SUPPLEMENTARY MATERIAL

MULTIPLE IMPUTATION METHODS

The results of complete case analyses can be biased[1]. The cumulative effect of missing data in several variables often leads to exclusion of a substantial proportion of the original sample, causing a loss of precision and power. Multiple imputation methods can be used to handle datasets with missing values. The risk of bias depends on the reasons why the data are missing. Missing data are seldom completely random. They are usually related, directly or indirectly, to other subject or disease characteristics, including the outcome under study[2]. If it is plausible the data are missing at random, but not completely at random, analyses on complete cases may be biased[3]. This bias can be overcome by using multiple imputation, which allows for the uncertainty about missing data by creating several plausible imputed datasets and appropriately combining their results. We have done this using the ICE procedure in Stata[4-6]. The first stage is to create multiple copies of the dataset with missing values replaced by imputed ones (we have created 10 copies). Missing values are sampled from their predictive distribution based on the observed data. The imputation procedure accounts for uncertainty in predicting missing values by injecting appropriate variability into the multiple imputed values. In the second stage regression models are fitted to each of the imputed datasets and averaged together to give overall estimated associations. Standard errors are calculated using Rubins Rules. We have included all predictor variables in the multiple imputation process, together with the outcome variable as this carries information about missing values of the predictors.

INTERNAL VALIDATION

Model building in prognostic studies is usually performed using automatic stepwise variable selection procedures. However, stepwise methods have a number of disadvantages[3,7,8], where their power to select true variables is limited and estimates of predictive validity and fit may be overly optimistic. It has been suggested that of the final significant variables included in a prognostic model using backwards selection only half may be true risk factors that would be replicated by other studies[8]. To overcome these limitations, it has been suggested that bootstrapping combined with automatic backward regression can be used to provide information on model stability[7,8].

For internal validation of the regression models we therefore use a combination of multiple imputation and bootstrapping. Firstly, missing data is imputed using the ICE procedure in Stata and 10 imputed datasets created. Missing values are sampled from their predictive distribution based on the observed data. The imputation procedure accounts for uncertainty in predicting missing values by injecting appropriate variability into the multiple imputed values. Using the 'micombine' procedure in Stata the full regression models including all predictor variables are fitted to each of the imputed datasets and averaged together to give overall estimated associations with standard errors calculated using Rubins Rules[6]. Second, 200 bootstrap samples are then randomly drawn with replacement (e.g. when a patient is randomly selected their data is taken from each of the 10 imputed datasets). An automatic backward selection procedure is then applied to each of the 200 bootstrap samples of 10 imputed datasets using a Wald test with a stopping rule of α = 0.157. This conservation p-value is comparable with the more complex Akaike Information Criterion (AIC)[7]. Variables retained in the final regression model are those consistently selected across the re-samples at least 70% of the time.

REFERENCES

1. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338(jun29_1):b2393.

2. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* 2009;338:b604.

3. Harrell FE, Jr. *Regression modeling strategies with applications to linear models, logistic regression, and survival analysis*: New York: Springer; 2001.

4. Royston P. Multiple imputation of missing values. *The Stata Journal.* 2004;4(3).

5. Royston P. Multiple imputation of missing values: update. *The Stata Journal.* 2005;5(2).

6. Royston P. Multiple imputation of missing values: Update of ice. *The Stata Journal.* 2005;5(4).

7. Vergouw D, Heymans M, Peat G, et al. The search for stable prognostic models in multiple imputed data sets. *BMC Medical Research Methodology.* 2010;10(1):81.

8. 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(1):33.

Repeated measures Analysis of covariance (ANCOVA) models to identify predictors of the average OHS between 1 and 5-years follow up

Variable	Univariable Δ Coef (95% CI)	P-value	Percentage retained in model	Multivariable (70% cut-off) Δ Coef (95% CI)	P-value
Patient variables					
Baseline Total Oxford Hip Score (10 units)	3.68 (3.16, 4.20)	<0.001	100.0%	2.68 (2.16, 3.21)	< 0.001
Year	0.02 (-0.10, 0.13)	0.77	100.0%	0.01 (-0.11, 0.13)	0.88
Age			-		
<50	-1.44 (-3.92, 1.03)	0.25	-		
50-60	-0.96 (-2.27, 0.35)	0.15	59.5%	-1.87 (-3.22, -0.53)	0.006
60-70	0.00 (0.00, 0.00)	-	92.5%	0.00 (0.00, 0.00)	-
70-80	-0.37 (-1.20, 0.46)	0.38	98.0%	-1.49 (-2.37, -0.61)	0.001
80+	-2.29 (-3.69, -0.88)	0.001	100.0%	-3.81 (-5.29, -2.33)	< 0.001
BMI (10 units)	-1.14 (-2.05, -0.22)	0.02	98.0%	-1.54 (-2.45, -0.64)	0.001
Gender			-		
Male	0.00 (0.00, 0.00)	-	-		
Female	-1.30 (-2.13, -0.48)	0.002	38.0%		
Hip For Surgery			-		
Left	0.00 (0.00, 0.00)	-	-		
Right	-0.16 (-0.99, 0.67)	0.71	15.0%		
Hip Indicated			-		
Unilateral	0.00 (0.00, 0.00)	-	-		
Bilateral	-0.48 (-3.01, 2.06)	0.71	16.5%		
Occupation			-		
Heavy manual	0.00 (0.00, 0.00)	-	-		
Light manual	-0.85 (-2.50, 0.80)	0.31	38.5%		
Office / professional	2.26 (0.96, 3.57)	0.001	58.0%		
Housewife	-0.92 (-2.18, 0.33)	0.15	54.5%		
Unemployed / Retired	-0.15 (-1.04, 0.73)	0.74	43.5%		
No. of Co-existing Diseases	-1.06 (-1.43, -0.69)	<0.001	98.0%	-0.90 (-1.27, -0.54)	< 0.001
Concomitant therapy used			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	-0.73 (-2.02, 0.56)	0.27	22.0%		
Centre number			-		
1	0.00 (0.00, 0.00)	-	-		
2	-0.93 (-2.23, 0.38)	0.16	38.0%		
3	0.31 (-0.69, 1.32)	0.54	34.5%		

