

## Corona: Konzepte für eine verbesserte statistische Datenbasis

**Wilfried Grossmann**  
Universität Wien

**Peter Hackl**  
Wirtschaftsuniversität Wien  
Masarykova Univerzita Brno  
Ehemals Statistik Austria

**Josef Richter**  
Ehemals  
Wirtschaftskammer Österreich  
Universität Innsbruck  
Universität Linz

---

### Abstract

Die Corona-Statistiken und deren Interpretation stehen im Fokus des öffentlichen Interesses. Die Daten spielen bei Maßnahmen zur Eindämmung der Pandemie eine zentrale Rolle. Betrachtet man den Status quo der veröffentlichten Corona-Datenbasis, so klafft zwischen den berechtigten Wünschen und den Notwendigkeiten einer evidenzbasierten Entscheidungsfindung auf der einen Seite und der Wirklichkeit auf der anderen Seite eine bedauerliche Diskrepanz. Es gilt, deutlich verbesserte statistische Grundlagen zur Einschätzung der Situation, zur Information der Bevölkerung und für weiterführende Analysen zu schaffen. Dazu will der Beitrag konkrete Schritte vorschlagen.

Im Abschnitt 2 werden die vorhandenen administrativen Daten beschrieben und nach den für die Amtliche Statistik geltenden Qualitätskriterien analysiert. Im Abschnitt 3 wird zuerst ein *Statistisches Corona Register* definiert und dann – als entscheidender Schritt – ein *Erweitertes Statistisches Corona Register* konzipiert, das Pandemiedaten mit demographischen und sozialstatistischen Daten zusammen führt. Ein solches Register böte umfassende Analysemöglichkeiten und könnte als Vorstufe für ein Statistisches Data Warehouse für das epidemiologische Geschehen angesehen werden. In Abschnitt 4 werden einige Analysemöglichkeiten mit dem vorgeschlagenen *Erweiterten Statistischem Corona Register* aufgezeigt. Abschnitt 5 diskutiert Fragen des Berichtswesens. Abschnitt 6 fasst die Ergebnisse zusammen und gibt einige Empfehlungen für Verbesserungen im Informationsmanagement von Verwaltung und Statistik bei der Bewältigung neuartiger unerwarteter Aufgabenstellungen.

Für die Implementierung eines *Erweiterten Statistischem Corona Registers* ist die intensive Kooperation des Gesundheitsbereiches mit der Amtlichen Statistik unumgänglich. Dies deshalb, weil die Statistik über viele der benötigten Grunddaten verfügt, ein reichhaltiges Inventar an Methoden zur Integration von Datenquellen kennt, eine hohe Kompetenz im Umgang mit personenbezogenen Daten besitzt und viel Erfahrung mit einem Berichtswesen hat, das hohen Standards entspricht.

*Keywords:* Corona, Datenqualität, Statistisches Register, Datenintegration.

---

## 1. Einleitung

Nach dem Ausbruch von COVID-19 rückten Corona-Statistiken und deren Interpretation in den Fokus des öffentlichen Interesses. Die Statistiken zu den COVID-19 Erkrankungen spielen bei Maßnahmen zur Eindämmung der Pandemie eine zentrale Rolle. Die zuständigen Stellen, allen voran das Ministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz, im Weiteren kurz Gesundheitsministerium, publizieren eine Reihe von Statistiken zum Corona-Geschehen. Am täglich aktualisierten *AGES-Dashboard* ([AGES 2020a](#)) werden Statistiken zu Testfällen, Infektionen, Erkrankungen, Hospitalisierungen und Todesfällen zur Verfügung gestellt. Daneben gibt es noch ein *Impf-Dashboard* ([Gesundheitsministerium 2021b](#)) und eine Fülle von Dokumenten mit häufig geänderten Definitionen, Erklärungen und Verhaltensregeln. Man gewinnt den Eindruck, dass viele Modifikationen im Berichtswesen nicht proaktiv nach einem vorausschauenden Konzept erfolgten. Im März 2021 wurde eine Gesamtdokumentation publiziert, die als Leitfaden dienen soll ([Gesundheitsministerium 2021c](#)).

Viele Daten sind als Open Data ([Gesundheitsministerium 2021f](#)) verfügbar und werden von unterschiedlichen Interessensgruppen für verschiedene Analysen verwendet. Sie sind ihrer Natur nach Administrative Daten und von hoher Relevanz für Entscheidungen in Politik und öffentlicher Gesundheit sowie für die Berichterstattung in den Medien. Nur in beschränktem Umfang sind diese Daten für die Analyse von sozioökonomischen und demographischen Auswirkungen von COVID-19 geeignet, da mit Ausnahme von Alter, Geschlecht und regionalen Merkmalen keine wirtschaftlichen und sozialen Indikatoren vorhanden sind.

Ziel dieser Arbeit ist es, einige Ideen zur Schaffung einer umfassenden Datenbasis und insbesondere für das Zusammenführen der Daten über das Pandemiegeschehen mit sozioökonomischen Daten darzustellen. Dazu werden zunächst die vier administrativen Datenquellen dargestellt, vom Standpunkt der Qualitätskriterien für die Amtliche Statistik analysiert und Mängel in den derzeit vorhandenen Daten aufgezeigt. In Abschnitt 3 werden ein Konzept eines *Statistischen Corona Registers* sowie eine Ergänzung dieses Registers mit demographischen und sozialstatistischen Daten zu einem *Erweiterten Statistischen Corona Register* dargestellt. Weiter werden Überlegungen für eine Informationsplattform präsentiert, die umfassend und auf hohem Qualitätsniveau die Information der Öffentlichkeit erlauben und die Basis für wissenschaftliche Analysen bereitstellen könnte. Die Konzepte eines solchen Systems, das neben medizinischen Aspekten auch sozioökonomische Gegebenheiten berücksichtigt und für das Gesundheitssystem als Ganzes von eminenter Bedeutung ist, sollten mittelfristig als ein Statistisches Data Warehouse für das Gesundheitswesen realisiert werden. In Abschnitt 4 werden einige Analysemöglichkeiten mit diesem *Erweiterten Statistischen Corona Register* behandelt, die über die derzeit möglichen Untersuchungen hinausgehen. Abschnitt 5 diskutiert Fragen des Berichtswesens und der Veröffentlichung der Daten. Abschnitt 6 fasst die Ergebnisse zusammen und gibt einige Empfehlungen für Verbesserungen im Management von solchen für Verwaltung und Statistik neuartigen Aufgabenstellungen.

Die durch die Pandemie gestellten Herausforderungen für die Amtliche Statistik sind nicht nur in Österreich ein zentrales Thema. Die Stellungnahmen der Deutschen Arbeitsgemeinschaft für Statistik ([Deutsche Arbeitsgemeinschaft Statistik 2021](#)) oder der Royal Statistical Society ([Royal Statistical Society 2021](#)) machen dies deutlich. Sie diskutieren die Bedeutung von Daten und Statistik als Grundlage für Entscheidungen und weisen auch auf die Verantwortung der wissenschaftlichen statistischen Gesellschaften für eine bessere *Statistical literacy* in der Bevölkerung hin. In den Arbeiten von [Cook and Gray \(2020\)](#) und [Ljones \(2020\)](#) werden vom strategischen Standpunkt die kurz- und langfristigen Aufgaben von nationalen Statistikämtern zur besseren Bewältigung solcher Ereignisse detailliert beschrieben. Beispiele zur besseren Nutzung von sozioökonomischen Daten zur Erklärung des Pandemiegeschehens findet man unter anderem in [Adhikari, Pantaleo, Feldman, Ogedegbe, Thorpe, and Troxel \(2020\)](#) oder [Kissam \(2020\)](#). Die statistischen Herausforderungen für internationale statistische Organisationen werden in [Fu, Hereward, MacFeely, Me, and Wilmoth \(2020\)](#) behandelt, und [Ashofteha and Bravo \(2020\)](#) analysieren die Problematik internationaler Vergleiche von

COVID-19 Daten. Wie in den verschiedenen offiziellen Publikationen werden in diesem Beitrag die Termini COVID-19 und SARS-COV-2 synonym verwendet.

## 2. Informationen zu COVID-19: Probleme des Status quo

Zur Erfassung des epidemiologischen Geschehens werden vom Gesundheitsministerium drei Register geführt, in denen Daten zu COVID-19 Infektionen, zu COVID-19 Tests und zu durchgeführten Impfungen dokumentiert werden. Die Register-Einträge sind fallbezogen und informieren über jeweils ein Infektions-, Test- oder Impf-Ereignis. Die rechtlichen Grundlagen für diese Register sind im Epidemiegesetz (BGBl 1950/186 idF BGBl I 2021/64 2021) niedergelegt. Dokumente des Gesundheitsministeriums über Falldefinitionen (Gesundheitsministerium 2021d), Teststrategien (Gesundheitsministerium 2021e) und Nachverfolgung von Kontaktpersonen (Gesundheitsministerium 2021a) informieren über die Details und die Durchführung der Bestimmungen. Bedingt durch die Neuartigkeit von SARS-COV-2, das zunehmende Wissen über die Erkrankung und die sich ändernden, unterschiedlichen politischen Strategien bei der Bekämpfung der Pandemie (insbesondere was die Testmöglichkeiten und Impfungen betrifft), wurden diese Dokumente im Laufe der Pandemie immer wieder modifiziert. In der hier gegebenen Beschreibung wird von den Versionen mit Stand 15.6.2021 ausgegangen. Eine Gesamtdokumentation ist in Gesundheitsministerium (2021c) zu finden. Als Datenquelle für die Beschreibung werden die in Open Data (Gesundheitsministerium 2021f) angebotenen, öffentlich zugänglichen Daten verwendet. Die relativ ausführliche Wiedergabe der Definitionen und der Verwaltungsvorgänge erfolgt, weil ohne Kenntnis der genauen Eigenschaften der Administrativen Daten eine Beurteilung der auf ihnen aufbauenden statistischen Ergebnisse unmöglich ist. Dieser Abschnitt ist in fünf Teile gegliedert. Abschnitt 2.1 beschreibt als wichtigste Datenquelle das *Register der anzeigepflichtigen Krankheiten*, Abschnitt 2.2 das *Testregister*, Abschnitt 2.3 das *Impfregister* und Abschnitt 2.4 die Daten über die Kapazitäten des Gesundheitssystems; falls es vom inhaltlichen Standpunkt notwendig erscheint, werden Mängel und wünschenswerte Ergänzungen aufgezeigt. Abschnitt 2.5 gibt eine Bewertung der Qualität aus Sicht der statistischen Qualitätskriterien.

### 2.1. Register der anzeigepflichtigen Krankheiten (EMS)

Für die Beurteilung des Infektionsgeschehens liefert das *Register der anzeigepflichtigen Krankheiten*, das als *Epidemiologisches Meldesystem* (EMS) vom Gesundheitsministerium geführt wird, die wichtigsten Informationen. Das Dokument über Falldefinitionen (Gesundheitsministerium 2021d) definiert die für die Aufnahme in das EMS wesentlichen Kriterien. Die folgenden fünf Kriterien werden unterschieden:

- Klinische Kriterien: Jede Person mit mindestens einem der folgenden Symptome: Husten, Fieber, Kurzatmigkeit, plötzliches Auftreten einer Störung bzw. Verlust des Geschmacks- oder Geruchssinns. In einer Fußnote werden auch noch weniger spezifische Symptome wie Kopfschmerzen, Muskelschmerzen, Gliederschmerzen, Abgeschlagenheit, Müdigkeit, Erbrechen und/oder Durchfall genannt.
- Diagnostisches Bildgebungskriterium: Radiologische Hinweise auf COVID-19 kompatible Läsionen.
- Labordiagnostisches Kriterium - PCR: Nachweis von SARS-COV-2 spezifischer Nukleinsäure in einer klinischen Probe mittels PCR. Dieser Test gilt international als der Goldstandard zum Nachweis von SARS-COV-2. Details findet man in Abschnitt 7.3.1 der Gesamtdokumentation (Gesundheitsministerium 2021c).
- Labordiagnostisches Kriterium - Antigen: Nachweis von SARS-COV-2 spezifischem Antigen in einer klinischen Probe mittels Antigentests. Dieser Test kann in verschiedenen

Formen durchgeführt werden und hat den Vorteil, dass die Auswertung der Tests rasch und ohne Labor verfügbar wird. Der Nachteil ist die geringere Genauigkeit. Details über die Durchführung und über die Qualität der verschiedenen Formen des Tests findet man in [Gesundheitsministerium \(2021e\)](#).

- Epidemiologische Kriterien: Bestehen eines Kontaktes mit einer an SARS-COV-2 erkrankten Person. Es werden Kontaktpersonen der Kategorie I und II unterschieden (siehe Dokument „Behördliche Vorgangsweise bei SARS-COV-2 Kontaktpersonen: Kontaktpersonennachverfolgung“ ([Gesundheitsministerium 2021a](#))).

