

## SOEPpapers

on Multidisciplinary Panel Data Research

# Ein interdisziplinäres Studienkonzept zur Dynamik von COVID-19 auf der Basis der prospektiv erhobenen Daten der Kohorten des Sozio-oekonomischen Panels (SOEP)

Ulrich Rendtel, Reinhard Meister, Jan Goebel, Antje Gößwald, Markus G. Grabka, Jens Hoebel, Martin Schlaud,  
Carsten Schroeder, Gert G. Wagner, Sabine Zinn, Stefan Liebig und Thomas Lampert

## **SOEPPapers on Multidisciplinary Panel Data Research** at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPPapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPPapers are available at <http://www.diw.de/soeppapers>

### **Editors:**

Jan **Goebel** (Spatial Economics)  
Stefan **Liebig** (Sociology)  
David **Richter** (Psychology)  
Carsten **Schröder** (Public Economics)  
Jürgen **Schupp** (Sociology)  
Sabine **Zinn** (Statistics)

Conchita **D'Ambrosio** (Public Economics, DIW Research Fellow)  
Denis **Gerstorff** (Psychology, DIW Research Fellow)  
Katharina **Wrohlich** (Gender Economics)  
Martin **Kroh** (Political Science, Survey Methodology)  
Jörg-Peter **Schräpler** (Survey Methodology, DIW Research Fellow)  
Thomas **Siedler** (Empirical Economics, DIW Research Fellow)  
C. Katharina **Spieß** (Education and Family Economics)  
Gert G. **Wagner** (Social Sciences)

ISSN: 1864-6689 (online)

German Socio-Economic Panel (SOEP)  
DIW Berlin  
Mohrenstrasse 58  
10117 Berlin, Germany

Contact: [soeppapers@diw.de](mailto:soeppapers@diw.de)



# Ein interdisziplinäres Studienkonzept zur Dynamik von COVID-19 auf der Basis der prospektiv erhobenen Daten der Kohorten des Sozio-oekonomischen Panels (SOEP)\*

Ulrich Rendtel<sup>1</sup>, Reinhard Meister<sup>2</sup>, Jan Goebel<sup>3</sup>, Antje Gößwald<sup>4</sup>, Markus G. Grabka<sup>3</sup>, Jens Hoebel<sup>4</sup>, Martin Schlaud<sup>4</sup>, Carsten Schroeder<sup>1, 3</sup>, Gert G. Wagner<sup>3, 5</sup>, Sabine Zinn<sup>3</sup>, Stefan Liebig<sup>1, 3</sup> und Thomas Lampert<sup>4</sup>

Berlin, im Juli 2020

- 1 Freie Universität Berlin
- 2 Beuth Hochschule, Berlin
- 3 Sozio-oekonomisches Panel (SOEP), Berlin
- 4 Robert Koch-Institut, Abteilung für Epidemiologie und Gesundheitsmonitoring, Berlin
- 5 MPI für Bildungsforschung, Berlin

## Abstract

Die Weltgesundheitsorganisation (WHO) hat jüngst Richtlinien für Studien, die Basisdaten für gesundheitspolitische Entscheidungen im Pandemiefall liefern können, veröffentlicht. Durch die Zusammenarbeit zwischen unterschiedlichen Disziplinen kann es gelingen, evidenzbasierte Entscheidungen durch qualitativ hochwertige, längsschnittliche Befragungsdaten aus den Sozial- und Wirtschaftswissenschaften in Verbindung mit wiederholten Labortests zum Virus- und Antikörpernachweis sowie Erhebungen des Gesundheitsstatus mit soliden statistischen Analysen zu unterfüttern. Der hier skizzierte Studienplan und das erhebungstechnische Konzept zur Erforschung des Umfangs und der Dynamik von SARS-CoV-2-Infektionen und COVID-19-Erkrankungen unter Nutzung der haushaltsbasierten Multi-Kohortenstudie Sozio-oekonomisches Panel (SOEP) erfüllen die Anforderungen der WHO zur Messung der Prävalenz von COVID-19 in vollem Umfang und bieten darüber hinaus weitreichende Analysemöglichkeiten zu gesundheitlichen und sozio-ökonomischen Auswirkungen der Pandemie.

---

\* Diese Projektidee wurde von einer Reihe von Forschenden aus ganz unterschiedlichen Bereichen entwickelt und wird entsprechend unterschützt. Die Beteiligten kommen aus unterschiedlichen Fächern: Epidemiologie, Umfrageforschung, medizinische Biometrie, Pharmakologie, Sozial-, Verhaltens- und Wirtschaftswissenschaften. Unter den Unterstützern treffen sich aber auch unterschiedliche Forschungsinstitutionen: staatliche Forschungsinstitute, unabhängige Forschungsinstitute, Universitäten und Unternehmen. Einige externe Kollegen, die bei der Konzeption des Projekts mitgeholfen haben, sind Ulrich Mansmann, Biometrie und Statistik LMU München; Michael Hölscher, Tropenmedizin und Virologie München; Gerd Antes, Cochrane Freiburg; Timo Schmid, Statistik FU Berlin; Harald Binder und Martin Wolkewitz, Med. Biometrie und Epidemiologie, Uni Freiburg; Jan Beyersmann, Mathematische Biometrie; Statistik Univ. Ulm; Brigitte Keller-Stanislawski, PEI Langen; Iris Pigeot, Hajo Zeeb und Wolfgang Ahrens (alle Bips Bremen [Leibniz Gemeinschaft], NAKO); sowie die Mitveranstalterinnen und -veranstalter des Kolloquiums „Berliner Kolloquium Statistische Methoden in der empirischen Forschung“ aus Berlin Angelika Schaffrath-Rosario, RKI; Christine Müller-Graf, BfR; Johannes Rauh, IQTIG Vitaly Belik Statistik und Veterinär-Epidemiologie FU Berlin; Geraldine Rauch Medizinische Biometrie Charité; Ulrike Grömping Beuth Hochschule für Technik Berlin; Patrick Schlömer, Bayer AG; Friedrich Leverkus, Pfizer; Thomas Bregenzer, Cytel, und Peter Jähni, PJ Statistics. Innerhalb des SOEP haben Hannes Kröger und Jürgen Schupp wichtige Hinweise zu diesem Papier gegeben.

## Einleitung

Die durch das neue Coronavirus (SARS-CoV-2) ausgelöste Pandemie stellt weltweit ein akutes Gesundheitsproblem dar, das uns noch auf absehbare Zeit beschäftigen wird. Derzeit kann allein unser Immunsystem eine Infektion so bekämpfen, dass diese gut überstanden wird und im besten Fall und zumindest für einen bestimmten Zeitraum Immunität eintritt. Umso wichtiger sind verlässliche Schätzungen der Ausbreitung und Dynamik von SARS-CoV-2-Infektionen sowie der Coronavirus-Disease-2019 (COVID-19), der durch das Virus ausgelösten Krankheit.

Die aktuellen Maßnahmen zur Eindämmung der COVID-19-Pandemie wurden in Deutschland meist auf Basis der Zahl gemeldeter COVID-19-Fälle und der Zahl der mit dieser Erkrankung Verstorbenen getroffen. Dies gilt auch für die Regelungen zur Lockerung der landesweit verfügbaren Beschränkungen, die sich an der Anzahl neu gemeldeter Fälle pro 100 Tsd. Einwohner pro Landkreis orientieren.

Die gemeldeten Fälle bilden aber nur einen, vermutlich kleinen Teil des Infektionsgeschehens ab, da die Infektion oftmals ohne Symptome verläuft. Asymptomatisch verlaufende Infektionen bleiben oftmals unerkannt, da in der Regel nur bei vorhandenen Symptomen oder bei nachgewiesenem Kontakt zu einer infizierten Person auf die Infektion getestet wird. Für die Ermittlung der Risiken für die Bevölkerung ist eine auf einer Zufallsstichprobe basierende Erfassung der gesamten Bevölkerung notwendig, da nur so die Dunkelziffer der Infektionen ermittelt werden und so das Ansteckungsgeschehen valide und unverzerrt für die gesamte in Deutschland lebende Wohnbevölkerung empirisch analysiert werden kann. Dabei ist es wichtig, diese Risiken demographisch und sozial zu differenzieren, um die Bevölkerungsgruppen zu identifizieren, die am stärksten betroffen bzw. gefährdet sind. Dies wird auch in der jüngst veröffentlichten Richtlinie<sup>1</sup> der WHO (2020) betont.

