

Published in final edited form as:

Ann Thorac Surg. 2013 January ; 95(1): 197–203. doi:10.1016/j.athoracsur.2012.08.074.

Differential Case Ascertainment in Clinical Registry vs. Administrative Data and Impact on Outcomes Assessment in Pediatric Heart Surgery

Sara K. Pasquali, MD, MHS^{1,3,*}, Eric D. Peterson, MD, MPH^{2,3}, Jeffrey P. Jacobs, MD⁴, Xia He, MS³, Jennifer S. Li, MD, MHS^{1,3}, Marshall L. Jacobs, MD⁵, J. William Gaynor, MD⁶, Jennifer C. Hirsch, MD, MS⁷, Samir S. Shah, MD, MSCE⁸, and John E. Mayer, MD⁹

¹Department of Pediatrics, Duke University School of Medicine, Durham, NC

²Department of Medicine, Duke University School of Medicine, Durham, NC

³Duke Clinical Research Institute, Durham, NC

⁴Department of Surgery, University of South Florida, All Children's Hospital, St. Petersburg FL

⁵Department of Pediatric and Congenital Heart Surgery, Cleveland Clinic, Cleveland, OH

⁶Department of Surgery, Children's Hospital of Philadelphia, Philadelphia, PA

⁷Department of Cardiac Surgery, University of Michigan Medical School, Ann Arbor, MI

⁸Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH

⁹Department of Cardiovascular Surgery, Children's Hospital Boston, Boston, MA

Abstract

Background—Administrative datasets are often used to assess outcomes and quality in pediatric heart surgery; however their accuracy regarding case ascertainment is unclear. We linked patient data (2004–2010) from the STS Congenital Heart Surgery Database (clinical registry), and Pediatric Health Information Systems Database (administrative database) from hospitals participating in both to evaluate differential coding/classification of operations between datasets, and subsequent impact on outcomes assessment.

Methods—Eight individual benchmark operations and the RACHS-1 categories were evaluated. The primary outcome was in-hospital mortality.

Results—The cohort included 59,820 patients (33 centers). There was a >10% difference in the number of patients identified between datasources for half of the benchmark operations. Negative predictive value of the administrative (vs. clinical) data was high (98.8–99.9%); positive predictive value was lower (56.7–88.0%). Overall agreement between datasources in RACHS-1 category assignment was 68.4%. These differences translated into significant differences in outcomes assessment, ranging from an underestimation of mortality associated with truncus arteriosus repair by 25.7% in the administrative vs. clinical data (7.01% vs. 9.43%, $p=0.001$), to an

© 2012 The Society of Thoracic Surgeons. Published by Elsevier Inc. All rights reserved.

Corresponding author: Sara Pasquali MD, C.S. Mott Children's Hospital, 1540 E. Hospital Drive, Ann Arbor, MI 48105, Phone: 734-232-8594, fax: 734-936-9470, pasqualisk@gmail.com.

*Dr. Pasquali's current affiliation: Department of Pediatrics, University of Michigan Medical School, Ann Arbor, MI

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

overestimation of mortality associated with VSD repair by 31.0% (0.78% vs. 0.60%, $p=0.1$). For the RACHS-1 categories, these ranged from an underestimation of Category 5 mortality by 40.5%, to an overestimation of Category 2 mortality by 12.1%; these differences were not statistically significant.

Conclusions—This study demonstrates differences in case ascertainment between administrative and clinical registry data for children undergoing heart surgery, which translated into important differences in outcomes assessment.

Keywords

congenital heart disease; health policy; outcomes

Introduction

Administrative datasets are often used to assess outcomes and quality of care for adults and children undergoing heart surgery (1–4). Current guidelines from the Agency for Healthcare Research and Quality (AHRQ), and National Quality Forum (NQF) advocate comparison of outcomes and case volume across hospitals performing congenital heart surgery using administrative data (2,3). These datasets are widely available, and utilize data already being collected for hospital billing purposes. Diagnoses and procedures are captured from the International Classification of Diseases, 9th Revision (ICD-9) codes listed on the hospital bill. In pediatric heart surgery, the use of administrative datasets poses challenges because ICD-9 codes do not address the full spectrum of congenital heart defects and operations. For example, there is no ICD-9 code for the Norwood operation. In addition, coding professionals may have limited knowledge regarding the complexities of congenital heart defects and operations, and typically have little interaction with physician care-givers. Accurate assessment of cases and associated outcomes is important to patients, physicians, payers, and policy makers, particularly in this era of public reporting and “pay for performance.” Recently, several clinical registries have been developed by surgical and medical subspecialty societies amid concerns regarding coding of diagnoses and procedures, adjustment for case-mix and patient risk, and capture of all relevant cases within administrative datasources (1,5).

