# Should We Question the External Validity of Database Studies? A Comparative Analysis of Demographics

Alexander Beletsky, B.A., Yining Lu, B.A., Bhavik H. Patel, B.S., B.A., Jorge Chahla, M.D., Ph.D., Gregory L. Cvetanovich, M.D., Brian Forsythe, M.D., Brian J. Cole, M.D., and Nikhil Verma, M.D.

**Purpose:** To define the external validity of national and institutional databases for common sports medicine procedures. **Methods:** Patient demographic data including age, sex, body mass index (BMI), and 4 racial categories were aggregated between 2007 and 2016 across 2 databases for 4 common sports medicine procedures: anterior cruciate ligament reconstruction, arthroscopic rotator cuff repair (RCR), partial meniscectomy (PMx), and both arthroscopic and open shoulder stabilization. The first database of interest was a prospectively collected institutional database. The second was the National Surgical Quality Improvement Program (NSQIP) database. Two-sample *t* tests were performed to examine mean differences (MDs) in age and BMI, and $\chi^2$ testing was used to test differences in sex and race. **Results:** A total of 7,019 institutional and 108,881 NSQIP patients were examined. The NSQIP cohort was significantly older (MD, 1.40 years), included more female patients (42.60% female patients vs 35.67% female patients), and showed a different racial distribution compared with the institutional data (all $P < .0001$). The NSQIP PMx cohort (MD, 7.38 years) was significantly older and the NSQIP RCR cohort (MD, 1.97 years) was significantly younger than their institutional counterparts (all $P < .0001$). The NSQIP anterior cruciate ligament reconstruction cohort (MD, 2.53) showed a greater average BMI ($P < .0001$). The NSQIP RCR cohort (41.8% female patients vs 33.3% female patients) and PMx cohort (46.0% female patients vs 37.9% female patients) also included more female patients. Race was distributed variably between databases for each procedure code (all $P < .0001$). **Conclusions:** Significant differences in age, BMI, sex, and race distributions were observed between an institutional database and the NSQIP database. This study underlines the importance of defining the generalizability of database research, particularly when significant demographic differences between databases may underlie differences in postoperative outcomes. **Level of Evidence:** Level III, cross-sectional study.

*See commentary on page 2695*

*From the Division of Sports Medicine, Midwest Orthopaedics at Rush, Rush University Medical Center (A.B., Y.L., B.H.P., J.C., B.F., B.J.C., N.V.), Chicago, Illinois; and Division of Sports Medicine, Department of Orthopaedics, The Ohio State University Wexner Medical Center (G.L.C.), Columbus, Ohio, U.S.A.*

*The authors report the following potential conflicts of interest or sources of funding: B.F. receives paid royalties from Elsevier; research support from Arthrex; owns stock or stock options in Jace Medical; receives fellowship support from Smith and Nephew and Össur; and is a consultant for and receives research support from Stryker. B.J.C. receives research support from Aesculap/B.Braun, National Institutes of Health (NIAMS and NICHD), and Zimmer; is on the editorial or governing board of the* American Journal of Orthopedics, American Journal of Sports Medicine, Cartilage, Journal of Shoulder and Elbow Surgery, *and* Journal of the American Academy of Orthopaedic Surgeons*; receives IP royalties and research support from Arthrex; is a board or committee member of the Arthroscopy Association of North America and International Cartilage Repair Society; receives other financial or material support from Athletico, JRF Ortho, and Smith and Nephew; receives IP royalties from Elsevier; receives publishing royalties from* Operative Techniques in Sports Medicine*; receives stock or stock options from Ossio; and receives stock or stock options and research support from Regentis. N.V. receives research support from Arthrex and DJ Orthopaedics; receives publishing royalties and material support from* Arthroscopy*; receives consultant fees from Orthospace; and receives publishing royalties from Vindico Medical-Orthopedics Hyperguide. Full ICMJE author disclosure forms are available for this article online, as supplementary material.*

*Paper previously presented as a podium presentation at: Arthroscopy Association of North America Annual Meeting, Orlando, Florida, May 2–4, 2019.*

*Received January 16, 2019; accepted May 10, 2019.*

*Address correspondence to Nikhil Verma, M.D., Midwest Orthopaedics at Rush, Rush University Medical Center, 1611 W Harrison St, Ste 300, Chicago, IL 60612, U.S.A. E-mail: nikhil.verma@rushortho.com*



*0749-8063/1965/$36.00*

*https://doi.org/10.1016/j.arthro.2019.05.020*

From 2000 to 2010, publications in the field of orthopaedics increased greater than twofold (384.6 articles per year; from 2,889 to 6,909) due in large part to retrospective studies using clinical registry and insurance claims data.[1-5] Database studies are of particular interest in our current health care environment because policy makers, administrators, and physicians may use national and private-payer databases to evaluate trends in surgical management, optimize patient outcomes, and risk stratify surgical candidates, respectively.[6-8] With continued growth of database utilization in orthopaedic research, it is becoming increasingly important for researchers and clinicians alike to recognize and understand the strengths, weaknesses, and limitations of database studies.[2,3,6]

Despite offering large sample sizes, database studies have important considerations including (1) highly significant *P* values that may lack clinical significance, (2) the potential for residual confounding by uncoded variables, (3) the risk of user input errors, (4) recording bias due to financial and other incentives, and (5) inconsistent handling of missing variables across studies.[6,8-10] Of equal concern is the concept of external validity, or the idea that a study's generalizability is dependent on the patient population and other factors potentially impacting data collection.[11,12] Although the external validity of national data sets relative to multiple randomized controlled trials has previously been compared for sports medicine procedures, as well as across databases in total hip arthroplasty, the external validity of national data sets relative to high-volume institutional data sets for sports medicine procedures has not been established.[12,13] This is particularly important in determining whether differences in specific demographic factors, such as age, sex, and body mass index (BMI), may prohibit the generalization of results from a given study to particular practice scenarios.

