

Systematic review of variables used in NCDS and BCS70 adult sweeps and literature reviews of scientific contribution of cohort studies: Protocol

Dylan Kneale¹ (D.Kneale@ucl.ac.uk)

Aase Villadsen² (A.Villadsen@ucl.ac.uk)

Claire Stansfield (C.Stansfield@ucl.ac.uk)

James Thomas¹ (James.Thomas@ucl.ac.uk)

¹Evidence for Policy and Practice Information and Coordinating Centre, UCL Institute of Education, University College London, 20 Bedford Way, London, WC1H 0AL, United Kingdom

²Centre for Longitudinal Studies UCL Institute of Education, University College London, 20 Bedford Way, London, WC1H 0AL, United Kingdom

This protocol should be cited as:

Kneale D, Villadsen A, Stansfield C, Thomas J (2018) Systematic review of variables used in NCDS and BCS70 adult sweeps and literature reviews of scientific contribution of cohort studies: Protocol

London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

For further details, please contact:

Dylan Kneale, Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre), Social Science Research Unit (SSRU), UCL Institute of Education, University College London, 20 Bedford Way, London, WC1H 0AL, United Kingdom

tel: 0207 612 6020

email: D.Kneale@ucl.ac.uk

<http://eppi.ioe.ac.uk/>

1. Background

This protocol describes a series of reviews that will be carried with the purpose of:

- (i) Conducting a **systematic review** of cohort data usage on a set of variables/question areas collected at adult sweeps for the BCS70 and NCDS
- (ii) Undertaking a series of **literature reviews** on the scientific contribution of the NCDS and BCS70 studies on a number of priority areas
- (iii) **Background to the NCDS and BCS70 cohort studies**

NCDS: The 1958 birth cohort or the National Child Development Study (NCDS) began as a study of Perinatal Mortality focussing on just over 17 000 births in a single week in March 1958. Initially developed in response to concerns about persistent levels of stillbirth and neonatal deaths, the study has continued to collect data at important developmental milestones throughout childhood, adolescence and into adulthood and older age. Adult data collection occurred when cohort members were aged 23 (1981), 33, 42, 46, 50 and 55 years old. A further sweep of data collection is scheduled in 2018. The latest available sweep was collected in 2013, and collected data from 9,100 cohort members, two-thirds of whom participated through online surveys and the remainder through telephone surveys. Anonymised data from all the surveys are made available to the research community through the UK Data Service. Data have been collected on a wide range of topics on health and social circumstances including detailed partnership, housing, employment, childbearing, and educational histories; health data have been collected on a range of physical and mental health indicators including risk behaviours. Data have also been collected from cohort members' partners and, from a sample of the cohort, their children. Surveys have also been fielded to other subsamples of the cohort, including the Biomedical Survey at age 44 years (2002) when nurse-interviewers gathered measures of: vision, hearing, lung function, blood pressure, pulse; weight, standing and sitting height, waist and hip size and mental health; samples of blood and saliva were also taken facilitating genetic and genomic epidemiology. Some of these data, and particularly life course histories were collected with the intention of facilitating cross-cohort comparisons with the BCS70 study.

BCS70: The 1970 British Birth Cohort Study (BCS70) began as the British Births Survey and shared an original focus with the NCDS in exploring maternal characteristics associated with neonatal morbidity. Collecting data from birth through childhood and adolescence, sweeps of data were collected in adulthood at ages 26, 30, 34, 38, 42 and 46 years old. The latest sweep of data collected started in 2016 and is expected to finish in December

2017. The latest completed data collection (age 42) successfully engaged with over 9,800 cohort members from an original sample of over 17,000 cohort members. Wave non-response was notably high at age 26, when the survey was completed through a postal questionnaire, with approximately 9,000 returning a survey; the later 2008 sweep also had a low response rate with data collected in a short telephone survey from approximately 8,900 cohort members. Data have been collected on a wide range of topics including a wide range of health indicators, stabilities and discontinuities and the development of health inequalities, as well as detailed histories charting transitions to adult life including educational transitions, employment histories, housing data, partnership formation and dissolution, and childbearing and parental histories. The latest sweep is also collecting data in a nurse visit on cohort members' physical and cognitive health, and will also allow for the undertaking of genetic and genomic epidemiological studies. Data have been collected from cohort members, although the age 34 also collected data from the children of a subsample of the cohort members with other surveys fielded periodically to other subsamples of cohort members.