Several previous studies have suggested differences in coding and classification of operations between administrative and clinical data in children with heart disease, but have been limited by small sample size and limited spectrum of congenital heart defects and procedures examined (4,6–8). In adult cardiac surgery, it has been shown that differences in coding of procedures between administrative and clinical datasets have led to differences in reported outcomes associated with these procedures (4). To date, a similar evaluation has not been carried out in the pediatric cardiac population.

The purpose of this study was to utilize linked data from the Society of Thoracic Surgeons Congenital Heart Surgery Database (STS-CHS; a clinical registry), and the Pediatric Health Information Systems Database (PHIS; an administrative database) from hospitals participating in both databases to compare case ascertainment between datasources, and the subsequent impact of any miscoding or misclassification on outcomes assessment for children undergoing heart surgery.

Patients and Methods

Data Source

Data on 62,052 (90%) eligible patients 0–18 years undergoing heart surgery (with or without cardiopulmonary bypass) at 33 hospitals participating in both the STS-CHS and PHIS Databases from 2004–2010 were linked at the individual patient level using the method of “indirect identifiers”, as previously described (9–11). The linked dataset contains information entered into both the PHIS and STS-CHS Databases for each patient, and ensures that any differences identified cannot be explained by the individual databases containing information on different patients.

The PHIS Database is a large administrative database that collects demographic information, ICD-9 diagnosis and procedure codes, in-hospital outcomes, and resource utilization data from 41 US children’s hospitals. The STS-CHS Database is the largest existing pediatric heart surgery registry, and collects pre-operative, operative, and outcomes data on all children undergoing heart surgery at >100 participating centers. Diagnoses and procedures are coded by clinicians and affiliated data managers using the International Pediatric and Congenital Cardiac Code (IPCCC) (12). This research was not considered human subjects research by the Duke Institutional Review Board in accordance with the Common Rule (45 CFR 46.102(f)).

Study Population

From the linked dataset, patients with missing (n=62 STS-CHS, and n=1567 PHIS) or discrepant (n=103) in-hospital mortality status, or discharge date (n=500) between databases were excluded. These exclusions were applied to eliminate the possibility that any differences in outcome identified might be related to differences in coding of outcomes themselves between the databases, rather than differences in coding/classification of cases. Of note, overall there was 99.83% agreement between databases for the mortality outcome.

The operations evaluated included eight previously described benchmark operations of varying levels of complexity: ventricular septal defect (VSD) repair, tetralogy of Fallot (TOF) repair (excluding those with pulmonary atresia), complete atrioventricular canal (CAVC) repair, arterial switch operation (ASO), arterial switch operation and ventricular septal defect (ASO+VSD) repair, Fontan operation (any type), truncus arteriosus repair, and Norwood operation (13). In the STS-CHS (clinical registry) data, the operation of interest was identified through assessment of the operation coded as the primary procedure for the index (first) operation of the admission (14). In the PHIS data, the two most commonly utilized methods to identify procedures in administrative datasets were evaluated. In Method 1, the individual ICD-9 procedure code for the operation of interest was used. In Method 2, the Risk Adjustment in Congenital Heart Surgery, version 1 (RACHS-1) methodology was used (15). As previously described, this method is based primarily on the type of procedure but also employs combinations of inclusionary and exclusionary ICD-9 diagnosis and procedure codes with the aim of more precisely identifying the procedure of interest; and then subsequently classifies the procedures into categories based on mortality risk (15). For Method 2, additional inclusion/exclusion codes on the STS-CHS side were applied for TOF and CAVC repair to attempt to match the RACHS-1 algorithms as closely as possible.

In addition to individual operations, we evaluated categories of operations, as it is possible that grouping of operations of similar risk may potentially mitigate miscoding of individual operations. For this portion of the analysis, we identified and grouped operations by RACHS-1 category both in the administrative and clinical data (15). RACHS-1 was the only risk stratification system evaluated because it is the only methodology that has been adapted for use with both types of data.

Outcomes

The primary outcome was in-hospital mortality. As described above, only patients with concordant mortality status between datasets were included.

Analysis

The operation performed for each patient was ascertained: 1) based on the data coded in the administrative dataset, and 2) based on the same patient's data as coded in the clinical registry. The clinical registry data was used as the reference based on previous studies suggesting greater accuracy of clinical vs. administrative data (6–8). Thus, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the administrative (vs. clinical) data were calculated. PPV is the probability that a patient with a certain operation coded in the administrative dataset had that operation performed as assessed in the clinical registry data. NPV is the probability that a patient who did not have a specific operation coded in the administrative dataset, did not have that operation performed as assessed in the clinical registry data.