For the aforementioned reasons, the purpose of this study was to define the external validity of national and institutional databases for common sports medicine procedures. We hypothesized that (1) significant mean differences (MDs) with respect to age, sex, BMI, and race would exist between databases; (2) significant differences in BMI and race would persist when stratifying by Current Procedural Terminology (CPT) code; and (3) significant differences in age and sex would become insignificant after CPT stratification.

## Methods

### Cohort Establishment

This study was approved by our institutional review board based on the analysis of retrospective, deidentified data. Two databases were used in this study. The first is the database of the American College of Surgeons (ACS)—National Surgical Quality Improvement Program (NSQIP), a national clinical registry used across a variety of disciplines that aggregates data across multiple participating hospitals nationwide.[14-16] The second database is an institutional database used for a variety of case series and cohort studies.[17-22] Cohorts from both databases were established using CPT codes to cover a broad range of common sports medicine procedures, inclusive of anterior cruciate ligament reconstruction (ACLR) (code 29888), partial meniscectomy (PMx) (code 29880 or 29881), arthroscopic rotator cuff repair (RCR) (code 29827), and both arthroscopic soft-tissue and open bony shoulder stabilization (SS) procedures (code 29806). The ACS-NSQIP cohort was constructed using data from July 2007 to June 2016. This database includes prospectively collected, deidentified data gathered by trained operators at participating hospitals using a systematic sampling process.[23] The NSQIP database contains over 5.5 million patient cases, with the 2016 report compiling 150 coded variables inclusive of demographic, intraoperative, and postoperative variables.[24] The institutional database was similarly queried for the aforementioned procedures between July 2007 and June 2017 using an electronic data collection service (Outcome Based Electronic Research Database; Universal Research Solutions, Columbia, MO). This database is a prospectively collected, single-institution clinical registry from an orthopaedic group based in Chicago, Illinois.

### Variables Examined and Statistical Analysis

All statistical analyses were conducted with RStudio software (version 1.0.143; R Foundation for Statistical Computing, Vienna, Austria). Demographic variables of interest were deidentified for comparison of 4 variables across the prospective databases: sex, age, BMI, and race. All BMI measurements were coded for in the NSQIP database. In the institutional database, BMI was either gathered from a form or directly calculated from measurements of height (in square meters) and weight (in kilograms). Race was categorized into 4 distinct categories: American Indian, Asian, African American, and white. "American Indian" was the sum of "American Indian or Alaska" and "American Indian or Alaskan Native" responses; "Asian" included both those who responded "Asian" and those who responded "Native Hawaiian or Pacific Islander." Other variables of interest including smoking status, medical comorbidities, orthopaedic-specific variables (i.e., duration of symptoms and injury mechanism), and other racial categories (i.e., Hispanic) were unavailable in 1 or both compared databases and thus were excluded from the analysis. Power analysis revealed minimum cohort sizes of 2,976 to provide 90% power in detecting a year difference in age with a type I error rate of 1%. Continuous variables including age and BMI were compared across prospective studies using the Student *t* test. Categorical variables including sex and race were

**Table 1.** Overall Trends in Age, Sex, BMI and Race Between Institutional and NSQIP Databases Without Stratification

| Variable | Institutional (n = 7,019) | NSQIP (n = 108,881) | *P* Value |
|---|---|---|---|
| Age, mean (SD), yr | 48.45 (17.10) | 49.85 (15.51) | $5.02 \times 10^{-5}$ (Student *t* test) |
| BMI, mean (SD) | 31.16 (7.11) | 31.33 (7.00) | .0486 (Student *t* test) |
| Female sex, n (%) | 2,504 (35.67) | 46,383 (42.60) | $2.21 \times 10^{-26}$ ($\chi^2$ test) |
| Race, n (%)* | | | $4.29 \times 10^{-59}$ ($\chi^2$ test) |
| AI | 10 (0.15) | 1,046 (1.22) | |
| Asian | 177 (2.69) | 4,171 (4.83) | |
| AA | 932 (14.14) | 8,009 (9.27) | |
| White | 5,471 (83.02) | 73,210 (84.68) | |

AA, African American; AI, American Indian; BMI, body mass index; NSQIP, National Surgical Quality Improvement Program; SD, standard deviation.

*The race variable includes a total of 6,590 institutional and 86,469 NSQIP patients, excluding racial groups that were not coded for similarly across the 2 databases.

examined between databases using the Pearson $\chi^2$ test. After initial analysis, the data were stratified by CPT code and the aforementioned analysis was repeated. All statistical tests were 2-tailed, with the level of significance set at $\alpha = .01$. We used a greater threshold for significance in the context of comparing 2 highly powered, large databases that could allow for the detection of small differences without clinical significance.[11,13,25,26] Thus, when comparing *P* values, we examined magnitudes of differences with thresholds greater than twofold being suggestive of important clinical differences.[13]

## Results

### Overall Cohort

We identified a total of 108,881 patients (63,545 with PMx, 24,354 with RCR, 16,222 with ACLR, and 4,760 with SS) in the NSQIP database and a total of 7,019 patients (3,478 with PMx, 2,469 with RCR, 625 with ACLR, and 447 with SS) in the institutional database. Overall cohort comparisons showed a significant difference in age (48.45 ± 17.10 years vs 49.85 ± 15.51 years, $P < .0001$) between the single-institution and NSQIP patients, without a significant difference in BMI ($P = .048$). The results of $\chi^2$ testing revealed significantly different distributions of female and male patients between databases (42.60% female patients vs 35.67% female patients, $P < .0001$), as well as significant differences in racial distributions between the 2 cohorts ($P < .0001$) (Table 1). The high-volume institutional database showed a greater proportion of African American patients (14.14% vs 9.27%), whereas the NSQIP data set had greater percentages of American Indian (1.22% vs 0.15%), Asian (4.83% vs 2.69%), and white (84.68% vs 83.02%) patients (Fig 1).

