschreibung von Antibiotika – präziser und schneller getroffen werden. Es ist bekannt, dass Antibiotika nur wirksam sind, wenn Erkrankungen durch bakterielle Erreger verursacht sind. Für die Behandlungsentscheidung ist daher nur relevant, ob eine bakterielle Infektion vorliegt. Die Wahrscheinlichkeit hierfür kann mittels datengetriebener Verfahren bereits vor der Auswertung endgültiger Laborbefunde abgeschätzt werden.

Bei solchen Vorhersageproblemen ist im Vorhinein nicht bekannt, welche personenbezogenen Informationen zu einer guten Vorhersage beitragen. Daher muss in umfangreichen Anwendungsbeispielen evaluiert werden, welche Daten und die Verknüpfung welcher Informationen einen substanziellen Zugewinn für die Vorhersage bringen.

Mit datenbasierten Entscheidungen zunehmenden Antibiotikaresistenzen entgegenwirken

Die hier vorgestellte, vom Europäischen Forschungsrat geförderte Studie am DIW Berlin zeigt am Beispiel von Dänemark, welchen Mehrwert das Kombinieren von Administrativdaten für die Bewältigung eines der großen Public Health Probleme unserer Zeit haben kann.⁸ Grundlage für die Analyse bilden umfangreiche administrative Personendaten, die für Forschungszwecke mit Labordaten zur Diagnose von Harnwegsinfektionen verknüpft wurden.⁹

Bakterielle Harnwegsinfektionen werden in der Regel effektiv mit Antibiotika behandelt. PatientInnen können jedoch auch Symptome einer Harnwegsinfektion vorweisen, wenn sie nicht an einer bakteriellen Infektion erkrankt sind. In diesem Fall ist eine Behandlung mit Antibiotika nicht wirksam.

Neben Atemwegsinfektionen stellen Harnwegsinfektionen den Hauptverschreibungsgrund für Antibiotika in der Bevölkerung dar. Während die SARS-Cov-2 Pandemie die Welt noch im Griff hat, setzt sich der vorübergehend weniger beachtete Trend zunehmender antibiotikaresistenter Bakterien fort. Antibiotika haben zwar seit ihrer Entdeckung Millionen von Menschenleben gerettet. Hauptsächlich aufgrund mangelnder finanzieller Anreize sind jedoch seit geraumer Zeit kaum neue Wirkstoffe entwickelt worden, sodass die Behandlungsmöglichkeiten aufgrund der Resistenzen im Zeitverlauf weniger werden. Die durch resistente Erreger

verursachten Kosten werden für das Jahr 2050 mit 100 Billionen US-Dollar prognostiziert.¹⁰

Als Hauptgrund für die Zunahme an resistenten Erregern werden übermäßige Antibiotikabehandlungen gesehen. Jede Antibiotikaeinnahme trägt zur Zunahme der Anzahl an resistenten Erregern und damit zu einem gesamtgesellschaftlichen Problem bei. Daher wurden in den letzten Jahrzehnten viele Initiativen und Politikmaßnahmen initiiert und evaluiert, um Antibiotikaverschreibungen effizienter zu gestalten. Hierbei ist in der Regel das Ziel, Anreizprobleme auf Seiten der ÄrztInnen oder PatientInnen zu lösen.¹¹

Das eigentliche Hauptproblem ist jedoch, dass Behandlungsentscheidungen unter Unsicherheit getroffen werden müssen. Grund hierfür ist die mehrtägige prozessbedingte Verzögerung der für eine genaue Diagnostik nötigen Laboranalysen. Datenbasierte Vorhersagen können hier helfen, die vorübergehende Unsicherheit zu verringern und schnelle Entscheidungen zu verbessern. Da nicht überall die gleiche Fülle an Daten zur Verfügung steht, ist zunächst ein besseres Verständnis dafür nötig, welcher Umfang an Informationen verwendet werden müsste, um relevante Verbesserungen erreichen zu können.

