
# An Evaluation of the Childhood Family Structure Measures from the Sixth Wave of the British Household Panel Survey

**Marco Francesconi**
*ISER*
*University of Essex*

**ISER Working Papers**
**Number 2002-25**

# Institute for Social and Economic Research

*The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates the following centres:*

- ESRC Research Centre on Micro-social Change. Established in 1989 to identify, explain, model and forecast social change in Britain at the individual and household level, the Centre specialises in research using longitudinal data.

- ESRC UK Longitudinal Centre. This national resource centre was established in October 1999 to promote the use of longitudinal data and to develop a strategy for the future of large-scale longitudinal surveys. It was responsible for the British Household Panel Survey (BHPS) and for the ESRC's interest in the National Child Development Study and the 1970 British Cohort Study

- European Centre for Analysis in the Social Sciences. ECASS is an interdisciplinary research centre which hosts major research programmes and helps researchers from the EU gain access to longitudinal data and cross-national datasets from all over Europe.

The British Household Panel Survey is one of the main instruments for measuring social change in Britain. The BHPS comprises a nationally representative sample of around 5,500 households and over 10,000 individuals who are reinterviewed each year. The questionnaire includes a constant core of items accompanied by a variable component in order to provide for the collection of initial conditions data and to allow for the subsequent inclusion of emerging research and policy concerns.

Among the main projects in ISER's research programme are: the labour market and the division of domestic responsibilities; changes in families and households; modelling households' labour force behaviour; wealth, well-being and socio-economic structure; resource distribution in the household; and modelling techniques and survey methodology.

BHPS data provide the academic community, policymakers and private sector with a unique national resource and allow for comparative research with similar studies in Europe, the United States and Canada.

BHPS data are available from the Data Archive at the University of Essex
http://www.data-archive.ac.uk

Further information about the BHPS and other longitudinal surveys can be obtained by telephoning +44 (0) 1206 873543.

*The support of both the Economic and Social Research Council (ESRC) and the University of Essex is gratefully acknowledged. The work reported in this paper is part of the scientific programme of the Institute for Social and Economic Research.*

**Acknowledgement:**
I thank John Ermisch, who got me started on this research, Nick Buck, Sarah Cheesbrough, Heather Joshi, Heather Laurie, Peter Lynn and John Rigg for advises, comments and suggestions. I also thank Judy Egerton for excellent research assistance. Financial support from the Economic and Social Research Council is gratefully acknowledged.

**Readers wishing to cite this document are asked to use the following form of words:**

> **Francesconi, Marco (November 2002) 'An Evaluation of the Childhood Family Structure Measures from the Sixth Wave of the British Household Panel Survey', *Working Papers of the Institute for Social and Economic Research*, paper 2002-25. Colchester: University of Essex.**

For an on-line version of this working paper and others in the series, please visit the Institute's website at: http://www.iser.essex.ac.uk/pubs/workpaps/

Institute for Social and Economic Research
University of Essex
Wivenhoe Park
Colchester
Essex
CO4 3SQ UK
Telephone: +44 (0) 1206 872957
Fax: +44 (0) 1206 873151
E-mail: iser@essex.ac.uk
Website: http://www.iser.essex.ac.uk

**ABSTRACT**

This paper performs an evaluation of the data collected in the sixth wave of the BHPS on childhood family structure. After comparing such data with a large number of studies using external sources, we find that the BHPS overestimates the proportion of people who report an experience of life in a non-intact family during childhood by about 12%. Although an explanation based on recall error that deteriorates with the age of the BHPS informants is possible, this overestimation is likely to be accounted for by non-ignorable attrition that may affect most of the comparison studies using longitudinal data. Conversely, comparisons with other independent measurements from the BHPS itself reveal that the wave-6 data underestimate the proportion of young people having experienced part of their childhood in a non-intact family by about 8%. The probability of disagreement between these two statistics is strongly associated with poor interview characteristics, which may affect the comparison measure more than the wave-6 measure. In general, however, despite such differences, there is a substantial degree of similarity between the family structure information collected in the sixth wave of the BHPS and the host of highly diverse records against which it has been compared.

## 1. Introduction

This paper undertakes a statistical evaluation of the data collected in the sixth wave (1996) of the British Household Panel Survey (BHPS) on family structure during childhood. Two methods have been implemented for this exercise. The first uses a pairwise record match to independent measurements of family structure. For this purpose, we employ data from the birth cohort studies (e.g., the National Child Development Study), official statistics and other national or special surveys (e.g., the General Household Survey and the Avon Longitudinal Study of Parents and Children). The second method uses independent measurements from the BHPS itself.

The main motivation for this work lies in the increasing interest that researchers and policy makers have devoted to the issue of family structure during childhood and its impact on children's early and later outcomes. Hypotheses drawn from a number of social sciences and concerned with the potential effect of childhood family structure on children's success or failure have been systematically tested using several different data sources. However, what we know is disproportionately based on data from the United States (Haveman and Wolfe, 1994; McLanahan and Sandefur, 1994; Duncan and Brooks-Gunn, 1997; Mayer 1997; and references therein). In Britain, data limitations have somewhat constrained the number of contributions, although we have recently observed a growing number of studies in this area (for a survey of studies up to the middle of the 1990s, see Rodgers and Pryor (1998); for more recent studies, see, among others, Hobcraft (1998), Gregg and Machin (1999), Joshi et al. (1999) and Ermisch and Francesconi (2001a)).

The 1946 National Study of Health and Development (NSHD) has generated several early contributions concerned with the effects of family dissolution on health-related and behavioural outcomes (e.g., Douglas and Blomfield, 1958). But it has only been after the rise of divorce rates through the 1970s and 1980s that we have witnessed an extensive interest in intergenerational links through childhood family experiences in all realms of social research. For

this purpose, the 1958 National Child Development Study (NCDS) and the 1970 British Cohort Study (BCS) have featured prominently (see Appendix A for more information on these surveys). The large sample sizes and the rich set of information on several life aspects are unambiguous advantages of the three cohort studies. For more recent cohorts, however, the BHPS data are likely to be a better reflection of contemporary trends of family structure. In fact, the incidences of lone parenthood and stepfamilies are now much higher than among parents of the 1946 or 1958 cohorts and, perhaps, even among those of the 1970 cohort. In addition, as discussed in many existing studies, childhood family structure is a complex variable to construct with data from the birth cohort surveys (e.g., Ferri (1976) and Hobcraft (1998)). On the contrary, the variables collected by the BHPS and analysed here can be readily used.

Independent records coming from data sources other than the cohort studies are by and large less systematic. We contrast the BHPS data with official divorce statistics for England and Wales (Haskey, 1997) as well as with 6 other national or special (local) surveys, that is, the OPCS Omnibus Survey, the Avon Longitudinal Study of Parents and Children (ALSPAC), the West of Scotland Twenty-07 Study, the data from the age/sex registers of eight general practices in North and East London, the General Household Survey (GHS), and the National Survey of Sexual Attitudes and Lifestyles (NSSAL). (Appendix A contains a brief description of such data sources).

The empirical strategy used in this paper is straightforward. In section 2, we present the childhood family structure variables collected in the sixth wave of the BHPS. They provide us with a set of estimates of the proportion of children who were born between 1894 and 1980 (of which 90% were born after 1920 and 50% after 1950) and experienced life in a non-intact family by their sixteenth birthday (note that 1894 is the year of birth of the eldest individuals in wave 6 who report valid information on family structure, while 1980 is the year of birth of the 16-year olds in 1996).

Section 3 discusses some methodological issues that are relevant in evaluating the reliability of our data, the results of which are reported in Section 4. This exercise is based on comparisons with independent estimates reported in a selection of studies employing the data sources mentioned above. In this comparison, we try to include as many studies as possible from a number of social sciences. In this way, agreement about the reliability of the BHPS estimates is likely to diminish, but we possibly widen the perspectives within which such measures ought to be assessed, and eventually collected in future waves. Section 5 reports the results from another evaluation, which is performed using data that come from a special BHPS sample of young adults. These are selected from the first seven waves (1991-1997) and can be matched with their mother in at least one survey year over the same period. Such data have already been used to study intergenerational links between family structure and several child outcomes in early adulthood (e.g., Ermisch and Francesconi (2001a) and Ermisch et al. (2002)).

Not only do the BHPS family structure measures refer to a nationally representative sample of individuals from several birth cohorts, but they are also extremely easy to extract and to analyse. Therefore, if they are shown to provide us with reliable information, more substantive research will be likely to stem from this source in the near future. This is even more pressing, given that the BHPS will gather again family structure information in future waves.

## 2. The childhood family structure measures in the sixth wave of the BHPS

For the first time in wave 6 (1996) of the BHPS, respondents were asked three questions concerning their family structure during childhood. The questions are:

D40. *Did you live with both your biological mother and biological father from the time you were born until you were 16?* (variable = FBPAR16)

D41 (asked if D40=no*). How old were you when you stopped living with both your biological parents?* (variable = FLVHMAG)

D42 (asked if D40=no). Showcard 12. *Please take a look at this showcard and tell me who you were living with when you were aged 14?* The showcard specifies the following mutually exclusive possibilities: natural mother and father, adoptive mother and father, mother and stepfather, father and stepmother, mother and no father figure, father and no mother figure, grandparent(s), aunt/uncle/own siblings, other relative, step parent, other non relative, institution, employer's premises, living independently, other (variable = FWHR14).

It should be noted that the BHPS does not collect information on the *reason* why respondents did not live with both biological parents until their sixteenth birthday. We therefore cannot distinguish family disruptions due to parental divorce from those due to bereavement. This may be important for comparisons by birth cohort, if individuals born, say, before 1920, experienced a high incidence of parental death, while those born, say, in the 1970s faced a high rate of family dissolution due to parental separation. It is also important for our comparison purposes in the next sections. In fact, some of the studies used for comparison do distinguish between parental death and divorce, whereas others do not. Some others combine children who experienced bereavement with those who lived all their childhood with both biological parents, and, finally, a minority of studies does not state the procedure used to construct their family structure variables.

Because FBPAR16, FLVHMAG and FWHR14 were first collected in the sixth wave of the BHPS, an obvious question is whether attrition and nonresponse lead to a non-representative sample or not. Of the 10264 individuals interviewed in the first wave, 7430 (72%) were re-interviewed in the sixth wave, while another 2008 have joined the survey by wave 6. In general, the 2834 individuals who are in wave 1 but not in wave 6 seem to be randomly distributed in the sample of the first wave. A two-sample Kolmogorov-Smirnov test cannot reject (at one-percent level or less) the hypothesis of equality of the distributions for these individuals on several observable characteristics, such as gender, age, ethnic origin, marital status, region of residence,

education, labour force status, housing tenure, occupation and socio-economic group. Moreover, the distributions of the 2008 individuals, who are in the sixth but not in the first wave, are similar (along the same observable characteristics listed above) to those of the 2834 individuals who have left the panel. Using weights produces virtually identical comparable results.

Of the 9438 individuals interviewed in wave 6, 309 (3.3%) have missing information on questions D40 and D42. Of the remaining 9129, other 55 do not provide usable information on question D41, which means that only 3.8% of the individuals have missing records on childhood family structure. Of the 364 individuals with missing information, two-thirds are men, almost three-fifths were born before 1958, 40% live in the South (including London), nearly 60% are either employed or self-employed, 70% are owner-occupiers, and 40% are in the lowest socio-economic groups. Therefore, there is not a clear under- or over-representation of specific groups of the population.

In this paper we analyse FBPAR16 and, to a lesser extent, FLVHMAG. Table 1 reports the mean and standard deviation of FBPAR16 (labelled $\mu_{BHPS}$ and $\sigma_{BHPS}$, respectively) for all individuals and for men and women separately, by selected birth cohorts. The figures by cohort will be useful for the comparison analysis in the next section. Looking at all birth cohorts (top panel of Table 1), about 19% of individuals have lived part of their childhood in a non-intact family. The proportion is slightly higher for men, albeit the difference is not statistically significant. To account for differential attrition, individual nonresponse, and the inclusion of new entrants between waves 1 and 6, we have also used weighted data by applying the BHPS cross-sectional respondent weights for wave 6. These results are reported in the last two columns of Table 1. The means from the weighted data for the entire sample are very close to (and statistically insignificantly different from) the means from the unweighted data. Because this is also true when we partition the sample by birth cohort, we henceforth focus our attention on the unweighted data.

Before turning to the evaluation analyses, it is interesting to chart the trends in $\mu_{BHPS}$. Figure 1 plots this proportion for thirteen birth cohort categories, which include all individuals born before 1921 in one group and the 12 remaining groups of individuals born in 5-year windows between 1921 and 1980 (this is the "benchmark" line in Figure 1). "Benchmark-3" is similarly defined, with the difference that it refers to 14 groups, the first of which includes all individuals born before 1919, the last contains individuals born in 1979 and 1980 and the 12 remaining groups include individuals born in 5-year windows between 1919 and 1978. Regardless of how we stratify the sample by birth cohort, the picture in Figure 1 is clear. The proportion of children who ever experienced a family disruption before their sixteenth birthday has been declining from 0.18-0.19 for those in the earliest cohorts to 0.13-0.14 for those born at the end of the 1940s and the beginning of the 1950s. It has steadily increased ever since and almost doubled for individuals born in the late 1970s. The experience of growing up in a non-intact family is therefore not new. But the risk has become greater for the most recent cohorts, and the main reason of family disruption may have perhaps changed over time, from parental death for the earlier cohorts to marital breakdown for the later cohorts. Although it is not the objective of this paper, the BHPS data, as mentioned above, would not allow us to analyse such a potential shift. Using the "benchmark" categorisation defined above, Figure 2 presents the same temporal patterns as those shown in Figure 1 but distinguishing boys from girls. Despite some sizeable differences among men and women from the early and late birth cohorts, the overall trends by gender are remarkably similar.

**3. Some methodological issues**

The wave-6 BHPS information is retrospective, and can thus suffer from recall bias. (For a review of the reliability of recall data, see Dex (1995)). It can be argued that respondents are unlikely to forget their childhood family structure altogether, but their interviews may contain

time errors (i.e., errors in the timing or dating of family disruptions). For example, individuals might be inaccurate in remembering the correct age at which they started living with their mother and stepfather, or the exact family type they experienced when they were aged 14. This will then affect primarily FLVHMAG and, to a greater extent, FWHR14. But FBPAR16, which is the variable of major interest here, will probably be largely unaffected. There is, however, no instrument that would allow us to check for the potentially different magnitude of these two recall errors. And that is beyond the scope of this paper.

Indeed our chief goal is to establish whether the family structure information over the entire childhood (ages 0-15) is reasonably reliable for all BHPS respondents. For this purpose we cannot use standard validation methods (Biemer et al., 1991). For instance, reinterviews of the same individuals with status or response recall cannot be applied, simply because childhood family structure was collected for the first time in the sixth wave and no additional related information has been elicited in subsequent interviews. For the same reason, test-retest methods, whereby individuals are asked exactly the same questions at two points in time, cannot be implemented. The main analysis reported in the next section is instead based on pairwise record match to independent measurements of family structure. There are at least three problems with this method (Biemer et al., 1991; Dex, 1995). First, the family structure measurements used for comparison in this paper are typically not across the same population and the same time period. Second, these independent records are also not necessarily error free. Indeed, the childhood family structure variables constructed with data, say, from the birth cohort surveys are likely to suffer from non-ignorable attrition and must use complex information from a wide range of partnership arrangements (Ferri, 1984; Hobcraft, 1998; Cheesbrough, 2001), with ample opportunities for coding and processing errors. Third, we have no information about the error rates in such records. These problems imply that there is no independent record that qualifies as the obvious measurement for comparison. Evaluation of the wave-6 family structure data using

this method is made even more difficult by the widely scattered amount of existing evidence on family structure available from other public-use data sources. In some circumstances, the pieces of information needed for comparison are not provided (e.g., Elliot and Richards (1991) and Rodgers (1990)). In others, the reported statistics are based on the mother as the unit of observation rather than on the child (e.g., Kiernan, 1996).

