



Beschreibung des «Public use file» (PUF) zum Datensatz der Swiss National Cohort (SNC)

Neuchâtel, 2022

Herausgeber: Bundesamt für Statistik (BFS)

Auskunft: info.dem@bfs.admin.ch

Inhalt: Claudia Berlin ISPM, Universität Bern

Themenbereich: 01 Bevölkerung

Originaltext: Deutsch

Layoutkonzept: Sektion DIAM

Download: www.statistik.ch

Copyright: BFS, Neuchâtel 2022
Wiedergabe unter Angabe der Quelle
für nichtkommerzielle Nutzung gestattet

Beschreibung des «Public use file» (PUF) zum Datensatz der Swiss National Cohort (SNC)

Die Swiss National Cohort (SNC) ist eine bevölkerungsbasierte Kohorte der Schweiz. Die SNC basiert auf den Volkszählungsdaten der Jahre 1990 und 2000 sowie den registerbasierten Zensusinformationen, die bis 2018 u.a. mit der Todesursachenstatistik und den Emigrationsdaten verknüpft wurden. Aufgrund fehlender Personenidentifikatoren in den jeweiligen Statistikdatensätzen bis 2010 wurde probabilistisches Linkage verwendet, um die Datensätze miteinander zu verknüpfen. Seit 2010 enthalten die verwendeten Datensätze die AHV-Nummer (in pseudonymisierter Form), die für das Verknüpfen der verschiedenen Datensätze verwendet wird.

Der Hauptdatensatz der SNC umfasst derzeit mehr als 12 Millionen Personen und enthält zahlreiche Informationen zu jeder Person aus verschiedenen Datenquellen des BFS wie den Volkszählungen, STATPOP und BEVNAT. Die Hauptbereiche sind Angaben zur Person (z.B. Geschlecht, Geburtsdatum, Zivilstand, Staatsangehörigkeit, Bildung, Religion, Anzahl Personen im Haushalt, Wohngemeinde), und sofern die Person verstorben ist, auch Angaben wie das Todesdatum und die Todesursache sowie Begleiterkrankungen. Zahlreiche Variablen wiederholen sich, da sie für jeden Erhebungszeitpunkt vorhanden sind, aber im Laufe der Zeit ändern können. Andere Variablen wie z.B. Bildung und Religion wurden nicht immer erfasst und sind deshalb nur für gewisse Erhebungszeitpunkte vorhanden. Weitere Angaben beschreiben, seit wann eine Person in der Kohorte geführt wird und wann sie diese verlassen hat.

Der Hauptdatensatz der SNC enthält nur eine Auswahl an verfügbaren Variablen. Es besteht die Möglichkeit, zusätzliche Variablen aus den Volkszählungen (1990, 2000), den registerbasierten Zensusinformationen (seit 2010 jährlich verfügbar), dem Gebäude- und Wohnungsregister, der Statistik der Lebendgeburten, Todgeburten, Hochzeiten, Scheidungen sowie der Todesursachenstatistik hinzuzufügen. Auch Angaben aus den Strukturhebungen seit 2011 können für einen Teil der im Hauptdatensatz der SNC enthaltenen Personen ergänzt werden.

Das Ziel des Public Use Files ist es, einen ersten Eindruck des SNC Datensatzes zu vermitteln und enthält insgesamt eine Auswahl von nur 9 Variablen. Die Variablenkategorien wurden teilweise verändert, um mögliche Rückschlüsse auf Einzelpersonen auszuschließen. So wurden z.B. die Angaben zur genauen Todesursache in nur 3 Kategorien zusammengefasst und das Alter anstatt das Geburtsdatum verwendet. Als Referenzdatum für das Public Use File haben wir die Volkszählung 2000 gewählt. Somit besteht die Grundgesamtheit nur aus den 7 280 246 Personen, die im Zensus 2000 erfasst wurden und im SNC Hauptdatensatz enthalten sind.

Variablenbeschreibung

- id: nicht sprechende Personennummer
- age_start: Alter beim Eintritt in die Kohorte
- age_stop: Alter beim Austritt aus der Kohorte (durch Tod, Emigration oder Ende des Beobachtungszeitraums)
- sex: Geschlecht
- v0_civil: Zivilstand im Jahr 2000
- v0_nat_bin: Staatsangehörigkeit (Kategorie) im Jahr 2000
- gross_reg: Wohnsitz nach Grossregion im Jahr 2000
- v0_urban: Urbanisationslevel des Wohnsitzes im Jahr 2000
- cause_death: Todesursache