Defining and Identifying Only Children

Research decisions regarding the concept and measurement with survey data

CLS working paper number 2022/3

By Jenny Chanfreau and Alice Goisis
Disclaimer

This working paper has not been subject to peer review.

CLS working papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Any opinions expressed here are those of the author(s) and not those of the UCL Centre for Longitudinal Studies (CLS), the UCL Social Research Institute, University College London, or the Economic and Social Research Council.

How to cite this paper


Acknowledgement: This work was supported by the Economic and Social Research Council grant number ES/S002103/1.
Abstract

Despite increasing interest in the circumstances and outcomes of only children and increasing family complexity, the conceptualisation of only children has received limited scholarly attention. We raise issues involved in defining and identifying only children in social survey data, reflecting on the decisions researchers need to take. Illustrating the discussion with descriptive analyses of four British large-scale birth cohorts, we show it is possible to identify groups of individuals who correspond to different definitions of only children. Prevalence estimates obtained from the survey data show a similar trend over time as estimates from official fertility statistics. Researchers need to consider the choice of the most appropriate measure for a given research question, but data currently collected in many surveys limit how accurately any indicator of only children can reflect the chosen definition.
Introduction

With shrinking family sizes and one-child families a small but growing family form in many global north populations, including Southern European, East Asian and Eastern European countries (Breton and Prioux 2009; Frejka 2008; Frejka, Jones, and Sardon 2010), researchers have become increasingly interested in the circumstances and outcomes of only children (e.g. Baranowska-Rataj, Barclay, and Kolk 2017; Beaujouan and Solaz 2019; Keenan, Barclay, and Goisis 2022; Laybourn 1990; Mancillas 2006; Rainer and Siedler 2012). Considering this increase in scholarly interest, in the context of growing family complexity since the mid-20th century, there has been surprisingly little discussion of the conceptualisation of only children in the literature.

At first glance, identifying or defining an only child appears straight-forward, as the dictionary definitions of ‘Only child’ below indicate:

- a child who has no sisters or brothers (Cambridge);
- a person who has no siblings (Collins);
- a person who never had a brother or sister (Merriam-Webster).

Yet closer reflection reveals added complexity. Is siblingship about sharing both parents or either? Or is it the experience of growing up without other children in the home – including social (i.e. step, adoptive or foster) siblings? If so, what about social and half sibling(s) growing up in another household? The lived experience of being an only child or a sibling may differ over the life course. A large age gap, or never co-residing, may mean an individual effectively grows up experiencing an ‘only child’ childhood but in adulthood the relevance of having a shared parent may come
to the fore in matters relating to care (Chanfreau and Goisis forthcoming) or inheritance.

Identifying individuals in social surveys who are ‘only children’ thus requires decisions about which siblings to consider (full, half, social, co-resident/non-resident) and the conceptualisation guided by the research question and theoretical framework. Research testing whether only children’s social skills are detrimentally affected by growing up without siblings (Downey and Condron 2004) may favour a co-residence definition. Research testing whether only children benefit from a concentration of parental resources (Blake 1989; Downey 1995) might usefully reflect on which resources are of interest. A focus on parental time suggests a co-residence definition while an interest in financial resources might additionally account for half and step siblings living elsewhere. For research on care for ageing parents, the most appropriate definition might focus on being a given parent’s only since step siblings and half siblings who do not share that parent may not be expected to be involved in caregiving. In practice, decisions are often restricted by the information available in data.

