

## Appendix

### *Principles and Process of Propensity Score Matching*

Propensity scores were generated using a logistic model (with the outcome being the allocation to either TKR or UKR) and represent the probability that each patient has of receiving UKR on the basis of his/her patient and surgical factors. By matching patients on the basis of the propensity score, two balanced cohorts are created for comparison (although individual patients were not matched with one another as they are in a case-control study)<sup>1</sup>. This form of matching allows effects to be estimated with use of univariable models, covariates having been adjusted for in the matching process<sup>2</sup>.

Patients were matched at a ratio of three TKRs to one UKR, within a caliper of 0.2 standard deviation of the logit of the propensity score<sup>3</sup>. The logistic model used to create the propensity scores included patient factors (age, sex, comorbidities, and Index of Multiple Deprivation [IMD] rank), surgical factors (cementless or cemented fixation and thromboprophylaxis, although detailed implant information including brand, component size, and whether the implant had a fixed or mobile bearing was not available), and surgeon/unit factors (proportion of a surgeon's knee replacements that were UKRs [termed "usage" of UKR], duration of follow-up, whether the operation was performed by a consultant or a trainee, and whether the unit was public, private, or an independent-sector treatment center [where NHS patients are treated by private providers]). Patients were matched for the overall practice size of their surgeon (number of TKRs and UKRs performed each year). Matching was repeated for each group when patient-reported outcomes were compared; for these comparisons, patients were also matched for preoperative OKS and EQ-5D scores. Comorbidities were assessed using the Charlson comorbidity index<sup>4</sup>, the American Society of Anesthesiologists (ASA) score<sup>5</sup>, and any patient-reported conditions from the patient-reported outcomes database. Deprivation was assessed using the IMD, which is a U.K. governmental measure that assesses the level of deprivation in multiple domains (such as employment, income, and access to education) in small geographical areas<sup>6</sup>. The IMD is expressed as a rank from most deprived (1) to least deprived (32,482).

### *Propensity Score Creation*

Matching was performed using propensity score methods in order to adjust for baseline differences in patient characteristics and to address the issue of confounding by indication. Each comparison (TKR versus low, medium, and high-volume UKR) was carried out separately so the contribution of each variable to the propensity score was subtly different in each. Table E-1 presents the results of the logistic model used to determine the propensity scores for the high-volume group (by way of illustration). The odds ratio in the table represents the probability of a knee being assigned to UKR (values of >1) or total knee replacement (values of <1). All values were included in the propensity score calculation as has been advised by previous authors<sup>7</sup>.

### *Model Diagnostics for Matching*

The assessment of the matching procedures used in the study included two procedures:

1. Assessment of the power of the propensity scores to predict whether the patient would be treated with UKR or TKR was performed by plotting a receiver operating characteristics (ROC) curve. In this ROC curve, each value of propensity score was plotted; the vertical axis represents sensitivity and the horizontal axis represents 1 – specificity. The area under the curve (AUC) was 0.83 (Fig. E-1).

2. Before and after matching, the distributions of the propensity scores in each group were estimated using kernel density plots. These were used to determine the accuracy of the matching process (Figs. E-2 and E-3). ■

## References

1. Liddle AD, Judge A, Pandit H, Murray DW. Adverse outcomes after total and unicompartmental knee replacement in 101,330 matched patients: a study of data from the National Joint Registry for England and Wales. *Lancet*. 2014 Oct 18;384(9952):1437-45.
2. Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Stat Med*. 2014;33(7):1242-58.
3. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*. 2011 May;46(3):399-424. Epub 2011 Jun 8.
4. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-83.
5. Saklad M. Grading of patients for surgical procedures. *Anesthesiology*. 1941;2:281-4.
6. Noble M, Wright G, Dibben C, Smith GAN, McLennan D, Anttila C, Barnes H, Mokhtar C, Noble S. The English indices of deprivation 2004. London: Queen's Printer and Controller of Her Majesty's Stationery Office; 2003.
7. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol*. 2006 Jun 15;163(12):1149-56. Epub 2006 Apr 19.