Subanalysis was performed to evaluate the sensitivity of our initial results to methodology used in the STS-CHS (clinical) data. In the STS-CHS data, the operation of interest was initially identified through assessment of the index (first) operation of the admission since this is the standard STS-CHS Database methodology (14). Because the methodology in the administrative data may not necessarily specify the first operation of the admission (in cases where there are multiple operations), in the subanalysis we instead considered any operation during the admission in the STS-CHS data and evaluated whether this impacted our results regarding case ascertainment.

For the operations identified in either datasource, in-hospital mortality rates were calculated. In order to take into account overlap between groups, binomial models with group indicators for patients identified in either or both datasources were constructed. Wald tests were used to compare mean values. All analyses were performed using SAS version 9.2 (SAS Institute Inc., Cary, NC). A p-value <0.05 was considered statistically significant.

Results

The overall cohort included 59,820 patients from 33 centers. Median age at surgery was 6.4 months (interquartile range 32.0 days–3.4 years); 55% were male. Included centers were diverse geographically (33.3% South, 33.3% Midwest, 21.2% West, 12.2% Northeast) with a wide range of average annual volume (median 351, range 111–891 cardiovascular cases/year).

Case Ascertainment

There were differences in case ascertainment between administrative and clinical registry data for all operations examined (Table 1). Using Method 1 (individual ICD-9 procedure codes in the administrative data), differences between datasources ranged from 11.9% more Fontan operations to 104.4% more VSD repairs identified in the administrative data relative to the clinical registry. Across operations, the administrative data (compared to the clinical registry data) had a high NPV (range 99.3%–100.0%), but lower PPV (range 45.4%–84.2%). Thus, it is highly likely that a patient without a certain operation coded in the administrative data truly did not have that operation performed as assessed in the clinical registry data (high NPV). Conversely, the relatively lower PPV indicates that many of the patients coded as having a certain operation performed in the administrative data are false positives, such that these patients did not have this operation performed as assessed in the clinical registry.

Method 2 (involving combinations of ICD-9 diagnosis and procedures codes to attempt to more precisely identify the operation of interest in the administrative data) was associated with smaller differences between datasources in the number of operations identified, compared with Method 1 (Table 1). However, even with Method 2, there was still a >10% difference between datasources in the number of operations identified for half of the benchmark operations (Table 1). While Method 2 was associated with higher PPV for the administrative data compared with Method 1, it remained below 80% for 6 of the 8 benchmark operations. Similar to Method 1, NPV was high for Method 2.

In subanalysis, we found similar results to those described above when we included any operation during the admission (in cases where there were multiple operations) in the STS-CHS data in the analysis, compared with the initial approach where the first (index) operation of the admission was considered. Specifically, the differences (comparing the results of the subanalysis to the main analysis for Method 2) across operations in PPV ranged from 0.1–1.4%, and NPV ranged from 0–0.2%.

Case ascertainment for groups of procedures of similar risk (the RACHS-1 categories; Table 2) was also evaluated. A greater number of operations were unable to be classified in the RACHS-1 system in the administrative vs. clinical registry data (n= 10,953 vs. 10,047, 9.0% difference). Across the RACHS-1 categories, differences between datasources ranged from 18.8% fewer operations categorized in Category 6 in the administrative data relative to the clinical registry, to 135.5% more operations categorized in Category 5. The smallest difference between datasources was for RACHS-1 Category 4. NPV (range 87.3%–100.0%) of the administrative data was high across all RACHS-1 categories, while PPV (range 23.3%–82.7%) was lower. PPV was <80% for 5 of the 6 RACHS-1 categories. Overall, the percent agreement between the administrative and clinical registry data regarding RACHS-1 category was 68.4%. Examination of percent agreement as a function of hospital volume and case mix did not reveal any apparent associations (Figures 1,2).

Impact on Outcomes Assessment

In-hospital mortality rates were subsequently calculated for the operations identified in the administrative and clinical registry data. As described above, only patients with concordant mortality status between datasets were included in order to eliminate the possibility that any differences in outcome identified might be related to differences in coding of outcomes themselves, rather than differences in coding/classification of the operations. Using Method 1, there were significant differences in the in-hospital mortality rates associated with the benchmark operations identified in the administrative vs. clinical registry data (Table 3). Using Method 2, these differences in mortality were smaller, yet ranged from an underestimation of mortality associated with truncus arteriosus repair of 25.7% in the administrative data relative to the clinical registry, to an overestimation of mortality associated with VSD repair of 31.0%. Given the relatively low mortality rates for most operations, the absolute differences in mortality rates between datasources were small in most cases (<1%), except for truncus arteriosus repair and the Norwood operation.

For the RACHS-1 categories (Table 4), the differences in in-hospital mortality rates ranged from an underestimation of mortality for RACHS-1 category 5 by 40.5% in the administrative data relative to the clinical registry, to an overestimation of mortality for RACHS-1 category 2 by 12.1%. The smallest difference was observed for RACHS-1 category 3. None of these differences were statistically significant.

