DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 54, ARTICLE 18, PAGES 559–576
PUBLISHED 18 MARCH 2026
https://www.demographic-research.org/Volumes/Vol54/18/
DOI: 10.4054/DemRes.2026.54.18

Data Description

Harmonised fertility histories in four British longitudinal cohort studies

Aase Villadsen

Samantha Parsons

Alice Goisis

Contents

Demographic Research: Volume 54, Article 18
Data Description

Harmonised fertility histories in four British longitudinal cohort studies

Aase Villadsen[1]

Samantha Parsons[2]

Alice Goisis[2]

Abstract

BACKGROUND
Since World War II, Britain has witnessed significant societal changes, including in relation to fertility. Robust longitudinal and cross-cohort research requires data harmonisation to create comparable fertility measures to understand the predictors and consequences of these changes across generations.

OBJECTIVE
This paper describes newly created datasets on fertility histories that have been derived and harmonised across four longitudinal British studies of cohorts born between 1946 and 1990. The consistency with national statistics on fertility are examined.

METHODS
The birth cohorts are: National Survey of Health and Development, born in 1946; the 1958 National Child Development Study; the 1970 British Cohort Study; Next Steps, born in 1989/90.

RESULTS
The harmonised datasets include information on the cohort members' biological children at each survey sweep during childbearing age, such as whether they have children and if so how many, age of eldest and youngest child, number of boys and girls, and children by a previous partner. Additional variables related to non-biological children have been derived where possible. The percentage of female cohort members who have had at least one live birth and the number of children show a good degree of consistency with national statistics on fertility.

[1] University College London (UCL), London. UK. Email: a.villadsen@ucl.ac.uk.
[2] University College London (UCL), London. UK.

CONTRIBUTION
The harmonised datasets are well placed to improve fertility research across several generations born in Britain between 1946 and 1990. The possibility of linking these data with other variables from these cohort studies facilitates cross-cohort examination of fertility as an outcome, as a predictor, or as a covariate.

1. Introduction

Since World War II, Britain has experienced significant changes in fertility (Berrington, Stone, and Beaujouan 2015; Ermisch 2021), marked by the post-war baby boom, declining birth rates, and the postponement of childbearing from the 1970s. Advances in contraception, shifting gender roles, increased education, changes in social norms, economic uncertainty, and evolving family structures have contributed to these trends (Mills et al. 2011). Yet, many questions remain about the causes and consequences of these important demographic shifts and low fertility levels (Graham 2021). To address these knowledge gaps requires high-quality data, spanning several generations, that accurately captures fertility along with information on other aspects of individuals' lives.

Britain has a unique series of national cohort studies that follow the lives of multiple generations of people born in 1946, 1958, 1970, and 1989/90 and collect rich data on multiple domains across the life course, including fertility. Through interviews at different ages, cohort members (CM) have reported on each child they have had, usually including the date of birth, sex, and year of birth. However, in their raw format the fertility data are not straightforward to use and to compare across the different studies. There is a lack of harmonised summary measures on cohort members' fertility, such as whether they have had any children and, if so, at what age and how many across their entire reproductive lifespan. To overcome these limitations and facilitate research on fertility within and across the four British cohort studies, we have created new datasets containing harmonised fertility measures, which are now deposited and available to data users (UCL Centre for Longitudinal Studies 2025). The primary aim of this paper is to describe these new datasets. The secondary aim is to examine their consistency with official statistics on fertility.

2. Data and methods

2.1 The cohort studies

The source data were collected as part of four separate cohort studies on those born in Britain between 1946 and 1989/90, which have followed large samples of individuals across time – most since their birth. The following briefly outlines each of the cohort studies.