Maschinelles Lernen kann Laborergebnisse prognostizieren

Durch Methoden des maschinellen Lernens wird im vorliegenden Fall vorhergesagt, mit welcher Wahrscheinlichkeit der Laborbefund eines Patienten oder einer Patientin zum Zeitpunkt der Probenentnahme bakterielle Erreger enthält. Hierzu werden zunächst große Datenmengen vergangener PatientInnen verwendet, für die verlässliche individuelle Laborbefunde vorliegen. Werden diese Laborbefunde mit anderen individuellen Administrativdaten verknüpft, können statistische Zusammenhänge zwischen Laborbefund und den weiteren vorhandenen persönlichen Informationen, zum Beispiel medizinische Behandlungshistorien, aufgedeckt werden. Die Verknüpfung der Laborergebnisse mit einer Fülle an weiteren individuellen Informationen aus anderen Administrativdaten ist daher für die Vorhersagemethode entscheidend. Bleiben Behandlungssituationen im Allgemeinen über die Zeit vergleichbar, können die in den Daten gefundenen Korrelationen verwendet werden, um Vorhersagen über Laborergebnisse aktueller PatientInnen zu treffen bevor diese vorliegen.

In der vorliegenden Studie werden hierfür sogenannte Ensemble Methoden, die zu den gängigsten Methoden maschinellen Lernens gehören, verwendet (*Extreme Gradient*

⁸ Vgl. Shan Huang, Michael A. Ribers und Hannes Ullrich (2021): The Value of Data for Prediction Policy Problems: Evidence from Antibiotic Prescribing. DIW Diskussionspapier Nr. 1939 (online verfügbar). Diese Forschung wird im Horizon 2020 Rahmenprogramm für Forschung und Innovation (Projekt Nr. 802450) durch die Europäische Union gefördert.

⁹ Im Rahmen dieser Forschung beleuchtete der DIW Wochenbericht 19/2019 bereits den möglichen Beitrag datenbasierter Vorhersagen für Antibiotikaverschreibungen. Dort wurde die Wichtigkeit, diagnostische Informationen der ÄrztInnen in Entscheidungen mit einzubeziehen, dargelegt, vgl. Michael A. Ribers und Hannes Ullrich (2019b): Künstliche Intelligenz und Daten können bei der Eindämmung von Antibiotikaresistenzen helfen. DIW Wochenbericht Nr. 19, 335–341 (online verfügbar). Im aktuellen Beitrag liegt der Fokus auf der Quantifizierung des Datenbedarfs, um Verbesserungen zu erreichen.

¹⁰ Vgl. Jim O'Neill (2016): Tackling drug-resistant infections globally: final report and recommendations. Review on Antimicrobial Resistance. HM Government and Wellcome Trust: UK.

¹¹ Vgl. Esmi Charani, Jonathan Cooke und Alison Holmes (2010): Antibiotic Stewardship Programmes – What's Missing? *Journal of Antimicrobial Chemotherapy*, 65(11), 2275–2277; Lina Maria Ellegård, Jens Dietrichson und Anders Anell (2018): Can Pay-for-Performance to Primary Care Providers Stimulate Appropriate Use of Antibiotics? *Health Economics*, 27(1), e39–e54; Janet Currie, Wanchuan Lin und Juanjuan Meng (2014): Addressing Antibiotic Abuse in China: An Experimental Audit Study. *Journal of Development Economics*, 110, 39–51.

