**Online Supplement**

**Appendix 1 – Multiple Imputation Models and Checks**

The aim of multiple imputation is to yield valid analysis results by restoring sample representativeness with regard to the defined target population (Mostafa *et al.* 2021). In our analysis the target population can be summarised as all people born in the cohort year who are alive and living in Great Britain at the age we analyse *and* have at least one living parent. The first condition for defining the target population is straightforward for the 1958 and 1970 cohorts where identification of cohort members who have died or emigrated is available in a separate ‘response’ dataset. For the 1946 cohort this information is not available and, therefore, we exclude cohort members who never participated in the survey after the age of 16. The second condition for the target population is more complex because information about cohort members parents’ deaths is not available through external sources and is conditional on the cohort members having participated in a previous sweep where this was asked. This introduces some noise since some cohort members will likely be included for imputation despite no longer having a living parent. This particularly affects the later ages when the experience of parental death is more common (i.e. ages 55 and 63 in our analysis). On the other hand, relaxing this condition would not be appropriate as individuals known not to have a living parent to potentially provide help or care for would be included and have help/caregiving values imputed.

We therefore excluded individuals known to have died or emigrated and those known to have no living parent (as reported prior to or at the sweep analysed, see Table A1.1 below) prior to imputation. As Table A1.1 shows, non-participation in the wave is the primary reason for reduction in sample size, accounting for 48-53 per cent of the missing data, with item missingness accounting for a minor reduction in complete case numbers (after exclusion of those without a living parent). Following the steps outlined by Carpenter and Kenward (2012) we ran chained equations imputations including all variables in the substantive models (the outcomes and covariates outlined in the Data section) plus auxiliary variables. The candidate auxiliary variables used are specified in the missing data user guide for the 1958 study (Silverwood *et al.* 2020), and we used similar measures for the other cohorts, including auxiliary variables that were significantly related to missingness and the substantive model outcome. We created 50 imputed data sets separately for each age and sample specification (all individuals for parent-care provision and well-being analyses; helpers/caregivers for analysis of care intensity). The pre-imputation sample sizes available for complete case analysis at ages 38, 42, 50, 55 and 63 were 6,332; 6,986; 4,073; 3,148 and 364, respectively; the respective analysis sample sizes following multiple imputation were: 17,255; 16,703; 12,775; 11,339 and 2,364. This means that we impute the outcome variable for a large proportion of our MI samples, especially for the 1946 cohort.

Although the amount of imputed cases, due to a combination of the long-running nature of the studies and the uncertainty around the exclusion of cohort members without a living parent, is of some concern, the alternative of complete case analysis could also introduce bias. Best practice advice on missing data suggest using techniques such as MI to improve the plausibility of the missing at random assumption (given the covariates and auxiliary variables Graham 2009, Jeličić, Phelps and Lerner 2009, Mostafa *et al.* 2021) and simulation suggests that correctly specified MI can reduce bias and improve efficiency under the missing at random assumption, even at high proportions of missing data (up to 90%; Madley-Dowd *et al.* 2019). The rich availability of variables measured at the time of the cohort members birth or in in childhood, and thus prior to non-response, that can be included in the imputation models is a strength of the cohort data. We discuss the checks we have run for all datasets, and additional checks specifically for the 1946 data further below.

We checked the plausibility of the imputed values using summary statistics for variables included in the analysis model and compared these with results prior to imputation (Tables A1.2 – A1.4). The summary statistics show that for the most part the imputations did not substantially change the sample proportions or averages for key analysis variables, including the proportions providing no assistance, help or care, or on average the intensity (weekly hours). The one exception is that for the 1946 cohort the imputation slightly reduces the proportion of the sample providing care to a parent both for the caregiving analysis (0.32 imputed compared with 0.35 observed) and the intensity analysis (0.59 imputed compared with 0.67 observed). Similarly, the distribution of hours spent providing care changed with a somewhat higher proportion providing 20 hours or more care a week (0.19 imputed, 0.15 observed), reflecting the higher proportion in the imputed data who provide ‘other’ (e.g. spousal) informal care only. The change is plausible given that the 1946 cohort was asked first whether they provide care to someone due to disability, illness or old age, followed by a question about the recipients of care which allowed us to identify parent-carers. This suggests to us that the change in the proportion reflects a rebalancing between parent-care and other care (including spousal care, which tends to be more intensive). Further, the small magnitude of these differences does not indicate cause for concern.

Across all cohorts, the imputations slightly reduced the proportion of cohort members with a mother who stayed in education past compulsory schooling age, increased the proportion with separated parents and reduced the proportion with both parents (rather than either) alive. These differences are to be expected since attrition in the cohort studies is related to social class and social disadvantage.