In this article, we test whether only children can reliably be identified using four British large-scale birth cohorts. Although the analyses focus on specific UK datasets, we reflect on the decisions researchers need to take when identifying only children using any social survey data and discuss how the issues raised are neither unique to the UK context nor these specific datasets.
Data and possible definitions

We analyse data from four British cohort studies, the design of which is a distinct strength for identifying only children and analysing their characteristics and circumstances. They have collected detailed maternal fertility history information and/or documented the residence of children other than the cohort member in the childhood home, and they have also followed cohort members through to adulthood, with the longest-running cohort now covering the life course from birth to age 70. Unlike many surveys of adults, which do not ask respondents if they have siblings but focus instead on own children and/or members of the household at the time of data collection (see for example Understanding Society or European Social Survey), the birth cohorts allow the researcher to identify only children using childhood data whilst also making it possible to investigate their outcomes beyond childhood.\textsuperscript{1}

The National Survey of Health and Development (NSHD; Douglas, Wadsworth, and Kuh 2015), has followed a subsample of the individuals born in a given week in 1946 (5,362 single births of the initially surveyed 13,687 births). The 1958 National Child Development Study (NCDS; CLS 2020) and the 1970 British Cohort Study (BCS70; Butler, Bynner, and CLS 2016) follow cohorts of initially approximately 17,000 people born in a particular week in 1958 and 1970, respectively. The 2001 Millennium Cohort Study (MCS) has regularly surveyed a representative sample of nearly 19,000 individuals born between September 2000 and January 2002 (CLS 2017).

For each cohort, we outline two methods for identifying only children which are as comparable as possible across the datasets: the co-residence and the shared

\textsuperscript{1} For Stata code that can be used to derive an indicator to identify only children in these datasets please see Goisis, Alice and Chanfreau, Jenny (2022). Identifying Only Children in Four British Birth Cohort Studies, 2022. [Data Collection]. Colchester, Essex: UK Data Service. 10.5255/UKDA-SN-855087
mother definition. The first is intended to capture the experience of growing up without other children in the home while the latter draws on fertility history data commonly used in demographic research. The 1946 cohort study does not include twins/triplets but the other cohorts do and for both definitions we treat cohort members who are twins/triplets as having siblings.

*Co-residence* – In the 1946 cohort, siblings living with the cohort child are identifiable through a household grid at the time the cohort member was aged 15. This method may misclassify those who are the youngest child in the household as only children if older siblings have left the family home. In the 1958 cohort, siblings are identifiable through direct questions at age 16 about older and younger siblings with the same mother (the wording implies living anywhere). Although the question might also fit under the ‘shared mother’ heading, given that children tend to grow up living with their mother (even if the parents separate) we consider it under the ‘co-residence’ method as it is likely siblings reported under this question shared the maternal home with the survey child. In the 1970 cohort, we used reports of any older or younger siblings in the age 10 sweep household grid, including step or adoptive siblings. In the 2001 cohort, we included any full, half, step or adoptive siblings reported as living in the household at any sweep up to age 11.

*Shared mother* – The 1946 cohort included questions regarding the interval between the birth of the survey child and the mother’s previous and next births respectively, only children are thus identified by the mother reporting that neither birth interval is applicable. In rare cases of sibling stillbirth or neonatal death, this method might misclassify a cohort member as having grown up with a sibling. In the 1958 cohort data we define as only children those who are the mother’s first child and without
subsequent live births reported by the mother at the age 11 survey. For the 1970 cohort we consider as only children cohort members whose mother reported no prior live (and surviving) births and no subsequent births by the time the cohort member was aged 5. The study did not collect maternal fertility histories at the age 10 sweep but did ask the parent respondent (96% were the mother) about any other family members in the household or living elsewhere and their relationship to the cohort member. We update the only child measure to age 10 based on younger siblings reported by the natural mother, on the assumption that these are her own children. In the 2001 cohort we identify only children based on the mother reporting neither any other prior or subsequent live births nor own children living elsewhere, at any sweep up to age 11.