Kasten 2

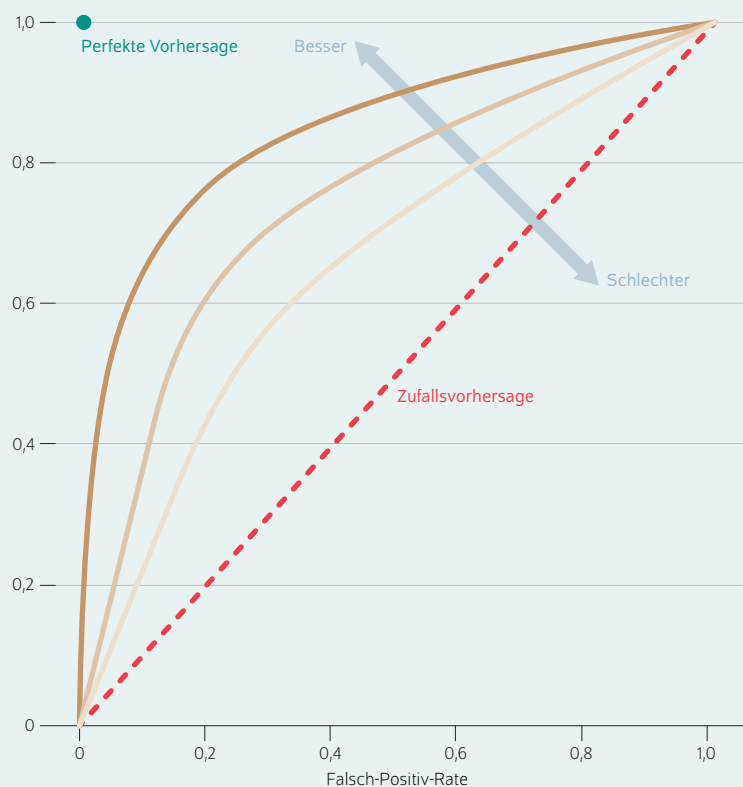
Die ROC-Kurve als Gütekriterium für binäre Klassifikationsprobleme

Die Receiver Operating Characteristic Kurve (ROC-Kurve) vermittelt einen Überblick über die Vorhersagequalität eines Klassifizierungsalgorithmus, indem sie die Richtig-Positiv-Rate gegen die Falsch-Positiv-Rate abträgt (Abbildung). Die Richtig-Positiv-Rate gibt an, mit welcher Wahrscheinlichkeit eine tatsächlich bakterielle Testprobe korrekt als solche, also positiv, klassifiziert wird. Die Falsch-Positiv-Rate gibt an, mit welcher Wahrscheinlichkeit eine nicht-bakterielle, also negative, Testprobe fälschlicherweise als bakteriell klassifiziert wird.

Wenn Beobachtungen mit einem Münzwurf als positiv oder negativ gekennzeichnet werden, spricht man von einer Zufallsklassifikation. Dabei wird für jede Richtig-Positiv-Rate dieselbe Falsch-Positiv-Rate erreicht. Das heißt, für jede Anzahl an korrekt positiv klassifizierten Beobachtungen werden genauso viele Beobachtungen fälschlicherweise als positiv klassifiziert, was durch die Diagonale im ROC-Raum abgebildet wird. Ein Klassifizierungsalgorithmus ist umso besser, je weiter oberhalb seine ROC-Kurve über der einer Zufallsklassifikation liegt: Dann hat der Algorithmus eine höhere Richtig-Positiv-Rate für eine gegebene Falsch-Positiv-Rate. Der bestmögliche Klassifizierungsalgorithmus ist immer korrekt, das heißt seine Richtig-Positiv-Rate beträgt eins und seine Falsch-Positiv-Rate null. Die Fläche unter der ROC-Kurve fasst die zweidimensionale ROC-Kurve in einem Wert zusammen. Die Zufallsklassifikation hat eine Fläche von 0,5 unter der ROC-Kurve. Je größer der Wert ist, desto besser ist die Klassifizierung. Die maximal mögliche Fläche unter der ROC-Kurve beträgt eins.

Abbildung

Beispielhafte Illustration: ROC-Kurve als Maß für Vorhersagequalität
Korrekt-Positiv-Rate



Anmerkungen: Beispielhafte Illustration. Die Korrekt-Positiv-Rate (Falsch-Positiv-Rate) bezeichnet den Anteil der korrekt (fälschlicherweise) positiv klassifizierten Fälle. Je weiter links oben die ROC-Kurve verläuft, desto besser ist die Vorhersage. Die Diagonale ist die ROC-Kurve einer reinen Zufallsklassifikation (Münzwurf).

Quellen: Eigene Darstellung.

