

**Erfolgreiche Registerforschung
in Österreich**

Welchen Mehrwert generiert
die reglementierte Öffnung
von Registerdaten für die
wissenschaftliche Forschung?

Eine Darstellung anhand
von drei Beispielen

Lukas Schmoigl
Thomas König

Erfolgreiche Registerforschung in Österreich

Welchen Mehrwert generiert die reglementierte
Öffnung von Registerdaten für die
wissenschaftliche Forschung?

Eine Darstellung anhand von drei Beispielen

Lukas Schmoigl, Thomas König

WIFO Working Papers 630/2021
Juni 2021

Inhalt

Der Nutzen eines strukturierten Zugangs zu Registern der öffentlichen Verwaltung für Gesellschaft und Politik ist weitreichend, geht jedoch oft in der öffentlichen Diskussion unter. Der Arbeitsauftrag für den vorliegenden Bericht bestand also darin, eine Potentialanalyse durchzuführen, d. h., drei konkrete Beispiele auszuformulieren, in denen wissenschaftlicher und gesellschaftlicher Mehrwert von der Forschung zugänglichen Registern dargelegt und veranschaulicht werden kann. Im Zentrum stehen dabei konkrete gesellschaftliche Probleme, die mit der Forschung adressiert wurden, sowie tatsächliche und mögliche Impacts der Ergebnisse der Forschung auf Politik, Gesellschaft und den Wissenschaftsstandort Österreich. Dies soll illustrieren, warum der Zugang zu Registerdaten für die Wissenschaft von großer Bedeutung ist, wo Wissenslücken bestehen und was mittels des Zugangs zu Registerdaten noch erforscht werden könnte.

E-Mail: lukas.schmoigl@wifo.ac.at

2021//W/0

© 2021 Österreichisches Institut für Wirtschaftsforschung.

Medieninhaber (Verleger), Hersteller: Österreichisches Institut für Wirtschaftsforschung
1030 Wien, Arsenal, Objekt 20 | Tel. (43 1) 798 26 01-0 | <https://www.wifo.ac.at>
Verlags- und Herstellungsort: Wien

Die WIFO Working Papers beruhen nicht notwendigerweise auf einer abgestimmten Position des WIFO. Die Autorinnen und Autoren wurden über die Richtlinien der Österreichischen Agentur für wissenschaftliche Integrität (ÖAWI) zur Guten Wissenschaftlichen Praxis informiert, insbesondere bezüglich der Dokumentation aller Elemente, die für eine Nachvollziehbarkeit der Ergebnisse notwendig sind.

Kostenloser Download: <https://www.wifo.ac.at/wwa/pubid/67191>

Einleitung

Ein wesentliches Merkmal der öffentlichen Verwaltung ist es, dass sie zur Erfüllung ihrer Aufgaben Informationen über ihre BürgerInnen sammelt – egal, ob es sich dabei um Meldedaten, Einkommen, Krankheitsverteilung oder Mobilitätsverhalten handelt. Es kann sich um sehr sensible Individualdaten, wie personenbezogene Merkmale, oder aber auch um unproblematische Informationen handeln, die keine Rückschlüsse auf einzelne Personen oder Unternehmen zulassen. Der primäre Zweck der Datenerhebung ist von Fall zu Fall verschieden, und auch seine Rechtfertigung kann sehr verschieden gestaltet sein – die Varianten reichen von routinemäßigen (und auch gesetzlich vorgeschriebenen) Erfassungen (wie etwa Arbeitsmarktdaten) bis hin zu Listen, die anlassbezogen erstellt werden (wie etwa die im Zuge der COVID-19 Pandemie erstellten Datensätze¹).

Die Digitalisierung, die in den letzten zwei Jahrzehnten massive Veränderungen in allen Aspekten des sozialen Zusammenlebens herbeigeführt hat, hat der öffentlichen Verwaltung einerseits neue Möglichkeiten in Bezug auf Datenerhebungen gegeben. Sogenannte Register und Verwaltungsdaten befinden sich zunehmend in digitalisierter Form. Dadurch ist es auch leichter möglich, diese Daten verfügbar zu machen beziehungsweise auch zu verknüpfen. Eine Folge davon ist eine erhöhte Aufmerksamkeit vonseiten der Zivilgesellschaft bezüglich des potenziellen Missbrauchs solcher (insbesondere personenbezogener) Daten vonseiten staatlicher Akteure, aber auch bezüglich der Sicherheit der Datenbestände gegenüber illegalem Zugriff („hacken“). Die verantwortlichen staatlichen Stellen tun gut daran, solche Sorgen ernst zu nehmen und sowohl in der Verwendung hohe ethische Maßstäbe zu entwickeln, als auch die technische Sicherheit der Datenbestände auf höchstem Niveau zu garantieren.

Eine ganz andere Folge von digitalisierten Verwaltungsdaten ist aber auch, dass sie für wissenschaftliche (Zweit-)Verwendung aufbereitet werden können. Es handelt sich hierbei also um eine von staatlicher Seite bewusst erlaubte Weiterverwendung, die freilich denselben ethischen Maßstäben unterliegen muss und für die es klare (juristische) Regeln und Restriktionen und technische Rahmenbedingungen gibt, damit kein Missbrauch erfolgen kann. Wenn das – etwa im Rahmen gesetzlicher Vorgaben, die ihrerseits durch den demokratischen Entscheidungsprozess legitimiert sind – sichergestellt ist, so stellen diese Verwaltungsdatenbestände (hier im Weiteren als Registerdaten bezeichnet) eine bedeutende Ressource für die wissenschaftliche Analyse dar – und zwar sowohl im Bereich der (akademischen) Grundlagenforschung als auch in der angewandten Forschung und Auftragsforschung.