Ziele entsprechender epidemiologisch motivierter bevölkerungsrepräsentativer Studien sind aktuell:

- a) Ermittlung der aktuellen Anzahl und des Anteils der mit SARS-CoV-2 infizierten Personen: Da bislang nur sehr selektiv in der Bevölkerung auf SARS-CoV-2 (mit PCR Rachenabstrich) getestet wird, u.a. aufgrund des wahrscheinlich relativ hohen Anteils asymptomatischer Infektionen, ist die Anzahl der tatsächlich mit dem Corona Virus infizierten Personen in der Bevölkerung nicht bekannt. Weiterhin ändert sich die Anzahl der Infizierten ständig als Folge von Ansteckungen bzw. Maßnahmen des sogenannten Social-Distancing. Nach ersten Ergebnissen der weit beachteten Heinsberg-Studie (Streeck et al. 2020) könnte die Anzahl der Infizierten um den Faktor 5 höher als die Anzahl der gemeldeten Fälle liegen. Allerdings ist diese Studie nicht repräsentativ für sämtliche Regionen Deutschlands, sondern nur aussagekräftig für den Kreis Heinsberg bzw. den Hot Spot in der Gemeinde Gangelt Ende März 2020.
- b) Messung des Effekts von Lockerungsmaßnahmen von verfügbaren Beschränkungen auf das Ansteckungsgeschehen: Da die Lockerungsmaßnahmen regional und temporal unterschiedlich gehandhabt werden, ist die zeitliche Entwicklung lokaler Ansteckungsraten von Interesse. Dies spricht dafür, neben bundesweit repräsentativen Studien auch solche durchzuführen, die Aussagen über die Situation in bestimmten Regionen, Kreisen und Gemeinden erlauben.

---

<sup>1</sup> "The seroepidemiological investigation for COVID-19 virus infection is a population-based, age-stratified prospective study" (S. 5): (WHO), Population-based age-stratified seroepidemiological investigation protocol for COVID-19 virus infection, Version 1.1, March 17 2020: <https://www.who.int/publications-detail/population-based-age-stratified-seroepidemiological-investigation-protocol-for-covid-19-virus-infection>.

- c) Rekonstruktion der sozialen Bedingungen des Infektionsgeschehens: Bei den Ansteckungen im häuslichen Umfeld wurde zunächst davon ausgegangen, dass wahrscheinlich alle Personen innerhalb eines Haushalts mit dem Virus infiziert werden, wenn ein Haushaltsmitglied am Corona Virus erkrankte. Diese Vermutung wird durch die Heinsberg-Studie nicht bestätigt und eröffnet neue Spekulationen über den Ansteckungsprozess. Insbesondere die Rolle der Kinder im Haushalt scheint hier wichtig zu sein. Die Einstellung wie auch die schrittweise Wiedereröffnung des Schul- und KITA-Betriebs steht hiermit in einem engen Zusammenhang.
- d) Infektionsgeschehen in Anstalten: Die Verbreitung des Corona Virus im Anstaltsbereich (Alten- und Pflegeheime) stellt ein besonderes Problem dar, da ältere Menschen zu den COVID-19-Hochrisikogruppen zählen. Allerdings wird der Anstaltsbereich traditionell in bevölkerungsrepräsentativen Surveys nicht in Umfragen erhoben.
- e) Ermittlung der Risiken entlang der kompletten Infektions- und Krankheitsverläufe: Auf Basis der bisherigen Daten sind lediglich die Verläufe bzw. Statussequenzen ab Einlieferung ins Krankenhaus dokumentiert. Damit können nur für einen Ausschnitt der kompletten Statussequenz die Risiken in Abhängigkeit zu individuellen oder strukturellen Merkmalen eingeschätzt werden. Um das Infektionsgeschehen vollständig beurteilen und entsprechende Maßnahmen ergreifen zu können, ist aber die Berücksichtigung der vollständigen Sequenz „ohne Ansteckung“, „mit Ansteckung“, „mit Symptomen“, „mit Krankenhausaufenthalt“, „mit Heilung“ oder „an COVID-19 verstorben“ notwendig.

Bisher sind entsprechende Studien, auch für bislang Nichterkrankte, nur an Brennpunkten (Hot Spots) der COVID-19-Erkrankung durchgeführt worden, so in Gangelt (Heinsberg Studie, vgl. Streeck et al. 2020) oder – laufend – in München (vgl. Radon et al. 2020) und Kupferzell im Hohenlohekreis (vgl. Lampert und Santos-Hövenner 2020). Die Ergebnisse dieser Studien sind strenggenommen nur für den jeweiligen Brennpunkt aussagekräftig. Gleiches gilt für geplante Städtestudien z.B. in Halle<sup>2</sup> oder Magdeburg. Eine ähnlich begrenzte Verallgemeinerbarkeit haben Studien, die zwar durchaus in der Fläche durchgeführt werden – wie etwa die aktuelle Erhebung der Nako-Gesundheitsstudie –, aber aufgrund einer geringen regionalen Abdeckung bei der Stichprobenerstellung sehr wahrscheinlich zu einer hohen Selektivität mit Blick auf die Grundgesamtheit (d.h. der allgemeinen Wohnbevölkerung in Deutschland) führt. So ist die Stichprobe der NAKO-Gesundheitsstudie auf acht Regionen begrenzt<sup>3</sup> und die Testung erfolgt vor Ort in den 18 Studienzentren durch medizinisches Personal, was eine zusätzliche Hürde für die Teilnahme darstellt und zu Selbst-Selektivität führen dürfte.<sup>4</sup> Ergänzend werden Studien an Spezialpopulationen durchgeführt, z.B. an Blutspendern, deren Ergebnisse nur begrenzt verallgemeinert werden können. Mit diesem Diskussionspapier möchten wir erläutern, welche weitergehenden Möglichkeiten die Nutzung des Sozio-oekonomischen Panels (SOEP) als bevölkerungsrepräsentativer Multi-Kohorten-Stichprobe in Verbindung mit ergänzenden regelmäßigen diagnostischen Tests

---

<sup>2</sup> Für die Stadt-Studie „Evaluation der Testqualität von Antikörper-Nachweisen (Serologie) und längsschnittliche Erhebung des Antikörperstatus bezüglich SARS-CoV-2 in einer repräsentativen Bevölkerungsstichprobe der Stadt Halle 04/2020“, verantwortlich durchgeführt vom Diakoniewerk Halle, liegt bislang kein finales Studienprotokoll vor.

<sup>3</sup> Zu den Unterschieden im Stichprobendesign von NAKO-Gesundheitsstudie und SOEP vgl. Kröger et al. (2010) sowie zur NAKO selbst Ahrens et al. (2020)

<sup>4</sup> Bei der NAKO-Gesundheitsstudie kommt die Altersbegrenzung dazu (aktuell sind Personen im Alter von 73plus nicht in der Stichprobe). Bei dem jüngsten Vorschlag von Schnell et al. (2020) ist ein hoher Grad an Selektivität der Stichprobe zu erwarten, da hier vorgeschlagen wird, die Probanden sollten „bei ihrem Hausarzt“ einen Termin vereinbaren und sich dort testen lassen.

und der prospektiven Dokumentation des Gesundheitszustands der Teilnehmenden für die Analyse der obenstehenden Fragestellungen eröffnet.

Im Folgenden wird zunächst die essentielle Frage nach sinnvollen und zuverlässigen diagnostischen Tests diskutiert. Unser Augenmerk gilt dabei insbesondere auch der aktuellen Entwicklung solcher Tests, die von den Probanden selbst und ohne medizinisches Personal durchgeführt werden, da gerade zuverlässige Tests zur Selbstbeprobung neue Möglichkeiten für flächendeckende, repräsentative Bevölkerungsstichproben eröffnen. Im zweiten Abschnitt geben wir dann zunächst einen Überblick über das Design und die Historie der SOEP-Studie, um danach die Möglichkeiten vorzustellen, welche das SOEP für die Beantwortung der oben genannten Forschungsfragen bietet. Wir gehen dabei auch auf den besonders erfolgskritischen Aspekt bei der Durchführung längsschnittlich angelegter Studie ein, der Sicherung der Teilnahmemotivation über einen längeren Zeitraum. Die Zusammenfassung im vierten Abschnitt enthält auch Schlussfolgerungen und einem Ausblick auf mögliche klinisch-epidemiologische Anschlussstudien.