Auf der Basis dieser Kriterien werden die drei Fälle *Verdachtsfall*, *Wahrscheinlicher Fall* und *Bestätigter Fall* unterschieden. Der Begriff *Wahrscheinlicher Fall* kommt anscheinend nur in diesem Dokument und auf der Homepage des Gesundheitsministeriums vor und scheint in der Praxis keine Rolle zu spielen. Es soll hier kurz die – soweit erkennbare – weitere Vorgehensweise bei Vorliegen eines der fünf Kriterien beschrieben werden. Die Interpretation der Falldefinitionen folgt den Erläuterungen zu den Meldepflichten in Abschnitt 4 der Gesamtdokumentation ([Gesundheitsministerium 2021c](#)).

- Fälle/Personen mit klinischen Kriterien: Verdachtsfälle werden entweder von Ärztinnen und Ärzten und bestimmten Personengruppen an eine Bezirksverwaltungsbehörde gemeldet oder sie melden sich selbst bei einer Hotline. In jedem Fall sind diese Fälle dann zunächst im EMS erfasst, und es wird ein behördlicher PCR-Test durchgeführt. Bei positivem PCR-Test werden sie als *Laborbestätigter Fall* im EMS registriert.
- Fälle/Personen mit positivem PCR-Test: Bei positiv bestätigtem Test werden diese Personen im EMS als *Laborbestätigter Fall* registriert. Dies gilt unabhängig davon, ob sie klinische Kriterien erfüllen oder nicht (sogenannte asymptomatische Fälle). Auch für Personen, an denen der PCR-Test im Rahmen einer Hospitalisierung durchgeführt wird, fallen unter dieses Kriterium.
- Fälle/Personen mit Epidemiologischen Kriterien: Diese Personen werden aufgrund ihres Kontaktes zu einem bereits registrierten Fall behördlich identifiziert; das [Gesundheitsministerium \(2021a\)](#) beschreibt die Vorgangsweise für diese Fälle, insbesondere ihre Absonderung und die eventuelle Aufnahme in das EMS als *Laborbestätigter Fall*.
- Fälle/Personen mit positivem Antigentest: Wurde der Antigentest von einer registrierten Teststelle durchgeführt, so wird dieser Fall behördlich gemeldet und ein PCR-Test behördlich verordnet. Je nach Ergebnis des Tests wird dieser Fall im EMS als *Laborbestätigter Fall* registriert. Schwieriger ist es, Fälle zu identifizieren, die einen Antigentest zur Eigenanwendung durchgeführt haben. Diesen Personen wird empfohlen, sich bei der Gesundheits-Hotline zu melden, die dann die weiteren Erfassungsschritte vorgibt.
- Fälle/Personen mit positiven Diagnostischen Bildgebungskriterien: In den vorhandenen Dokumenten findet sich kein Hinweis, wie mit diesen Fällen weiter verfahren wird. Vermutlich wird auch in diesem Fall ein PCR-Test durchgeführt und entschieden, ob es sich um einen *Laborbestätigten Fall* handelt.

Es gibt also fünf unterschiedliche Möglichkeiten, die zur Erfassung eines Falles im EMS führen können. Zur Beschreibung des Status im Verlauf einer COVID-Erkrankung unterscheidet das EMS zwischen *Aktiven Fällen*, *Genesenen Fällen* und *Todesfällen*. Ein *Aktiver Fall* kann entweder ein *Aktiver Fall in häuslicher Pflege*, ein *Aktiver Fall auf Normalstation* oder ein *Aktiver Fall auf Intensivstation* sein ([AGES 2020b](#)). Ein nicht hospitalisierter und nicht verstorbener *Aktiver Fall* wird 21 Tage nach dem ersten positiven Test automatisch zu einem *Genesenen Fall*. Zu beachten ist, dass dies auch für asymptomatische Fälle gilt, das sind Fälle, bei denen ein positiver PCR-Test vorliegt, aber keines der klinischen Kriterien zutrifft. Zu

den Todesfällen ist anzumerken, dass die Kausalität von COVID-19 als Todesursache nicht untersucht wird. Wesentlich ist nur, dass zwischen dem Status *Aktiver Fall* und dem Status *Todesfall* kein Status *Genesener Fall* vorgelegen hat.

### *Verfügbare Daten*

Auf der Grundlage der im EMS gespeicherten Informationen veröffentlicht die Österreichische Agentur für Gesundheit und Ernährungssicherheit (AGES) auf ihrem *AGES-Dashboard* (AGES 2020a) täglich die Anzahl der *Laborbestätigten Fälle*, wobei nach Status der Krankheit und den Wohnbezirken gegliedert wird (in Wien nur für die Stadt insgesamt). Falls eine Person keinen Wohnort in Österreich hat, wird der Aufenthaltsort der Person in Österreich (zum Zeitpunkt der Erfassung) herangezogen. Bei der Datenbereitstellung erfolgt somit der Schritt von der Einzelbetrachtung für administrative Zwecke zur Gesamtbetrachtung der Pandemiesituation. Zeitreihen werden in Open Data (Gesundheitsministerium 2021f) und auch am *AGES-Dashboard* zur Verfügung gestellt. Ein Nachteil der veröffentlichten Zeitreihen besteht darin, dass die für jeden Tag ausgewiesenen Daten nicht nach dem Geschlecht oder dem Alter der Erkrankten gegliedert sind. Interessieren diese Verteilungen, müssen sie aus den täglich veröffentlichten Gesamtzahlen der Fälle für die Bundesländer, die von verschiedenen Institutionen und Experten gesammelt werden, berechnet werden. Ein Repository, das solche Information anbietet, ist zum Beispiel das Github-Repo *Statistik Austria* (2020b).

### *Inhaltliche Beurteilung der Daten*

Vom Standpunkt der statistischen Analyse ist diese Datenlage nicht zuletzt deswegen so unbefriedigend, da kein Bezug zu anderen sozioökonomischen Daten vorhanden ist. Nach § 4 (4) des Epidemiegesetzes sind aber im EMS noch eine Reihe von weiteren Datenkategorien vorhanden:

1. *Daten zur Identifikation von Erkrankten, einer Erkrankung Verdächtigen, Gebissenen, Verstorbenen oder Ausscheidern (Name, Geschlecht, Geburtsdatum, Wohnsitz, soweit vorhanden Telefonnummer und E-Mail-Adresse, Sozialversicherungsnummer und Bereichsspezifisches Personenkennzeichen (§ 9 E-GovG, BGBl. I Nr. 10/2004)),*
2. *gegebenenfalls Sterbedaten (Datum, Todesursache, Autopsiestatus),*
3. *die für die anzeigepflichtige Krankheit relevanten klinischen Daten (Vorgeschichte und Krankheitsverlauf) sowie die in § 24c Abs. 2 Z 2 GTelG 2012 genannten Angaben und Labordaten sofern für die Zwecke des Abs. 2 erforderlich auch negative Testergebnisse auf SARS-CoV-2,*
4. *Daten zum Umfeld des Erkrankten, einer Erkrankung Verdächtigen, Gebissenen, Verstorbenen oder Ausscheiders, soweit sie in Bezug zur anzeigepflichtigen Erkrankung stehen, sowie Daten zur Identifikation von Kontaktpersonen (Name, Telefonnummer, E-Mail-Adresse, Wohnsitz) und*
5. *Daten zu den getroffenen Vorkehrungsmaßnahmen.*

Dass die genannten Merkmale zur persönlichen Identifikation nach Ziffer 1. und Ziffer 4. nicht publiziert werden, ist aus Datenschutzgründen selbstverständlich. Der Grund, dass andere Merkmale, insbesondere jene nach den Ziffern 3. bis 5., nicht einmal in aggregierter Form zur Verfügung gestellt werden, ist nicht so einsehbar und ist vielleicht historisch bedingt. Bei einer der zahlreichen, anlassbezogenen Änderungen des Epidemiegesetzes hätte man auch die Erfassung einiger für Analysen wesentlicher, soziodemographischer Daten vorsehen können. Die vorliegenden Daten sind Administrative Daten, die vermutlich zur Bewältigung der Aufgaben der Verwaltung geeignet sind. Dass administrative Daten nicht immer unmittelbar auch den



Anforderungen der Statistik entsprechen, ist ein bekanntes Phänomen. Aus den vielen statistischen Arbeiten, die auf der Nutzung von Verwaltungsdaten aufbauen, ist freilich ebenso bekannt, dass oft geringe Modifikationen die Eignung der Daten für statistische Zwecke verbessern. Die folgenden Änderungen würden eine verbesserte Qualität der Corona-Statistiken ermöglichen:

1. Bedingt durch die unterschiedlichen Arten der Meldungen kommt es oft zu Verzögerungen in der Erfassung der Fälle. Es sollte daher zwischen vorläufigen und endgültigen Daten zum Zeitpunkt der Veröffentlichung unterschieden werden.
2. Die unterschiedlichen Kriterien, die zur Eintragung in das EMS gemäß den Fällen a) bis e) führten, sollten in dieser Detaillierung im Register vermerkt sein.
3. Neben dem Wohnort sollte auch der Ort der Erfassung des Falles angeführt werden. In den Erläuterungen zum EMS in Abschnitt 8.4 der Gesamtdokumentation ([Gesundheitsministerium 2021c](#)) wird zwar auch von Wohnort und gegenwärtigem Aufenthaltsort gesprochen, aber ob das realisiert wird, ist nicht klar.
4. Für Krankheitsfälle, die bei ausländischen Touristen, Asylsuchenden, Obdachlosen, Saisonarbeiter und 24-Std.-PflegerInnen mit nicht-österreichischer Nationalität, die keinen Wohnsitz in Österreich haben, wäre eine „Extraregion“ zu schaffen. Damit könnten alle Fälle, für die in Österreich Information anfällt, entsprechend klassifiziert werden.
5. Aus konzeptionellen Gründen sollten in das *Register der anzeigepflichtigen Krankheiten* auch jene Personen (und die ihnen zugeordneten Fälle) aufgenommen werden, die ihren Wohnsitz in Österreich haben, die aber im Ausland erkrankt bzw. getestet wurden. Wie aus vielen Fällen der Nutzung von Verwaltungsdaten bekannt, ist es schwierig, diese Information aus dem Ausland vollständig und in der erforderlichen Qualität zu erhalten. Ausnahme sind vermutlich hospitalisierte Fälle. Bei der Interpretation von Ergebnissen nach dem Wohnsitz ist daher mit einer systematischen Untererfassung zu rechnen.

## 2.2. Testregister

Ein zentrales Element der Bekämpfung der Pandemie ist die Teststrategie. Sie soll die Verbreitung der Krankheit kontrollieren, das Funktionieren des Gesundheitssystems sicherstellen, Krankheitsausbrüche unter speziellen Bedingungen identifizieren (z.B. in Schulen), und auch länderübergreifende Infektionen verhindern. Grundsätzlich kann es sich bei den Tests entweder um behördlich veranlasste Tests (von Verdachtsfällen, Kontaktpersonen), um spezifische Screeningprogramme für bestimmte Gruppen (z.B. in Alten- und Pflegeheimen, medizinischen Einrichtungen, Schulen), oder um bevölkerungsweite Screeningprogramme handeln ([Gesundheitsministerium 2021e](#)). Derzeit werden primär PCR-Tests und verschiedene Arten von spezifischen Antigentests verwendet. Zu beachten ist, dass eine Person freiwillig beliebig viele Tests durchführen lassen kann. Als behördlich veranlasste Tests werden PCR-Tests verwendet, die derzeit den Goldstandard für die Identifikation von Krankheitsfällen darstellen. Da diese Tests in Labors ausgewertet werden, erfolgt in der Regel eine automatische Meldung eines positiven Testergebnisses an das Gesundheitsministerium. Bei den spezifischen und den bevölkerungsweiten Screeningprogrammen können entweder PCR-Tests oder spezifische Antigentests verwendet werden. Antigentests werden in Teststraßen, Apotheken, Betrieben und als Tests in Eigenanwendung (Selbsttests) durchgeführt. Die weitere Bearbeitung der Fälle mit positiven Testergebnisse erfolgt nach den in den vorigen Unterabschnitten beschriebenen Kriterien b) und d). Nach § 5b des Epidemiegesetzes ist ein Register für Screeningprogramme (im Weiteren als *Testregister* bezeichnet) vorgesehen. Jeder Test ist in diesem *Testregister* ein Fall. Nach der derzeitigen Rechtslage sind Daten über Testfälle mit negativen Testergebnissen zu löschen, sobald die getestete Person über das Ergebnis informiert wurde. Damit wird

das *Testregister* praktisch zu einem Verzeichnis von aktuell durchgeführten Tests und ist für weitere statistische Analysen wenig hilfreich.

### *Verfügbarkeit von Daten*

Die auf Open Data ([Gesundheitsministerium 2021f](#)) veröffentlichten Daten zu den Tests liefern nur tägliche Gesamtzahlen der durchgeführten PCR-Tests und Antigentest. Für Apotheken und Betriebstestungen werden ebenfalls die Gesamtzahl der Tests eines Tages in den Bundesländern veröffentlicht. Auch Informationen über die Schultestungen werden nur summarisch - jeweils für eine Woche - bereitgestellt. Derzeit (30.6.2021) gibt es auf Open Data ([Gesundheitsministerium 2021f](#)) keine aktuellen Daten über Schultestungen. Aus Gründen des Datenschutzes dürfte es schwierig sein, die Ergebnisse für die einzelnen Testfälle zu publizieren.