### Age and BMI Stratified by CPT Code

Table 2 displays mean and standard deviation values for age and BMI after the 2 cohorts were stratified by CPT code. With respect to age, PMx patients showed the most significant difference when mean values were

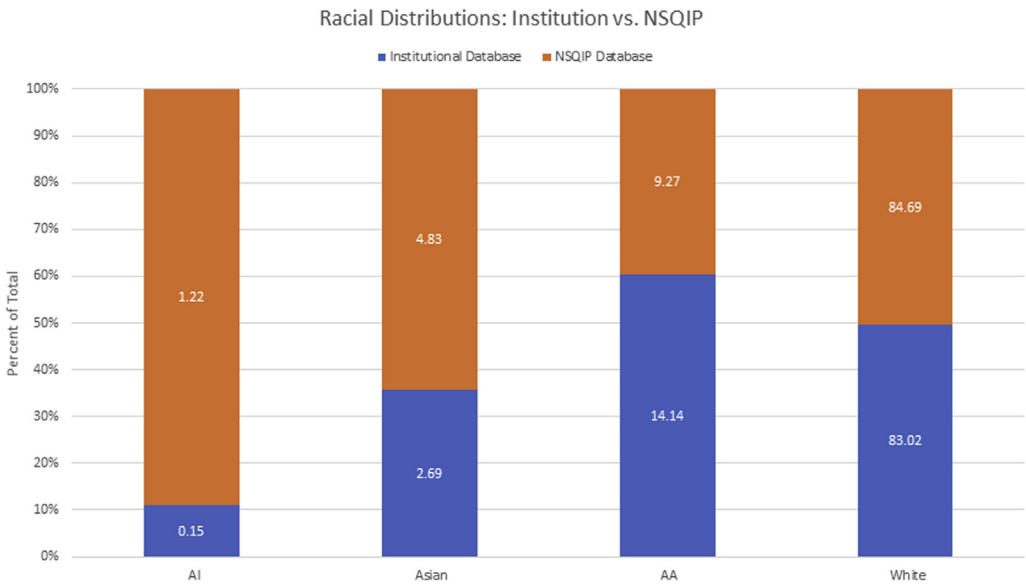


**Fig 1.** Racial distributions for institutional versus National Surgical Quality Improvement Program (NSQIP) databases. Significant differences ($P < .0001$) were observed during $\chi^2$ testing across the 4 listed categorical variables. AA, African American; AI, American Indian.

**Table 2.** Weighted *t* Tests for Age and BMI Stratified by Procedure Between Institutional and NSQIP Databases

| Procedure | Institutional | NSQIP | *P* Value (Weighted *t* Test) |
|---|---|---|---|
| Age, yr | | | |
| ACLR | 30.46 (12.22) | 32.16 (10.71) | $6.97 \times 10^{-4*}$ |
| RCR | 60.38 (10.00) | 58.41 (10.92) | $1.06 \times 10^{-26*}$ |
| PMx | 45.07 (14.87) | 52.45 (12.15) | $1.48 \times 10^{-123*}$ |
| SS | 31.55 (15.05) | 32.27 (12.03) | .03 |
| BMI | | | |
| ACLR | 26.27 (5.56) | 28.80 (6.01) | $6.84 \times 10^{-26*}$ |
| RCR | 31.35 (6.19) | 31.30 (6.60) | .71 |
| PMx | 31.46 (7.76) | 32.20 (7.34) | .28 |
| SS | 27.51 (5.40) | 28.19 (5.70) | .01 |

NOTE. Data are presented as mean (standard deviation).

ACLR, anterior cruciate ligament reconstruction; BMI, body mass index; NSQIP, National Surgical Quality Improvement Program; PMx, partial meniscectomy; RCR, rotator cuff repair; SS, shoulder stabilization.

*$P < 10^{-10}$.

compared between the institutional patients and their NSQIP counterparts (45.07 ± 14.87 years vs 52.45 ± 12.15 years, $P < .0001$). The RCR cohort (60.38 ± 10.00 years vs 58.41 ± 10.92 years, $P < .0001$) and ACLR cohort (30.46 ± 12.22 years vs 32.16 ± 10.71 years, $P < .0001$) also displayed significant MDs, with only the SS cohort showing an insignificant difference in age ($P = .03$).

With respect to BMI, ACLR patients exhibited a significant difference between the institutional and NSQIP cohorts (26.27 ± 5.56 vs 28.80 ± 6.01, $P < .0001$). However, neither RCR patients ($P = .71$), PMx patients ($P = .28$), nor SS patients ($P = .01$) showed significant MDs with respect to BMI (Table 2).

### Sex and Race Stratified by CPT Code

Significant differences in the distribution of sex existed across the institutional and NSQIP databases for RCR patients (33.3% female patients vs 41.8% female patients, $P < .0001$) and PMx patients (37.9% female patients vs 46.0% female patients, $P < .0001$) (Table 3). However, we observed no significant differences when examining the ACLR cohort (39.2% female patients vs 35.8% female patients) and SS cohort (27.1% female patients vs 24.6% female patients) (Fig 2).