Aims

1. Conduct a systematic review of cohort data usage on a set of variables/question areas collected at adult sweeps for the BCS70 and NCDS
2. Undertake a series of literature reviews on the scientific contribution of the NCDS and BCS70 studies on a number of priority areas

3. Methods

3.1 Approach to conducting the review

This study will mirror the approach piloted in the earlier exercise exploring data usage in the Millennium Cohort Study (Kneale et al., 2016). For both sweeps, a systematic review of cohort data usage will be undertaken that prioritises the exploration of ten key variables or measures that have been collected at adult sweeps of both cohorts. Each systematic review will be produced through using a relatively simple search string that will be fielded across a number of databases. All included citations will then be screened initially on the basis of study and sweep relevance to create a comprehensive database of studies that use the BCS70 and NCDS adult sweeps using EPPI-Reviewer systematic reviewing software (Thomas et al., 2010). This comprehensive database will be further screened at a second stage to produce the full systematic review. The aim of each systematic review is to:

1. Identify studies that have used data from adult sweeps of the NCDS and BCS70 (only studies that included primary analyses of the data will be included) and carry out basic scientometric analyses of usage.
2. Identify studies that use target variables/measures as the focus of their analyses - i.e. those that use the target variables as the dependent or main independent variables of interest (studies that use the variables of interest as part of battery of controls in a regression analysis are not the focus of interest here)
3. Provide a descriptive account of the patterns of usage across the variables of interest and further scientometric analysis
4. Synthesise the published evidence with reference to specific research questions for selected areas of interest

Further details on the literature reviews is contained in the appendix.

3.2 Topic areas for review

The systematic review of topic areas aims to:

- (i) explore patterns of usage for approximately ten variables/measures/topic areas;
- (ii) identify features of measures/questions/variables that are associated with frequent or under-utilisation of the data and
- (iii) address targeted research questions based on these questions/variables.

While some of the variables or measures suggested may reflect relatively narrow domains, broader cohort study contributions are addressed through undertaking the accompanying literature reviews.

Priority areas and focussed research questions are included below:

Overall domain	Individual Measures	NCDS	BCS70	Notes	Focussed research question
Attitudes, participation and values	Political alignment and voting behaviour	☑	☑	These questions have considerable potential to explore longitudinal shifts in values and exploring age, cohort and period effects, although it is unknown the extent to which these questions have been used to these ends.	Are the cohort studies representative in terms of political affiliation?

Mental wellbeing	Mental wellbeing: WEMWBS	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Many of these measures have been fielded in successive surveys. Contrasting the usage of these measures can help to inform the design of future surveys in terms of fielding different instruments measuring related or similar underlying constructs.	The relationship between Mental health and employment: evidence from the 1958 and 1970 birth cohort studies
	Mental wellbeing: CASP	<input checked="" type="checkbox"/>			
	Mental wellbeing: GHQ	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
	Mental wellbeing: Malaise inventory	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
Housing Histories	Housing moves, tenure, cost, conditions.	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Housing tenure has been collected relatively consistently across adult sweeps. This analysis could explore the type of granularity needed in collecting tenure data in future surveys, as well as understanding the extent to which housing tenure is used as a focal variable (as opposed to part of a battery of controls in regression analyses).	
Physical activity	Frequency of exercise (focus on MVPA)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Both studies ask cohort members about partaking in exercise, which can be used in some cases to explore whether CMs meet PA guidelines. Differences in the way in which this information is asked can inform on the way in which future surveys should collect physical activity data.	Gender differences in exercise patterns during adulthood
Alcohol consumption	Alcohol consumption: CAGE measure	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	The way in which measures of problematic alcohol consumption is measured in both cohort studies changed from the CAGE scale to AUDIT after 2004. This analysis could examine whether	Changes in correlates of problematic alcohol consumption across adulthood
	Alcohol consumption: AUDIT measure	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		

