Journal of Pediatric Surgery 57 (2022) 9–16

Contents lists available at ScienceDirect

Journal of Pediatric Surgery

journal homepage: www.elsevier.com/locate/jpedsurg.org

The influence of decreasing variable collection burden on hospital-level risk-adjustment

Andrew Hu[a,*], Marie Iwaniuk[b], Vanessa Thompson[b], Catherine Grant[b], Alaina Matthews[b], Claudia Byrd[b], Jacqueline Saito[c], Bruce Hall[b,d], Mehul V. Raval[a]

[a] Division of Pediatric Surgery, Department of Surgery, Northwestern University Feinberg School of Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, 633 N. Saint Clair St, 20th Floor, Chicago, IL 60011, USA
[b] Division of Research and Optimal Patient Care, American College of Surgeons, Chicago, IL, USA
[c] Division of Pediatric Surgery, Department of Surgery, Washington University School of Medicine, St. Louis, MO, USA
[d] Department of Surgery, Washington University School of Medicine, and BJC Healthcare, St. Louis, MO, USA

ARTICLE INFO

Article history:
Received 24 July 2021
Revised 30 September 2021
Accepted 4 October 2021

Keywords:
Risk-adjustment
Quality improvement
NSQIP pediatric
Congenital malformation
Variable elimination
Outcomes
Pediatric surgery

ABSTRACT

Background: Risk-adjustment is a key feature of the American College of Surgeons National Surgical Quality Improvement Program-Pediatric (NSQIP-Ped). Risk-adjusted model variables require meticulous collection and periodic assessment. This study presents a method for eliminating superfluous variables using the congenital malformation (CM) predictor variable as an example.

Methods: This retrospective cohort study used NSQIP-Ped data from January 1st to December 31st, 2019 from 141 hospitals to compare six risk-adjusted mortality and morbidity outcome models with and without CM as a predictor. Model performance was compared using C-index and Hosmer-Lemeshow (HL) statistics. Hospital-level performance was assessed by comparing changes in outlier statuses, adjusted quartile ranks, and overall hospital performance statuses between models with and without CM inclusion. Lastly, Pearson correlation analysis was performed on log-transformed ORs between models.

Results: Model performance was similar with removal of CM as a predictor. The difference between C-index statistics was minimal (≤ 0.002). Graphical representations of model HL-statistics with and without CM showed considerable overlap and only one model attained significance, indicating minimally decreased performance ($P = 0.058$ with CM; $P = 0.044$ without CM). Regarding hospital-level performance, minimal changes in the number and list of hospitals assigned to each outlier status, adjusted quartile rank, and overall hospital performance status were observed when CM was removed. Strong correlation between log-transformed ORs was observed ($r \geq 0.993$).

Conclusions: Removal of CM from NSQIP-Ped has minimal effect on risk-adjusted outcome modelling. Similar efforts may help balance optimal data collection burdens without sacrificing highly valued risk-adjustment in the future.

Level of evidence: Level II prognosis study.

1. Introduction

Since 2008, the American College of Surgeons (ACS) National Surgical Quality Improvement Program-Pediatric (NSQIP-Ped) has served as the premiere resource for risk-adjusted patient- and hospital-level pediatric surgical outcomes [1–4]. This multi-center, international clinical registry has established preoperative risk assessment tools and launched multiple successful quality improvement initiatives [5–9]. The program continues to evolve by sampling high-risk procedures, developing targeted data collection for specific cases, and honing its risk-adjusted modeling [10,11]. The validity of NSQIP-Ped is contingent upon a rigorous chart review process prospectively collecting over 120 precisely specified variables across 6 surgical specialties from nearly 150 hospitals. This significant workload falls upon trained data collection personnel known as Surgical Clinical Reviewers (SCRs) who each collect data from approximately 1400 cases annually [12]. The cost of training and employing these abstractors can be significant with some institutions employing multiple SCRs. As NSQIP-Ped continues to adapt and expand to reflect nuances and changes within children's surgery, the data collection burden faced by SCRs must be

Abbreviations: American college of surgeons, (ACS); Anesthesiologist physical status classification system, (ASA); Confidence interval, (CI); Concordance statistic, (C-index); Congenital malformation, (CM); Hosmer-lemeshow, (HL); National Surgical Quality Improvement Program-Pediatric, (NSQIP-Ped); Odd ratios, (ORs); Operating room, (OR); Pearson's correlation coefficient, (PCC); Procedure targeted cases, (PTC); Surgical clinical reviewer, (SCR); Surgical site infection, (SSI).

* Corresponding author.
E-mail address: andrew.hu1@northwestern.edu (A. Hu).

https://doi.org/10.1016/j.jpedsurg.2021.10.007
0022-3468/

assessed. Existing variables with significant collection burden and with limited influence on NSQIP-Ped outcome models should be eliminated to ensure manageable SCR workload.

The presence of congenital malformations (CM) is a known risk-factor for children undergoing surgery [13–16]. Within NSQIP-Ped, CM has been captured using a categorical variable designed to identify patients with a structural, functional, or genetic abnormality present at birth or diagnosed before the age of 4 years. SCRs are instructed to review a list of CMs based on administrative diagnosis codes for inclusion. This broad list encompasses a plethora of rare, complex, and often eponymous conditions with significant burden to SCRs [17]. Given the comprehensive nature of this composite variable and the varying risk associated with each of its included conditions, its effect on risk-adjustment is unclear.