Comment

This study demonstrates important differences in case ascertainment between administrative and clinical registry data for children undergoing heart surgery, which translated into significant differences in outcomes assessment. Previous analyses of smaller cohorts have suggested differences in case ascertainment between datasources in this population. In one study, 4,918 records in the Metropolitan Atlanta Congenital Defects Program database of patients identified as having congenital heart disease based on ICD-9 codes were recoded following medical record review using IPCCC codes (the same system used in the STS-CHS Database) (6). The sensitivity of the surveillance data/ICD-9 codes for tetralogy of Fallot was 83%, 100% for transposition of the great arteries, and 95% for hypoplastic left heart syndrome; the false positive rates were 2%, 49%, and 11%, respectively. In another study, differences between 373 administrative records in Wisconsin's birth defects database, and diagnoses coded based on a review of the medical record were examined. Only 52% of cases had an exact match for diagnosis (7). Finally, a review of ICD-9 codes for children born at a Minnesota hospital in 2001 revealed that of those coded with cardiac defects (n=66), only 41% of codes accurately reflected the diagnosis upon medical record review (8).

The present study supports these findings in a broad population of 59,820 children undergoing heart surgery across 33 centers. For individual operations, our analysis suggests that Method 2 (which is based on the RACHS-1 methodology) improves the accuracy of case ascertainment compared with the use of individual ICD-9 procedure codes in Method 1. Based on these results, the use of individual ICD-9 procedure codes to identify the operation of interest in outcomes and quality analyses of congenital heart surgery is not advised. Although Method 2 was associated with improved accuracy, it is important to note that this method was still associated with a >10% difference in the number of operations identified between datasources for half of the benchmark operations.

We also found differences between datasets in ascertainment of groups of operations of similar risk (the RACHS-1 categories), suggesting that grouping of procedures into larger categories does not necessarily mitigate all potential miscoding/misclassification. More operations were unable to be classified in the RACHS-1 system in the administrative vs. clinical data. These and other differences in case ascertainment may impact analyses of hospital case volumes. Of those operations that could be classified, there were also differences in category assignment, with an overall percent agreement between the administrative and clinical registry data regarding RACHS-1 category of 68.4%. Our finding of consistent differences in case ascertainment across varying hospital surgical volume and case mix suggests that these differences may be more related to the limitations of the ICD-9 system and coding methodology itself rather than any particular risk stratification system or other hospital-specific factors. Further investigation will be required to evaluate whether coding changes in ICD-10 or 11 will translate into improved case ascertainment (16).

Previous pediatric studies did not evaluate the impact of differential coding of diagnoses and procedures on outcomes assessment. In adult cardiac surgery, miscoding of procedures in an administrative dataset was found to result in an overestimation of mortality associated with isolated coronary artery bypass surgery relative to a clinical dataset of 40% (4). The methodology used in the present study ensures that any differences in outcome are related to differences in case ascertainment/coding, rather than the patient population included in either dataset or how outcomes were coded (17). We found that differences in case ascertainment between datasets translated into significant differences in in-hospital mortality rates. This was partially mitigated by evaluating groups of procedures of similar risk as opposed to individual procedures; however, clinically important differences in outcome remained in some cases even with this approach.

These data may have implications for the use of administrative data for outcomes assessment in this population, including evaluation of case volumes, outcomes, and risk stratification. Further evaluation of how these findings impact assessment and ranking of performance on a hospital level is necessary, given that current AHRQ and NQF guidelines recommend using administrative data to compare outcomes and case volume across hospitals performing congenital heart surgery (2,3).

Finally, it is important to recognize that administrative datasets contain valuable resource utilization data, medication data, and information on other non-cardiac diagnoses/procedures that may not be collected in disease or procedure-specific registries. Linking clinical and administrative databases together can capitalize on the strengths and mitigate some of the weaknesses of these different datasources, and allow analyses not possible with either dataset alone (9–11).

Limitations

This study is subject to several limitations. While we chose to use the clinical registry as the reference based on previous studies, it is possible that different results may be obtained through manual abstraction of the medical record as the “gold standard” to assign diagnoses and procedures. However, the similarities of our findings with previous studies using medical record review suggest this is unlikely (6–8). In addition, the consistent results between our subanalysis and the main results suggest that the differences in case ascertainment observed between datasources are not related to the STS-CHS methodology regarding evaluation of the first (index) operation. We have previously shown that our methodology to link information between datasources is successful for the vast majority of patients (>90%); however, we were unable to link data on every patient at every center. Finally, this analysis is based on data from the 33 included hospitals. Although differences in case ascertainment in our study seemed to be consistent across these hospitals, it is possible that our results may not be generalizable to all US centers.