© DIW Berlin 2021

Boosting).¹² Hierbei werden die Vorhersagen vieler einzelner Methoden kombiniert. In der vorliegenden Studie werden 95 594 akute Behandlungssituationen berücksichtigt, denen keine Antibiotikabehandlung oder Diagnostik unmittelbar vorausgehen. Aus diesen werden für 42 480 Behandlungssituationen anhand vorausgehender Daten Laborergebnisse vorhergesagt und im Anschluss ihrem tatsächlichen Laborergebnis gegenübergestellt.

Um den Zugewinn der Vorhersagequalität durch das sukzessive Verknüpfen zusätzlicher Personendaten zu messen, werden die verfügbaren Daten in fünf Teilsegmente

aufgeteilt. Diese Segmente werden nacheinander dem Algorithmus zur Verfügung gestellt. Dadurch kann evaluiert werden, welche Informationen die datenbasierte Vorhersage merklich verbessern.

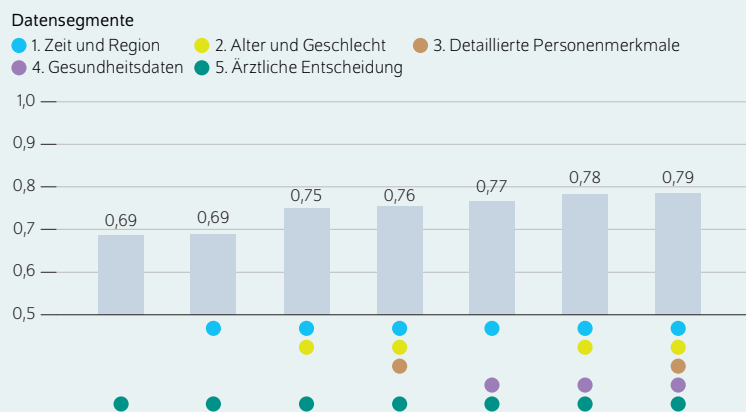
Das erste Datensegment enthält die Zeit und Gemeinde der Behandlungssituation, um die zeitlichen Schwankungen und Variation der Testauswahl verschiedener Regionen zu dokumentieren. Das zweite Datensegment enthält als einfache persönliche Eigenschaften Alter und Geschlecht der behandelten Person. Alter und Geschlecht sind zwar auch für ÄrztInnen leicht zu erheben. Dennoch können diese Informationen Vorhersagen stark verbessern, da ein Algorithmus beispielsweise Altersgruppen differenzierter abbilden kann. Des Weiteren werden die Korrelationen anhand

¹² Vgl. Trevor Hastie, Robert Tibshirani und Jerome Friedman (2009): The elements of statistical learning: data mining, inference, and prediction. Springer Series in Statistics, New York.

Abbildung 1

Vorhersagequalität unter Nutzung verschiedener Datenkombinationen

Qualitätsmaß Fläche unter der ROC-Kurve



Anmerkungen: Die Fläche unter der ROC-Kurve ist ein Qualitätsmaß für Vorhersagen im Vergleich zum tatsächlichen Laborbefund. Sie nimmt Werte zwischen 0,5 und eins an, wobei ein höherer Wert für eine bessere Vorhersagequalität steht.

Quellen: Dänische Administrativdaten, eigene Berechnungen.

© DIW Berlin 2021

Die Vorhersagequalität wird durch einfache demografische Merkmale und Gesundheitsdaten erheblich verbessert.

Demografische Daten und Gesundheitsdaten erhöhen Vorhersagequalität

Zur Beurteilung der Genauigkeit der datenbasierten Vorhersagen einer bakteriellen Infektionsursache wird die sogenannte *Fläche unter der ROC-Kurve* betrachtet. Diese Zahl stellt ein etabliertes Qualitätsmaß für binäre Vorhersagen im Vergleich zum tatsächlichen Laborbefund dar (Kasten 2). Sie nimmt Werte zwischen 0,5 und eins an, wobei ein höherer Wert für eine bessere Vorhersagequalität steht. Ein Wert von eins wird nur dann erreicht, wenn die Vorhersagen perfekt sind, das heißt der Algorithmus die Wahrheit immer richtig vorhersagt.