In Österreich ist die Debatte zur wissenschaftlichen Weiterverwendung von Registerdaten in den letzten Jahren zunehmend intensiv geführt worden. Zwei Gründe sind dafür ausschlaggebend: erstens wird vonseiten der Forschungscommunity darauf hingewiesen, dass viele andere Länder in Europa und in der Welt den Zugang zu Verwaltungsdatensätzen bereits mehr oder weniger klar geregelt haben, was sich u. a. darin ausgewirkt hat, dass dieser Zugang zu neuen Forschungsfragen und spannenden Ergebnissen geführt hat. Ausdruck des Diskurses in Österreich ist der informelle Zusammenschluss der *Plattform Registerforschung*, in welcher sich mehr als 150 mit der Plattform assoziierte Personen – meist WissenschaftlerInnen – zusammengetan haben, um den Austausch zu diesem Thema zu fördern.²

¹ <https://www.data.gv.at/covid-19/> (diese und alle weiteren Websites wurden letztmalig abgerufen am 27.10.2020)

² <http://registerforschung.at/>

Zweitens hat die europäische Datenschutz-Grundverordnung von 2018 es notwendig gemacht, den Bereich der wissenschaftlichen Weiterverwendung von Registerdaten grundsätzlich zu regeln, was durch Novellierung des Forschungsorganisationsgesetzes (FOG) auch passiert ist. Mit der Novelle wurde ein bedingter Rechtsanspruch wissenschaftlicher Einrichtungen auf Bereitstellung von Daten aus bundesgesetzlich vorgesehenen Registern geschaffen. Das FOG enthält Sonderdatenschutzrecht für den Wissenschafts- und Forschungsbereich und schafft in der Umsetzung der DSGVO Rahmenbedingungen für die Verarbeitung von personenbezogenen Daten u. a. zu Forschungszwecken. Den Regelungen zur Registerforschung zufolge sind wissenschaftliche Einrichtungen bei Einhaltung angemessener Garantien berechtigt, von Verantwortlichen, die bundesgesetzlich vorgesehene Register führen, die Bereitstellung von (personenbezogenen und nicht personenbezogenen) Daten aus diesen Registern zu verlangen – sofern diese Register auch in einer eigenen Verordnung angeführt sind.

Eine solche Verordnung ist noch nicht erlassen, und einer der Gründe dafür ist, dass im Augenblick noch ein genauerer Überblick fehlt, welche Datenbestände überhaupt als Register im Sinne des FOG gelten; um wie viele Register es sich handelt; und auf welcher Rechtsgrundlage diese Register beruhen. Klar ist hingegen, dass die Regelung im FOG eine große Chance für die österreichische Forschungspolitik darstellt, durch den geregelten Zugang zu Verwaltungsdatensätzen der Forschung endlich ein anderen Ländern vergleichbares Datenangebot zur Verfügung zu stellen. Aber nicht nur für die wissenschaftliche Forschung ist der prospektive Zugang relevant. Registerdaten sind auch eine wesentliche Voraussetzung, um wichtige Fragen der Verwaltung und der Politik durch methodisch rigorose Verfahren, welche die wissenschaftliche Forschung – insbesondere im Bereich der Sozial- und Wirtschaftswissenschaften, aber auch in der Gesundheitswissenschaft – entwickelt hat, beantworten zu können. Es handelt sich hierbei um Fragen der Effizienz von staatlichen Regularien, Wirksamkeit von politischen Maßnahmen in unterschiedlichen Politikfeldern, die von Gesundheit bis Arbeitsmarkt reichen.

Der Nutzen eines strukturierten Zugangs zu Registern der öffentlichen Verwaltung für Wissenschaft, Gesellschaft und Politik ist weitreichend, geht jedoch oft in der öffentlichen Diskussion unter. Der Arbeitsauftrag für den vorliegenden Bericht bestand also darin, eine Potenzialanalyse durchzuführen: das heißt, drei konkrete Beispiele auszuformulieren, in denen wissenschaftlicher und gesellschaftlicher Mehrwert von der Forschung zugänglichen Registern dargelegt und veranschaulicht werden kann. Diese drei Beispiele bilden daher auch den Gegenstand der folgenden Seiten. Im Zentrum stehen dabei konkrete gesellschaftliche Probleme, die mit der Forschung adressiert wurden, sowie tatsächliche und mögliche Impacts der Ergebnisse der Forschung auf Politik, Gesellschaft und den Wissenschaftsstandort Österreich. Dies soll illustrieren, warum der Zugang zu Registerdaten für die Wissenschaft von großer Bedeutung ist, wo Wissenslücken bestehen und was mittels des Zugangs zu Registerdaten noch erforscht werden könnte.

Die Ausarbeitung der drei Beispiele wurde an den beiden Instituten WIFO und IHS durchgeführt, ihrerseits maßgebliche institutionelle Akteure in der oben bereits genannten *Plattform Registerforschung*. Als Forschungsinstitute an der Schnittstelle zwischen angewandter empirischer und Grundlagenforschung sind beide Einrichtungen ganz besonders auf die Verwendung von Registerdaten angewiesen, nicht zuletzt, weil deren Verwendung auch von den auftraggebenden Einrichtungen (Ministerien, staatliche Behörden etc.) verstärkt gefordert wird, und weil es dem wissenschaftlichen Selbstverständnis der ForscherInnen entspricht, ihre Analysen mit Daten durchzuführen, die bestmöglich die gesellschaftliche Realität abbilden.

Der Jargon der folgenden drei Beispiele ist bewusst so ausgerichtet, dass auch Personen außerhalb des Fachkreises die dargestellte Forschung verstehen können und auf technische sowie wissenschaftliche Fachbegriffe wird tendenziell verzichtet. Das Zielpublikum variiert dabei von der interessierten Öffentlichkeit, PolitikerInnen, JournalistInnen, aber auch Leitungen von Forschungsstätten, die ein Interesse an der Verfügbarkeit von Registerdaten haben.

Das Beispiel „Mit Nudging zu weniger Arbeitslosigkeit“ wurde kurz nach der Ausarbeitung für den vorliegenden Bericht unter Zustimmung des Auftraggebers bereits auf der Website der *Plattform Registerforschung* veröffentlicht.³ Dort befinden sich auch noch andere veröffentlichte Beispiele der Plattform, die bereits vor Auftragserteilung dieses Projekts entstanden sind (diese sind kurz im Anhang des Berichts aufgelistet). Das Beispiel „Wer zitiert werden will, braucht Registerdaten“ wurde unter Verwendung von noch unveröffentlichten Ergebnissen einer Analyse freundlicherweise von Martin Halla zur Verfügung gestellt. Vielen Dank! Das Beispiel „Mehr Evidenz in einer Pandemie“ basiert auf Hinweisen von Thomas Czypionka, der uns auch sonst noch wichtige Hinweise gegeben hat.