The purpose of this study was to illustrate an approach to variable assessment and potential elimination from NSQIP-Ped risk-adjusted, hospital-level modeling using CM as an example.

2. Materials and methods

2.1. Data source and study population

This study was a retrospective cohort study which reviewed the NSQIP-Ped database for eligible cases with operation dates between January 1 and December 31, 2019. The NSQIP-Ped clinical registry prospectively collects data from 141 participating hospitals on patients less than 18 years of age undergoing general, neurosurgical, urologic, otolaryngologic, plastic, and orthopedic procedures. Cases were chosen by systematic sampling on an 8-day cycle from a defined list of procedures with a maximum threshold of 35 cases selected per cycle. In addition to systematic sampling, NSQIP-Ped also collected additional data for Procedure Targeted Cases (PTC) such as appendectomy and cerebrospinal fluid shunt procedures. Over 150 patient-level and procedure-specific data elements were abstracted by SCRs from the electronic medical record and with follow-up phone calls to ensure capture of 30-day outcomes including mortality and morbidity. Data integrity was ensured by audits excluding sites that have greater than 5% disagreement rates from semi-annual reports used for hospital-level comparison. This study was deemed not human subjects research by the Ann & Robert H. Lurie Children's Hospital Institutional Review Board.

2.2. Predictor variable definitions

The primary predictor variable of interest in this study was the categorical CM variable which codes for the presence of a structural, functional, or genetic abnormality present at birth or acquired before 4 years of life. A provided "Collect List" of malformations is supplied to SCRs by NSQIP-Ped (Appendix A). The CM variable is binary, such that the presence of any of the conditions in the "Collect List" would result in the patient being labelled as having a CM. Some conditions captured by CM overlap with other patient-level predictor variables such as cardiac risk factors, esophageal/gastric/intestinal disease, structural pulmonary/airway abnormalities, developmental delay, neuromuscular disorder, structural central nervous system abnormality, cerebral palsy, hematologic disorder, and seizure disorder. The definitions for these overlapping variables are provided in Appendix B.

2.3. Risk-adjustment models

Risk-adjustment models examined different combinations of patient-level and procedure-related predictors and were computed for mortality and morbidity outcomes ($n = 112$). Eligible predictor variables ($n = 40$) were selected and included in logistic regression models using a forward stepwise selection method. Predictor variables were then used in a hierarchical model and odds ratios (ORs) were calculated for each participating NSQIP-Ped hospital. These ORs represent a ratio of the odds that an outcome would occur at the specified-hospital compared to the odds of the same outcome occurring at an estimated "average" NSQIP-Ped hospital, assuming a similar case-mix of patients and procedures.

2.4. Hospital quality improvement performance assessments

Each hospital was assigned an overall performance assessment status to determine areas for potential quality improvement. These were determined using the outlier status and adjusted quartile rank assigned to each hospital. Outlier status was defined by the 95% Confidence Interval (95% CI) of the ORs such that ORs with a 95% CI entirely above 1.0 were assigned as "High Outliers" whereas ORs with a 95% CI entirely below 1.0 were assigned as "Low Outliers". ORs with a 95% CI that overlap 1.0 were not assigned an outlier status. In addition to outlier status, hospitals were also ranked via adjusted quartiles, which were based on percentile ranks of adjusted ORs. Performance assessment was defined such that a hospital that was assigned either a "Low Outlier" status or a 1st adjusted quartile rank was assigned an "Exemplary" status whereas a hospital assigned a "High Outlier" status or a 4th adjusted quartile was assigned a "Needs Improvement" status. Hospitals that were not assigned an outlier status and were in the 2nd or 3rd adjusted quartile ranks were assigned an "As Expected" status for the hospital performance assessment.

2.5. Selection of risk-adjusted models

Of 112 outcome models assessed, approximately 47 models utilized CM as a patient-level predictor. Six models that contained CM as a significant patient-level predictor were identified for further analysis: (1) morbidity in all patients in all surgeries, (2) return to operating room (OR) in all patients in all surgeries, (3) sepsis in all patients in all surgeries, (4) morbidity in all neonates in all surgeries (5) morbidity in all patients in urology surgeries, (6) surgical site infection (SSI) in pediatric patients in abdominal surgeries. These models were chosen as CM was ranked higher in the stepwise logistic regression compared to other models, they allowed for distinction of rates between specific patient populations (*e.g.*, neonates, urologic patients), and based on clinical guidance.

2.6. Model comparison

To assess the effect of CM on risk adjustment, two models were generated for each of the six aforementioned outcomes that found CM to be a significant patient-level predictor. We evaluated each set of models (with CM; without CM) based on the following five statistical criteria: (1) concordance statistics, (2) Hosmer-Lemeshow (HL) goodness of fit statistics, (3) comparison of hospital-level outlier status and adjusted quartile rank based on ORs, (4) comparison of hospital-level Performance Assessments, and (5) Pearson correlation analysis of log-transformed ORs. Each assessment has been described in greater detail below.