With respect to the stratification of race by CPT code, racial distributions differed for each CPT code of interest, although the distribution of *P* values was highly variable (ranging from $P = 5.73 \times 10^{-50}$ to $P = .0071$) (Table 4). The results of $\chi^2$ testing for the RCR cohort ($P < .0001$) showed the greatest magnitude of significant differences in categorical distributions, followed by the PMx cohort ($P < .0001$). Both the ACLR cohort ($P = .0007$) and SS cohort ($P = .007$) showed significant differences although less in magnitude than either the RCR or PMx cohort. The greatest differences by racial category between the institutional and NSQIP cohorts observed included the following: 6% (80.1% vs 86.1%) and 5.9% (90.9% vs 85.0%) differences in the percentage of white patients in the RCR and SS cohorts, respectively; a 9.1% difference (18.5% vs 9.4%) in the percentage of African American patients treated in the RCR cohort; a 2.3% difference (1.3% vs 3.6%) in the percentage of Asian patients treated in the RCR cohort; and a 2.5% difference (0.2% vs 2.7%) in the percentage of American Indian patients treated in the ACLR cohort (Table 4, Fig 3).

## Discussion

The main findings of this study were that significant differences with respect to age, BMI, sex, and race exist between a prospectively collected institutional database and the ACS-NSQIP national database, both of which represent highly used databases well represented in the literature. Overall, the NSQIP cohort was older, included a greater percentage of female patients, and had a different racial distribution without significant differences in the average BMI overall. When data were stratified by CPT code, the greatest MDs in age were observed for the PMx cohort (45.07 years vs 52.45 years), and the ACLR cohort was the only procedural cohort to show significant MDs with respect to BMI. Large differences (>5%) in the distribution of sex were particularly apparent for the PMx and RCR cohorts. Each of the 4 procedural cohorts (PMx, RCR, ACLR, and SS) exhibited significant differences in racial distributions. The greatest differences in cohort percentages by racial category were observed for the RCR (6%) and SS (5.9%) cohorts for white patients, the

**Table 3.** Results of $\chi^2$ Testing for Sex Stratified by Procedure (CPT Code) Between Institutional and NSQIP Databases

| | Institutional | | NSQIP | | | |
|---|---|---|---|---|---|---|
| Procedure | Male | Female | Male | Female | $\chi^2$ Statistic | *P* Value |
| ACLR | 380 (60.8) | 245 (39.2) | 10,429 (64.3) | 5,793 (35.7) | 2.97 | .085 |
| RCR | 1,648 (66.7) | 821 (33.3) | 14,174 (58.2) | 10,180 (41.8) | 67.1 | $2.54 \times 10^{-16}$ |
| PMx | 2,161 (62.1) | 1,317 (37.9) | 34,309 (54.0) | 29,236 (46.0) | 38.8 | $4.76 \times 10^{-10}$ |
| SS | 326 (72.9) | 121 (27.1) | 3,591 (75.4) | 1,169 (24.6) | 2.67 | .102 |

NOTE. Data are presented as number (percentage).

ACLR, anterior cruciate ligament reconstruction; CPT, Current Procedural Terminology; NSQIP, National Surgical Quality Improvement Program; RCR, rotator cuff repair; PMx, partial meniscectomy; SS, shoulder stabilization.

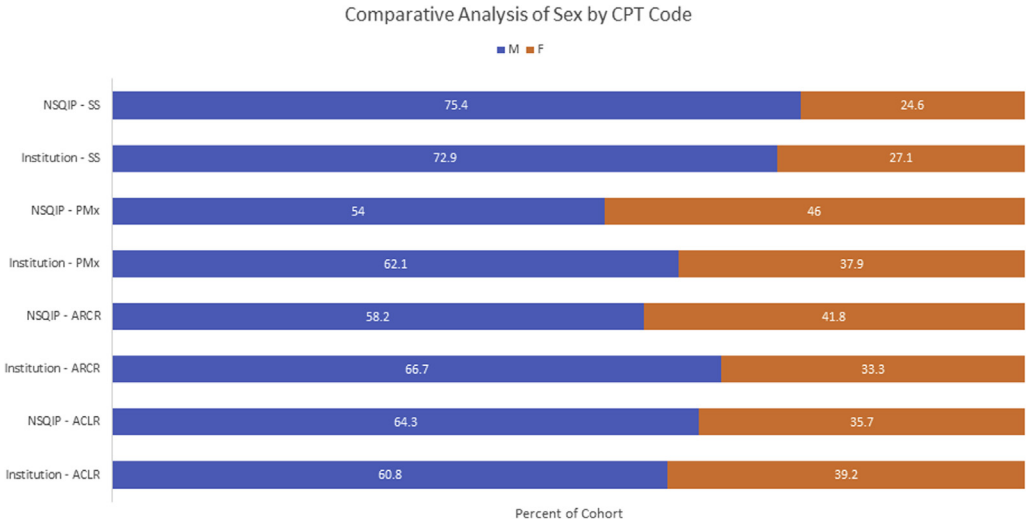


**Fig 2.** Distribution of sex by Current Procedural Terminology (CPT) code for institutional versus National Surgical Quality Improvement Program (NSQIP) databases. ACLR, anterior cruciate ligament reconstruction; ARCR, arthroscopic rotator cuff repair; F, female; M, male; PMx, partial meniscectomy; SS, shoulder stabilization.

RCR cohort (9.1%) for African American patients, the RCR cohort (2.3%) for Asian patients, and the ACLR cohort (2.5%) for American Indian patients. Taken together, the important demographic differences our study identifies serve to better define the important factors one must consider prior to translating the results of database research into practice.