Conclusions

This study demonstrates differences in case ascertainment between administrative and clinical registry data for children undergoing heart surgery, which translated into important differences in outcomes assessment. Further study is needed to investigate the implications of our findings on evaluation and ranking of hospital performance.

Acknowledgments

We acknowledge Matt Hall, PhD for assistance with analysis and interpretation of PHIS data.

Funding Sources: National Heart, Lung, and Blood Institute (1K08HL103631, PI: Pasquali; 1RC1HL099941, co-PI's: J Jacobs, Li).

Dr. Shah: National Institute of Allergy and Infectious Diseases (K01AI73729), Robert Wood Johnson Foundation Physician Faculty Scholar program.

References

1. Likosky DS. Lessons learned from the Northern New England Cardiovascular Disease Study Group. *Prog Pediatr Cardiol.* 2012; 33:53–56.
2. Agency for Healthcare Research and Quality. [Accessed 3/22/2012] Pediatric Quality Indicators. Available at: http://www.qualityindicators.ahrq.gov/modules/PDI_TechSpec.aspx
3. National Quality Forum. [Accessed 3/22/2012] Pediatric Heart Surgery Mortality. Available at: <http://www.qualityforum.org/MeasureDetails.aspx?actid=0&SubmissionId=349#e=1&s=n&so=a&p=2&k=0339>

4. Shahian DM, Silverstein T, Lovett AF, et al. Comparison of clinical and administrative data sources for hospital coronary artery bypass graft surgery report cards. *Circulation*. 2007; 115:1518–1527. [PubMed: 17353447]
5. Jacobs ML, Jacobs JP, Franklin RC, et al. Databases for assessing the outcomes of the treatment of patients with congenital and pediatric cardiac disease-the perspective of cardiac surgery. *Cardiol Young*. 2008; 18:101–115. [PubMed: 19063780]
6. Strickland MJ, Riehle-Colarusso TJ, Jacobs JP, et al. The importance of nomenclature for congenital cardiac disease: implications for research and evaluation. *Cardiol Young*. 2008; 18:92–100. [PubMed: 19063779]
7. Cronk CE, Malloy ME, Pelech AN, et al. Completeness of state administrative databases for surveillance of congenital heart disease. *Birth Defects Res A Clin Mol Teratol*. 2003; 67:597–603. [PubMed: 14703780]
8. Frohnert BK, Lussky RC, Alms MA, et al. Validity of hospital discharge data for identifying infants with cardiac defects. *J Perinatol*. 2005; 25:737–742. [PubMed: 16163368]
9. Pasquali SK, Jacobs JP, Shook GJ, et al. Linking clinical registry data with administrative data using indirect identifiers: Implementation and validation in the congenital heart surgery population. *Am Heart J*. 2010; 160:1099–1104. [PubMed: 21146664]
10. Pasquali SK, Li JS, He X, et al. Comparative analysis of antifibrinolytic medications in pediatric heart surgery. *J Thorac Cardiovasc Surg*. 2012; 143:550–557. [PubMed: 22264414]
11. Pasquali SK, Li JS, He X, et al. Perioperative methylprednisolone and outcome in neonates undergoing heart surgery. *Pediatrics*. 2012; 129:e385–e391. [PubMed: 22271697]
12. Franklin RC, Jacobs JP, Krogmann ON, et al. Nomenclature for congenital and paediatric cardiac disease: historical perspectives and The International Pediatric and Congenital Cardiac Code. *Cardiol Young*. 2008; 18:70–80. [PubMed: 19063777]
13. Jacobs JP, O'Brien SM, Pasquali SK, et al. Variation in outcomes for benchmark operations. *Ann Thorac Surg*. 2011; 92:2184–2191. [PubMed: 22115229]
14. Jacobs, JP. Executive Summary: The STS Congenital Heart Surgery Database-Fourteenth Harvest. Durham, NC: The STS and Duke Clinical Research Institute;
15. Jenkins KJ, Gauvreau K, Newburger JW, et al. Consensus-based method for risk adjustment for surgery for congenital heart disease. *J Thorac Cardiovasc Surg*. 2002; 123:110–118. [PubMed: 11782764]
16. [Accessed 4/11/2012] HHS announces intent to delay ICD-10 compliance date. Available at: <http://www.hhs.gov/news/press/2012pres/02/20120216a.html>
17. Welke KF, Diggs BS, Karamlou T, Ungerleider RM. Comparison of pediatric cardiac surgical mortality rates from national administrative data to contemporary clinical standards. *Ann Thorac Surg*. 2009; 87:216–223. [PubMed: 19101301]