Wird vom Algorithmus nur die Verschreibungsentscheidung zum Zeitpunkt der Probenentnahme für Vorhersagen verwendet (Datensegment 5), kann bereits eine Fläche unter der ROC-Kurve von 0,69 erreicht werden (Abbildung 1). Die zusätzlichen Informationen Zeit und Region der Probenentnahme (Datensegment 1) verändern diesen Wert nicht merklich. Fügt man einfache Personenmerkmale wie das Alter und Geschlecht (Segment 2) hinzu, wächst die Fläche unter der ROC-Kurve auf 0,75 an. Zusätzliche, detaillierte persönliche Eigenschaften (Segment 3) erhöhen das Qualitätsmaß nur geringfügig auf 0,76.

Verwendet man nur Gesundheitsdaten gemeinsam mit Daten zu Verschreibungsentscheidung, Zeit und Ort (Segmente 1, 4 und 5), beläuft sich das Qualitätsmaß auf 0,77. Ein geringfügig höherer Wert von 0,78 kann durch Hinzufügen einfacher Personenmerkmale (Segmente 1, 2, 4 und 5) erreicht werden. Der Algorithmus kann so zusätzliche Interaktionen zwischen Alter, Geschlecht und Gesundheitsinformationen für die Vorhersage nutzen. Mit allen verfügbaren Datensegmenten, insgesamt 1266 Variablen, steigt das Qualitätsmaß auf 0,79 an.¹⁴

Erhebliches Verbesserungspotenzial für Behandlungsentscheidungen durch verknüpfte Daten

Für die Evaluierung des Verbesserungspotentials datenbasierter Vorhersagen reicht es nicht aus, ausschließlich die Güte der Vorhersagen zu messen. Darüber hinaus ist ein Vergleich zur bestehenden Verschreibungspraxis notwendig.

Um eine Evaluierung zu ermöglichen, verwendet die Studie ein Gedankenexperiment. Hierbei wird angenommen, dass die Verwendung eines Antibiotikums einen direkten positiven Heilungsnutzen für eine behandelte Person mit bakterieller Infektion hat. Dieser Heilungsnutzen ist bei Personen ohne bakterielle Infektion nicht vorhanden. Darüber hinaus geht jede Antibiotikaverwendung mit indirekten gesellschaftlichen Kosten in Form steigender Resistenzen einher. Da die Gewichtung dieser Nutzen und Kosten unbekannt ist,

von Beobachtungen aller ÄrztInnen geschätzt, während einzelne ÄrztInnen nur ihre eigenen PatientInnen beobachten können.

Das dritte Segment beinhaltet detailliertere persönliche Informationen wie beispielsweise Ausbildung, Beruf, Einkommen, Haushaltzusammensetzung oder Migrationshintergrund der behandelten Person. Die Verknüpfung dieser Daten mit Gesundheitsdaten ist weniger üblich und mit höheren Kosten hinsichtlich der Privatsphäre verbunden. Das vierte Datensegment enthält detaillierte Gesundheitsdaten der behandelten Person wie die Diagnose- und Behandlungshistorie in allgemeinärztlichen Praxen und Krankenhäusern.

Das fünfte Segment beinhaltet die ärztliche Verschreibungsentscheidung zum Zeitpunkt der Probenentnahme. Dies erscheint zunächst paradox, da das Ziel ist, diese Entscheidungen anhand der datenbasierten Vorhersagen zu verbessern. Gibt man aber die Verschreibungsentscheidung an den Algorithmus, der zum Ziel hat, den Vorhersagefehler zu minimieren, werden weitere in ärztlichen Entscheidungen enthaltene Informationen herausgefiltert. Der Informationsgehalt ist oft so groß, dass Vorhersagen, die auf diese Information verzichten, nicht besser abschneiden als die ÄrztInnen selbst.¹³

¹³ Vgl. Ribers und Ullrich (2019a), a. a. O.