				changes in the use of data on problematic alcohol consumption was observed, and provide insight into the implications of changing established measures.	
Age proofing the studies	Pensions and savings	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	The cohort studies represent an increasingly important source of information on ageing trajectories in the UK. Focussing on the way in which information is collected in these areas and understanding how this information is used and the perceived strengths and limitations could be useful for future survey design and ensure that the cohort studies are relevant for ageing researchers.	
Age proofing the studies	Adult cognitive assessments: vocabulary test in BCS70 and word-list recall exercises in NCDS	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
	Surveys of CM children: British Ability Scale	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Data have been collected from a subsample of CM children in both studies, although it is unclear the extent to which a smaller sample of children of diverse ages has posed a challenge in terms of usage among researchers. Exploration of this area also helps to understand the contribution of the cohort studies to the educational literature (an omission of the previous MCS research)	
	Imagine you are 60 qualitative essays	<input checked="" type="checkbox"/>		This was the second qualitative exercise fielded to NCDS survey members. Other studies have tended not to include similar exercises and	

				usage of these qualitative data are unclear, and whether their inclusion has expanded on the disciplines using NCDS data.	
--	--	--	--	---	--

3.3 How we will identify sources and further notes on synthesis

Inclusion criteria: Screening criteria will cover population (i.e. NCDS/BCS70 studies) and the sweeps(s) of interest (likely adult sweeps post-2001). The tightly defined population will facilitate the first round of title and abstract screening and the results will be used in initial scientometric analyses. Some further keywording may be possible. In the second round of screening, we will focus on studies where the variables of interest were variables were either the dependent variables or were the focus of the review (i.e. we will not seek out studies where the variables were used as control variables in models). Inclusion criteria around dates can also be imposed reflecting the year in which the particular sweep(s) of data were deposited. The title and abstract inclusion/exclusion criteria are outlined in Table 1.

Table 1: Inclusion/Exclusion criteria for title and abstract screening

<p>Exclude 1 - Duplicate</p> <p>Exclude 2 - not a 1958/70/WTCCC study <i>Exclude if not NCDS or BCS70 or Wellcome Trust Case Control Study (WTCCC)</i></p> <p>Exclude 3 - cohort profile <i>Exclude if publication about cohorts profile but no primary analysis</i></p> <p>Exclude 4 - No primary analysis <i>Exclude if review or systematic review with no primary analysis. Meta-analysis using results from existing studies go here. Meta-analysis with individual level pooled data is classed as primary analysis. Book of abstracts also go here.</i></p> <p>Exclude 5 - News item <i>Exclude if news item</i></p> <p>Exclude 6 - GWAS study <i>GWAS (genome-wide association study) - WTCCC not stated GWAS method studies also go here</i></p> <p>Exclude 7 - Previous Cohort Reviews</p> <p>Exclude 8 - Commentary or review of cohorts</p> <p>Include on NCDS 1958</p> <p>Include on BCS70</p> <p>Include for both NCDS and BCS70</p>

Use this also for when suspect a cohort study but unclear which e.g. British Birth Cohort in Abstract

Search protocol: To be developed with UCL Institute of Education's Information Scientist. We anticipate fielding a relatively simple search exhaustively across a number of databases to assemble a bibliographic database that can be interrogated further in later stages. Searches will be conducted across 65 databases, with individual search parameters developed for each. An example from the search on Scopus is recreated in the appendix.

Data collection and analysis: The relevant full-text studies will be retrieved but are not expected to be quality assessed unless addressing a specific research question.