In the 1970 cohort the proportion of only children is somewhat higher after imputation (e.g. 0.098 after imputation compared with 0.075 in the observed data at age 38). In the 1946 cohort the proportion of only children at age 63 is somewhat higher in both the observed data (0.151) and the imputed data (0.150) for the caregiving analysis than the proportion observed in the childhood data at age 11 (0.14). However, this could well be because of the difference in target sample since the 63 observed sample is restricted to those with at least one living parent and the imputation sample similarly excludes respondents who have previously reported that both their parents have died. Overall, we conclude for both the 1970 and the 1946 cohorts, although the imputed estimates of the proportion of only children differ somewhat from the observed data at the same age (1970) or in childhood (1946), substantively it is of the expected order of magnitude and thus not implausible. In addition, it is important to bear in mind that, although vastly implausible imputed values may be an indication of problems with the imputation model, the aim of MI is not to fill in the values that would have been recorded had the missing respondent taken part but instead to “produce valid analytic results in the presence of missing data” and therefore “it may be more important that relationships between variables are preserved during the imputation process” (Nguyen, Carlin and Lee 2017, p. 4).

With this in mind, we also compared our substantive caregiving models run on imputed and on the observed data and the substantive conclusions were not altered. Tables A1.5 and A1.6 show the coefficients for only child from the fully adjusted models reported in Tables 2 and 3 in the main manuscript alongside the estimates obtained from the same model specification run on observed data (complete cases). The point estimates and confidence intervals are of a similar magnitude and generally in the same direction both in the complete cases and multiple imputation models. The exception is the intensity analysis where at age 42 and 50 the sign of the point estimates for only children coefficient reverses when comparing the complete cases analysis and MI analysis. However, at both ages the confidence intervals for these estimates include zero in both models. Thus, our substantive interpretation holds that among adult children who provide assistance the intensity of assistance does not differ by sibling status at these ages.

Finally, for the 1946 cohort, because of the particularly high levels of imputed values, we also used posterior predictive checking to assess the adequacy of imputation models with respect to regression coefficients for the analysis models (Tables A1.7 and A1.8). The posterior predictive p-values did not indicate problems with the imputation models (for a detailed example of using posterior predictive checking for MI models see Nguyen, Carlin and Lee 2017).

**Table A1.1 Analysis sample: Comparison of complete cases and multiple imputation sample sizes**

|  |  |  |  |
| --- | --- | --- | --- |
| Cohort | 1970 | 1958 | 1946 |
| Age | 38 | 42 | 50 | 55 | 63 |
| Participated in Wave | 8,874 | 9,841 | 9,790 | 9,137 | 2,662 |
| Reported at least 1 living parent 1 | 8,167 | 9,213 | 7,093 | 5,530 | 459 |
| Non-missing on sibling status & caregiving status | 7,161 | 8,037 | 5,845 | 4,503 | 435 |
| Complete cases: Non-missing on all covariates | 6,332 | 6,986 | 4,073 | 3,148 | 364 |
| Original sample size 2 | 18,637 | 18,637 | 18,558 | 18,558 | 5,362 |
| Emigrated or died 3 | 1,346 | 1,406 | 2,752 | 2,945 | 732 |
| Reported no living parent 1 | 36 | 528 | 3,031 | 4,274 | 2,266 |
| MI sample | 17,255 | 16,703 | 12,775 | 11,339 | 2,364 |

*Notes: 1 Reported at any data collection sweep up to and including the sweep analysed. Questions asked at ages: 30, 38 & 42 (1970); 42, 46, 50 & 55 (1958) and 53 & 63 (1946). 2 Present in the outcome file and non-missing on cohort member sex. 3 Emigration or cohort member death not available for 1946 cohort; we exclude individuals who have not participated in the study since age 16.*