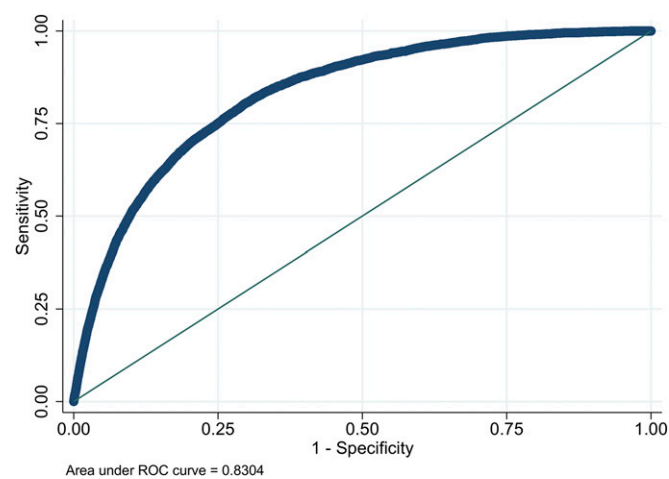


Fig. E-1

ROC curve of the propensity scores to determine the power with which the model predicts treatment allocation.

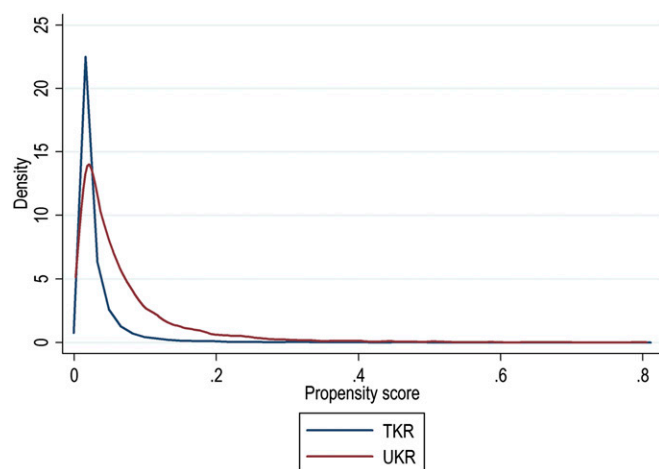


Fig. E-2

Distribution of propensity scores in each group before matching.

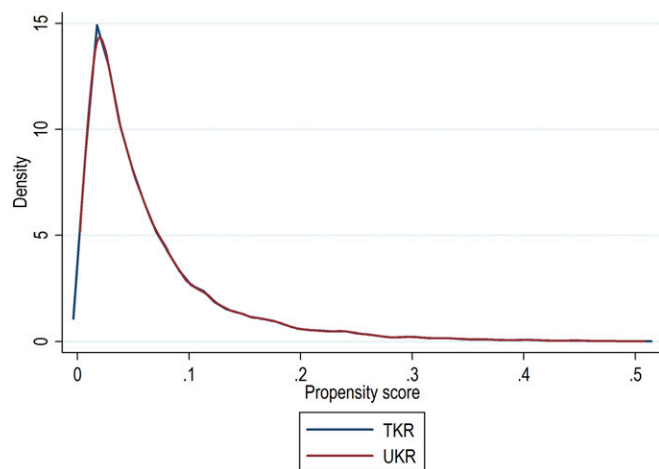


Fig. E-3

Distribution of propensity scores in each group after matching.

**TABLE E-1 Logistic Model for Creation of Propensity Score for the High-Volume Group**

Exposure	Odds Ratio	Z	P > Z	95% CI	
Age (by year)	0.95	−36.68	<0.001	0.95	0.95
Sex (male)	1.33	10.73	<0.001	1.26	1.40
Unit type					
Public	Reference group	Reference group	Reference group	Reference group	Reference group
Private	0.83	−4.37	<0.001	0.76	0.90
Independent-sector treatment center	0.88	−1.98	0.048	0.78	1.00
ASA (by group)	0.83	−7.16	<0.001	0.79	0.87
Fixation (cemented)	0.97	−0.77	0.444	0.90	1.05
Ethnicity (overall)	1.00	0.13	0.895	0.96	1.05
Index of Multiple Deprivation rank	1.00	8.01	<0.001	1.00	1.00
Charlson comorbidity index (by group)	1.02	0.62	0.532	0.97	1.07
Chemical thromboprophylaxis (overall)	0.90	−14.71	<0.001	0.89	0.92
Mechanical thromboprophylaxis (overall)	0.95	−3.75	<0.001	0.93	0.98
Surgery performed by consultant	1.73	15.98	<0.001	1.62	1.85
Time since surgery	1.00	−7.19	<0.001	1.00	1.00
Constant	0.03	−16.30	<0.001	0.02	0.05