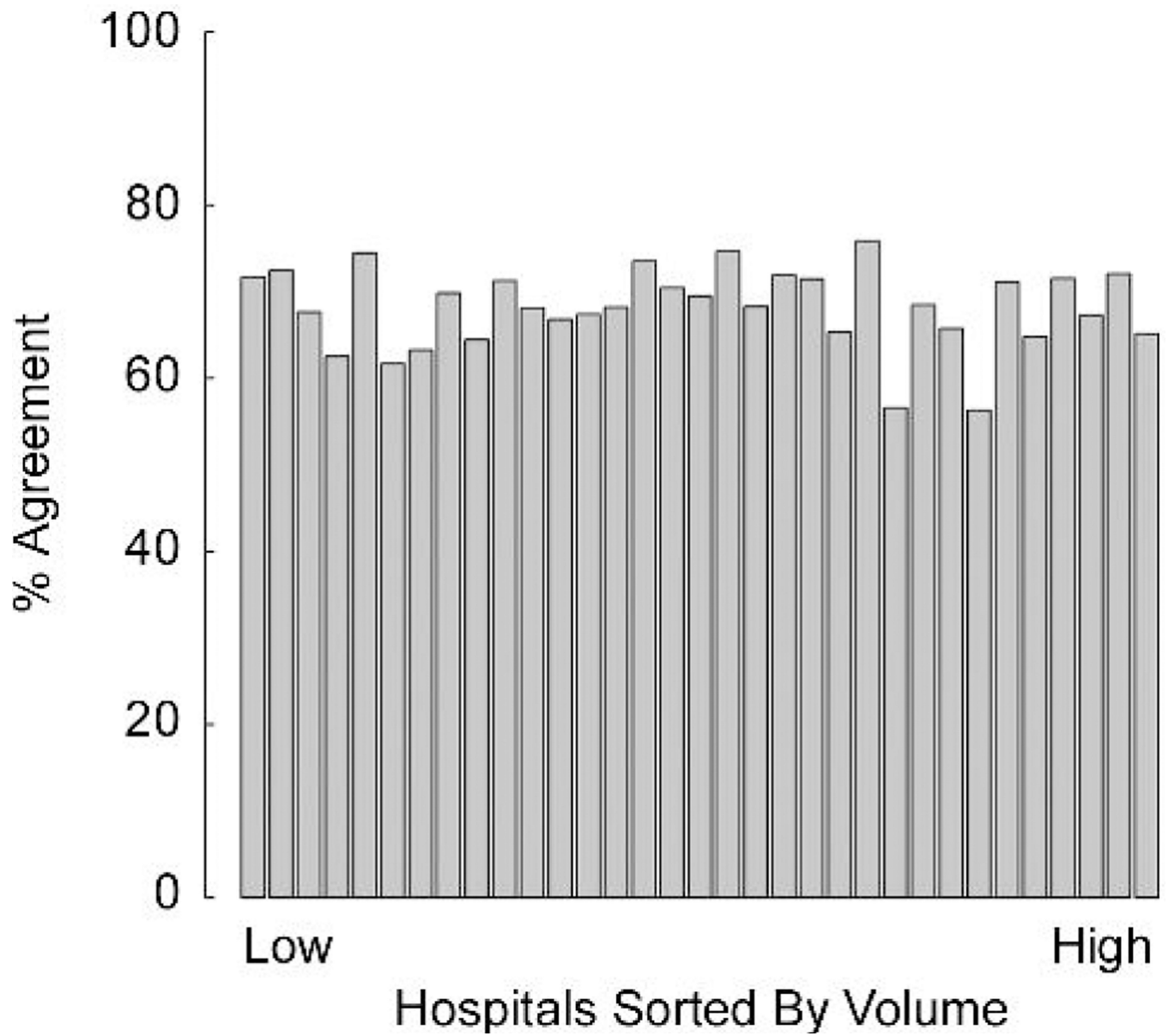


Figure 1. Overall percent agreement between administrative and clinical registry data for RACHS-1 category as a function of hospital volume.