- **MRC National Survey of Health and Development (1946)** (Wadsworth et al. 2006). The NSHD is the oldest and longest-running British birth cohort study. It began with a maternity survey of 13,687 births in England, Scotland, and Wales during one week in 1946. A socially stratified sample of 5,362 singleton babies born to married parents was selected for follow-up, with the latest survey conducted at age 77.
- **National Child Development Study (1958)** (Power and Elliott 2006). The NCDS tracks over 17,000 people born in England, Scotland, and Wales in a single week in March 1958. Starting at birth, follow-up surveys have been conducted from age 7 through age 62.
- **British Cohort Study (1970)** (Sullivan et al. 2023). The BCS follows about 17,000 people born in England, Scotland, and Wales during one week in April 1970. The study began at birth, with follow-ups conducted from age 5 to 51.
- **Next Steps (1989/90)** (Wu et al. 2024). The youngest cohort with fertility data, this study follows around 16,000 young people in England who were in Year 9 (age 14) in 2004, born mainly in 1989 and 1990. Annual surveys ran until age 20, with further follow-ups at ages 25 and 32.

2.2 Why harmonisation of the fertility measures is necessary

Most interviews across the cohorts were conducted face to face, though some sweeps used postal, telephone, or mixed modes. Alongside extensive data on family background, education, health, and cognition, fertility information was collected from young adulthood onwards. However, the format and method of collecting fertility data vary both across sweeps within a cohort and between cohorts, making the data difficult to use for within- or cross-cohort analyses. For example, in the older cohort born in 1946 the early sweeps used the simple approach of asking cohort members whether they had any children and the total number. In a later sweep of the 1946 cohort, details of each individual child were also collected (age at birth, sex of child), and in subsequent sweeps cohort members reported any new children since the last survey participation. We see similar variation in methodological approach between sweeps in the 1958 and 1970 cohorts, where all children ever are reported in the first two fertility sweeps, and from the third fertility sweep onwards only new children are reported. The way the fertility data was collected requires data users to first add up all children reported in previous sweeps with the information on their characteristics collected in subsequent sweeps before adding the new children. In the 1989/90 cohort a more consistent approach was taken, which collected various information on each child that cohort members considered themselves a parent to in a dedicated child grid, which was then carried forward to age 32, when new children were added. However, common to all cohorts is that none contain derived variables at each sweep that tally up summary measures, such as whether cohort members have any biological children, at what age, and the total number they have had by the end of their reproductive lives (or last sweep for the 1989/90 cohort).

2.3 Harmonised fertility variables in the new datasets

By combining and recoding the available information on cohort members' fertility, we have created a harmonised longitudinal dataset for each cohort, with fertility data covering survey sweeps from when cohort members were young adults – generally in their early to mid-20s – until the early 50s, which for most marks the end of the reproductive window (or age 32 for the 1989/1990 cohort study). The derived variables provide a summary of the cohort member's fertility at each sweep, as well as at the end of their reproductive lives (or last available sweep for the 1989/90 cohort). The focus was on live births rather than terminated pregnancies, miscarriages, or stillbirths. As families have become increasingly complex across the cohorts, we enriched the datasets with harmonised information on the cohort members' non-biological (step, adopted, and foster) children residing in the household and partnership status, as these ready-to-use summary measures were also lacking in the cohorts. In contrast to the individual datasets, these new data resources offer data users ready-to-use, simplified, comparable (within and across cohorts), and comprehensive measures of the cohort member's fertility and related family variables.