Notwithstanding these limitations, there are many studies that one can choose from. Table 2 reports a total of 35 articles (some published and some not), of which 26 use birth cohort data and the remaining 9 employ data from other sources, including official divorce statistics and government social surveys. Each article may contain one or more relevant statistics, and may or may not distinguish between men and women. For our purposes, we shall focus on 59 statistics, which form the basis of our comparison exercise. Again, proportions and standard deviations are denoted $\mu_j$ and $\sigma_j$, respectively, where $j$ denotes the comparison survey (e.g., NCDS or ALSPAC). Not all the statistics reported in Table 2 are used in the analysis. The reason for doing so is comparability. For instance, of the two statistics shown in the case of Douglas's (1970) study, we shall only consider the bottom figure, as the top one is not directly comparable to the BHPS records (see also Table 3 and the discussion below).

The selection of the studies listed in Table 2 is driven by two basic principles: (a) equitable representation of the different data sources — the greater number of studies using the NCDS simply reflects the fact that more research has been conducted with that source (interestingly, Rodgers and Pryors (1998) list a total of more than 40 papers concerned with about 70 child outcomes, only two of which have used BCS data) — and different disciplines and authors; and (b) clear comparability with the measures that are available from the BHPS. Even though elements of arbitrariness cannot be ruled out, our choice therefore has not been driven by strategic considerations.

Two additional pieces of information displayed in Table 2 are the child's age at which family structure is measured ('Child's age'), and whether the figures include or exclude parental death ('Parental loss information'). For example, three statistics are shown in the case of Ferri's (1984) study, with the top figure (0.080) being measured when children were aged 7, the middle one (0.114) when children were aged 11 and the bottom one (0.164) when children were aged 16. Thus, a correct comparison of these statistics with the BHPS data requires us to use FLVHMAG rather than FBPAR16. Similarly, in the case, say, of Douglas's (1970) study, we show two figures, 0.052 and 0.115, which differ by parental loss information. The first (denoted with 'D') refers to the case in which parental deaths are excluded, while the second (denoted with 'DD') includes both divorces (or separations) and deaths or other reasons for family dissolution. A few studies, which do not explicitly indicate whether deaths are included in their definition of non-intact family, are denoted with '?'. As shown in the next section, these two pieces of information are of critical importance to our comparison exercise.

Our second method to evaluate the wave-6 family structure variables is to use independent measurements within the BHPS itself (Section 5). For a subgroup of (young) individuals who coreside with their parents in at least one of the available panel years, a measure of their childhood family structure can be derived using their parents' fertility and marital histories collected in wave 2 (1992). A problem of using this as a comparison measure is that it is difficult to know whether the inconsistencies are a result of recall errors in the wave-6 reports by children or an amalgamation of data handling errors and recall errors in the wave-2 reports by parents. Although this issue cannot be fully resolved, we will attempt to identify observable characteristics that correlate with inconsistent reports.

## 4. Comparing the BHPS data with the evidence found from other data sources

### *4.1 Birth cohort studies*

From the studies reported in Table 2, we observe large differences in $\mu_j$ even within data source, that is, the proportion of children who ever lived in a non-intact family by their sixteenth birthday across papers given *j*=NSHD, NCDS (including second generation NCDS) or BCS. This is particularly apparent in the case of the 26 (first generation) NCDS statistics, with the lowest proportion being 0.037 (for men only), and the greatest being 0.168 (for men and women combined), reported in the works by Dearden (1998) and Hobcraft (1998) respectively. As noted in the table, however, the former study does not indicate whether family disruptions due to parental death are included in the definition of non-intact family, while the latter includes them. A potential explanation for this huge discrepancy can thus be the different definitions of family structure. Although this is may be at work, it is not the whole story. In fact, the figure obtained by Hobcraft after excluding disruptions due to death is reduced only to 0.124, which is still at least 3 times larger than the Dearden's statistics and 2.5 times larger than the proportion reported in Chase-Lansdale et al. (1995), which excludes parental deaths too (0.048). More forcible reasons for these and other differences across papers using NCDS data should therefore be related to other factors, such as differential attrition rates, differential item non-response and missing values on different (dependent and independent) variables, and different sample selection criteria. These are factors that cannot be possibly accounted for in our analysis. In any case, this large heterogeneity warrants some caution in comparing the NCDS figures in Table 2 to the BHPS statistics in Table 1.

Another indication of the considerable dispersion across NCDS statistics is given by standard inequality measures. For example, the coefficient of variation (CV) is 0.305, suggesting that the standard deviation in $\mu_{NCDS}$ is more than 30% of its overall mean. Similar degrees of variation emerge also among studies that disaggregate the NCDS sample by gender of the child (e.g., Dearden (1998) versus Gregg and Machin (1999)).

The dispersion in μ among papers using BCS data is smaller after removing the statistics reported in Wadsworth et al. (1983 and 1985), which refer to children aged 0-5 and appear to be larger than those found with BHPS data. The corresponding CV computed on the 6 relevant statistics is 0.18. Nonetheless, the BCS studies in Table 2 show quite diverse figures. Machin et al. (1997) report that 17.3% of children in the BCS are in a non-intact family by age 16. This figure increases to 21.2% in the paper by Ely et al. (1999) and increases even further to 27% in the study by Cheesbrough (2001). In all three cases, $\mu_{BCS}$ is computed when the BCS children were aged 16, although the latter study emphasises the importance of family transitions from birth up to age 16. Finally, the within-study variation for the 7 statistics from the NSHD (omitting the two reported in Rodgers (1994) that exclude parental death) and for the 4 statistics from the second generation NCDS is considerably lower (CV=0.10 and 0.014, respectively).

Besides the ample heterogeneity in μ, it is noteworthy noticing that the statistics reported in the studies using the birth cohort surveys (Table 2) are generally smaller than the corresponding statistics found with the BHPS (Table 1). This "overestimation" may be an artefact of the non-ignorable attrition that affects the birth cohort data, whereby children in non-intact families are less likely to be followed in such surveys. It may also be due to different definitions of childhood family structure and different sample selections. While accounting for the first possibility is beyond the aims of this paper, the extent to which different selections and definitions influence our comparison instead will be partly evaluated in a later subsection.

Before doing so, we concentrate on pairwise comparisons. As a means of assessing whether the differences between the proportions from the BHPS and the other corresponding samples with known or unknown variances are significant, we use standard *t*-test methods (Hoel, 1984). Table 3 reports the *p*-values of such tests, where each of the relevant proportions reported in Table 2 is compared with the corresponding proportion from the BHPS. In most cases, the

BHPS proportions are listed in Table 1, in others, however — as for the comparison with the proportions reported in Ferri (1984), Wadsworth et al. (1983 and 1985) and O'Connor et al. (1998) — they are not due to space limitations. (They are presented in the notes of Table 3). Furthermore, when the comparison studies listed in Table 2 do not report standard deviations, we arbitrarily assign three separate values to $\sigma_j$, specifically, one that equals the corresponding $\sigma_{BHPS}$, another that is one-half of $\sigma_{BHPS}$, and another that is twice as large. Tests in which $\sigma_j$ is left unknown deliver *p*-values that are remarkably similar to those obtained when $\sigma_j$ is assumed to be equal to $\sigma_{BHPS}$, and are therefore not reported.

4.1.1 NSHD statistics

Only two out of the nine NSHD statistics are significantly different from the corresponding BHPS records when $\sigma_{NSHD}=\sigma_{BHPS}$ (see the third column of Table 3), namely those in Kuh and Maclean (1990) and Rodgers (1994), which both refer to women. Three additional significant differences emerge if $\sigma_{NSHD}$ is more conservatively set to be half of $\sigma_{BHPS}$ (last column), whereas no difference is detected when $\sigma_{NSHD}$ is twice the value of $\sigma_{BHPS}$ (second column). Perhaps unsurprisingly, therefore, the value of the standard deviation does matter. Thus, the validation of the BHPS figures against those available from the NSHD produces mixed results. The BHPS statistics for all individuals with no distinction by gender and those for men do not seem to be far off the mark of the NSHD statistics. But those for women show some appreciable departures.

It can be argued that $\mu_{BHPS}$ for women born in 1946 (0.204, see Table 1) is large simply because of measurement or recall errors. That errors in recall are relevant in this instance may be at odds with some of the existing evidence showing that marital and fertility histories — but not necessarily childhood family structure histories — are more accurately reported by women than by men (see, among others, Fikree, et al. (1993) and Rendall et al. (1999)). Measurement errors

may be nonetheless at work. One (partial) way to ascertain this possibility is to recompute the proportion $\mu$ for women born in the two adjacent years, take the average and perform the test again. For women born in 1945 and 1947, the average $\mu_{BHPS}$ is substantially lower and equal to 0.162. Assuming equal standard errors, we cannot reject the hypothesis that this proportion is equal to that reported in Kuh and Maclean (1990) ($p$=0.127). Notice that if the exercise is repeated for women born between 1945 and 1947 (i.e., including 1946), then $\mu_{BHPS}$=0.177, which is again significantly different from the figure in Kuh and Maclean ($p$<0.016). This finding suggests that measurement errors may play a part in explaining the departure of the BHPS statistics for women from the existing statistics derived from the NSHD. How far this issues is consequential for substantive studies concerned with, say, the effect of childhood family structure on child outcomes is not known and left for future research.

Notice also that the two proportions — computed either on the 1945 and 1947 cohorts or on the 1945-1947 cohorts in the BHPS — are never statistically close to the proportions reported in Rodgers (1994). But this should not be a concern given that Rodgers's family structure figures exclude parental death.

4.1.2 NCDS and second generation NCDS statistics

This represents the largest set of comparison studies. From the 18 works listed in Table 2, we use 30 statistics for the validation exercise in Table 3. Of these, only five show statistically different records from the corresponding BHPS figures (when the unknown $\sigma_{NCDS}$ is assumed to be equal to $\sigma_{BHPS}$). But only two of such departures are possibly relevant, that is, the two for women reported in Ní Bhrolcháin, et al. (1994) and Ní Bhrolcháin, et al. (2000). In fact, the remaining three statistics — the two in Dearden (1998) and that in Chase-Lansdale et al. (1995) — are not comparable to the wave-6 data because they exclude (or are not clear about) parental

loss due to death (see Table 2). As for the two departures emerging from the Ní Bhrolcháin et al.'s studies, it should be noticed that in both such studies family structure is measured when children were aged 7-16 (and thus lived with both natural parents at, and presumably up to, age 7) rather than over their entire childhood as in the BHPS. This selection may render the comparison less stringent.

Again, the results are affected by the value of $\sigma_{NCDS}$, in that if we set it to be half of the relevant standard deviations observed in the BHPS, a total of 11 significant differences emerge, whereas if we set it to be twice as large as $\sigma_{BHPS}$, no difference is detected.

For 14 (i.e., 67%) of the 21 comparisons with NCDS statistics that account for parental death (those labelled 'DD' in Table 2), $\mu_{BHPS}>\mu_{NCDS}$ (we exclude here the three statistics from the second generation NCDS, as they are remarkably close to those found with the BHPS data, and also because they come from a relatively homogeneous pool of authors). However, this is true for only eight (i.e., just over 50%) of the 15 records that measure family structure from birth to the child's sixteenth birthday. So, what at first glance seemed to be an overestimation issue is much less of a problem once differences in sample selection and family structure definition are accounted for.

Interestingly, the statistics for women show the largest departures again, as in the case of the NSHD studies. Part of the overestimation, therefore, may be due to measurement error. To check this, we recompute the average $\mu_{BHPS}$ for women born in 1957 and 1959 and perform the test against the two Ní Bhrolcháin et al.'s studies (despite their different definition of family structure) once again. The new average $\mu_{BHPS}$ is equal to 0.140 and, assuming equal standard errors, we cannot reject the hypothesis that this proportion is equal to that reported either in Ní Bhrolcháin et al. (1994) with $p$=0.105 or in Ní Bhrolcháin et al. (2000) with $p$=0.171. If the proportion is computed over the years between 1957 and 1959 (i.e., including 1958), then

$\mu_{BHPS}$=0.156. While we still cannot reject the hypothesis that this new proportion is equal to that found by Ní Bhrolcháin et al. (2000) with $p$=0.062, we can in the other case ($p$=0.035). This indicates that the presence of measurement error may explain part of the differences in the BHPS records of childhood family structure as compared to the NCDS figures for women.

In any case, both the BHPS retrospective information and the constructed family histories for the NCDS respondents and their children are strikingly close to each other. Given the considerable heterogeneity amongst the NCDS studies in Table 2 (in terms of, for example, research objectives, selected samples and relevant variables for analysis) and their relatively large number, this can be taken as suggestive evidence that the wave-6 variables provide satisfactorily accurate pieces of information to determine childhood family structure (at least for people born in the late 1950s). As the rest of this section illustrates, this seems to be the case also for people born in more recent years.

#### 4.1.3 BCS statistics

Of the eight BCS statistics shown in Table 2, only one is found to be significantly different from the corresponding BHPS records, when standard deviations are assumed to be equal. This emerges in the case of men in the study by Machin et al. (1997), which however does not provide sufficient information to establish whether or not parental deaths are a source of family disruption and is thus not fully comparable to the wave-6 statistics. But with a secular rise in life expectancy, the chance of living in a bereaved family by age 16 is arguably lower for individuals born in 1970 than, say, for individuals born in 1958 or 1946. For example, the death rates of men aged 35-44 has decreased from 2.1 per 1000 at the beginning of the 1950s to 1.5 per 1000 at the beginning of the 1970s (Office for National Statistics (2001), Tables 5.3 and 5.18). This therefore suggests that excluding family disruptions due to parental death may not lead to substantially different figures of childhood family structure for individuals born in 1970 or later.

We investigate this possibility, albeit only partially, with the following approximation. Consider the statistics reported in Wadsworth et al. (1983 and 1985), which refer to family structure over the first five years of life and exclude family disruption due to death. The difference between these and the corresponding BHPS figures (obtained using FLVHMAG and reported in the notes of Table 3), which include parental deaths, is 0.015. If, for simplicity, this differential is assumed to apply uniformly across child developmental stages, we will have another 0.015 difference for ages 6-10 and another 0.015 difference for the ages 11-15. By the end of childhood, therefore, the difference can be of the order of 4.5 percentage points. Adding this figure to the Machin et al.'s statistics of Table 2 would lead to new proportions that are not significantly different from those found with the wave-6 data ($p$=0.198 and 0.835 for men and women respectively). This exercise confirms the salience of the family type definitions, and illustrates that inclusion or exclusion of bereaved families in the definition of non-intact family may lead to substantially different conclusions in our comparison analysis even for birth cohorts as recent as that of the 1970 BCS. Thus, a clear (and feasible) improvement upon the current design will be to obtain information also on the *reasons* of dissolution in future BHPS waves.

*4.2 Official statistics and other studies*

From this set of studies we use 12 statistics for our comparison analysis. A glance at Table 2 and Appendix A reveals that the seven data sources are highly heterogeneous in their sampling scheme, sample sizes, geographical coverage, child's age at which family structure was measured and definition of family structure. Despite these differences, the coefficient of variation in $\mu$ is only about 0.22 (almost 70% less than the dispersion previously found among BCS studies).