## 1 Coronavirus-Tests

Im Prinzip gibt es zwei Möglichkeiten, relevante Informationen über die Infektion zu erhalten: durch Tests auf das akute Vorhandensein des Virus (Direktnachweis) und serologische Tests, bei denen Antikörper, also die Immunantwort auf das Virus nach einer akuten Infektion, bestimmt werden.

Tests auf akute Infektion: Der klassische Test auf eine akute Infektion ist der „PCR-Test“ (polymerase chain reaction). Man kann den PCR-Test, bei dem Virus-RNA durch ein Enzym, die DNA-Polymerase, wiederholt synthetisiert wird und so eine messbare Menge an DNA entsteht, wegen seiner hohen analytischen Sensitivität<sup>5</sup> und Spezifität<sup>6</sup> von jeweils nahezu 100% labortechnisch gesehen als Goldstandard für die Diagnose einer akuten COVID-19-Infektion bezeichnen. Praktisch fällt die Sensitivität jedoch geringer aus, weil nur in einem relativ kleinen Zeitfenster bei Infizierten die Viren im Rachen nachweisbar sind und liegt nach einer Meta-Analyse zwischen 71% und 98% (Arevalo-Rodriguez et al. 2020). Die Kosten für einen solchen Test sind deutlich höher als die serologischer Tests, doch dafür liefert die PCR zudem Hinweise auf die Ausgangsmenge an Virus in der Probe und den Virustyp. Allerdings zeigt der Test nur eine akute Virus-Infektion an, und zwar über 7 bis 14 Tage nach dem Einsetzen der Symptome. Sobald eine Immunreaktion einsetzt, verringert sich die Viruslast und die Sensitivität des Tests sinkt.

Für die klinische Diagnostik von SARS-CoV-2 mittels PCR-Test sind Abstriche durch den Mund oder die Nase erforderlich, die am zuverlässigsten von geschultem medizinischem Fachpersonal durchgeführt werden. Erste Erfahrungen mit der Probenentnahme durch Studienteilnehmende selbst zeigen, dass

---

<sup>5</sup> Die Sensitivität eines Tests ist seine Fähigkeit tatsächliche medizinisch relevante Merkmale korrekt zu identifizieren. Ein Test mit 100%-iger Sensitivität würde alle Probanden mit akuter COVID-19-Infektion erkennen. Ein Test mit 80% Sensitivität erkennt 80% der Infektionen (richtig-positiv), aber 20% Infizierte blieben unentdeckt (falsch-negativ).

<sup>6</sup> Die Spezifität eines Tests bezieht sich auf seine Fähigkeit, die Patienten ohne das relevante Merkmal korrekt zu identifizieren. Ein Test mit 100%-iger Spezifität klassifiziert alle Probanden ohne COVID-19-Infektion korrekt als nicht infiziert. Ein Test mit 80%-Spezifität identifiziert 80% der nicht-infizierten Probanden als Testnegativ (richtig-negativ), aber 20% der Probanden ohne Infektion würden fälschlicherweise als Testpositiv (falsch-positiv) klassifiziert.

die Kombination mehrerer Abstriche aus Mund und Nase eine fast ebenso hohe Sensitivität aufweist wie ein durch Fachpersonal gewonnener Rachenabstrich.<sup>7</sup> PCR-Tests aus Speichelproben scheinen nach derzeitigem Sachstand einen höheren Anteil falsch negativer Ergebnisse aufzuweisen als Rachenabstriche. So beschreiben verschiedene Studien, dass im Speicheltest nur knapp 85% der im Rachenabstrich PCR-positiven Fälle erkannt werden (Pasomsub et al. 2020, Williams et al. 2020).

Tests auf die Immunreaktion des Körpers: Weltweit beschäftigen sich im Moment viele Experten auf dem Gebiet der Virenforschung damit, serologische Tests zur Detektion von Antikörpern gegen das Coronavirus hinsichtlich ihrer Handhabung zu vereinfachen und ihre Qualität zu verbessern. Inzwischen sind mehrere kommerzielle Testsysteme auf IgG-Antikörper mit Sensitivitätswerten von  $\geq 94\%$  und Spezifitätswerten  $\geq 99,6\%$  verfügbar. Die Ergebnisse solcher Tests geben allerdings nur an, ob sich das Immunsystem mit dem Virus auseinandergesetzt hat. Dabei kann es durch Antikörper, die der Körper gegen Viren ausgebildet hat, die denen des Corona-Virus sehr ähnlich sind wie zum Beispiel Grippeviren, zu einer gewissen Anzahl von falsch positiven Ergebnissen kommen. Von medizinischem Personal durchgeführte Blutentnahmen und Selbstbeprobung mittels Blutstropfen aus der Fingerkuppe können in laborbasierten Tests vergleichbare Testgüten aufweisen. Soforttests, die wie bei einer Insulin-Messung von Studienteilnehmenden selbst durchgeführt werden, weisen derzeit noch große Ungenauigkeiten durch geringere Sensitivitäts- und Spezifitätswerten auf.

Grundsätzlich können die Schätzungen für die Infektionsrate in der Population für Messungenauigkeiten korrigiert werden. Das heißt bei der statistischen Auswertung mit Blick auf die Population ist keine 100prozentige Sensitivität und Spezifität notwendig (die es auch niemals geben kann). Es reicht, wenn das Ausmaß der fehlerhaften Messungen hinreichend genau bekannt ist. Freilich stellt sich ein Ethikproblem, wenn Probanden das Ergebnis ihres persönlichen Tests mitgeteilt bekommen. Probanden müssen somit entsprechend aufgeklärt werden (vgl. dazu z. B. Lampert und Santos-Hövenner 2020, Abschnitt 4.6); darauf wird weiter unten im Text näher eingegangen. Bereits hier sei angemerkt, dass das Ethikproblem sich auch bei nicht-fehlerhaften Ergebnissen stellt, solange nicht bekannt ist, ob der positive Antikörpernachweis eine Immunität erzeugt und wie lange diese ggf. anhält.

Für die Untersuchung der SOEP-Stichprobe ist aus verschiedenen Gründen, z.B. mit Blick auf den zu erwartenden Aufwand und die damit verbundenen Kosten sowie die Belastung für die Studienteilnehmenden, einer Selbstbeprobung der Vorrang einzuräumen (siehe auch Abschnitt 3). Hierfür bietet sich die Dried Blood Spots (DBS)-Analyse an, die auf einer Blutprobe basiert, die in Form kleiner Tropfen auf einem Filterpapier gewonnen und getrocknet per Post an ein Labor geschickt wird. Dieses Verfahren ist bereits seit den 1960ern – seit dem von Robert Guthrie entwickelten Test zum Screening von Neugeborenen auf Anzeichen einer Phenylketonurie – bekannt. DBS sind jedoch nicht nur für das Neugeborenen-Screening geeignet, sondern qualifizieren sich in besonderem Maße auch für die serologische Diagnostik von Infektionskrankheiten oder Autoimmunerkrankungen, für die es bereits viele Zulassungen mit dieser Probennahmetechnik gibt. Die DBS-Analysen sind zudem hervorragend für Screening- und Monitoringzwecke nutzbar, da die eigenständige Sammlung von Blutproben zu jeder Zeit und an jedem Ort möglich ist.

Ein Vorteil der getrockneten Blutstropfen ist vor allem der minimal invasive Eingriff zur Gewinnung der Probe, für den kein Fachpersonal benötigt wird. Die sehr geringe Blutmenge (es genügt bereits

---

<sup>7</sup> Dr. Janine Michel, RKI, ZBS1; persönliche Kommunikation.

ein Tropfen von ca. 30 µL, für eventuell Nachtestungen sollten jedoch nach Möglichkeit mehrere Tropfen gewonnen werden) ermöglicht auch die einfache Beprobung von Kindern. Außerdem sind IgG-Antikörper gegen SARS-CoV-2 in dem getrockneten Blut über mindestens Wochen stabil. Die Probe kann so unkompliziert bei Raumtemperatur transportiert werden und stellt ein sehr geringes Infektionsrisiko dar, was den normalen Postversand als freigestellte medizinische Probe ermöglicht. Zur Langzeitlagerung empfiehlt sich das Einfrieren in Plastiktüten bei -20°C, wobei nur minimaler Platz benötigt wird.

