Check for updates

POPULATION STUDY ARTICLE

Predicting prolonged length of stay in hospitalized children with respiratory syncytial virus

Krista Wollny[1,2,3 ✉], Tona Pitt[4], Darren Brenner[1,3,5] and Amy Metcalfe[1,3,6,7]

BACKGROUND: Respiratory syncytial virus (RSV) is the most common cause of lower respiratory tract infections in children. This study aimed to predict the prolonged length of stay in children admitted to hospital with RSV.
METHODS: Children aged <2 years with RSV in the National Inpatient Sample (NIS) were included in the analyses. The primary outcome was prolonged length of stay (≥90th percentile). Logistic regression models were developed using data from 2016; internal validation was completed using a bootstrapped sample. Data from 2017 were used to validate out-of-sample discrimination and calibration of the models.
RESULTS: The sample included 9589 children; 1054 had prolonged length of stay (≥7 days). Children who were younger, transferred from another hospital, and required intubation during admission had a higher risk of prolonged length of stay. The prediction model included age, transport, intubation, comorbidities, hospital location, and teaching status. The area under the receiver operating characteristic curve was 0.73, demonstrating good predictive ability. The model performed similarly in external validation.
CONCLUSIONS: Variables that predict the prolonged length of stay for RSV include younger age, transport, intubation, comorbidities, hospital location, and teaching status. This can be used to predict children who will have a prolonged length of stay when hospitalized for RSV.

IMPACT:

- There are no recommended treatments for RSV; medical care involves supportive treatment such as oxygen delivery, hydration, and antipyretics.
- The clinical course is difficult to predict, partially attributable to the supportive nature of care and the sparsity of evidence-based therapies for this population.
- A prediction model was developed, demonstrating variables that predict prolonged length of stay in RSV hospitalizations, including age, interhospital transport, intubation, comorbidities, hospital location, and teaching status.
- The model was developed with a sample size of 9589 that is representative of all hospitalizations in the United States.

INTRODUCTION

Respiratory syncytial virus (RSV) is the most common cause of lower respiratory tract infections in children.[1] Worldwide, there are approximately 34 million new cases of RSV-related lower respiratory tract infections per year in children under the age of 5 years, with 10% requiring hospitalization.[2] RSV presents with signs and symptoms of upper and/or lower respiratory tract infection, varying in severity from rhinorrhea, nasal congestion, fever, and myalgia to wheezes, prolonged expiration, and respiratory failure.[3] The primary management for RSV is supportive care, including respiratory support, antipyretics, and hydration.[3] Respiratory support ranges from humidified oxygen therapy to endotracheal intubation and mechanical ventilation, where a

breathing tube is inserted into the airway and a machine breathes for the child.[4] Children at risk for severe respiratory disease may be given palivizumab, an immunoprophylaxis medication, but this is restricted to certain populations such as preterm infants (born before 29 weeks gestational age or before 34 weeks gestational age with chronic lung disease) and children with severe heart disease.[1] Palivizumab is given to prevent RSV and is unable to treat the viral infection once it has been contracted.[1] The hospital course for RSV is challenging to predict, partially attributable to the supportive nature of care and the sparsity of evidence-based therapies for this population.[5]

The length of stay during hospital admission for RSV is estimated to range between 3 and 7 days, which is determined

[1]Alberta Children's Hospital Research Institute, Calgary, AB, Canada. [2]Faculty of Nursing, University of Calgary, Calgary, AB, Canada. [3]Department of Community Health Sciences, University of Calgary, Calgary, AB, Canada. [4]Department of Paediatrics, University of Calgary, Calgary, AB, Canada. [5]Department of Oncology, University of Calgary, Calgary, AB, Canada. [6]Department of Obstetrics and Gynecology, University of Calgary, Calgary, AB, Canada. [7]Department of Medicine, University of Calgary, Calgary, AB, Canada. ✉email: krista.wollny@ucalgary.ca

by the severity of illness and hospital-level factors that impact efficiency.[5–7] Prolonged length of stay has been defined as greater than or equal to the 90th percentile length of stay for a particular patient population or group.[8] Prolonged length of stay results in an increased risk of hospital-related adverse effects and increased costs to the healthcare system, patient, and family.[6,7,9]