**Table A1.2 Summary statistics of analysis variables for imputed and raw data – 1970 Cohort**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cohort: 1970 | Age 38 – Caregiving | Age 38 - Intensity | Age 42 - Caregiving | Age 42 - Intensity |
|  | Observed |  | MI |  | Observed |  | MI |  | Observed |  | MI |  | Observed |  | MI |  |
|  | Prop | se | Prop | se | Prop | se | Prop | se | Prop | se | Prop | se | Prop | se | Prop | se |
| No help/care | 0.645 | 0.005 | 0.620 | 0.006 |   |   |   |   | 0.509 | 0.005 | 0.489 | 0.006 |   |   |   |   |
| Helper | 0.223 | 0.005 | 0.236 | 0.005 | 0.624 | 0.009 | 0.616 | 0.010 | 0.316 | 0.005 | 0.322 | 0.006 | 0.643 | 0.007 | 0.631 | 0.009 |
| Carer | 0.132 | 0.004 | 0.144 | 0.004 | 0.376 | 0.009 | 0.384 | 0.010 | 0.175 | 0.004 | 0.189 | 0.005 | 0.357 | 0.007 | 0.369 | 0.009 |
| *Obs* | *8,556* |  | *17,255* |  | *3,003* |  | *11,702* |  | *9,234* |  | *16,703* |  | *4,539* |  | *12,003* |  |
| Sibling | 0.925 | 0.003 | 0.902 | 0.003 | 0.921 | 0.005 | 0.894 | 0.004 | 0.924 | 0.003 | 0.904 | 0.003 | 0.917 | 0.004 | 0.890 | 0.005 |
| Only child | 0.075 | 0.003 | 0.098 | 0.003 | 0.079 | 0.005 | 0.106 | 0.004 | 0.076 | 0.003 | 0.096 | 0.003 | 0.083 | 0.004 | 0.110 | 0.005 |
| *Obs* | *7,758* |  | *17,255* |  | *2,850* |  | *11,702* |  | *8,127* |  | *16,703* |  | *4,018* |  | *12,003* |  |
| Male | 0.474 | 0.005 | 0.513 | 0.004 | 0.454 | 0.009 | 0.527 | 0.005 | 0.481 | 0.005 | 0.514 | 0.004 | 0.917 | 0.004 | 0.531 | 0.005 |
| Female | 0.526 | 0.005 | 0.487 | 0.004 | 0.546 | 0.009 | 0.473 | 0.005 | 0.519 | 0.005 | 0.486 | 0.004 | 0.083 | 0.004 | 0.469 | 0.005 |
| *Obs* | *8,838* |  | *17,255* |  | *3,285* |  | *11,702* |  | *9,313* |  | *16,703* |  | *4,018* |  | *12,003* |  |
| Mother stayed in school | 0.388 | 0.005 | 0.355 | 0.004 | 0.304 | 0.008 | 0.315 | 0.004 | 0.382 | 0.005 | 0.358 | 0.004 | 0.324 | 0.007 | 0.327 | 0.004 |
| *Obs* | *8,161* |  | *17,255* |  | *3,020* |  | *11,702* |  | *8,580* |  | *16,703* |  | *4,247* |  | *12,003* |  |
| Parental separation | 0.207 | 0.004 | 0.333 | 0.004 | 0.220 | 0.007 | 0.397 | 0.005 | 0.222 | 0.004 | 0.334 | 0.004 | 0.223 | 0.006 | 0.378 | 0.004 |
| *Obs* | *8,700* |  | *17,255* |  | *3,219* |  | *11,702* |  | *9,153* |  | *16,703* |  | *4,527* |  | *12,003* |  |
| Different region | 0.258 |  | 0.242 | 0.005 | 0.157 | 0.007 | 0.162 | 0.009 | 0.266 | 0.005 | 0.247 | 0.006 | 0.196 | 0.006 | 0.197 | 0.008 |
| *Obs* | *8,191* |  | *17,255* |  | *3,037* |  | *11,702* |  | *8,565* |  | *16,703* |  | *4,244* |  | *12,003* |  |
| Both alive | 0.745 | 0.005 | 0.665 | 0.005 | 0.657 | 0.009 | 0.540 | 0.011 | 0.676 | 0.005 | 0.641 | 0.005 | 0.623 | 0.007 | 0.591 | 0.009 |
| Mother alive | 0.180 | 0.004 | 0.207 | 0.006 | 0.263 | 0.008 | 0.264 | 0.011 | 0.241 | 0.004 | 0.272 | 0.005 | 0.297 | 0.007 | 0.319 | 0.009 |
| Father alive | 0.075 | 0.003 | 0.128 | 0.005 | 0.080 | 0.005 | 0.195 | 0.012 | 0.083 | 0.003 | 0.087 | 0.003 | 0.081 | 0.004 | 0.090 | 0.006 |
| *Obs* | *8,167* |  | *17,255* |  | *2,696* |  |  |  | *9,213* |  | *16,703* |  | *4,534* |  | *12,003* |  |
|  |  Mean | se |  Mean | se |  Mean | se |  Mean | se |  Mean | se |  Mean | se |  Mean | se |  Mean | se |
| Maternal age at birth | 26.025 | 0.059 | 25.944 | 0.043 | 26.995 | 0.107 | 26.178 | 0.055 | 25.655 | 0.055 | 25.777 | 0.043 | 26.145 | 0.083 | 26.014 | 0.053 |
| *Obs* | *8,161* |  | *17,255* |  | *3,020* |  | *11,702* |  | *8,580* |  | *16,703* |  | *4,247* |  | *12,003* |  |
| Help/care hours |   |   |   |   | 3.470 | 0.100 | 3.619 | 0.093 |   |   |   |   | 3.437 | 0.090 | 3.606 | 0.103 |
| *Obs* |  |  |  |  | *3,285* |  | *11,702* |  |  |  |  |  | *4,528* |  | *12,003* |  |
| Malaise |   |   |   |   |   |   |   |   | 1.850 | 0.022 | 1.964 | 0.022 | 1.951 | 0.033 | 2.039 | 0.034 |
| *Obs* |  |  |  |  |  |  |  |  | *8,130* |  | *16,703* |  | *3,966* |  | *12,003* |  |
| Wellbeing (WEMWB) |   |   |   |   |   |   |   |   | 49.250 | 0.094 | 48.585 | 0.094 | 48.810 | 0.135 | 48.331 | 0.137 |
| *Obs* |  |  |  |  |  |  |  |  | *7,650* |  | *16,703* |  | *3,735* |  | *12,003* |  |

