

# Stichproben für die COVID-19 Forschung \*

Rainer Schnell<sup>1</sup>, Menno Smid<sup>2</sup>, Horst Müller-Peters<sup>3</sup>, and Anke Müller-Peters<sup>4</sup>

<sup>1</sup>Universität Duisburg-Essen

<sup>2</sup>infas

<sup>3</sup>Technische Hochschule Köln

<sup>4</sup>SMARTmed-Test

13. Mai 2020

Basierend auf der Kritik bestehende Datenerhebungen zur Prävalenz von COVID-19 Infektionen<sup>1</sup> wird eine neue Datenerhebung zur Bestimmung des Anteils von Personen in der Gesamtbevölkerung, bei denen eine Infektion nachgewiesen werden kann, vorgeschlagen.

Mit dieser Studie können Aussage über Gesamtdeutschland gemacht werden. Im Gegensatz zu allen anderen durchgeführten und geplanten Studien basiert diese Studie auf einer bundesweiten echten Zufallsstichprobe aus Einwohnermelderegistern und berücksichtigt auch Alten- und Pflegeheime.

## 1 Seroprävalenzsurvey

Das Ziel des Seroprävalenzsurveys ist die Schätzung des Anteils der Personen, die Antikörper zum Nachweis bereits durchgemachter Infektionen besitzen. Die beabsichtigte Methode des Nachweises ist der Roche-Test (ElecSys<sup>®</sup> Anti-SARS-CoV-2-Antikörpertest).<sup>2</sup>

---

\* Wir danken Prof. Dr. Ralf Münnich, PD Dr. med. Günther Heller, Dr. Ansgar Schmitz-Veltin, Dr. med. Katrin Schnell, Prof. Dr. Hans Kiesel und Philip Hoecker für Diskussionen.

<sup>1</sup>Klingwort und Schnell (2020), Schnell und Smid (2020b), Schnell und Smid (2020a)

<sup>2</sup>Nach Herstellerangaben mit einer Spezifität von 99.81% (ab 14 Tage nach Infektion) und einer Sensitivität 100% (Roche 2020).

## 1.1 Hauptstichprobe

Zur Ziehung von echten Zufallsstichproben der Gesamtbevölkerung in Deutschland gibt es kaum eine Alternative zu Einwohnermeldestichproben (Gabler und Hoffmeyer-Zlotnik 1997; Kauermann und Küchenhoff 2010; Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute 2014). Die Qualität der Melderegister überwiegt den einzigen Nachteil dieser Stichproben, die lange Vorbereitungszeit, bei weitem.

Vorgeschlagen wird eine Einwohnermeldestichprobe, bei der aus jedem Bundesland eine Stichprobe von Gemeinden zufällig ausgewählt wird. Innerhalb jedes Bundeslandes erfolgt die Ziehung der Gemeinden proportional zur Einwohnerzahl ("probability proportional to size", PPS).<sup>3</sup> Da innerhalb von Gemeinden mit Klumpeneffekten unbekannter Größe gerechnet werden muss, sollte die Zahl der Gemeinden vergleichsweise groß gewählt werden. Wir schlagen insgesamt 300 Gemeinden vor. Innerhalb einer Gemeinde werden jeweils 100 Personen untersucht.<sup>4</sup> Die Einwohnermeldestichprobe garantiert eine räumlich weite Streuung über die Gemeinde, pro Haushalt wird nur eine Person untersucht. Dies vermeidet weitere Klumpungseffekte, die bei anderen Formen der Auswahl und der Durchführung als Haushaltsstichprobe sonst unvermeidlich wären.

Die Größe der Stichprobe hängt von der geschätzten Prävalenz in der Population sowie von der gewünschten Irrtumswahrscheinlichkeit für den Fehler der 1. und den Fehler der 2. Art ( $\alpha$  bzw.  $\beta$ ) ab. Wählt man diese beiden Fehler wie üblich mit 0.05 und 0.2, dann benötigt man noch die vermutete Größe des Parameters (des Anteilswerts), den man eigentlich schätzen möchte. Nimmt man hierfür 1% an, dann liegt die Größe der Unsicherheit der Schätzung (das Konfidenzintervall) bei naiver Betrachtung bei 0.89–1.11%. Das ist für die meisten Zwecke ausreichend.