The concordance statistic (or C-index) is a measure of distinction. In this case, it is a measure of the model's ability to accurately distinguish a case with an outcome event from a case without an outcome event. Values range from 0.5 to 1.0, with 0.5 indicating that a model's performance is not better than random chance and 1.0 indicating perfect prediction. Values that are > 0.8 indicate an effective model [18].

HL statistics are used in logistic regression models to determine if observed event rates match the expected event rates of

Table 1
Descriptive statistics for each risk-adjusted outcome assessed from The American College of Surgeons National Surgical Quality Improvement Project-Pediatric.

Risk-Adjusted Model	Hospitals (*n*)	Patient Records (*n*)	Events (*n* [%])
Morbidity in all patients in all surgeries	141	13,2881	6460 [4.86%]
Morbidity in neonate patients in all surgeries	133	9,146	1194 [13.05%]
Return to OR[a] in all patients in all surgeries	141	132,881	4318 [3.25%]
SSI[b] in pediatric patients in abdominal surgeries	136	47,405	1750 [3.69%]
Morbidity in all patients in urology surgeries	137	14,245	590 [4.14%]
Sepsis in all patients in all surgeries	141	132,881	973 [0.73%]

[a] Operating room
[b] Surgical Site Infection

subgroups within the population. The null hypothesis of the HL-test is that observed event rates are the same as the expected rate whereas the alternative hypothesis is that these two event rates differ. A nonsignificant HL statistic reflects good calibration of the outcomes model. One limitation, especially in larger data sets, is that small inconsequential divergences may often achieve statistical significance without significant associated clinical utility [18,19]. To avoid issues related to this limitation, we reviewed the graphical representation of the HL-statistics using 20 sequential risk category groups for each outcome model with CM included and excluded as a predictor in the model.

Hospital-level quality performance was estimated by each included model using a hierarchical modeling approach (random intercept, fixed slope) with hospital being the random factor. This produces an estimate of hospital performance in terms of a risk-adjusted OR, which is the odds of an event occurring at a particular hospital compared to the odds of the event occurring at a statistically estimated average hospital, given the same procedure and case-mix. Within each model, an OR was calculated for each hospital and hospital performance was determined using parameters associated with the risk-adjusted OR. As previously described, each hospital was assigned an outlier status ("High Outlier" vs. "Low Outlier") based on the 95% CI, an adjusted quartile rank, and Performance Assessment Status ("Needs Improvement," "Exemplary," or "As Expected"). Outlier status and adjusted quartile rank was assessed by comparing the number and cohort of hospitals included within each outlier status category and quartile rank for each model (with or without CM) within each of the six outcomes. The average change in adjusted percentile and adjusted quartile was calculated. For each outcome model, Performance Assessment Status was evaluated by determining the change in the number and cohort of hospitals within the "Needs Improvement" and "Exemplary" categories.

Lastly, Pearson correlation analyses were conducted on hospitals' log-transformed ORs between each set of models (with and without CM) for each of the six aforementioned outcomes to determine if the OR results from each set of models had good agreement. Prior to correlation analysis, ORs were log-transformed to normalize the distribution of ORs. Therefore, Pearson's correlation coefficient (PCC; r) was calculated between the log-transformed hospital ORs from a model with CM included as a predictor and the log-transformed hospital ORs from the corresponding model without CM included as a predictor. As a measure of linear association between continuous variables, PCC values range from -1.0 to +1.0 with values close to -1 or +1 indicating a strong negative or positive correlation, respectively [20]. Significance was declared at $P < 0.05$ for these analyses.

3. Results

The descriptive statistics including the number of hospitals, number of records used, and number of events recorded for each of the six outcomes are presented in Table 1.

3.1. Model performance

For all six outcomes, model performance statistics were similar between models regardless of inclusion of CM (Table 2). The largest difference between C-index statistics for a set of models with and without CM was 0.002 for the "SSI in pediatric patients in abdominal surgeries" outcome. The remaining five outcomes had smaller C-index differences (≤ 0.001) between models with and without CM. Thus, the risk adjusted models for each outcome, regardless of CM inclusion, were able to discriminate events at a similar probability level.

HL-test statistics were similar for each set of models. For four of the six outcomes, removing CM as a predictor from the risk-adjusted model did not alter the significance of the HL-test (Table 2). For the "Return to OR in all patients in all surgeries" outcome, the HL-test was significant when CM was removed from the risk-adjusted model. While this shift may indicate a decrease in model performance, it is important to note that the shift in significance of the HL-test was minimal ($P = 0.058$ With CM vs. $P = 0.044$ Without CM). For the "SSI in pediatric patients in abdominal surgeries" outcome model, removal of CM suggested significant improvement in the model's performance ($P = 0.034$ with CM vs. $P = 0.171$ without CM).

The graphical representations of the HL-test statistics for each of the six outcomes modeled with and without CM used as a predictor are presented in Fig. 1. Each graphical symbol represents data for one of the 20 sequential risk categories constructed for each outcome model (with/without CM included as a predictor). The center line represents the line of best fit in which all predicted rates are equal to observed rates (slope = 1; intercept = 0). For each of the six outcomes, there is considerable overlap between the 20 sequential risk category data points (derived from regressing predicted rates on observed rates) for models that included CM as a predictor and for models that excluded CM. These results suggest that removing CM had little impact on the predicted model values and, subsequently, the goodness of fit for each outcome model. The largest variation between sequential risk category data points occurred for the "Morbidity in neonate patients in all surgeries." However, the HL-statistics for this outcome were relatively similar between the model that included CM (10.424; $P = 0.237$) and the model that excluded CM (7.932; $P = 0.440$) as a predictor. These results, combined with an identical C-statistic for both models (0.729), suggest that the removal of CM as a predictor did not significantly alter the model's prediction capability for this outcome.