**Table A1.3 Summary statistics of analysis variables for imputed and raw data – 1958 Cohort**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cohort: 1958 | Age 50 - Caregiving | Age 50 - Intensity | Age 55 - Caregiving | Age 55 - Intensity |
|  | Observed |  | MI |  | Observed |  | MI |  | Observed |  | MI |  | Observed |  | MI |  |
|  | Prop | se | Prop | se | Prop | se | Prop | se | Prop | se | Prop | se | Prop | se | Prop | se |
| No help/care | 0.447 | 0.006 | 0.431 | 0.006 |   |   |   |   | 0.366 | 0.007 | 0.350 | 0.006 |   |   |   |   |
| Helper | 0.369 | 0.006 | 0.373 | 0.006 | 0.668 | 0.008 | 0.657 | 0.009 | 0.377 | 0.007 | 0.365 | 0.007 | 0.591 | 0.008 | 0.566 | 0.009 |
| Carer | 0.184 | 0.005 | 0.196 | 0.004 | 0.332 | 0.008 | 0.343 | 0.009 | 0.257 | 0.006 | 0.285 | 0.006 | 0.409 | 0.008 | 0.434 | 0.009 |
| *Obs* | *7,093* |  | *12,775* |  | *3,925* |  | *9,607* |  | *5,463* |  | *11,339* |  | *3,507* |  | *9,339* |  |
| Sibling | 0.928 | 0.003 | 0.926 | 0.002 | 0.916 | 0.005 | 0.920 | 0.003 | 0.934 | 0.004 | 0.929 | 0.003 | 0.924 | 0.005 | 0.923 | 0.003 |
| Only child | 0.072 | 0.003 | 0.074 | 0.002 | 0.084 | 0.005 | 0.080 | 0.003 | 0.066 | 0.004 | 0.071 | 0.003 | 0.076 | 0.005 | 0.077 | 0.003 |
| *Obs* | *5,888* |  | *12,775* |  | *3,260* |  | *9,607* |  | *4,591* |  | *11,339* |  | *2,947* |  | *9,339* |  |
| Male | 0.498 | 0.006 | 0.516 | 0.004 | 0.491 | 0.008 | 0.520 | 0.005 | 0.496 | 0.007 | 0.520 | 0.005 | 0.488 | 0.008 | 0.522 | 0.005 |
| Female | 0.502 | 0.006 | 0.484 | 0.004 | 0.509 | 0.008 | 0.480 | 0.005 | 0.504 | 0.007 | 0.480 | 0.005 | 0.512 | 0.008 | 0.478 | 0.005 |
| *Obs* | *7,147* |  | *12,775* |  | *3,979* |  | *9,607* |  | *5,567* |  | *11,339* |  | *3,567* |  | *9,339* |  |
| Mother stayed in school | 0.291 | 0.006 | 0.257 | 0.004 | 0.261 | 0.007 | 0.232 | 0.004 | 0.313 | 0.006 | 0.259 | 0.004 | 0.295 | 0.008 | 0.241 | 0.005 |
| *Obs* | *6,744* |  | *12,775* |  | *3,753* |  | *9,607* |  | *5,248* |  | *11,339* |  | *3,371* |  | *9,339* |  |
| Parental separation | 0.062 | 0.003 | 0.079 | 0.003 | 0.055 | 0.004 | 0.082 | 0.003 | 0.055 | 0.003 | 0.080 | 0.003 | 0.044 | 0.004 | 0.081 | 0.003 |
| *Obs* | *6,157* |  | *12,775* |  | *3,412* |  | *9,607* |  | *4,808* |  | *11,339* |  | *3,077* |  | *9,339* |  |
| Different region | 0.285 | 0.006 | 0.267 | 0.006 | 0.204 | 0.007 | 0.196 | 0.006 | 0.302 | 0.006 | 0.277 | 0.006 | 0.240 | 0.007 | 0.232 | 0.008 |
| *Obs* | *7,130* |  | *12,775* |  | *3,713* |  | *9,607* |  | *5,086* |  | *11,339* |  | *3,286* |  | *9,339* |  |
| Both alive | 0.403 | 0.006 | 0.378 | 0.006 | 0.345 | 0.008 | 0.329 | 0.007 | 0.318 | 0.006 | 0.277 | 0.005 | 0.288 | 0.008 | 0.266 | 0.009 |
| Mother alive | 0.453 | 0.006 | 0.471 | 0.006 | 0.537 | 0.008 | 0.546 | 0.008 | 0.524 | 0.007 | 0.546 | 0.006 | 0.580 | 0.008 | 0.594 | 0.010 |
| Father alive | 0.144 | 0.004 | 0.151 | 0.004 | 0.118 | 0.005 | 0.125 | 0.006 | 0.157 | 0.005 | 0.177 | 0.005 | 0.132 | 0.006 | 0.140 | 0.007 |
| *Obs* | *7,093* |  | *12,775* |  | *3,925* |  | *9,607* |  | *5,530* |  | *11,339* |  | *3,530* |  | *9,339* |  |
|  |  Mean | se |  Mean | se |  Mean | se |  Mean | se |  Mean | se |  Mean | se |  Mean | se |  Mean | se |
| Maternal age at birth | 26.172 | 0.059 | 26.612 | 0.048 | 26.841 | 0.081 | 27.032 | 0.057 | 25.659 | 0.062 | 26.323 | 0.050 | 26.114 | 0.078 | 26.649 | 0.057 |
| *Obs* | *6,762* |  | *12,775* |  | *3,766* |  | *9,607* |  | *5,263* |  | *11,339* |  | *3,382* |  | *9,339* |  |
| Mean help/care hours |   |   |   |   | 4.222 | 0.137 | 4.571 | 0.145 |   |   |   |   | 5.827 | 0.198 | 6.674 | 0.231 |
| *Obs* |  |  |  |  | *3,979* |  | *9,607* |  |  |  |  |  | *3,362* |  | *9,339* |  |
| Meal Malaise | 1.442 | 0.023 | 1.503 | 0.023 | 1.452 | 0.030 | 1.522 | 0.027 |   |   |   |   |   |   |   |   |
| *Obs* | *7,063* |  | *12,775* |  | *3,943* |  | *9,607* |  |  |  |  |  |  |  |  |  |
| Mean CASP-6 |   |   |   |   |   |   |   |   | 18.585 | 0.055 | 18.373 | 0.055 | 18.570 | 0.068 | 18.358 | 0.068 |
| *Obs* |  |  |  |  |  |  |  |  | *5,567* |  | *11,339* |  | *3,567* |  | *9,339* |  |