The most important corollary of the aforementioned data is the translation of the results of our highly powered database comparison into clinical significance. Numerous previous studies have reported that even minor differences in patient demographic characteristics may exert significant effects on patient outcomes, including small differences in age and BMI impacting mean patient-reported outcome (PRO) scores, patient satisfaction, or complications such as SS revision, redislocation, and anterior cruciate ligament retear.[12,27-29] Although we did not find any significant differences with respect to BMI between databases for the overall cohort, we did find significant MDs in age ($P < .0001$), with the PMx cohort showing the greatest MD and $P$ value (MD, 7.38 years; $P = 1.48 \times 10^{-123}$) followed by the RCR cohort (MD, 1.97 years; $P = 1.06 \times 10^{-26}$) and ACLR cohort (MD, 1.7 years; $P = 6.97 \times 10^{-4}$). These differences in age may be explained in part by regional population variations but also may be a result of practice differences between the academic hospitals comprising the NSQIP database and the single center included in our study. Age has previously been shown to influence preoperative and postoperative changes in mean PRO scores, as well as complications, in arthroscopic PMx patients.[30-32] In addition, age has been shown to be predictive of unanticipated hospital admission in RCR patients[33-35] and of minimal clinically important difference achievement and ligament retear in ACLR patients.[36,37] Although the MDs in age for the RCR and ACLR cohorts were both less than 2 years, the MD in age between databases for the PMx cohort was 7.38 years. This finding suggests that the older NSQIP PMx cohort may be uniquely susceptible to both poor outcomes (i.e., meniscal retear and decreased PRO improvements) compared with the younger institutional PMx cohort.

**Table 4.** Results of $\chi^2$ Testing for Race Stratified by Procedure (CPT Code) Between Institutional and NSQIP Databases

| | Institutional | | | | NSQIP | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Procedure | AI | Asian | AA | White | AI | Asian | AA | White | $\chi^2$ Statistic | $P$ Value |
| ACLR | 1 (0.2) | 32 (6.0) | 70 (13.2) | 427 (80.6) | 308 (2.7) | 944 (8.0) | 1,295 (11.0) | 9,215 (78.3) | 16.96 | .00072 |
| RCR | 2 (<0.1) | 31 (1.3) | 433 (18.5) | 1,870 (80.1) | 184 (0.9) | 757 (3.6) | 1,945 (9.4) | 17,944 (86.1) | 231.77 | $5.73 \times 10^{-50}$ |
| PMx | 2 (<0.1) | 103 (3.2) | 416 (12.6) | 2,786 (84.2) | 520 (1.1) | 2,326 (4.6) | 4,423 (8.8) | 43,031 (85.5) | 95.69 | $1.31 \times 10^{-20}$ |
| SS | 5 (1.1) | 11 (2.6) | 23 (5.4) | 388 (90.9) | 34 (1.0) | 144 (4.2) | 337 (9.8) | 2,920 (85.0) | 12.09 | .0071 |

NOTE. Data are presented as number (percentage). The race variable includes a total of 6,590 institutional and 86,469 NSQIP patients, excluding racial groups that were not coded for similarly across the 2 databases.

ACLR, anterior cruciate ligament reconstruction; CPT, Current Procedural Terminology; NSQIP, National Surgical Quality Improvement Program; PMx, partial meniscectomy; RCR, rotator cuff repair; SS, shoulder stabilization.

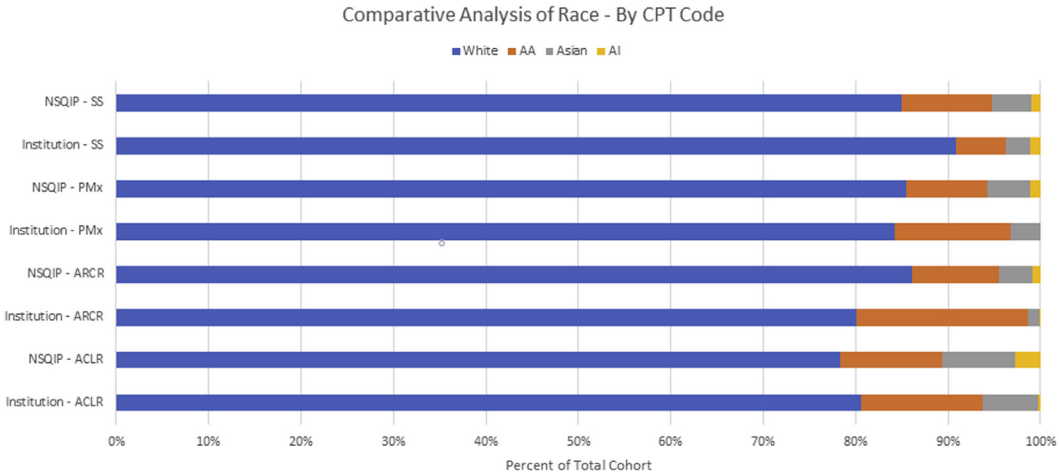


**Fig 3.** Racial distributions by Current Procedural Terminology (CPT) code for institutional versus National Surgical Quality Improvement Program (NSQIP) databases. AA, African American; ACLR, anterior cruciate ligament reconstruction; ARCR, arthroscopic rotator cuff repair; AI, American Indian; PMx, partial meniscectomy; SS, shoulder stabilization.

With respect to the RCR cohort, unanticipated readmission was specifically studied in an NSQIP population found to be significantly younger, with variable sex and race distributions, compared with the institutional cohort used in our study. Examination of readmission rates at our single institution is warranted to understand whether age is similarly predictive of readmission despite the demographic differences observed between databases. This example underlines the importance of careful demographic examination between study cohorts prior to the application of a study's findings to a specific clinical practice.