3.2. Hospital level performance and correlation

Hospital level performance status, dependent on outlier and adjusted quartile rank, remained either the same or changed minimally across the six outcomes models with CM removal. Regarding outlier status, the same number of "Low Outliers" were detected for all models regardless of CM inclusion/exclusion, except

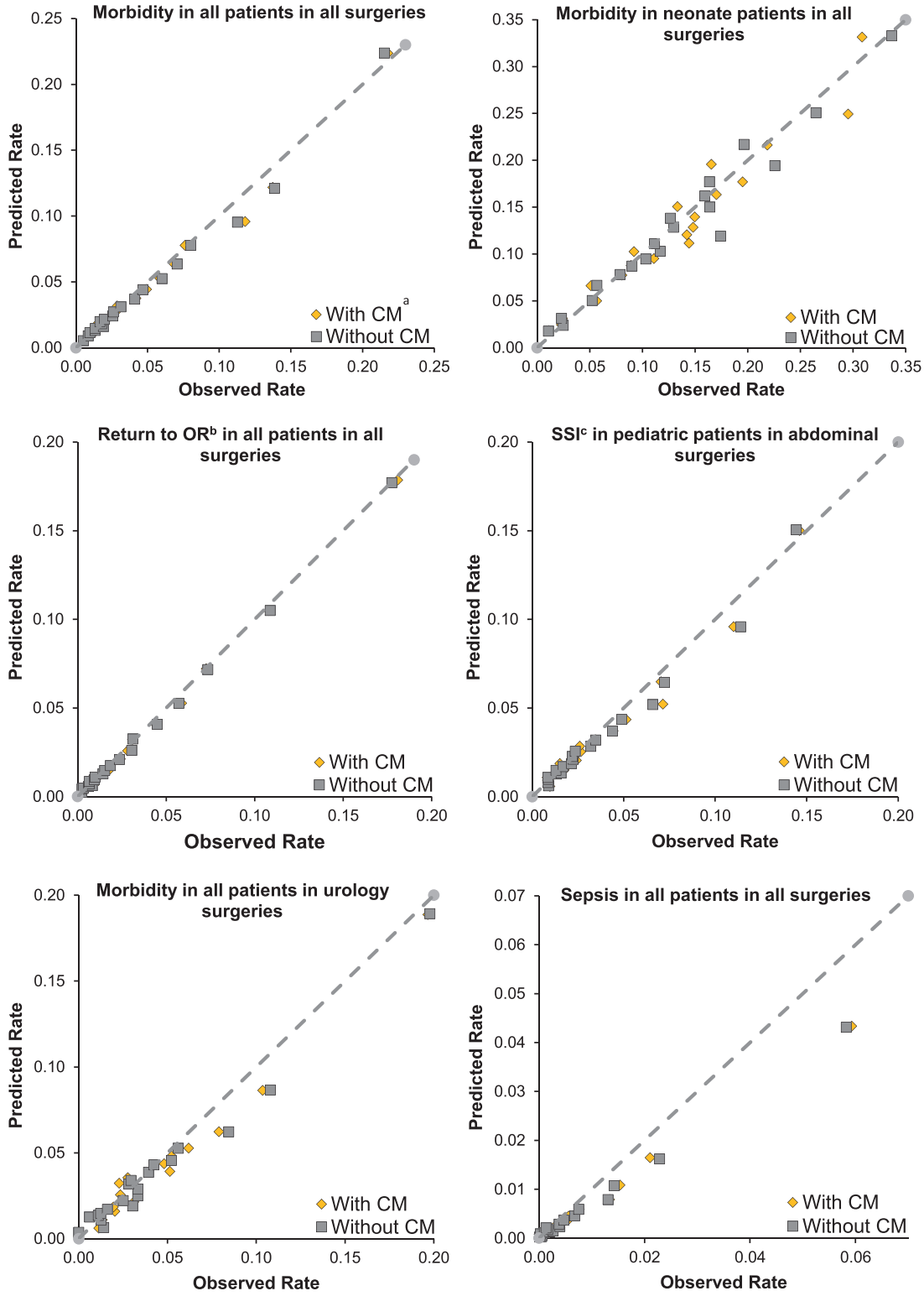

Fig. 1. Graphical representation of Hosmer-Lemeshow statistics for risk-adjusted outcomes assessed from the American College of Surgeons National Surgical Quality Improvement Project-Pediatric with and without congenital malformation. Graphical representation of Hosmer-Lemeshow statistics for the six outcome models with Congenital Malformation (CM; diamond (◊); yellow) and without CM (square (□); gray) used as a predictor in each model. Each symbol represents data for one of the 20 sequential risk categories constructed for these statistical analyses. The center line (dashed; gray) represents the line of best fit in which all predicted rates are equal to observed rates (slope = 1; intercept = 0) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).
[a]Congenital Malformation, [b]Operating room, [c]Surgical Site Infection

Table 2
Analyses of risk-adjusted models with and without the congenital malformation (CM) variable in the American College of Surgeons National Surgical Quality Improvement Project-Pediatric.