**Table A1.4 Summary statistics of analysis variables for imputed and raw data – 1946 Cohort**

| Cohort: 1946 | Age 63 - Caregiving |  | Age 63 - Intensity |
| --- | --- | --- | --- |
|  | Observed |  | MI |  |  | Observed |  | MI |  |
|  | Prop | se | Prop | se |  | Prop | se | Prop | se |
| No (parent) care | 0.650 | 0.021 | 0.679 | 0.024 |  |   |   |   |   |
| Parent care | 0.350 | 0.021 | 0.321 | 0.024 | Parent care | 0.668 | 0.029 | 0.591 | 0.039 |
|  |   |   |   |   | Other care | 0.332 | 0.029 | 0.409 | 0.039 |
| *Obs* | *511* |  | *2,364* |  | *Obs* | *268* |  | *2,121* |  |
|  |   |   |   |   | Hours of care: 0-4 | 0.384 | 0.031 | 0.380 | 0.034 |
|  |   |   |   |   | 5-9 | 0.318 | 0.030 | 0.301 | 0.027 |
|  |   |   |   |   | 10-19 | 0.149 | 0.023 | 0.148 | 0.024 |
|  |   |   |   |   | 20+ | 0.149 | 0.023 | 0.170 | 0.031 |
|  |  |  |  |  | *Obs* | *242* |  | *2,121* |  |
| Sibling | 0.849 | 0.013 | 0.851 | 0.008 | Sibling | 0.829 | 0.017 | 0.858 | 0.008 |
| Only child | 0.151 | 0.013 | 0.149 | 0.008 | Only child | 0.171 | 0.017 | 0.142 | 0.008 |
| *Obs* | *724* |  | *2,364* |  | *Obs* | *497* |  | *2,121* |  |
| Male | 0.504 | 0.018 | 0.552 | 0.010 | Male | 0.474 | 0.022 | 0.550 | 0.011 |
| Female | 0.496 | 0.018 | 0.448 | 0.010 | Female | 0.526 | 0.022 | 0.450 | 0.011 |
| *Obs* | *772* |  | *2,364* |  | *Obs* | *529* |  | *2,121* |  |
| Mother's age <=24 | 0.416 | 0.019 | 0.334 | 0.010 | Mother's age <=24 | 0.372 | 0.022 | 0.314 | 0.011 |
| 25-29 | 0.369 | 0.018 | 0.331 | 0.010 | 25-29 | 0.396 | 0.023 | 0.333 | 0.011 |
| 30-34 | 0.176 | 0.015 | 0.222 | 0.009 | 30-34 | 0.187 | 0.018 | 0.232 | 0.010 |
| 35+ | 0.040 | 0.007 | 0.113 | 0.007 | 35+ | 0.045 | 0.010 | 0.120 | 0.007 |
| *Obs* | *683* |  | *2,364* |  | *Obs* | *470* |  | *2,121* |  |
| Social class I | 0.062 | 0.009 | 0.052 | 0.005 | Social class I | 0.062 | 0.011 | 0.050 | 0.005 |
| II | 0.166 | 0.013 | 0.147 | 0.007 | II | 0.159 | 0.016 | 0.143 | 0.008 |
| III NM | 0.198 | 0.014 | 0.158 | 0.008 | III NM | 0.212 | 0.018 | 0.157 | 0.008 |
| III M | 0.251 | 0.016 | 0.291 | 0.009 | III M | 0.253 | 0.019 | 0.296 | 0.010 |