Table 3 shows that four out of the 12 comparisons with the wave-6 data exhibit significant differences — three of which are however probably inconsequential because they

come from studies that either are unclear on the way they treat parental deaths or exclude them from the definition of non-intact family. The only significant difference that uses a definition of family structure comparable to that of the BHPS is found in relation to the work by Sweeting et al. (1998). In their paper, they analyse data from the West of Scotland Twenty-07 Study, which samples individuals born in 1971-1972 located in and around Glasgow city (see Appendix A). If we restrict our attention to the subsample of people who were born in 1971-1972 (as in the Twenty-07 Study) and currently (i.e., wave 6) live in Scotland, $\mu_{BHPS}$=0.239, which is virtually identical to the proportion reported in Table 1. After recomputing our equality tests, we cannot however reject the hypothesis that the new BHPS proportion is equal to the Sweeting et al.'s proportion. But this is so simply because the corresponding BHPS sample size drops to 46 (the standard deviation remains around the values reported in Table 1), making any inferential judgement difficult. Sweeting and West (1995) note that after the first sweep collected in 1987, the two follow-up samples collected in 1990 and 1993 suffer from differential attrition rates with underrepresentation of individuals from some social class backgrounds who may face a higher risk of living in a disrupted family. This may explain at least part of the departure from the BHPS records.

*4.3 Regression-based evaluation*

Our evaluation so far rests on pairwise comparisons between the wave-6 figures and the statistics obtained from independent external sources. Another way of assessing the quality of the BHPS measures is to compare them to the existing statistics *altogether*. To do this, we estimate the following

$$\mu_j = \beta \mu_{BHPS} + \varepsilon_j, \qquad (1)$$

using weighted least squares (WLS) regressions, where the weights are given by the sample size of each study (for the papers that separate men and women, sample sizes have been doubled), and $\beta$ is a parameter to be estimated. A value of $\beta$=1 suggests that the statistics from the BHPS and those from the other studies are identical, while a value of $\beta$<(or >)1 indicates that the BHPS figures are greater (or smaller) than the statistics from the other studies.

Figure 3 reports the results from the estimation of (1) when all 59 statistics are used, and visually displays them, along with the regression prediction (solid line) and the 45 degree line (dashed line), which indicates when $\beta$=1. (The size of each observation as displayed in the figure is proportional to the size of the sample from which the corresponding proportion is taken or computed). The estimated value of $\beta$ is 0.848 (*t*-ratio=27.58), and this is significantly different from 1 ($p$<0.0001), with a 95% confidence interval of [0.787, 0.910]. (Similar results are obtained if we estimate (1) by ordinary least squares without weighting). This finding suggests that the statistics obtained from the sixth wave of the BHPS are on average 15% higher than those reported in all the other studies listed in Table 3.

To determine which of the comparison statistics are closer to the BHPS figures, we re-estimate equation (1) on statistics grouped by source. The results from this exercise are given in Figures 4-7. The WLS estimate of $\beta$ is lowest amongst the NSHD studies ($\beta$=0.710) and largest amongst the more diverse group of the 'other' studies ($\beta$=0.977). This is the only group of statistics for which the hypothesis that $\beta$=1 cannot be rejected at any significance level ($p$=0.782). These findings confirm our previous overestimation result, whereby the proportion of people who report an experience of life in a non-intact family during childhood tends to be larger in the BHPS (wave 6) than in all the other comparison sources used here (possibly with the exclusion of the 'other' studies).

They also confirm the conjecture that there is a correlation between degree of departure and starting collection dates of the comparison studies, i.e., $\beta$ decreases monotonically as we move backward from the most recent studies (e.g., ALSPAC) to the 1958 NCDS (first generation) or the 1946 NSHD. Because the time gap between 1996 (year in which the BHPS collected the family structure information) and childhood family experiences is greater for those born in 1946 than for those born in 1970, recall error affecting the BHPS reports of older people may be responsible. This is a possibility that cannot be easily disregarded with our retrospective data (Dex, 1995).

However, as already argued before, an explanation of the overestimation based solely on recall error seems implausible for at least four reasons. First, there is substantial evidence suggesting that demographic (marital and fertility) histories — but, admittedly, not childhood family structure — can be collected with a reasonable degree of reliability by retrospective method (Dex, 1995), and especially for women (Peters, 1988). Second, it may not be reasonable to assume that recall always deteriorates with the passage of time. In fact, the elapse of time seems to make little difference to the recall of salient events (Cannell and Henson, 1974), and family structure during childhood can arguably be one of such events. Third, it is quite possible that the studies based on the birth cohort data suffer from non-ignorable attrition, whereby children in non-intact families are less likely to be followed by the surveys. Fourth, analyses of the birth cohort data have to construct family structure variables using complex information from a wide range of partnership arrangements, making (sometime unspecified) assumptions on how to fill gaps, particularly at early ages (e.g., Ferri, 1984; Machin et al. 1997; Hobcraft, 1998; Cheesbrough, 2001). The retrospective wave-6 information of the BHPS, instead, requires minimal data handling, thereby greatly reducing the likelihood of processing error.

This discussion underlies the possibility that overestimation and measurement (or recall) error bias are correlated. To assess how much of the overestimation is due to different age

selections and definitions of childhood family structure, we re-estimate equation (1) after excluding the comparison studies for which child's age or parental loss information are not consistent with those of the wave-6 data. The results of this WLS regression are displayed in Figure 8. The estimated value of $\beta$ — obtained from 33 (rather than 59) statistics — is now 0.875 ($t$-ratio=31.25), which is again significantly different from 1 ($p$<0.0002). Trimming the less comparable studies out of the analysis, therefore, does reduce the overestimation rate from 15% to about 12%. But this means that, at most, only 20% of the overestimation can be due to differences in ages at which family structure is measured or in the definitions of family structure itself. The main sources of departure must then be found along the other directions outlined above (e.g., non-ignorable differential attrition in longitudinal surveys and heterogeneity in sample selection criteria).

## 5. A comparison with other BHPS measures

### *5.1 The comparison sample*

Our last assessment of the wave-6 family structure variables relies on a comparison with other independent measures derived from the BHPS itself. This is done by analysing a special sample of young adults selected from the first seven waves of the BHPS, 1991-1997. This and similar samples have already been used in substantive research on the impact of childhood family structure and parental employment on children's outcomes (e.g., education and childbearing experiences) in early adulthood (Ermisch and Francesconi, 2001a and 2001b; Ermisch et al., 2001). This sample consists of 1286 young adults (i.e., aged 16 or over) who were born between 1970 and 1981, who did not have any serious health problems or disabilities, and who could be matched with one or both of their biological, adoptive or step-parents interviewed in at least one year during the first seven waves of the survey.

The second wave (1992) of the BHPS contains retrospective information on complete fertility, marital and cohabitation histories for all adult panel members in that year. This information provides the basis for the family structure measure in this special sample, i.e., whether the young adult spent time in a non-intact family during his/her childhood (ages 0-15). The variable is denoted ENIF. A child is defined as being brought up in an intact family if he/she lived continuously with both biological (or adoptive) parents, up to his/her sixteenth birthday. Thus, according to this definition, a child would have spent sometime in a non-intact (or single-mother) family if he/she ever lived with a biological or adoptive mother who was not cohabiting nor married before the child's sixteenth birthday, either because of partnership dissolution, or parental death, or because the child was born outside of a live-in partnership and the mother did not cohabit or marry within one year of the birth (ENIF=1). Further information on this sample can be found in Ermisch and Francesconi (2001a).

*5.2 Raw comparisons*

Let $\mu_{ENIF}$ denote the proportion of young adults in this special sample who ever lived in a non-intact family during childhood according to our definition, and $\mu_{FBPAR16}$ be the proportion for the same sample but obtained from FBPAR16, the wave-6 variable. The first row in panel (a) of Table 4 reports these proportions over the entire childhood (ages 0-15). It also shows the number of individuals who are classified as having lived in a non-intact family according to our definition of family structure (ENIF=1) but report to have lived all their childhood in an intact family (FBPAR16=0), expressed as a proportion of the total population in the sample. Let $\eta_{01}$ denote this proportion. Similarly, it shows $\eta_{10}$, for individuals for whom FBPAR16=1 and ENIF=0. The sum $\eta_{01}+\eta_{10}=\eta$ provides us with the proportion of young adults who are not correctly matched on the variables FBPAR16 and ENIF.

We observe that almost 24% of young adults in our special sample experienced life in a single-mother family by their sixteenth birthday. Using FBPAR16, we find that 22% of the sample did so before they reach age 16. The difference of 1.8 percentage points is not statistically significant at any conventional level. Of the 1286 observations, only 74 (or 5.7%) are mismatched, an arguably small fraction. The fifth column of Table 4 reports the kappa statistics of agreement, $\kappa$, which cannot exceed one by construction. In describing the relative strength of agreement associated with kappa statistics, Landis and Koch (1977, page 165) suggest that values of $\kappa$ above 0.81 indicate an 'almost perfect' agreement, while the agreement reflected by values between 0.61 and 0.80 is 'substantial' (the agreement is 'moderate' if $0.41 \leq \kappa \leq 0.60$, 'fair' if $0.21 \leq \kappa \leq 0.40$, 'slight' if $0.00 \leq \kappa \leq 0.20$ and 'poor' if $\kappa \leq 0$). These cut-off values are arbitrary but offer benchmarks for comparisons. In Table 4, a value of $\kappa$ of 0.838 suggests that the agreement between the childhood family structure measure from the sixth wave and the corresponding measure from our special sample is 'almost perfect'. Although we cannot expect this to hold for all individuals who cannot be matched with their parents (and this remains untestable), it is encouraging to observe such a high level of agreement for young adults born from 1970 onwards.

The last two columns show the results from equality tests at two points of the distributions of ENIF and FBPAR16, namely at the mean (*t*-tests) and at the median (sign rank tests). At both points the two distributions are significantly different. So, despite the substantial degree of similarity of the family structure information between the two samples, there are systematic differences that are potentially important to identify. This will be the goal of the next subsection.

The rest of panel (a) reports the same set of statistics broken down by developmental stage (ages 0-5, 6-10 and 11-15). To construct the proportions by developmental stage from the

wave-6 data, we used the information given in FLVHMAG. The proportions $\mu_{FBPAR16}$ and $\mu_{ENIF}$ are always close to each other, and their difference is never statistically significant at any conventional level for each stage (both at the mean and at the median). The total proportion of mismatched cases, $\eta$, appears to decrease by age, from 0.054 if the child is aged 0-5 to 0.032 when the child is aged 11-15. This trend suggests that young adults may have a better recollection of family disruptions when such disruptions have occurred in the proximate past. But there is no reason to expect this trend to be observed for the family structure variable derived from the special sample of young adults, ENIF, because this is derived from parents' retrospective family histories. Table 4 also shows that, on average, $\eta_{01} \geq \eta_{10}$, which indicates that ENIF tends to pick up more family disruptions than FBPAR16 (or FLVHMAG) does. Finally, the kappa statistics reveal 'substantial' to 'almost perfect' agreement between the two sets of variables also by developmental stage. Interestingly, according to both ENIF and FLVHMAG, of the children who spent some time in a single-mother family, approximately 45-46% had this family experience below the age of 6. Therefore, a partial explanation of why FBPAR16 overestimates the proportion of individuals who lived in a non-intact family during their childhood as compared to the evidence reported using birth cohort data (see Section 4) may be linked to the fact that the birth cohort studies do not collect detailed information on family transitions between their first and second sweeps — e.g., between birth and age 7 in the NCDS (e.g., Ní Bhrolcháin et al. (2000)), and between birth and age 5 in the BCS (e.g., Machin et al. (1997)).

Panels (b)-(f) of Table 4 show comparisons by various characteristics of the child (ages 0-15 only). Although the kappa statistics reveal that ENIF and FBPAR16 have an 'almost perfect' agreement in most cases, they appear to be different for women, never married individuals and young people with GCSE/O-level qualifications. For both ENIF and FBPAR16,

the table also reports the Pearson test for independence between two variables and the Kruskal-Wallis test for independence in the panels of Table 4 in which the child's characteristic is ordered, i.e., age and education. Women are less likely than men to have spent (at least part of) their childhood in a non-intact family, but this difference is not statistically significant (panel (b)). There is some variation in non-intact family structure rates with education and marital status. Young people with no educational qualification (panel (d)) and those in cohabiting unions (panel (e)) are significantly more likely to come from a disrupted family. Conversely, no statistically significant patterns by either age or employment status are evident (panels (c) and (f), respectively). Importantly, both ENIF and FBPAR16 show the same bivariate patterns across panels (b)-(f).

*5.3 Inconsistent reports in a multivariate analysis*

Knowing who are the "mismatched" individuals (i.e., those contributing to $\eta_{01}$ and $\eta_{10}$) is potentially relevant for survey designers and analysts. Children whose family structure information (either that reported in wave 6 or that contained in our special sample) is problematic may differ in some systematic way from children whose information is unambiguously determined. By considering $\eta$ ($=\eta_{01}+\eta_{10}$) in Table 4, we can detect that women, older and less educated individuals, unemployed and cohabiting people are more likely to be characterised by a mismatch (higher $\eta$). To test whether these bivariate associations are robust in a multivariate framework, we estimate the probability of mismatch using logistic regressions. For each individual *i* in family (or living with mother) *j*, the dependent variable, $U_{ij}$, is binary taking value one if ENIF=1 *and* FBPAR16=0 or ENIF=0 *and* FBPAR16=1, and taking value zero otherwise. The multivariate analysis assumes that (see Maddala (1983), pages 22-27):

$$\Pr(U_{ij} = 1) = 1 - F(-\alpha_0 - \alpha_1 \mathbf{C}_{ij} - \alpha_2 \mathbf{M}_{ij} - \alpha_3 \mathbf{Z}_{ij}), \quad (2)$$

where $\alpha = \{\alpha_0,...,\alpha_3\}$ is a conformable vector of parameters to be estimated, **X**={**1**,**C**,**M**,**Z**} is the vector of regressors (including a constant term), and $F(\cdot)$ is the logistic cumulative distribution, so that $1 - F(-\alpha\mathbf{X}) = \exp(\alpha\mathbf{X})/[1+\exp(\alpha\mathbf{X})]$. We distinguish three sets of explanatory variables, which are sequentially added to the regressions. In the first, denoted by **C**, we only include characteristics of the child (e.g., age, ethnic origin and education). They are presumably important in affecting how individuals respond to the wave-6 questions. The second, denoted by **M**, contains characteristics of the child's mother (e.g., her age at the child's birth and her education). These are possibly more relevant for the construction of ENIF. The last set of variables, denoted by **Z**, involves characteristics and conditions of the interviews (as reported by the interviewer) for the child in wave 6 — which may be relevant for FBPAR16 — and for the mother in wave two when marital and fertility histories were collected — which may be relevant for ENIF.

Table 5 reports the results from such regressions. A cursory examination of the tests reported at the bottom of the table reveals that specification [3] provides a substantially better fit than the other two specifications. Although all three specifications can predict 94-96% of cases correctly, the Pseudo $R^2$ in specification [3] is more than six times larger than that in specification [1] and almost 2.5 times more than that in specification [2]. The area under the receiver operating characteristic (ROC) curve — which is an index of the predictive power of the model, with a value of 0.5 indicating no predictive power and a value of 1 indicating a perfectly accurate model (Hanley and McNeil, 1982) — is 0.85 under specification [3], a substantially greater value than that obtained under the other two specifications. Finally, the Pearson $\chi^2$ goodness-of-fit tests — which test observed against expected number of responses using cells defined by the covariate patterns of the model (Hosmer and Lemeshow, 1989) — show that all three models fit the data well, but specification [3] outperforms the previous two. This evidence

already suggests that the variables **Z** add considerable explanatory power in estimating (2). We then shall concentrate on the results from this specification.