The objective in presenting these two methods is twofold. First, we gauge the accuracy of the identification of only children in the survey data by comparing the prevalence estimates obtained from the surveys with estimates from official statistics. Although not directly identifiable, we can derive comparable estimates of only children using cohort fertility statistics by birth year of the woman published by the Office for National Statistics (ONS 2020) for women born approximately a generation prior to the survey cohorts. To compare the official statistics with the survey estimates, we selected the years of birth that matched the mean maternal age at first birth (MAFB) for births occurring in the years corresponding to the survey birth cohorts (ONS 2019). The ONS data provides a useful alternative source to survey data to estimate the prevalence of only children as it is based on official birth registrations.
The second objective for comparing the two definitions is to assess whether the proportion of children identified in the surveys differs depending on the definition used or data available. Drawing on additional detail available in the 2001 cohort, we first show how the prevalence of individuals without siblings differs across ages in childhood depending on whether the analyst has access to cross-sectional or longitudinal data. Then we disaggregate the two methods introduced above to demonstrate that the strengths or limitations of a given definition/measure depend on the research focus. The objective is not to present an exhaustive list of definitions of only childhood, nor to recommend a particular measure, but to illustrate a variety of possibilities and highlight some of the decisions researchers need to reflect on in researching only children.

Results

Table 1 sets out the prevalence estimates of only children in each study, using the two methods of identification: the co-residence and the shared mother definitions. The prevalence of only children was highest in the 1946 cohort, lower among those born in 1958 and 1970 and somewhat higher again among those born in 2001.

The ONS data on the proportions of women who had one child only confirm the trend of first decreasing prevalence, from 21% of women born in 1920 to 16% and 14% among women born in 1933 and 1946, followed by an upturn more recently to 18% of women born in 1974. Adjusting the ONS statistics to account for the proportion of women in each cohort without children, gives an estimated percentage of mothers with one child only, comparable with the survey estimates of the percentage of first-borns who are only children. Finally, adjusting the ONS statistics for the average
family size among women with children in each cohort, gives an estimated prevalence of only children.

Overall, the survey estimates are broadly of a comparable magnitude to the ONS fertility data. Although the prevalence estimates of only children for the 1946 cohort are higher based on the survey data (13.6%) than based on the ONS data (11%) and in the 1970 cohort are higher for the ‘shared mother’ definition in the survey data (10.5%) compared with ONS data (6%), for the other cohorts and the co-residence definition in the 1970 cohort the prevalence estimates are very similar for the two data sources.\(^2\) Despite the limitations involved when estimating only children based on the ONS data, we are reassured by the similarity of the general pattern over time as well as the similarity of the substantive magnitude of the estimates derived from the surveys and the ONS data respectively that it is possible to accurately identify only children in the British birth cohorts.

\(^2\) Maternal age at birth amongst women who had an only child was 3-4 years higher, on average, than MAFB among women who had more than one child (results not shown). As a sensitivity analysis, we estimated the results using MAFB obtained from the 1958, 1970 and 2001 surveys (instead of ONS estimates) which allows us to obtain separate estimates between women who had an only child and women who had more than one child. The ONS estimated prevalence rates of only children were highly similar when using MAFB obtained using survey data: 8%, 6% and 10% for the 1958, 1970 and 2001 cohorts respectively when using MAFB for women who had an only child compared with 7%, 7% and 9% when using MAFB among those who had two or more children. This sensitivity analysis is not possible for the 1946 cohort study since maternal age is only provided in 5-year categories.
Table 1 Prevalence estimates of only children at age 10/11 from four cohort studies and ONS fertility estimates

<table>
<thead>
<tr>
<th>Surveys:</th>
<th>Only Child: % of all children</th>
<th>Only Child: % of first-borns</th>
<th>Co-residence definition</th>
<th>Shared mother definition</th>
<th>Co-residence definition</th>
<th>Shared mother definition</th>
<th>Co-</th>
<th>ONS Data</th>
<th>Mean age at 1st birth</th>
<th>% of women with 1</th>
<th>% of mothers with 1</th>
<th>Only as % of children</th>
</tr>
</thead>
<tbody>
<tr>
<td>1946</td>
<td>13.6</td>
<td>33.1</td>
<td>13.6</td>
<td>31</td>
<td>13.6</td>
<td>31</td>
<td>13.6</td>
<td>4,154</td>
<td>1920</td>
<td>26</td>
<td>21</td>
<td>27</td>
</tr>
<tr>
<td>1958</td>
<td>6.3</td>
<td>15.9</td>
<td>6.8</td>
<td>19.0</td>
<td>6.8</td>
<td>19.0</td>
<td>6.8</td>
<td>13,606</td>
<td>1933</td>
<td>25</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>1970</td>
<td>6.9</td>
<td>16.5</td>
<td>10.5</td>
<td>25.0</td>
<td>10.5</td>
<td>25.0</td>
<td>10.5</td>
<td>13,836</td>
<td>1946</td>
<td>24</td>
<td>14</td>
<td>15</td>
</tr>
</tbody>
</table>