### *Inhaltliche Beurteilung der Daten*

Die Erstellung der im folgenden skizzierten Übersichtstabellen würde zusätzliche Informationen zur Analyse des Testgeschehens bieten:

- Ergebnisse der PCR-Tests: Die Gesamtzahl der PCR-Tests ist bekannt, und die Zahl der positiven Tests als *Laborbestätigte Fälle* ist im EMS vorhanden. Die Ergebnisse könnten auch regional und nach Geschlecht gegliedert werden.
- Ergebnisse der Gurgeltests: Die Ergebnisse dieser Tests sollten ähnlich wie die Ergebnisse der PCR-Tests dargestellt werden (eventuell als Untergruppe ausgewiesen).
- Ergebnisse der Antigentests: Für diese Tests könnten ebenfalls die positiven und negativen Ergebnisse, gegliedert nach dem Screeningprogramm (Gruppen- und bevölkerungsweites Screening), nach der Region, nach dem Alter und nach dem Geschlecht dargestellt werden.
- Darstellung der Multiplizität der Tests: Da die überwiegende Anzahl der bevölkerungsweiten Tests von den Gesundheitskassen finanziert wird, liegen für jede Person Informationen über die Zahl der verrechneten Tests vor. Daraus könnten man die Verteilungen der Testhäufigkeiten für bestimmte Altersgruppen und Regionen bestimmen. Eine derartige Darstellung scheint mit dem Datenschutz vereinbar zu sein. Sie würde einen Einblick in die Akzeptanz der Testprogramme durch die Bevölkerung geben. Zu bedenken ist allerdings, dass in diesen Daten Personen, die sich von ELGA, den elektronischen Gesundheitsakten ([ELGA 2021](#)), abgemeldet haben, nicht erfasst werden.

## **2.3. Impfreister**

Das *Impfreister* ist nach Epidemiegesetz § 4 Z 3a [BGBl 1950/186 idF BGBl I 2021/64 \(2021\)](#) ein Auszug aus dem zentralen Impfreister. Die Einträge in das Register sind in § 24c Z 2 Gesundheitstelematikgesetz ([BGBl 111/2012 idF BGBl I 2021/34 \(2021\)](#)) definiert:

- a) *zum Impfstoff (Klassifikation, Handelsname, Hersteller, Zulassungsnummer, Chargennummer, Verfallsdatum, Serialisierungsnummer, Pharmazentralnummer und Anato-misch-Therapeutisch-Chemische Zuordnung),*
- b) *zur verabreichten Impfung (Datum der Verabreichung, Dosierung und Dosis, angewandtes Impfschema, Impfpfehlung und Zuordnung zu Impfprogrammen),*
- c) *zur Bürgerin/zum Bürger (Name, Geburtsdatum, Geschlecht, Wohnadresse, Angaben zur Erreichbarkeit, Angaben zu einer allfälligen Vertretung, Sozialversicherungsnummer, bereichsspezifisches Personenkennzeichen Gesundheit, Gemeindecodex, Titerbestimmung, impfrelevante Vorerkrankungen und besondere Impfindikationen) sowie*

d zum impfenden bzw. speichernden Gesundheitsdiensteanbieter (Name, Rolle, Berufsadresse und Datum der Speicherung).

Die Abmeldung vom *Impfregister* ist nicht möglich.

#### *Verfügbarkeit von Daten*

Die Daten zu den Tests werden täglich am *Impf-Dashboard* publiziert und sind auch auf Open Data ([Gesundheitsministerium 2021f](#)) verfügbar.

#### *Inhaltliche Beurteilung der Daten*

Da die Struktur der Daten in einem langen Verhandlungsprozess entwickelt wurde, scheinen diese Daten vom inhaltlichen und rechtlichen Standpunkt unproblematisch.

## **2.4. Daten zum Gesundheitssystem**

Eine zentrale Aufgabe im Management einer Pandemie ist die Sicherstellung eines funktionierenden Gesundheitssystems. Im Fall von SARS-COV-2 hängt diese Funktion zentral von der Kapazität der Spitalsbetten, insbesondere der Betten auf Intensivstationen, ab. Diese vereinfachte Betrachtung ist unter der Annahme zulässig, dass die verfügbare Bettenkapazität automatisch auch die verfügbaren Personalressourcen impliziert.

#### *Verfügbarkeit von Daten*

In Open Data ([Gesundheitsministerium 2021f](#)) und am *AGES-Dashboard* werden neben den *Aktiven Fällen auf Normalstationen* und *Aktiven Fällen auf Intensivstationen* auch die Kapazitäten der freien Betten zum Download zur Verfügung gestellt. Am *AGES-Dashboard* ist die Zeitreihe länger.

#### *Inhaltliche Beurteilung der Daten*

Betrachtet man die Zeitreihen dieser Zahlen während der Pandemie in den einzelnen Bundesländern, so fällt auf, dass die Summenbedingung Gesamtzahl Betten = Anzahl Freie Betten + Anzahl Belegte Betten erst ab dem Frühjahr 2021 zutrifft. Betrachtet man die am *AGES-Dashboard* ([AGES 2020a](#)) abrufbare Zeitreihe der Hospitalisierungen seit dem 1.4.2020, so zeigt sich erst ab November 2020 der Beginn einer Konsolidierung bezüglich dieser Summenbedingung. Erst ab März 2021 scheint diese Konsolidierung in allen Bundesländern abgeschlossen. Die Ursachen für die schlechte Qualität dieser administrativen Daten dürften vielfältig und in den einzelnen Bundesländern auch unterschiedlich sein. So gibt es im Burgenland einige offensichtliche Ausreißer in der Kapazität, in Salzburg gibt es Perioden, in denen vermutlich keine Meldungen erfolgten. Offensichtlich fehlte vor der Berichterstattung eine Plausibilitätsprüfung der Angaben. Erst nach einer verstärkten Diskussion der Kapazitätsauslastung der Spitäler in den Medien wurde mit einer Bereinigung der Daten begonnen.

## **2.5. Die Informationssituation im Lichte statistischer Qualitätskriterien**

Die vom Gesundheitsministerium und nachgelagerten Dienststellen wie der AGES veröffentlichten (aggregierten) Daten haben ohne Zweifel den Charakter von Ergebnissen Amtlicher Statistik und werden auch als solche von der Öffentlichkeit wahrgenommen. Zudem ist das Gesundheitsministerium in dieser Rolle nach der Bestimmung des Bundesstatistikgesetz § 3 (19) ein Organ der Bundesstatistik. Amtliche Statistik sollte den *Fundamental Principles of Official Statistics* der Vereinten Nationen ([United Nations 2013](#)) entsprechen. Diese Prinzipien wurden für die europäischen Statistiken, die auf der Grundlage europäischer Rechtsnormen erstellt werden, noch detaillierter ausgearbeitet, um Indikatoren ergänzt und im *Verhaltenskodex für Europäische Statistiken* ([Eurostat 2017](#)) niedergelegt. Ohne Zweifel sind auf die



Corona-Statistiken die *Fundamental Principles* und sinngemäß die Grundsätze des *Verhaltenskodex* anzuwenden. Die in diesen beiden Dokumenten niedergelegten Kriterien und insbesondere der definierte Qualitätsrahmen und die Indikatoren zu deren Beurteilung bieten ein geeignetes Raster, um die Probleme der derzeitigen Datensituation zu charakterisieren.

### *Vertrauenswürdigkeit der Corona-Statistiken*

Ein zentrales Element ist die Vertrauenswürdigkeit von Statistiken als Voraussetzung ihrer Akzeptanz bei den Nutzern. Prinzip 2 der *Fundamental Principles* stipuliert das Verwenden der besten statistischen Methoden, und Prinzip 3 das Offenlegen der Informationen über Datenquellen und statistische Prozesse. Analoge Grundsätze (7, 8, 15) finden sich im *Europäischen Verhaltenskodex*. Die folgenden Ausführungen zeigen, dass bei der Produktion und Veröffentlichung der Corona-Statistiken wesentliche Standards nicht eingehalten werden.

### *Prozesse*

Die Grundsätze 7 (Anwendung einer soliden statistischen Methodik) und 8 (Verwendung geeigneter statistischer Verfahren) des *Europäischen Verhaltenskodex* zielen auf das Sicherstellen einer hohen Qualität der statistischen Produktion.

Die wichtigste Grundlage für die Corona-Statistiken sind die Informationen, die in das EMS eingespeist werden. Bei der Erstellung dieser Statistiken handelt es sich somit um die Nutzung von Verwaltungsdaten für statistische Zwecke. Über entsprechende Verfahren gibt es in Österreich viel spezifisches Knowhow, das etwa bei der *Registerbasierten Volkszählung 2011* ([Statistik Austria 2015](#)) aber auch in anderen Bereichen zur Anwendung kommt. [Eurostat \(2020\)](#) bietet einen umfangreichen Überblick über Erfahrungen mit der Verwendung von Verwaltungsdaten und stellt methodische Standards zur Verfügung.

Vieles vom Knowhow bei der Nutzung von Verwaltungsdaten sollte bei der Erstellung der Corona-Statistiken berücksichtigt werden: Mit Indikator 7.1 wird überprüft, ob die verwendete Methodik internationalen Standards, Leitlinien und vorbildlichen Praktiken Rechnung trägt. Beispielsweise sollten Standards für das Messen und Beurteilen der Qualität von administrativen Daten und den auf deren Basis berechneten Statistiken angewendet ([Eurostat 2020](#)) und damit zur Verbesserung der Datensituation genutzt werden. Die Beurteilung, inwieweit das bei der Produktion der Corona-Statistiken der Fall ist, ist mangels wichtiger Metadaten nur teilweise möglich.

Nach den Erläuterungen am *AGES-Dashboard* ([AGES 2020b](#)) erfolgt – wie bereits erwähnt – derzeit die regionale Zuordnung der laborbestätigten Fälle auf der Basis des Wohnortes der Person. Bei Fehlen eines registrierten Wohnsitzes wird der Bezirk des Aufenthaltes zum Zeitpunkt der Labordiagnose herangezogen (so z.B. bei ausländischen Touristen, Asylsuchenden, Obdachlosen, Saisonarbeiter, 24-Std.-PflegerInnen mit nicht-österreichischer Nationalität). Diese Vorgangsweise stellt eine unzulässige Vermischung von Zuordnungskriterien dar.

Zur Verbesserung der Methodik und der Wirksamkeit der angewendeten Methoden sind entsprechend Indikator 7.7 Kooperationen, auch mit der Wissenschaft, anzustreben. Die Notwendigkeit, bei allen statistischen Arbeiten zu Corona die Zusammenarbeit mit der jeweiligen Amtlichen Statistik zu suchen, wird in zahlreichen Publikationen betont (z.B. [di Genaro Splendore \(2020\)](#)). Das Fehlen von Expertinnen und Experten aus dem Bereich der amtlichen und der methodischen Statistik in der Coronavirus Task-Force und im COVID Prognose-Konsortium ist ein ernstzunehmendes Manko.

Ein erhebliches Problem für Nutzerinnen und Nutzer der Corona-Statistiken liegt darin, dass die publizierten Statistiken immer wieder korrigiert werden, und dass es zu den meisten Änderungen weder eine Dokumentation noch ausreichende Erklärungen gibt. Es wird auch nicht zwischen vorläufigen und endgültigen Statistiken unterschieden. In diesem Zusammenhang sei der Indikator 8.5 erwähnt: Revisionen der statistischen Ergebnisse sollen nach standardisierten und transparenten Verfahren erfolgen. Die zahlreichen Revisionen und die häufige Erwähnung

von Nachmeldungen und Datenbereinigungen legen nahe, dass viele der veröffentlichten Daten als vorläufige Ergebnisse anzusehen sind, ohne dass dies entsprechend dokumentiert wird. Ob auch gleichzeitig an einem für viele Analysen unverzichtbaren, gesicherten Datenbestand gearbeitet wird, ist unbekannt.

Insbesondere die Dokumentation der verwendeten Verfahren ist sehr mangelhaft. Über die angewendeten Validierungsprozesse liegen keine ausreichenden Informationen vor. Auch darüber, ob für das Editieren und Imputationen geeignete Verfahren eingesetzt werden, sind keine Informationen verfügbar.

### *Statistische Produkte*

Die Grundsätze zu den statistischen Produkten zielen darauf ab, dass die angebotenen Statistiken dem Nutzerbedarf entsprechen und in Einklang mit europäischen Qualitätsstandards stehen.