For each priority measure/variable we will provide a descriptive account of:

- **Analytical conventions** including how the data were used (e.g. as dependent or independent variables), the type of analyses undertaken, the number and which sweeps were included, cross-cohort comparability, and where appropriate the granularity of the analyses
- **Reporting conventions** e.g. whether particular thresholds are used in the reporting of the data and whether the thresholds allow comparison with guidelines
- **Ascertain impact** based on estimated citation count and also in terms of capacity development (e.g. use of the data within PhD theses)
- **The main findings** and contribution to the literature
- **Strengths and limitations** of the measure/variable as described by authors
- **Strengths and limitations** of the overall data as described by authors
- **Further scientometric analyses** around types of publications, influential papers, collaborative teams and influential authors, institutions and impact across disciplines
- **Recommendations for future research** which could inform future survey development

Data synthesis: Synthesis of the results obtained from the data extraction will be conducted narratively. Further details of the scientometric analyses are outlined in the next section.

In the case of addressing more focussed research questions, our synthesis methods will extend the narrative, configurative approach and we will provide a descriptive account of the main recurring themes. We will follow five stages: (a) initial coding the text by producing preliminary textual descriptions of studies and their findings in a tabular format; (b) further inductive coding of the textual summaries and identifying key preliminary themes and their recurrence across studies; (c) developing a framework for

arranging groupings and clusters of studies according to the themes and exploration of these within and between the studies; (d) further generation of analytical themes through attempting to develop a common rubric to describe these findings; (e) consideration of the completeness and applicability of evidence, the robustness of the analysis methods and the quality of evidence in terms of its relevance to the research question.

3.4 Scientometric analyses of NCDS and BCS70 usage

Identification of all studies using NCDS and BCS70 data provides an opportunity to undertake scientometric analyses of the contribution of NCDS and BCS70 studies. Scientometric analysis involves the quantitative (and often semi-automated) analysis of publications (Serenko et al., 2010). Scientometric analysis can explore descriptive patterns around publication type (e.g. journal article, report, book chapter etc.), discipline, productive institutions, keywords and the most frequently published journals using BCS70 and NCDS data. It can build on descriptive patterns to create networks exploring networks of multinational collaborations or the most impactful institutions based on citation analysis (although analyses beyond descriptive analyses may not be possible for all grey literature not appearing on indexed databases). All citations will be screened first to exclude duplicate publications and those not meeting inclusion criteria based on adult sweeps of the cohort studies and using EPPI-Reviewer. Descriptive analyses will then be conducted which may progress to generating networks of authors, institutions, countries etc. using Table 2 Net, and networks will be visualised using Gephi and/or VOSviewer (all open source freeware). This initial stage of the review will:

- Create accurate estimates of the number of publications using adult sweeps of the cohort
- Create descriptive tables of information on publication type, discipline, year, etc.
- Aim to produce more comprehensive networks for a subsample of publications (i.e. those indexed on large databases) that allow for analyses of keywords, institutions, countries and citation counts/impact

3.5 A note on the production of the literature reviews: approach and focus

In addition to the systematic reviews, literature reviews are planned exploring the scientific contribution of the BCS70 and NCDS on a number of issues. Although not exhaustive in nature, these literature reviews have an added advantage of including influential studies that were conducted on childhood sweeps of the data alone. The topics will include:

- Family transitions (partnership formation and dissolution and childbearing)

- Attitudes, values, participation and identity
- Mental health and wellbeing
- Social mobility
- Health

These literature reviews aim to provide a thorough overview of the evidence uncovered by the cohort studies for each topic. Some additional literature may be included to help contextualise the contribution of the studies. Each literature review is intended to provide an accessible resource to academics and policy-making and non-academic audiences alike, through presenting a critical review of cohort study achievements and identifying potential future contributions. Included studies will span a breadth of disciplines, although to maintain focus on major scientific contributions, as well as to ensure that the reviews are produced within the scope of the project, included studies will be focussed on peer reviewed literature.