Studies have evaluated variables that are associated with the severity of disease or death caused by RSV. Associated comorbidities include preterm birth, congenital heart disease, neuromuscular disease, chronic lung disease, airway anomalies, asthma, trisomy 21, and conditions that impact immune function.[10–12] Prolonged length of stay for RSV has been associated with several patient-level variables, including younger age, prematurity, co-infection, co-morbidities, and interventions such as intubation and mechanical ventilation.[5,6,8,13–15]

Prediction models are used to classify patients as being at risk (or higher probability) for an outcome, which can often have a greater clinical impact than traditional logistic regression models.[16] A model to predict a prolonged length of stay in children with RSV would help clinicians identify these patients and prepare the family for a longer length of stay. Logistic regression models have been built to predict the severity of RSV[17] or hospitalization for RSV.[18] To our knowledge, a model has not been built to predict the prolonged length of stay in children with RSV.

The overarching aim of this study is to predict the prolonged length of stay in children admitted to the hospital with RSV. The specific research objectives are: to develop and internally validate a prediction model for a prolonged length of stay in children admitted to hospital with RSV; and to externally validate the prediction model using temporal validation.

METHODS
This report adheres to the Transparent Reporting of a multivariate prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines.[19]

Data source
We examined the National Inpatient Sample (NIS), which is a sample obtained through stratified, self-weighted, single-stage cluster design, approximating 20% of all hospitalizations in the United States.[20] Data from 2016 were used to develop and internally validate the prediction model. Data from 2017 were used for temporal external validation. As the NIS contains publicly available de-identified data, this study was exempt from ethics review by the Conjoint Health Research Ethics Board at the University of Calgary.

Sample
Hospitalizations were included if the patients were under the age of 2 years at admission to the hospital and had a primary diagnosis of RSV in 2016 or 2017. The Tenth Revision of the International Statistical Classification of Diseases and Related Health Problems Clinical Modification (ICD-10-CM) diagnosis codes were used to identify RSV diagnosis. The codes used were: J12.1 (RSV pneumonia), J20.5 (acute bronchitis due to RSV), J21.0 (acute bronchiolitis due to RSV), and B97.4 (RSV).[21] It is estimated that 75–90% of children admitted to hospital with RSV are ≤12 months old,[22] so including children up to the age of 2 years should capture the population of interest and will be consistent with the literature in the field. Hospitalizations were excluded if the length of stay was <1 day to ensure we only captured children who were admitted to the hospital. Hospitalizations were also excluded if the children died during their hospital admission or were transferred out to an acute care hospital, as including these may bias the results.

Variables
The primary outcome was prolonged length of stay. Prolonged length of stay was determined by finding the 90th percentile length of stay from the data. The patient-level variables that we included in the initial model were age, sex, race, comorbidities, transfer from another site, household income, health insurance status, and intubation. Age was categorized into <1 month, 1 month to <12 months, and ≥12 months. Sex was included in

the initial model despite the conflicting evidence to support male sex being associated with RSV hospitalization and severity.[11,12,23–26] Race was included as others have found an association between the non-Hispanic white race and RSV hospitalization.[23] The comorbidities we included in the model include congenital cardiac disease, chronic lung disease, neuro-muscular disease, trisomy 21, and immunosuppression. See Supplemental Information for the ICD-10-CM codes used to identify these variables. These comorbidities were chosen based on the literature and availability in the NIS database. Others have found that children who required transfer from an acute care hospital have increased odds of severe RSV,[10] so we included this variable. Household income (in quartiles) and health insurance (government insurance, private insurance, self-pay, or other) were included in the initial model, as these relate to the child's socioeconomic status and access to healthcare. Intubation and mechanical ventilation have been found to be associated with increased length of stay for children hospitalized for RSV, so intubation was also included.[13] The hospital-level variables that were considered for the model included urban or rural hospital status and teaching or non-teaching hospital status.[20] It is anticipated that the urban and teaching hospitals would have higher acuity patients, resulting in a longer length of stay.[27]