¹⁴ Ähnliche Werte für die Fläche unter der ROC-Kurve werden auch in anderen Vorhersagemodellen für politische Entscheidungsprobleme erreicht. Vgl. zum Beispiel Jon Kleinberg et al. (2018): Human decisions and machine predictions. Quarterly Journal of Economics, 133(1), 237–293.

ist eine Bewertung von Verschreibungsänderungen in vielen Fällen nicht direkt möglich. Im vorliegenden Gedankenexperiment wird von einer Verbesserung gesprochen, wenn bei sinkender Verschreibungszahl sichergestellt wird, dass mindestens genauso viele bakteriell Infizierte erkannt und behandelt werden wie unter rein ärztlicher Entscheidung.

Werden nur die ärztlichen Verschreibungsentscheidungen für die algorithmischen Vorhersagen (Segment 5) verwendet, bleibt die Verschreibungsanzahl nahezu unverändert im Vergleich zur ärztlichen Entscheidung (Abbildung 2). Auch die Hinzunahme von Zeit und Region der Probenentnahme führt nur zu einem geringem Gesamtrückgang von 1,2 Prozent. Sind darüber hinaus Daten zu Alter und Geschlecht verfügbar, können Verschreibungen um insgesamt 5,2 Prozent reduziert werden. Zusätzliche detaillierte persönliche Eigenschaften führen zu einer nur geringfügig stärkeren Reduktion von insgesamt 5,4 Prozent im Vergleich zur ärztlichen Entscheidung.

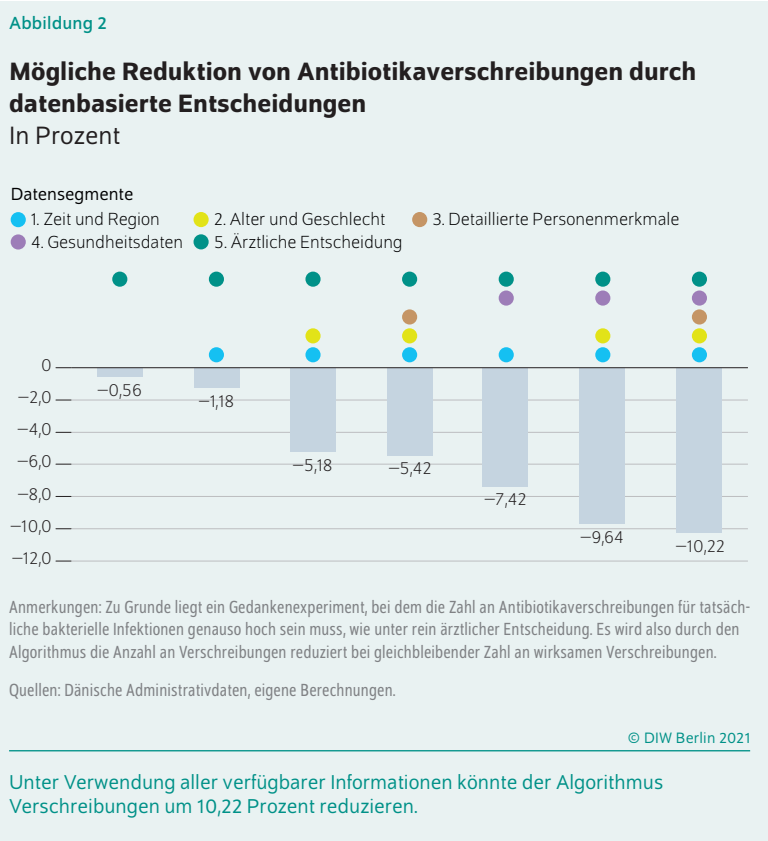
Sind detaillierte Gesundheitsdaten in Verbindung mit Datum und Praxis der Probenentnahme verfügbar, könnten Verschreibungen bereits um 7,4 Prozent sinken. Zu einer deutlich stärkeren Senkung von insgesamt 9,6 Prozent führt das weitere Hinzufügen von Alter und Geschlecht, zusätzlich zu den Gesundheitsdaten.