There are several noteworthy features of the estimates in Table 5. First, there are only a few characteristics of young people, **C**, that are significant predictors of mismatch. Older individuals and individuals of specific ethnic origins (notably, Bangladeshi/Pakistani) tend to have higher risks of mismatch. Other individual characteristics — such as education, employment status, number of siblings, marital status, housing tenure and region of residence (the last three sets of variables are not reported in the table) — do not have any significant impact. Second, maternal social status and education do not have a systematic pattern of association with mismatch between reported and constructed family structure information. But mothers' age at child's birth appears to play a relevant part: children with very young mothers (aged 21 or less at the time of their birth) are almost 4 times more likely to have a mismatch. Third, characteristics and conditions of the child's (wave 6) interview seem to be inconsequential, as do those of the mother's interview in the marital/fertility history section of wave 2. However, if, in the second wave, the mother did not cooperate with the interviewer and her interview was influenced by others, the chances of observing mismatched family structure information increase by about 7 and 9 times, respectively. This means that the probability of mismatch increases, on average, by 35 and 46 percentage points respectively, soaring from 5.7% to 40.5% and 51.6%.

In sum, it appears that only a few observable individual- and family-specific characteristics have a part in the way family structure information is reported (either by young adults or by their parents). These include child's and mother's age and ethnicity. But characteristics and conditions of the interviews, especially of mothers, turn out to be the single most important predictors for sorting young people into the unmatched group. The interviewer's competence to promote desirable conditions for disclosure of potentially sensitive data (e.g.,

avoiding the interference of other individuals) needs not be relevant to the entire BHPS questionnaire, but appears to be decisive to collect high-quality family structure retrospective information. In long-term follow-up studies such as the BHPS, this feature is therefore of paramount importance.

## 6. Conclusions

This paper has performed an evaluation of the data collected in the sixth wave of the BHPS on family structure during childhood. Two methods have been used for this evaluation. The first uses a pairwise record match to independent measurements of family structure. For this purpose, statistics obtained from a number of substantive studies using external data sources — for instance, official divorce data and birth cohort surveys — and spanning a range of social sciences provide the comparison records. The second method uses independent measurements within the BHPS itself.

The first method shows that, in relation to external sources, the BHPS overestimates the proportion of people who report an experience of life in a non-intact family during childhood by about 12%. There is also evidence that the overestimation is larger for records obtained from the 1946 NSHD data than the more recent 1970 BCS or 1991-1992 ALSPAC (the dates here refer to the first year of data collection). Although an explanation based on recall error that deteriorates with the age of the BHPS respondents cannot be dismissed, part of the overestimation is likely to be accounted for by non-ignorable attrition in the birth cohort and other longitudinal surveys, whereby children in non-intact families are less likely to be followed by such studies. Interestingly, circumvention of this and other related problems (e.g., differences in sample selection criteria and rates of item non-response) is relevant for an appropriate implementation of our method of evaluation.

From the second method, there is evidence that the wave-6 data underestimate the proportion of young people having experienced part of their childhood in a non-intact family by about 8%. The likelihood of disagreement between these two statistics is strongly and positively associated with poor interview characteristics that seem to affect the comparison measure more than the wave-6 measure.

In general, despite such differences, there is a substantial degree of similarity between the family structure information collected in the sixth wave of the BHPS and the host of highly diverse records against which it has been compared. Even though it is elicited retrospectively and can be recalled with error, there are at least three reasons to use such information in future substantive analyses. First, the BHPS measures are easy to use. Rather than processing complex information from a range of partnership arrangements (as most of the studies using birth cohort information do), the wave-6 variables are straightforward to operationalise because they directly come from respondents. Second, they refer to a broad range of birth cohorts covering a significant part of the twentieth century (the last cohort being born in 1980 at the time of the wave-6 collection). This may be relevant for analysing whether childhood family structure has a long-lasting effect on socio-economic and health outcomes not only in early adulthood but also later in life (see, among others, Hauser and Sweeney (1997)). Third, as the BHPS remains broadly representative of the population of Britain as it changes over time and as it continues to collect childhood family structure information, its use looms more urgently in substantive research to come. In future waves, collecting data on the reason why people had experience of life in a non-intact family will undoubtedly represent a considerable improvement upon the currently available information.

## Appendix A

### The British Household Panel Survey

This is a multi-purpose study of a nationally representative sample of the population of Great Britain living in private households in 1991, when the first wave of the BHPS was collected. The achieved wave-1 sample covered 5500 households and corresponds to a response rate of about 74% of the effective sample size. At wave 1, about 92% of eligible adults, i.e. approximately 10300 individuals, provided full interviews. The same individuals have been reinterviewed each successive year, and if they split off from their original households to form new households all adult members of these households are also interviewed. Similarly, children in the original households are interviewed when they reach 16 years of age. Thus the sample remains broadly representative of the population of Britain as it changes over time. Of those interviewed in the first wave, 88% were successfully reinterviewed in wave 2 (autumn 1992), and subsequent wave-on-wave response rates have consistently been around 95% or more. The core questionnaire elicits information about income, labour market behaviour, housing conditions, household composition, education and health at each yearly interview. Information on changes (e.g., employment, household membership ad receipt of each income source), which have occurred with the households in the period between interviews, is also collected. The second wave (1992) obtained retrospective information on complete fertility, marital, cohabitation and employment histories for all adult panel members in that year, and the third wave (1993) collected detailed job histories. The sixth wave (1996) collected the childhood family structure data that is analysed in this paper. Further information may be obtained at:
http://www.iser.essex.ac.uk/bhps/.

### The National Survey of Health and Development

This originated as a study of childbirth. Interviews were conducted with women who gave birth during the week 3-9 March 1936 in Great Britain. Subsequently, a follow-up at age 2 years was carried out, and this excluded multiple and illegitimate births from the original sample of 13687 and included all the births to nonmanual workers' and agricultural families and one-quarter of the remaining births to manual workers' families. The resulting sample (N=5362) constituted the NSHD, and was the target for further follow-up studies which occurred at frequent intervals throughout childhood and early adulthood, and less often since. The most recent contacts were home interviews at age 36 (1982-1983), age 43 (1989-1990) and age 53 (1999-2000). Data collected during childhood included socio-economic and demographic family characteristics, educational behaviour, histories of the children's growth, intellectual development, illness and behaviour problems, and measures of classroom behaviour and personality. In adulthood, histories of employment, housing, marriage, fertility and illness have been collected by personal interview. Measures of body shape, blood pressure and respiratory functions, and assessments of certain symptoms have been obtained, particularly at the most recent contacts. For further details, see Atkins et al. (1981) and Wadsworth et al. (1992). Additional information can be found at:
http://www.ucl.ac.uk/epidemiology/.

### The National Child Development Study

This survey originated in the Perinatal Mortality Survey, which was originally designed to examine the social and obstetric factors associated with still birth and death in early infancy. A total of 17414 mothers, representing 98% of all births in the week 3-9 March 1958, were interviewed for the original study. The children were subsequently followed up through their school years at ages 7 (N=15468), 11 (N=15503) and 16 (N=14761) and were traced and

interviewed on three occasions during adulthood at ages 23 (N=12537), 33 (N=11407) and 41-42 (N=12901). The available information covers a wide range of topics including medical, demographic, social and psychological, educational and economic aspects of their life histories. During the 1991 survey, a special study was also undertaken of the children of one in three cohort members, including assessments of the behavioural and cognitive development of approximately 5000 children. There have also been surveys of subsamples of the cohort, the most recent at age 37 when the basic skills of a representative sample of 10 per cent of cohort members were assessed. Further information about this and previous sweeps of, and related studies to, the NCDS may be found in Bynner et al. (2001) and at:
http://www.data-archive.ac.uk/ (SN:4396, 33004), and
http://www.cls.ioe.ac.uk/Cohort/Ncds/mainncds.htm, and
http://www.cls.ioe.ac.uk/Cohort/Ncds2000/mainncds00.htm.

**The British Cohort Study**

This survey originated in the British Births Survey, which was a study of the health of infants (and their mothers) at birth and during the first week of life. The study comprised 17198 children (more than 98% of all births) born throughout the United Kingdom (including Northern Ireland) during the week 5-11 April 1970. Since then there have been five attempts to gather information from the full cohort. With each successive attempt, the scope of enquiry has broadened from a strictly medical focus at birth, to encompass physical and educational development at the age of five (1975 interview, N=13135), physical, educational and social development at the ages of ten (1980, N=14940) and sixteen (1986, N=11628), and then to include economic development and other wider factors at 26 years (1996, N=9003) and 29 years (1999-2000, N=11261). Data have been collected from a number of different sources, and in a variety of ways. In the birth survey, information was collected by means of a questionnaire that was completed by the midwife present at the birth, and supplementary information was obtained from clinical records. In 1975 and 1980, parents of the cohort members were interviewed by Health Visitors, and information was gathered from head and class teachers (who completed questionnaires), the school health service (which carried out medical examinations on each child), and the subjects themselves (who undertook tests of ability). In both 1975 and 1980, the cohort was augmented by the addition of immigrants to Britain who were born in the target week in 1970. Subjects from Northern Ireland, who had been included in the birth survey, were dropped from the study in all subsequent sweeps. In the 1986 survey, sixteen separate survey instruments were employed, including parental questionnaires, school class and head teacher questionnaires and medical examinations (including measurement of height, weight and head circumference). The cohort members completed questionnaires, kept two four-day diaries (one for nutrition and one for general activity), and undertook some educational assessments. Further information on the BCS is available in Bynner et al. (2001) and at
http://www.data-archive.ac.uk/ (SN:4396, 33229), and
http://www.cls.ioe.ac.uk/Cohort/Bcs/mainbcs.htm, and
http://www.cls.ioe.ac.uk/Cohort/Ncds2000/mainncds00.htm.

**The OPCS Omnibus Survey**

This is a regular, multi-purpose survey. In recent years it has gone into the field in eight months of the year. The Omnibus started operating commercially in 1990, and was set up originally to meet the needs of Government departments for a survey that used short and simple sets of questions, had greater statistical reliability than private sector omnibus surveys and a properly designed random sample. Each month a representative sample of approximately 2000 adults aged 16 or over living in private households is interviewed on a variety of topics. Each month's

questionnaire consists of two elements: core questions, mainly covering demographic information, are always asked together with non-core questions that vary from month to month. This are specially designed to address ad-hoc issues. In the stepfamily module of questions used in Haskey (1994), respondents were asked whether they — and their partners — had dependent stepchildren living with them in their family. The questions were included in successive months: October-December 1990, April-June 1991, and March-June 1992. Some additional questions were included in the surveys of these latter four months. In particular, respondents were also asked whether they — or their partners — had any dependent children, including stepchildren, living *outside* the family. Further details are available in Haskey (1994) and at http://www.data-archive.ac.uk/.

**The West of Scotland Twenty-07 Study**

This is a three-sweep longitudinal study of individuals born in 1971-72, located in Central Clydeside Conurbation, a predominantly urban area in and around Glasgow city. The aim of the study of to examine the relationship between social position and health, and data relating to life circumstances, beliefs, behaviours and health have therefore been collected. The study began in 1987, when individuals were aged 15-16, and at that date a response rate of 65% (excluding those who had moved house prior to first contact) of the issued samples was obtained, giving a total of 1009 respondents (482 males and 527 females). At this baseline stage, separate interviews were also conducted with the parents of 995 respondents. Of the baseline participants, 908 (430 males and 478 females) were re-interviewed in 1990, at the age of 18-19. Finally, 806 (372 males and 434 females) also completed a postal questionnaire in 1993, at ages 21-22. The numbers in these two follow-up samples represent response rates of 90% and 80% of baseline respondents respectively, with substantially greater losses among respondents from manual social class backgrounds. For further details, see Sweeting and West (1995) and references therein.

**The North-East London GP Registers**

This is a cross-sectional survey of 529 girls aged 15-20 years whose names were drawn from general practitioners (GPs) age/sex registers of eight general practices in North and East London. The areas covered vary considerably in social class, and educational and employment opportunities. The practices also vary in size and in the range of facilities available to patients. In each practice a list of all the girls aged between 15 years and 19 years and 11 months was compiled. This list was given to the GPs and practice secretaries who excluded families in which it was known that the girl or her ‘key’ relative had serious learning difficulties, or could not speak or read sufficiently good English for the interview. The girls and their key relatives (92% biological or surrogate mothers) were sent separate letters explaining the aims of the study. The semi-structured interview was held with each girl and her key relative together in the girl’s home. Furhter information can be found in Monck et al. (1994).

**The Avon Longitudinal Study of Parents and Children**

This is a prospective longitudinal study of women, their partners and an index child. A key aim of ALSPAC, formerly the Avon Longitudinal Study of Pregnancy and Childhood, is to understand the ways in which the physical and social environment interact, over time, with the genetic inheritance to affect the child’s health, behaviour and development. The study is designed to link together information from a variety of sources including hands-on examination of the children, questionnaires completed by parents, health records, assays of biological samples and specific measurements of the environment in the home, and to use these data to test hypotheses on the causes and prevention of childhood ailments and disorders. Prospective data

from early pregnancy and from maternal as well as child DNA permit intergenerational studies. The study design includes all pregnant women living the geographical area of Avon, England, who were to deliver their baby between April 1, 1991 and December 31, 1992. Avon includes both urban and rural areas and is demographically similar to Britain as a whole. It is estimated that 85-90% of the eligible population took part. The 14,000 children who form the basis of this survey reached the age of seven years in 1998-1999, at which point a half-day examination of each child assessed various (mainly physical) characteristics. At age 8, a further half day of tests is being undertaken, mainly assessing cognitive and behavioural attributes. The results of these two assessments are the end point of a number of specific studies and form a baseline for future data collections. Future strategies involve examinations of all children and their parents every year. Further information can be obtained at:
http://www.alspac.bris.ac.uk/AlspacExt/Default.shtm.

**The National Survey of Sexual Attitudes and Lifestyles**
This cross-sectional survey was carried out in 1990-1991 on a nationally representative sample of the British population. The sample consists of 18876 persons aged between 16 and 59. The aims of the survey are to understand the epidemiology of HIV and other sexually transmitted diseases, to assess existing patterns and temporal changes in sexual behaviour in the general population, and to associate these with demographic characteristics and attitudes to different lifestyles and knowledge of possible associated risks. A wide range of topics are investigated by the survey, and cover general health, family circumstances and current marital status, demographic characteristics and a highly detailed information on sexual practices (including past sexual histories) and attitudes towards different sexual lifestyles, knowledge of AIDS and HIV, experience of infertility, abortion and miscarriage, contraception use and psychosocial factors influencing sexual behaviour such as source of information about sex (eg. family, friends and school). Detailed information is provided in the comprehensive report on the study by Johnson et al. (1994). Further information is available at
http://www.data-archive.ac.uk/ (SN:3434)

**The General Household Survey**
This is a multi-purpose continuous cross-sectional survey, which collects information on a range of topics from people living in private households in Great Britain. The survey started in 1971 and has been carried out annually since then, except for breaks in 1997-1998 (when the survey was reviewed) and 1999-2000 (when the survey was redeveloped). The main aims of the GHS is to gather data on a wide range of core topics, comprising household and family information (including marriage, cohabitation and fertility) housing, employment, education, income and health. A sample of over 12000 addresses is selected each year from the Postcode Address File. All adults aged 16 and over are interviewed in each responding household. Demographic and health information is also collected about children (aged less than 16 years) in the household. For 2000-2001, the survey response rate was 67%, which yielded a sample of 19266 people of all ages in 8221 households. Further information is in Walker et al. (2001). Additional details are available at:
http://www.data-archive.ac.uk/ (SN:4518).

**Official divorce statistics**
These come directly from the courts. For each divorce — decree absolute — granted in England and Wales, the Office for National Statistics receives from the court awarding the decree a completed form giving statistical information on the couple, their marriage, and any children in their family. The court records the date of birth of each child in the couple's family who was

aged under 18 when the divorce petition was filed, and the number of children who were aged 18 or over on the same date. (See Haskey (1990) for further details). These statistics provide the numbers of children by age who experienced divorce in their family at a particular point in time. Haskey (1997) uses data from the Labour Force Survey to obtain estimates of the number of children at each age who were living in married couple families during the same time period (1994-1995). Combining the data from the two sources allows one to compute the (period-specific) probabilities of children experiencing divorce in their family at each age.