The range of derived measures included in the harmonised datasets are listed in Box 1. There are four main groupings of variables, relating to: (1) biological children (live births), (2) non-biological children, (3) all children (biological and non-biological), and (4) children combined with partnership status. Additional variables are included, such as research case ID (to enable linkage to other variables from these cohorts), sex of cohort member, sweep participation, and weights to adjust for survey design in cohorts where this is applicable. We used the approach of deriving as many of these variables as possible in each cohort and their respective sweeps, rather than reducing the list to the lowest common denominator. The NSHD was especially affected by lack of information (particularly on non-biological children), whereas a larger number of harmonised variables could be derived in the NCDS, BCS, and Next Steps, which shared many identical survey questions. Coverage of the harmonised target variables across cohorts and sweeps is shown in Table 1. Also seen in this table is the availability of a number of variables that flag inconsistencies in the data. These inconsistencies varied by the cohort and between sweeps within cohorts as data collection methods for fertility changed over time. Flagged variables include where more biological children were reported as living in the household (reported in the household grid) than reported as pregnancies, or where study members reported having more biological children in a previous sweep than in a later sweep. The data guide provides further details of data inconsistencies for each of the cohorts (Villadsen, Parsons, and Goisis 2025).

In addition, we have made available the annotated Stata code for variable derivation for all cohorts on https://github.com/CLS-Data/Fertility-histories-in-four-UK-cohort-studies, which allows data users to inspect the derivation process, as well as adapt the code for creation of further and alternative fertility variables.

Box 1: List of harmonised fertility variables for each sweep in each of the cohorts

CM biological children
Whether has had any bio children
Number of bio children
Flag: More biological children reported at previous age than at current age
Number of bio children in HH
Flag: More bio children reported in HH grid than in pregnancy data
Number of bio children not in HH
Number of bio children had with a previous partner
Have had any bio children with a previous partner
Age in years of eldest bio child
Age in years of youngest bio child
Age in years of CM at birth of eldest bio child
Age in years of CM at birth of youngest bio child
Number of bio children who are boys
Number of bio children who are girls

CM non-biological children
Whether has any non-bio children in HH
Number of non-bio children in HH
Number of adopted children in HH
Number of fostered children in HH
Number of step-children in HH
Age in years of eldest non-bio child
Age in years of youngest non-bio child
Number of non-bio children who are boys
Number of non-bio children who are girls

CM biological or non-biological children
Whether has any children (bio or non-bio)
Number of children (bio or non-bio)
Age in years of eldest child (bio or non-bio)
Age in years of youngest child (bio or non-bio)
Number of children who are boys (bio or non-bio)
Number of children who are girls (bio or non-bio)

CM partner and/or children
Whether has a partner in HH
Marital status
Whether has live-in partner/spouse partner and/or any bio children
Whether has live-in partner/spouse partner and/or any bio or non-bio children

Other variables
CM research case identifier
Sex of CM
Birth year of CM
Birth month of CM
Whether CM took part in survey sweep
Interview year
Interview month
Survey weights (in applicable cohorts)