Risk-Adjusted Model	C-Index[a]		HL-Test Statistic[b]	
	With CM	Without CM	With CM (*P*-value)	Without CM (*P*-value)
Morbidity in all patients in all surgeries	0.783	0.782	24.7346 ($P = 0.002$)	32.0407 ($P = 0.001$)
Morbidity in neonate patients in all surgeries	0.729	0.729	10.4245 ($P = 0.237$)	7.9316 ($P = 0.440$)
Return to OR[c] in all patients in all surgeries	0.817	0.816	15.0768 ($P = 0.058$)	15.8687 ($P = 0.044$)
SSI[d] in pediatric patients in abdominal surgeries	0.768	0.766	16.6491 ($P = 0.034$)	11.5705 ($P = 0.171$)
Morbidity in all patients in urology surgeries	0.791	0.792	14.3484 ($P = 0.073$)	12.8402 ($P = 0.118$)
Sepsis in all patients in all surgeries	0.885	0.884	20.6934 ($P = 0.008$)	16.2131 ($P = 0.039$)

[a] C-index is a measure of the model's ability to predict a patient's outcome based on the predictors included in the logistic regression model
[b] HL is the Hosmer-Lemeshow goodness of fit test statistic which is used to determine if observed event rates match the expected event rates of subgroups within the population used in logistic regression modeling. A nonsignificant result reflects good calibration of the outcome model.
[c] Operating room
[d] Surgical Site Infection

Table 3
Summary of changes in number of hospitals included for outlier status and 1st and 4th adjusted quartile rank between outcomes models when congenital malformation was removed as a Predictor.

		Differences in Outlier Status[a]				Difference in Adjusted Quartiles Rank[b]			
		With CM[c]	Without CM	Outlier Status		With CM	Without CM	Adjusted Quartile Rank	
CM Outcome Models	Hospitals (*n*)	High/Low Outlier (*n*)	High/Low Outlier (*n*)	High Outlier Change[f] (*n*)	Low Outlier Change[f] (*n*)	$4^{th}/1^{st}$ Quartile (*n*)	$4^{th}/1^{st}$ Quartile (*n*)	4th Quartile Change[g] (*n*)	1st Quartile Change[g] (n)
Morbidity in all patients in all surgeries	141	26/15	26/15	0	-1, +1	29/26	29/27	0	.+1
Morbidity in neonate patients in all surgeries	133	6/0	5/0	-1	0	13/11	13/11	0	-2, +2
Return to OR[d] in all patients in all surgeries	141	13/6	12/6	-1	0	25/19	26/20	.+1	-1, +2
SSI[e] in pediatric patients in abdominal surgeries	136	14/3	14/3	0	0	20/18	21/18	.+1	0
Morbidity in all patients in urology surgeries	137	5/0	4/0	-1	0	13/12	16/13	.+3	-1, +2
Sepsis in all patients in all surgeries	141	28/7	28/8	0	.+1	26/26	26/26	0	0

[a] Outlier Status assignment is based on the 95% Confidence Interval (CI) for the model-adjusted Odds Ratio.
[b] Adjusted quartile ranks are based on percentile ranks of the adjusted Odds Ratios.
[c] Congenital Malformation.
[d] Operating Room.
[e] Surgical Site Infection.
[f] These values represent the shift in number of individual hospitals identified as high or low outliers when CM was removed from the model.
[g] These values represent the shift in number of individual hospitals identified as 4^{th} or 1^{st} quartile when CM was removed from the model.

for the "Sepsis in all patients in all surgeries" outcome in which the number of "Low Outliers" increased from 7 to 8 with CM removal. In this model, one new hospital was added to the cohort of hospitals identified as "Low Outliers" when CM was removed. In the "Morbidity in all patients in all surgeries" model, removal of CM resulted in the same number of hospitals identified as "Low Outlier" hospitals. While the same number of "Low Outlier" hospitals were identified, the removal of CM in this model resulted in the retention of 14 hospitals, loss of 1 hospital, and addition of 1 new hospital in the "Low Outlier" hospital cohort. In the remaining four models, the hospitals identified as "Low Outliers" were identical regardless of whether the model contained CM as a predictor (Table 3).

The number of "High Outliers" detected were the same for models regardless of CM inclusion/exclusion for the following three outcomes: (1) morbidity in all patients in all surgeries, (2) SSI in pediatric patients in abdominal surgeries, and (3) sepsis in all patients in all surgeries. For these three outcomes, the same cohort of hospitals were identified as "High Outliers" for each model regardless of CM inclusion/exclusion. For the remaining three outcomes, removing CM as a model predictor resulted in the same cohort of hospitals identified as "High Outliers" with a decrease of one "High Outlier" hospital per outcome model (Table 3).