Möchte man zu einem späteren Zeitpunkt das Ergebnis der Stichprobe mit dem Ergebnis einer anderen Stichprobe vergleichen, z.B. wenn man eine Zunahme um einen bestimmten Wert nachweisen möchte, dann müssen die statistischen Schwankungen zwischen den Stichproben berücksichtigt werden. Möchte man eine Veränderung um 0.5% nachweisen, ergeben sich die Stichprobengrößen der Abbildung 1. Nimmt man eine Prävalenz von 3% an und möchte eine Zunahme um 0.5% nachweisen, werden in jeder Stichprobe mindestens 30.000 untersuchte Personen notwendig. Steigt der Anteil in der Population,

---

<sup>3</sup>Aufgrund der Probleme der Verbundgemeinden erfolgt die tatsächliche Auswahl von LAU-1 Einheiten ( $n=1268$ ), wobei in größeren LAU-1 Einheiten mehrere Gemeinden gezogen werden.

<sup>4</sup>Wir rechnen aufgrund der Bedeutung der Thematik und der öffentlichen Unterstützung mit einer geringeren Nonresponseproblematik (Schnell 1997) als in Deutschland derzeit üblich (Schnell 2019). Um eine Ausschöpfung von 30.000 tatsächlich untersuchten Personen zu erreichen, werden trotzdem mindestens 60.000 Personen kontaktiert werden müssen. Um die angestrebten 50% Ausschöpfung zu erreichen, empfehlen wir eine Incentivierung in Höhe von 10 Euro für jede untersuchte Person und eine intensive telefonische und schriftliche Nonresponse-Nachbearbeitung sowie eine lokale Öffentlichkeitsarbeit.

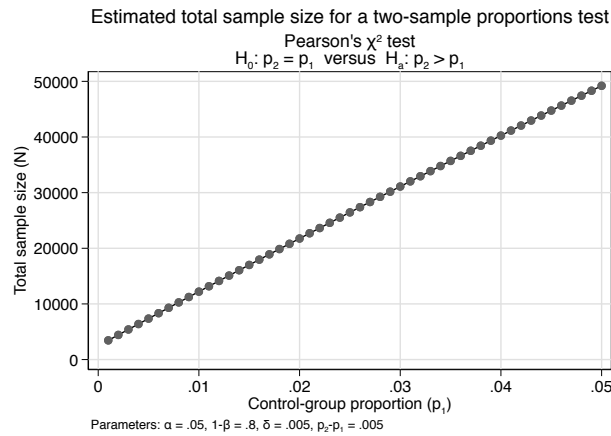


Abbildung 1: Stichprobengrößen für unabhängige Stichproben(Querschnitte)

vergrößert sich die notwendige Stichprobengröße. Angesichts der vielen unbekanntenen Größen erscheint uns eine Stichprobe der genannten Größenordnung unvermeidlich.

Für einen späteren Vergleich der Ergebnisse zu einem anderen Untersuchungszeitpunkt stellt sich die Frage, ob die Untersuchung als ein Längsschnitt oder mehrfache Querschnitte durchgeführt werden sollte. Bei einer Längsschnittuntersuchung werden dieselben Personen erneut untersucht, bei mehreren Querschnitten andere Personen. Zwar sind die Stichprobengrößen für einen Längsschnitt etwas kleiner (siehe Abbildung 2), aber der geringe Gewinn wird durch die Probleme einer erneuten Kontaktierung und vor allem des Aufwands für die untersuchten Personen aufgewogen. Wir empfehlen daher die Durchführung als Querschnittsuntersuchung.

## 1.2 Durchführung

Die Erstellung der Einwohnermelderegisterstichproben nimmt in Deutschland in der Regel mehrere Monate in Anspruch. Angesichts des besonderen öffentlichen Interesses nehmen wir an, dass sich die Stichprobe innerhalb von 6 Wochen in allen ausgewählten Gemeinden ziehen lässt. Erfahrungsgemäß gelingt dies ca. 10% nicht, daher müssen Ersatzgemeinden bereits zuvor ausgewählt werden. Damit müssen ca. 330 Gemeinden kontaktiert werden.