Table 1: Fertility variables available for each cohort, by sweep

Variable name	Variable label	NSHD								NCDS					BCS70							NS	
		Age 19	Age 20	Age 22	Age 26	Age 31	Age 36	Age 43	Age 53	Age 23	Age 33	Age 42	Age 46	Age 50	Age 26	Age 30	Age 34	Age 38	Age 42	Age 46	Age 51	Age 25	Age 32
sex*	sex of cohort member									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
cmbyear*	Birth year of CM									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
cmbmonth*	Birth month of CM									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
survey		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
intyear	Interview year				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
intmonth	Interview month				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
partner	Whether has a partner in HH									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
marital	Marital status	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
anybiochildren	Whether has had any bio children	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
biochild_tot	Number of bio children	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
biototal_flag	More biological children reported at previous age than at current age		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓
biochildhh_total**	Number of bio children in HH						(✓)	(✓)		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
biochild_extra_flag	Flag: More bio children reported in HH grid than in pregnancy data										✓	✓	✓	✓		✓	✓	✓					
biochildnonhh_total	Number of bio children not in HH									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
biochildprev_total	Number of bio children had with a previous partner										✓					✓	✓	✓	✓	✓	✓	✓	✓
biochildprevany	Have had any bio children with a previous partner										✓				✓	✓	✓	✓	✓	✓	✓	✓	✓
biochildy_eldest	Age in years of eldest bio child				✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
biochildy_youngest	Age in years of youngest bio child				✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
cmageybirth_eldest	Age in years of CM at birth of eldest bio child				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Variable name	Variable label	NSHD								NCDS					BCS70							NS	
		Age 19	Age 20	Age 22	Age 26	Age 31	Age 36	Age 43	Age 53	Age 23	Age 33	Age 42	Age 46	Age 50	Age 26	Age 30	Age 34	Age 38	Age 42	Age 46	Age 51	Age 25	Age 32
cmageybirth_youngest	Age in years of CM at birth of youngest bio child				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
biochildboy_total	Number of bio children who are boys				✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
biochildgirl_total	Number of bio children who are girls				✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
anynonbio	Whether has any non-bio children in HH									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
nonbiochild_tot	Number of non-bio children in HH									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
adopt_tot	Number of adopted children in HH										✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
foster_tot	Number of fostered children in HH									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
step_tot	Number of stepchildren in HH									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
nonbiochildy_eldest	Age in years of eldest non-bio child									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
nonbiochildy_youngest	Age in years of youngest non-bio child									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
nonbiochildboy_total	Number of non-bio children who are boys									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
nonbiochildgirl_total	Number of non-bio children who are girls									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
anychildren	Whether has any children (bio or non-bio)						✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
children_tot	Number of children (bio or non-bio)						✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
childy_eldest	Age in years of eldest child (bio or non-bio)									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
childy_youngest	Age in years of youngest child (bio or non-bio)									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
childboy_total	Number of children who are boys (bio or non-bio)									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Variable name	Variable label	NSHD								NCDS					BCS70							NS	
		Age 19	Age 20	Age 22	Age 26	Age 31	Age 36	Age 43	Age 53	Age 23	Age 33	Age 42	Age 46	Age 50	Age 26	Age 30	Age 34	Age 38	Age 42	Age 46	Age 51	Age 25	Age 32
childgirl_total	Number of children who are girls (bio or non-bio)									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
partnerchildbio	Whether has live-in partner/spouse and/or any bio children	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
partnerchildany	Whether has live-in partner/spouse and/or any bio or non-bio children									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
cflag	Mismatched information in anybiochildren and biochild_tot		✓				✓	✓															
cflag_19	Biological children reported at previous age(s), not at current age		✓	✓	✓	✓																	
cnflag	Fewer biological children reported at current age than previous age(s)				✓	✓	✓	✓															
cgflag	Mismatched information in anybiochildren and biochildgirl_total				✓	✓	✓																
cbflag	Mismatched information in anybiochildren and biochildboy_total				✓	✓	✓																

Note: In the datasets, variable names have a suffix that identifies the sweep. * Time invariant variable, included once in each dataset rather than for each age sweep. ** In the NSHD, total number of biological children in household (biochildhh_total) cannot be separated from non-biological children.

2.4 Sample sizes of harmonised datasets

The NSHD fertility dataset includes eight sweeps from ages 19 to 53 (1965–1999). Sample sizes range from 3,034 (males = 1,471; females = 1,563) to 3,897 (males = 2,021; females = 1,876).

The NCDS fertility dataset covers five sweeps from ages 23 to 50 (1981–2008). Sample sizes range from 9,532 (males = 4,643; females = 4,889) to 12,535 (males = 6,266; females = 6,269).

The BCS fertility dataset includes seven sweeps from ages 26 to 51 (1996–2021). Sample sizes range from 8,016 (males = 3,802; females = 4,210) to 11,260 (males = 5,471; females = 5,789).

The Next Steps fertility dataset uses surveys at ages 25 and 32 (2015 and 2022), with sample sizes of 7,705 (males = 3,427; females = 4,278) and 7,279 (males = 3,154; females = 4,125). Because this is a younger cohort, complete fertility histories cannot yet be constructed; future sweeps will provide the necessary data.