For adjusted quartiles, the same number of 1st quartile hospitals was found in 3 models. For "SSI in pediatric patients in abdominal surgeries" and "Sepsis in all patients in all surgeries", there was no change in hospital quartile rank, however in "Morbidity in neonate patients in all surgeries", 2 hospitals lost their 1st quartile rank and 2 gained it. The remaining 3 models saw a net gain of one 1st quartile hospital with "Morbidity in all patients in all surgeries" seeing the addition of one hospital and the remaining 2 models losing one and gaining 2 1st quartile hospitals (Table 3).

Regarding changes in 4th quartile hospitals, the models "Return to OR in all patients in all surgeries" and "SSI in pediatric patients in abdominal surgeries" saw the addition of one 4th quartile hospital while "Morbidity in all patients in all surgeries", "Morbidity in neonate patients in all surgeries" and "Sepsis in all patients in all surgeries" saw no change in the number of 4th quartile ranked hospitals. The greatest change was seen in the "Morbidity in all patients in urology surgeries" which, upon removal of CM, added three 4th quartile hospitals (Table 3).

The average change in adjusted percentiles ranged from 0.504 to 1.602 rank positions with the smallest and largest average changes occurring for the "sepsis in all patients in all surgeries" and "morbidity in neonate patients in all surgeries" models, re-

Table 4
Summary of changes in number of hospitals included for needs improvement and exemplary performance assessment status[a] between outcome models when CM[b] was Removed as a Predictor

Outcome Model		With CM		Without CM		Performance Assessment Status	
	Hospitals (*n*)	NI[e] (*n*)	Exemplary (*n*)	NI (*n*)	Exemplary (*n*)	Changes[f] in NI (*n*)	Changes[f] in Exemplary (*n*)
Morbidity in all patients in all surgeries	141	30	26	30	27	0	.+1
Morbidity in neonate patients in all surgeries	133	13	11	13	11	0	-2, +2
Return to OR[c] in all patients in all surgeries	141	25	19	26	20	.+1	-1, +2
SSI[d] in pediatric patients in abdominal surgeries	136	20	18	21	18	.+1	0
Morbidity in all patients in urology surgeries	137	13	12	16	13	.+3	-1, +2
Sepsis in all patients in all surgeries	141	29	26	29	26	0	0

[a] Hospital Performance was assigned based upon outlier status and adjusted quartiles. Low outliers or 4th quartile hospitals were considered. "Needs Improvement" while high outliers or 1st quartile hospitals were considered "Exemplary". Outlier status and quartile were determined using hospital adjusted odds ratios.[b]Congenital Malformation.
[c] Operating Room.
[d] Surgical Site Infection.
[e] Needs Improvement.
[f] These values represent the shift in number of individual hospitals identified as Needs Improvement or Exemplary Performance Assessment Status when CM was removed from the model.

Table 5
Correlation analyses of log-transformed odds ratios between risk-adjusted models with and without the congenital malformation variable in the American College of Surgeons National Surgical Quality Improvement Project-Pediatric.

Risk-Adjusted Models	Correlation Coefficient
Morbidity in all patients in all surgeries	0.999
Morbidity in neonate patients in all surgeries	0.993
Return to OR[a] in all patients in all surgeries	0.998
SSI[b] in pediatric patients in abdominal surgeries	0.999
Morbidity in all patients in urology surgeries	0.997
Sepsis in all patients in all surgeries	1.000

[a] Operating room.
[b] Surgical Site Infection.

spectively. The average change in adjusted quartiles ranged from 0.000 to 0.068 quartile positions with the smallest and largest average changes occurring for the "sepsis in all patients in all surgeries" and "morbidity in neonate patients in all surgeries" models, respectively (Appendix C).

Considering both outlier status and adjusted quartile rank, changes in overall performance status were minimal. "Morbidity in all patients in urology surgeries" had 3 additional hospitals labelled as "Needs Improvement" (NI) with the removal of CM. "Return to OR in all patients in all surgeries" and "SSI in pediatric patients in abdominal surgeries" each added 1 NI hospital with the remaining 3 models having no change in NI assessment. Regarding exemplary assessment, the same number of exemplary hospitals was found in 3 models with "Morbidity in neonate patients in all surgeries" losing and gaining 2 exemplary hospitals. "SSI in pediatric patients in abdominal surgeries" and "Sepsis in all patients in all surgeries" had no change in the number of exemplary hospitals. The remaining 3 models had a net change of 1 additional exemplary hospital with "Morbidity in all patients in all surgeries" adding 1 exemplary hospital and the remaining two adding 2 and losing 1 exemplary hospital (Table 4). For all six outcome models, the percentage of hospitals in which Performance Assessment statuses shifted significantly was less than 4.36% which suggests that removing CM as a predictor in all six outcome models has a relatively small effect on hospital Performance Assessment status.

Lastly, Pearson correlation analyses were conducted on log-transformed ORs between each set of models for each of the six outcomes (Table 5). All correlation coefficients (r) were greater than or equal to 0.993 for the six of the outcomes of interest. These results suggest that there was good agreement between log-transformed ORs calculated from risk-adjusted models regardless of the inclusion of CM as a predictor variable.