Verfahren zur Konsultation der Nutzerinnen und Nutzer und zur Überwachung der Relevanz der Statistiken sollten vorgesehen werden (Indikator 11.1). Hinsichtlich der Corona-Statistiken wurde die Kooperation bisher nur mit einem eng definierten Kreis von Nutzerinnen und Nutzern gesucht. Ein Konsultationsmechanismus mit unterschiedlichen Stakeholdern zur Sicherstellung der Relevanz der Corona-Statistiken wurde bisher nicht etabliert. Er würde das Bewusstsein der Gesundheitsstatistiker für die Bedürfnisse der Nutzer und Nutzerinnen schärfen und sie sorgfältiger und transparenter mit Änderungen im Statistik-Portfolio umgehen lassen. Derzeit stehen die möglichst aktuelle Beobachtung des Infektionsgeschehens, der Belastung des Gesundheitswesens und der Inanspruchnahme von Maßnahmen wie das Testen und die Impfungen im Mittelpunkt. Diese Schwerpunktsetzung im Interesse der politischen Entscheidungsträger und der breiten Öffentlichkeit ist verständlich. Der Aufbau eines gesicherten, kohärenten Datenbestandes für weiterführende Analysen (siehe dazu Abschnitte 3.2 und 3.3) dürfte derzeit keinen hohen Stellenwert haben.

Ob Maßnahmen eingesetzt werden, mit denen die Basisdaten, die vorläufigen Ergebnisse und die statistischen Produkte regelmäßig evaluiert und validiert werden (Indikator 12.1), ist wegen der Unvollständigkeit der Metadaten nur schwer zu beurteilen.

Mit dem Indikator 13.2 des Grundsatzes 13 (Aktualität und Pünktlichkeit) wird überprüft, ob für die Veröffentlichung der Statistiken ein Standardzeitpunkt bekanntgegeben wird. Corona-Statistiken werden von mehreren Institutionen veröffentlicht, die zu unterschiedlichen Tageszeiten an die Öffentlichkeit treten. Dass sich die involvierten Institutionen nicht auf einen einheitlichen Meldezeitpunkt für die Basiszahlen der Erkrankung zu einigen vermochten, stiftet Verwirrung. Manchmal divergieren die absoluten Werte und die angebotenen 24 Stunden Vergleiche erheblich. Die angekündigten Zeitpunkte der Publikation werden meist eingehalten.

Ein wichtiges Qualitätskriterium betrifft die Kohärenz und Vergleichbarkeit der Statistiken. Der Indikator 14.2 sieht vor, dass Statistiken, über einen ausreichenden Zeitraum betrachtet, vergleichbar sein sollen. Die Änderungen im Angebot der Corona-Statistiken, die immer wieder zu beobachten sind und sowohl die Inhalte als auch die Berichterstattung betreffen, sind für Nutzer und Nutzerinnen ein großes Problem. Der zentralen Forderung, dass die Statistiken auf der Grundlage von einheitlichen Standards in Bezug auf den Geltungsbereich, die Definitionen, die Einheiten und die Klassifikationen zu erstellen sind (Indikator 14.3), wird nicht Rechnung getragen. Es werden zwar Hinweise auf die Ursachen von Divergenzen angeboten, diese Hinweise dokumentieren aber nur, dass der Forderung nach einem koordinierten Vorgehen nicht entsprochen wird.

Gleiches gilt für den Indikator 14.4, wonach Statistiken aus verschiedenen Quellen und von verschiedener Periodizität zu vergleichen und miteinander in Einklang zu bringen sind. So werden bei den von der AGES bereitgestellten Daten – im Unterschied zu den Meldungen des Gesundheitsministeriums – die laborbestätigten Fälle, Todesfälle etc. nicht zum Meldezeitpunkt dargestellt, sondern nach dem Diagnose- bzw. Sterbedatum.

Besonders ausgeprägt sind die Unterschiede, was die publizierte Zahl der Genesenen betrifft.

Am 20. April 2021 beispielsweise berichtete das Gesundheitsministerium 560.492 Genesene, während die AGES nur einen Wert von 541.283 Genesenen auswies.

### *Publikationen*

Ein wesentlicher Indikator (6.7) dafür, ob Unparteilichkeit und Objektivität gegeben ist (Grundsatz 6), ist darin zu sehen, dass alle Nutzerinnen und Nutzer gleichzeitigen und gleichberechtigten Zugang zu statistischen Daten haben. Im Falle der Corona-Statistiken wird diesem Grundsatz nicht ausreichend Rechnung getragen. Es ist nicht transparent, wem wann welche Daten zur Verfügung gestellt werden.

Der *Verhaltenskodex* verlangt im Grundsatz 15, dass die Statistiken und die entsprechenden Metadaten in einer Weise präsentiert und archiviert werden, die eine korrekte Interpretation und aussagekräftige Vergleiche erlaubt: Nach Indikator 15.5 sollte die Dokumentation nach einem standardisierten Metadaten-Systemen erfolgen und nach Indikator 15.6 sollten die Nutzerinnen und Nutzer über die Methodik der statistischen Prozesse informiert werden. Einer der größten Mängel der derzeitigen Situation liegt darin, dass Erläuterungen an unterschiedlichen Stellen publiziert werden und es für die Nutzer und Nutzerinnen sehr mühsam ist, die verwendeten Definitionen abzurufen. Auch Informationen über die Qualität der statistischen Produkte in Bezug auf die Qualitätskriterien werden nicht angeboten, wie es nach Indikator 15.7 Standard sein sollte.

Die Bereitstellung der Ergebnisse erfolgt vor allem mit Hilfe moderner Informationstechnologien. Die Beschriftung mancher Grafiken ist freilich unzureichend und erschwert die Interpretation. So fehlen z.B. in den Grafiken „Fälle pro 100.000 und Fälle absolut nach Altersgruppen und Geschlecht Österreich“ am *AGES-Dashboard* ([AGES 2020a](#)) Angaben, auf welchen Zeitraum sich die Angaben beziehen.

Dem Gesundheitsministerium und den vielen an der Datengenerierung beteiligten Stellen ist zugute zu halten, dass die Corona-Pandemie eine neue, unbekannte und nicht-triviale Herausforderung darstellt. Zusammenfassend muss aber gesagt werden, dass die Corona-Statistiken die Standards Amtlicher Statistiken bei weitem nicht erfüllen.

## **3. Schritte zu einem Statistischen Informationssystem Corona**

Die Fragestellungen, die auf der Basis der Registerdaten behandelt werden können, beschränken sich auf das eigentliche Pandemiegeschehen. Eine Reihe von Analysen mit den öffentlich verfügbaren Daten ([Gesundheitsministerium 2021f](#)), viele aufschlussreiche Grafiken und Anmerkungen zur Datenqualität findet man bei [Neuwirth \(2021a\)](#) und [Neuwirth \(2021b\)](#).

Es gibt allerdings wichtige Fragen, die auf Basis der oben erwähnten Registerdaten nicht beantwortet werden können. Solche Fragen betreffen beispielsweise die Determinanten der Ausbreitung von COVID-19, deren Folgen für Gesellschaft und Wirtschaft, und die Auswirkungen von Maßnahmen zur Eindämmung von COVID-19. Für solche Analysen sind neben den Pandemiedaten auch weitere Gesundheitsdaten und sozioökonomische Daten notwendig.

Die vordringliche Aufgabe zur Verbesserung und Vertiefung der Corona-Berichterstattung ist es, die Register des Gesundheitsministeriums durch Integration der relevanten weiteren Daten in ein umfassendes, personenbezogenes Corona-Register (im Folgenden als *Erweitertes Statistische Corona Register* bezeichnet) zu erweitern. Für diese Aufgabe ist die Kooperation mit der Statistik Austria unumgänglich, die über viele der für das Register relevanten Daten verfügt und vieles von der benötigten, methodischen Kompetenz einbringen kann.

Der Abschnitt ist in drei Teile gegliedert: 3.1 behandelt die Integration der administrativen Register in ein *Statistisches Corona Register*, 3.2 die Erstellung des *Erweiterten Statistischen Corona Registers*, und im dritten Teil werden weitere Ausbaumöglichkeiten skizziert.

### 3.1. Von administrativen Daten zum Statistischen Corona Register

Der Übergang zu einem *Statistischen Register* ist nicht nur ein technischer Schritt, sondern es erfolgt ein Wechsel in der analytischen Ausrichtung, ein echter Paradigmenwechsel. An die Stelle der Verwaltung von Einzelfällen im Gesundheitswesen tritt das Ziel, die Voraussetzungen zu schaffen, dass einerseits der Öffentlichkeit aussagekräftige Aggregatsdaten zur Verfügung gestellt und andererseits weiterführende Analysen zum Pandemiegeschehen ermöglicht werden.

Die Eigenschaften der in den drei administrativen Registern (*Register der anzeigepflichtigen Krankheiten*, *Testregister*, *Register der COVID-19 Impffälle*) vorliegenden Daten wurden im Abschnitt 2 ausführlich diskutiert. §4a des Epidemiegesetzes behandelt zwar das Thema eines Statistik-Registers zur Pandemie, doch aus den Ausführungen in Ziffern §4 (1) und (3), auf die Bezug genommen wird, ergibt sich, dass dieses Statistische Register nur ein – bis auf Alter und Geschlecht pseudonymisierter – Auszug aus dem *Register der anzeigepflichtigen Krankheiten*, aber kein statistisches Register im Sinn der Amtlichen Statistik ist, wie es etwa von Eurostat definiert ist ([Eurostat-CROS 2019](#)):

*A Statistical register is a register created for statistical purposes, typically by integrating data from several sources, both statistical and administrative. This may be done by linking records by means of common identifiers, or by using other matching techniques.*

Voraussetzung für eine tiefere Analyse der Zusammenhänge zwischen der Pandemie und sozialen und wirtschaftlichen Faktoren ist ein Wechsel von der fallorientierten Betrachtung der Administrativen Register zu einer personenorientierten Betrachtung in den Statistischen Registern.

Für die Transformation von Administrativen Registern in ein *Statistisches Corona Register* sind die folgenden Schritte von zentraler Bedeutung:

- Zusammenführung der drei Register.
- Validierung aller administrativen Daten.
- Übergang zu einer pseudonymisierten, personenbezogenen Betrachtung.
- Datenbereinigungen durch Plausibilitätskontrollen und Ergänzung von Information. Einige Beispiele dafür wurden bereits im Abschnitt 2 als wünschenswerte Verbesserungen der Administrativen Register angeführt.
- Regelmäßige Updateprozeduren, in denen auch vorläufige Daten durch endgültige ersetzt werden.
- Systematische Dokumentation aller Schritte.

Das *Statistische Corona Register* sollte auch Basisdaten über den Verlauf der Krankheiten für Personen enthalten. Im einfachsten Fall kann die Darstellung einer individuellen Krankengeschichte durch eine Folge von sogenannten Episoden beschrieben werden. Zum Beispiel kann eine Person zunächst als Kontaktfall im *Epidemiologischen Meldesystem* aufscheinen und nach Ablauf der für den Kontaktfall wesentlichen Zeitperiode nicht mehr infiziert und damit auch nicht mehr meldepflichtig sein (Episode 1). Danach werden für die Person zwei negative Antigentests durchgeführt (Episode 2, Episode 3). Anschließend identifiziert ein wegen klinischer Kriterien behördlich verordneter PCR-Test die Person als Krankheitsfall (Episode 4), und sie bleibt bis zur Genesung ein Fall in häuslicher Pflege (Episode 5). Um einen derartigen, individuellen Krankenverlauf zu beschreiben und zu analysieren, ist ein konzeptionelles Datenmodell notwendig, das in der Situation des obigen Beispiels folgende Struktur hat:

**Personendaten:** ID, Alter, Geschlecht, Wohnort, weitere Personenmerkmale

**Episoden:**

**Episode1:** Kontaktfall (Kriterium c); Beginn Kontaktfall: dd.mm.yy;  
Ende Kontaktfall: dd.mm.yy, Details Kontaktfall

**Episode2:** Antigentest (Kriterium d); Beginn Antigentest: dd.mm.yy;  
Ende Antigentest: dd.mm.yy; Details: negativ

**Episode3:** Antigentest (Kriterium d); Beginn Antigentest: dd.mm.yy;  
Ende Antigentest: dd.mm.yy; Details: negativ

**Episode4:** PCR-Test (Kriterium c); Beginn PCR-Test: dd.mm.yy;  
Ende PCR-Test: dd.mm.yy; Details: Behördlich verordnet, positiv

**Episode5:** Krankheitsfall (Kriterium c); Beginn Krankheitsfall c): dd.mm.yy;  
Ende Krankheitsfall c): dd.mm.yy; Details: Fall in häuslicher Pflege

Die hier gegebene Darstellung ist eine einfach gehaltene Beschreibung des Krankheitsverlaufes zur Durchführung grundlegender statistischer Analysen. Will man komplexere Zusammenhänge zwischen einzelnen Episoden des Krankheitsverlaufes und den medizinischen Interventionen erfassen, sollte man Modelle für Temporale Informationssysteme in der Medizin verwenden (Combi, Keravnou-Papailiou, and Shahar 2010).