Notes: ¹ Survey data includes England, Scotland and Wales (1946, 1958, 1970) and England, Scotland, Wales and Northern Ireland (2001). 2001 cohort weighted using Sweep 5 analytical weight for whole-UK analyses to adjust for sample design and attrition; other cohorts unweighted. ² ONS data birth year of mother chosen based on the average maternal age at first births occurring in the years of the cohort studies.
The relative similarity of the two sets of survey prevalence estimates shows there is considerable overlap between the co-residence and shared mother definitions. The exception being that the ‘shared mother’ method in the 1970 cohort appears to overestimate the prevalence of only children (at 10.5%, compared with 6.9% based on the co-residence method and 6% based on ONS data). This is likely due to the study not updating the detailed maternal birth history at the age 10 sweep and the derivation relying on other questions to update the only child indicator to age 10.

We may interpret the broadly similar estimates for the two survey methods as indicating that sibling status derived from maternal fertility history is a reasonable proxy for co-residence if detailed household grid information is unavailable (and vice versa). This may be context specific, as in the UK children usually reside with their mother, and therefore also with the siblings with whom they share a mother. However, lack of detail also partly limited the match between conceptual definition and its operationalisation (e.g. the 1958 cohort siblings question specified shared mother rather than co-residence).

Because the discussion of the co-residence derivation indicates that timing matters in the operationalisation, we next explore how the age at data collection might affect the identification of only children in survey data. Conceptually only childness can be thought of as a stable characteristic; an individual growing up without siblings is an only child from birth and throughout the life course. However, especially for analysts using household grid information in cross-sectional data, the age of observing an absence of siblings should balance minimising the chance of subsequent births of younger siblings against the likelihood of older siblings still co-residing. Drawing on the 2001 cohort with detailed household grid information at every data collection
sweep, we compare the cross-sectional estimates derived solely using information collected at the given sweep with a longitudinal derivation that also uses information collected at prior sweeps (Figure 1). Here we identify only children as those without either co-resident (full, half or social) siblings or siblings with the same mother but living elsewhere. The initial proportion of only children (at age 9 months) relates to singleton first-borns without co-resident social siblings, approximately 41% of the cohort.

**Figure 1 Percentage of only children at different ages using the MCS (2001 cohort)**

![Graph showing percentage of only children at different ages using the MCS (2001 cohort)](image)

Notes: Cross-sectional n 18,550 (9m); 15,589 (3y); 15,246 (5y); 13,857 (7y); 13,286 (11y); 11,725 (14y); 10,622 (17y). Longitudinal n 7,838. All analyses weighted for non-response. * Indicates sweeps where the mother was asked about own children living elsewhere.

After the first sweep, the proportion of children without siblings declines with age as younger siblings are born, with the largest reductions seen up to age 7. At each age, the cross-sectional estimate results in a higher prevalence of only children, but the difference between the two methods is relatively small up to age 11. In adolescence, the cross-sectional estimate of the prevalence of only children rises again, indicating
misclassification of youngest siblings whose older siblings have left the home as ‘only children’. Considering siblings reported at prior sweeps means the proportion of only children based on the longitudinal method decreases with age, but the change is minimal after age 11 as age gaps of more than 11 years are rare. In fact, among firstborns who had a co-resident sibling at the age 14 sweep less than 2% had an age gap of 12 years or more to their (oldest) younger sibling.