Mit allen verfügbaren Daten ist eine Reduktion von 10,2 Prozent im Vergleich zu den von Ärzten getroffenen Entscheidungen erreichbar. Wären alle Vorhersagen perfekt, könnten 39 Prozent weniger Antibiotika verschrieben werden. Somit könnte unter Verwendung aller Daten gut ein Viertel dieser maximal möglichen Reduktion erreicht werden. Die Anzahl der Verschreibungen für tatsächliche bakterielle Infektionen bleibt per Konstruktion des Gedankenexperiments gleich, sodass die Reduktion von Antibiotika allein bei den Fehlverschreibungen erfolgt.

Der direkte Vergleich zwischen der Vorhersage und der Arztentscheidung zeigt, dass ein reiner Fokus auf die Vorhersagequalität (Fläche unter der ROC-Kurve) nicht ausreicht. Allein das Hinzufügen von Alter und Geschlecht oder der Gesundheitsdaten genügt, um nahezu die Vorhersagequalität zu erreichen, die unter Verwendung aller Daten möglich ist. Bei der Gegenüberstellung von datenbasierter Verschreibungsregel und tatsächlicher Arztentscheidung führen jedoch auch die Kombination demografischer Daten mit Gesundheitsdaten noch zu deutlicheren Verbesserungen. Für die Evaluierung derartiger Anwendungsbeispiele ist also der Vergleich der tatsächlichen Entscheidungen mit den Vorhersagen zentral.

Fazit: Verknüpfte Daten können erheblichen Mehrwert generieren, aber Evaluation der Anwendungen nötig

Die vorliegende Studie zeigt, dass anhand verknüpfter Administrativdaten und maschinellem Lernen wertvolle Vorhersagen über Laborergebnisse generiert werden können. Diese



haben das Potenzial, ärztliche Verschreibungen vor dem Hintergrund des Problems zunehmender antibiotikaresistenter Erreger zu verbessern. Die Verknüpfung von Daten kann darüber hinaus für ganz unterschiedliche Anwendungen sinnvoll sein. Ein sicherer Zugriff auf pseudonymisierte, verknüpfte Administrativdaten für Forschungszwecke ist Grundvoraussetzung, um genau solche Anwendungsgebiete zu identifizieren.

Für das konkrete Problem der Antibiotikaverschreibungen konnte mittels hochwertiger verknüpfter Administrativdaten gezeigt werden, dass zwar bereits einfache Informationen für eine hohe Vorhersagequalität ausreichen, detailliertere Daten aber zu deutlicheren Verbesserungen von Entscheidungen führen könnten. Welche Anforderungen bei anderen Beispielen von Daten, Methoden und Umsetzung erfüllt sein müssen, um ausreichend großen Nutzen zu stiften, muss mittels problemspezifischer, wissenschaftlich begleiteter Auswertungen untersucht werden. Ein solcher Ansatz steht im Kontrast zu den negativen Beispielen großer, teils gescheiterter Digitalprojekte aus der Privatwirtschaft.¹⁵ Die Diskussion und Abwägung ethischer Fragen stellt einen weiteren wichtigen Teil der Evaluierung von Anwendungsbeispielen dar.

¹⁵ IBM's Watson Health beispielsweise versprach visionäre Anwendungen in der Medizin, war mangels Evidenz in der Praxis jedoch nur eingeschränkt nutzbar und lieferte wenig Mehrwert, vgl. Casey Ross und Mario Aguilar: Inside the fall of Watson Health: How IBM's audacious plan to 'change the face of health care' with AI fell apart (online verfügbar; abgerufen am 8. März 2021).

Vor dem Hintergrund der SARS-CoV-2 Pandemie haben sich Bemühungen für eine stärkere Digitalisierung im Gesundheitsbereich in Deutschland beschleunigt. Nicht zuletzt hat die Pandemie die Relevanz hochwertiger Daten in das Bewusstsein der breiten Öffentlichkeit, Verwaltung und Politik gerufen. Diese Dynamik könnte Anwendungen im Gesundheitswesen und darüber hinaus beflügeln, um besser informiert gesellschaftliche Herausforderungen zu meistern. Das neu gegründete Netzwerk Universitätsmedizin kann hierfür, unter Einbindung quantitativ arbeitender Wirtschafts- und Sozialwissenschaften, eine Plattform werden.¹⁶ Ein weiterer Baustein könnte auch die gerade eingeführte elektronische Patientenakte sein. Dabei ist noch unklar, ob

in der aktuellen Ausgestaltung Daten für Forschungszwecke verfügbar gemacht werden können.