³ <https://www.registerforschung.at/erfolgsbeispiele>

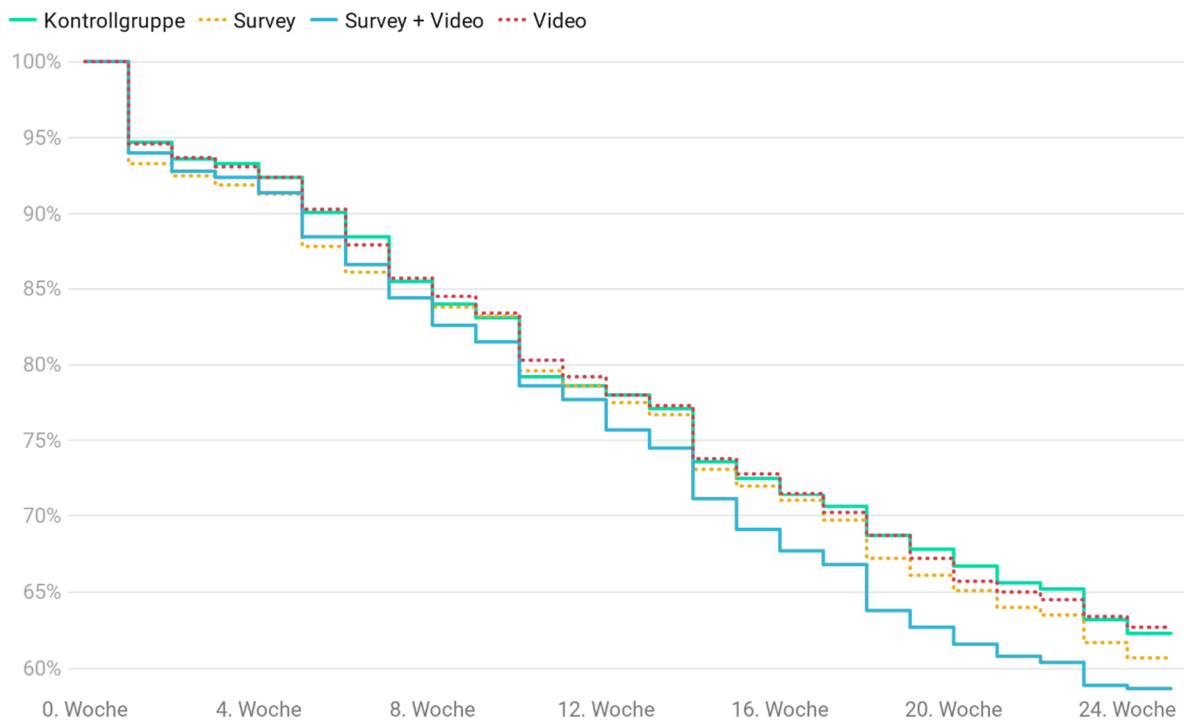
Beispiel 1: Mit Nudging zu weniger Arbeitslosigkeit?

Ein innovatives Experiment zeigt, ob die bloße Bereitstellung von Informationen jungen Erwachsenen bei der Arbeitssuche helfen kann. Ein Forschungsprojekt mit der potenziellen Implikation, mit wenig Kostenaufwand Arbeitslosigkeit verkürzen zu können – getrieben von Registerdaten.

Gerade für junge Leute ist die Jobsuche oft schwierig. Gleichzeitig kann sich längere Arbeitslosigkeit zu Beginn der beruflichen Laufbahn negativ auf die gesamte spätere Karriere auswirken. Daher stehen junge Erwachsene im Zentrum arbeitsmarktpolitischer Maßnahmen. Im Fokus sind dabei meist berufliche Weiter-Qualifikationen. Doch auch kleine Interventionen können potenziell eine große Wirkung erzielen. Etwa sind Arbeitssuchende oft nicht gut informiert, welche Strategien bei der Arbeitssuche erfolgreich sind. So liefern beispielsweise persönliche Kontakte die vielversprechendsten Aussichten für eine gelungene Suche. Wenn man nun versuchen würde, Arbeitssuchende über diese und andere Strategien aufzuklären, würden sie dann schneller einen Job finden?

Abbildung 1: Verbleib in Arbeitslosigkeit für Personen mit niedriger formaler Bildung über die Wochen nach dem Nudging für Kontrollgruppe und verschiedene Treatments (Intention to Treat).

Lesebeispiel: Von allen Personen, die das Video und den Survey erhalten haben, waren nach 18 Wochen noch 63,8 Prozent ohne Arbeit, bei jenen die weder Video noch Survey erhalten haben, waren es 68,7 Prozent.



Quelle: Daten aus Mühlböck et al. (2020), eigene Darstellung

Genau diese Frage haben sich die AutorInnen in dem Paper „Information, Reflection, and Successful Job Search: A Nudging Experiment“ (Mühlböck et al., 2020) gestellt. Mittels eines Video-Clips und einer Online-Umfrage sollten arbeitssuchende junge Erwachsene über Möglichkeiten und erfolgreiche Strategien der Jobsuche informiert werden sowie über ihre eigenen Suchanstrengungen reflektieren. Diese kleinen und zwanglosen Anstöße („Nudges“) wurden per E-Mail

an zufällig ausgewählte Personen verschickt, die sich beim Arbeitsmarktservice (AMS) arbeitssuchend gemeldet hatten. Weitere arbeitssuchende Personen fungierten als Kontrollgruppe, um die Veränderung über die Zeit in den verschiedenen Gruppen zu beobachten und feststellen zu können, ob diese Nudges helfen, schneller einen neuen Job zu finden.