Die ausgewählten Personen werden angeschrieben und mit Informationen versehen. Dazu gehört auch ein schriftlicher Fragebogen zur Haushaltszusammensetzung, beruflicher Exposition und potentiellen Vorerkrankungen. Die ausgewählten Personen werden um Rücksendung der bereits adressierten und frankierten Unterlagen gebeten. Die Personen

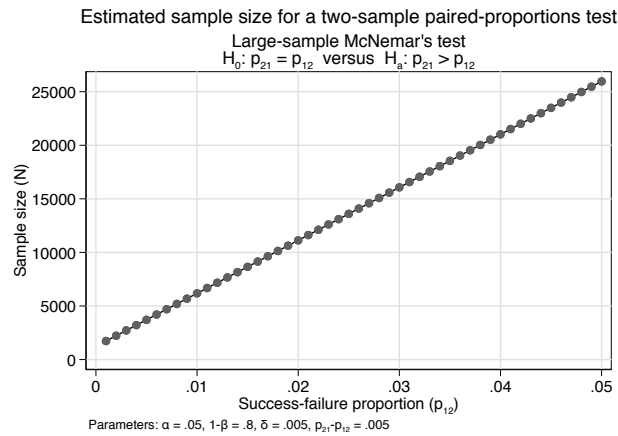


Abbildung 2: Stichprobengrößen für abhängige Stichproben (Längsschnitt)

werden aufgefordert, ihren Hausarzt aufzusuchen und um eine Blutentnahme zu bitten. Für den Fall, das eine Person keinen Hausarzt besitzt, wurde im Anschreiben ein nahegelegener Arzt mit Kontaktdaten benannt. Der die Blutprobe entnehmende Arzt übersendet die Blutprobe mit einer Identifikationsnummer an das Zentrallabor. Das Labor berichtet das Ergebnis an den Arzt. Der Arzt wird für seine Tätigkeit gemäß der Gebührenordnung entlohnt, der Patient erhält eine Aufwandsentschädigung.

Wir gehen von gesamten Fallkosten in Höhe von 75 Euro pro Fall aus. Bei einer bundesweiten Stichprobe mit 30.000 Fällen entspräche dies 2.25 Mio Erhebungskosten.

Sollte man eine Einschränkung auf ein Bundesland wünschen, so würden wir im Falle von NRW eine Stichprobe von 10.000 Fällen empfehlen. Die Ergebnisse wären dann nur auf NRW verallgemeinerbar, würden aber der großen Heterogenität der 396 Gemeinden in NRW Rechnung tragen. Die Zahl der ausgewählten Gemeinden sollte daher 100 nicht unterschreiten. Die Kosten der Stichprobe lägen dann bei 750.000 Euro.

### 1.3 Altersheime

Altersheime und Pflegeheime gelten als Hotspots für Infektionen. Um einen Überblick über das Infektionsgeschehen zu erhalten, erscheint die Durchführung einer getrennten Erhebung in Altersheimen sinnvoll. Geht man von ca. 11.500 Altenheime und Pflegeheime in Deutschland mit ca. 820.000 Personen in vollstationärer Pflege aus,<sup>5</sup> dann entspräche dies zwar lediglich ca. 0.985% der Bevölkerung, aber einem deutlich größeren Anteil der

<sup>5</sup>Destatis Genesis, auf der Grundlage der Pflegestatistik 2017.

mutmaßlichen Risikopopulation. 95.4% der Heimbevölkerung ist im Alter von 60 und mehr.