| IV | 0.179 | 0.014 | 0.197 | 0.008 | IV | 0.174 | 0.016 | 0.198 | 0.009 |
| V | 0.053 | 0.008 | 0.058 | 0.005 | V | 0.057 | 0.010 | 0.059 | 0.005 |
| n/a | 0.091 | 0.010 | 0.097 | 0.006 | n/a | 0.083 | 0.012 | 0.096 | 0.006 |
| *Obs* | *772* |  | *2,364* |  | *Obs* | *529* |  | *2,121* |  |
| Mother stayed in school | 0.403 | 0.018 | 0.369 | 0.010 | Mother stayed in school | 0.389 | 0.021 | 0.362 | 0.010 |
| *Obs* | *772* |  | *2,364* |  | *Obs* | *529* |  | *2,121* |  |
| Parental separation | 0.086 | 0.010 | 0.100 | 0.006 | Parental separation | 0.068 | 0.011 | 0.097 | 0.006 |
| *Obs* | *771* |  | *2,361* |  | *Obs* | *528* |  | *2,121* |  |
| No qualification | 0.267 | 0.016 | 0.387 | 0.010 | No qualification | 0.281 | 0.020 | 0.404 | 0.011 |
| Up to Olevel | 0.308 | 0.017 | 0.271 | 0.009 | Up to Olevel | 0.300 | 0.020 | 0.265 | 0.010 |
| Up to Alevel | 0.310 | 0.017 | 0.254 | 0.009 | Up to Alevel | 0.298 | 0.020 | 0.245 | 0.009 |
| Degree or above | 0.115 | 0.012 | 0.088 | 0.006 | Degree or above | 0.121 | 0.014 | 0.086 | 0.006 |
| *Obs* | *746* |  | *2,364* |  | *Obs* | *513* |  | *2,121* |  |
| Married | 0.754 | 0.016 | 0.747 | 0.020 | Married | 0.739 | 0.020 | 0.716 | 0.024 |
| Cohabiting | 0.050 | 0.008 | 0.065 | 0.015 | Cohabiting | 0.050 | 0.010 | 0.071 | 0.014 |
| Not living with a partner | 0.196 | 0.015 | 0.188 | 0.015 | Not living with a partner | 0.211 | 0.018 | 0.213 | 0.020 |
| *Obs* | *720* |  | *2,364* |  | *Obs* | *502* |  | *2,121* |  |
| Different region | 0.423 | 0.019 | 0.410 | 0.021 | Different region | 0.374 | 0.023 | 0.383 | 0.020 |
| *Obs* | *676* |  | *2,364* |  | *Obs* | *444* |  | *2,121* |  |
| Mother alive | 0.728 | 0.021 | 0.740 | 0.026 |   |   |   |   |   |
| Father alive | 0.141 | 0.016 | 0.159 | 0.025 |   |   |   |   |   |
| Both alive | 0.130 | 0.016 | 0.101 | 0.017 |   |   |   |   |   |
| *Obs* | *453* |  | *2,364* |  |  |  |  |  |  |
|  |  Mean | se |  Mean | se |  |  Mean | se |  Mean | se |
| Mean GHQ | 2.107 | 0.151 | 2.146 | 0.126 | Mean GHQ | 2.409 | 0.226 | 2.479 | 0.182 |
| *Obs* | *514* |  | *2,364* |  | *Obs* | *276* |  | *2,121* |  |