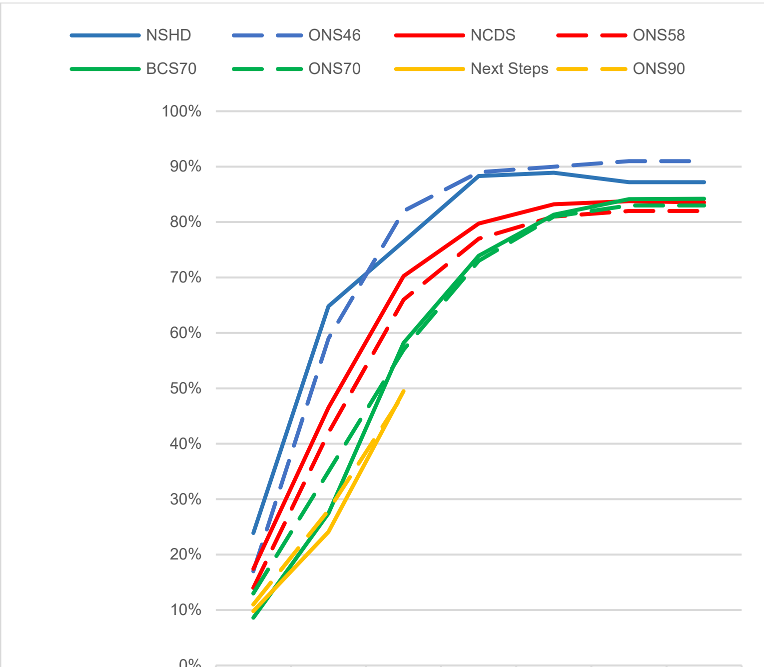
2.5 How our data compare to national statistics on fertility rates

To quality check the fertility variables in our datasets, we present comparisons between estimates from our derived datasets and those from national statistics on childbearing for all women in England and Wales born in the same years as our cohort members (Office for National Statistics 2024). Similar national statistics are not available for men (Dudel and Klüsener 2019) so these comparisons are limited to female cohort members. Specifically, we compared the proportion with at least one live birth, and the number of children.

For comparisons of any live births we use 5-year intervals from age 20 to 50. Although cohort members were not always interviewed at these specific ages, we are able to work out for any age whether they had had a child by using the derived variable on the age of cohort members at the birth of their eldest biological child. By using the same age points, we can not only compare fertility to national statistics but also observe trends across generations of cohorts. For number of children, we are not able to use a similar 5-year interval approach as our derived data only includes the number of children had by each age-specific sweep. Sampling weights are applied to all analyses of NSHD, as these are available across all survey sweeps and adjust for the sampling of only babies born to married babies in the follow-up surveys. Survey weights are not available in the NCDS or BCS, as these were nationally representative samples inviting all parents of babies born in one particular week in 1958 and 1970 respectively. In BCS, attrition weights are available at age 51 but are not applied to the analyses to retain consistency between sweeps and with the other cohorts. In the Next Steps analyses a weight is used to adjust for the sampling design, and although attrition weights are available these are not used in order to be consistent with the other cohorts. Researchers who wish to work with the fertility data should carefully consider the use of weights depending on whether analysis involves single cohorts or comparison across multiple cohorts.

Figure 1 shows the proportion of women with at least one live birth in the cohorts and in national statistics, by age. We generally see a good degree of consistency between the two, although with a tendency for the cohort figures to track slightly above those of national figures at most ages, except for Next Steps, where the reverse is seen. Figure 2 shows the number of children born to women in the cohorts and in national statistics, by age. Again, the figures are largely consistent, although there is a tendency for female cohort members to have slightly fewer children than in the national data, especially in the BCS, but also in the NSHD and NCDS in the later sweeps.