Aktuell laufen Studien zur Akzeptanz und zum Verständnis der Anleitung zur Selbstbeprobung, deren Ergebnisse zeitnah vorliegen werden. Gleiches gilt für den Vergleich der Ergebnisse von IgG-Tests mit Serum aus venösem Blut und mit Trockenblut (aus der Fingerkuppe gewonnen). Nach vorläufigen Ergebnissen der noch unveröffentlichten Studien soll eine sehr gute Vergleichbarkeit gegeben sein. Solche Vergleiche sind u.a. auch im Rahmen der Studie Corona Monitoring lokal des Robert Koch-Instituts geplant (Lampert und Santos-Hövener 2020).

## 2 Das Sozio-oekonomische Panel (SOEP)

Das SOEP ist eine multidisziplinäre Multi-Kohortenstudie, die im jährlichen Rhythmus seit 1984 in Deutschland durchgeführt und zu diesem Zweck im Rahmen der Leibniz-Gesellschaft am Deutschen Institut für Wirtschaftsforschung (DIW Berlin) gefördert wird. Derzeit werden für rund 49 Tsd. Personen in verschiedenen Teilstichproben (im SOEP Core und dem SOEP Innovation Sample), die insgesamt bevölkerungsrepräsentativ sind, jährlich Daten erhoben. Eine Altersbeschränkung existiert nicht, während z. B. in der NAKO-Gesundheitsstudie nur Volljährige einbezogen sind, die beim ersten Untersuchungszeitpunkt nicht älter als 69 Jahre waren. Derzeit umfasst das SOEP ca. 11 Tsd. Personen in Ostdeutschland und 38 Tsd. in Westdeutschland. Davon sind 10 Tsd. unter 18 Jahren, 18,5 Tsd. zwischen 18 und 49 Jahren und 20,5 Tsd. Personen älter als 50.

Zentrales Merkmal des SOEP ist, dass über alle Haushaltsmitglieder Informationen im Längsschnitt vorliegen. Alle erwachsenen Haushaltsmitglieder sowie im Haushalt lebenden Kinder ab dem 11. Lebensjahr werden direkt befragt. Informationen über im Haushalt lebende Kinder unter 11 Jahren werden über die Eltern ermittelt. Im Rahmen des bei Umzügen erfolgenden zeitlichen Weiterverfolgungskonzepts des SOEP werden auch Umzüge in Alters- und Pflegeheime erfasst. Allerdings erlauben die wenigen selektiven Fälle im SOEP keine Verallgemeinerung auf die knapp 1 Million umfassende Bevölkerungsgruppe in Anstalten. (Derartige Informationen können im Moment nur dem Mikrozensus entnommen werden.) Eine Nachverfolgung geschieht auch, wenn Personen den ursprünglichen Haushalt verlassen. Die SOEP-Erhebungen werden in der Regel interviewergestützt (Face-to-Face) durchgeführt.

Im Zeitraum von April bis Juni 2020 wurde bei rund 10.000 SOEP-Haushalten eine telefonische Befragung zur Corona-Krise durchgeführt (SOEP-CoV) (vgl. Kühne et al. 2020). Die Befragung beschäftigte sich mit dem Thema, wie die Corona-Krise (insbesondere die landesweit verfügbaren Beschränkungen) den Alltag der Menschen in Deutschland geprägt hat und wie diese mit der Situation umgegangen sind. Außerdem wird mit den SOEP-CoV Daten untersucht, welche mittel- und längerfristigen wirtschaftlichen und sozialen Folgen die Pandemie hat. Dabei geht es zum Beispiel um die Auswirkungen auf die



gesundheitlichen und ökonomischen Ungleichheiten zwischen verschiedenen Bevölkerungsgruppen, um die psycho-sozialen Folgen für Einzelne sowie um den gesamtgesellschaftlichen Zusammenhalt. Die Befragung lieferte zudem Informationen über die grundsätzliche Teilnahmebereitschaft an Infektions- und Immunitäts-Tests. Insgesamt haben ca. 90% aller Befragten ihre Bereitschaft für eine Teilnahme an einer Testung (Rachenabstrich und Antikörpertest) angezeigt. Ebenso zeigten sich 90% der befragten Eltern bereit, ihre Kinder im Rahmen einer Selbstbeprobung testen zu lassen.

### 3 Potentiale einer Coronavirus-Testung im Rahmen des Sozio-oekonomischen Panels

Der bereits vorliegende Datenbestand des SOEP ist von sehr großem Wert, für eine Testung der SOEP Stichprobe auf COVID-19 sowie Antikörper gegen COVID-19. Einerseits sind bereits Informationen über Personen und ihre Haushalte verfügbar, die in den vorangegangenen Erhebungswellen zur sozio-ökonomischen Situation, den individuellen oder haushaltsbezogenen Lebensumständen sowie den subjektiven Befindlichkeiten (inklusive selbstberichteten Gesundheitszuständen) für den jeweiligen Erhebungszeitpunkt erfragt wurden (Giesselmann et al. 2019). Damit kann auf eine retrospektive Erhebung von Informationen verzichtet und die damit verbundenen methodischen Probleme und Verzerrungen vermieden werden. Der Einfluss von Lebensumständen und Gesundheitsstatus auf das Auftreten von Infektionen sowie deren Dynamik kann mittels statistischer Modellierung über den Lebenslauf ermittelt werden.

Zudem bietet ein laufendes, prospektives Panel die Möglichkeit, Folgeeffekte bestimmter Ereignisse in der Zukunft identifizieren zu können – da die Haushalte und ihre Personen auch in weiteren SOEP-Befragungswellen über die nächsten Jahre hinweg regelmäßig befragt werden. Damit können auch langfristige Effekte einer Corona-Infektion, wie etwa Immunität durch Antikörper und deren zeitliche Entwicklung oder Krankheitsverläufe und -folgen, beobachtet werden. Für SOEP-Befragte, die eine Infektion durchgemacht haben, können gezielt Fragen gestellt werden (zum Ansatz des „dependent interviewing“ vgl. Kühne et al. 2020: 198).

Darüber hinaus eröffnet das Design einer Haushaltsbefragung die Möglichkeit, Infektionen bei Familienmitgliedern und Virusübertragungen innerhalb des Haushalts genauer zu rekonstruieren (vgl. Döhla et al. 2020) – wodurch man beispielsweise die aktuell diskutierte Frage nach der Rolle von Kindern in Infektionsketten zuverlässiger als bislang beantworten könnte.

Insgesamt bietet das SOEP somit die Möglichkeit einer prospektiven Studie („serial sampling“ in der Terminologie der WHO bzw. der Medizin), und zwar nahezu exakt wie dies die WHO empfiehlt: d.h. im Haushaltskontext einschließlich der Kinder und Hochaltriger und als Längsschnittstudie so früh wie möglich im Verlauf der Pandemie.<sup>8</sup>

#### Durchführung der Tests

Im Rahmen der Testung aller Personen der SOEP-Stichprobe ist es möglich, zunächst anhand eines Tests auf Antikörper zu überprüfen, ob die Person schon einmal mit dem Corona-Virus infiziert war. Dies wird erwartungsgemäß nach aktuellem Stand in mehr als 98 Prozent der Fälle nicht zutreffen. Daran anschließend soll eine Testung auf eine akute Infektion erfolgen. Im Kontext der üblichen Abläufe der SOEP-Erhebungen ist die Abwicklung der Test auf Antikörper und der Rachenabstrich über

---

<sup>8</sup> WHO, a. a. O., S. 5f.

einen medizinischen Dienst nicht zielführend. Dafür sind zwei Argumente entscheidend: Ein „logistisches“ und ein „surveymethodologisches“. Der Einsatz eines medizinischen Dienstes oder medizinisch ausgebildeten Personals ist aufgrund der hohen regionalen Streuung der SOEP-Stichprobe nicht realistisch (weil zu teuer und langwierig). Eine Testung in medizinischen Zentren, wie dies beispielsweise im Rahmen der NAKO-Gesundheitsstudie vorgenommen wird, würde einen zu hohen Aufwand für die Befragten bedeuten. Zudem wäre die Koordinierung der Testtermine zwischen medizinischem Dienst/Personal und Befragten – letztlich auch aufgrund der räumlichen Distanzen – zu aufwändig und somit nicht umsetzbar – auch aufgrund der dadurch entstehenden hohen Kosten.