With respect to BMI, we did find a significant MD in the ACLR cohort (MD, 2.53; $P < .0001$), a finding that has particular significance given that greater BMI has been linked to worse outcomes after ACLR.[38,39] Studies have yet to comparatively examine outcomes of ACLR between national and single-institution data sets.[40] An important limitation of such study designs is inconsistent coding of variables across databases, making the analysis of outcomes more challenging. Future studies examining ACLR outcomes, particularly when comparing multiple data sets, should consider BMI an important variable that can affect both outcomes and generalizability. In addition, important differences with respect to the distribution of sex existed for both the RCR (MD, 8.5%; $P < .0001$) and PMx (MD, 8.1%; $P < .0001$) cohorts. Female sex has been identified as a risk factor for decreased functional outcomes after arthroscopic partial RCR in irreparable rotator cuff tears, as well as for unexpected hospital readmission[33,35]; however, other studies have reported insignificant relations between female sex and outcomes after RCR.[41,42] It is interesting to note that the association with unanticipated readmission was examined in an NSQIP cohort, which our study shows contains disproportionately more female patients who may be more susceptible to worse outcomes. Further examination of the role of female sex regarding outcomes after RCR is warranted given inconsistent relation reporting in the literature. With respect to PMx, decreased postoperative knee function and a longer time to maximal medical improvement after PMx have been reported for female patients.[43,44] It is important to note that differences in the distribution of sex may underlie any differences in outcomes after PMx between high-volume institutions and national data sets. Further studies comparing outcomes across databases should consider using matched-cohort designs that control for the important demographic factors that may influence outcome attainment.

The observed differences in racial distributions are likely due to geographic variations in patient populations between the single-institution and national databases used in this study. Stratifying by procedure identified the RCR cohort as the main driver of significant differences in overall cohort racial distributions ($P = 5.73 \times 10^{-50}$). The PMx ($P = 1.31 \times 10^{-20}$), ACLR ($P = .0007$), and SS ($P = .007$) cohorts also showed significantly different racial distributions; however, differences in the magnitude of association suggest the RCR cohort to have the greatest clinical significance, followed by the PMx cohort. Within the RCR cohort, the NSQIP cohort contained 6.1% more white patients and 2.3% more Asian patients whereas the institutional cohort contained 3.8% more African American patients. The impact of race on outcomes after RCR, PMx, and SS is not well described in the literature and represents an important area of future study to better ascertain the role of observed demographic differences between databases. An interesting finding was that the ACLR cohort showed 2.3 more white patients per 100 in the institutional database than in the NSQIP database. Recent literature has suggested that white patients

are less likely to achieve the minimal clinically important difference in the International Knee Documentation Committee score after ACLR, although the highly powered 10-year MOON (Multicenter Orthopaedic Outcomes Network) ACLR cohort did not show any specific race as a risk factor for poor outcomes after ACLR.[36,45] Longitudinal follow-up data from the MOON shoulder instability cohort may help better ascertain the role of race regarding outcomes after SS procedures.[46] Similarly designed highly powered cohort studies that follow up patients longitudinally are indicated for both RCR and PMx populations to better ascertain the role of race regarding outcomes after common sports medicine procedures.

Previous studies have compared NSQIP and randomized controlled trial data for common sports medicine procedures, identifying important demographic differences between trial and national data.[12] In addition, other studies have reported differences in total knee arthroplasty rates based on sex, race, and geographic region within the United States and found important differences in patient demographic characteristics when comparing international ACLR cohorts with United States–based cohorts.[47,48] Our study adds to the growing body of necessary literature that better defines the scope of generalizability of database studies. We additionally identify important potential areas for future research into the role of demographic factors regarding outcomes after orthopaedic sports medicine procedures and encourage future research to use study designs that control for the possible impact of variable demographic characteristics on study findings.

### Limitations

This study does have limitations to consider. First, our comparison relies on 2 specific databases—1 institutional, the other a clinical registry. Health claims–based databases are gaining increased attention in the health literature, in part from insurance companies but also from national discharge databases and Medicare. Our study does not assess sports medicine research using health claims data and instead uses national data from participating hospitals via the NSQIP database. Second, both databases are highly powered (institution, n = 7,019; NSQIP, n = 108,881). As a result, we report large *P* values, making it difficult to conceptualize and compare magnitudes of significance. We attempted to address this important limitation through (1) a smaller threshold of significance ($\alpha = .01$) and (2) a comparative analysis focusing on the magnitude of difference between *P* values. Methodologies connecting statistical significance to clinical significance, such as an anchor- or distribution-based calculation of the minimal clinically important difference, were unable to be used given a lack of PRO data.[49] Nonetheless, our results must be interpreted within the scope of clinical relevance to avoid generating unnecessary conclusions. Third, our analysis of race was limited by differences in the coding of variable categories between databases, limiting our analysis to races with concurrent coding between the NSQIP data set and our institutional data set. Similarly, we were unable to include data on operative side because of variable limitations associated with our institutional electronic registry. Finally, all database studies may be subject to the effect of residual confounding if uncoded variables exert important effects on observed relations. This effect is more concerning in studies using regression models, however, and less concerning in studies such as ours that define variable distributions to test for significant differences between groups.[10]

## Conclusions

Significant differences in age, BMI, sex, and race distributions were observed between an institutional database and the NSQIP database. This study underlines the importance of defining the generalizability of database research, particularly when significant demographic differences between databases may underlie differences in postoperative outcomes.