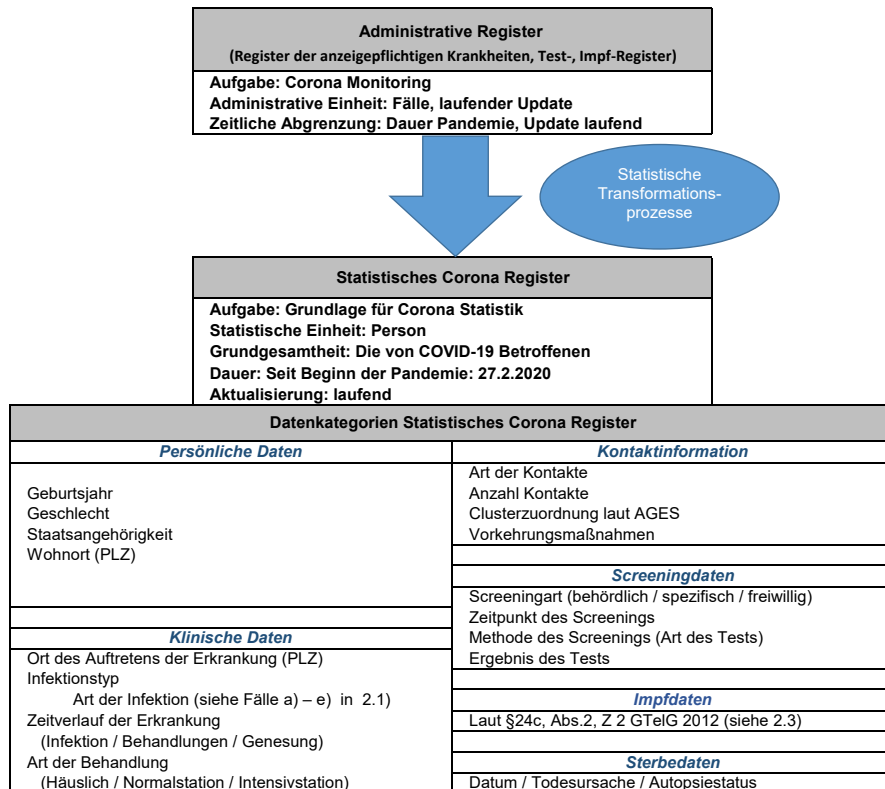


Abbildung 1: Schematische Darstellung der Zusammenhänge zwischen Administrativen Registern und einem Statistischem Corona Register (Adaptiert von: Eurostat (2019))

Aus den obigen Ausführungen ergibt sich ein Vorschlag für eine Grobstruktur des *Statistischen Corona Registers*, wie in Abbildung 1 dargestellt. Die Grundgesamtheit des *Statistischen Corona Registers* ist durch die betroffenen Personen bestimmt. Es ist zu beachten, dass dieses Register nur Informationen aus dem EMS laut Abschnitt 8.4 der Gesamtdokumentation (Gesundheitsministerium 2021c) und Daten über durchgeführte Tests der Österreichischen Gesundheitskasse zusammen führt. Möglicherweise liegt hier eine Untererfassung vor, falls Personen, die sich von der Elektronischen Österreichischen Gesundheitsakte (ELGA 2021) abgemeldet haben, nicht erfasst sind.



Eine Zusammenführung der Daten sollte innerhalb des Gesundheitsbereiches möglich sein. Welche Merkmale einbezogen werden sollen, wäre in Abwägung zwischen den Erfordernissen von statistischen und wissenschaftlichen Analysen einerseits und den durch den Datenschutz gegebenen Randbedingungen andererseits zwischen den Expertinnen und Experten der unterschiedlichen Interessensgruppen zu erarbeiten.

Dieses *Statistische Corona Register* sollte die alleinige Quelle für die Veröffentlichung von Daten sein. Publiziert werden sollten, ähnlich wie im *AGES-Dashboard* (AGES 2020a), aggregierte Daten, etwa gegliedert nach österreichischen Bezirken, Altersgruppen und Geschlecht. Bereitzustellen wären sowohl Informationen über Fälle wie auch personenbezogene Resultate. Neben dem *Statistischen Corona Register* sind natürlich noch eine Reihe weiterer Daten über das Pandemiegeschehen von Interesse, die entweder nicht für jeden Krankheitsfall erhoben wurden oder nicht erkrankte Personen betreffen.

Zur ersten Gruppe gehören Daten über das Virus und Daten über Clusterfälle. Daten zur Art des Virus wurden erst nach dem Auftreten der britischen Variante im Jänner 2021 systematisch untersucht, allerdings nicht flächendeckend bei allen Infizierten. Details findet man in AGES (2021). Daten über Clusterfälle werden in AGES (2020c) publiziert, allerdings sind diese Tabellen nicht sehr nutzerfreundlich gestaltet.

Für die Daten, welche die Gesamtbevölkerung betreffen, würde eine Darstellung, die ähnlich dem *Impf-Dashboard* aufgebaut ist und die in Abschnitt 2.2 angeführten Vorschläge aufgreift, wertvolle zusätzliche Information liefern.

Auf Bezirksebene aggregierte und publizierte Daten können in jedem Fall mit den öffentlich verfügbaren Ergebnissen aus der *Abgestimmten Erwerbsstatistik* und der *Arbeitsstättenzählung 2018*, den Bevölkerungsdaten und Daten über Gemeindetypen verknüpft werden (Statistik Austria 2018a). Eine solche Zusammenführung auf einer aggregierten Ebene würde eine Reihe von Untersuchungen über die Entwicklung der Pandemie erlauben. Ergebnisse einer derartigen Untersuchung für Oberösterreich findet man in beispielsweise Nimmervoll (2021).

### 3.2. Das Erweiterte Statistische Corona Register

Ein entscheidender Schritt, um detailliertere Auswertungen zu ermöglichen, wäre der Aufbau eines *Erweiterten Statistischen Corona Registers*, das Daten des *Statistischen Corona Registers* und soziodemographische Daten zusammenführt. Eine derartige Zusammenführung könnte unter Verwendung des *Bereichsspezifischen Personenkennzeichens Statistik* erfolgen. Dieses von der Statistik Austria gemeinsam mit der österreichischen Datenschutzbehörde entwickelte Konzept zur Durchführung der *Registerbasierten Volkszählung 2011* würde auch sicherstellen, dass diese Datenintegration nur für statistische, analytische Zwecke und nicht für administrative Aufgaben genutzt wird.

Die Kooperation mit der Statistik Austria beim Aufbau und Betrieb eines *Erweiterten Statistischen Corona Registers* kann auch sicherstellen, dass Metadaten nach einheitlichen Kriterien erstellt werden und die Qualität des Registers und seiner Inhalte gesichert ist. Für die Nutzer und Nutzerinnen müssen detaillierte Metadaten sowohl bezüglich der einzelnen Variablen als auch für alle Prozesse bereitstehen. Von der Notwendigkeit, detaillierte Metadaten zu erstellen und die verwendeten Konzepte im Detail zu analysieren, können auch bedeutende Qualitätsverbesserungen in den administrativen Registern wie dem *Register der anzeigepflichtigen Krankheiten* erwartet werden (Eurostat 2017).

Da die Prozesse zur Integration von Daten in ein solches Register stets von Metadaten geleitet werden, ist eine periodische Aktualisierung des Datenbestandes problemlos möglich, solange nicht in den Konzepten der Rohdaten Modifikationen vorgenommen werden. Der große Vorteil des *Erweiterten Statistischen Corona Registers* ist es, dass es die Nutzung eines einmal erstellten Datenbestands für unterschiedlichste Analysen erlaubt. Insbesondere könnten, wie in der Biometrie üblich, mittels statistischer Modelle (zum Beispiel Verallgemeinerte Lineare Modelle oder Methoden des Machine Learning) individuelle Risikoanalysen durchgeführt wer-

den. Es ist zu beachten, dass diese Analysen nur zum Verständnis des Pandemiegeschehens, aber nicht als Instrument für die Identifikation von Verdachtsfällen verwendet werden.

Das *Erweiterte Statistische Corona Register* integriert die Daten des *Statistischen Corona Registers*, den Datenbestand der *Registerbasierten Volkszählung 2011* und Daten der jährlich erstellten *Abgestimmten Erwerbsstatistik* zu einem Datenkörper. Die Grobstruktur ist in Abbildung 2 dargestellt.

Erweitertes Statistisches Corona Register	
Aufgabe: Soziodemographische Analysen	
Statistische Einheit: Person	
Grundgesamtheit: Erweiterte Wohnbevölkerung, Österreich	
Zeitliche Abgrenzung: Stichtagsdaten, Periodendaten	
Datenkategorien Statistisches Corona Register (siehe Abbildung 1)	
Datenkategorien Registerbasierte Volkszählung, Abgestimmte Erwerbsstatistik	
Demographische Merkmale	Bildungsmerkmale
Geburtsjahr	Höchste abgeschlossene Ausbildung
Geschlecht	Ausbildungsfeld höchste abgeschlossene Ausbildung
Staatsangehörigkeit	Merkmale der Pendelzielstatistik
Familienstand	Pendeltyp
Wohnort (PLZ)	Erwerbspendlerinnen und Erwerbspendler
Wohnort (PLZ) ein Jahr vor Stichtag	Schülerpendlerinnen, Schülerpendler und Studierende
Herkunftsland	Haushalts- und Familienmerkmale
Information zur Erwerbstätigkeit	Privathaushalte
Aktueller Erwerbsstatus	Stellung im Haushalt
Beruf	Haushaltstyp
Stellung im Beruf	Haushaltsgröße
ÖNACE des Unternehmens	Anstaltshaushalte
Unternehmensstruktur	Wohnverhältnisse
Arbeitsort (PLZ)	Nutzfläche in m <sup>2</sup> pro Bewohnerin/Bewohner
	Anzahl der Räume
	Anzahl der Räume pro Bewohnerin/Bewohner

Abbildung 2: Datenkategorien Erweitertes Statistisches Corona Register (Statistisches Corona Register, Registerbasierte Volkszählung und jährlich erstellte Abgestimmte Erwerbsstatistik)

### *Einheiten, Grundgesamtheit, zeitliche Abgrenzung des Erweiterten Corona Registers*

Die statistische Einheit ist die Person. Das Register sollte als Grundgesamtheit die gesamte Wohnbevölkerung in Österreich zu einem Stichtag umfassen. Damit kann erreicht werden, dass von der Pandemie betroffene Personen oder Personengruppen jeweils in Relation zur Gesamtbevölkerung bzw. zu wohldefinierten Gruppen der Bevölkerung analysiert werden können. Zusätzlich ist eine Erweiterung der Grundgesamtheit um Personen, die keinen Hauptwohnsitz in Österreich haben, jedoch hier erwerbstätig sind, vorzusehen. Dieses Konzept wird als *Erweiterte Wohnbevölkerung* bezeichnet. Personen mit Wohnsitz in Österreich, die im Ausland beschäftigt sind, wären gesondert zu kennzeichnen. Die letztgenannte Erweiterung ist besonders wesentlich, wie folgendes Beispiel aus der letzten Volkszählung zeigt: In der Referenzwoche, 25.10. bis 31.10.2011, waren rund 144.500 Personen in Österreich erwerbstätig, ohne zur Wohnbevölkerung gezählt zu haben ([Statistik Austria 2018a](#)).

Die zeitliche Abgrenzung des Registers ist durch den Zeitraum vom Beginn der Pandemie bis zu einem Stichtag definiert. In einem ausgebauten System wären auch andere Zeiträume frei wählbar.

Die in dem oben vorgestellten Konzept des *Erweiterten Statistischen Corona Registers* vorgesehene Verknüpfung von Stichtagsinformation mit Daten, die sich auf einen längeren Zeitraum beziehen, schafft eine Reihe von spezifischen Problemen. Einige sollen skizziert werden:

- Im Datensatz der administrativen Quellen können Personen enthalten sein, die zum Stichtag einer Volkszählung nicht in ihr registriert wurden, weil sie verstorben oder in Österreich keinen Wohnsitz (mehr) haben. Diese Personengruppe wäre besonders zu

identifizieren. Für das Merkmal Wohnsitz können entweder für spezielle Fragestellungen zusätzlich die gespeicherten Angaben zum Wohnsitz oder das Merkmal Wohnort ein Jahr vor dem Stichtag für Personen mit Wohnsitz in Österreich herangezogen werden. Wie schon erwähnt, wären Personen, die keinen Hauptwohnsitz in Österreich haben, jedoch hier erwerbstätig sind, gesondert zu identifizieren, und es wäre eine *Restregion* zu bilden, der sie zuzuordnen wären.

- Auch andere Merkmalsausprägungen können sich im Zeitraum zwischen einer Infektion und dem Stichtag geändert haben. Dies gilt insbesondere für Merkmale - wie dem Erwerbsstatus -, die direkt vom Pandemiegeschehen beeinflusst werden. Es würde viele Analysen ermöglichen, wenn aus den Rohdaten der *Abgestimmten Erwerbsstatistik* eine Variable Erwerbsstatus zum Zeitpunkt der Erkrankung konstruiert und in das System aufgenommen wird.
- Bei allen Auswertungen und Analysen muss die verbleibende Grundproblematik der Verknüpfung von stichtagsbezogenen Angaben mit zeitraumbezogenen Daten berücksichtigt werden. Der mögliche Erkenntnisgewinn aus der Verknüpfung übersteigt aber die je nach spezifischer Auswertung sehr kleinen oder größeren Beschränkungen bei weitem.