## References


Atkins, E., Cherry, N., Douglas, J.W.B., Kiernan, K.E. and Wadsworth, M.E.J. (1981) The 1946 British birth cohort. In *Prospective Longitudinal Research* (eds. S.A. Mednick and A.E. Baert), pp. 25-30. Oxford: Oxford University Press.

Biemer, P.B., Groves, R.M. Lyberg, L.E., Mathiowetz, N.A. and Sudman, S. (eds.) (1991) *Measurement Errors in Surveys*. New York: Wiley.

Bynner, J., Butler, N., Ferri, E., Shepard, P. and Smith, K. (2001) The design and conduct of the 1999-2000 surveys of the National Child Development Study and the 1970 British Cohort Study. Institute of Education, CLS Working Paper 1.

Cannell, C.F. and Henson, R. (1974) Incentives, motives, and response bias. *Annals of Economic and Social Psychology*, **2**, 526-595.

Chase-Lansdale, P.L., Cherlin, A.J. and Kiernan, K.E. (1995) The long-term effects of parental divorce on the mental health of young adults: a developmental perspective. *Child Development*, **66**, 1614-1634.

Cheesbrough, S. (2001) Childhood family disruption and outcomes in young adulthood: evidence from the 1970 British Cohort Study. Ph.D. dissertation, University of Southampton, Department of Social Statistics.

Cherlin, A.J., Kiernan, K.E. and Chase-Lansdale, P.L. (1995) Parental divorce in childhood and demographic outcomes in young adulthood. *Demography*, **32**, 299-318.

Dearden, L. (1998) Ability, families, education and earnings in Britain. Institute for Fiscal Studies Working Paper No. W98/14.

Dex, S. (1995) The reliability of recall data: a literature review. *Bulletin de Methodologie Sociologique*, **49**, 58-89.

Douglas, J.W.B. (1970) Broken families and child behaviour. *Journal of the Royal College of Physicians London*, **4**, 203-210.

Douglas, J.W.B. and Blomfield, J.M. (1958) *Children under Five*. London: Allen & Unwin.

Duncan, G.J. and Brooks-Gunn, J. (eds.) (1997) *Consequences of Growing up Poor*. New York: Russell Sage Foundation.

Dunn, J., Deater-Deckard, K., Pickering, K., O'Connor, T.G., Goodman, R. and Golding, J. (1998) Children's adjustment and prosocial behavior in step-, single-parent, and non-stepfamily settings: findings from a community study. *Journal of Child Psychology and Psychiatry*, **39**, 1083-1095.

Elliott, J. and Richards, M. (1991) Parental divorce and the life chances of children. *Family Law*, **21**, 481-484.

Ely, M., Richards, M.P.M., Wadsworth, M.E.J. and Elliot, B.J. (1999) Secular changes in the association of parental divorce and children's educational attainment — Evidence from three British birth cohorts. *Journal of Social Policy*, **28**, 437-455.

Ermisch, J. and Francesconi, M. (2001a) Family structure and children's achievements. *Journal of Population Economics*, **14**, 249-270.

Ermisch, J. and Francesconi, M. (2001b) *The effect of parents' employment on children's lives*. London: Family Policy Study Centre for the Joseph Rowntree Foundation.

Ermisch, J, Francesconi, M. and Pevalin, D.J. (2002) Childhood parental behaviour and young people's outcomes. University of Essex, ISER Working Paper No. 2002-12, July.

Ferri, E. (1984) *Stepchildren: A National Study*. Windsor: NFER-Nelson.

Fikree, F.F., Gray, R.H. and Shah, F. (1993) Can men be trusted? A comparison of pregnancy histories reported by husbands and wives. *American Journal of Epidemiology*, **138**, 237-242.

Gregg, P. and Machin, S. (1999) Childhood disadvantage and success or failure in the labour market. In *Youth Employment and Joblessness in Advanced Countries* (eds. D. Blanchflower and R. Freeman), pp. 247-288. Cambridge, Mass.: National Bureau of Economic Research.

Green, G., Macintyre, S., West, P. and Ecob, R. (1990) Do children of lone parents smoke more because their mothers do? *British Journal of Addiction*, **85**, 1497-1500.

Hanley, J.A. and McNeil, B.J. (1982) The meaning and use of the area under a receiving operating characteristic (ROC) curve. *Radiology*, **143**, 29-36.

Haskey, J. (1990) Children in Families Broken by Divorce. *Population Trends*, **61**, 34-43.

Haskey, J. (1994) Stepfamilies and stepchildren in Great Britain. *Population Trends*, **76**, 17-28.

Haskey, J. (1997) Children who experience divorce in their family. *Population Trends*, **87**, 5-10.

Hauser, R.M. and Sweeney, M.M. (1997) Does poverty in adolescence affect the life chances of high school graduates? In *Consequences of Growing Up Poor* (eds. G.J. Duncan and J. Brooks-Gunn), pp. 541-595. New York: Russell Sage Foundation.

Haveman, R. and Wolfe, B. (1994) *Succeding Generations: On the Determinants of Children's Attainments*. New York: Russell Sage Foundation.

Hobcraft, J. (1998) Intergenerational and life-course transmission of social exclusion: influences of childhood poverty, family disruption, and contact with the police. London School of Economics, CASEpaper 15, November.

Hoel, P.G. (1984) *Introduction to Mathematical Statistics*. Fifth Edition. New York: John Wiley & Sons.

Hope, S., Power, C. and Rodgers, B. (1998) The relationship between parental separation in childhood and problem drinking in adulthood. *Addiction*, **93**, 505-514.

Hosmer, D.W., Jr. and Lemeshow, S. (1989) *Applied Logistic Regression*. New York: John Wiley.

Johnson, A.M., Wadsworth, J., Wellings, K. and Field, J. (1994) *Sexual Attitudes and Lifestyles*. Oxford: Blackwell Scientific Publications.

Joshi, H., Cooksey, E.C., Wiggins, R.D., McCulloch, A., Verropoulou, G. and Clarke, L. (1999) Diverse family living situations and child development: a multilevel analysis comparing longitudinal evidence from Britain and the United States. *International Journal of Law, Policy and the Family*, **13**, 292-314.

Joshi, H. and Verropoulou, G. (2000) *Maternal Employment and Child Outcomes*. London: The Smith Institute.

Joshi, H.E. and McCulloch, A. (2002) Child development and family resources: evidence from the second generation of the 1958 British birth cohort. *Journal of Population Economics*, **15**, 283-304.

Kiernan, K.E. (1992) The impact of family disruption in childhood and transitions made in young adult life. *Population Studies*, **46**, 213-234.

Kiernan, K.E. (1996) Lone motherhood, employment and outcomes for children. *International Journal of Law, Policy and the Family*, **10**, 233-249.

Kiernan, K.E. (1997) The legacy of parental divorce: social, economic and demographic experiences in adulthood. London School of Economics, CASEpaper 1, October.

Kiernan, K.E. and Hobcraft, J. (1997) Parental divorce during childhood: age at first intercourse, partnership and parenthood. *Population Studies*, **51**, 41-55.

Kuh, D. and Maclean, M. (1990) Women's childhood experience of parental separation and their subsequent health and socioeconomic status in adulthood. *Journal of Biosocial Science*, **22**, 121-135.

Landis, J.R. and Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159-174.

Machin, S., Montgomey, S.J. and Thomas, J. (1997) Childhood family disruption and juvenile delinquency: evidence from the 1970 British Cohort Study. Unpublished paper, University College London, June.

Maclean, M. and Wadsworth, M.E.J. (1988) The interests of children after parental divorce: a long-term perspective. *International Journal of Law and the Family*, **2**, 155-166.

Mayer, S.E. 1997. *What Money Can't Buy: Family Income and Children's Life Chances*. Cambridge, Mass.: Harvard University Press.

McLanahan, S.S. and Sandefur, G. (1994) *Growing up with a Single Parent*. Cambridge: Harvard University Press.

Monck, E. Graham, P., Richman N. and Dobbs, R. (1994) Adolescent girls I: self-reported mood disturbance in a community population. *British Journal of Psychiatry*, **165**, 760-769.

Ní Bhrolcháin, M., Chappell, R. and Diamond, I. (1994) Scolarité et autres caractéristiques socio-démographiques des enfants de mariages rompus [Educational and socio-demographic outcomes among children of disrupted and intact marriages.] *Population*, **6**, 1585-1612.

Ní Bhrolcháin, M., Chappell, R., Diamond, I. and Jameson, C. (2000) Parental divorce and outcome for children: evidence and interpretation. *European Sociological Review*, **16**, 67-91.

O'Connor, T.G., Hawkins, N., Dunn, J., Thorpe, K.J., Golding, J. and the ALSPAC Study Team. 1998. Family type and depression in pregnancy: factors mediating risk in a community sample. *Journal of Marriage and the Family*, **60**, 757-770.

Office for National Statistics (2001) *Annual Abstract of Statistics*. London: The Stationary Office.

Peters, H.E. (1988) Retrospective versus panel data in analysing lifecycle events. *Journal of Human Resources*, **23**, 488-513.

Rendall, M.S., Clarke, L., Peters, E.H., Ranjit, N. and Verropoulou, G. (1999) Incomplete reporting of men's fertility in the United States and Britain: a research note. *Demography*, **36**, 135-144.

Rodgers, B. (1990) Adult affective disorder and early environment. *British Journal of Psychiatry*, **157**, 539-550.

Rodgers, B. (1994) Pathways between parental divorce and adult depression. *Journal of Child Psychology and Psychiatry*, **35**, 1289-1308.

Rodgers, B., Power, C. and Hope, S. (1997) Parental divorce and adult psychological distress: evidence from a national birth cohort. *Journal of Child Psychology and Psychiatry*, **38**, 867-872.

Rodgers, B. and Pryor, J. (1998) *Divorce and Separation: The Outcomes for Children*. York: Joseph Rowntree Foundation.

Sweeting, H. and West, P. (1995) Distribution of basic information from the 1993 postal follow-up of the Twenty-07 Study youth cohort. Glasgow, MRC Medical Sociology Unit Working Paper No. 48.

Sweeting, H., West, P. and Richards, M. (1998) Teenage family life, lifestyles and life changes: associations with family structure, conflicts with parents and joint family activity. *International Journal of Law, Policy and the Family*, **12**, 15-46.

Verropoulou, G., Joshi, H. and Wiggins, R.D. (2002) Migration, family structure and children's well-being: a multi-level analysis of the second generation of the 1958 birth cohort study. *Children and Society*, **16**, 219-231.

Wadsworth, J., Burnell, I., Taylor, B. and Butler, N. (1983) Family type and accident in preschool children. *Journal of Epidemiology and Community Health*, **37**, 100-104.

Wadsworth, J., Burnell, I., Taylor, B. and Butler, N. (1985) The influence of family type on children's behaviour and development at five years. *Journal of Child Psychology and Psychiatry*, **26**, 245-254.

Wadsworth, M.E.J., Mann, S.L., Rodgers, B., Kuh, D.J.L., Hilder, W.S. and Yusuf, E.J. (1992) Loss and representativeness in a 43 year follow up of a national birth cohort. *Journal of Epidemiology and Community Health*, **46**, 300-304.

Walker, A., Maher, J., Coulthard, M., Goddard, E. and Thomas, M. (2001) *Living in Britain: Results from the 2000 General Household Survey*. London: The Stationary Office.

Table 1
Proportion of children who lived in a non-intact family before age 16 by birth cohort and gender (FBPAR16)

| Birth cohort and gender | Unweighted data | | | Weighted data | | Comparative study |
|---|---|---|---|---|---|---|
| | $\mu_{BHPS}$ | $\sigma_{BHPS}$ | N | $\mu_{BHPS}$ | N | |
| Birth cohort: All [a] | | | | | | |
| All | 0.191 | 0.393 | 9129 | 0.192 | 9109 | |
| Men | 0.195 | 0.396 | 4220 | 0.196 | 4211 | |
| Women | 0.188 | 0.391 | 4909 | 0.189 | 4898 | |
| Birth cohort: 1931-74 | | | | | | |
| All | 0.183 | 0.387 | 6678 | 0.186 | 6665 | |
| Men | 0.184 | 0.388 | 3135 | 0.186 | 3130 | NSSAL |
| Women | 0.183 | 0.386 | 3543 | 0.185 | 3535 | |
| Birth cohort: 1946 | | | | | | |
| All | 0.168 | 0.375 | 167 | 0.160 | 167 | |
| Men | 0.116 | 0.322 | 69 | 0.116 | 69 | NSHD |
| Women | 0.204 | 0.405 | 98 | 0.191 | 98 | |
| Birth cohort: 1958 | | | | | | |
| All | 0.154 | 0.361 | 169 | 0.144 | 169 | |
| Men | 0.118 | 0.325 | 76 | 0.126 | 76 | NCDS |
| Women | 0.183 | 0.388 | 93 | 0.159 | 93 | |
| Birth cohort: 1970 | | | | | | |
| All | 0.256 | 0.438 | 164 | 0.285 | 163 | |
| Men | 0.282 | 0.453 | 85 | 0.300 | 85 | BCS |
| Women | 0.228 | 0.422 | 79 | 0.269 | 78 | |
| Birth cohort: 1971-72 | | | | | | |
| All | 0.240 | 0.428 | 367 | 0.264 | 364 | Twenty-07 |
| Men | 0.250 | 0.434 | 184 | 0.289 | 184 | Study |
| Women | 0.230 | 0.422 | 183 | 0.241 | 180 | |
| Birth cohort: 1974-80 | | | | | | Omnibus Survey, |
| All | 0.247 | 0.431 | 1166 | 0.242 | 1159 | Official divorce |
| Men | 0.271 | 0.445 | 573 | 0.267 | 569 | statistics, |
| Women | 0.224 | 0.417 | 593 | 0.217 | 590 | London GP registry |
| Birth cohort: 1979-80 | | | | | | 2nd generat. NCDS, |
| All | 0.256 | 0.437 | 308 | 0.229 | 308 | ALSPAC, 2000 GHS |
| Men | 0.289 | 0.455 | 159 | 0.251 | 159 | London GP registry, |
| Women | 0.221 | 0.417 | 149 | 0.205 | 149 | Official divorce stats. |

*Source*: Wave 6 of the British Household Panel Survey (1996).
*Note*: N=number of individuals. These figures refer to family structure over the entire childhood (ages 0 to 15). FBPAR16 is the wave 6 family structure measure and denotes "lived with both natural parents". Weighted figures are obtained using the BHPS cross-sectional enumerated individual weights for wave 6.
[a] Years of birth range from 1894 (with 1 individual) to 1980 (with 158 individuals).

