**Supplementary file**

**Multiple imputation methods**

As each study collected data on a different set of confounders, combining the two studies together resulted in a high proportion of missing data. Within the literature methodology has been developed to combine data from multiple data sources to adjust for a wider of variables[1-4](#_ENREF_1). Heymans et al used the method of Multivariate Imputation by Chained Equations (MICE) to combine data from three controlled trials to identify predictors of chronic low back pain[3](#_ENREF_3). As missing values are sampled from the posterior predictive distribution based on the observed data, MICE is based on a Bayesian approach[5](#_ENREF_5). Jackson et al describe the use of multiple imputation methods within a Bayesian graphical modelling framework, to combine data sets with different sets of covariates, in order to describe the association of low birth weight on environmental exposure to air pollution, and incorporate adjustment for important confounders not available within each dataset[1](#_ENREF_1). The performance of these two different approaches (Fully Bayesian graphical Model (FBM) versus MICE) has been compared using both simulation[6](#_ENREF_6) and cases studies[7](#_ENREF_7). Even with extreme levels of missing data both the FBM and MICE approaches have similar performance and correct most of the bias with a non-hierarchical (simple) data structure. Our preference was to use MICE to combine data sources, for reasons of transparency as the methodology is easier to understand and more familiar, ease of implementation in standard statistical software, and the computational time taken to run the models being substantially faster.

Multiple imputation allows for the uncertainty about missing data by creating several different plausible imputed datasets and combining their results[5](#_ENREF_5). Use of this method makes the assumption that the data are missing at random (MAR). This is plausible as the reason behind the missing data is that not all covariates were collected in each study[3](#_ENREF_3). In the first stage, a number of copies of the dataset are created (we created 50 copies), each of which has the missing values replaced by imputed values[5](#_ENREF_5), [8](#_ENREF_8), [9](#_ENREF_9) with 50 cycles of regression switching. Imputations are made by drawing from the posterior predictive distribution of each variable requiring imputation. We included all of the covariates (as listed earlier) together with the outcome variable (12-month post-operative OHS) in the imputation model as this carries information about the missing values of predictors[5](#_ENREF_5). Continuous variables were transformed to be approximately normally distributed before imputation where necessary. Linear regression was used to impute continuous variables and logistic, ordinal and multinomial regression for categorical variables. The distributions of complete case and imputed variables are described in the table below, confirming consistency between the original and imputed variables. In the second stage, a statistical model is fitted to each of the imputed datasets separately, and the results are averaged together to get a single estimate of the association. Standard errors are calculated using Rubin’s rules, which account for the variability between results of imputed datasets, reflecting the uncertainty associated with imputing missing data[5](#_ENREF_5).

The methodology we used to combine data from the cohorts has been used before in other settings and has the advantage of transparency in accordance with reporting guidelines[5](#_ENREF_5) where the extent of missing data is presented, the method familiar within the methodological literature, and its relative ease of implementation within standard statistical software. However, use of these methods require us to make the assumption data are MAR, which is plausible in the context of this study as the reason for missing data is due to variables not being collected in the study. Plausibility is further enhanced by inclusion of a wide range of covariates to ensure enough variables predictive of missing information are included. Concerns may be raised over the extent of missing data in this study where even if the MAR assumption is valid, as the proportion of missing observations on a variable increases the reliability of the regression coefficient for that covariate may diminish[8](#_ENREF_8). Reassurance is provided from other studies in the literature using these methods where using both simulation and case studies these methods have been shown to be effective even with extreme levels of missing data (over 90% missing)[6](#_ENREF_6), [7](#_ENREF_7).

**REFERENCES**

1 Jackson CH, Best NG, Richardson S. Bayesian graphical models for regression on multiple data sets with different variables. *Biostatistics*. 2009; 10: 335-51.

2 Molitor NT, Best N, Jackson C, Richardson S. Using Bayesian graphical models to model biases in observational studies and to combine multiple data sources: Application to low birth-weight and water disinfection by-products. 2008.

3 Heymans M, van Buuren S, Knol D, van Mechelen W, de Vet H. Variable selection under multiple imputation using the bootstrap in a prognostic study. *BMC Medical Research Methodology*. 2007; 7: 33.

4 Vergouw D, Heymans M, Peat G, Kuijpers T, Croft P, de Vet H*, et al.* The search for stable prognostic models in multiple imputed data sets. *BMC Medical Research Methodology*. 2010; 10: 81.

5 Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG*, et al.* Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009; 338: b2393.

6 Mason A, Richardson S, Best N. A comparison of fully Bayesian and two-stage imputation strategies for missing covariate data. *5th Annual Bayesian Biostatistics Conference*: University of Texas January 23-25, 2012.