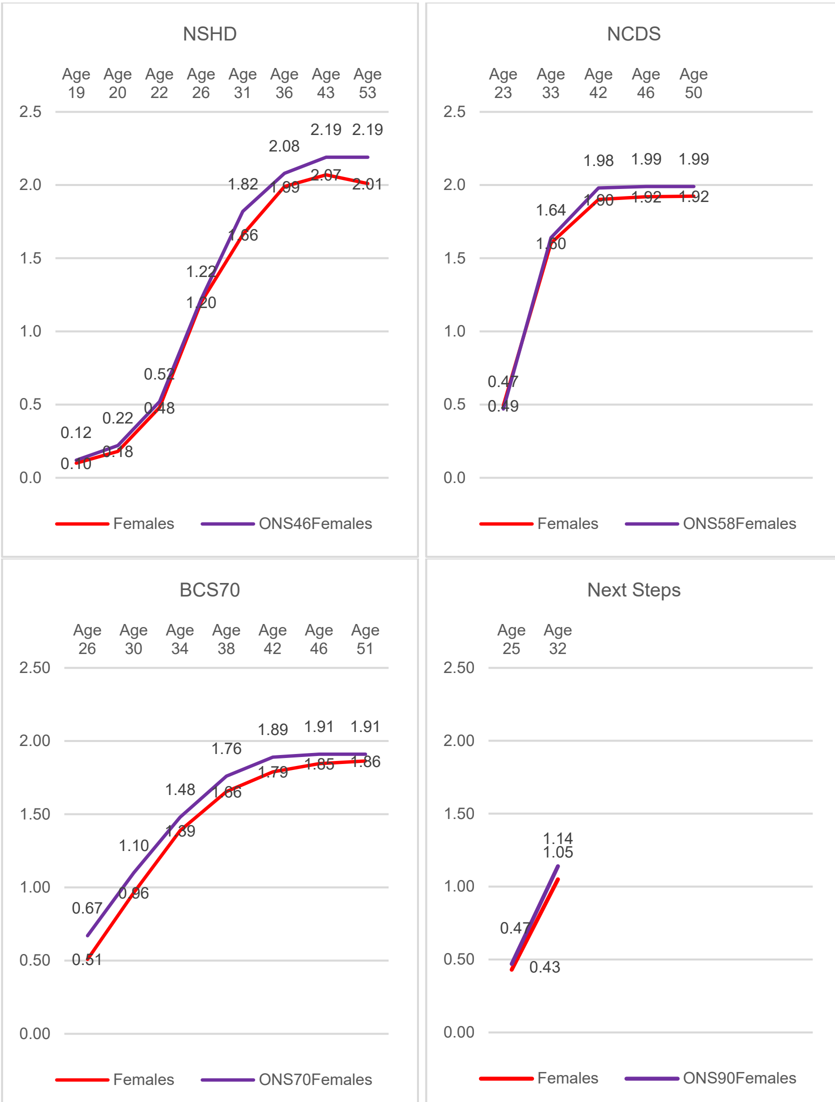
Figure 1: Percentage of women with at least 1 live birth, by age, comparing cohort data with ONS national statistics



	Age 20	Age 25	Age 30	Age 35	Age 40	Age 45	Age 50
NSHD	23.9%	64.8%	76.5%	88.3%	88.9%	87.2%	87.2%
ONS46	17.0%	59.0%	82.0%	89.0%	90.0%	91.0%	91.0%
NCDS	17.4%	46.5%	70.2%	79.7%	83.2%	83.8%	83.5%
ONS58	14.0%	42.0%	66.0%	77.0%	81.0%	82.0%	82.0%
BCS70	8.6%	27.4%	58.2%	73.9%	81.3%	84.1%	84.2%
ONS70	13.0%	35.0%	57.0%	73.0%	81.0%	83.0%	83.0%
Next Steps	9.8%	24.1%	49.5%				
ONS90	11.0%	28.0%	49.0%				

Note: Within each cohort, sample sizes vary across age points due to attrition over time. ONS statistics are published figures from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/datasets/childbearingforwomenbornindifferentyearsreferencetable.

