International Journal of Epidemiology, 2020, 367–367e
doi: 10.1093/ije/dyz185
Advance Access Publication Date: 20 September 2019
Cohort Profile

Cohort Profile

Cohort Profile Update: The Stockholm Birth Cohort Study (SBC)

Ylva B Almquist [1]*, Alessandra Grotta,[1] Denny Vågerö,[1] Sten-Åke Stenberg[2] and Bitte Modin[1]

[1]Department of Public Health Sciences, Centre for Health Equity Studies (CHESS), Stockholm University, Stockholm, Sweden and [2]Swedish Institute for Social Research, Stockholm University, Stockholm, Sweden

*Corresponding author. Department of Public Health Sciences, Centre for Health Equity Studies (CHESS), Stockholm University, SE-106 91 Stockholm, Sweden. E-mail: ylva.almquist@su.se

Editorial decision 29 July 2019; Accepted 19 August 2019

The original cohort

The Stockholm Birth Cohort Study (SBC) was set up in 2004/2005 to investigate how childhood living conditions and experiences are related to life-course health, occupation and income.[1] Briefly, the SBC was established by a probability matching of two longitudinal, anonymized datasets: the Stockholm Metropolitan Study (SMS) and the Swedish Work and Mortality Database (WMD).[1] The former dataset contains information on all individuals who were born in 1953 and lived in the greater metropolitan area of Stockholm in 1963 ($n = 15\ 117$). Register and survey follow-ups were carried out up until 1986 when the SMS was de-identified. The latter dataset consists of all individuals who were born before 1985 and lived in Sweden in 1980 and/or 1990 ($n = 9\ 428\ 526$). The WMD includes information on income, work, education, family formation, in-patient care and mortality for the period 1981–2009. The probability matching rendered it possible for researchers to follow 14 294 members of the SMS cohort from birth (1953) to the age of 56 (2009).

What is the reason for the new data collection?

The key that made it possible to perform updates of the WMD, and consequently of the SBC, was deleted in 2017. This deletion was decided upon by Statistics Sweden as part of adherence to the Personal Data Act of 1998, which only allows researchers to keep databases for a limited amount of time to study a set of beforehand-specified and ethically approved research questions. A new research program—'Reproduction of inequality through linked lives' (RELINK)—required a follow-up of the original cohort as well as addition of multigenerational linkages. In consequence, the entire procedure of probability matching with the SMS had to be re-done based on a new updated data register: RELINK53. RELINK53, created in 2017/2018, is defined as all individuals born in 1953 who lived in Sweden in 1960, 1965 and/or 1968 (hence, comprising the absolute majority of the members of the original SMS cohort), as well as their ascendant, contemporaneous and descendant family members ($n = 2\ 390\ 753$).

After RELINK53 had been created, it was linked to the SMS through a new probability matching in 2018/2019. This resulted in the establishment of the Stockholm Birth Cohort Multigenerational Study (SBC Multigen). Of the original 15 117 cohort members, 14 608 could be positively matched and thus included in the SBC Multigen. These individuals and their siblings constitute our Generation 1. For the cohort, there is also detailed sociometric information about childhood friendships between its members. SBC Multigen furthermore includes information about Generation 1's antecedents (Generation 0) and descendants (Generations 2 and 3), as well as the second

parent of the descendants in Generations 2 and 3. The structure of the data allows us to examine and contrast the life trajectories of individuals who share genetic and/or environmental factors.

The Regional Ethical Review Board in Stockholm approved the creation of RELINK53 as well as the probability matching to the SMS that resulted in SBC Multigen (no. 2017/34–31/5; 2017/684–32). Statistics Sweden, along with the other governmental agencies that were asked to provide data, approved of the new matching, extensions and data extractions.

What will be the new areas of research?