Taken together, although resulting in somewhat higher prevalence estimates than derivations based on longitudinal information, cross-sectional data collected between the ages of about 7 and 11 result in relatively more accurate identification of only children than data collected at younger or older ages.

Table 2 Definitions and characteristics of only children using the MCS (2001 cohort)

<table>
<thead>
<tr>
<th>Only child definition</th>
<th>No full sibling</th>
<th>Co-residence</th>
<th>No full/half siblings</th>
<th>No full/half social siblings</th>
<th>Shared parent</th>
<th>Father's only child</th>
<th>No known siblings</th>
<th>Sample overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>Gender: Girl</td>
<td>19.9</td>
<td>9.6</td>
<td>9.1</td>
<td>9.2</td>
<td>11.8</td>
<td>7.6</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Low income (9 months)</td>
<td>47.5</td>
<td>48.1</td>
<td>48.5</td>
<td>48.9</td>
<td>48.4</td>
<td>48.9</td>
<td>48.5</td>
<td></td>
</tr>
<tr>
<td>Mother's education:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree</td>
<td>28.3</td>
<td>36.2</td>
<td>36.8</td>
<td>36.7</td>
<td>37.0</td>
<td>37.6</td>
<td>38.1</td>
<td></td>
</tr>
<tr>
<td>Parental social class:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Managerial/professional</td>
<td>28.8</td>
<td>37.6</td>
<td>38.0</td>
<td>38.1</td>
<td>36.6</td>
<td>38.2</td>
<td>38.7</td>
<td></td>
</tr>
<tr>
<td>Both parents in</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>household (age 11)</td>
<td>31.6</td>
<td>39.6</td>
<td>40.0</td>
<td>41.2</td>
<td>43.3</td>
<td>40.2</td>
<td>60.6</td>
<td></td>
</tr>
<tr>
<td>Number of siblings</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.6</td>
<td>0.0</td>
<td>1.6</td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1 Here a small number of foster siblings are included under the umbrella term social sibling, along with step and adoptive siblings, although their co-residence with the cohort member is likely to be shorter. 2 Reported by biological father while co-resident with the cohort child, excludes siblings fathered by non-resident fathers. 3 No co-resident siblings or siblings with a shared (co-resident) parent living elsewhere. 3 Includes all co-resident (full, half or social siblings) and siblings with a shared parent living elsewhere at age 7 (the latest sweep that records non-residential siblings).

Finally, we examine the two definitions discussed previously in further detail in Table 2, again drawing on the 2001 study. First, we show three separate ‘co-residence’
definitions based on having no full siblings, no full or half siblings, or no full, half or social siblings in the household. Second, we use household grids in each sweep and additional questions about parents’ biological children living elsewhere to derive separate shared parent definitions for the mother and father. Finally, we combine the information into an indicator of having no siblings either co-resident or with a shared parent. All groups are compared to the overall sample, i.e. all cohort members in the age 11 sweep irrespective of sibling status. For each definition, we report the prevalence as well as headline socio-demographic characteristics and the number of siblings.

Definition (1), no full siblings, results in the highest prevalence estimate of only children at nearly 20%. Because this definition disregards co-resident half and social siblings as well as (full or half) siblings living elsewhere, only children in this group have on average 0.7 siblings. Compared with the overall sample and the other only child groups, this definition captures a disproportionately disadvantaged group. Therefore, research using data with information only about full biological siblings will likely over-estimate the size of the only child group and, at least in the UK, focus on a negatively selected sub-group. This might have implications when the outcomes of only children are compared to the outcomes of those growing up with siblings.

Definitions (2), (3) and (4) all result in similar prevalence estimates of only children, ranging between 9% and 10%. Except for the lower proportion of children living with both parents at age 11, the characteristics of these groups reflect the overall sample. Despite the definitional differences, siblings (identifiable in the data) are well accounted for under each version. Since 97% of the 11-year-olds in the sample
overall lived with their mother, this similarity across definitions reiterates the point regarding shared residence/mother overlap in the UK.