Table 2
Childhood family structure statistics from selected birth cohort studies

| Study | Proportion of all children in a non-intact family | | | Proportion of young men in a non-intact family | | | Proportion of young women in a non-intact family | | | Parental loss inform. | Child's age | Data source |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mu_j$ | $\sigma_j$ | $N_j$ | $\mu_j$ | $\sigma_j$ | $N_j$ | $\mu_j$ | $\sigma_j$ | $N_j$ | | | |
| *A. Birth cohort studies* | | | | | | | | | | | | |
| Douglas (1970) | 0.052 | | 4701 | | | | | | | D | 0-15 | NSHD |
| [Table 1 and pp. 203-204] | 0.115 | | 4701 | | | | | | | DD | 0-15 | |
| Maclean and Wadsworth (1988) | 0.066 [a] | | 3964 | 0.125 | | 2376 | | | | DD | 0-15 | NSHD |
| [Tables 1-4] | 0.141 | | 3964 | 0.117 | | 1654 | | | | DD | 0-15 | |
| | | | | 0.150 | | 1351 | | | | DD | 0-15 | |
| Kuh and Maclean (1990) | | | | | | | 0.062 | | 2548 | D | 0-15 | NSHD |
| [p. 123] | | | | | | | 0.117 | | 2548 | DD | 0-15 | |
| Rodgers (1994) [Table 2] | | | | 0.066 | | 1521 | 0.069 | | 1516 | D | 18 | NSHD |
| Ely et al. (1999) | 0.133 | | 5362 | | | | | | | DD | 15 | NSHD |
| [Table 1] | 0.122 | | 17414 | | | | | | | DD | 16 | NCDS |
| | 0.212 | | 17198 | | | | | | | DD | 16 | BCS |
| Ferri (1984) | 0.080 | | 14699 | | | | | | | DD | 7 | NCDS |
| [Table A3.1] | 0.114 | | 13860 | | | | | | | DD | 11 | |
| | 0.164 | | 11688 | | | | | | | DD | 16 | |
| Estaugh and Power (1991) | | | | 0.162 | | 3668 | 0.164 | | 3676 | DD | 0-16 | NCDS |
| [Table 1] | | | | | | | | | | | 0-16 | |
| Kiernan (1992) [Table 2] | | | | 0.115 | | 4546 [b] | 0.124 | | 4536 [b] | DD | 16 | NCDS |
| Ní Bhrolcháin et al. (1994) | | | | 0.048 | | 3926 | 0.057 | | 3940 | D | 7-16 | NCDS |
| [Table 1] | | | | 0.086 | | 3926 | 0.095 | | 3940 | DD | 7-16 | |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Chase-Lansdale et al. (1995) [Table 1] | 0.048 | 7966 | | | | | | | D | 7-16 | NCDS |
| Cherlin et al. (1995) [Table 1] | | | 0.104 | 0.369[c] | 5180[c] | 0.107 | 0.378[c] | 5209[c] | DD | 7-16 | NCDS |
| Kiernan (1997) | 0.111 | 9559 | | | | | | | D | 0-16 | NCDS |
| [Tables 1 and 2] | 0.112 | 10138 | | | | | | | D | 0-16 | |
| Rodgers et al. (1997) [Table 1] | | | 0.135 | | 4413 | 0.151 | | 4648 | DD | 0-16 | NCDS |
| Hobcraft (1998) | 0.124 | 11692 | | | | | | | D | 16 | NCDS |
| [Table 4][d] | 0.168 | 11692 | | | | | | | DD | 16 | |
| Dearden (1998) [Table A1][e] | | | 0.037 | 0.190 | 2597 | 0.055 | 0.228 | 2363 | ? | 16 | NCDS |
| Hope et al. (1998) [Table 1] | | | 0.129 | | 4606 | 0.144 | | 4892 | DD | 0-16 | NCDS |
| Gregg and Machin (1999) | | | 0.125 | | 5995 | 0.131 | | 5696 | DD | 7-16 | |
| [Tables 2a and 2b][f] | | | 0.115 | | 6381 | 0.118 | | 6135 | DD | 7-16 | NCDS |
| | | | 0.101 | | 6267 | 0.109 | | 6270 | DD | 7-16 | |
| Ní Bhrolcháin et al. (2000) [Table 1] | | | 0.092 | | 3926 | 0.102 | | 3940 | DD | 7-16 | NCDS |
| Wasdworth et al. (1983) | 0.086 | 12410 | | | | | | | D | 0-5 | BCS |
| [Table 1 and p. 101] | | | | | | | | | | | |
| Wasdworth et al. (1985) | 0.084 | 12356 | | | | | | | D | 0-5 | BCS |
| [Table 1][g] | | | | | | | | | | | |
| Machin et al. (1997) [Table 1] | | | 0.173 | | 3569 | 0.173 | | 3661 | ? | 16 | BCS |
| Cheesbrough (2001) | 0.270 | 9467 | 0.221 | | 2439 | 0.238 | | 3144 | DD | 0-16 | BCS |
| [Figures 3.3 and 3.4, Table 4.5] | | | | | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Joshi et al. (1999) [Table 1] | 0.260 | 0.440 | 1526 | | | | D | 5-17 | Second generation NCDS |
| Joshi and Verropoulou (2000) [Table A2.2] | 0.260<br>0.118[h]<br>0.142[i] | 0.439<br>0.323<br>0.349 | 1730 | | | | ? | 5-17 | Second generation NCDS |
| Joshi and McCulloch (2002) [Table 1] | 0.261 | | 1041 | | | | ? | 6-17 | Second generation NCDS |
| Verropoulou et al. (2002) [Table 1] | 0.253 | | 1472 | | | | ? | 0-17 | Second generation NCDS |
| *B. Official statistics and other studies* | | | | | | | | | |
| Green et al. (1990) [Table 1] | 0.117 | | 866 | | | | ? | 15 | Twenty-07 Study |
| Sweeting et al. (1998) [pp. 22-23] | 0.180 | | 984 | | | | DD | 15 | Twenty-07 Study |
| Monck et al. (1994) [Table 4] | | | | 0.235 | | 524 | ? | 15-20 | North-East London GP registers |
| Haskey (1994) [Table 1][j] | 0.245 | | 12764 | | | | ? | 0-16 | OPCS Omnibus Survey |
| Haskey (1997) [Tables 1 and 2] | 0.230 | | 13000[k] | | | | D | 0-15 | Official divorce statistics |
| Kiernan and Hobcraft (1997) [Table 1] | 0.172 | | 4548 | | | | DD | 0-16 | 1990-1991 NSSAL |

| | | | | | |
|---|---|---|---|---|---|
| Dunn et al. (1998) [Table 1] | 0.259 | 4071 | D | 7[l] | ALSPAC |
| O'Connor et al. (1998) | 0.145 | 11090 | D | 0[m] | ALSPAC |
| [Table 1 and p. 762] | 0.170 | 13088 | D | 0[m] | |
| Walker et al. (2001) [Table 3.8] [n] | 0.225 | 4916 | DD | 0-16 | 2000 GHS |

*Note*: The column labelled 'Parental loss information' contains the details on whether the reported figures refer to cases that include both divorce or separation and death or other reasons for family dissolution (denoted 'DD'), or to cases that exclude death (denoted 'D'), or to cases that cannot be determined with the information in the study (denoted '?'). The column labelled 'Child's age' reports the age or ages at which the reported family structure has been measured.

[a] This figure excludes family break-ups due to parental death.

[b] The exact sample sizes are not reported. These figures are from Table 3 and correspond to the largest sample sizes.

[c] Exact standard deviation for the variable pooling divorced parents and dead parents is not reported. These figures are obtained by assuming that 'parent divorced' and 'parent died' are independent and summing the two corresponding standard deviations reported in their Table 1. Similarly, the exact sample sizes are not reported. These figures are from Table 2 and correspond to the largest sample sizes.

[d] The sample size excludes missing cases.

[e] Statistics refer to young adults drawn from the 1991 sweep of the NCDS who report no father figure in 1974.

[f] The different proportions and sample sizes refer to different child outcomes.

[g] Children who live with neither parent or who experienced the death of one their parents are not included in this study.

[h] Proportion of children living with a single mother.

[i] Proportion of children living with a mother and a step-father.

[j] These statistics refer to the number of dependent children living in non-intact families in 1990-92. Dependent children are persons aged under 16, or aged 16-18 and in full-time education, in the family unit and living in the household.

[k] This figure is an average of the number of children of couples who divorced in 1994-95 and the number of children in marriage couple families in England and Wales reported in Haskey's (1997) Table 2 over the 35 birth cohorts displayed in his Table 1.

[l] Age range not reported, average age is 7.3 (S.D. is 2.7 years).

[m] All these statistics are measured during the period ranging between 18 and 32 weeks into maternal pregnancy. The top figures are computed using the largest numbers reported in the note of their Table 1. The bottom figures use the statistics discussed in page 762 of their published article.

[n] Figures are obtained from our own computations using the average number of dependent children by family type.

Table 3
Comparison results: *p*-values of the test of equality between the relevant BHPS statistics and the statistics from all other studies

| Study | $\sigma_j$ is known | $\sigma_j$ is not known and set equal to: | | |
|---|---|---|---|---|
| | | $2\times\sigma_{BHPS}$ | $\sigma_{BHPS}$ | $\sigma_{BHPS}\div 2$ |
| *A. Birth cohort studies* | | | | |
| Douglas (1970)[a] | | 0.363 | 0.073 | 0.001* |
| Maclean and Wadsworth (1988) | | | | |
| All | | 0.644 | 0.362 | 0.086 |
| Men | | | | |
| Top figure | | 0.908 | 0.819 | 0.660 |
| Middle figure | | 0.990 | 0.980 | 0.962 |
| Bottom figure | | 0.663 | 0.392 | 0.110 |
| Kuh and Maclean (1990) | | 0.346 | 0.037* | 0.000* |
| Rodgers (1994) | | | | |
| Men | | 0.522 | 0.207 | 0.018* |
| Women | | 0.145 | 0.001* | 0.000* |
| Ely et a. (1999) | | | | |
| NSHD | | 0.548 | 0.235 | 0.023* |
| NCDS | | 0.565 | 0.252 | 0.024* |
| BCS | | 0.521 | 0.200 | 0.012* |
| Ferri (1984)[b] | | | | |
| Top figure (age 7) | | 0.821 | 0.652 | 0.375 |
| Middle figure (age 11) | | 0.717 | 0.470 | 0.156 |
| Bottom figure (age 16) | | 0.857 | 0.721 | 0.485 |
| Estaugh and Power (1991) | | | | |
| Men | | 0.556 | 0.245 | 0.024* |
| Women | | 0.814 | 0.641 | 0.368 |
| Kiernan (1992) | | | | |
| Men | | 0.972 | 0.935 | 0.877 |
| Women | | 0.465 | 0.147 | 0.005* |
| Ní Bhrolcháin et al. (1994) | | | | |
| Men | | 0.669 | 0.395 | 0.099 |
| Women | | 0.276 | 0.031* | 0.000* |
| Chase-Lansdale et al. (1995) | | 0.057 | 0.000* | 0.000* |
| Cherlin et al. (1995) | | | | |
| Men | 0.742 | | | |
| Women | 0.055 | | | |
| Kiernan (1997)[a] | | 0.440 | 0.125 | 0.003* |
| Rodgers et al. (1997) | | | | |
| Men | | 0.820 | 0.651 | 0.379 |
| Women | | 0.692 | 0.431 | 0.126 |
| Hobcraft (1998) | | 0.590 | 0.284 | 0.036* |
| | | 0.801 | 0.617 | 0.328 |
| Dearden (1998) | | | | |
| Men | 0.000* | | | |
| Women | 0.000* | | | |
| Hope et al. (1998) | | | | |
| Men | | 0.883 | 0.770 | 0.569 |
| Women | | 0.629 | 0.337 | 0.062 |
| Gregg and Machin (1999)[a] | | | | |

| | | | | |
|---|---|---|---|---|
| Men | | 0.935 | 0.852 | 0.715 |
| Women | | 0.519 | 0.200 | 0.012* |
| Ní Bhrolcháin et al. (2000) | | | | |
| Men | | 0.763 | 0.490 | 0.180 |
| Women | | 0.316 | 0.047* | 0.000* |
| Wadsworth et al. (1983)[c] | | 0.750 | 0.526 | 0.215 |
| Wadsworth et al. (1985)[c] | | 0.718 | 0.472 | 0.160 |
| Machin et al. (1997) | | | | |
| Men | | 0.269 | 0.028* | 0.000* |
| Women | | 0.531 | 0.215 | 0.017* |
| Cheesbrough (2001) | | | | |
| All | | 0.838 | 0.685 | 0.429 |
| Men | | 0.567 | 0.222 | 0.020* |
| Women | | 0.916 | 0.835 | 0.688 |
| Joshi et al. (1999) | 0.884 | | | |
| Joshi and Verropoulou (2000)[a] | 0.883 | | | |
| Joshi and McCulloch (2002) | | 0.923 | 0.860 | 0.786 |
| Verropoulou et al. (2002) | | 0.953 | 0.913 | 0.859 |
| | | | | |
| *B. Official statistics and other studies* | | | | |
| Green et al. (1990) | | 0.009* | 0.000* | 0.000* |
| Sweeting et al. (1998) | | 0.199 | 0.022* | 0.001* |
| Monck et al. (1994) | | | | |
| Birth cohort: 1974-80 | | 0.777 | 0.660 | 0.585 |
| Birth cohort: 1979-80 | | 0.843 | 0.718 | 0.575 |
| Haskey (1994) | | 0.938 | 0.879 | 0.787 |
| Haskey (1997) | | | | |
| Birth cohort: 1974-80 | 0.489 | | | |
| Birth cohort: 1979-80 | 0.297 | | | |
| Kiernan and Hobcraft (1997) | | 0.321 | 0.139 | 0.077 |
| Dunn et al. (1998) | | 0.952 | 0.908 | 0.833 |
| O'Connor et al. (1998)[d] | | | | |
| Top figure (Table 2) | | 0.012* | 0.000* | 0.000* |
| Bottom figure (Table 2) | | 0.001* | 0.000* | 0.000* |
| Walker et al. (2001) | | 0.537 | 0.227 | 0.026* |

*Note*: $\sigma_{BHPS}$ refers to the standard deviation observed in the corresponding sample of the BHPS (see Table 1).

[a] Refers to the top figure in the relevant box of Table 2.

[b] The corresponding BHPS figures make use of FLVHMAG to select only the group of individuals who experienced life in a non-intact family by age 7 and by age 11 (not reported in Table 1). The BHPS $\mu$ and $\sigma$ for the first age group are 0.071 and 0.258, while for the second age group are 0.133 and 0.340, respectively.

[c] The corresponding BHPS figures make use of FLVHMAG to select only the group of individuals who experienced life in a non-intact family during the ages 0-5. The BHPS $\mu$ and $\sigma$ for this group of individuals are 0.101 and 0.301, respectively.

[d] The corresponding BHPS figures make use of FLVHMAG to select only the group of individuals who experienced life in a non-intact family at birth. The BHPS $\mu$ and $\sigma$ for this group of individuals are 0.065 and 0.247, respectively.

"*" indicates that the BHPS figures and the corresponding figures from the birth cohort studies are significantly different from each other ($p<0.05$).