3.3. Overlapping variables and other common variables

Overlapping variables were present throughout models regardless of CM inclusion. The CNS Abnormality and Hematologic Disorder variable was present in 4 and 3 outcome models, respectively, and these predictors were also present in each model subset (with and without CM). Esophageal/Gastric/Intestinal Disease was significant in 3 outcome models (and both model subsets – with and without CM) as well as the "morbidity in all patients in all surgeries" without CM model. The preoperative variables Cardiac Risk Factor and Developmental Delays were present in 2 outcome models (and their subsets - with and without CM) and Developmental Delay being additionally significant in "morbidity in all neonates in all surgeries" without CM. Both Neuromuscular Disorder and Pulmonary Abnormality were seen in 1 outcome model (and its subset - with and without CM) while Pulmonary Abnormality was also seen in the "morbidity in all neonates in all surgeries" without CM model. Seizure and Cerebral palsy were significant in the "morbidity in all patients in all surgeries" without CM model and the "sepsis in all patients in all surgeries" with CM model, respectively. Except for one outcome model, overlapping predictor variables either remained constant in rank or increased in significance with the removal of CM. The exception is that Hematologic Disorder decreased by one step in the hierarchical model for "morbidity in all patients in all surgeries". Common significant preoperative variables (seen in at least 5 models with and without CM) included American Society of Anesthesiologist Physical Status Classification System (ASA), Nutritional Support, Inpatient Surgery Classification, and the Presence of an Ostomy.

4. Discussion

The ability of NSQIP-Peds to accurately model hospital performance and reliably compare hospitals is contingent upon adequate hospital- and patient-level risk adjustment. Proper risk adjustment depends on accurate abstraction of patient- and case-specific variables. The abstraction of granular data is costly and burdensome. This study provides evidence that eliminating the CM composite variable has minimal effect on a variety of outcome models. Changes in hospital performance status were minimal and outlier status agreements were excellent when CM was removed from modeling. The correlation between models with and without CM was near perfect.

The finding that removing CM has a negligible effect on risk-adjusted performance modelling is not surprising given the contribution of redundant patient-level variables that are collected within NSQIP-Ped. As noted, these overlapping variables were common within all models and most increased in significance upon the

removal of CM. For example, cardiac risk factors include those with congenital structural defects and is known to carry elevated risk for worse outcomes in the postoperative period [21–23]. Patients with congenital structural cardiac defects contributing to their preoperative risk would also be considered to have a congenital malformation. Similarly, central nervous system abnormalities and the developmental delay variable contain conditions and data points that would similarly fall under the definition for CM. A patient with a CM that led to a significant systemic disease would likely have a higher ASA status which is a preoperative variable that is present in most NSQIP-Ped models.

With regards to the NSQIP-Ped literature, this is the first study that the authors are aware of that examines whether a variable can be removed without affecting model performance. However, there is precedence in studies examining adult populations in the ACS National Surgical Quality Improvement Program. One study found that excluding wound classification variables did not affect their model's ability to measure hospital performance. The authors note that this is likely due to inconsistencies in abstracting wound classification data and the redundant explanatory power present in other risk adjustment covariates [24]. A further study showed that procedure-specific hospital quality measures can be adequately risk adjusted with a limited model of the 5 most significant variables when compared with the full model containing all 21 variables considered significant by the stepwise regression model [25]. These findings show that constant evaluation of the variables being collected can yield significant results.

This study and those mentioned above serve as examples that have important implications for NSQIP-Ped and other clinical registries. They show that not all variables considered to be significant in risk-adjusted outcome modelling are truly necessary. Portions of their ability to risk-adjust may be compensated by other variables leading to redundancy in the collected variable. This concept applies especially to broad, complex, or loosely defined variables, such as CM, which are also typically labor-intensive to collect. While congenital malformations may play an important role in determining risk at the patient-level, our data demonstrates that the CM variable in its current format is not useful to collect for the NSQIP-Ped program. Removal of CM from NSQIP-Ped abstraction should allow for the development and implementation of more specific variables. Being more selective with data collection, and eliminating variables such as CM, will reduce the burden on SCRs. Furthermore, by eliminating unnecessary variables NSQIP-Ped may increase capacity to collect new variables or track more cases. This is especially pertinent as NSQIP-Ped transitions to collecting more procedure specific outcomes.

This study has certain limitations. First, this study only examines 6 models in which CM was significant. While these models were chosen as they were most likely to be representative and CM was ranked highest in significance, there is a small possibility that removal of CM may have different effects on other models containing specific procedures and patient populations that were not examined. Second, due to the wide variety of collected procedures, many variables are intentionally generic and may not offer enough granular detail for optimal risk-adjustment. Third, we did not examine the influence of removing CM as an eligible predictor for PTCs which may mean that any potential effects of this removal on procedure-specific hospital and modeling performance are still unknown. However, we hypothesized that significant effects related to the removal of CM on risk-adjusted modeling for PTC may be quite minimal due to the fact that oftentimes PTC models have eligible procedure-specific variables which may better capture the presence of CMs for risk-adjustment purposes. Regardless of the potential effects of CM removal on model and hospital performance for PTCs, the burden of abstraction would not justify inclusion of CM as it does not contribute to the primary object of NSQIP-Ped which is to provide hospital-level adjusted outcomes data. Finally, data submitted to NSQIP-Ped is mostly from large, dedicated children's hospitals which limits the generalizability of these findings.