Statistical analyses
Data analyses were performed using STATA IC 15 (Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC). Descriptive statistics were used to examine the data, taking into account the sampling design. Discharge weights were used to produce national estimates.[20] Logistic regression models were developed using the NIS data from 2016. The variables described above were assessed for association with the outcome using bivariate analysis ($p < 0.2$) to be included in the initial logistic regression model. All variables were dichotomized or categorized for interpretation purposes. Multicollinearity was assessed using the variance inflation factor.[28] Stepwise selection was used to remove variables, retaining variables significantly associated with length of stay ($p < 0.05$) in the model.

Comparison of model performance (internal validation)
To test the model's performance, each regression model was fit in a bootstrapped sample with the same number of subjects as the original sample. This method is preferred over split-sample or cross-validation designs, as it maintains a large sample size.[29] The bootstrap sample was drawn with replacement and was replicated 500 times to generate precise confidence intervals and optimism correction. The c-statistic, which, for dichotomous outcomes is equivalent to the area under the receiver operating characteristic (ROC) curve, was used to quantify the discriminative ability of each model. Discrimination quantifies how well a model can predict whether a child hospitalized with RSV would have a prolonged length of stay.[30] Calibration, the agreement between the predicted and the observed outcomes, was assessed using the Hosmer–Lemeshow goodness-of-fit test and calibration slopes.[30] Both discrimination and calibration are essential to assess for in prediction models because they quantify the model's performance, demonstrating whether or not the model can predict what it should.

Comparison of model performance (external validation)
The data from 2017 were used to test for the external validity of the models. Discrimination and calibration of the models in the 2017 sample were conducted using the same methods described above.

RESULTS
The total sample size was 9589 hospitalizations. The mean length of stay was 3.7 days (95% confidence interval (CI): 3.6–3.8). The prolonged length of stay (90th percentile of the length of stay) was 7 days, so the outcome was dichotomized into ≥7 days and <7 days. In general, hospitalizations with children <1 month had the highest proportion of the prolonged length of stay (20.1%), compared to those 1–12 months (10.8%) and those ≥12 months (7.9%). Children being transferred from another acute care hospital had a higher proportion of the prolonged length of stay (18.3%), compared to those not transferred (10.0%). Most hospitalizations that required intubation had a prolonged length of stay (94.1%). See Table 1 for the sample characteristics. The

Table 1. Sample demographics.