Das surveymethodologische Argument baut auf der Überlegung auf, dass die Teilnahme an einer Testung für die SOEP-Befragten möglichst niederschwellig sein muss, damit eine möglichst hohe Ausschöpfung auch erreicht werden kann. Die Koordinierung gesonderter Termin und Besuche im Haus bedeuten einen deutlich höheren Aufwand auch auf der Seite der Befragten. Dadurch wäre hier mit größeren Ausfällen zu rechnen. Der Vorschlag von Schnell und Smid (2020), die Befragten sollten eine Testung bei ihrem Hausarzt vornehmen, ist dabei keine wirkliche Alternative. Denn auch hier entsteht auf Seiten der Befragten ein Koordinierungs- und Zeitaufwand, um Termine abzustimmen und diese dann auch wahrzunehmen. Hinzu kommt, dass eine Testung – zur Ausschöpfung des analytischen Potentials des SOEP – für alle Haushaltsmitglieder vorzunehmen wäre, was gerade für Mehrpersonenhaushalte den Aufwand weiter erhöhen würde. Vor diesem Hintergrund – Durchführbarkeit und Kosten einerseits, Effekte auf die Teilnahmebereitschaft durch hohen Koordinierungs- und Zeitaufwand auf Seiten der Befragten andererseits – ist allein die Beprobung durch die Befragten selbst die realistische und zielführende Methode. Vor diesem Hintergrund ist jedoch sehr genau zu prüfen, inwieweit die verfügbaren Verfahren und Instrumente zur Selbstbeprobung hinreichend exakte Testergebnisse liefern.

Nach einer ersten Basismessung im zuvor beschriebenen Format könnte in regelmäßigen Abständen eine weitere Prüfung auf eine akute Corona-Infektion vorgenommen werden. Die Abstände sollten sich an den natürlichen immunologischen Prozessen der Antikörperentwicklung und den regelmäßigen Testungen systemrelevanter Personen (Polizei, Ärzte, Pfleger etc.) orientieren.<sup>9</sup>

#### Erfahrungen des SOEP mit Selbstbeprobungen

Das SOEP erhebt im Rahmen einer Selbstauskunft der Befragten seit 1984 gesundheitlich und medizinisch relevante Daten im Haushaltskontext (Schupp 2012); seit 2006 wird auch alle zwei Jahre die „Handgreifkraft“ als Indikator für körperliche Fitness physisch gemessen, vgl. Hank et al. (2008).

Das SOEP verfügt auch für das Einsammeln von Körpermaterial bereits einschlägige Erfahrungen. In einem Pretest-Panel wurden im Jahr 2008 für einige hundert Fälle von Erwachsenen Speichelproben für die Analyse bestimmter Loci im Genom mit Hilfe einer interviewergestützten Selbstbeprobung gewonnen (vgl. Schonlau et al. 2010). In der SOEP IS-Stichprobe wurden zudem 2015 (vgl. Gerstorf und

---

<sup>9</sup> In Großbritannien werden 100 Tsd. Probanden Test-Kits zugesandt, siehe <https://www.theguardian.com/world/2020/apr/30/randomised-coronavirus-test-of-100000-will-determine-end-date-of-uk-lock-down>. Diese Zusammenarbeit eines Universitätsinstituts mit einem Umfrageinstitut wird näher beschrieben unter <https://www.imperial.ac.uk/news/197217/home-testing-coronavirus-track-levels-infection/>

Schupp 2016:) sowie 2019 und 2020 ca. 2750 Speichelproben ebenfalls mit Hilfe einer interviewergestützten Selbstbeprobung (Swabs) gewonnen. Dabei wurden 2019/20 nicht nur Proben für Erwachsene gewonnen, sondern auch von Kindern, denen die Eltern die Probe entnahmen, um das Genom von Familien (Trios) analysieren zu können. Es liegen im SOEP somit hinreichende Erfahrungen bei der Sammlung von Speichelproben vor, einschließlich der Berücksichtigung der Forschungsethik und des nicht unkomplizierten Datenschutzes.<sup>11</sup>

Unabhängig von den oben bereits genannten Argumenten für eine Selbstbeprobung zeigen die bisherigen Erfahrungen des SOEP somit auch, dass die Einschaltung eines medizinischen Dienstes bzw. medizinischen Personals kein Muss bei der Erhebung von Bioproben in der SOEP-Stichprobe ist (vgl. auch Kühne et al. 2020: 197).

### Erwartete Anzahlen von infizierten und seropositiven Teilnehmenden

Legt man die Fallzahl von 49 Tsd. aktiven Mitgliedern der SOEP-Stichprobe zugrunde, so werden bei einer geschätzten Infektionsrate von 5 Prozent im Herbst 2020 voraussichtlich rund 2450 Stichprobenmitglieder mit dem Corona Virus infiziert sein. Diese Fallzahl erlauben eine präzise Schätzung der Infektionsrate in der Bevölkerung sowie differenziertere Analysen.

Bei einer Mortalitätsrate von 0,36 Prozent (Wert aus der Heinsberg-Studie) ist im SOEP in 2020 mit acht bis neun Probanden zu rechnen, die infolge einer COVID-19-Erkrankung in 2020 sterben. Ohne weitere Zusatzinformation über die COVID-19-Sterblichkeit kann auf Basis einer so geringen Fallzahl keine Aussage gemacht werden.<sup>12</sup>

Der Längsschnittansatz im SOEP und querschnittsorientierte Ansätze aus anderen Erhebungen sind methodisch gut vereinbar. Das Nebeneinander dieser beiden Messungen liefert einen Erkenntnisgewinn im Sinne präziserer Aussagen über das Infektionsgeschehen im Längsschnitt. Ohne die Befragung über einen Panel-Ansatz sind dagegen Aussagen über individuelle Ansteckungs- und Krankheitsverläufe nur bedingt und mit großer Unsicherheit möglich (z.B. durch retrospektive Befragungen, die allerdings von Erinnerungsproblem der Befragungspersonen betroffen sind).

### Teilnahmemotivation

Die Beteiligung an einer im Rahmen des SOEP durchgeführten Studie ist grundsätzlich freiwillig, wobei die „informierte Zustimmung“ eine Grundvoraussetzung für die Teilnahme an der Studie ist. Zudem müssen alle Aspekte des Schutzes individueller Daten berücksichtigt werden. Der entscheidende Unterschied zu neu aufgesetzten Stichproben (im Querschnitt oder Längsschnitt) besteht im Fall einer

---

<sup>11</sup> Die 2019er-Erhebung ist bislang nicht öffentlich dokumentiert. Es liegen detaillierte Unterlagen und entsprechende Probandenaufklärungen vor, die neben den Datenschützern des DIW Berlin und des Erhebungsinstituts KANTAR von der zentralen Ethikkommission der Max-Planck-Gesellschaft geprüft und für gut befunden wurden.

<sup>12</sup> Der Fragebogen über Verstorbene, der im SOEP eingesetzt wird (vgl. z. B. [https://www.diw.de/documents/publikationen/73/diw\\_01.c.622052.de/diw\\_ssp0611.pdf](https://www.diw.de/documents/publikationen/73/diw_01.c.622052.de/diw_ssp0611.pdf)), könnte zwar un schwer ab 2021 um Corona-spezifische Fragen ergänzt werden, aber SOEP-Analysen der Todesumstände von mit Corona infizierten Verstorbenen werden – wenn überhaupt – nur langfristig sowie selektiv bzgl. der Haushaltzusammensetzung möglich werden.

SOEP-Studie darin, dass die Haushalte bzw. Personen bereits über mehrere Erhebungswellen Erfahrungen mit der Studie selbst und vor allem mit dem Erhebungsinstitut haben, das die Feldarbeit durchführt. Damit besteht ein gewisses Vertrauensverhältnis und auch eine gewisse Identifikation mit dem SOEP, was wiederum die Bereitschaft zur Teilnahme an einer Beprobungsstudie erwartungsgemäß befördert. Zumindest können wir hier eine deutlich geringere Teilnahmeschwelle voraussetzen als im Fall einer Ersterhebung mit Erstkontakt zwischen der Forschungseinrichtung, dem Feldinstitut und den Befragten. Eine SOEP-basierte Beprobungsstudie kann also auf einen Vertrauensvorschluss auf Seiten der Befragten aufbauen, der in anderen Fällen unter normalen Umständen nicht besteht und erst aufgebaut werden müsste. Dies gilt insbesondere auch deshalb, weil die Einhaltung datenschutzrechtlicher Regelungen im SOEP den Befragten hinreichend kommuniziert wird, die Befragten mit der Einhaltung dieser Regelungen durch das SOEP auch hinreichende Erfahrungen haben und dementsprechend die Grundlage für die künftige Kooperationsbereitschaft schafft.

