COPYRIGHT © BY THE JOURNAL OF BONE AND JOINT SURGERY, INCORPORATED THIRUKUMARAN ET AL. NATURAL LANGUAGE PROCESSING FOR THE IDENTIFICATION OF SURGICAL SITE INFECTIONS IN ORTHOPAEDICS http://dx.doi.org/10.2106/JBJS.19.00661 Page 1

The following content was supplied by the authors as supporting material and has not been copy-edited or verified by JBJS.

Appendix Table 1: Distribution of characteristics between selected controls (from years 2015 and 2016), and controls from other years (2011 to 2014, and 2017)

	Selected controls (2015 and 2016)	Non-selected controls (2011 to 2014, 2017)	Total	p-value
Ν	1,407	5,890	7,297	
Age: Mean (SD)	45.9 (20.5)	46.6 (19.9)	46.5 (20.0)	0.4
Female: N (%)	689 (49)	2,838 (48)	3,527 (48)	0.6
Race: N (%)				0.8
White	1,231 (87)	5,132 (87)	6,363 (87)	
Black	124 (9)	518 (9)	642 (9)	
Other	52 (4)	240 (4)	292 (4)	
Average number of comorbidities: Median (IQR)	0 (0 to 1)	0 (0 to 1)	0 (0 to 1)	0.4

Notes:

- N Number of patients,
- Mean (SD) Mean (Standard deviation)
- N (%) Number (Column percentage)
- Comorbidities were identified using Elixhauser's comorbidity algorithm
- Median (IQR) Median (Interquartile range)
- P-values for chi-square tests for categorical variables and Kruskal-Wallis tests for continuous variables reporting mean and standard deviations (age), and equality of median tests for skewed continuous variables reporting median and interquartile range (average number of comorbidities)