SBC Multigen enables us to examine how social, economic and health inequalities are produced and reproduced across four generations of Swedes (Generations 0–3). More specifically, we intend to study the inheritance of e.g. social class, educational attainment, income, poverty, physical health, mental health, substance misuse, out-of-home care, criminality, family formation and divorce. We are also able to investigate the extent to which the cohort members' sibships and friendships may contribute to expanded knowledge about these processes. As a first step, we will look more closely into sibling and friend interdependence for a wide range of outcomes, comparing patterns and trajectories of advantage and disadvantage at various stages of the life course. Second, we will investigate the structural characteristics of sibships and friendships, and the relative importance of these two types of network for developmental outcomes in childhood, adolescence and adulthood. In the third and final part, we will explore the role of siblings and friends for the intergenerational reproduction of inequality.

Who is in the cohort?

As described in a previous paper, the original SBC consisted of all individuals in the SMS for which an appropriate match could be found using the WMD.[1] In a similar way, the SBC Multigen consists of all individuals in the SMS for which an appropriate match was found using RELINK53. Similar to the one designed in 2004/2005, the procedure in 2018/2019 was based on a two-step matching algorithm.[2] However, this time we were able to add several new variables to increase the reliability of the matching and, moreover, RELINK53 did not suffer from the same restrictions as the WMD (i.e. that the individuals had to be alive and resident in Sweden in 1980 and/or 1990).

In a first step, a 21 variable key was created using two demographic variables (gender, month of birth); eight variables from Census 1960 (number of apartments in the building, overcrowding, year of construction of the building, quality of housing, occupational status of the head of the household, number of children to the head of the household/spouse living in the household, number of parents to the head of the household/spouse living in the household, total number of other household members); three variables from Census 1970 (county, municipality, quality of housing); two variables from Census 1975 (marital status, occupation); and six variables from Census 1980 (county, municipality, marital status, socio-economic position, employment status, occupation). We subsequently identified pairs of subjects that could be uniquely matched according to the key. Matched pairs were then validated by further comparing (i) month and year of death (since the mortality follow-up in the SMS ended in 1984, validation was only possible for deaths occurring up until this point) and (ii) parents' years of birth. We were here able to match and validate 13 880 out of the 15 117 members of the SMS cohort.

Since the percentage of missing values for Census 1960 variables was higher in the SMS compared with RELINK53, a reduced 13-variable key was created by using variables from Census 1975, 1980 and 1985. This key was then used in the second step to match SMS cohort members who had missing values on the Census 1960 variables. Uniquely matched pairs were validated in the same way as in the first step, resulting in 728 additional cohort members being matched.

Results from the linkage procedure are presented in Table 1. In sum, the matching procedure resulted in 14 608 of the original SMS cohort members being included in the SBC Multigen, of which 7447 are males and 7161 are females. We can follow this cohort from birth (1953) and, currently, up until 2018 (age 65).

To assess the extent to which the unmatched individuals represent a random subset of the SMS, we computed means and medians (for continuous variables) and proportions (for categorical variables) for some main baseline characteristics (birth weight, cognitive ability at the age of 13, mean school grades at the age of 15), father's income and occupation, and family's receipt of social assistance) for matched and unmatched individuals. We applied t-test and chi-squared test to compare means and proportions between matched and unmatched subjects, for continuous and categorical variables respectively. For father's income, we tested equality of medians through a non-parametric test. Results suggest that unmatched individuals had lower cognitive ability, lower grades and lower paternal socio-economic status during childhood compared with matched subjects, thus indicating a less favourable situation than those included in the final cohort (see Table 2).