Da vermutlich Aussagen über Unterschiede zwischen den Ländern beabsichtigt sind, ist eine Schichtung nach Ländern unvermeidlich. Prinzipiell stehen zwei Möglichkeiten zur Ziehung zur Verfügung:

1. Ziehung über Melderegister oder
2. Ziehung über ein Heimverzeichnis.

Die Ziehung über die Melderegister würde keine Änderung des Stichprobenplans erfordern: Befindet sich eine gezogene Person in einem Heim, wird an dieser Anschrift eine Erhebung weiterer Fälle durchgeführt. Nur die Ausgangsperson ist und bleibt Bestandteil der Hauptstichprobe. Die Ausgangsperson und die Zusatzpersonen im Heim bilden die Heimstichprobe. Für die Hauptstichprobe entstehen keine weiteren Klumpeneffekte. Bei einem Bevölkerungsanteil von 0.985% werden in der Stichprobe im Mittel 295 Heimbewohner erwartet, wobei das 95% Konfidenzintervall (naiv) bei 254–330 liegt (99%: 252–342). In jedem Fall würden genügend Heime in die Stichprobe gelangen. Selbst bei Positivrate von 0.5% wäre das ungünstigste Konfidenzintervall ca. 0.25%–0.83%, was immer noch ausreichend erscheint.

Die Ziehung über ein Heimverzeichnis könnte analog der Ziehung im Altenheimsurvey (Klein und Gabler 1996) erfolgen. Proportional zu Heimbevölkerung eines Bundeslandes werden Heime proportional zur Größe der Heime gezogen.

Pro Heim wird stets die gleiche Zahl von Personen untersucht. Um die Klumpeneffekte und die Belastung des Personals nicht zu groß werden zu lassen, wäre eine Begrenzung der Fallzahl auf 10 Fälle ratsam.

Legt man die Größen vergleichbarer Studien zugrunde wären dies ca. 3200 Personen in 320 Heimen. Bei angenommenen Fallkosten von ca. 65 € pro Teilnehmer und 100 € Organisationskosten pro Heim ergäben sich Kosten von  $208.000 + 32.000 = 240.000$  €. Angesichts dieser Größenordnung wäre eine solche Studie in hohem Maß wünschenswert.

## **2 Datenbereitstellung**

Die Daten aus der Befragung und den Laborergebnissen werden zusammengeführt und als anonymisierter Mikrodatensatz aufbereitet und bereitgestellt. Die Veröffentlichung der Analyseergebnisse kann wenige Tage nach Abschluss der Feldarbeit erfolgen.

## Literatur

- Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (2014). *Stichproben-Verfahren in der Umfrageforschung: eine Darstellung für die Praxis*. 2. Aufl. Wiesbaden: Springer.
- Gabler, Siegfried und Jürgen H. P. Hoffmeyer-Zlotnik, Hrsg. (1997). *Stichproben in der Umfragepraxis*. Opladen: Westdeutscher Verlag.
- Kauermann, Göran und Helmut Küchenhoff (2010). *Stichproben: Methoden und praktische Umsetzung mit R*. Berlin: Springer.
- Klein, Thomas und Siegfried Gabler (1996). „Der Altenheimsurvey: Durchführung und Repräsentativität einer Befragung in den Einrichtungen der stationären Altenhilfe“. In: *ZUMA-Nachrichten* 20.38, S. 112–134.
- Klingwort, J. und R. Schnell (2020). „Critical Limitations of Digital Epidemiology: Why COVID-19 Apps Are Useless“. In: *under review*.
- Roche (2020). *Roche's COVID-19 antibody test receives FDA Emergency Use Authorization(...)* Techn. Ber. <https://www.roche.com/media/releases/med-cor-2020-05-03.htm>: press release, 3.5.2020.
- Schnell, R. (1997). *Nonresponse in Bevölkerungsumfragen: Ausmaß, Entwicklung und Ursachen*. Opladen: Leske+Budrich.
- Schnell, R. (2019). *Survey-Interviews: Methoden standardisierter Befragungen*. 2. Aufl. Wiesbaden: Springer VS.
- Schnell, R. und M. Smid (2020a). „Methodische Probleme und Lösungen für eine epidemiologische Corona-Forschung“. In: *Vortrag in der Reihe 'Soziologische Perspektiven auf die Corona-Krise', WZB Berlin*. 15. April 2020.
- Schnell, R. und M. Smid (2020b). „Methodological Problems and Solutions for Epidemiological Corona Research“. In: *under review*.