Personen mit niedriger formaler Bildung profitieren

Hier kommen die Registerdaten ins Spiel: Über den Sozialversicherungsstatus wurde für die folgenden 25 Wochen überprüft, ob sich die untersuchten Personen in Beschäftigung befinden oder nicht. Für Personen mit niedriger formaler Bildung (ISCED 0-2) – etwa ein Drittel der jungen Arbeitssuchenden – lassen sich dabei tatsächlich Effekte der Intervention festmachen, in der die Personen zunächst das Survey und danach das Video erhielten. So betrug am Ende des Beobachtungszeitraums die Arbeitslosigkeit in dieser Gruppe 58,7 Prozent, während sie in der Kontrollgruppe bei 62,3 Prozent lag. Für Personen mit höherem formalen Bildungsniveau lassen sich hingegen keine signifikanten Effekte erkennen. Dies legt nahe, dass besonders benachteiligte Personen von Informationen über erfolgreiche Suchstrategien profitieren können.

Ideen eine Chance geben

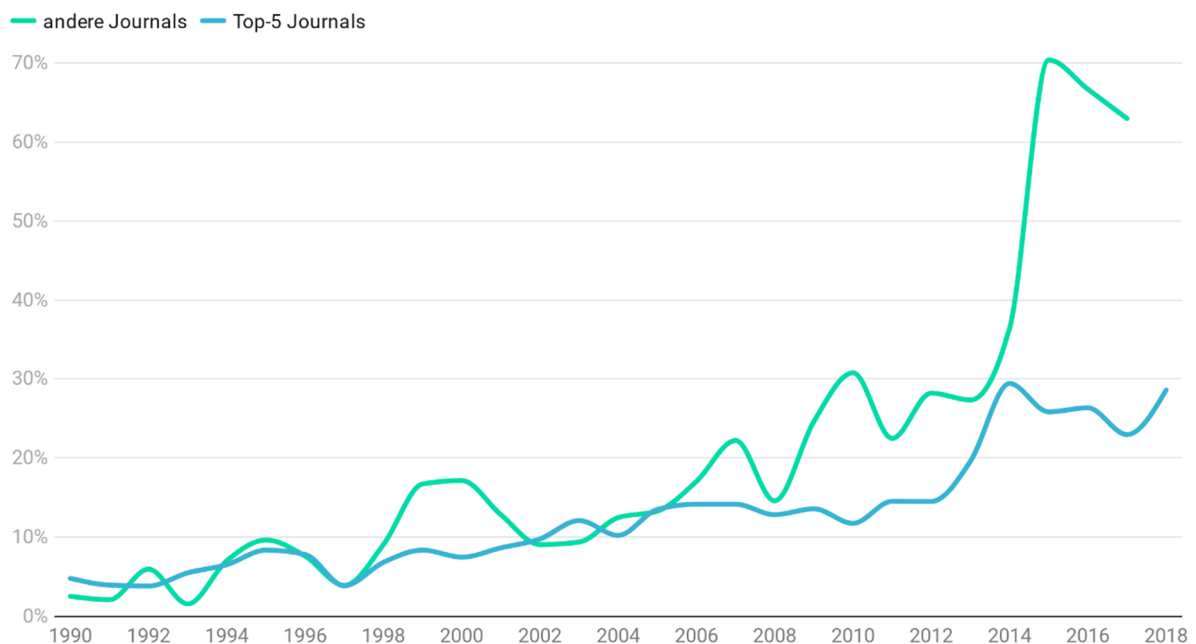
Auch wenn die Wirkung der Nudging-Intervention mit diesem ersten Experiment nicht vollständig gesichert ist, zeigt die Studie, wie dringend wir Forschung dieser Art benötigen. Ein Ziel von uns Forschern und Forscherinnen ist es, Wirkungsmechanismen zu entdecken, die genutzt werden können, um die Lebensbedingungen von Menschen zu verbessern. Neue Ideen, etwa ob kleine Stellschrauben etwas zum Besseren verändern, können dadurch getestet werden. Um zu überprüfen, ob diese Ideen effektiv sind, braucht es den Zugang zu einer Fülle von Daten.

Beispiel 2: Wer zitiert werden will, braucht Registerdaten

Administrative Daten stellen nicht nur eine wichtige Erkenntnisquelle für politische Entscheidungen dar, sondern sind für den Wissenschaftsstandort Österreich von entscheidender Bedeutung. Die Entwicklung der vergangenen Jahre zeigt: Wer sich im internationalen Forschungsumfeld bewähren will, muss zunehmend auf Registerdaten zurückgreifen.

Registerdaten haben viele bekannte Vorteile, die andere Datenquellen wie Surveys oder Experimente nicht aufweisen. Sie bilden oft fast vollständig die relevante Population ab und reduzieren oder eliminieren dadurch Stichprobenfehler. Bedingt durch die große Datenmenge, erlauben sie die Verknüpfung vieler spezifischer Merkmale und damit die Untersuchung speziellerer Forschungsfragen. Durch ihre systematische und oft vom beobachteten Individuum unabhängige Erhebung sind sie meist weniger fehlerbehaftet. Potenziell ermöglichen sie eine Beobachtung von Individuen im Längs- sowie Querschnitt. Nicht zuletzt können sie durch wenig Zusatzkosten bereitgestellt werden. Eine bislang unveröffentlichte Publikation von Halla et al. (2020) leuchtet ihren Wert innerhalb der Forschungscommunity aus. Durch eine Auswertung von Artikeln in international renommierten Journals zeigen sie, wie essenziell die Verwendung geworden ist, um publizieren zu können und zitiert zu werden.

Abbildung 2: Anteil von Artikeln mit Verwendung von Registerdaten an allen publizierten Artikeln in den jeweiligen Journalgruppen.