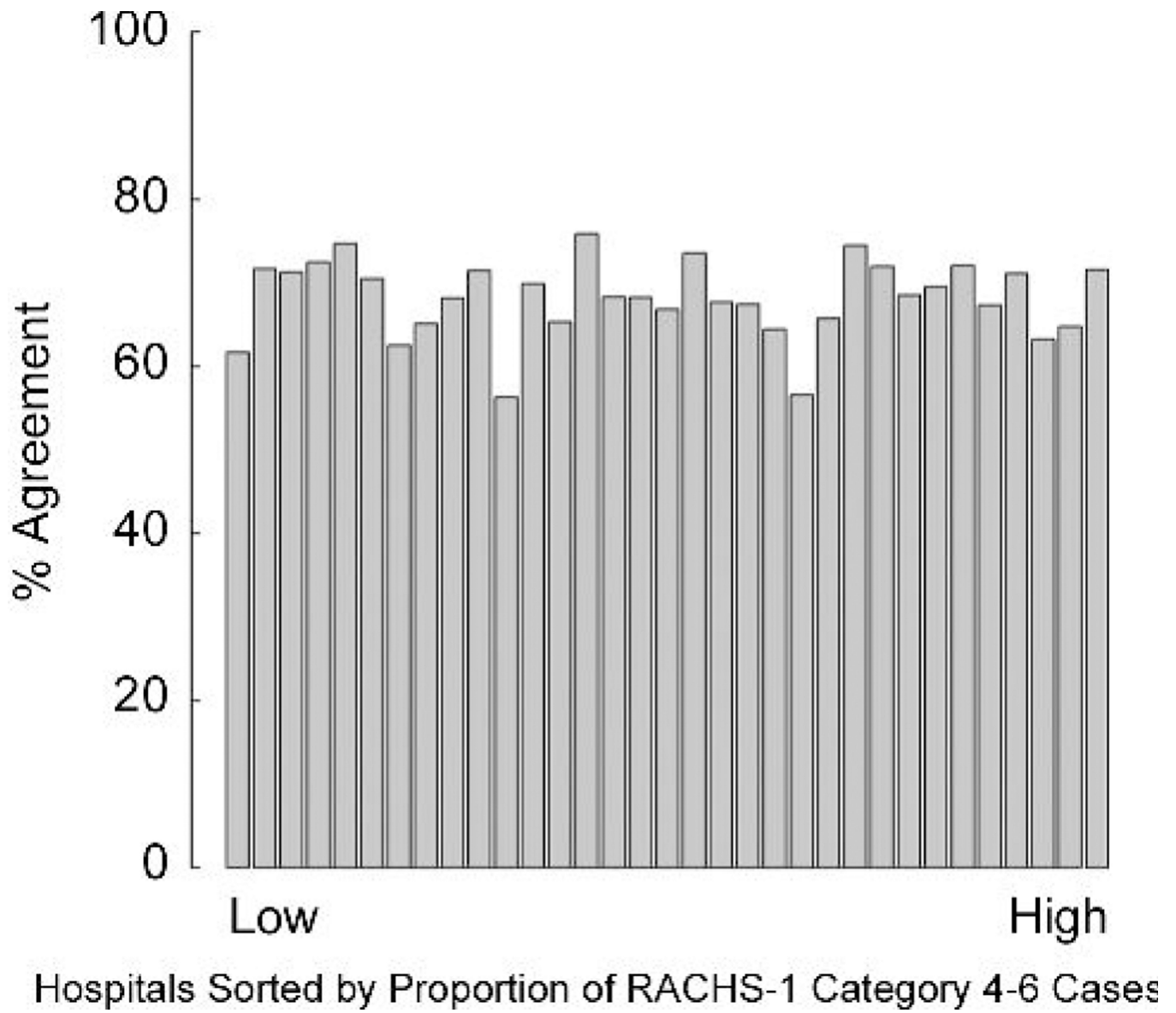


Figure 2. Overall percent agreement between administrative and clinical registry data for RACHS-1 category as a function of hospital case mix.

Table 1

Case ascertainment for benchmark operations in administrative vs. clinical registry data

Operation	No. of Operations Identified in Clinical Registry	No. of Operations Identified in Administrative Data	% Difference in No. of Operations	Sensitivity	Specificity	PPV	NPV
<i>Method 1 – Individual ICD-9 Procedure Codes</i>							
VSD Repair	5023	10266	+ 104.4%	92.7%	89.8%	45.4%	99.3%
TOF Repair	3138	3903	+ 24.4%	92.3%	98.2%	74.2%	99.6%
CAVC repair	2151	3676	+ 70.9%	92.2%	97.1%	54.0%	99.7%
ASO	1449	2415	+ 66.7%	97.5%	98.3%	58.5%	99.9%
ASO + VSD repair	575	829	+ 44.2%	82.6%	99.4%	57.3%	99.8%
Fontan	3116	3488	+ 11.9%	94.3%	99.0%	84.2%	99.7%
Truncus repair	456	602	+ 32.0%	95.4%	99.7%	72.3%	100.0%
Norwood	NA	NA	NA	NA	NA	NA	NA
<i>Method 2 – Additional Inclusionary/Exclusionary ICD-9 Diagnosis & Procedure Codes</i>							
VSD Repair	5023	5498	+ 9.5%	74.6%	96.8%	68.2%	97.7%
TOF Repair	2817	3338	+ 18.5%	85.5%	98.4%	72.2%	99.3%
CAVC repair	2079	3097	+ 49.0%	84.4%	97.7%	56.7%	99.4%
ASO	1449	1394	-3.8%	81.4%	99.6%	84.6%	99.5%
ASO + VSD repair	575	777	+ 35.1%	80.0%	99.5%	59.2%	99.8%
Fontan	3116	3055	-2.0%	86.3%	99.4%	88.0%	99.2%
Truncus repair	456	542	+ 18.9%	86.6%	99.8%	72.9%	99.9%
Norwood	2181	2024	-7.2%	68.5%	99.1%	73.8%	98.8%

Reference for comparisons: clinical registry data. NA=No data are listed for the Norwood operation for Method 1, as there is no individual ICD-9 procedure code for the Norwood operation.