Table 4: Comparing the wave-6 data with other statistics derived from a special BHPS sample of young adults

| Developmental stage and variables | $\mu_{ENIF}$ | $\mu_{FBPAR16}$ | $\eta_{01}$ | $\eta_{10}$ | $\kappa$ | Tests of equality of:[c] Means | Medians |
|---|---|---|---|---|---|---|---|
| *(a) All sample* | | | | | | | |
| 0-15[a] | 0.239 | 0.221 | 0.038 | 0.019 | 0.838** | 0.005† | 0.007† |
| 0-5[b] | 0.107 | 0.101 | 0.031 | 0.023 | 0.716* | 0.186 | 0.228 |
| 6-10[b] | 0.076 | 0.067 | 0.029 | 0.019 | 0.694* | 0.075 | 0.098 |
| 11-15[b] | 0.054 | 0.054 | 0.016 | 0.016 | 0.868** | 0.876 | 0.998 |
| *(b) By gender* (0-15)[a] | | | | | | | |
| Men | 0.251 | 0.236 | 0.034 | 0.019 | 0.856** | 0.096 | 0.133 |
| Women | 0.227 | 0.204 | 0.043 | 0.020 | 0.815** | 0.023† | 0.034† |
| Pearson $\chi^2$ (1 d.f.) | 1.057 | 1.995 | | | | | |
| *(c) By age* (0-15)[a] | | | | | | | |
| Less than 19 | 0.235 | 0.212 | 0.035 | 0.012 | 0.865** | 0.056 | 0.063 |
| [19-21) | 0.285 | 0.269 | 0.035 | 0.019 | 0.866** | 0.286 | 0.424 |
| [21-24) | 0.222 | 0.214 | 0.035 | 0.027 | 0.817** | 0.532 | 0.678 |
| 24 or more | 0.226 | 0.194 | 0.052 | 0.020 | 0.785* | 0.059 | 0.096 |
| Pearson $\chi^2$ (3 d.f.) | 3.809 | 4.826 | | | | | |
| Kruskal-Wallis $\chi^2$ (3 d.f.) | 3.806 | 4.823 | | | | | |
| *(d) By education* (0-15)[a] | | | | | | | |
| No qualification | 0.392 | 0.351 | 0.072 | 0.031 | 0.779* | 0.208 | 0.344 |
| Less than GCSE/O level | 0.272 | 0.248 | 0.040 | 0.016 | 0.855** | 0.259 | 0.453 |
| GCSE/O level | 0.260 | 0.231 | 0.044 | 0.016 | 0.838** | 0.012† | 0.019† |
| A level | 0.175 | 0.169 | 0.029 | 0.023 | 0.816** | 0.638 | 0.814 |
| Higher vocational qualifications | 0.239 | 0.234 | 0.027 | 0.021 | 0.868** | 0.740 | 0.995 |
| University or more | 0.169 | 0.157 | 0.024 | 0.012 | 0.867** | 0.567 | 0.875 |
| Pearson $\chi^2$ (5 d.f.) | 24.246‡ | 17.808‡ | | | | | |
| Kruskal-Wallis $\chi^2$ (5 d.f.) | 24.227‡ | 17.794‡ | | | | | |
| *(e) By marital status* (0-15)[a] | | | | | | | |
| Married | 0.118 | 0.092 | 0.026 | 0.000 | 0.861** | 0.159 | 0.497 |
| Cohabiting | 0.294 | 0.302 | 0.040 | 0.048 | 0.791* | 0.764 | 0.986 |
| Never married | 0.242 | 0.221 | 0.039 | 0.018 | 0.842** | 0.003† | 0.004† |
| Pearson $\chi^2$ (2 d.f.) | 8.174‡ | 12.095‡ | | | | | |
| *(f) By employment status* (0-15)[a] | | | | | | | |
| In work | 0.232 | 0.218 | 0.037 | 0.023 | 0.828** | 0.090 | 0.119 |
| Unemployed | 0.300 | 0.244 | 0.067 | 0.011 | 0.805* | 0.058 | 0.125 |
| Other | 0.241 | 0.222 | 0.033 | 0.014 | 0.868** | 0.091 | 0.144 |
| Pearson $\chi^2$ (2 d.f.) | 2.048 | 0.333 | | | | | |

*Note*: The sample size is 1286. The terms $\mu_{ENIF}$ and $\mu_{FBPAR16}$ denote the proportions of individuals who have experienced life in a non-intact family during childhood in the special sample of young adults and in the sixth wave of the BHPS respectively. The terms $\eta_{01}$ and $\eta_{01}$ denote the proportion of individuals for whom FBPAR16=0 and ENIF=1 and the proportion of individuals for whom FBPAR16=1 and ENIF=0, respectively.

[a] Statistics for the wave 6 data are constructed using FBPAR16.

[b] Statistics for the wave 6 data are constructed using FLVHMAG.

[c] Tests are t-tests for equality of means and sing tests for equality of medians in $\mu_{ENIF}$ and $\mu_{FBPAR16}$. The figures in the tables are *p*-values of the test of equality. A '†' indicates that the statistics are significantly different at the 0.05 level or less.

"**" denotes that agreement is 'almost perfect', "*" denotes that agreement is 'substantial' (see Landis and Koch (1977, p. 165)).

A '‡' indicates that the corresponding proportions are significantly different at the 0.05 level or less.

Table 5
Logistic regression analysis: factors predicting "mismatch" in family structure information (Odds ratios)

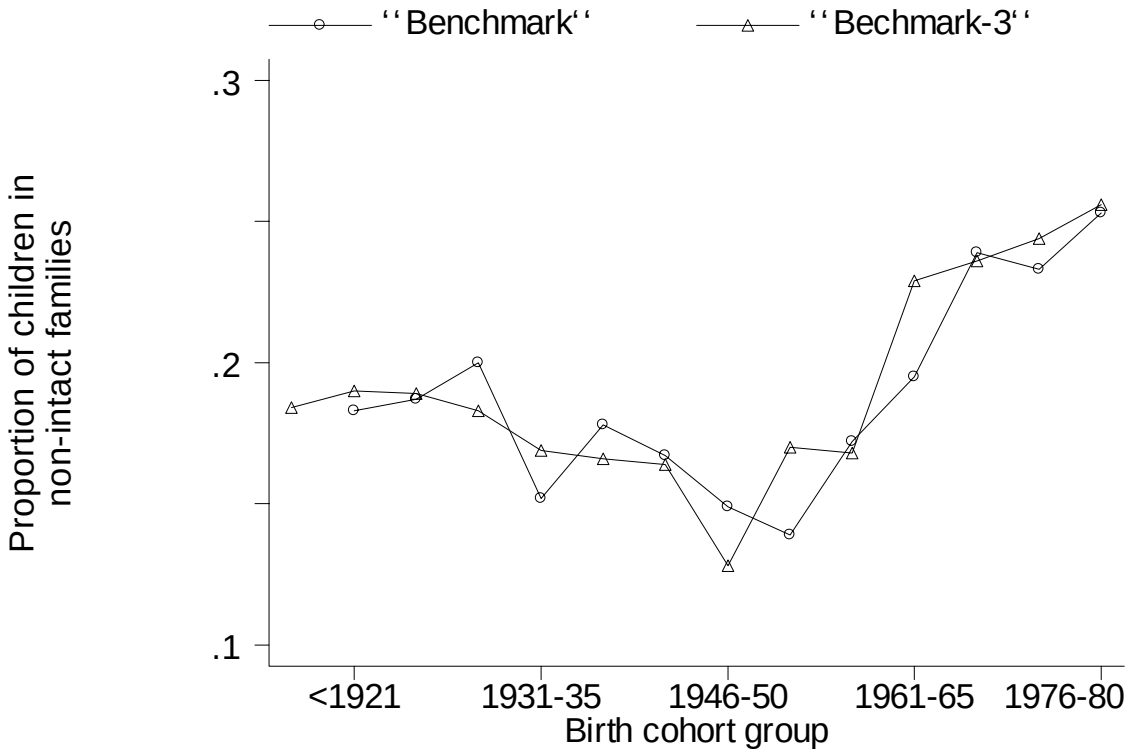
| Variable | Specification [1] | [2] | [3] | Mean |
|---|---|---|---|---|
| *Child's characteristics* (**C**) | | | | |
| Age | 1.096* | 1.122* | 1.173** | 20.982 |
| Female§ | 1.321 | 1.413 | 1.542 | 0.474 |
| Ethnic origin§ | | | | |
| White [=base] | | | | 0.931 |
| Black | 0.497 | 0.427 | 0.166 | 0.020 |
| Indian | 1.666 | 1.889 | 0.811 | 0.021 |
| Pakistani/Bangladeshi | 2.765 | 4.198 | 6.400** | 0.015 |
| Other | 2.237 | 2.075 | 1.346 | 0.013 |
| Number of siblings | 1.132 | 1.168 | 1.074 | 1.652 |
| Employment status§ | | | | |
| In work | 1.342 | 1.149 | 1.168 | 0.649 |
| Unemployed | 1.287 | 1.151 | 1.243 | 0.070 |
| Other (e.g., in education and on government training) [=base] | | | | 0.281 |
| Education§ | | | | |
| No qualification [=base] | | | | 0.075 |
| Less than GCSE/O level | 0.469 | 0.525 | 0.345 | 0.097 |
| GCSE/O level | 0.613 | 0.714 | 0.842 | 0.350 |
| A level | 0.490 | 0.572 | 0.608 | 0.267 |
| Higher vocational qualifications | 0.359* | 0.475 | 0.377 | 0.146 |
| University or higher degrees | 0.254* | 0.330 | 0.313 | 0.065 |
| *Mother's characteristics* (**M**) | | | | |
| Age at child's birth§ | | | | |
| Less than 22 | | 3.921** | 3.701* | 0.114 |
| Aged 22-34 [=base] | | | | 0.842 |
| More than 34 | | 2.217 | 2.741 | 0.044 |
| Education§ | | | | |
| No qualification [=base] | | | | 0.323 |
| Less than GCSE/O level | | 1.352 | 1.189 | 0.111 |
| GCSE/O level | | 2.010* | 2.349* | 0.208 |
| A level | | 1.881 | 1.632 | 0.065 |
| Higher vocational qualifications | | 0.468 | 0.433 | 0.220 |
| University or higher degrees | | 0.395 | 0.393 | 0.073 |
| Hope-Goldthorpe score of prestige over child's childhood | | 0.977 | 0.985 | 37.618 |
| *Interview characteristics and conditions (interviewer's report)* (**Z**) | | | | |
| Child (wave 6) | | | | |
| Child's interview was influenced by others§ | | | 1.420 | 0.038 |
| Child did not cooperate§ | | | 1.360 | 0.026 |
| Mother (wave 2) | | | | |
| Her interview was influenced by others§ | | | 9.054** | 0.028 |

| | | | | |
|---|---|---|---|---|
| She did not cooperate[§] | | | 7.097** | 0.043 |
| Marital/fertility section: | | | | |
| Partner was present[§] | | | 1.907 | 0.062 |
| Child(ren) was(were) present[§] | | | 0.412 | 0.026 |
| Other adults were present[§] | | | 1.195 | 0.076 |
| Mean of the dependent variable | | 0.057 | | |
| Wald $\chi^2$ [p-value] | 41.64 [0.0020] | 61.62 [0.0006] | 168.25 [0.0000] | |
| Correctly predicted (%) | 94.1 | 94.3 | 95.6 | |
| Pearson $\chi^2$ [p-value] | 1157.49 [0.644] | 1155.97 [0.821] | 1152.62 [0.968] | |
| Area under ROC curve | 0.674 | 0.740 | 0.849 | |
| Psuedo $R^2$ | 0.038 | 0.100 | 0.245 | |
| Number of young adults | | 1286 | | 1286 |

*Note*: The dependent variable takes value one if there is mismatch between ENIF and FBPAR16 (i.e., if ENIF=0 *and* FBPAR16=1 or ENIF=1 *and* FBPAR16=0) and takes value zero otherwise. Standard errors account for the fact that individuals may come from the same family (i.e., observations are repeated on siblings or half-siblings). Other variables included in specification [1] are dummy variables for child's region of residence (6), marital status (2) and housing tenure (2). In addition to such variables, specifications [2] and [3] include also dummy variables for mother's region of residence (6), mother's marital status (2) and mother's housing tenure (2) (all measured as of wave 2).
"§" indicates dummy variable(s).
"**" , "*" denote statistically significant at 0.01, 0.05 level respectively.

Figure 1

Proportion of children who ever lived in a non-intact family during childhood by birth cohort (unweighted data)



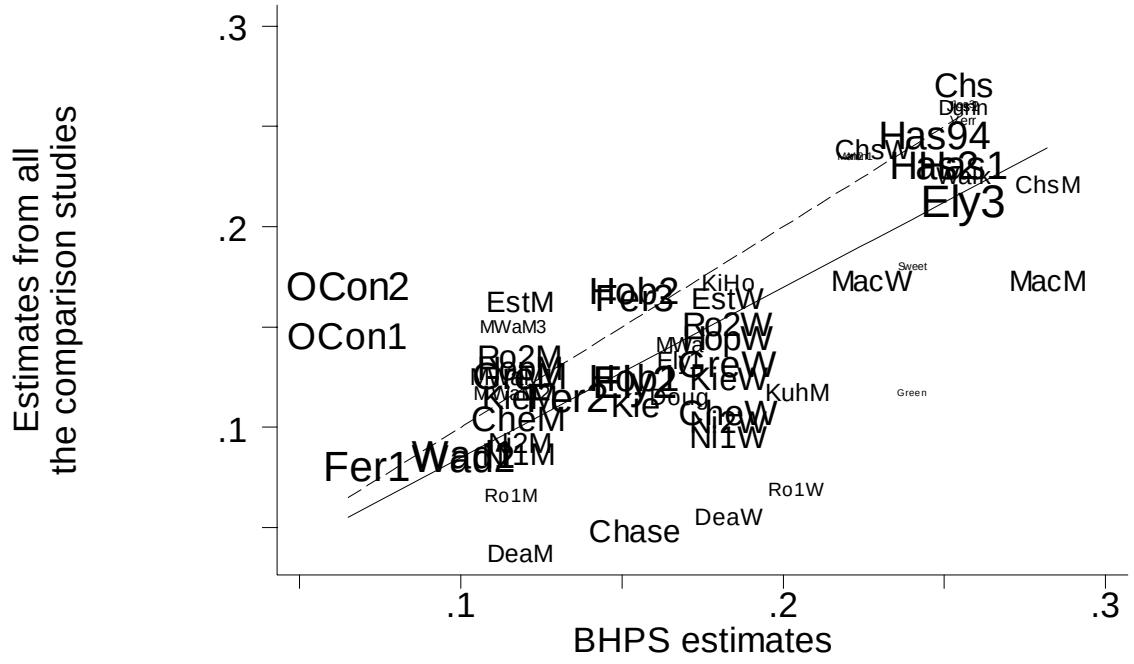

*Note*: "Benchmark" and "Benchmark-3" refer to the way in which birth cohorts have been grouped. Under "Benchmark", cohorts are grouped in thirteen categories, with all individuals born before 1921 in one group, and twelve groups of individuals who are born in 5-year windows between 1921 and 1980. "Benchmark-3" is defined analogously, with the differences that the first group refers to individuals born before 1919, the last group includes individuals born between 1979 and 1980, and the other twelve groups refer to individuals who are born in 5-year windows between 1919 and 1978.

Figure 2
Proportion of children who ever lived in a non-intact family during childhood by birth cohort and gender (unweighted data)


*Note*: Cohort groups are defined according to the "benchmark" case of Figure 1.

Figure 3
Weighted least-square regression for all estimates on BHPS statistics — Estimation prediction (solid line) and 45 degree line (dashed)

$$E(\mu_j) = 0.848\ \mu_{BHPS}, \qquad \overline{R}^2 = 0.928, \quad N=59$$
$$(0.031)$$

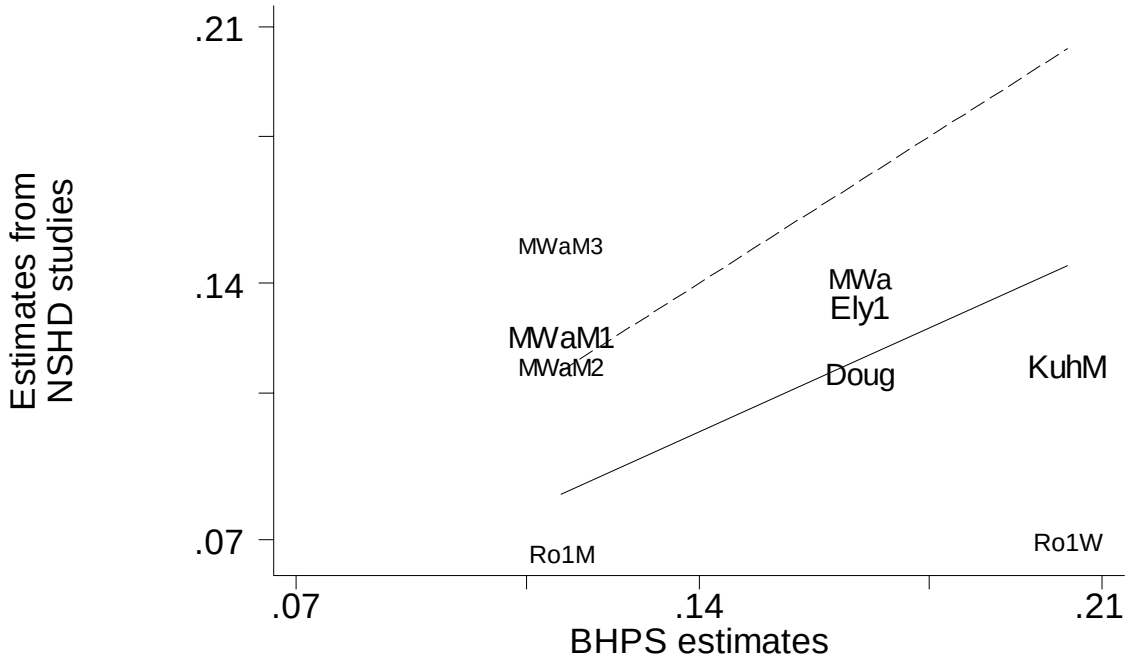



*Note*: Standard error in parentheses. Statistics (means) of each paper are weighted by the corresponding sample size (reported in Table 2). In the figure, the scaling of each statistic is proportional to the sample size from which it is taken or computed.