Table 1. Results of the matching procedure of the Stockholm Metropolitan Study (SMS) with the RELINK53 data register

	SMS Cohort	Matched	Validated	Not validated	Unmatched
Match I					
(gender, month of birth + 19 variables from censuses 1960, 1970, 1975, 1980)					
Verified by date of death (month and year)					
Yes			14 075		
No				9	
Verified by parents' year of birth					
Two parent's year of birth given: identical in both data sets			13 302		
One parents' year of birth given: identical in both data sets			585		
Not verified due to inconsistent parental date of birth				90	
No year of birth given				107	
Total result, Match I	15 117	14 084	13 880	204	1033
Match II					
(gender, month of birth + 11 variables from censuses 1970, 1975, 1980)					
Verified by date of death (month and year)					
Yes			736		
No				1	
Verified by parents' year of birth					
One parent's year of birth given: identical in both data sets			701		
Two parents' year of birth given: identical in both data sets			27		
Not verified due to inconsistent parental date of birth				6	
No year of birth given				3	
Total result, Match II	1274	737	728	9	500
Total	15 117		14 608		509

Table 2. Comparison of means, medians or proportions for selected variables between matched and unmatched subjects

	Matched individuals ($n = 14\,608$)	Unmatched individuals ($n = 509$)	*P*-value
Birth weight (kg), mean	3.5	3.5	0.096
	0.54	*0.54*	
IQ in 1966 (summary raw scores), mean	68	65	0.003
	18	*18*	
School grades in 1968 (summary score), mean	3.18	3.04	0.001
	0.77	*0.79*	
Father's income in 1963 (SEK 000s), median	24	22	<0.001
	21	*16*	
Working class in 1963, %	38	38	0.901
Managers, large entrepreneurs etc. in 1963, %	17	12	0.002
Means-tested social benefit 1982–83, %	8	10	0.084
Death before 1984, %	1	10	<0.001

Standard deviations in italics. SEK, Swedish Kronor.

Comparison with the 2004/2005 probability matching

When compared with the matching procedure performed in 2004/2005, we observed that we were still able to match 14 035 (98.2%) out of the 14 294 SMS cohort members previously linked, and that we were able to match a further 448 (68.3%) out of the 656 subjects for whom an appropriate match was not found before. Among these 448 subjects, 223 (49.8%) had at least one year of residence abroad between 1978 and 1984. Thus, we were able to confirm that a sizeable part of the loss observed during the previous matching was due to emigration. Moreover, we were able to track 125 (74.9%) out of the 167 individuals who had been previously lost because of death occurring before 1980.

Moreover, we have repeated the analyses from two published studies based on the 2004/2005 matching, using the new probability matching but with the same study sample restrictions.[3,4] The results show only marginal differences in the estimates compared with the previous studies (details available upon request).

What has been measured?

Baseline information available for the members of the SMS has previously been described.[1] Follow-up of these individuals was possible through the probability matching to RELINK53. RELINK53 includes information from several national registers[5–7] connected through personal identification numbers, a unique identifier for Swedish residents.[8] Moreover, data from the Total Population Register and the Multigenerational Register facilitated the creation of the multigenerational structure.[5] Below, we present a brief outline of the available data sources.

From the Cause of Death Register, we can access information on the underlying cause of death for Swedish residents and Swedish citizens who died abroad (from 1952). The National Patient Register provides information on in-patient care (from 1964) as well as out-patient care (from 2001), whereas detailed information about cancer diagnoses is found in the Cancer Register (from 1958). Obstetrical data, including birth defects, are derived from the Medical Birth Register (from 1973).[9] RELINK53 also contains information from The Social Assistance Register (from 1985), entries in out-of-home care collected from the Register for Children and Youth (from 1968), as well as actions recorded in the Register for Municipal Health Care (from 2007). The above-mentioned registers are administered by the National Board of Health and Welfare.

Data on education, income, occupation and other indicators of socio-economic living conditions are derived from the Longitudinal Integration Database for Health Insurance and Labour Market Studies, LISA (from 1990). LISA is administered by Statistics Sweden, from which we also have access to the Educational Registers that contain information on school grades and year of graduation (from 1973). Information on sickness leave and benefits (from 1955), together with the underlying diagnoses (from 1994), is retrieved from the Social Insurance Agency (Försäkringskassan). Moreover, we can access information on crimes and convictions using the National Crime Register (from 1973), administered by the Swedish National Police Board. From the Swedish Defence Recruitment Agency, we have military conscription data (from 1968). Finally, RELINK53 includes the Population and Housing Censuses from 1960, 1965, 1970, 1975, 1980, 1985 and 1990.