Variable	Description	2016 N = 9589			2017 N = 9705		
		LOS <7 days, N (%)	LOS ≥7 days, N (%)	p value	LOS <7 days, N (%)	LOS ≥7 days, N (%)	p value
Age	<1 month	545/ 682 (79.9%)	137/682 (20.1%)	<0.001	574/708 (81.1%)	134/708 (18.9%)	<0.001
	1–12 months	6447/7231 (89.2%)	784/7231 (10.8%)		6571/7321 (89.8%)	750/7321 (10.2%)	
	>12 months	1542/1675 (92.1%)	133/1675 (7.9%)		1536/1676 (91.7%)	140/1676 (8.4%)	
Sex	Male	4786/5370 (89.1%)	584/5370 (10.9%)	0.68	4965/5522 (89.9%)	557/5522 (10.1%)	0.086
	Female	3749/4219 (88.9%)	470/4219 (11.1%)		3715/4182 (88.8%)	467/4182 (11.2%)	
Transfer	Not transferred	7397/8220 (90.0%)	823/8220 (10.0%)	<0.001	7556/8354 (90.5%)	798/8354 (9.5%)	<0.001
	Transferred from acute care hospital	947/1159 (81.7%)	212/1159 (18.3%)		1100/1324 (83.1%)	224/1324 (16.9%)	
Comorbidities	Congenital cardiac disease	243/391 (62.2%)	148/391 (37.9%)	<0.001	261/419 (62.3%)	158/419 (37.7%)	<0.001
	Chronic lung disease	53/97 (54.6%)	44/97 (45.4%)	<0.001	77/130 (59.2%)	53/130 (40.8%)	<0.001
	Neuromuscular disease	1/5 (20.0%)	4/5 (80.0%)	<0.001	4/9 (44.4%)	5/9 (55.6%)	<0.001
	Trisomy 21	58/86 (67.4%)	28/86 (32.6%)	<0.001	75/121 (62.0%)	46/121(38.0%)	<0.001
	Immunosuppressed	14/21 (66.6%)	7/21 (33.3%)	<0.001	7/14 (50.0%)	7/14 (50.0%)	<0.001
Intubated	Intubated	13/ 222 (5.9%)	209/222 (94.1%)	<0.001	42/253 (16.6%)	211/253 (83.4%)	<0.001
	Not intubated	8522/9367 (91.0%)	845/9367 (9.0%)		8639/9452 (91.4%)	813/9452 (8.6%)	
Urban or rural hospital	Urban	7697/8729 (88.2%)	1032/8729 (11.8%)	<0.001	7824/8823 (88.7%)	999/8823 (11.3%)	<0.001
	Rural	838/860 (97.4%)	22/860 (2.6%)		857/882 (97.2%)	25/882 (2.8%)	
Teaching or non-teaching	Teaching	6452/7393 (87.3%)	941/7393 (12.7%)	<0.001	6760/7693 (87.9%)	933/7693 (12.1%)	<0.001
	Non-teaching	1245/1336 (93.2%)	91/1336 (6.8%)		1064/1130 (94.2%)	66/1130 (5.8%)	

Table 2. Model 1—β coefficients and odds ratios.

Variable	Description	β Coefficients	Odds ratio (95% CI)	p value
Age	<1 month	—	Baseline	—
	1–12 months	−0.81	0.45 (0.36–0.55)	<0.001
	>12 months	−1.15	0.32 (0.24–0.42)	<0.001
Transfer	Transferred from acute care hospital	0.45	1.57 (1.30–1.90)	<0.001
Comorbidities	Congenital cardiac disease	1.22	3.40 (2.61–4.42)	<0.001
	Chronic lung disease	1.85	6.38 (4.04–10.08)	<0.001
	Neuromuscular disease	0.85	2.33 (0.04–133.02)	0.682
	Trisomy 21	1.00	2.72 (1.61–4.61)	<0.001
	Immunosuppressed	0.86	2.37 (0.81–6.95)	0.115
Intubated	Intubated	4.85	127.84 (72.20–226.36)	<0.001
Hospital type	Rural	—	Baseline	—
	Urban, non-teaching	0.91	2.49 (1.53–4.06)	<0.001
	Urban, teaching	1.32	3.74 (2.40–5.82)	<0.001
Constant		−2.92		

mean total charges for the hospital stay was $16,709 USD (95% CI: $16,389–17,032) for hospitalizations <7 days and $100,052 USD (95% CI: $91,688–108,416) for hospitalizations lasting 7 days or longer.

The first model (Model 1) included age, transport from an acute care hospital, intubation, comorbidities (congenital cardiac disease, chronic lung disease, neuromuscular disease, trisomy 21, and immunosuppression), urban or rural status, and teaching status (Table 2). Using a classification cut-off of 0.2, Model 1 had an area under the ROC curve of 0.73 (95% CI: 0.71–0.75), demonstrating good predictive ability (Fig. 1a). The Hosmer–Lemeshow goodness-of-fit test was non-significant with a p value of 0.44, demonstrating agreement between the predicted and observed outcomes. The calibration slope was 1.0, confirming the agreement between the predicted and observed outcomes (Fig. 1b). The model's sensitivity was 33.7% and specificity was 96.2%; the positive predictive value was 52.4%, and the negative predictive value was 92.1%.