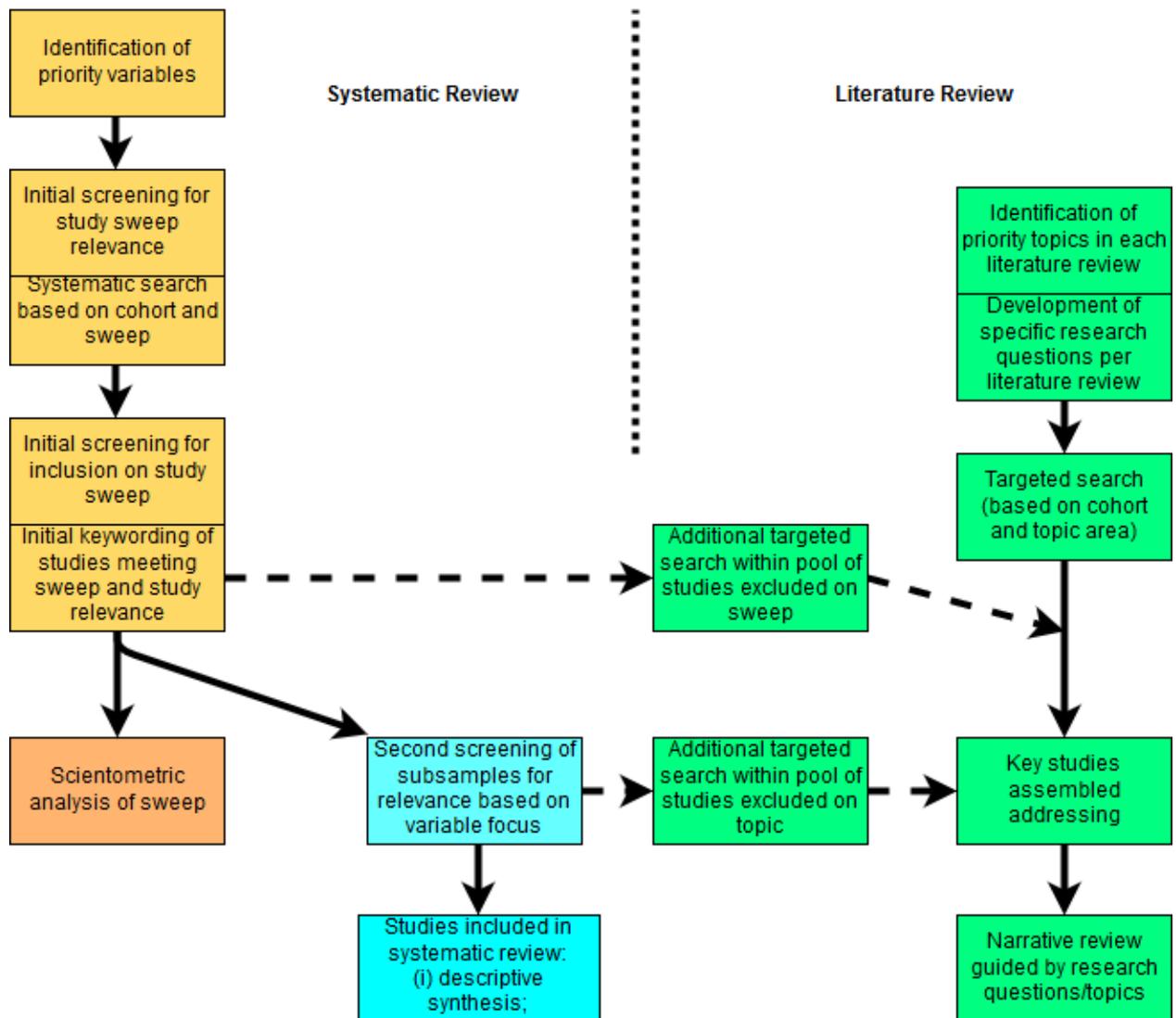
Each review will begin with a summary of which and when information supporting the topics have been collected in the NCDS and BCS70 studies. Production of each review is supported by senior subject experts who will provide input (informed by the summary of data that has been collected) around expected topics and usage, and potentially some known key outputs and papers that have emanated from the studies. This input will help to identify some of the key terms that could be included in searches and help to structure the reviews.

A basic search strategy incorporating study and subject limiters will be developed and operationalised on a small number of databases (target databases will include Scopus, PubMed, PsychInfo and Web of Science). Manual supplementary searches will be conducted through Google Scholar. Titles and abstracts will be imported into EPPI-Reviewer and screened for relevance and duplicates. Additional references identified through the systematic review (above) will also be added. The process of searching and screening will not be as comprehensive as is the case for a systematic review. Further details of the search protocol will be developed with UCL Institute of Education's Information Scientist once the data summaries and topics have been discussed.