Die nächste registerbasierte Volkszählung ist in Österreich für den Stichtag 31.10.2021 geplant. Nach Abschluss der Arbeiten wird für viele Variable ein besonders gut geprüfter, umfangreicher und konsistenter Datensatz zur Verfügung stehen. Bis zum Stichtag Ende Oktober 2021 stünden Daten aus dem *Statistischen Corona Register* für 20 Monate zur Verfügung.

Da die endgültigen Ergebnisse der Volkszählung 2021 nicht in nächster Zukunft bereitstehen werden, sollte der Aufbau des *Erweiterten Statistischen Corona Registers* mit aktuellsten, vorläufigen Daten in Angriff genommen werden, um Erfahrungen zu sammeln und möglichst früh mit den Analysen beginnen zu können.

### 3.3. Weitere mögliche Ausbaustufen

Generell ist ein Data Warehouse eine für Analysezwecke optimierte, zentrale Datenbank, die Daten aus mehreren Quellen zu einem gemeinsamen, konsistenten Datenbestand zusammenführt, um Auswertungen verschiedenster Art zu ermöglichen.

Im Unterschied zu kommerziellen Data Warehouses ist es die Aufgabe eines *Statistischen Data Warehouses*, ausschließlich als Basis für die Produktion statistischer Information zu dienen. Dabei gilt es, zahlreiche Herausforderungen zu meistern: Es sind Daten aus unterschiedlich strukturierten Registern und Datenbeständen, die nicht notwendigerweise für statistische Aufgaben konzipiert und erstellt wurden, zu integrieren.

Die Analyse der Corona-relevanten Daten zeigt eine hohe Komplexität der Inhalte und Zusammenhänge, die in den derzeit veröffentlichten Statistiken und Berichten nur mangelhaft berücksichtigt wird. Davon ausgehend wurde das *Erweiterte Statistische Corona Register* konzipiert. Um ein besseres Bild über das Pandemiegeschehen zu erhalten, sollen hier einige konzeptionelle Ideen für ein Statistisches Data Warehouse Corona dargestellt werden, die auch für andere Daten aus dem Gesundheitsbereich von Interesse sind.

Eine konzeptionelle Grobstruktur der Inhalte eines ausgebauten Data Warehouses Corona ist in Abbildung 3 zu finden. Der linke Block der Abbildung nennt die Datenquellen, die im Data Warehouse zusammengeführt werden. Die wesentlichen Datenquellen sind die administrativen Register zur Pandemie des Gesundheitsministeriums und soziodemographische Daten von der Statistik Austria. Daneben spielen Daten von anderen Datenanbietern eine wesentliche Rolle. Mit der *AGES, Gesundheit Österreich (Gesundheit Österreich 2020)*, dem *Austrian Corona Panel Project (ACPP) (Austria Corona Panel Project 2020)*, oder Mobilfunkdatenanbietern sollen hier nur die wichtigsten Institutionen genannt werden. Daten der *AGES* wurden bereits in Abschnitt 3.1 als Quelle zur besseren Beschreibung der Art des Virus und zur Beschreibung

von Clusterfällen genannt. Die Gesundheit Österreich als Forschungs- und Planungsinstitut für das Gesundheitswesen ist ein zentraler Informationsanbieter und das *Austrian Corona Panel Project* (ACPP) führt seit Beginn der Pandemie laufend Untersuchungen über soziale Veränderungen durch die Pandemie durch ([Austria Corona Panel Project 2020](#)). Einige Untersuchungen zur Veränderung im Mobilitätsverhalten durch Analyse von Mobilfunkdaten liegen ebenfalls vor. Da es langfristig auch um internationale Vergleiche des Managements der Krise und der Bewältigung ihrer Folgen geht, sollte auch die Integration von internationalen Daten angedacht werden.

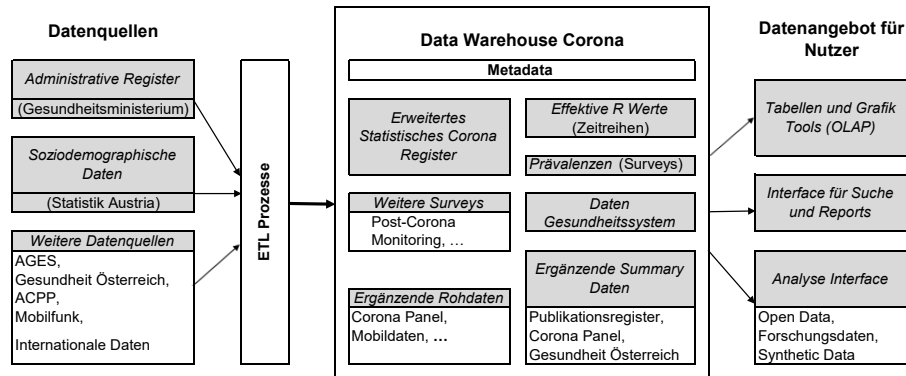


Abbildung 3: Grobstruktur eines Data Warehouses Corona

Der mittlere Block zeigt die zentralen Informationseinheiten des Data Warehouses Corona. Neben den im *Erweiterten Statistischen Corona Register* enthaltenen Datenbeständen wären besonders die aus diesem Register abgeleiteten Zeitreihen für Effektive Reproduktionsraten (R-Werte) und die Ergebnisse von Prävalenzstudien einzubeziehen. Effektive R-Werte werden derzeit von der AGES berechnet ([AGES 2020d](#)).

Für die Prävalenz gibt es drei österreichweite Studien, die Ende April, Ende Mai und Mitte November 2020 durchgeführt wurden ([Statistik Austria 2020a](#)). Weiter gab es sechs Testdurchgänge an Schulen.

Daten des Gesundheitssystems sollten hier ebenfalls in einheitlicher Form zusammengeführt werden. Der Block *Weitere Surveys* ist für die genannten Studien und für das Monitoring der sozialen und wirtschaftlichen Auswirkungen in der Zeit nach Corona vorgesehen (siehe auch [Cook and Gray \(2020\)](#)). Der Block *Weitere Rohdaten* ist für vertiefende Analysen von Interesse. Im Block *Weitere Summary Daten* sollte ein Publikationsregister enthalten sein. Die Ergebnisse des ACPP ([Austria Corona Panel Project 2021](#)) wären ebenfalls über das Data Warehouse verfügbar zu machen.

Die Überführung der Datenquellen in das System hat über geeignete Extraktions-, Transformations- und Lade-Prozeduren (ETL) zu erfolgen. Um einen homogenen Datenbestand zu erreichen, sind zahlreiche Vergleichbarkeitsprobleme zu lösen und Harmonisierungen vorzunehmen.

Für die Nutzer sind drei Schnittstellen vorgesehen: einerseits ein konfigurierbares Werkzeug zur Erstellung von Tabellen und Grafiken (OLAP Tools), ein Interface zum Suchen von Reports und Ergebnissen, die derzeit verstreut sind und nicht zentral verwaltet werden, und schließlich eine Schnittstelle für die Forschung mit Open Data ([Gesundheitsministerium 2021f](#)) für alle Interessierten, mit Forschungsdaten für registrierte Nutzer, und mit synthetische Daten, die den Anforderungen des Datenschutzes genügen.

Die Realisierung eines derartigen Projektes ist langfristig zu sehen und bedarf zunächst einmal der Entwicklung eines detaillierten Konzeptes. Über die Implementierung ist danach zu entscheiden. Wesentlich scheint aber, dass man sich von einer Zielvorstellung leiten lässt und schon rasch mit Verbesserungen in den Datengrundlagen beginnt.

## 4. Analysen mit dem Erweiterten Statistischen Corona Register

Das Infektionsgeschehen könnte auf der Grundlage der im *Erweiterten Statistischen Corona Register* verfügbaren Merkmale reich differenziert untersucht werden. Die Gliederung nach Alter, Familienstand, Geschlecht, Bildungsstand, Erwerbstatus, Staatsangehörigkeit, Geburtsland, Privathaushalt/Anstandshaushalt, Haushaltsgröße, Wohnverhältnisse (Nutzfläche/BewohnerIn), Haushaltstyp ist für alle Personen von Interesse. Bei den Erwerbstätigen kann nach Beruf, Stellung im Beruf, Wirtschaftszweig, Arbeitsort und Größe des Unternehmens unterschieden werden. Für Schülerinnen und Schüler bzw. Studentinnen und Studenten kann eine Differenzierung nach Art der Ausbildung, Pendler/Nicht-Pendler bzw. Wohnort/Arbeitsort von Interesse sein. Jede dieser Variablen kann das Risiko zu erkranken und den Verlauf der Krankheit beeinflussen.

Neben diesen soziodemographischen Gruppen sind auch noch die detaillierten Beschreibungen der Infektionszahlen nach den in 2.1 genannten Kriterien a) bis e) der Erkrankung und die Daten zur Art des Virus und die Clustertypen zu berücksichtigen.

Der gesicherte, konsistente Datenbestand mit einer großen Vielfalt an Variablen bietet der Forschung vielversprechende Möglichkeiten für weiterführende Analysen. Im einfachsten Fall sind das die Zuordnung aller infizierten Personen zu den durch die Merkmale in Abbildung 2 definierten Gruppen und die Visualisierung des Infektionsverlaufes in diesen Gruppen. In einem zweiten Schritt könnten Kennzahlen für den Verlauf der Epidemie in diesen Gruppen (z.B. Inzidenzen oder Indikatoren zur Dauer der Krankheit) berechnet und die Unterschiede zwischen den Gruppen hinsichtlich der Kennzahlen untersucht werden.

Von Interesse ist es auch, die Dynamik dieser Kennzahlen im Verlauf der Pandemie zu untersuchen. Bekanntlich war zu Beginn des Pandemiegeschehens für bestimmte Altersgruppen das Infektionsrisiko in Anstaltshaushalten höher als in Privathaushalten. Bestimmte Typen von Clustern traten in den ersten Phasen der Pandemie primär in einigen Wirtschaftszweigen und in bestimmten Berufsgruppen auf. Der Zusammenhang zwischen Clustern in Haushalten und den Wohnverhältnissen wäre auch eine Frage, die mit dem *Erweiterten Statistischen Corona Register* genauer untersucht werden könnte.

Der Einfluss der Impfungen auf die Entwicklung der Pandemie muss systematisch untersucht werden. Mit reichgegliederten Daten könnten auch wichtige Informationen über den Stand der Immunisierung gewonnen werden.

Betrachtet man den Gesamtverlauf der Pandemie, so sind neben den bisher berechneten Kennzahlen auch Analysen des zeitlichen Verlaufs von Interesse. So findet man in [Mitsa \(2010\)](#) eine systematische Darstellung von Methoden und Betrachtungsweisen für Temporal Data Mining. Mehr statistisch orientierte Arbeiten über Methoden basieren meist auf der Theorie der Modelle mit Zufallseffekten. Von Interesse wären auch Change Point Analysen, beispielsweise mit R-Paketen wie `changepoint` ([Killick, Haynes, and Eckley 2016](#)) oder `ecp` ([James, Zhang, and Matteson 2019](#)).

Eine systematische Durchleuchtung des vergangenen Infektionsgeschehens nach den hier skizzierten Dimensionen könnte den Entscheidungsträgern für zukünftige Pandemien wertvolle Unterlagen für zu setzende Prioritäten liefern. Beim Auftauchen neuer Fragestellungen können in vielen Fällen schneller adäquate Vorgangsweisen konzipiert werden. Das hohe Maß an Flexibilität ist ohne Zweifel ein entscheidender Vorteil eines solchen Registers.

Im Folgenden sollen beispielhaft drei Anwendungen angeführt werden.

### 4.1. Detailanalysen für Inzidenzen

Die Inzidenz ist eine wesentliche Kennzahl für den Vergleich von Entwicklungen und von regionalen Unterschieden. Unter Verwendung des *Erweiterten Statistischen Corona Registers* könnten verschiedene Inzidenzen bestimmt werden, die das Pandemiegeschehen in bestimmten Bevölkerungsgruppen auf regionaler Ebene detailliert beschreiben.



Zum Beispiel könnte die Differenzierung nach dem Erwerbsstatus wertvolle Einblicke gestatten. Es ist zu vermuten, dass Erwerbstätige und Lernende (Schülerinnen, Schüler und Studierende) einem von anderen Bevölkerungsgruppen verschiedenen Infektionsrisiko ausgesetzt sind. Für die Gruppe der Erwerbstätigen und Lernenden wäre deshalb eine Auswertung nach dem Wohnort und nach dem Arbeits- bzw. Schulort von Interesse.

Als Illustration wollen wir die Entwicklung der 7-Tage-Inzidenz in Wiener Neustadt betrachten, die im März 2021 für ein großes Interesse sorgte, da Wiener Neustadt ein Politischer Bezirk mit den höchsten Inzidenzwerten war. Bei der Interpretation der Inzidenzen ist zu bedenken: Mehr als die Hälfte der in Wiener Neustadt wohnhaften Erwerbstätigen pendeln in andere Politische Bezirke aus, und die Zahl der Einpendler und Einpendlerinnen aus anderen Politischen Bezirken ist größer als die Zahl der in Wiener Neustadt wohnhaften Erwerbstätigen. Bei den Schülerinnen, Schülern und Studierenden übertrifft die Zahl der Einpendler jene der in Wiener Neustadt wohnenden ([Statistik Austria 2018b](#)).