Quelle: Daten aus Halla et al. (2020), eigene Darstellung

Bedeutungsgewinn in internationalen Publikationen

Für die quantitative Erfassung haben Halla et al. (2020) Inhalt, Affiliation der AutorInnen und Zitationen von Publikationen in den fünf renommiertesten Journals⁴ und weiteren international bekannten Journals⁵ im Bereich der Ökonomie seit 1990 ausgewertet. Wie Abbildung 2 zeigt,

⁴ American Economic Review (AER), Econometrica (ECA), Journal of Political Economy (JPE), Quarterly Journal of Economics (QJE), Review of Economic Studies (RES)

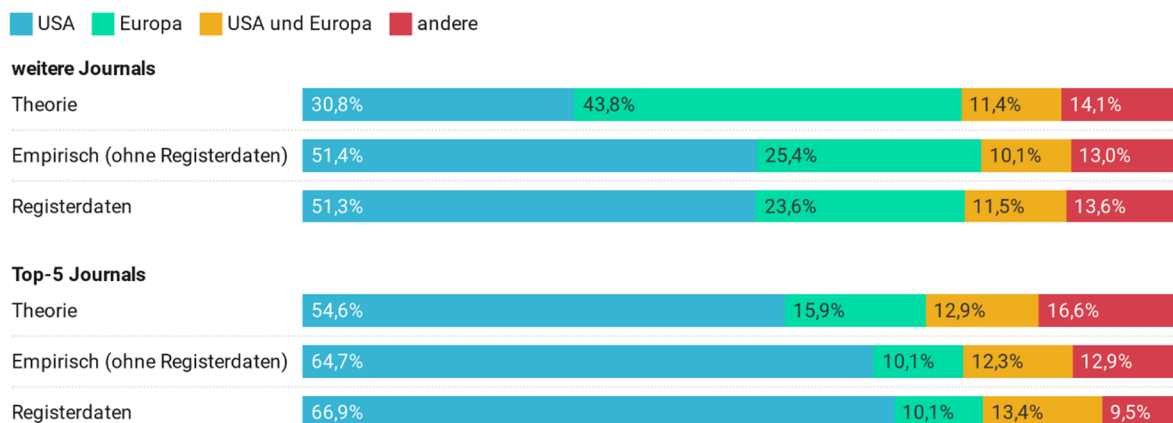
⁵ Applied Economics (AEJ-AE), the American Economic Journal: Economic Policy (AEJ-EP), Economic Journal (EJ), Journal of the European Economic Association (JEEA), Review of Economics and Statistics (ReStat).

hat die Registerforschung über die letzten drei Jahrzehnte einen gewaltigen Bedeutungsgewinn erfahren. Am Beginn der 90er Jahre lag der Anteil an publizierten Artikeln, die auf Registerforschung zurückgreifen, noch im niedrigen einstelligen Prozentbereich. Vor allem über das letzte Jahrzehnt sind diese Anteile stark gewachsen und erreichen nun jährliche Werte an die 30 Prozent (Top-5 Journals) beziehungsweise um 60 Prozent (andere Journals). Diese Erhöhung steht auch im Kontext einer grundsätzlich gewachsenen Anzahl an Literatur von empirischer Natur. Thematisch zeigt sich dabei ein speziellerer Fokus von publizierter Spitzenforschung mit administrativen Daten. Sozialversicherung und Wohlfahrtsstaat stellen die am häufigsten bearbeiteten Themenfelder dar – also jene Bereiche, die von hoher politischer und gesellschaftlicher Relevanz gekennzeichnet sind.

Europa unter Zugzwang

Die zunehmende Verwendung von Registerdaten scheint dabei die Dominanz der USA im internationalen Forschungsumfeld zu verfestigen. Wie Abbildung 3 zeigt, ist über Gruppen von Journals hinweg ein höherer Anteil von US-AutorInnen unter empirischen Artikeln mit oder ohne Registerbezug aufzufinden. Im tendenziell schrumpfenden Teil theoretischer Artikel nehmen dagegen europäische AutorInnen einen vergleichsweise größeren Teil ein. Innerhalb der Registerforschung in den Top-5 Journals ist dabei die Dominanz der USA am eklatantesten: 66,9 Prozent der AutorInnen kommen dort aus den Vereinigten Staaten. Hier steht Europa unter Zugzwang: Verschiebt sich der Schwerpunkt der Publikationen weiter in Richtung administrativer Daten, so ist davon auszugehen, dass sich auch der Anteil der ForscherInnen mit europäischer Affiliation in diesen Journals weiter verringern wird.

Abbildung 3: Geografische Verteilung der AutorInnenschaft nach Artikeltyp und Journaltyp



Quelle: Daten aus Halla et al. (2020), eigene Darstellung

Das gestiegene Interesse an Registerdaten schlägt sich nicht allein auf der Seite der Publikationen nieder. Denn auch zwischen dem Artikeltyp und der Anzahl der Zitationen scheint ein Zusammenhang gegeben. Tabelle 1 zeigt die Koeffizienten dreier statistischer Modelle, die Variablen der jeweiligen Publikationen auf die Anzahl der Publikationen regressieren. Im Mittel erzielen Publikationen unter Einsatz von Registerforschung durchgehend eine höhere Zahl an Referenzen gegenüber theoretischen und anderen empirischen Artikeln. Im Vergleich mit rein theoriebasierten Publikationen ist der Effekt in allen Modellen substantiell und statistisch abgesichert. So erreicht Registerforschung auch unter Berücksichtigung vieler anderer Einflussfaktoren im Schnitt noch immer über 22 Zitationen mehr als Publikationen anderen Typs.

Tabelle 1: Anzahl der Zitationen in Top-5 Journals

	(1)	(2)	(3)
Abhängige Variable: Anzahl an Zitationen			
Artikeltyp (Admin = Referenz)			
Theoretisch	-44,28***	-34,10***	-22,99***
Empirisch (non-admin)	-7,34**	-3,35	-0,24
Manuelle Identifikation Dummy	3,85	-10,61**	-12,06***
Journal (AER = Referenz)			
ECA		-11,65***	-17,75***
JPE		-7,91**	-17,23***
QJE		26,59***	10,28**
RES		-25,98***	-25,58***
Affiliation der AutorInnen (USA = Referenz)			
Europa			-7,37***
Europa und andere			-16,69***
andere			-19,06***
USA und Europa			9,40***
USA und andere			-4,25
USA, Europa und andere			-4,44
Anzahl der AutorInnen			6,31***
Anzahl der Seiten			0,85***
Anzahl der Referenzen			0,63***
Konstante	131,04***	125,95***	102,52***
FE Publikationsjahr	Ja	Ja	Ja
Durchschnittliche Zitationen	84	84	82
N	7.365	7.365	6.494
R ²	0,21	0,23	0,29

Quelle: Halla et al. (2020)

Anmerkungen: Die Anzahl der Zitationen ist die Frequenz, mit der ein Artikel im Web of Science von einer anderen Quelle zitiert wird. Ausreißer (Zitationen > 500) sind exkludiert. Die Sterne beziehen sich auf ein Signifikanzniveau von: * p < 0.10 ** p < 0.05 *** p < 0.01.