A second model (Model 2) was developed to test whether the specific comorbidities were highly predictive, or if the number of comorbidities a child had could replace these variables. There were 454 (4.7%) RSV hospitalizations where children had one comorbidity and 73 (0.7%) with two. This model included age, if the patient was transported, if the patient was intubated, the number of comorbidities, urban or rural status, and teaching status. Model 2 had an area under the ROC curve of 0.73 (95% CI: 0.71–0.75), demonstrating a similar predictive ability as Model 1 with a classification cut-off of 0.2. The Hosmer-Lemeshow goodness-of-fit test was non-significant with a p value of 0.38, demonstrating agreement between the predicted and observed outcomes. Again, the calibration slope was 1.0, confirming the agreement between the predicted and observed outcomes. Model 2 had similar sensitivity and slightly higher specificity than Model 1, with values of 33.3 and 96.3%, respectively.

A third model (Model 3) was developed to test how well the model performed without the variable *intubation*, including variables: age, if the patient was transported, the number of comorbidities, urban or rural status, and teaching status. Model 3 had an area under the ROC curve of 0.70 (95% CI: 0.69–0.71), finding slightly lower predictive ability than Models 1 and 2. The Hosmer-Lemeshow goodness-of-fit test was non-significant with a p value of 0.32, suggesting adequate model fit.

All models included sex, race, household income, and health insurance in the initial model development; however, these variables were not found to be predictors of increased length of stay for RSV hospitalizations.

All three models were externally validated using NIS data from 2017. See Table 1 for the sample characteristics for this data. The models performed similarly in the external validation analysis. The ROC curve for Model 1 using 2017 data was 0.74 (95% CI: 0.72–0.76), demonstrating that the model has a similar predictive ability in a different sample (Fig. 1c). The calibration slope was 1.0 for Model 1 fit with 2017 data, demonstrating agreement between the predicted and observed outcomes (Fig. 1d). The ROC curve for Model 3 using 2017 data was 0.74 (95% CI: 0.72–0,76), and the calibration slope was 1.0, demonstrating that when the variable *intubation* was removed, it did not change the model's predictive ability or the fit of the model in the new dataset.

Going back to the NIS data from 2016, there were 99 hospitalizations that had ≥95% predicted probability of having a prolonged length of stay, obtained from the second model. The majority of the patients (69.4%) were >1 month, but <1 year old. Only 7.1% of the total sample was under 1 month; however, this age group accounted for 19.4% of the hospitalizations with ≥95% predicted probability of prolonged length of stay. Most (75.7%) of the hospitalizations with ≥95% predicted probability of being prolonged had intubation occur. The majority (52.5%) had a patient with 1 comorbidity, and 9.1% had 2 comorbidities. Almost all of the hospitalizations occurred in urban hospitals (90.1%) and teaching hospitals (85.9%). The majority of hospitalizations with ≥95% predicted probability of prolonged length of stay did not involve a transfer from another acute care hospital (69.7%); however, a higher percentage of hospitalizations had patients who were transferred (30.3%) in this group than the overall sample (12.1%).