Für Anwendungen wie in dieser Studie bedarf es einer Infrastruktur zur Datenverknüpfung und -bereitstellung, die gesellschaftlich akzeptierte Datenschutz- und Ethikstandards einhält. Stünden solche Dienste der Wissenschaft und dem Gesundheitssektor zur Verfügung, könnten sich zahlreiche Möglichkeiten eröffnen, die gesundheitliche Versorgung deutlich zu verbessern. Perspektivisch könnte das Personal im Gesundheitsbereich hierdurch entlastet und die zwischenmenschliche Komponente wieder stärker in den Vordergrund gerückt werden.

16 Das Netzwerk Universitätsmedizin wurde während der SARS-CoV-2 Pandemie, unter anderem zur Zusammenführung von Forschungsdaten gegründet, vgl. Webseite (online verfügbar).

Shan Huang ist Doktorandin in der Abteilung Unternehmen und Märkte am DIW Berlin | shuang@diw.de

Michael Allan Ribers ist Wissenschaftlicher Mitarbeiter in der Abteilung Unternehmen und Märkte am DIW Berlin | mribers@diw.de

Hannes Ullrich ist Wissenschaftlicher Mitarbeiter in der Abteilung Unternehmen und Märkte am DIW Berlin | hullrich@diw.de

JEL: C10; C55; I11; I18; L38; O38; Q28

Keywords: antibiotic prescribing; prediction policy; administrative data; data combination

IMPRESSUM



DIW Berlin — Deutsches Institut für Wirtschaftsforschung e.V.

Mohrenstraße 58, 10117 Berlin

www.diw.de

Telefon: +49 30 897 89-0 Fax: -200

88. Jahrgang 31. März 2021

Herausgeberinnen und Herausgeber

Prof. Dr. Tomaso Duso; Prof. Marcel Fratzscher, Ph.D.; Prof. Dr. Peter Haan;
Prof. Dr. Claudia Kemfert; Prof. Dr. Alexander S. Kritikos; Prof. Dr. Alexander
Kriwoluzky; Prof. Dr. Stefan Liebig; Prof. Dr. Lukas Menkhoff; Dr. Claus
Michelsen; Prof. Karsten Neuhoff, Ph.D.; Prof. Dr. Carsten Schröder;
Prof. Dr. C. Katharina Spieß; Dr. Katharina Wrohlich

Chefredaktion

Dr. Gritje Hartmann; Dr. Anna Hammerschmid (Stellv.)

Lektorat

Mara Barschkett

Redaktion

Marten Brehmer; Rebecca Buhner; Claudia Cohnen-Beck; Kristina van
Deuverden; Petra Jasper; Sebastian Kollmann; Sandra Tubik

Vertrieb

DIW Berlin Leserservice, Postfach 74, 77649 Offenburg

leserservice@diw.de

Telefon: +49 1806 14 00 50 25 (20 Cent pro Anruf)

Gestaltung

Roman Wilhelm, DIW Berlin

Umschlagmotiv

© imageBROKER / Steffen Diemer

Satz

Satz-Rechen-Zentrum Hartmann + Heenemann GmbH & Co. KG, Berlin

Druck

USE gGmbH, Berlin

ISSN 0012-1304; ISSN 1860-8787 (online)

Nachdruck und sonstige Verbreitung – auch auszugsweise – nur mit
Quellenangabe und unter Zusendung eines Belegexemplars an den
Kundenservice des DIW Berlin zulässig (kundenservice@diw.de).

Abonnieren Sie auch unseren DIW- und/oder Wochenbericht-Newsletter
unter www.diw.de/newsletter