Wie bereits oben erwähnt, stellt aus einer surveymethodischen Perspektive die Beprobung außerhalb des befragten Haushalts eine besondere Hürde der Teilnahme dar. Damit sind für die Haushalte zusätzliche Kosten (in Form von Zeit oder auch Wegekosten) verbunden, die sehr wahrscheinlich zu einer systematischen Verzerrung der Stichprobe führen würden, da der zu erbringende Aufwand für unterschiedliche Haushaltstypen unterschiedlich hoch ist (regionale Nähe, Anzahl an Personen, Aufwand der Terminvereinbarung aufgrund eigener Erwerbstätigkeit etc.). Zudem ist davon auszugehen, dass die verschiedenen Haushalte auch unterschiedlich bereit sind, diesen Aufwand in Kauf zu nehmen (z.B. ist für ältere und weniger mobile Befragten mit einem höheren individuellen Aufwand zu rechnen als bei mobilen Gruppen). Vor diesem Hintergrund sollte auch die Beprobung möglichst niederschwellig, d.h. von Befragten selbst durchgeführt werden können. Nur dann ist eine erhöhte Teilnahmequote und eine mit Blick auf relevante Merkmale (wie Erwerbsstatus, Alter, Größe des Haushalts, Erreichbarkeit medizinischer Einrichtungen etc.) geringere Stichprobenselektivität zu erwarten.

Gleichwohl bedeuten gerade regelmäßige Tests für die SOEP-Studienteilnehmenden eine erhebliche zusätzliche Belastung. Wenn es gelingt, eine win-win-Situation für Probanden und Studieninstitution durch adäquate Anreize für die Teilnehmenden zu schaffen, kann nach unserer Erfahrung auch im Längsschnitt mit einer guten Compliance und geringen künftigen Panelausfällen gerechnet werden.

Ein besonders guter Anreiz zur Teilnahme an der Testung dürfte die Möglichkeit sein, dass die SOEP-Teilnehmenden ihren Antikörperstatus und damit vermutete Immunität gegen das SARS-CoV-2-Virus erfahren können. Da mittlerweile entsprechende Tests auch kommerziell angeboten werden, ergeben sich daraus für Haushalte, die an ihrem Immunisierungsstatus interessiert sind, auch monetäre Anreize, da die Testung im Rahmen der Studie für sie kostenfrei ist. Freilich ist – was die Übermittlung der Ergebnisse anbelangt – damit auch ein forschungsethisches Problem verbunden, das es zu adressieren gilt (durch eine getrennte Vorhaltung von Adressdaten und Testergebnissen, ein geeignetes Datenschutzkonzept und eine Datenschutzfolgeabschätzung).

Die meisten Teilnehmenden werden ein großes Interesse an ihren Testergebnissen (und denen ihrer Kinder) haben. Hierbei gilt es jedoch zu berücksichtigen, dass neben der (1) bislang unzureichenden Kenntnis über die durch eine Infektion erworbene Immunität und deren Dauer (2) selbst die besten Tests einen gewissen Anteil falsch positiver Testergebnisse aufweisen (und somit eine Spezifität, die kleiner als 100% ist, siehe Abschnitt 1 zu den genutzten Corona-Tests). Zum Beispiel kann eine Person, die positiv auf Antikörper gegen COVID-19 getestet wurde, nicht sicher sein, auch tatsächlich Antikörper zu haben und somit vor einer COVID-19-Infektion geschützt zu sein.

Wird indes vom Antikörper-Test eine zurückliegende Infektion übersehen (die Sensitivität liegt also unter 100%), ist das für das Alltagsverhalten der Betroffenen weniger gravierend, da sie sich vermutlich weiterhin vorsichtig verhalten werden. Gravierender ist auf jeden Fall das Problem, dass nicht bekannt ist, ob und wie lange eine korrekt erkannte Infizierung zu einer Immunität führt, die nicht nur den Betroffenen selbst schützt, sondern ihn auch nicht mehr infektiös sein lässt. Um diese Frage zu beantworten, wäre zumindest eine weitere Testung der als Antikörper positiv identifizierten Personen notwendig, z.B. in einer weiteren Selbstbeprobung der SOEP-Stichprobe zu einem späteren Zeitpunkt z.B. im Sommer 2021. Auch über eine Nachtestung positiver Proben auf Stichprobenbasis ließe sich die aus der SOEP-Stichprobe geschätzte bevölkerungsbezogene Seroprävalenz rechnerisch korrigieren.

Die skizzierte Studie muss diesen Problemen durch eine sorgfältige Teilnehmer-Aufklärung und Erläuterung der Testergebnisse Rechnung tragen.

### Nonresponse

Ein großer Vorteil der Benutzung des SOEP als Stichprobenbasis für eine freiwillige Studie zur Testung auf COVID-19 und Antikörper gegen COVID-19 ist die Möglichkeit, den zu erwartenden nicht zufälligen Ausfall in der Stichprobe statistisch durch eine geeignete Gewichtung bzw. Nonresponse-Modellierung über die Merkmale aus dem SOEP quantifizieren und somit kontrollieren zu können. Im Gegensatz zu Stichproben aus dem Einwohnermeldeamt, die lediglich Alter, Geschlecht und den Familienstand als Kontrollmerkmale benutzen können, ergibt sich hier die Möglichkeit, sehr spezifische Merkmale zu nutzen, wie zum Beispiel der Gesundheitsstatus der Person oder die Anzahl der Krankenhausbesuche. Auch die Größe des sozialen Umfelds kann als Kontroll- bzw. Korrekturvariable genutzt werden. Insgesamt gehen wir von einer Teilnahmequote von 65% aus, die sich erwartungsgemäß zwischen verschiedenen Bevölkerungsgruppen unterscheiden wird.

### Fallzahlplanung

Fallzahlplanung ist in medizinischen Studien obligatorisch. Generell gilt, je seltener ein Merkmal ist, umso höher muss die Fallzahl sein, um die Genauigkeit der Schätzwerte abzusichern. Bei den Hot-Spot-Stichproben des RKI wurde die Prävalenz des SARS-CoV-2-Virus nach Altersgruppen als Planungsrahmen genommen (vgl. Lampert und Santos-Hövenner (2020)). Der Stichprobenumfang von 2000 Probanden wurde als ausreichend angesehen. In der hier beschriebenen Studie sollen aber regionale Unterschiede berücksichtigt werden. Weiterhin sollen im Rahmen des prospektiven Designs auch seltenere Merkmale wie das Auftreten von Symptomen, ein Krankenhausaufenthalt sowie ein schwerer Verlauf analysiert werden. Hierzu bedarf es deutlich höherer Fallzahlen. Hinzu kommt, dass aufgrund der gewünschten Erfassung aller Haushaltsmitglieder mit Klumpungseffekten zu rechnen ist, die die effektive Power für Analysen verringern. Die maximale Fallzahl ist durch die Anzahl der 49 Tsd. SOEP-Mitglieder gegeben. Rechnet man mit einem Designeffekt von 1,6 durch die Klumpung von Personen in Haushalten, verbleibt eine effektive Fallzahl von ca. 31 Tsd. Personen.

Ein klassisches Argument für möglichst geringe Fallzahlen bei medizinischen Studien ist die Minimierung von Risiken für die Probanden, z.B. über Nebenwirkungen. Aus diesem Grund sollten nur so we-

nige Personen wie statistisch notwendig in die Studie einbezogen werden. In diesem Fall ist das Treatment, also Tests auf Antikörper bzw. auf eine akute Infektion, mit einem gesundheitlichen Vorteil für die Probanden verbunden, so dass sich von dieser Seite keine Einschränkungen begründen lassen. Es bleibt damit einzig das Kostenargument, welches den Umfang der Studien begrenzt.<sup>13</sup> Um hier das Feld je nach Teilnahmequote und vorhandenen finanziellen Ressourcen gut steuern zu können, ist es möglich die Stichprobe der SOEP-Haushalte zufällig in zum Beispiel drei gleichgroße Teilstichproben zu zerlegen (Tranchen), die je nach Bedarf nacheinander ins Feld gegeben werden können. Damit besteht der effektive Stichprobenumfang jeder Tranche aus ca. 10 Tsd. potentiellen Teilnehmern. Dabei sollten Personen, die schon im Rahmen der SOEP-CoV-Studie zu ihren Corona-Erfahrungen befragt wurden (vgl. Kühne et al. 2020), in jedem Fall in die Stichprobe gelangen, um Synergie-Effekte nutzen zu können.