The register data currently cover the years up until 2016–2018 (depending on the data source). Updates of the registers included in RELINK53 will be performed every second year, until 2024.

Table 3. Cohort members' ($n = 14\ 608$) vital/migration status up until 2017/2018 and number of individuals with ICD-10 diagnoses (in-patient care) from 1997 to 2016 (ages 44–63)

	Number of individuals	% of individuals
Cohort members		
Died before 11 February 2018	1446	9.8
Emigrated before 1 January 2017	1279	8.8
Diagnoses		
Certain infectious and parasitic diseases (A00-B99)	632	4.3
Neoplasms (C00-D48)	1918	13.1
Endocrine, nutritional and metabolic diseases (E00-E90)	517	3.5
Mental and behavioural disorders (F01-F99)	1024	7.0
Diseases of the nervous system (G00-G99)	612	4.2
Diseases of the eye and adnexa (H00-H59)	210	1.4
Diseases of the ear and mastoid process (H60-H95)	197	1.4
Diseases of the circulatory system (I00-I99)	1818	12.5
Diseases of the respiratory system (J00-J99)	816	5.6
Diseases of the digestive system (K00-K93)	2009	13.8
Diseases of the skin and subcutaneous tissue (L00-L99)	206	1.4
Diseases of the musculoskeletal system and connective tissue (M00-M99)	1595	10.9
Diseases of the genitourinary system (N00-N99)	1169	8.0
Pregnancy, childbirth and the puerperium (O00-O99)	40	0.3
Certain conditions originating in the perinatal period (P00-P96)	2	0.0
Congenital malformations, deformations and chromosomal abnormalities (Q00-Q99)	59	0.4
Symptoms, signs and abnormal clinical and laboratory findings, not classified elsewhere (R00-R99)	2007	13.7
Injury, poisoning and certain other consequences of external causes (S00-T98)	2101	14.4
External causes of morbidity and mortality (V01-Y98)	1643	11.3
Factors influencing health status and contact with health services (Z00-Z99)	854	5.9

What has it found? Key findings and publications

Cohort members—follow-up extension

Table 3 shows the cohort members' vital/migration status up until 2017 as well as the number of diagnosed individuals based on the 10th revision of the International Classification of Diseases (ICD), derived from data on in-patient care, covering the period from 1997 to 2016.

A four-generational data register

We have investigated the multigenerational structure of SBC Multigen (see Figure 1) and, apart from the cohort members born in 1953 ('A', $n = 14\ 608$) and their siblings ('C', $n = 28\ 592$), we have identified their mothers ('B', $n = 14\ 502$), fathers ('B', $n = 14\ 172$), partners ('H', $n = 13\ 112$), children ('D', $n = 24\ 929$), children's partners ('I', $n = 12\ 786$), grandchildren ('E', $n = 22\ 065$), siblings' partners ('J', $n = 26\ 807$), siblings' children ('F', $n = 51\ 509$), siblings' children's partners ('K', $n = 26\ 652$) and siblings' grandchildren ('G', $n = 47\ 466$). Regarding partners, it is worth noting that this refers to the child(ren)'s second biological parent(s).

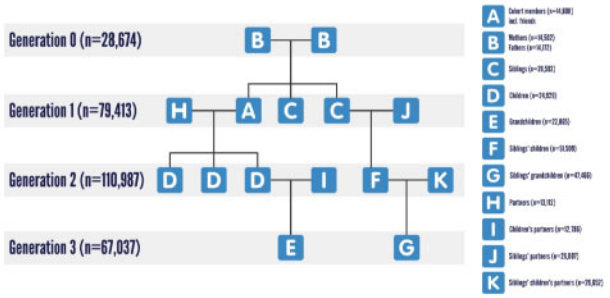

Figure 1. Description of the multigenerational data structure.