4	1.48 (0.28, 2.68)	0.02	44.5%		
5	-1.76 (-3.27, -0.25)	0.02	69.0%		
6	-1.18 (-2.44, 0.07)	0.06	62.0%		
7	2.22 (1.33, 3.12)	<0.001	45.0%		
Fixed flexion	0.03 (-0.01, 0.06)	0.12	47.0%		
SF36 Mental Health Score (10 units)	0.80 (0.50, 1.11)	<0.001	100.0%	0.76 (0.46, 1.07)	< 0.001
Surgical variables					
Grade of Operator			-		
Consultant, locum consultant, assoc. specialist/staff	0.00 (0.00, 0.00)	-	-		
Fellow, senior registrar, registrar, locum registrar	-1.22 (-2.13, -0.32)	0.008	19.5%		
Surgical Approach			-		
Anterolateral	0.00 (0.00, 0.00)	-	-		
Posterior	1.03 (0.07, 1.98)	0.04	17.5%		
Patient's position			-		
Supine	0.00 (0.00, 0.00)	-	-		
Lateral	-0.99 (-2.00, 0.03)	0.06	25.5%		
Lavage System (Acetabular)			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	-0.70 (-2.60, 1.21)	0.47	64.5%		
Cement Pressurisation (Acetabular)			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	-0.37 (-1.43, 0.68)	0.48	34.5%		
Type of cement (Socket)			-		
No Cement	0.00 (0.00, 0.00)	-	-		
simplex	0.09 (-0.73, 0.91)	0.84	43.0%		
cmw1	0.63 (-0.33, 1.58)	0.2	28.5%		
palacos r	-1.21 (-2.34, -0.09)	0.03	33.0%		
Cement pressurisations (Femur)			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	0.58 (-3.85, 5.01)	0.8	20.0%		
Type of cement (Femur)			-		
simplex	0.00 (0.00, 0.00)	-	-		
cmw1	1.49 (0.20, 2.79)	0.02	48.0%		

cmw3	0.23 (-0.88, 1.34)	0.68	27.5%		
palacos r	-1.26 (-2.49, -0.03)	0.04	22.0%		
palacos lv	-1.44 (-4.89, 2.01)	0.41	37.0%		
Stem size (mm offset)	0.15 (0.04, 0.26)	0.01	84.0%	0.17 (0.06, 0.28)	0.002
Femoral Head			-		
Stainless Steel	0.00 (0.00, 0.00)	-	-		
Ceramic - Zirconia/ Alumina	0.76 (-0.66, 2.17)	0.29	21.0%		
Head size			-		
22	0.00 (0.00, 0.00)	-	-		
26	1.11 (0.28, 1.93)	0.009	32.0%		
28	-0.15 (-0.97, 0.68)	0.73	29.0%		
Type of Polythene			-		
uhmwpe	0.00 (0.00, 0.00)	-	-		
duration	1.64 (0.83, 2.46)	<0.001	52.0%		
Hip Dislocation			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	-4.27 (-8.05, -0.50)	0.03	74.5%	-3.77 (-7.47, -0.07)	0.05
Acetabular cup inclination (10 degrees)	-0.11 (-0.81, 0.58)	0.75	6.0%		
Acetabular cup version (10 degrees)	0.82 (0.17, 1.47)	0.01	32.5%		
Duration of Operation (Log)	-1.90 (-3.10, -0.70)	0.002	57.5%		
R2		-			17.4%
Optimism		-			0.8%
Bias-Corrected R2		-			16.6%

Δ: Represents the average follow up OHS between 1, 2, 3, 4 and 5-years follow up.

Percentage: the proportion of times the variable was retained in the backward selection regression models using a P-value of 0.157 (inclusion frequency)

70% cut-off: Variables included in the final regression model are those that are retained in at least 70% of the backward selection regression models

Univariable – Each predictor in the model is adjusted for Baseline OKS only