### Terminierung der Erhebung

Grundsätzlich gilt, dass der Zeitpunkt des Einsatzes einer Längsschnittstudie möglichst früh nach Start einer Pandemie angesetzt werden sollte, da die Antikörpertests nur die Information liefern, dass von Beginn des Epidemie-Geschehens bis zum Messzeitpunkt eine Infektion aufgetreten ist.

Neben der einmaligen Testung empfiehlt die WHO die Erfassung und Modellierung des zeitlichen Verlaufs der Stadien der Virusinfektion sowohl durch den diagnostischen Virusnachweis als auch durch serologische Tests auf Antikörper.<sup>14</sup> Dementsprechend ist eine Wiederholung der Testungen angeraten. Eine Einbettung dieser Testungen im weiteren Panelverlauf ist möglich. In dem Fall bietet es sich an, die Testungen ggf. als Teil der regulären, interviewergestützten Befragungswellen einzubetten.

### Datennutzung

Der entscheidende Vorteil der skizzierten Nutzung der SOEP-Stichprobe zur Testung auf akute Infektion und Antikörper ist die Verknüpfungsmöglichkeit mit den bereits im Längsschnitt vorliegenden Informationen über die Personen und ihre Haushalte. Die Testergebnisse werden dem existierenden Datenbestand des SOEP hinzugefügt und damit Teil des regulären Datenangebots des SOEP für wissenschaftliche Zwecke. Dies bedeutet auch, dass sie für die Scientific Community nutzbar sind. Bei der Aufbereitung, Dokumentation und Bereitstellung dieser Daten kann auf die bestehenden Infrastrukturen und internen Datenproduktionsprozesse des SOEP zurückgegriffen werden. Da gerade durch die Verknüpfung von Sozial- und Gesundheitsdaten ein hohes datenschutzrechtliches Risikopotential besteht, muss sichergestellt werden, dass auch in der Nachnutzung der Daten keine Identifikation der Befragten vorgenommen wird. Das SOEP hat für derartige datenschutzrechtlich hoch sensible Forschungsdaten besondere Nutzungsmodelle entwickelt, wobei hoch sensible Informationen

---

<sup>13</sup> Einfache Fallzahlabeschätzungen basieren normalerweise auf der Binomialverteilung und einem einfachen Anteilsschätzer. Komplexere Schätzprozeduren, die etwa zusätzliche Stichproben im Rahmen eines Small-Area Modells (vgl. Schmid et al. 2017) benutzen, bleiben hierbei unberücksichtigt. Je nach Prognosequalität für das interessierende Merkmal können dabei hohe Genauigkeitsgewinne aus Zusatzinformationen gewonnen werden.

<sup>14</sup> WHO, a. a. O., S. 5: "If serial sampling is to be conducted it is best to initiate the investigation as quickly as possible."

nur vor Ort im Forschungsdatenzentrum des SOEP (FDZ SOEP) unter besonderen Sicherheitsvorkehrungen genutzt werden können. Genau dieses Modell wäre in dem vorliegenden Fall anzuwenden.

Was die Datenanalyse und die eingesetzten Methoden anbelangt, so hat in den letzten zwei Jahrzehnten eine stürmische Entwicklung stattgefunden, die basierend auf klassischen Ereigniszeit-Methoden so genannte Multi-State-Models betrachtet (vgl. Beyersmann, Allignol, Schumacher 2012, van Houwelingen, Putter 2012). Diese sind das ideale Mittel, um Infektions- und Krankheitsverlauf mit Hilfe getesteter Verfahren zu analysieren. Fragt man nicht nur, ob ein Ereignis eingetreten ist, sondern auch wann, ist damit ein enormer Gewinn an Informationsgehalt und damit an statistischer Power verbunden. Mittlerweile finden diese Modelle Anwendungen in vielen Gebieten sowohl in medizinischem Kontext als auch in den Wirtschafts- und Gesellschaftswissenschaften. Als Blaupause für die geplanten Analysen könnten die Arbeiten der Gruppe um Martin Schumacher, Universität Freiburg dienen, der für seine statistischen Analysen des Verlaufs von Infektionen mit Krankenhauskeimen höchste, auch internationale Anerkennung erhalten hat (vgl. Schumacher et al. 2013). Bei den SOEP-Daten kommt hinzu, dass die Wohngegend der Teilnehmenden (vom Postleitzahl-Bezirk bis zum Straßenabschnitt bzw. Wohnblock) verwendet werden kann, um so genannte georeferenzierte Analysen durchzuführen. Derartige Analysen können innerhalb des SOEP-FDZ am DIW Berlin durchgeführt werden. Es gibt in Deutschland auch eine Reihe von Forschenden, die auf dem zugehörigen Feld der Small-Area-Schätzmethoden aktiv sind (vgl. z.B. Schmid et al. 2017).

Eine Studie auf der Basis der SOEP-Kohorte kann dazu beitragen,

- zuverlässige georeferenzierte Schätzungen der Dynamik und des Stands der Pandemie zu liefern und
- Risikofaktoren für den Krankheitsverlauf zu ermitteln.

Da nicht damit gerechnet werden kann, dass in dieser Studie eine ausreichende Anzahl von schweren und fatalen Krankheitsverläufen beobachtet wird, bietet es sich an, ab Krankenhausaufenthalt eines Teilnehmenden einen nahtlosen Übergang in eine neue Studie, die dann auf der Krankenhausdokumentation beruht, zu veranlassen. Der Stichprobenumfang könnte dabei erhöht werden, indem eine weitere Anzahl von COVID-19-Patienten aus demselben Krankenhaus ausgewählt wird.

#### 4 Zusammenfassung und Ausblick

Mit einem seriellen prospektiven Panelansatz auf Basis einer bevölkerungsrepräsentativen Stichprobe lassen sich die Risiken einer Infektion mit SARS-CoV-2 und die Ursachen für unterschiedliche Infektionsrisiken messen. Die Testung der Probanden kann auf der Grundlage des aktuellen Entwicklungsstands der Testmethodik durch Versand von Sets zur Selbstbeprobung mit zuverlässiger, laborbasierter Analytik durchgeführt werden. Dabei bietet die seit 1984 laufende Multi-Kohortenstudie des Sozioökonomischen Panels (SOEP) gegenüber der Ziehung einer neuen Stichprobe wertvolle Vorteile: Es können alle Haushaltsmitglieder ohne Altersbeschränkung, inklusive Kinder, und ohne starke geographische bzw. regionale Klumpung getestet werden. Zudem stehen für alle SOEP-Teilnehmenden vielfältige, auch den Gesundheitsstatus betreffende Merkmale aus früheren und künftigen jährlichen Panelwellen zur Verfügung. Durch den prospektiven Ansatz ist es auch möglich, erst später im Verlauf

der Pandemie erworbene Immunitäten zu finden und ggf. ihre sozio-ökonomisch differentiellen Ursachen zu analysieren.

Der hier vorgestellte Ansatz ist mit den Anforderungen der WHO kompatibel (vgl. auch Abbott 2020) und geht durch die Nutzung der Information aus früheren und nachfolgenden Panelwellen noch darüber hinaus. Der Stichprobenumfang des SOEP von über 49 Tsd. Probanden ist für valide statistische Analysen ausreichend. Angesichts der deutlich gesteigerten Testkapazitäten in Deutschland sind derart viele Akut- und Antikörper-Tests zudem machbar.

Aufgrund der hohen Aktualität der Thematik ist mit einer sehr großen Teilnahmebereitschaft zu rechnen (ca. 65%). Das SOEP hat zudem praktische Erfahrungen mit der Gestaltung der forschungsethischen Notwendigkeiten und Aufklärungspflichten als auch für den Umgang mit sensitivem Material und Daten.

Eine Studie auf der Basis der prospektiven SOEP-Multikohorten-Studie kann dazu beitragen,

- Risikofaktoren für den Krankheitsverlauf zu ermitteln und
- Folgeeffekte von Erkrankungen im weiteren Lebensverlauf zu identifizieren.