**Table A1.5 Regression Summary - Coefficients for Only Child (ref: sibling) – Complete cases and MI**

|  |  |
| --- | --- |
|  | Multinomial logistic regression - Base category: No help |
|  | Complete cases: Fully adjusted |  | Multiple imputation: Fully Adjusted |  |
|  | Helper | Carer | Helper | Carer |
|  | RRR | SE | p | 95% CI | RRR | SE | p | 95% CI | RRR | SE | p | 95% CI | RRR | SE | p | 95% CI |
| 1970 Cohort  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Age 38 | 0.93 | 0.12 | 0.58 | 0.72 - 1.20 | 1.22 | 0.20 | 0.23 | 0.89 - 1.67 | 0.92 | 0.12 | 0.49 | 0.71 - 1.18 | 1.27 | 0.16 | 0.06 | 0.99 - 1.64 |
| Age 42 | 1.28 | 0.14 | 0.02 | 1.04 - 1.59 | 1.25 | 0.17 | 0.10 | 0.96 - 1.63 | 1.21 | 0.12 | 0.05 | 1.00 - 1.46 | 1.30 | 0.15 | 0.02 | 1.03 - 1.63 |
| 1958 Cohort  |  |  |  |  |  |  |  |  |   |  |  |  |  |  |  |  |
| Age 50 | 1.44 | 0.21 | 0.01 | 1.08 - 1.92 | 1.46 | 0.26 | 0.03 | 1.03 - 2.06 | 1.57 | 0.16 | 0.00 | 1.28 - 1.93 | 1.45 | 0.21 | 0.01 | 1.09 - 1.93 |
| Age 55 | 1.54 | 0.29 | 0.02 | 1.07 - 2.22 | 1.47 | 0.31 | 0.07 | 0.98 - 2.21 | 1.63 | 0.24 | 0.00 | 1.22 - 2.16 | 1.78 | 0.26 | 0.00 | 1.34 - 2.36 |
|  | Binary logistic Regression - Base category: No parent care  |  |
|  | Complete cases: Fully adjusted |  | Multiple imputation: Fully Adjusted |  |
|  |  | Carer |  | Carer |
|  |   |   |   |   | OR | SE | p | 95% CI |   |   |   |   | OR | SE | p | 95% CI |
| 1946 Cohort  |   |   |   |   |  |  |  |  |   |   |   |   |  |  |  |  |
| Age 63 |   |   |   |   | 1.90 | 0.64 | 0.06 | 0.98 - 3.69 |   |   |   |   | 2.11 | 0.67 | 0.02 | 1.13 - 3.93 |
| Sample n |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Age 38 | 6,332 |  |  |  |  |  |  |  | 17,255 |  |  |  |  |  |  |  |
| Age 42 | 6,986 |  |  |  |  |  |  |  | 16,703 |  |  |  |  |  |  |  |
| Age 50 | 4,194 |  |  |  |  |  |  |  | 12,775 |  |  |  |  |  |  |  |
| Age 55 | 3,230 |  |  |  |  |  |  |  | 11,339 |  |  |  |  |  |  |  |
| Age 63 | 364 |   |   |   |   |   |   |   | 2,364 |  |  |  |  |  |  |  |
| *Notes: Fully adjusted models all control for respondent gender, maternal age at respondent’s birth, maternal education, and paternal occupational class, parental separation in childhood, cohort member’s level of qualification, occupation (1970 and 1958 cohorts only), and cross-regional moves between childhood and adulthood and which parent is alive.*  |

**Table A1.6 Regression Summary - Coefficients for Only Child (ref: sibling) – Complete cases and MI**

|  |  |
| --- | --- |
|  | Linear regression: Hours per week spent helping/caring – Among Helpers and carers |
|  | Complete cases | Multiple imputation |
|  | Coeff. | SE | p | 95% CI | Coeff. | SE | p | 95% CI |
| 1970 Cohort  |  |  |  |  |  |  |  |  |
| Age 38 | 0.48 | 0.36 | 0.18 | -0.23 - 1.19 | 0.08 | 0.34 | 0.81 | -0.59 - 0.76 |
| Age 42 | -0.20 | 0.31 | 0.52 | -0.81 - 0.41 | 0.49 | 0.38 | 0.20 | -0.26 - 1.24 |
| 1958 Cohort  |  |  |  |  |   |  |  |  |
| Age 50 | -0.50 | 0.59 | 0.40 | -1.65 - 0.66 | 0.33 | 0.50 | 0.51 | -0.64 - 1.30 |
| Age 55 | 1.39 | 0.92 | 0.13 | -0.42 - 3.19 | 1.46 | 0.86 | 0.09 | -0.24 - 3.15 |
|  | Ordinal logistic regression: Banded hours per week spent helping/caring – Among carers |
|  | Complete cases | Multiple imputation |
|  | OR | SE | p | 95% CI | OR | SE | p | 95% CI |
| 1946 Cohort  |  |  |  |  |  |  |  |  |
| Age 63 | 0.63 | 0.24 | 0.23 | 0.29 - 1.35 | 0.66 | 0.21 | 0.18 | 0.36 - 1.21 |
| Sample n |  |  |  |  |   |  |  |  |
| Age 38 | 2,060 |  |  |  | 11,702 |  |  |  |
| Age 42 | 3,359 |  |  |  | 12,003 |  |  |  |
| Age 50 | 2,347 |  |  |  | 9,607 |  |  |  |
| Age 55 | 2,017 |  |  |  | 9,339 |  |  |  |
| Age 63 | 186 |   |   |   | 2,121 |  |  |  |
| *Notes: Fully adjusted models control for respondent gender, maternal age at respondent’s birth, maternal education, and paternal occupational class, parental separation in childhood, cohort member’s level of qualification, occupation (1970 and 1958 cohorts only), cross-regional moves since childhood and which parent is alive (1970 and 1958 cohorts only; the smaller sample size and older age of the 1946 cohort meant that after the sample restrictions the imputation model did not converge when the indicator of which parent was alive was included).*  |