Wenn auch nur ein Bruchteil des Infektionsgeschehens dem Arbeitsort bzw. dem Ort des Schulbesuchs zuzurechnen ist, wird – wie das Beispiel Wiener Neustadt zeigt – die regionale Auswertung ausschließlich nach dem Wohnort sehr problematisch. Sie lässt kaum eine valide Aussage über das Infektionsgeschehens zu. Es müssen also sowohl Wohnort als auch Arbeitsort bzw. Schulstandort betrachtet werden. Ohne diese Differenzierung kann das Pandemiegeschehen nicht richtig verstanden werden.

## 4.2. Analysen zur Auslastung des Gesundheitssystems

Für eine Darstellung der Auslastung des Gesundheitssystems stehen, wie in Abschnitt 2.4 ausgeführt, erst seit etwa einem Jahr nach Auftreten der Pandemie plausible Daten zur Verfügung. Für die Auslastung des Gesundheitssystems ist vor allem die Information über die Dauer der Erkrankung und die Dauer der unterschiedlichen Behandlungsphasen (Häusliche Pflege, Normalstation, Intensivstation) von Interesse. Dadurch könnte die Veränderung des Krankheitsverlaufes durch die Varianten des Virus besser dargestellt werden. Insbesondere würden dann die unterschiedlichen Dauern der Phasen der Behandlung und die jeweiligen aktuellen Anteile der Infizierten in den drei Phasen der Behandlung sichtbar werden. Eine grobe, erste Abschätzung kann man zwar den Graphiken im Corona-Blog von [Neuwirth \(2021a\)](#) entnehmen, doch genauere Daten zur Beurteilung wären notwendig.

Auf der Grundlage der im *Erweiterten Statistischen Corona Register* verfügbaren soziodemographischen Merkmale könnte die Notwendigkeit der Inanspruchnahme von Spitalpflege und Intensivstation nach all den in Abschnitt 3.3 genannten Merkmalen und in Kombination dieser Merkmale analysiert werden. Es kann davon ausgegangen werden, dass die Notwendigkeit, das Gesundheitssystem stark in Anspruch zu nehmen, keineswegs nur vom Alter, sondern ebenso auch von Merkmalen wie dem Erwerbsstatus, dem Familienstand und der Haushaltsgröße bestimmt wird.

Die Anteile der Hospitalisierten könnten systematisch nach unterschiedlichen Bevölkerungsgruppen untersucht und dargestellt werden. Besonders naheliegend wäre ein Vergleich nach Altersgruppen und Haushalts- und Familienmerkmalen, insbesondere die Unterscheidung nach Privathaushalt und Anstaltshaushalt. Ebenso könnte die Schwere des Krankheitsverlaufs nicht nur nach dem Alter, sondern zusätzlich nach Merkmalen wie Erwerbsstatus, Beruf, Bildung, etc. untersucht werden.

## 4.3. Untersuchungen zur Akzeptanz von Maßnahmen in der Bevölkerung

Die Integration der Informationen aus dem *Testregister* in das *Erweiterte Statistische Corona Register* kann fundierte Einblicke in die Bereitschaft der einzelnen Bevölkerungsgruppen liefern, sich einem Test zu unterziehen. Das *Austria Corona Panel Project* (ACPP) der Universität Wien hat in seinem Blog 108: „Wer geht testen?“ eine Analyse von 1376 Befragten des Panels zur Häufigkeit der Nutzung des Testangebots vorgelegt, welche die Bedeutung einiger

soziodemographischer Merkmale für die Testbereitschaft darstellt ([Austria Corona Panel Project 2021](#)). Eine systematischere Auswertung aus den Testdaten könnte dieser Untersuchung aber deutlich mehr Gewicht verleihen. Auf Basis einer solchen Datengrundlage könnte rasch und flexibel untersucht werden, wie weit in den einzelnen Bundesländern die nach Gruppen differenzierenden Angebote zum Testen und zur Impfung von den verschiedenen soziodemographischen Gruppen genutzt werden. Derartige Informationen könnten auch dazu verwendet werden, gezielte Informationskampagnen und Werbemaßnahmen zu planen.

## 5. Vorschläge zur Berichterstattung

Die für die Berichterstattung zur Pandemie zuständigen Stellen, allen voran das Gesundheitsministerium, publizieren eine Reihe von Statistiken zum Corona-Geschehen, unter anderem das täglich aktualisierte *AGES-Dashboard* ([AGES 2020a](#)). Dabei steht die laufende Dokumentation des Infektionsgeschehens (Infektionen, Erkrankungen, Todesfällen), der Impfungen und der Auswirkungen auf das Gesundheitssystem (Hospitalisierungen, Spitalsauslastung) im Mittelpunkt des Interesses. Alle diese Statistiken sind von hoher Relevanz für Entscheidungen in Politik und öffentlicher Gesundheit, für die Einschätzung der sozioökonomischen und demographischen Auswirkungen von SARS-COV-2, sowie für die Berichterstattung in den Medien, aber auch als Basis für wissenschaftliche Analysen Corona-bezogener Fragestellungen, wie sie etwa auf der Datenplattform der Gesundheit Österreich GmbH ([Gesundheit Österreich 2020](#)) und im Corona-Blog des ACPP ([Austria Corona Panel Project 2020](#)) der Universität Wien zur Verfügung gestellt werden.

Wie im Abschnitt 2 ausgeführt, weisen die verfügbaren Informationen zum Pandemiegeschehen allerdings erhebliche Mängel auf. Dies gilt nicht nur für Österreich, sondern auch für andere Länder ([Ashofteha and Bravo 2020](#)). Zuallererst ist das Fehlen von Erklärungen wie Definitionen der veröffentlichten Statistiken, Beschreibungen der Erhebungsverfahren, Angaben zu Validierung und Verarbeitung der Daten und Genauigkeit und Vergleichbarkeit der Statistiken zu nennen, Metadaten, die für das Verständnis und die korrekte Interpretation durch Nutzerinnen und Nutzer ganz wesentlich sind. Dass eine Verbesserung der Information der Öffentlichkeit über das Pandemiegeschehen eine wesentliche Voraussetzung für mehr Vertrauen der Öffentlichkeit sowohl in die Statistiken als auch in die von den Statistiken abgeleiteten Maßnahmen ist, wurde von den Autoren mehrfach argumentiert ([Hackl 2021](#)), ([Grossmann 2021](#)).

Das im Kapitel 3.2 dargestellte *Erweiterte Statistische Corona Register* erfüllt die Voraussetzungen einer gesicherten und kontrollierten Datenbasis für unterschiedlichste Berichte und Analysen über das Pandemiegeschehen. Das *Erweiterte Statistische Corona Register* ermöglicht es, derzeit publizierte Berichte und Analysen in Inhalt und Qualität zu verbessern und weiterführende und vertiefende Berichte und Analysen zu erstellen. Das Register soll ein geeignetes und flexibles Datenangebot für die mediale Information der Öffentlichkeit und für wissenschaftliche Zwecke bereitstellen.

- Auf der Basis des *Erweiterten Statistischen Corona Registers* kann die Information für die Öffentlichkeit durch ein Informationsportal erfolgen, das einerseits einige vorgefertigte Standardtabellen anbietet, es andererseits Nutzerinnen und Nutzern ermöglicht, selbst Tabellen zu erstellen. Eine derartige Funktionalität ist Teil des Standardangebots der Statistik Austria.
- Für die Wissenschaft bietet das *Erweiterten Statistischen Corona Register* ein umfassendes Datenangebot, mit dem Analysen für eine weite Palette von Themen möglich sind.

Wie schon früher betont, ist die Qualität der für die Öffentlichkeit berichteten Statistiken, insbesondere die Konsistenz der verwendeten Merkmale im Zeitverlauf und die Verfügbarkeit

von Metainformation, von entscheidender Bedeutung für ihre Akzeptanz und das Vertrauen bei Nutzerinnen und Nutzern. Die Vielzahl von Anwendungen und Visualisierungen, die auf der Website von Open Data ([Gesundheitsministerium 2021f](#)) abrufbar sind, demonstrieren das Interesse und die Nachfrage nach Antworten auf unterschiedlichste Fragestellungen zum Pandemiegeschehen. Diese Ergebnisse demonstrieren aber auch die Problematik der Verwendung von Pandemiedaten, deren genaue Definitionen und Erhebungsverfahren nicht dokumentiert sind: Bedenkt man die in den früheren Abschnitten angesprochenen Datenprobleme, wird klar, dass die Interpretation der berichteten Ergebnisse meist nicht so einfach ist, wie es manche Überschrift erscheinen lässt. Die Verfügbarkeit des *Erweiterten Statistischen Corona Registers* eröffnet eine neue Dimension von wissenschaftlichen Analysen hinsichtlich sowohl der Themen als auch der Qualität der Ergebnisse.

Die Realisierung des *Erweiterten Statistischen Corona Registers* erfordert neben der umfassenden Planung die Kooperation mit allen Eignern von relevanten Daten und das Implementieren der entsprechenden Software. Dieser Prozess wird einige Zeit in Anspruch nehmen und wird in mehreren Stufen ablaufen. Dabei soll die Priorität der Generierung von Berichten und Analysen mit besonderer Wichtigkeit für die Eindämmung der Pandemie sichergestellt sein.

## 6. Zusammenfassung und Ausblick

Der Ausbruch von COVID-19 hat einen großen Bedarf an Corona-Statistiken für Planung und Beurteilung von Maßnahmen ausgelöst, aber auch massive Mängel in den verfügbaren Daten und veröffentlichten Corona-Statistiken offenbart, die gravierende, negative Auswirkungen haben. Diese Erfahrungen bieten die Möglichkeit, für die derzeitigen und vor allem für künftige neue Herausforderungen wichtige Schlüsse zu ziehen. Die wichtigsten Lehren sind:

- Auf internationaler Ebene herrscht Übereinstimmung darüber ([Ljones 2020](#)), dass die statistischen Institutionen bei Auftreten neuer Probleme schneller und proaktiv reagieren müsste.
- Eine Gesundheitsberichterstattung ohne Berücksichtigung von sozialstatistischen Daten ist unvollständig ([Royal Statistical Society 2021](#)).
- Neue politische Maßnahmen zur Bewältigung von Problemen wie jenen der Pandemie müssen durch ein statistisches Monitoring begleitet werden.
- Professionalität in den datengenerierenden Prozessen ist sicherzustellen. Aus den Erfahrungen mit der Pandemie sollte gelernt werden, wie Verwaltungsprozesse besser organisiert werden können.
- Es muss eine zentrale Informationsstelle geben, die das Einhalten des Grundsatzes des gleichen Zugangs zu den Daten für alle garantiert. Auch die anderen Grundsätze Amtlicher Statistik sind zu beachten.
- Die Statistik muss eine zentrale Rolle in der ganzheitlichen Betrachtung der Probleme haben und in den Gremien eingebunden sein.
- Die Erfahrung der Statistik in der Dokumentation von Metadaten und im Umgang mit Änderungen in den Spezifikationen ist zu nutzen.
- Es sind nicht nur Daten notwendig, sondern es müssen auch geeignete Analysemethoden zur Verfügung stehen. Die methodische Statistik verfügt über einen reichen Fundus von Ansätzen.
- Die Amtliche Statistik und die Statistischen Gesellschaften haben einen erzieherischen Auftrag zu erfüllen, um die *Statistical Literacy* zu verbessern ([Deutsche Arbeitsgemeinschaft Statistik 2021](#)), ([Royal Statistical Society 2021](#)).

- Am Beispiel der Nutzung von Verwaltungsdaten zur Gewinnung von statistischer Information über die Pandemie sollte in der Öffentlichkeit das Risiko für den Datenschutz bei derartigen Anwendungen gegenüber den Mehrkosten bei einer Nichtnutzung solcher Daten diskutiert werden.

Die dringend notwendige Verbesserung und Vertiefung der Corona-Berichterstattung sollen in einem ersten Schritt durch Zusammenführen der Register des Gesundheitsministeriums in ein umfassendes, personenbezogenes Corona-Register erreicht werden. In der Folge würde die Ausweitung des Corona-Registers durch Integration von sozioökonomischen, weiteren Gesundheits- und anderen Daten in ein *Erweitertes Statistische Corona Register* erlauben, neben einer vertieften Beschreibung auch weitergehende Analysen des Pandemiegeschehens durchzuführen. Für die Schaffung eines solchen Registers sind verschiedene organisatorische Lösungen und Kooperationsformen denkbar.

Für die Implementierung von erweiterten Registern ist aber jedenfalls die Zusammenarbeit der Gesundheitsexperten mit der Bundesanstalt Statistik Austria unumgänglich, die über ein reichhaltiges Inventar an Methoden zur Zusammenführung und Nutzung von verschiedenen Datenquellen verfügt. Statistik Austria ist vor allem Inhaber vieler Daten, die in das vorgeschlagene *Erweiterte Statistische Corona Register* aufzunehmen wären. Statistik Austria ist zudem gut mit Europäischen und internationalen Organisationen vernetzt, die an einem konzertierten Plan für eine internationale statistische Aufarbeitung der Daten zur Pandemie arbeiten (Fu *et al.* 2020).