Um eine Abschätzung der klinischen Pandemiefolgen vornehmen zu können, bedarf es Informationen, die nur durch einen nahtlosen Übergang in eine Anschlussstudie, bei der im Fall einer Erkrankung, die stationär behandelt werden muss, weitere COVID-19-Krankenhauspatienten hinzugezogen werden. Auch Informationen anderer klinisch-epidemiologischer Studien sollten im Stil eines metanalytischen Vorgehens herangezogen werden. Eine Voraussetzung dafür ist die transparente Beschreibung aller erfassten Befunde und Kovariablen einer SOEP-Studie. Wenn es damit gelingt, ein Multi-State-Model zu erstellen, lassen sich viele wichtige Fragen bezüglich erwarteter Verweilzeiten z.B. im Krankenhaus, auf der Intensivstation, Abhängigkeiten von Komorbiditäten etc. und letztendlich die Wahrscheinlichkeit für fatale Ausgänge der Infektion mit soliden statistischen Methoden schätzen.

## 5. Literaturangaben

Abbott, A. (2020). Huge Cohort Studies Pivot to Investigate the Coronavirus. *Nature*, 582, 326.

Ahrens, W., Greiser, K.H., Linseisen, J., Pischon, T. & Pigeot, I. (2020). Erforschung von Erkrankungen in der NAKO Gesundheitsstudie. Die wichtigsten gesundheitlichen Endpunkte und ihre Erfassung. *Bundesgesundheitsblatt* 63 (4), 376-84. [doi: 10.1007/s00103-020-03111-0](https://doi.org/10.1007/s00103-020-03111-0)

Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, Zambrano-Achig P, Del Campo R, Ciapponi A, Sued O, Martínez-García L, Rutjes A, Low N, Perez-Molina JA, Zamora J. (2020). False-negative results of initial RT-PCR assays for COVID-19: a systematic review. *medRxiv* 2020.04.16.20066787. [doi:10.1101/2020.04.16.20066787](https://doi.org/10.1101/2020.04.16.20066787).

Beyersmann, J., Allignol, A. and Schumacher, M. (2012). *Competing Risks and Multistate Models with R*. Springer, New York.



- Döhla, M. et al. (2020). SARS-CoV-2 in environmental samples of quarantined households. Bonn: medRxiv preprint (posted June 2, 2020), [doi: 10.1101/2020.05.28.20114041](https://doi.org/10.1101/2020.05.28.20114041).
- Gerstorff, S. und Schupp, J. (2016). SOEP Wave Report 2015. Berlin: <https://www.econstor.eu/handle/10419/148021>
- Giesselmann, M., Bohmann, S., Goebel, J., Krause, P., Liebau, E., Schacht, D., Richter, D., Schröder, C. Schupp, J. & Liebig, S. (2019). The Individual in Context(s): Research Potentials of the Socio-Economic Panel Study (SOEP) in Sociology. *European Sociological Review*, 35(5), 738-755. [doi: 10.1093/esr/jcz029](https://doi.org/10.1093/esr/jcz029).
- Hank, K., Juerges, H., Schupp, J. und Wagner, G.G. (2009). Isometrische Greifkraft und sozialgerontologische Forschung: Ergebnisse und Analysepotentiale des SHARE (2004) und SOEP (2006), *Zeitschrift für Gerontologie und Geriatrie*, 42 (2), 117–126. [doi: 10.1007/s00391-008-0537-8](https://doi.org/10.1007/s00391-008-0537-8)
- KoCo19 (2020): Prospektive COVID-19 Kohorte München (KoCo19) <http://www.klinikum.uni-muenchen.de/Abteilung-fuer-Infektions-und-Tropenmedizin/de/news/Aktuelles/KoCo19.html> (2. Juni 2020)
- Kröger, H. Schupp, J. and Behrens, J. (2014). Unlocking further potential in the National Cohort study (NaKo) through comparability with the German Socio-Economic Panel, RatSWD Working Paper No. 237. [https://www.ratswd.de/dl/RatSWD\\_WP\\_237.pdf](https://www.ratswd.de/dl/RatSWD_WP_237.pdf)
- Kühne, S., Kroh, M., Liebig, S. and Zinn, S. (2020). The Need for Household Panel Surveys in Times of Crisis: The Case of SOEP-CoV, *Survey Research Methods* 14 (2), 195-203. [doi: 10.18148/srm/2020.v14i2.7748](https://doi.org/10.18148/srm/2020.v14i2.7748)
- Lampert, T. and Santos-Hövenner, C. (2020). Studienprotokoll Entwurf (Stand 15.5.2020) „Seroepidemiologische Studie zur Verbreitung von SARS-CoV-2 in der Bevölkerung an besonders betroffenen Orten in Deutschland – CORONA-MONITORING lokal“, Berlin: [https://www.rki.de/DE/Content/Gesundheitsmonitoring/Studien/cml-studie/Fachinformationen/Studienprotokoll.pdf?\\_\\_blob=publicationFile](https://www.rki.de/DE/Content/Gesundheitsmonitoring/Studien/cml-studie/Fachinformationen/Studienprotokoll.pdf?__blob=publicationFile) .
- Pasomsub, E. et al. (2020). Saliva sample as a non-invasive specimen for the diagnosis of coronavirus disease-2019 (COVID-19): a cross-sectional study. *Clinical Microbiology and Infection* [doi: 10.1016/j.cmi.2020.05.001](https://doi.org/10.1016/j.cmi.2020.05.001)
- Portal Embryotox: [www.embryotox.de](http://www.embryotox.de) (6.Juni2020)
- Radon, Katja, Elmar Saathoff et al. 2020), Protocol of a population-based prospective COVID-19 cohort study Munich, Germany (KoCo19), München: medRxiv preprint. [doi: 10.1101/2020.04.28.20082743](https://doi.org/10.1101/2020.04.28.20082743)
- Schmid, T., Bruckschen, F., Salvati, N. and Zbiranski, T.(2017). Constructing socio-demographic indicators for National Statistical Institutes using mobile phone data: Estimating literacy rates in Senegal *Journal of the Royal Statistical Society: Series A*, 180, 1163-1190. doi: [10.1111/rssa.12305](https://doi.org/10.1111/rssa.12305)
- Schnell, R., & Smid, M. (2020). Methodological Problems and Solutions in Sampling for Epidemiological COVID-19 Research. *Survey Research Methods*, 14 (2), 123-129. [doi: 10.18148/srm/2020.v14i2.7749](https://doi.org/10.18148/srm/2020.v14i2.7749)

Schonlau, Matthias, Thomas Dohmen, Armin Falk, Christian Montag, Martin Reuter, Matthias Schonlau, Juergen Schupp, Nico A. Siegel, Uwe Sunde, Gert G. Wagner und Bernd Weber (2010):, Collecting Genetic Samples in Population Wide (Panel) Surveys: Feasibility, Nonresponse and Selectivity, *Survey Research Methods*, 4 (2), 121-126. [doi: 10.18148/srm/2010.v4i2.3959](https://doi.org/10.18148/srm/2010.v4i2.3959)

Schumacher M, Allignol A , Beyersmann J, Binder N, Wolkewitz M (2013). Hospital-acquired infections—appropriate statistical treatment is urgently needed! *International Journal of Epidemiology*, 42 (5), 1502-1508. [doi: 10.1093/ije/dyt111](https://doi.org/10.1093/ije/dyt111)

Schupp, J. (2012): Das Sozio-oekonomische Panel (SOEP). *Bundesgesundheitsblatt*, 55 (6-7), 767-774. [doi: 10.1007/s00103-012-1496-1](https://doi.org/10.1007/s00103-012-1496-1)

Streeck, H. et al. (2020): Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event, medRxiv 2020.05.04.20090076. [doi: 10.1101/2020.05.04.20090076](https://doi.org/10.1101/2020.05.04.20090076).

van Houwelingen, H. C. and Putter, H. (2012). *Dynamic Prediction in Clinical Survival Analysis*. Monographs on Statistics and Applied Probability 123. CRC Press, Boca Raton, FL.

Williams E, Bond K, Zhang B, Putland M, Williamson DA (2020). Saliva as a non-invasive specimen for detection of SARS-CoV-2. *Journal of Clinical Microbiology*, [doi: 10.1128/JCM.00776-20](https://doi.org/10.1128/JCM.00776-20).

WHO (2020): Population-based age-stratified seroepidemiological investigation protocol for COVID-19 virus infection, Version 1.1, March 17 2020 <https://www.who.int/publications-detail/population-based-age-stratified-seroepidemiological-investigation-protocol-for-covid-19-virus-infection> (12. Mai 2020).