The total number of unique individuals included in this population amounts to 285 340. This number is not equal to the sum of the individuals in each of the previous categories, since some of the individuals are represented in more than one category.

Figure 2 illustrates the distribution of birth years for each of the four generations.

What are the main strengths and weaknesses?

Thanks to the creation of RELINK53, it is possible to follow the individuals included in the SMS from birth to retirement age. Complete follow-up on mortality, health and socio-economic living conditions is available. Assessment of all the outcomes is performed longitudinally, providing the data necessary to examine trajectories and sequencing of socio-economic and health status. Moreover, the availability of several types of information at the same point in time makes it possible to look at patterns of outcomes, adopting a person-oriented rather than a variable-oriented approach.[10]

The possibility to place life-course events in one generation within the broader social and historical context of four family generations represents the main novel feature of the updated cohort. Another rare feature is the sociometric data that enable us to map out the childhood-school class friendships between the cohort members. Thus, our

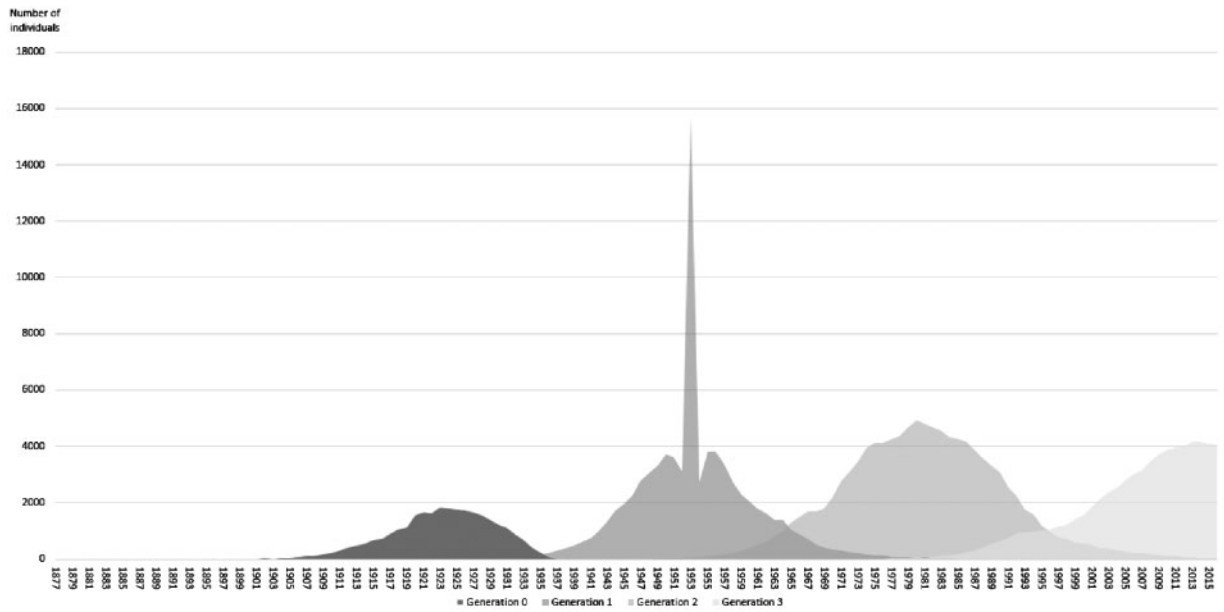

Figure 2. Distribution of birth years for Generation 0 ($n = 28\ 674$), Generation 1 ($n = 79\ 413$), Generation 2 ($n = 110\ 987$) and Generation 3 ($n = 67\ 037$) across the period 1877–2016.

data are an invaluable source of material to generate new knowledge on the extensively studied, but still not completely understood, mechanisms behind the reproduction of social, economic and health-related inequalities. To the best of our knowledge, there are currently no other multigenerational cohorts with such comprehensive baseline and follow-up data.