## References

1. Lee KM, Ryu MS, Chung CY, et al. Characteristics and trends of orthopedic publications between 2000 and 2009. *Clin Orthop Surg* 2011;3:225-229.
2. Pugely AJ, Martin CT, Harwood J, Ong KL, Bozic KJ, Callaghan JJ. Database and registry research in orthopaedic surgery: Part I: Claims-based data. *J Bone Joint Surg Am* 2015;97:1278-1287.
3. Pugely AJ, Martin CT, Harwood J, Ong KL, Bozic KJ, Callaghan JJ. Database and registry research in orthopaedic surgery: Part 2: Clinical registry data. *J Bone Joint Surg Am* 2015;97:1799-1808.
4. Cunningham BP, Harmsen S, Kweon C, et al. Have levels of evidence improved the quality of orthopaedic research? *Clin Orthop Relat Res* 2013;471:3679-3686.
5. Sugrue CM, Joyce CW, Sugrue RM, Carroll SM. Trends in the level of evidence in clinical hand surgery research. *Hand (N Y)* 2016;11:211-215.
6. Pugely AJ, Bozic KJ. *Editorial Commentary:* Rising interest in "big data" in arthroscopy: Is the juice worth the squeeze? *Arthroscopy* 2017;33:232-233.
7. Weinreb JH, Yoshida R, Cote MP, O'Sullivan MB, Mazzocca AD. A review of databases used in orthopaedic surgery research and an analysis of database use in *Arthroscopy: The Journal of Arthroscopic and Related Surgery*. *Arthroscopy* 2017;33:225-231.
8. Patel AA, Singh K, Nunley RM, Minhas SV. Administrative databases in orthopaedic research: Pearls and pitfalls of big data. *J Am Acad Orthop Surg* 2016;24:172-179.
9. Whitehouse SL, Bolland BJ, Howell JR, Crawford RW, Timperley AJ. Mortality following hip arthroplasty—Inappropriate use of National Joint Registry (NJR) data. *J Arthroplasty* 2014;29:1827-1834.

10. Hailer NP. Orthopedic registry research—Limitations and future perspectives. *Acta Orthop* 2015;86:1-2.
11. Steckler A, McLeroy KR. The importance of external validity. *Am J Public Health* 2008;98:9-10.
12. Saltzman BM, Cvetanovich GL, Bohl DD, Cole BJ, Bach BR Jr, Romeo AA. Comparisons of patient demographics in prospective sports, shoulder, and national database initiatives. *Orthop J Sports Med* 2016;4:2325967116665589.
13. Bedard NA, Pugely AJ, McHugh MA, Lux NR, Bozic KJ, Callaghan JJ. Big data and total hip arthroplasty: How do large databases compare? *J Arthroplasty* 2018;33:41-45 e43.
14. Gowd AK, Cvetanovich GL, Liu JN, et al. Management of chondral lesions of the knee: Analysis of trends and short-term complications using the National Surgical Quality Improvement Program database. *Arthroscopy* 2018;35:138-146.
15. Kazaure HS, Roman SA, Rosenthal RA, Sosa JA. Cardiac arrest among surgical patients: An analysis of incidence, patient characteristics, and outcomes in ACS-NSQIP. *JAMA Surg* 2013;148:14-21.
16. Cvetanovich GL, Chalmers PN, Levy DM, et al. Hip arthroscopy surgical volume trends and 30-day postoperative complications. *Arthroscopy* 2016;32:1286-1292.
17. Gowd AK, Garcia GH, Liu JN, Malaret MR, Cabarcas BC, Romeo AA. Comparative analysis of work-related outcomes in hemiarthroplasty with concentric glenoid reaming and total shoulder arthroplasty. *J Shoulder Elbow Surg* 2019;28:244-251.
18. Saltzman BM, Cotter EJ, Stephens JP, et al. Preoperative tibial subchondral bone marrow lesion patterns and associations with outcomes after isolated meniscus allograft transplantation. *Am J Sports Med* 2018;46:1175-1184.
19. Cvetanovich GL, Weber AE, Kuhns BD, et al. Hip arthroscopic surgery for femoroacetabular impingement with capsular management: Factors associated with achieving clinically significant outcomes. *Am J Sports Med* 2018;46:288-296.
20. Cvetanovich GL, Gowd AK, Liu JN, et al. Establishing clinically significant outcome after arthroscopic rotator cuff repair. *J Shoulder Elbow Surg* 2019;28:939-948.
21. Gowd AK, Lalehzarian SP, Liu JN, et al. Factors associated with clinically significant patient-reported outcomes following primary arthroscopic partial meniscectomy. *Arthroscopy* 2019;35:1567-1575.e3.
22. Garcia GH, Liu JN, Wong AC, et al. The shoulder self-administered motion evaluation has excellent patient reliability and reproducibility on both physician and repeat follow-up testing. *Orthopedics* 2018;41:e820-e826.
23. American College of Surgeons. *ACS NSQIP participant use data file* Chicago: American College of Surgeons, 2019;Vol 2019, 2019.
24. American College of Surgeons National Surgical Quality Improvement Program. *User guide for the 2016 ACS NSQIP participant use data file (PUF)*. Chicago: American College of Surgeons, 2017.
25. Bohl DD, Russo GS, Basques BA, et al. Variations in data collection methods between national databases affect study results: A comparison of the nationwide inpatient sample and National Surgical Quality Improvement Program databases for lumbar spine fusion procedures. *J Bone Joint Surg Am* 2014;96:e193.
26. Bohl DD, Basques BA, Golinvaux NS, Baumgaertner MR, Grauer JN. Nationwide Inpatient Sample and National Surgical Quality Improvement Program give different results in hip fracture studies. *Clin Orthop Relat Res* 2014;472:1672-1680.
27. Wasserstein DN, Sheth U, Colbenson K, et al. The true recurrence rate and factors predicting recurrent instability after nonsurgical management of traumatic primary anterior shoulder dislocation: A systematic review. *Arthroscopy* 2016;32:2616-2625.
28. Clement ND, Walker LC, Merrie K, et al. Which patients are satisfied with their overall outcome but dissatisfied with their return to recreational activities after total knee arthroplasty? *Knee* 2019;26:258-266.
29. Kaeding CC, Pedroza AD, Reinke EK, Huston LJ, , MOON Consortium, Spindler KP. Risk factors and predictors of subsequent ACL injury in either knee after ACL reconstruction: Prospective analysis of 2488 primary ACL reconstructions from the MOON cohort. *Am J Sports Med* 2015;43:1583-1590.
30. Liebensteiner MC, Nogler M, Giesinger JM, Lechner R, Lenze F, Thaler M. Cartilage degeneration and not age influences the health-related quality of life outcome after partial meniscectomy. *Knee Surg Sports Traumatol Arthrosc* 2015;23:26-31.
31. Hupperich A, Salzmann GM, Niemeyer P, et al. What are the factors to affect outcome and healing of meniscus bucket handle tears? *Arch Orthop Trauma Surg* 2018;138:1365-1373.
32. Abram SGF, Judge A, Beard DJ, Price AJ. Adverse outcomes after arthroscopic partial meniscectomy: A study of 700 000 procedures in the national Hospital Episode Statistics database for England. *Lancet* 2018;392:2194-2202.
33. Galasso O, Riccelli DA, De Gori M, et al. Quality of life and functional results of arthroscopic partial repair of irreparable rotator cuff tears. *Arthroscopy* 2017;33:261-268.
34. Jenssen KK, Lundgreen K, Madsen JE, Kvakestad R, Dimmen S. Prognostic factors for functional outcome after rotator cuff repair: A prospective cohort study with 2-year follow-up. *Am J Sports Med* 2018;46:3463-3470.
35. Gil JA, Durand WM, Johnson JP, Goodman AD, Owens BD, Daniels AH. Unanticipated admission following outpatient rotator cuff repair: An analysis of 18,061 cases. *Orthopedics* 2018;41:164-168.
36. Nwachukwu BU, Chang B, Voleti PB, et al. Preoperative Short Form Health Survey score is predictive of return to play and minimal clinically important difference at a minimum 2-year follow-up after anterior cruciate ligament reconstruction. *Am J Sports Med* 2017;45:2784-2790.
37. Kaeding CC, Aros B, Pedroza A, et al. Allograft versus autograft anterior cruciate ligament reconstruction: Predictors of failure from a MOON prospective longitudinal cohort. *Sports Health* 2011;3:73-81.
38. Group MK, Spindler KP, Huston LJ, et al. Ten-year outcomes and risk factors after anterior cruciate ligament reconstruction: A MOON longitudinal prospective cohort study. *Am J Sports Med* 2018;46:815-825.