In exploring the scientific contribution, these reviews will provide a critical narrative synthesis that 'takes stock' and evaluates what is of value from the body of work. The degree of conceptual innovation that studies contributed is of key interest in these reviews. The reviews will also aim to identify where the data have been used to attempt to resolve competing schools of thought. Key questions/criteria reflecting

conceptual innovation for each review will be developed and will be shared with the subject experts for input.

Figure 1: The link between the literature and systematic reviews



4. References

- GOUGH, D., THOMAS, J. & OLIVER, S. 2012. Clarifying differences between review designs and methods. *Systematic reviews*, 1, 1.
- KNEALE, D., PATALAY, P., KHATWA, M., STANSFIELD, C., FITZSIMONS, E. & THOMAS, J. 2016. Piloting and producing a map of Millennium Cohort Study Data usage: Where are data underutilised and where is granularity lost? . London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.
- SERENKO, A., BONTIS, N., BOOKER, L., SADEDDIN, K. & HARDIE, T. 2010. A scientometric analysis of knowledge management and intellectual capital academic literature (1994-2008). *Journal of Knowledge Management*, 14, 3-23.
- SNILSTVEIT, B., OLIVER, S. & VOJTKOVA, M. 2012. Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. *Journal of development effectiveness*, 4, 409-429.
- THOMAS, J., BRUNTON, J. & GRAZIOSI, S. 2010. EPPI-Reviewer 4.0: software for research synthesis. London: Institute of Education.

5. Appendix

Appendix 1 - Scopus search strategy

```
(( ( TITLE-ABS-KEY ( "National child development study" OR "National child development survey" OR "NCDS58" OR "1958 British Birth Cohort" OR "perinatal mortality survey" OR "longitudinal NCDS" OR "58C Cohort" OR "1958 cohort" OR "1958 national birth cohort" ) ) OR ( TITLE-ABS-KEY ( ( cohort* OR longitudinal ) AND 1958 AND ( "Britain" OR "British" OR uk OR u.k . OR "United Kingdom" ) ) ) OR ( TITLE-ABS-KEY ( ( "national child development" OR ( ncds AND NOT "non communicable" ) ) AND ( 1958 OR "Birth cohort*" OR "Born cohort*" OR "Britain" OR "British" OR uk OR "United Kingdom" OR "U.K." OR "longitudinal cohort*" OR "Population cohort*" OR "nationwide cohort*" OR "national cohort*" ) ) ) OR ( TITLE-ABS-KEY ( ( "birth cohort*" OR "born cohort*" OR "british cohort*" OR "UK cohort*" OR "U.K. cohort*" OR "longitudinal cohort*" OR ( ( "follow up" OR "followed up" ) AND ( british OR uk OR britain OR "United Kingdom" OR "U.K." ) ) AND 1958 ) ) ) OR ( FUND-ALL ( "National child development study" OR "National child development survey" OR "NCDS58" OR "1958 British Birth Cohort" OR "perinatal mortality survey" OR "longitudinal NCDS" OR "58C Cohort" OR "1958 cohort" OR "1958 national birth cohort" ) ) OR ( ALL ( "National child development study" OR "National child development survey" OR b58c OR bc1958 OR 1958bc OR "NCDS58" OR "1958 British Birth Cohort" OR "perinatal mortality survey" OR "longitudinal NCDS" OR "58C Cohort" OR "1958 cohort" OR "1958 national birth cohort" OR "1958 United Kingdom birth cohort" ) AND TITLE-ABS-KEY ( longitudinal OR observational OR cohort OR cohorts OR population ) ) OR ( ALL ( ( wtccc AND "cohort" AND 1958 ) OR ( "Wellcome Trust Case Control Consortium" AND "cohort" AND 1958 ) ) ) OR ( ALL ( ( wtccc AND "birth cohort" ) OR ( "Wellcome Trust Case Control Consortium" AND "birth cohort" ) ) ) ) )
```