7 Best N, Mason A, Richardson S, McCandless L. Bayesian approaches for combining multiple data sources to adjust for missing confounders. *4th International Joint Meeting of the Institute of Mathematical Statistics and the International Society for Bayesian Analysis*: Utah January 5-7, 2011.

8 Royston P. Multiple imputation of missing values. *The Stata Journal*. 2004; 4.

9 Marshall A, Altman D, Holder R, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Medical Research Methodology*. 2009; 9: 57.

**Distribution of variables in the EUROHIP and EPOS datasets before (complete case) and after (imputed) imputation**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Missing** | **Complete Case** | **Imputed** |
| Body Mass Index (BMI) | 141 (6.7%) | 27.6 (4.7) | 27.6 (4.7) |
| Age at operation | 17 (0.8%) | 67.6 (10.3) | 67.6 (10.3) |
| Sex | 74 (3.5%) |   |   |
|  Male |   | 821 (40.2%) | 843 (39.8%) |
|  Female |   | 1223 (59.8%) | 1,275 (60.2%) |
| SF36 mental health score pre-op  | 1268 (59.9%) | 70.9 (19.8) | 69.7 (20.3) |
| Fixed flexion | 951 (44.9%) |   |   |
|  0 |   | 611 (52.4%) | 983 (46.4%) |
|  1-9 |   | 115 (9.9%) | 187 (8.8%) |
|  10-19 |   | 249 (21.3%) | 444 (21.0%) |
|  20-29 |   | 118 (10.1%) | 262 (12.4%) |
|  30-59 |   | 64 (5.5%) | 196 (9.3%) |
|  60+ |   | 10 (0.9%) | 46 (2.2%) |
| No. of Comorbidities | 1238 (58.5%) |   |   |
|  0 |   | 185 (21.0%) | 540 (25.5%) |
|  1 |   | 462 (52.5%) | 1,082 (51.1%) |
|  2 |   | 178 (20.2%) | 382 (18.0%) |
|  3+ |   | 55 (6.3%) | 114 (5.4%) |
| Analgesic/NSAID medication use | 166 (7.8%) |   |   |
|  No |   | 424 (21.7%) | 478 (22.6%) |
|  Yes |   | 1528 (78.3%) | 1,640 (77.4%) |
| Education | 1368 (64.6%) |   |   |
|  None |   | 400 (53.3%) | 1,181 (55.7%) |
|  College / University |   | 350 (46.7%) | 937 (44.3%) |
| ASA grade | 1375 (64.9%) |   |   |
|  1 |   | 117 (15.8%) | 339 (16.0%) |
|  2 |   | 472 (63.5%) | 1,355 (64.0%) |
|  3+ |   | 154 (20.7%) | 425 (20.1%) |
| Years of Hip Pain | 1273 (60.1%) |   |   |
|  <1 |   | 84 (9.9%) | 217 (10.2%) |
|  1-2 |   | 244 (28.9%) | 618 (29.2%) |
|  3-5 |   | 258 (30.5%) | 636 (30.0%) |
|  6+ |   | 259 (30.7%) | 647 (30.6%) |
| Pre-op expectation of less pain | 1267 (59.8%) |   |   |
|  No |   | 691 (81.2%) | 1,682 (79.4%) |
|  Yes |   | 160 (18.8%) | 436 (20.6%) |
| K&L grade pre-operative | 1332 (62.9%) |   |   |
|  0,1,2 |   | 112 (14.3%) | 292 (13.8%) |
|  3 |   | 262 (33.3%) | 686 (32.4%) |
|  4 |   | 412 (52.4%) | 1,140 (53.8%) |
| No. of joints with OA | 1271 (60.0%) |   |   |
|  0 |   | 278 (32.8%) | 680 (32.1%) |
|  1 |   | 240 (28.3%) | 595 (28.1%) |
|  2 |   | 143 (16.9%) | 366 (17.3%) |
|  3 |   | 75 (8.9%) | 192 (9.1%) |
|  4+ |   | 111 (13.1%) | 285 (13.5%) |
| Surgical Approach | 1206 (56.9%) |  |   |
|  Anterolateral |   | 591 (64.8%) | 1,407 (66.4%) |
|  Posterior |   | 321 (35.2%) | 711 (33.6%) |
| Stem size (mm offset) | 853 (40.3%) |  |   |
|  31 |   | 59 (4.7%) | 188 (8.9%) |
|  38 |   | 535 (42.3%) | 902 (42.6%) |
|  44 |   | 650 (51.4%) | 991 (46.8%) |
|  50 |   | 21 (1.7%) | 37 (1.7%) |

Numbers represent mean (SD) or median (inter-quartile range) for continuous variables, and number (percentage) for categorical.