Definition (5) uses information about other children fathered by the biological father of the cohort child, resulting in an only child prevalence estimate of nearly 12%. The information is collected while the father and cohort member co-resided, so any siblings reported as living elsewhere are most likely half siblings from the father’s previous relationship. If the father later left the survey household, any siblings he fathered in a subsequent relationship are not included. The definition clearly disregards siblings since this group has 0.6 siblings on average, more than the shared mother definition (4). This occurs because it neglects co-resident half or social siblings following maternal re-partnering. Definition (6), results in the lowest prevalence estimate of only children, at below 8%. By definition, this version excludes all identifiable siblings irrespective of shared residence or parent(s), and thus in some sense is the ‘purest’ derivation of only children. Nonetheless, depending on the research question this may or may not be the most appropriate definition to use. Importantly, since the cohort member is likely to continue living with the mother after parental separation, all definitions overlook non-resident siblings following paternal re-partnering. Depending on the research question, and the extent of shared care, this omission might affect the results.³

A key finding is that, among children born around 2001 in the UK, being an only child is strongly related to parental separation. Across all our definitions, only children are

---

³ In the overall sample among 11-year-olds living with their mother, 67% had some contact with the non-resident father and among those who did: 63% had contact at least weekly and 73% reported overnight stays (including 40% staying overnight ‘often’). If there are social or half siblings living with the non-resident father, children who are their mother’s only child may nonetheless grow up with the social experience of having siblings.
much less likely to reside with both biological parents at age 11 than is the case in the sample overall (32-43% compared with 61%). Therefore, the lack of information about half and social siblings living elsewhere after the father has left the survey household highlights the need for data to reflect the complexity of family life to allow for flexible definitions of only children, and accurate derivation of measures that reflect the relevant definition.

Conclusion

In this paper we have presented different definitions and derivations of only children based on survey data using four British large-scale birth cohorts. We emphasise that the choice of the most appropriate measure is not merely a question of data quality and accuracy but also of conceptual fit for a given research focus. Tailoring the definition to the research question, can improve our understanding of whether, and if so why, only children’s outcomes and experiences differ from those of individuals who grow up with siblings.

However, data limitations clearly impinge on how accurately any derivation can reflect the chosen definition. We suggest the data issues discussed here are neither specific to the UK context nor to the datasets we have used for illustrative purposes. For example, the Survey of Health, Ageing and Retirement in Europe (SHARE) includes retrospective information about respondents’ co-residential siblings at age 10, separating (full) biological siblings from half, adoptive, step and foster siblings (included as a single category), thus the issue of identifying the shared parent and lack of information about non-resident siblings may be relevant for analysis.
More complete parental fertility histories, including follow-up after separation, and/or taking a network approach to collecting family information, would better reflect the increasing complexity of family life and thus improve the flexibility and quality of measures to match any chosen definition of only children. The limited information about (both biological and social) fathers’ fertility means there is likely an under-counting of both younger half siblings and older step siblings living elsewhere. This suggests that the prevalence of only children that can currently be derived with survey data may be somewhat overestimated, although the extent to which this is a limitation depends on the research question and definition used. As shared care post-separation becomes more common, data limitations regarding both the existence of and the amount of contact with non-resident siblings might become increasingly problematic.

The analysis presented in this paper shows it is possible to accurately identify groups of individuals who correspond to different definitions of only children based on information about co-residence and/or maternal fertility history collected at or up to around age 10/11. Researchers using cross-sectional data collected in childhood around this age may thus be reasonably confident about the accuracy of identifying only children using household grid information. It may also be appropriate to specify this age in retrospective questions about the presence of siblings. The comparison of the prevalence estimates obtained from the survey data with official cohort fertility statistics reassuringly show a similar trend over time. Nonetheless, as family life continues to diversify some changes to data collection practices would improve not only the identification of only children but also our understanding of the variety of sibling experiences among individuals with siblings.
References