Figure 2: Number of biological children born to women in the cohort datasets compared with ONS national statistics



Note: Within each cohort ,sample sizes vary across age points due to attrition over time. ONS statistics are published figures from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/datasets/childbearingforwomenbornindifferentyearsreferencetable.

Methodological differences between national statistics and the cohorts might contribute to the small differences we observe. National statistics include all women in England and Wales regardless of their birthplace being in the United Kingdom or abroad, whilst the cohorts are all born in Wales, England, or Scotland, with the exception of Next Steps, who were recruited around the age of 14 and 15 in England but could be born elsewhere in the United Kingdom or abroad. Attrition in the cohorts may also play a role, especially affecting respondents with lower socioeconomic status (Silverwood 2024), who tend to have a higher level of fertility (Berrington, Stone, and Beaujouan 2015). Indeed, in additional analyses in which we applied the available attrition weights in the BCS at age 51 (the only sweep which includes weights in BCS70), we see a higher number of biological children for females (2.02) than the unweighted number (1.86). Similarly, in Next Steps where an attrition weight is available, applying this results in a higher number of children born to females both at age 25 (0.57) and at age 32 (1.19), compared to results not weighted for attrition at age 25 (0.43) and age 32 (1.05).

2.6 Data availability and access

These new harmonised datasets can be accessed via the UK Data Service (UKDS). Users need to register with the UKDS, and data for NCDS, BCS, and Next Steps can be downloaded for free once the End User Licence has been accepted by the user. The NSHD fertility data can be downloaded once the Special Licence application form has been reviewed by UKDS and approved by the NSHD Data Sharing Committee. For the NCDS, BCS, and Next Steps, other data from these cohorts are also available via the UKDS and can be easily linked to the harmonised fertility datasets using individual IDs. Data users wishing to obtain additional variables for the NSHD alongside fertility data will need to apply for access via the MRC Unit for Lifelong Health and Ageing at UCL (LHA), which manages the NSHD. We refer to the data user guide for these datasets for full documentation (Villadsen, Parsons, and Goisis 2025).

3. Discussion and conclusion

In this paper we have presented and described new datasets that we have created and deposited for four British cohort studies, which include harmonised fertility variables. The aim of these new datasets is to facilitate the measurement and comparison of fertility within and across the four British cohort studies.

3.1 Strength and limitations

The strengths of these data include the large and longitudinal samples, allowing the examination of fertility across the lifespan in multiple generations. As demonstrated, the data have a good degree of consistency with national data on fertility. A significant advance is that the datasets include both males and females, unlike much existing data on fertility which is limited to women, and which has left a gap in our research and knowledge of men's fertility. Our derived fertility data can be linked with a rich range of other variables available in these cohorts, facilitating a broad range of research on fertility in single-cohort studies or in cross-cohort examinations. Three of the four cohorts are birth cohorts, so measures from early childhood can be linked with the fertility data to examine early predictors of fertility or as control variables. For those interested in cross-cohort examinations, harmonisation work on various other domains has been carried out with data made available to the research community on UKDS. Overall, these fertility data are valuable to further our understanding of late 20th century fertility trends, such as examining the early characteristics of those postponing childbearing or those who remain childless, or examining the consequences of fertility – for example, the impact of lifetime childlessness on health, wellbeing, and socioeconomic outcomes.