5. Conclusion

Removal of the CM variable from NSQIP-Ped does not negatively impact the ability to model patient outcomes nor does it affect the ability to model and compare hospital-level outcomes. More specific variables estimating the influence of congenital malformations on risk-adjusted outcomes are likely necessary. Future efforts to decrease existing data collection burden might emulate these methods, recovering more capacity for salient variable addition and expansion.

Previous communication

None

Financial support

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Declarations of Competing Interest

None

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jpedsurg.2021.10.007.

References

[1] Raval MV, Pawlik TM. Practical guide to surgical data sets: national surgical quality improvement program (NSQIP) and pediatric NSQIP. JAMA Surg 2018;153:764–5.
[2] Dillon P, Hammermeister K, Morrato E, et al. Developing a NSQIP module to measure outcomes in children's surgical care: opportunity and challenge. Semin Pediatr Surg 2008;17:131–40.
[3] Bruny JL, Hall BL, Barnhart DC, et al. American College of Surgeons national surgical quality improvement program pediatric: a beta phase report. J Pediatr Surg 2013;48:74–80.
[4] Raval MV, Dillon PW, Bruny JL, et al. Pediatric American College of Surgeons national surgical quality improvement program: feasibility of a novel, prospective assessment of surgical outcomes. J Pediatr Surg 2011;46:115–21.
[5] Stey AM, Kenney BD, Moss RL, et al. A risk calculator predicting postoperative adverse events in neonates undergoing major abdominal or thoracic surgery. J Pediatr Surg 2015;50:987–91.
[6] Patel VA, Khaku A, Carr MM. Pediatric thyroidectomy: NSQIP-P analysis of adverse perioperative outcomes. Ann Otol Rhinol Laryngol 2020;129:326–32.
[7] Schoel L, Maizlin II, Koppelmann T, et al. Improving imaging strategies in pediatric appendicitis: a quality improvement initiative. J Surg Res 2018;230:131–6.
[8] Anandalwar SP, Cameron DB, Graham DA, et al. Association of intraoperative findings with outcomes and resource use in children with complicated appendicitis. JAMA Surg 2018;153:1021–7.
[9] Skarsgard ED, Bedford J, Chan T, Whyte S, Afshar K. ACS national surgical quality improvement program: targeting quality improvement in Canadian pediatric surgery. J Pediatr Surg 2014;49:682–7.
[10] Cameron DB, Melvin P, Graham DA, et al. Extended versus narrow-spectrum antibiotics in the management of uncomplicated appendicitis in children: a propensity-matched comparative effectiveness study. Ann Surg 2018;268:186–92.
[11] Malik AT, Yu E, Kim J, Khan SN. Intensive care unit admission following surgery for pediatric spinal deformity: an analysis of the ACS-NSQIP pediatric spinal fusion procedure targeted dataset. Global Spine J 2020;10:177–82.
[12] Anderson KT, Bartz-Kurycki MA, Austin MT, et al. Room for "quality" improvement? Validating national surgical quality improvement program-pediatric (NSQIP-P) appendectomy data. J Pediatr Surg 2019;54:97–102.
[13] Raval MV, Dillon PW, Bruny JL, et al. American College of Surgeons national surgical quality improvement program pediatric: a phase 1 report. J Am Coll Surg 2011;212:1–11.

[14] Mery CM, De León LE, Rodriguez JR, et al. Effect of gastrointestinal malformations on the outcomes of patients with congenital heart disease. Ann Thorac Surg 2017;104:1590–6.
[15] Bojanić K, Pritišanac E, Luetić T, et al. Malformations associated with congenital diaphragmatic hernia: Impact on survival. J Pediatr Surg 2015;50:1817–22.
[16] Stolwijk LJ, Keunen K, de Vries LS, et al. Neonatal surgery for noncardiac congenital anomalies: neonates at risk of brain injury. J Pediatr 2017;182:335–41 .e331.
[17] . ACS NSQIP Pediatric Variables & DefinitionsACS NSQIP Pediatric Operations Manual. American College of Surgeons; 2020.
[18] Hosmer D, Lemeshow S. Applied logistic regression. 2nd ed. New York, NY: John Wiley & Sons, Ltd; 2000.
[19] Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: the Hosmer-Lemeshow test revisited. Crit Care Med 2007;35:2052–6.
[20] Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018;18:91–3.
[21] Saettele AK, Christensen JL, Chilson KL, Murray DJ. Children with heart disease: risk stratification for non-cardiac surgery. J Clin Anesth 2016;35:479–84.
[22] Taylor D, Habre W. Risk associated with anesthesia for noncardiac surgery in children with congenital heart disease. Paediatr Anaesth 2019;29:426–34.
[23] García H, Cervantes-Luna B, González-Cabello H, Miranda-Novales G. Risk factors for nosocomial infections after cardiac surgery in newborns with congenital heart disease. Pediatr Neonatol 2018;59:404–9.
[24] Ju MH, Cohen ME, Bilimoria KY, et al. Effect of wound classification on risk adjustment in American College of Surgeons NSQIP. J Am Coll Surg 2014;219:371–81 .e375.
[25] Dimick JB, Osborne NH, Hall BL, Ko CY, Birkmeyer JD. Risk adjustment for comparing hospital quality with surgery: how many variables are needed? J Am Coll Surg 2010;210:503–8.