COMMENT

The variables included in this prediction model are consistent with the literature. Others have found that younger age,[12,13,31] transport from another site,[10] and comorbidities, including trisomy 21,[14] congenital cardiac disease[13,32], and bronchopulmonary dysplasia,[12,31] have been associated with increased severity of RSV. Other studies have found factors that have moderate associations with severity of illness such as maternal smoking during pregnancy,[25] siblings ≥2 years of age,[22,25] and daycare exposure[22]; we were unable to include these variables in the model due to limitations in the availability of such data. Intubation for RSV has been associated with an increased length of stay in other cohorts,[10,13] which makes sense from a clinical perspective. Children who require intubation have a higher severity of illness, requiring the highest level of respiratory support. Unfortunately,
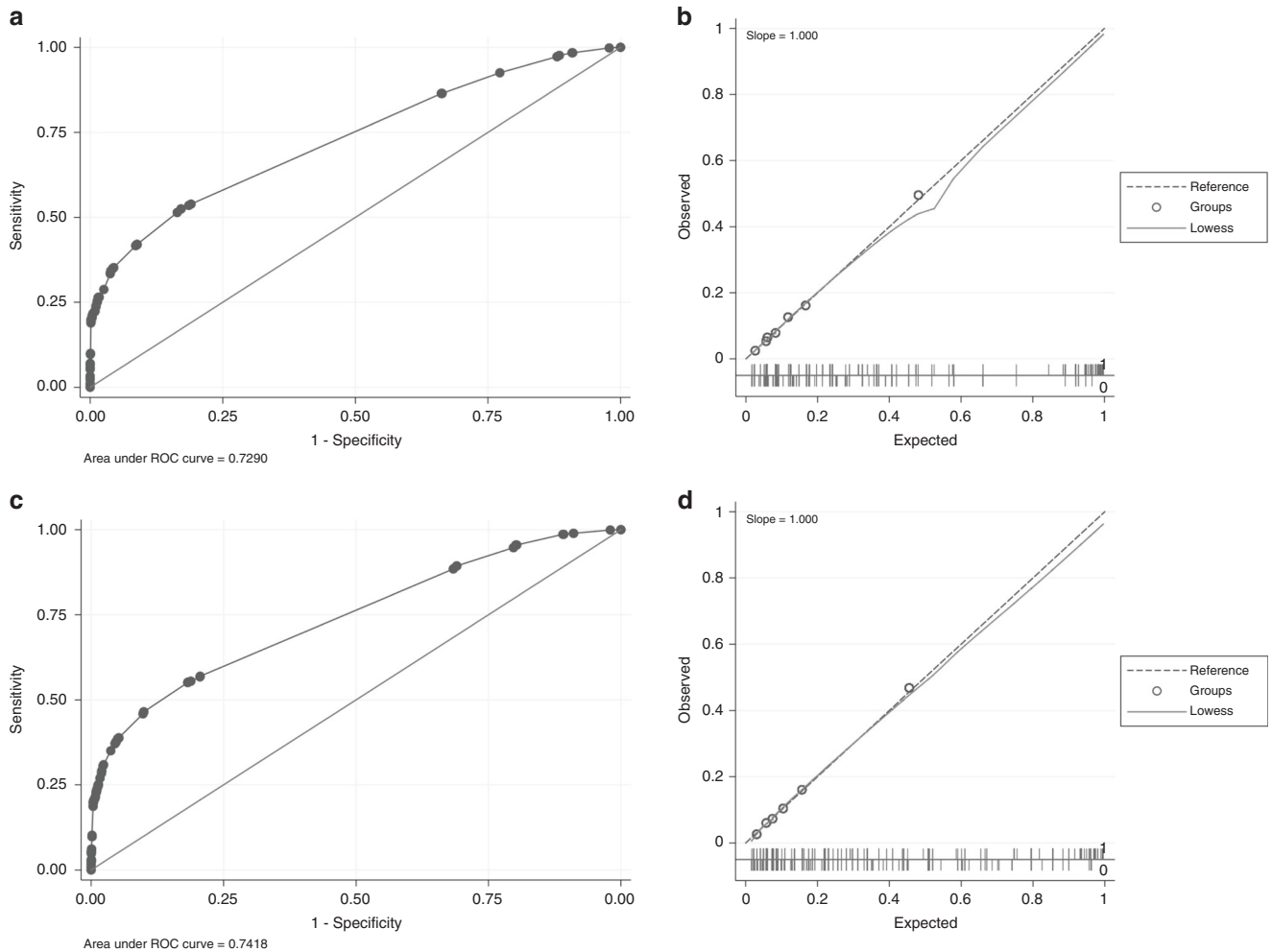
a