**Table A1.7 Results of** **posterior predictive checking for logistic regression coefficients (Caregiving model)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | MI analysis dataset | Replications |  |
|   | Coefficient | Mean of Coefficients | PPP |
| Only Child (Ref=Sibling)  | 0.747 | 0.759 | 0.514 |
| Female (Ref=Male) | 0.977 | 0.970 | 0.478 |
| Mother's age at birth = 25-29 (Ref: <25) | 0.449 | 0.443 | 0.486 |
| 30-34 | 0.184 | 0.136 | 0.443 |
| 35+ | -0.943 | -0.759 | 0.623 |
| Parental social class: II (Ref: I) | -0.139 | -0.155 | 0.463 |
| III non-manual | -0.018 | -0.036 | 0.464 |
| IV manual | -0.120 | -0.146 | 0.453 |
| V | -0.225 | -0.224 | 0.508 |
| VI | -0.232 | -0.261 | 0.468 |
| Not available/ Unemployed. | -0.184 | -0.142 | 0.547 |
| Mother stayed in school (Ref: Left) | -0.125 | -0.143 | 0.464 |
| Parental separation = Yes (Ref: No) | -0.563 | -0.642 | 0.398 |
| Highest Qual=Up to O-level (Ref: None) | 0.767 | 0.878 | 0.659 |
| Up to A-level | 0.379 | 0.469 | 0.621 |
| Degree or above | 1.197 | 1.274 | 0.577 |
| Partnership: Cohabiting (Ref=Married) | -0.052 | 0.034 | 0.581 |
| Not living with a partner | 0.580 | 0.574 | 0.495 |
| Region different than age 11 (Ref=Same) | -0.618 | -0.637 | 0.465 |
| Which parent alive = Father (Ref: Mother) | -0.989 | -1.054 | 0.413 |
| Both | -0.834 | -0.843 | 0.495 |

*Notes: Based on 2,000 replications. A posterior predictive p-value (PPP) close to 0 or 1 indicates systematic differences between the analysis dataset and replications, potentially suggesting problems with the model*

**Table A1.8 Results of posterior predictive checking for ordinal logistic regression coefficients (Intensity model)**

|  |  |  |  |
| --- | --- | --- | --- |
| Intensity | MI analysis dataset | Replications |  |
|   | Coefficient | Mean of Coefficients | PPP |
| Sibling status = 1, Only | -0.419 | -0.380 | 0.541 |
| Sex = 1, Female | 0.914 | 0.934 | 0.542 |
| Mother's age at birth = 2, 25-29 | 0.327 | 0.317 | 0.467 |
| Mother's age at birth = 3, 30-34 | 1.208 | 1.284 | 0.574 |
| Mother's age at birth = 4, 35+ | 0.690 | 0.612 | 0.467 |
| 1946 father social class in 1950 = 2, II | 0.214 | 0.178 | 0.432 |
| 1946 father social class in 1950 = 3, III non-manual | 0.333 | 0.326 | 0.485 |
| 1946 father social class in 1950 = 4, manual | 0.462 | 0.468 | 0.509 |
| 1946 father social class in 1950 = 5, IV | 0.589 | 0.611 | 0.534 |
| 1946 father social class in 1950 = 6, V | 0.770 | 0.779 | 0.513 |
| 1946 father social class in 1950 = 7, Not applicable/unemployed | 0.900 | 0.923 | 0.530 |
| CM mother school beyond min age = 1, Yes | 0.093 | 0.084 | 0.498 |
| parental\_separation = 1 | 0.527 | 0.545 | 0.520 |
| Highest Qual 26-43 = 1, Up to Olevel | -0.022 | -0.059 | 0.453 |
| Highest Qual 26-43 = 2, Up to Alevel | -0.384 | -0.488 | 0.387 |
| Highest Qual 26-43 = 3, Degree or above | 0.077 | 0.039 | 0.467 |
| Partnership status 63 = 2, Cohabiting | 0.050 | 0.123 | 0.555 |
| Partnership status 63 = 3, Not living with a partner | 0.258 | 0.289 | 0.530 |
| Region age 69 different than age 11 = 1, Yes | 0.126 | 0.119 | 0.482 |

*Notes: Based on 2,000 replications. A posterior predictive p-value (PPP) close to 0 or 1 indicates systematic differences between the analysis dataset and replications, potentially suggesting problems with the model*

**Appendix References**

Carpenter, J. and Kenward, M. 2012. *Multiple Imputation and its Application*. Hoboken: Wiley, Hoboken.

Graham, J.W. 2009. Missing Data Analysis: Making It Work in the Real World. *Annual review of psychology*, **60**, 1, 549-576.

Jeličić, H., Phelps, E. and Lerner, R.M. 2009. Use of Missing Data Methods in Longitudinal Studies: The Persistence of Bad Practices in Developmental Psychology. *Developmental psychology*, **45**, 4, 1195-1199.

Madley-Dowd, P., Hughes, R., Tilling, K. and Heron, J. 2019. The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology*, **110**, 63-73.

Mostafa, T., Narayanan, M., Pongiglione, B., Dodgeon, B., Goodman, A., Silverwood, R.J. and Ploubidis, G.B. 2021. Missing at random assumption made more plausible: evidence from the 1958 British birth cohort. *Journal of clinical epidemiology*, **136**, 44-54.

Nguyen, C.D., Carlin, J.B. and Lee, K.J. 2017. Model checking in multiple imputation: An overview and case study. *Emerging themes in epidemiology*, **14**, 1, 8-8.

Silverwood, R., Narayanan, M., Dodgeon, B. and Ploubidis, G. 2020. Handling missing data in the National Child Development Study: User guide. In UCL Centre for Longitudinal Studies, London.