Are Database Studies Reproducible and Reliable? An Update

Ayobami S. Ogunsola, MD, MPH[1]; Michael C. Marinier, BS[1]; Jacob M. Elkins, MD, PhD[1]
[1]University of Iowa Department of Orthopedics and Rehabilitation, Iowa City, IA
Email of Presenting Author: michael-marinier@uiowa.edu

Disclosures: Ayobami S. Ogunsola (N) Michael C. Marinier (N), Jacob M. Elkins (3B: Depuy Synthes; 5: DePuy Synthes, IronMind; 8: Editorial Board –J Arthroplasty)

INTRODUCTION: Research studies greatly influence treatment guidelines, as orthopaedic surgeons should practice evidence-based medicine. However, studies follow an evidence hierarchy, which ranks their strength of evidence based on data collection method and presence of randomization or controls. Database studies, like the ACS-NSQIP database, have gained popularity in orthopedics, but generally do not achieve high levels of evidence. While establishing causation is challenging in cross-sectional studies, consistency and reproducibility of findings are often overlooked. This study aims to determine the reproducibility of ACS–NSQIP arthroplasty studies on smoking and its complications by employing identical dataset and statistical methods. Furthermore, this study aims to enhance the reliability and applicability of database studies. Pilot data for this research was originally presented at ORS 2023, and this study aims to expand on the original data.

METHODS: A comprehensive PubMed search, including terms “arthroplasty”, “smoking”, “complications”, and “ACS-NSQIP”, was used to identify relevant studies published between 2011 and 2022. Once studies were identified, each study’s methods were reproduced by a trained epidemiologist and statistician based on the original authors’ reported methodology. If a required step in analysis (i.e. data cleaning, handling of missing values) was not specifically stated in the original publication, the task was reproduced at our statistician’s discretion. To determine reproducibility, the adjusted odds ratios (aORs) and p-values ($\alpha = 0.05$) were compared between the original publication and the reanalyzed dataset.

RESULTS SECTION: The initial search generated 43 studies, 22 of which were selected for full-text review. After full text-review, 11 studies met the inclusion criteria and datasets were obtained for repeat analysis. Among the 11 analyzed studies, there were 268 aORs reanalyzed: 12.69% of the original studies’ aORs crossed 1 upon reanalysis, and 12.83% changed statistical significance (Table I). Additionally, the average magnitude change of each individual aOR was 17.22% across all studies, and the N included in the analyses varied by up to 47.84%.

DISCUSSION: This study aimed to improve the reliability and consistency of evidence by reanalyzing published database studies. This study includes an additional 6 studies than was previously presented. Across 11 commonly cited studies, approximately one of eight objective conclusions changed whether the exposure (smoking) was harmful or protective to the outcome. Additionally, 12.83% of the compared results had changes in statistical significance. The variability between the original and reproduced results are likely secondary to both systems and individual issues. For example, with one reproduced study including just over 50% of the original study’s N there are likely preferences in each statisticians’ data cleaning and handling of missing values. Furthermore, systems issues, such as how institutions store and report their data to registries, likely contribute to the overall data aggregation.

SIGNIFICANCE/CLINICAL RELEVANCE: Overall, the variability between original studies and this reproduced data indicates that orthopaedic surgeons should take heed of the level of evidence of database studies and seek higher validity studies when available. This research fills a crucial gap in assessing the reliability and consistency of evidence in orthopaedic surgery.

IMAGES AND TABLES:

Table I. Summary of reproduced adjusted odds ratios (aORs). Specifically highlighter are those results that changed whether an aOR varied from protective to non-protective (aOR crossed 1), individual magnitude change of each aOR, significance change of aORs (with $\alpha = 0.05$), and changes in the N between original and reproduced data.

	aORs				*aOR Significance**		*Comparison of N*			
Study Number	*Original Study (N)*	*Crossed 1 (N)*	*Changed Sign (%)*	*Average Magnitude Change (%)*	*Changed Significance (N)*	*Changed Significance (%)*	*Original (N)*	*Reproduced (N)*	*Difference (N)*	*Difference (%)*
1	15.00	4.00	26.67	9.40	5.00	33.33	78191	83736	5545	7.09
2	7.00	2.00	28.57	4.99	5.00	71.43	67897	35413	32484	47.84
3	3.00	0.00	0.00	47.00	2.00	66.67	2502	2476	26	1.04
4	3.00	0.00	0.00	17.30	NA**	NA**	8237	8790	553	6.71
5	42.00	3.00	7.14	2.10	6.00	14.29	5068	5198	130	2.57
6	3.00	0.00	0.00	13.67	1.00	33.33	8776	7945	831	9.47
7	40.00	14.00	35.00	5.35	4.00	10.00	2088	2208	120	5.75
8	40.00	5.00	12.50	17.01	5.00	12.50	169406	169406	0	0.00
9	56.00	2.00	3.57	26.34	5.00	8.93	10112	9788	324	3.20
10	35.00	4.00	11.43	39.49	1.00	2.86	120742	167402	46660	38.64
11	24.00	0.00	0.00	6.79	0.00	0.00	210075	210064	11	0.01
Total	**268.00**	**34.00**	**12.69%**	**17.22%**	**34.00**	**12.83%**	**683094**	**702426**	**19332**	**2.83%**

*determined by $p < 0.05$
**study did not provide p-values