Table 2

Case ascertainment for RACHS-1 categories in administrative vs. clinical registry data

RACHS-1 Category	No. of Operations Identified in Clinical Registry	No. of Operations Identified in Administrative Data	% Difference in No. of Operations	Sensitivity	Specificity	PPV	NPV
1	6519	5755	-11.7%	73.0%	98.1%	82.7%	96.7%
2	18581	17218	-7.3%	70.8%	90.2%	76.5%	87.3%
3	16510	17960	+8.8%	72.4%	86.1%	66.6%	89.1%
4	5636	5834	+3.5%	52.3%	94.7%	50.5%	95.0%
5	31	73	+135.5%	54.8%	99.9%	23.3%	100.0%
6	2496	2027	-18.8%	61.9%	99.2%	76.2%	98.4%

Reference for comparisons: clinical registry data.

Table 3

In-hospital mortality for benchmark operations in administrative vs. clinical registry data

Operation	Mortality for Operations Identified in Clinical Registry	Mortality for Operations Identified in Administrative Data	Absolute Difference in Mortality	Relative Difference in Mortality	P-Value
<i>Method 1 – Individual ICD-9 Procedure Codes</i>					
VSD Repair	0.60% (30/5023)	2.14% (220/10266)	+1.55%	+258.8%	<0.0001
TOF Repair	1.12% (35/3138)	1.82% (71/3903)	+0.70%	+63.1%	<0.0001
CAVC repair	2.00% (43/2151)	2.39% (88/3676)	+0.39%	+19.8%	0.09
ASO	1.66% (24/1449)	2.73% (66/2415)	+1.08%	+65.0%	<0.001
ASO + VSD repair	3.48% (20/575)	4.58% (38/829)	+1.11%	+31.8%	0.10
Fontan	1.41% (44/3116)	1.66% (58/3488)	+0.25%	+17.8%	0.03
Truncus repair	9.43% (43/456)	7.64% (46/602)	-1.79%	-19.0%	<0.01
Norwood	NA	NA	NA	NA	NA
<i>Method 2 – Additional Inclusionary/Exclusionary ICD-9 Diagnosis & Procedure Codes</i>					
VSD Repair	0.60% (30/5023)	0.78% (43/5498)	+0.2%	+31.0%	0.12
TOF Repair	1.06% (30/2817)	1.17% (39/3338)	+0.1%	+9.7%	0.57
CAVC repair	1.92% (40/2079)	1.91% (59/3097)	-0.0%	-1.0%	0.94
ASO	1.66% (24/1449)	1.43% (20/1394)	-0.2%	-13.4%	0.42
ASO+ VSD repair	3.48% (20/575)	4.12% (32/777)	+0.6%	+18.4%	0.36
Fontan	1.41% (44/3116)	1.31% (40/3055)	-0.1%	-7.3%	0.39
Truncus repair	9.43% (43/456)	7.01% (38/542)	-2.4%	-25.7%	0.001
Norwood	15.82% (345/2181)	14.08% (285/2024)	-1.7%	-11.0%	<0.01

Reference for comparisons: clinical registry data. NA=No data are listed for the Norwood operation for Method 1, as there is no individual ICD-9 procedure code for the Norwood operation.

Table 4

In-hospital mortality for RACHS-1 categories in administrative vs. clinical registry data

RACHS-1 Category	Mortality for Operations Identified in Clinical Registry	Mortality for Operations Identified in Administrative Data	Absolute Difference in Mortality	Relative Difference in Mortality	P-Value
1	1.06% (69/6519)	0.92% (53/5755)	-0.14%	-13.0%	0.34
2	0.98% (182/18581)	1.10% (189/17218)	+0.12%	+12.1%	0.12
3	2.94% (486/16510)	2.91% (522/17960)	-0.04%	-1.3%	0.76
4	6.41% (361/5636)	6.26% (365/5834)	-0.15%	-2.3%	0.60
5	32.26% (10/31)	19.18% (14/73)	-13.1%	-40.5%	0.09
6	15.18% (379/2496)	14.16% (287/2027)	-1.03%	-6.8%	0.08

Reference for comparisons: clinical registry data.