Die Realisierung des *Erweiterten Statistischen Corona Registers* würde einen Meilenstein in der Berichterstattung über das Pandemiegeschehen bedeuten. Auf der Basis des *Erweiterten Statistischen Corona Registers* können – unter Einhaltung der in der Amtlichen Statistik üblichen Standards – Berichte und Analysen generiert werden, die in qualitativer Hinsicht die großen Anforderungen erfüllen, die angesichts der hohen Verantwortung aller mit der Pandemie befassten Personen erwartet werden müssen.

## Danksagung

Die Autoren danken dem Herausgeber für das unkomplizierte Prozedere und ihm und drei anonymen Gutachtern für die Vielzahl von hilfreichen Kommentaren und Vorschlägen zur Verbesserung des Manuskripts. Frau Mag. Viktoria Haidinger hat uns wertvolle Hilfe bei der Interpretation der einschlägigen Gesetzestexte geleistet.

## Literatur

Adhikari S, Pantaleo N, Feldman J, Ogedegbe O, Thorpe L, Troxel A (2020). “Assessment of Community-Level Disparities in Coronavirus Disease 2019 (COVID-19) Infections and Deaths in Large US Metropolitan Areas.” doi:10.1001/jamanetworkopen.2020.16938. Accessed on 2021-07-03.

AGES (2020a). “AGES Dashboard COVID19.” Accessed on 2021-07-03, URL <https://covid19-dashboard.ages.at/>.

AGES (2020b). “AGES Erläuterungen.” Accessed on 2021-07-03, URL <https://covid19-dashboard.ages.at/basisinfo.html>.

AGES (2020c). “Epidemiologische Abklärung Covid 19.” Accessed on 2021-07-03, URL <https://www.ages.at/themen/krankheitserreger/coronavirus/epidemiologische-abklaerung-covid-19/#>.

- AGES (2020d). “Epidemiologische Parameter des COVID 19 Ausbruches.” Accessed on 2021-07-03, URL <https://www.ages.at/en/wissen-aktuell/publikationen/epidemiologische-parameter-des-covid19-ausbruchs-oesterreich-20202021/>.
- AGES (2021). “SARS-CoV-2-Varianten in Österreich.” Accessed on 2021-07-03, URL <https://www.ages.at/themen/krankheitserreger/coronavirus/sars-cov-2-varianten-in-oesterreich/>.
- Ashofteha A, Bravo J (2020). “A Study on the Quality of Novel Coronavirus (COVID-19) Official Datasets.” *Statistical Journal of the IAOS*, **36**(2), 291–301. doi:10.3233/SJI-200674. Accessed on 2021-07-03.
- Austria Corona Panel Project (2020). “Panelumfrage zur Coronakrise.” Accessed on 2021-07-03, URL <https://viecer.univie.ac.at/coronapanel/>.
- Austria Corona Panel Project (2021). “Blog 108 Wer geht Testen?” Accessed on 2021-07-03, URL <https://viecer.univie.ac.at/corona-blog/corona-blog-beitraege/blog108>.
- BGBI 111/2012 idF BGBI I 2021/34 (2021). *Gesundheitstelematikgesetz*. Accessed on 2021-07-03, URL <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20008120>.
- BGBI 1950/186 idF BGBI I 2021/64 (2021). *Epidemiegesetz*. Accessed on 2021-07-03, URL <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10010265>.
- Combi C, Keravnou-Papailiou E, Shahar Y (2010). *Temporal Information Systems in Medicine*. Springer. ISBN 978-1-4419-66542-4.
- Cook L, Gray A (2020). “Official Statistics in the Search for Solutions for Living with COVID-19 and Its Consequences.” *Statistical Journal of the IAOS*, pp. 253-278, 2020, **36**(2), 253 – 278. Accessed on 2021-07-03, URL <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji200671>.
- Deutsche Arbeitsgemeinschaft Statistik (2021). “Daten und Statistik als Grundlage für Entscheidungen: Eine Diskussion am Beispiel der Corona-Pandemie.” Accessed on 2021-07-03, URL [https://www.dagstat.de/fileadmin/dagstat/documents/DAGStat\\_Covid\\_Stellungnahme.pdf](https://www.dagstat.de/fileadmin/dagstat/documents/DAGStat_Covid_Stellungnahme.pdf).
- di Gennaro Splendore L (2020). “COVID-19 Unprecedented Situation, Unprecedented Official Statistics.” *Journal of Official Statistics*, **36**(2), 229 – 235. Accessed on 2021-07-03, URL <https://doi.org/10.2478/jos-2020-0012>.
- ELGA (2021). “Elektronische Gesundheitsakte.” Accessed on 2021-07-03, URL <https://www.elga.gv.at/elga-die-elektronische-gesundheitsakte/elga-im-ueberblick/>.
- Eurostat (2017). *Verhaltenskodex für Europäische Statistiken*. Eurostat. Accessed on 2021-07-03, URL <https://ec.europa.eu/eurostat/documents/4031688/9394019/KS-02-18-142-DE-N.pdf/27ca19ca-e349-45f8-bbd4-4d78a33601ae>.
- Eurostat (2019). “Quality Guidelines on Frames for Social Statistics.” Accessed on 2021-07-03, URL [https://ec.europa.eu/eurostat/cros/content/quality-guidelines-frames-social-statistics-qgffs-revised-final-version-151\\_en](https://ec.europa.eu/eurostat/cros/content/quality-guidelines-frames-social-statistics-qgffs-revised-final-version-151_en).
- Eurostat (2020). *ESS Vision 2020 ADMIN (Administrative Data Sources)*. Eurostat. Accessed on 2021-07-03, URL [https://ec.europa.eu/eurostat/cros/content/ess-vision-2020-admin-administrative-data-sources\\_en](https://ec.europa.eu/eurostat/cros/content/ess-vision-2020-admin-administrative-data-sources_en).



- Eurostat-CROS (2019). “Statistical Register.” Accessed on 2021-07-03, URL [https://ec.europa.eu/eurostat/cros/content/statistical-register\\_en](https://ec.europa.eu/eurostat/cros/content/statistical-register_en).
- Fu H, Hereward M, MacFeely S, Me A, Wilmoth J (2020). “How COVID-19 is Changing the World: A Statistical Perspective from the Committee for the Coordination of Statistical Activities.” *Journal of Official Statistics*, **36**(4), 851–860. doi:10.3233/SJI-200759. Accessed on 2021-07-03.
- Gesundheit Österreich (2020). “Datenplattform COVID.” Accessed on 2021-07-03, URL <https://datenplattform-covid.goeg.at/EMS>.
- Gesundheitsministerium (2021a). “Behördliche Vorgangsweise bei SARS-CoV-2 Kontaktpersonen: Kontaktpersonennachverfolgung (Stand: 25.06.2021).” Accessed on 2021-07-03, URL <https://www.sozialministerium.at/Informationen-zum-Coronavirus/Coronavirus---Fachinformationen.html>.
- Gesundheitsministerium (2021b). “Corona-Schutzimpfung in Österreich.” Accessed on 2021-07-03, URL <https://info.gesundheitsministerium.gv.at/>.
- Gesundheitsministerium (2021c). “COVID-19-Pandemie. Bestandsaufnahme, Handlungsrahmen.” Accessed on 2021-07-03, URL [https://www.sozialministerium.at/Informationen-zum-Coronavirus/Neuartiges-Coronavirus-\(2019-nCov\).html](https://www.sozialministerium.at/Informationen-zum-Coronavirus/Neuartiges-Coronavirus-(2019-nCov).html).
- Gesundheitsministerium (2021d). “Falldefinitionen.” Accessed on 2021-07-03, URL <https://www.sozialministerium.at/Informationen-zum-Coronavirus/Coronavirus---Fachinformationen.html>.
- Gesundheitsministerium (2021e). “Österreichische Teststrategie.” Accessed on 2021-07-03, URL <https://www.sozialministerium.at/Informationen-zum-Coronavirus/Coronavirus---Fachinformationen.html>.
- Gesundheitsministerium (2021f). “Österreichisches COVID-19 Open Data Informationsportal.” Accessed on 2021-07-03, URL <https://www.data.gv.at/covid-19/>.
- Grossmann W (2021). “Wissen wir denn genug über dieses Virus?” Accessed on 2021-07-03, URL <https://www.diepresse.com/5954875/wissen-wir-denn-genug-uber-dieses-virus>.
- Hackl P (2021). “Pandemiestatistiken: Sünden von gestern?” Accessed on 2021-07-03, URL <https://www.diepresse.com/5932316/pandemiestatistiken-sunden-von-gestern>.
- James NA, Zhang W, Matteson DS (2019). “ecp: An R Package for Nonparametric Multiple Change Point Analysis of Multivariate Data.” R package version 3.1.2, URL <https://cran.r-project.org/package=ecp>.
- Killick R, Haynes K, Eckley IA (2016). *changeoint: An R package for Changepoint Analysis*. R package version 2.2.2, URL <https://CRAN.R-project.org/package=changeoint>.
- Kissam E (2020). “The Impact of the COVID-19 Pandemic on California Farmworkers: Better Local Data Collection and Reporting Will Improve Strategic Response.” *Journal of Official Statistics*, **36**(4), 867–898. doi:10.3233/SJI-200763. Accessed on 2021-07-03.
- Ljones O (2020). “Pandemic and Official Statistics; some Comments on Recent COVID-19 Experiences.” *Statistical Journal of the IAOS*, **36**(2), 285–290. doi:10.3233/SJI-200672. Accessed on 2021-07-03.
- Mitsa T (2010). *Temporal Data Mining*. CRC Press. ISBN 978-1-4200-8976-9.

- Neuwirth E (2021a). “Erich Neuwirths COVID-19-Analysen.” Accessed on 2021-07-03, URL <https://just-the-covid-facts.neuwirth.priv.at/2021/04/26/covid-19-oesterreich-aktuelle-daten/>.
- Neuwirth E (2021b). “Erich Neuwirths COVID-19-Analysen.” Accessed on 2021-07-03, URL <http://www.neuwirth.priv.at/covid-test/>.
- Nimmervoll L (2021). “Corona-Studie: Im ländlichen Raum deutlich höhere Fallzahlen.” Accessed on 2021-07-03, URL <https://www.derstandard.at/story/2000124112426/corona-studie-im-laendlichen-raum-deutlich-hoehere-fallzahlen>.
- Royal Statistical Society (2021). “Our work on the Covid-19 pandemic.” Accessed on 2021-07-03, URL <https://rss.org.uk/policy-campaigns/policy/covid-19-task-force/>.
- Statistik Austria (2015). “Registerzählung 2011, Standard-Dokumentation Meta-informationen, Bearbeitungsstand: 27.04.2015.” Accessed on 2021-07-03, URL [https://www.statistik.at/web\\_de/dokumentationen/menschen\\_und\\_gesellschaft/Bevoelkerung/index.html](https://www.statistik.at/web_de/dokumentationen/menschen_und_gesellschaft/Bevoelkerung/index.html).
- Statistik Austria (2018a). “Abgestimmten Erwerbsstatistik und Erwerbsstatistik der Registerzählung 2011, Standard-Dokumentation Metainformationen; Bearbeitungsstand: 10.09.2018.” Accessed on 2021-07-03, URL [https://www.statistik.at/web\\_de/dokumentationen/menschen\\_und\\_gesellschaft/Bevoelkerung/index.html](https://www.statistik.at/web_de/dokumentationen/menschen_und_gesellschaft/Bevoelkerung/index.html).
- Statistik Austria (2018b). “Blick auf die Gemeinde, Wiener Neustadt (Stadt); Abgestimmte Erwerbsstatistik 2018; Erwerbs- und Schulpendler/-innen nach Entfernungskategorie.” Accessed on 2021-07-03, URL <https://www.statistik.at/blickgem/ae2/g30401.pdf>.
- Statistik Austria (2020a). “COVID-19 Prävalenzstudien.” Accessed on 2021-07-03, URL [https://www.statistik.at/web\\_de/statistiken/menschen\\_und\\_gesellschaft/gesundheit/covid19/index.html](https://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/gesundheit/covid19/index.html).
- Statistik Austria (2020b). “statistik / coronaDAT.” Accessed on 2021-07-03, URL <https://github.com/statistik/coronaDAT>.
- United Nations (2013). *Fundamental Principle of Official Statistics*. United Nations. Accessed on 2021-07-03, URL <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>.

**Affiliation:**

Wilfried Grossmann  
Facultät für Informatik  
Universität Wien  
1090 Wien, Austria  
E-mail: [wilfried.grossmann@univie.ac.at](mailto:wilfried.grossmann@univie.ac.at)

Peter Hackl  
Löfflerweg 9  
7100 Neusiedl/See, Austria  
E-mail: [peter.g.hackl@gmail.com](mailto:peter.g.hackl@gmail.com)

Josef Richter  
Windschutzstr. 14  
1140 Wien, Austria  
E-mail: [josef.richter@aon.at](mailto:josef.richter@aon.at)