A limitation of the SBC Multigen is that the follow-up of the cohort members after 1986 is entirely based on register data and, for the descendants, there is only register information available. Although such information is often less biased by selection, it offers a rather aerial view of social, economic and health-related living conditions. Moreover, the cohort originates from the Stockholm metropolitan area which may restrict the generalizability of their—and their descendants'—life courses.

Can I get hold of the data? Where can I find out more?

The linked data files are kept at the Department of Public Health Sciences, at Stockholm University. They can only be accessed on site in Stockholm, in accordance with strict confidentiality agreements between Stockholm University, Statistics Sweden and the National Board of Health and Welfare. If you are interested in addressing a research question within our research programme and in using our data, a steering committee will evaluate your request and decide on data to extract for your purpose. For submitting a new project, please contact Y.B.A. (ylva.almquist@su.se).

Profile in a nutshell

- The Stockholm Birth Cohort (SBC), following a cohort of 14 294 individuals born in 1953, was established through a probability matching of two anonymous datasets in 2004/2005: The Stockholm Metropolitan Study (SMS) and the Swedish Work and Mortality Database (WMD).
- In 2017/2018, WMD was replaced by RELINK53 in order to extend the follow-up of the cohort and create a multigenerational data structure. The new probability matching between the SMS and RELINK53 resulted in The Stockholm Birth Cohort Multigenerational Study (SBC Multigen), established in 2018/2019.
- The SBC Multigen includes 14 608 of the original cohort members and their siblings (Generation 1), parents (Generation 0), children and siblings' children (Generation 2), as well as grand-children and siblings' grand-children (Generation 3). The population also includes the second parent of Generations 2 and 3.
- We have detailed information on the life course of the cohort members, from birth (1953) to retirement age (2018). The register follow-up, which is available for all generations, covers data on: causes of death, in-patient care, out-patient care, cancer, birth records, social assistance, out-of-home care, health care, education, income, occupation, housing, sickness leave and benefits, crimes and convictions, and military conscription.
- The linked data files are kept at the Department of Public Health Sciences, Stockholm University. If you are interested in addressing a research question using our data, a steering committee will evaluate your research proposal. For submitting a new project, please contact Y.B.A. (ylva.almquist@su.se).

Funding

This work was supported by the Swedish Research Council for Health, Working Life and Welfare (Grant No. 2016–07148).

Conflict of interest: None declared.

References

1. Stenberg S-Å, Vågerö D. Cohort profile: the Stockholm birth cohort of 1953. *Int J Epidemiol* 2006;**35**:546–8.
2. Stenberg S-Å, Vågerö D, Österman R, Arvidsson E, Von Otter C, Janson C-G. Stockholm birth cohort study 1953-2003: a new tool for life course studies. *Scand J Public Health* 2007;**35**:104–10.
3. Almquist Y, Modin B, Östberg V. Childhood social status in society and school: implications for the transition to higher levels of education. *Br J Sociol Educ* 2010;**31**:31–46.
4. Almquist YB, Jackisch J, Forsman H *et al*. A decade lost: does educational success mitigate the increased risks of premature death among children with experience of out-of-home care? *J Epidemiol Community Health* 2018;**72**:997–1002.
5. Ludvigsson JF, Almqvist C, Bonamy A-K *et al*. Registers of the Swedish total population and their use in medical research. *Eur J Epidemiol* 2016;**31**:125–36.
6. Ludvigsson JF, Andersson E, Ekbom A *et al*. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;**11**:450.
7. Barlow L, Westergren K, Holmberg L, Talbäck M. The completeness of the Swedish Cancer Register–a sample survey for year 1998. *Acta Oncol* 2009;**48**:27–33.
8. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol* 2009;**24**:659–67.
9. Cnattingius S, Ericson A, Gunnarskog J, Kallen B. A quality study of a medical birth registry. *Scand J Soc Med* 1990;**18**:143–8.
10. Bergman LR, Trost K. The person-oriented versus the variable-oriented approach: Are they complementary, opposites, or exploring different worlds? *Merrill-Palmer Q* 2006;**52**:601–32.