39. Group M, Allen CR, Anderson AF, et al. Surgical predictors of clinical outcomes after revision anterior cruciate ligament reconstruction. *Am J Sports Med* 2017;45:2586-2594.
40. Cvetanovich GL, Chalmers PN, Verma NN, Cole BJ, Bach BR Jr. Risk factors for short-term complications of anterior cruciate ligament reconstruction in the United States. *Am J Sports Med* 2016;44:618-624.
41. Ravindra A, Barlow JD, Jones GL, Bishop JY. A prospective evaluation of predictors of pain after arthroscopic rotator cuff repair: Psychosocial factors have a stronger association than structural factors. *J Shoulder Elbow Surg* 2018;27:1824-1829.
42. Namdari S, Donegan RP, Chamberlain AM, Galatz LM, Yamaguchi K, Keener JD. Factors affecting outcome after structural failure of repaired rotator cuff tears. *J Bone Joint Surg Am* 2014;96:99-105.
43. Fabricant PD, Rosenberger PH, Jokl P, Ickovics JR. Predictors of short-term recovery differ from those of long-term outcome after arthroscopic partial meniscectomy. *Arthroscopy* 2008;24:769-778.
44. Rosenberger PH, Dhabhar FS, Epel E, Jokl P, Ickovics JR. Sex differences in factors influencing recovery from arthroscopic knee surgery. *Clin Orthop Relat Res* 2010;468:3399-3405.
45. Bram JT, Talathi NS, Patel NM, DeFrancesco CJ, Striano BM, Ganley TJ. How do race and insurance status affect the care of pediatric anterior cruciate ligament injuries? [published online November 28, 2018]. *Clin J Sport Med*, https://doi.org/10.1097/JSM.0000000000000706.
46. Kraeutler MJ, McCarty EC, Belk JW, et al. Descriptive epidemiology of the MOON shoulder instability cohort. *Am J Sports Med* 2018;46:1064-1069.
47. Skinner J, Weinstein JN, Sporer SM, Wennberg JE. Racial, ethnic, and geographic disparities in rates of knee arthroplasty among Medicare patients. *N Engl J Med* 2003;349:1350-1359.
48. Magnussen RA, Granan LP, Dunn WR, et al. Cross-cultural comparison of patients undergoing ACL reconstruction in the United States and Norway. *Knee Surg Sports Traumatol Arthrosc* 2010;18:98-105.
49. Harris JD, Brand JC, Cote MP, Faucett SC, Dhawan A. Research pearls: The significance of statistics and perils of pooling. Part 1: Clinical versus statistical significance. *Arthroscopy* 2017;33:1102-1112.