Die Verwendung von Registerdaten hat auch Einfluss auf die KoautorInnenenschaft: Ein Paper, das Registerdaten verwendet, hat im Schnitt 0,3 mal mehr AutorInnen. Dies ist vermutlich zum einen der Tatsache geschuldet, dass die Aufbereitung von administrativen Daten hohe methodische und institutionelle Kenntnisse verlangt. Zum anderen ist es auch Indiz dafür, dass durch die Verfügbarkeit von Daten wichtige Kooperationen im wissenschaftlichen Umfeld entstehen können, die zur Beteiligung österreichischer AutorInnen an international renommierten Publikationen führen. Ein ungleicher und teils schwieriger Zugang zu Daten steht dem im Wege und stellt einen Wettbewerbsnachteil für einzelne Länder, akademische Einheiten und Forschende dar. Sie kommen schwieriger an Fördergelder, schaffen es weniger häufig in anerkannte Fachzeitschriften und werden seltener zitiert.

Beispiel 3: Mehr Evidenz in einer Pandemie

Im Rahmen der COVID-19-Pandemie hat sich die Verfügbarkeit von Registerdaten für die wissenschaftliche Forschung als besonders relevant herausgestellt. Viele Fragen, die sich gestellt haben, konnten nicht beantwortet werden, weil die Daten zwar grundsätzlich angelegt sind, aber entweder nicht darauf zugegriffen werden durfte oder konnte, oder weil es keine entsprechende Expertise in der Aufbereitung der Datensätze gab, oder weil es nicht möglich war, zwei (oder mehr) Datensätze miteinander zu verknüpfen.

Im Folgenden wird ein Blick ins Ausland geworfen, und zwar konkret auf zwei Themen, die deutlich machen können, wie die Verfügbarkeit von Registerdaten für die wissenschaftliche Forschung sehr rasch den Kenntnisstand im Bereich der Gesundheitspolitik erweitern kann – was insbesondere in Zeiten einer Pandemie von höchster Wichtigkeit ist. Die beiden Themen sind erstens die Frage, wie man die Faktoren genauer bestimmen kann, welche bei einer Erkrankung mit COVID-19 zum Tod führen; und zweitens, wie die im Zuge der Pandemie beobachteten Veränderungen im sensiblen Bereich der Schwangerschaft und Geburt durch Verwendung von Registerdaten zu neuen Aufschlüssen führen, die ihrerseits gesundheitspolitische Maßnahmen ergeben können.

Gründe für den tödlichen Verlauf einer Erkrankung an COVID-19

Im Sommer 2020 publizierte eine Gruppe britischer WissenschaftlerInnen eine Studie mit dem Titel „Factors associated with COVID-19-related death using OpenSAFELY“ in der Zeitschrift *Nature*. Die Studie basiert auf einer extensiven statistischen Analyse eines großen Datensatzes, nämlich rund 40 Prozent der englischen Bevölkerung. Die Datenbasis beruht auf der Plattform OpenSAFELY, die ihrerseits auf die Einträge von allgemeinmedizinischen Arztpraxen des National Health Service (NHS) zurückgreift. Diese Daten wurden mit der Todesstatistik der englischen Statistikbehörde (ONS) verknüpft und dabei jene Todesfälle identifiziert, die als von COVID-19 verursacht gezählt werden.

Insgesamt umfasst die Studienpopulation mehr als 17 Millionen Individuen, nämlich all jene Personen in England, von denen in der Vergangenheit ein Krankenverzeichnis beim NHS angelegt worden ist. Von dieser Population sind im Untersuchungszeitraum der Studie (1. Februar bis 6. Mai 2020) 10.926 Personen an COVID-19 verstorben. Aufgrund dieser großen Zahl können die StudienautorInnen konkret nach dem Zusammenhang von verschiedenen demografischen, gesundheitlichen und sozialen Faktoren und einem tödlichen Verlauf einer COVID-19-Erkrankung fragen. Bestätigt wird dabei, dass das Lebensalter eine sehr starke Assoziation mit dem Risiko auf einen tödlichen Verlauf aufweist; ebenso ist es mit dem Geschlecht (Männer haben ein höheres Risiko als Frauen). Die Zugehörigkeit zu einer ethnischen Minderheit steigert das Risiko ebenfalls beträchtlich; dabei konnte aber nicht geklärt werden, ob dies wiederum mit dem beruflichen Umfeld zu tun hat. Eindeutig ist auch der soziale Status ausschlaggebend: Personen, die der Gruppe des Fünftels der sozial Benachteiligten angehören, haben zu Personen aus der Gruppe des Fünftels der sozial am höchsten Stehenden ein um den Faktor 1,79 höheres Risiko, bei Erkrankung an COVID-19 auch daran zu sterben.

Das Studiendesign beziehungsweise der dafür verwendete Datensatz erlaubt keine Interpretation dahingehend, dass zwischen demografischen und gesundheitlichen Faktoren auf der einen Seite und der Wahrscheinlichkeit eines tödlichen Verlaufs einer COVID-19-Erkrankung auf der anderen ein kausaler Zusammenhang bestünde. Außerdem kann die Studie auch auf eine Reihe von relevanten Fragen keine Antwort geben, etwa welche Berufsfelder und welche anderen sozialen Faktoren mit dem erhöhten Risiko eines tödlichen Verlaufs assoziiert sind – denn

dazu lagen zum Zeitpunkt der Auswertung keine Daten vor. Das ist deshalb schade, weil die Studie selbst recht deutlich zeigt, dass Vorerkrankungen (also gesundheitliche Faktoren) nur zum Teil erklären konnten, warum das Risiko eines tödlichen Verlaufs in bestimmten demografischen Gruppen (ethnischen Minderheiten und sozial Benachteiligten) höher ist.