*Legend*: Doug=Douglas (1970); Ro1M=Rodgers (1994) (men); Ro1W=Rodgers (1994) (women); Ely1=Ely et al. (1999) (NSHD); MWa=Maclean and Wadsworth (1988); MWaM1=Maclean and Wadsworth (1988) (men, top figure); MWaM2= Maclean and Wadsworth (1988) (men, middle figure); MWaM3= Maclean and Wadsworth (1988) (men, bottom figure); KuhM=Kuh and Maclean (1990); Ely2=Ely et al. (1999) (NCDS); Fer1=Ferri (1984) (top figure); Fer2=Ferri (1984) (middle figure); Fer3=Ferri (1984) (bottom figure); KieM=Kiernan (1992) (men); KieW=Kiernan (1992) (women); CheM=Cherlin et al. (1995) (men); CheW=Cherlin et al. (1995) (women); Kie=Kiernan (1997); Hob1=Hobcraft (1998) (top figure); Hob2=Hobcraft (1998) (bottom figure); DeaM=Dearden (1998) (men); DeaW=Dearden (1988) (women); GreM=Gregg and Machin (1999) (men); GreW=Gregg and Machin (1999) (women); Ni2M=Ní Bhrolcháin et al. (2000) (men); Ni2W=Ní Bhrolcháin et al. (2000) (women); Chase=Chase-Lansdale et al. (1995); Ro2M=Rodgers et al. (1997) (men); Ro2W=Rodgers et al. (1997) (women); EstM=Estaugh and Power (1991) (men); EstW=Estaugh and Power (1991) (women); Ni1M=Ní Bhrolcháin et al. (1994) (men); Ni1W=Ní Bhrolcháin et al. (1994) (women); HopM=Hope et al. (1998) (men); HopW=Hope et al. (1998) (women); Ely3=Ely et al. (1999) (BCS); MacM=Machin et al. (1997) (men); MacW=Machin et al. (1997) (women); Wad1=Wadsworth et al. (1983); Wad2=Wadworth et al. (1985); Chs=Cheesbrough (2001) (all); ChsM=Cheesbrough (2001) (men); ChsW=Cheesbrough (2001) (women); Jos1=Joshi et al. (1999); Jos2=Joshi and Verropoulou (2000); Jos3=Joshi and McCulloch (2002); Verr=Verropoulou et al. (2002); Green=Green et al. (1990); Sweet=Sweeting et al. (1998); Mon1=Monck et al. (1994) (birth cohorts 1974-80); Mon2=Monck et al. (1994) (birth cohorts, 1979-80); KiHo=Kiernan and Hobcraft (1997); Has94=Haskey (1994); Has1=Haskey (1997) (birth cohorts=1974-80); Has2=Haskey (1997) (birth cohorts=1979-80); Dunn=Dunn et al. (1998); Walk=Walker et al. (2001); OCon1=O'Connor et al. (1998) (top figure); OCon2=O'Connor et al. (1998) (bottom figure).

Figure 4
Weighted least-square regression for the NSHD estimates on the relevant BHPS statistics — Estimation prediction (solid line) and 45 degree line (dashed)

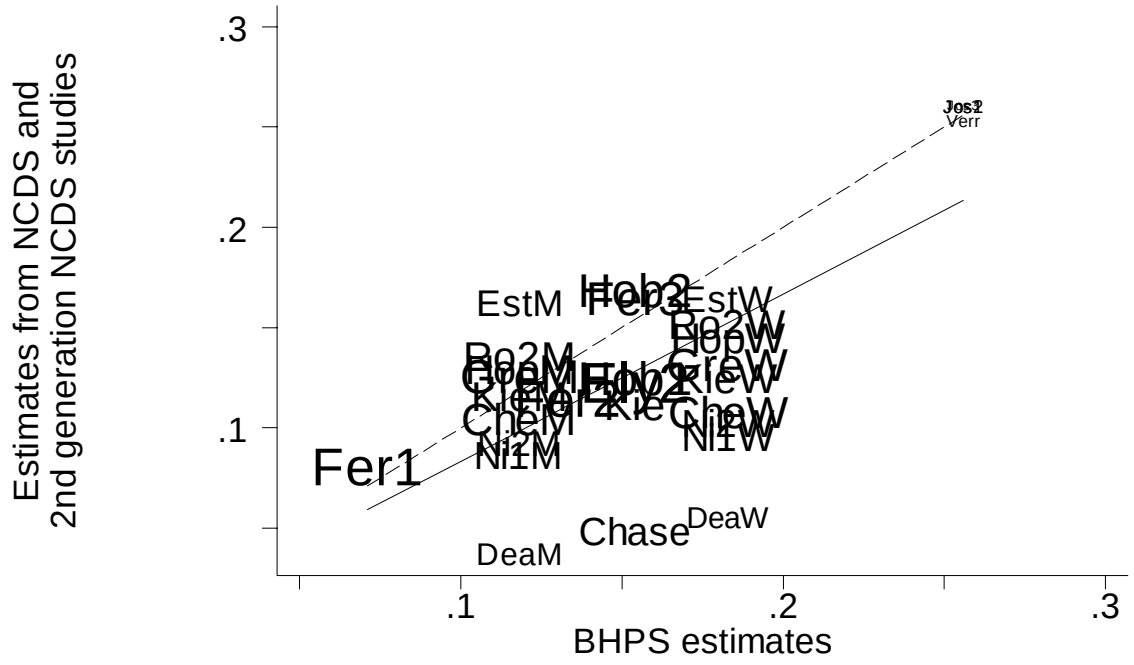
$$E(\mu_{NSHD}) = \underset{(0.082)}{0.710}\ \mu_{BHPS}, \qquad \bar{R}^2 = 0.890, \qquad N=9$$




*Note*: See note and legend of Figure 3.

Figure 5
Weighted least-square regression for the NCDS and second generation NCDS estimates on the relevant BHPS statistics — Estimation prediction (solid line) and 45 degree line (dashed)

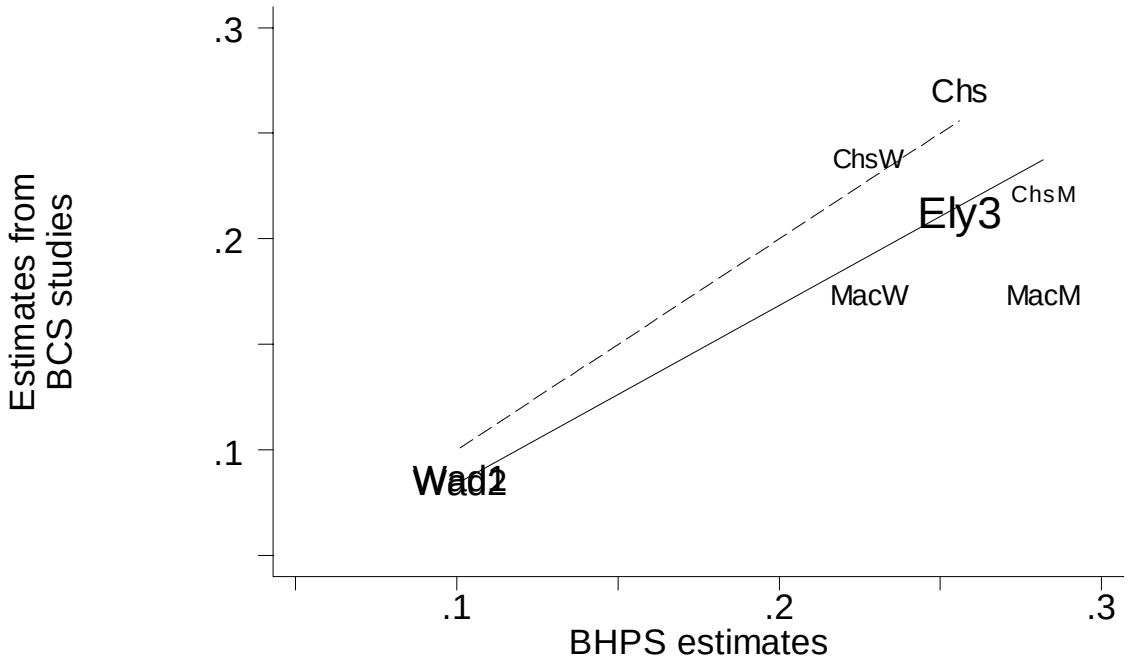
$$E(\mu_{NCDS}) = \underset{(0.044)}{0.833}\ \mu_{BHPS}, \qquad \bar{R}^2 = 0.922, \quad N=30$$




*Note*: See note and legend of Figure 3.

Figure 6
Weighted least-square regression for the BCS estimates on the relevant BHPS statistics — estimation prediction (solid line) and 45 degree line (dashed)

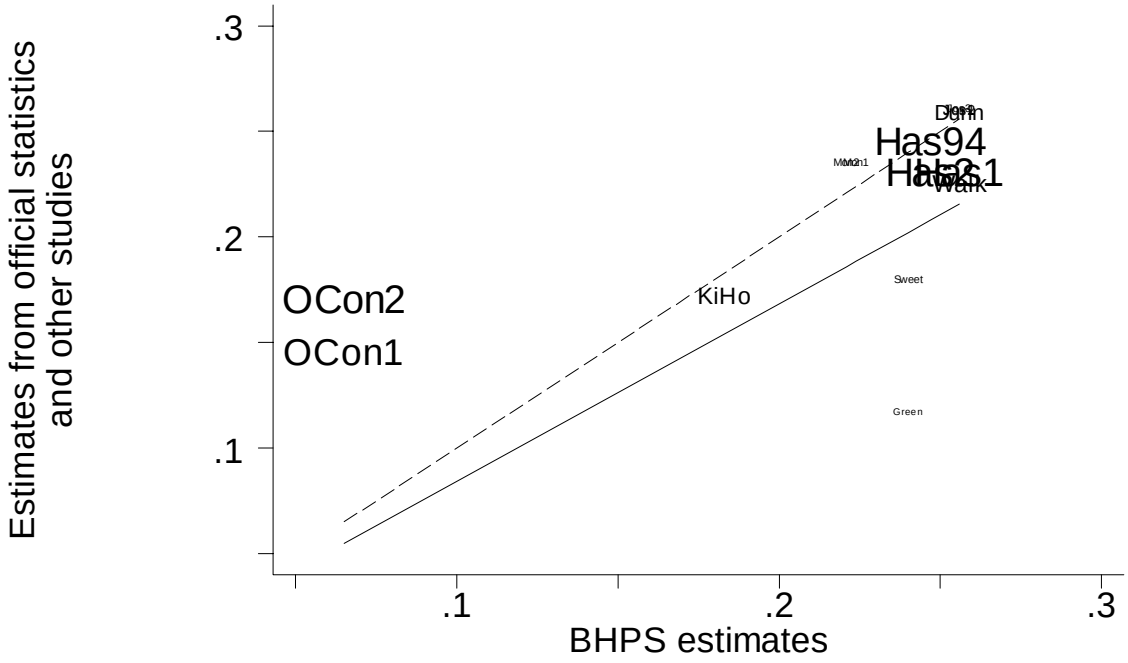
$$E(\mu_{BCS}) = \underset{(0.054)}{0.842}\ \mu_{BHPS}, \qquad \overline{R}^2 = 0.968, \quad N=8$$




*Note*: See note and legend of Figure 3.

Figure 7
Weighted least-square regression for the estimates from official divorce statistics and other (non-birth-cohort) studies on the relevant BHPS statistics — estimation prediction (solid line) and 45 degree line (dashed)

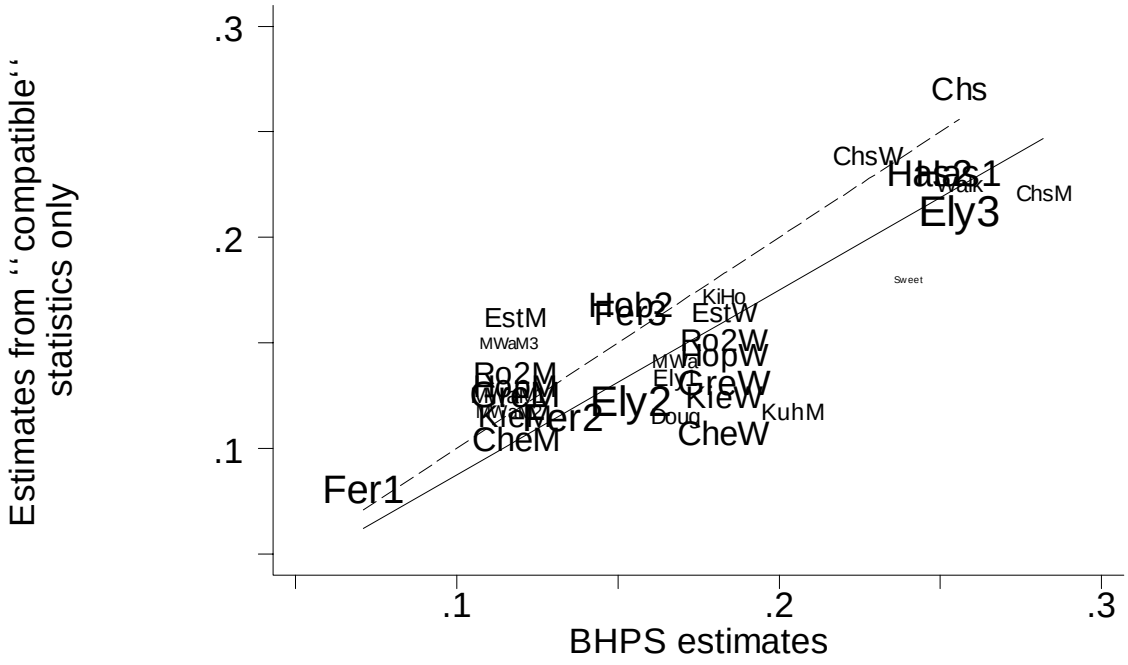
$$E(\mu_{OTHER}) = \underset{(0.081)}{0.977}\ \mu_{BHPS}, \qquad \bar{R}^2 = 0.924, \qquad N=12$$




*Note*: See note and legend of Figure 3.

Figure 8
Weighted least-square regression for the estimates from all studies in Table 2 with comparable information to the BHPS data ("compatible" studies) on the relevant BHPS statistics — estimation prediction (solid line) and 45 degree line (dashed)

$$E(\mu_j) = \underset{(0.028)}{0.875}\ \mu_{BHPS}, \qquad \overline{R}^2 = 0.967, \qquad N=33$$




*Note*: See note and legend of Figure 3.