Our data also have some limitations that must be highlighted. Data on fertility is not always collected at the same age points in each cohort so this can limit some cross-cohort examinations. In addition, the coverage of our harmonised variables differs somewhat between the cohorts due to lack of information in the original surveys, especially affecting the older NSHD cohort where fewer questions were asked. Nor were we able to derive variables for number of children who had died, because children reported in a prior sweep were not asked about in a later sweep. Only non-biological children who were resident in the household in a given sweep were counted, as their information was obtained from the household grid/survey, and therefore we do not have a cumulative measure for these. Other limitations of our data relate to our approach of providing relatively simple summary measures at the level of the cohort members rather than for each individual child; e.g., we only include age of the cohort member at the birth of the eldest and youngest child. Researchers who wish to obtain information at the child level, including additional and specialised information that is available in some cohorts in some sweeps (e.g., contact with absent children and children's contact with non-resident natural parents) would need to revisit the raw data to draw out this specific information. The annotated code we have made available for deriving our datasets can be adapted to derive further fertility-related variables that meet the specific needs of researchers.

3.2 Conclusion

The harmonised fertility datasets on four birth cohorts in Britain improve the measurement of fertility (as an outcome, predictor, or control variable) in within-cohort analyses, and are well placed to facilitate cross-cohort research on fertility, such as understanding the drivers and consequences of changes over time in the timing and number of children born to women and men. Reassuringly, these data show good consistency with national statistics on fertility.

4. Acknowledgements

This work was undertaken as part of the Centre for Longitudinal Studies Resource Centre funded by the ESRC (ES/W013142/1). We would also like to thank our colleagues at Unit for Lifelong Health and Ageing at UCL for their collaboration around the NSHD data.

References

Berrington, A., Stone, J., and Beaujouan, E. (2015). Educational differences in timing and quantum of childbearing in Britain: A study of cohorts born 1940–1969. *Demographic Research* 33(26): 733–764. doi:10.4054/DemRes.2015.33.26.

Dudel, C. and Klüsener, S. (2019). Estimating men's fertility from vital registration data with missing values. *Population Studies* 73(3): 439–449. doi:10.1080/00324728.2018.1481992.

Ermisch, J. (2021). English fertility heads south. *Demographic Research* 45(29): 903–916. doi:10.4054/DemRes.2021.45.29.

Graham, E. (2021). Theory and explanation in demography: The case of low fertility in Europe. *Population Studies* 75(2): 133–155. doi:10.1080/00324728.2021.1971742.

Mills, M., Rindfuss, R.R., McDonald, P., and te Velde, E. (2011). Why do people postpone parenthood? Reasons and social policy incentives. *Human Reproduction Update* 17(6): 848–860. doi:10.1093/humupd/dmr026.

Office for National Statistics (2024). Childbearing for women born in different years [electronic resource]. Newport, Wales: Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/datasets/childbearingforwomenbornindifferentyearsreferencetable.

Power, C. and Elliott, J. (2006). Cohort profile: 1958 British birth cohort (National Child Development Study). *International Journal of Epidemiology* 35(1): 34–41. doi:10.1093/ije/dyi183.

Silverwood, R. (2024). Handling missing data in the CLS cohort studies. London: UCL Centre for Longitudinal Studies.

Sullivan, A., Brown, M., Hamer, M., and Ploubidis, G.B. (2023). Cohort profile update: The 1970 British Cohort Study (BCS70). *International Journal of Epidemiology* 52: e179–e186. doi:10.1093/ije/dyac148.

UCL Centre for Longitudinal Studies (2025). Harmonised fertility histories in the National Child Development Study, 1970 British Cohort Study and next steps, 1981–2024 [data collection]. UK Data Service.

Villadsen, A., Parsons, S., and Goisis, A. (2025). Fertility histories in four UK cohort studies: User guide (Version 1). London: UCL Centre for Longitudinal Studies.

Wadsworth, M., Kuh, D., Richards, M., and Hardy, R. (2006). Cohort profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). *International Journal of Epidemiology* 35(1): 49–54. doi:10.1093/ije/dyi201.

Wu, A.F.-W., Henderson, M., Brown, M., Adali, T., Silverwood, R.J., Peycheva, D., and Calderwood, L. (2024). Cohort profile: Next steps – the longitudinal study of people in England born in 1989–90. *International Journal of Epidemiology* 53: dyae152. doi:10.1093/ije/dyae152.