Dennoch – und insbesondere für unsere Zwecke – ist die vorliegende Studie bemerkenswert. Ihr Vorteil liegt darin, dass sie auf eine so große Grundgesamtheit zurückgreifen kann und dies auch noch in einem vergleichsweise atemberaubend schnellen Tempo. Zwar ist diese Studie nicht die erste, die sich mit dem Zusammenhang von sozialen und gesundheitlichen Faktoren und einem tödlichen Verlauf einer COVID-19-Erkrankung beschäftigt. Aber sie stellt alle anderen Studien hinsichtlich ihrer Größe in den Schatten.⁶

Traditionelle Studien zum Zusammenhang von Todesursache und gesundheitlichen sowie sozialen Faktoren beruhen immer nur auf kleinen Samples von PatientInnen und die tatsächlichen Fallzahlen sind dann immer sehr klein. Bei einer so großen Grundgesamtheit, wie sie in der Plattform OpenSAFELY zur wissenschaftlichen Analyse der COVID-19-Pandemie eingerichtet wurde, ist es möglich, vollständig pseudonymisierte Individualdaten so zu analysieren, dass eine Aussagekraft erreicht wird, die von kleineren Studien nicht erwartet werden kann. Außerdem kann sehr viel genauer nach einzelnen Faktoren und Subsamples gefragt werden, für die dann immer noch ausreichend große Fallzahlen vorliegen (selbst wenn die Studie aufgrund der Verteilung der Studienpopulation nicht unbedingt vollständige Repräsentativität der Gesamtbevölkerung Englands in Anspruch nehmen darf).

Klar ist aber auch, dass eine solche Studie eine erhebliche Vorleistung in Form einer belastbaren Datenplattform braucht, und auch ein ForscherInnenteam, welches sich mit vereinten Kräften auf die Ausarbeitung konzentrieren kann, und nicht zuletzt die Bereitschaft der Politik, diese Daten zur Verfügung zu stellen.

Die Effekte der Pandemie auf Frühgeburten

Die Auswirkungen der Pandemie zu verstehen, ist die eine Sache. Aber in gewissem Sinne erlauben die Pandemie und die damit einhergehenden Änderungen in unserem Leben – Verhalten, Physical Distancing etc. – auch, dass die Ursachen bestimmter Probleme näher betrachtet werden. Ein besonders markantes Beispiel, welches auch bereits mediale Aufmerksamkeit erhalten hat, betrifft Frühgeburten. In vielen industrialisierten Ländern stellen Frühgeburten ein persistentes Problem dar: die damit verbundenen Gefahren für die Mutter wie auch das Neugeborene sind erheblich größer, als wenn die Geburt nach einer biologisch normalen Schwangerschaftsperiode passiert.

Bereits im Frühjahr 2020, kurz nach Eintreten der in Europa zu unterschiedlichen Zeitpunkten ausgerufenen Maßnahmen und Restriktionen zur Eindämmung der COVID-19-Pandemie, stellten ForscherInnen fest, dass die Zahl der Frühgeburten rückläufig zu sein schien. Dieses Ergebnis haben sowohl ForscherInnen in Dänemark und den Niederlanden als auch in Irland zu lebenden Neugeborenen empirisch überprüft. Möglich war das durch den für ForscherInnen geltenden (und genau festgelegten) Zugang zu den entsprechenden Registerdatensätzen; in Dänemark etwa durch die Danish Neonatal Screening Biobank (DNSB). So war es auch möglich, mit Gesamtpopulationen beziehungsweise mit sehr großen Samples zu arbeiten.

⁶ Dies bestätigt auch der entsprechende Eintrag im Novel Coronavirus Research Compendium (NCRC) der Johns Hopkins University, das COVID-19-Studien kuratiert und nach ihrer wissenschaftlichen Qualität beurteilt; siehe <https://ncrc.jhsph.edu/research/opensafely-factors-associated-with-covid-19-death-in-17-million-patients/>

In der dänischen Untersuchung wurden etwa 31.180 Kinder analysiert, welche jeweils zwischen 12. März und 14. April der Jahre 2015 bis 2020 geboren wurden. In der irischen Studie wurden die Lebendgeburten einer Klinik (University Maternity Hospital Limerick, UMHL) zwischen 2001 und 2020 jeweils im Zeitraum zwischen Januar und April analysiert (30.705 Fälle) und dabei die Fälle mit besonders niedrigem Gewicht als Proxy für Frühgeburten gezählt. Im Fall der niederländischen Studie wurden 56.720 Lebendgeburten zwischen 9. März und 16. Juli 2020 verglichen mit den insgesamt 1.599.547 Lebendgeburten in den knapp zehn Jahren davor (ab 9. Oktober 2010).

Das Ergebnis dieser drei voneinander unabhängigen und auch nach unterschiedlichen Verfahren durchgeführten Untersuchungen ist eindeutig: die statistische Verteilung des Gestationsalters (d. i. die Dauer der Schwangerschaft zum Zeitpunkt der Geburt) hat sich nach Ausbruch der Pandemie in allen drei Ländern deutlich verändert, und zwar deshalb, weil die Zahl der Frühgeburten stark abgenommen hat. In allen drei Untersuchungen sind die Ergebnisse statistisch signifikant.

Die Untersuchungen haben einige mediale Aufmerksamkeit erhalten.⁷ Gründe dafür, warum es zu der Reduktion von Frühgeburten während des Lockdowns im Frühjahr 2020 gekommen ist, konnten empirisch noch nicht eruiert werden. Eine generelle (und naheliegende) Vermutung lautet, dass die alltägliche, praktische Hygiene und andere verhaltensrelevante Aspekte ausschlaggebend sind. In der Erklärung von Frühgeburten wurden schon bisher Entzündungen bei der Mutter als ein ausschlaggebender Faktor gesehen. Der Lockdown hat in allen drei Ländern zu massiven Verhaltensänderungen geführt, physische Kontakte reduziert und neue Betonung auf Hygiene gelegt. Es liegt der Zusammenhang nahe, dass damit auch die entsprechenden Belastungen der schwangeren Mütter insgesamt reduziert wurden. Umgekehrt wäre es möglich, dass mehr Totgeburten für die Veränderung in der Verteilung verantwortlich sind, da diese Angabe aus den Daten nicht zu ersehen sind (es werden nur lebend geborene Kinder erfasst). Auch andere, wie zum Beispiel soziodemografische Faktoren sind bisher nicht berechnet worden.

Weiterführende Untersuchungen sind nun bemüht, mögliche Zusammenhänge empirisch zu überprüfen und damit ein besseres Verständnis darüber zu bekommen, warum es in Vor-Pandemie-Zeiten so häufig zu Frühgeburten gekommen ist, und mit welchen Maßnahmen sie sich zukünftig besser vermeiden lassen.⁸ Die Prävention von Frühgeburten ist eine gesundheitspolitische Priorität, und die Registerdaten können im Rahmen eines quasi-natürlichen Experiments verwendet werden, um klare Hinweise zu identifizieren, wie Frühgeburten potenziell in Zukunft viel besser vermieden werden können.

Wie die AutorInnen der niederländischen Studie betonen:

„International collaborative efforts are needed to collate evidence from across the globe to further substantiate these findings and to study the underlying mechanisms.“ (Hedermann et al. 2020)

Damit Österreich hier seinen Beitrag liefern kann, ist der entsprechende Datenzugang unter Wahrung strenger datenschutzrechtlicher Maßgaben erforderlich. Doch nicht nur für die internationale Kollaboration wäre dies zweckdienlich. Regionale Unterschiede in den Gründen, die

⁷ Siehe etwa die beiden ausführlichen Berichte in der New York Times; online: <https://www.ny-times.com/2020/07/19/health/coronavirus-premature-birth.html> und <https://www.ny-times.com/2020/10/15/health/covid-premature-births-lockdown.html>.

⁸ Siehe <https://www.ipopstudy.com/>

zu einer Frühgeburt führen können, sind durchaus möglich; um die Erkenntnisse der Forschung auf Österreich anzuwenden, ist es umso dringender notwendig, dass auch hierzulande die entsprechenden juristischen und technischen Rahmenbedingungen geschaffen werden, um Registerforschung zum Nutzen der Öffentlichkeit durchführen zu können.

Literatur

- Been, J. V., Burgos Ochoa, L., Bertens, L. C. M., Schoenmakers, S., Steegers, E. A. P., Reiss, I. K. M. 2020. "Impact of COVID-19 Mitigation Measures on the Incidence of Preterm Birth: A National Quasi-Experimental Study." *The Lancet Public Health* 5 (11): e604–11. [https://doi.org/10.1016/S2468-2667\(20\)30223-1](https://doi.org/10.1016/S2468-2667(20)30223-1).
- Halla, M., Artner, A., & Stumpfner, S. (2020, unveröffentlicht). How Did Administrative Data Change Economic Research Output – A Descriptive Analysis. Keynote der Konferenz der NOeG 2020.
- Hedermann, G., Hedley, P. L., Bækvad-Hansen, M., Hjalgrim, H., Rostgaard, K., Poorisrisak, P., Breindahl, M., et al., 2020. "Danish Premature Birth Rates during the COVID-19 Lockdown." *Archives of Disease in Childhood – Fetal and Neonatal Edition*, August. <https://doi.org/10.1136/archdischild-2020-319990>.
- Mühlböck, M., Kalleitner, F., Steiber, N., & Kittel, B. (2020). Information, Reflection, and Successful Job Search: A Nudging Experiment. Reflection, and Successful Job Search: A Nudging Experiment (April 15, 2020).
- Philip, R. K., Purtill, H., Reidy, E., Daly, M., Imcha, M., McGrath, D., O'Connell, N. H., Dunne, C. P., 2020. "Reduction in Preterm Births during the COVID-19 Lockdown in Ireland: A Natural Experiment Allowing Analysis of Data from the Prior Two Decades." *MedRxiv*, June, 2020.06.03.20121442. <https://doi.org/10.1101/2020.06.03.20121442>.
- Williamson, E. J., Walker A. J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C. E., Curtis, H. J., et al., 2020. "Factors Associated with COVID-19-Related Death Using OpenSAFELY." *Nature*, July. <https://doi.org/10.1038/s41586-020-2521-4>.

Anhang

Zusätzliche Erfolgsbeispiele der Plattform Registerforschung

Mit jedem Meter Daten näher zur Evidenz

Naturkatastrophen wie Überschwemmungen und Lawinen können für die BewohnerInnen Österreichs existenzbedrohend sein. Sie haben aber nicht nur unmittelbar Einfluss auf das Leben der Menschen in den betroffenen Regionen. Gefahrenpotenzial und Schutzinvestitionen wirken sich auf regionaler Ebene langfristig auf Wirtschaft und Bevölkerung aus. Durch innovative methodische Ansätze und den Zugang zu kleinräumigem Datenmaterial können diese Auswirkungen gemessen werden.

Sinabell, F., Fritz, O., Puwein, W., & Streicher, G. (2009). Eine volkswirtschaftliche Analyse der Wildbach- und Lawinenverbauung. WIFO Studies.

Der Knick im Steuertarif und seine Folgen

Ein Forschungsprojekt zeigt mittels Verknüpfung von Daten der Lohnzettel und Arbeitnehmerveranlagungen, wie österreichische ArbeitnehmerInnen ihr zu versteuerndes Einkommen beeinflussen, um ihre Steuerlast zu reduzieren. So hilft Registerforschung, um bei zukünftigen Steuerreformen mögliche Ausweicheffekte besser abschätzen zu können.

Paetzold, J. (2019). How do taxpayers respond to a large kink? Evidence on earnings and deduction behavior from Austria. *International Tax and Public Finance*, 26(1), 167–197.