Area under ROC curve = 0.7290

b

c

Area under ROC curve = 0.7418

d

Fig. 1 Internal and external validation. a ROC curve for Model 1 (internal validation); **b** calibration slope for Model 1 (internal validation); **c** ROC curve for Model 1 (external validation); **d** calibration slope for Model 1 (external validation).

one limitation of using data from NIS is that the variables are not time-related, and we were unable to determine when during the hospital stay the child was intubated. We developed Model 3 excluding the variable *intubation* to provide a model that only included variables that would be known at the time of admission. Interestingly, all three models performed similarly in external validation, suggesting that removing the variable *intubation* from the model did not impact the model's performance.

The internal validation demonstrated good predictive ability and discrimination of the models. All three models had high specificity, as they were able to accurately predict >95% of the patients who did not have a prolonged length of stay. High specificity is clinically relevant, as it provides information about the hospitalizations that will most likely have a length of stay <7 days. Clinicians can use this information to identify the children who will most likely have a shorter, less-costly length of stay, reassuring families and the healthcare team. Conversely, the models accurately predicted approximately one-third of patients who actually had a prolonged length of stay, demonstrating low sensitivity. This may be due to unmeasured variables that are not captured in the NIS database, such as birth history and pre-existing conditions. As with any administrative database, information is only captured if it was documented during the admission.[33] Other studies have demonstrated that the validity of certain conditions, such as congenital anomalies, varies based on the database used, which can bias the analyses.[34] The benefit of using an administrative dataset for this study is the low cost and labor

associated with the collection, and the ability to have a large sample size. To address this concern, future research could pursue externally validating the model using clinical data.

An interesting finding from the descriptive analyses includes the estimated mean cost difference between children admitted for RSV with and without a prolonged length of stay ($100,000 vs $17,000 USD). The increased costs would not only be due to a longer hospitalization; this likely reflects the increased acuity, requiring additional interventions such as intubation and intensive care admission, that may be associated with a prolonged length of stay. Predicting the patients who are at risk for a prolonged length of stay should be of interest to hospital administrators and unit managers, based on these stark differences in estimated mean costs.

This study has several strengths, including large sample size and external validation conducted in a temporally different sample. The large sample size ensured there was adequate power to develop and validate the models. The external validation demonstrated very similar results to the internal validation measures, increasing the transportability of the models.[30,35] This suggests the models are generalizable to the population that this sample is reflective of (hospitalizations in the United States). Future studies could explore the generalizability of the model in another population, separate from the NIS database.

There are limitations when using an administrative database. We are relying on the validity of the ICD-10 codes to identify the diagnoses and procedures desired; however, if the hospitalization

record was not coded properly, there is a risk for both selection and misclassification bias. When possible, we used ICD-10 codes that had been validated in other studies. A recent study evaluated ICD-10 codes for RSV surveillance and found them to have low sensitivity (6%, 95% CI: 3–12%) but high specificity (99.8%, 95% CI: 99.6–99.9%).[21] This may lead to an underestimation of the true prevalence of RSV in the population, and patients who were admitted to hospital with RSV may be excluded; however, we are confident that most of the hospitalizations identified as having RSV are true positive cases.

Missing data is another concern when using the NIS.[36] Race, for example, is missing in approximately 8% of cases because not all sites provide this data. To maintain a large sample size and avoid listwise deletion, we coded any of the missing data on race to "unknown." We were also unable to determine if a child was represented twice in this database, as the NIS contains individual admissions, not individually identified patients. Based on prior research demonstrating low rates of readmission for pediatric lower respiratory tract infections,[37] we hypothesize that the number of patients admitted for RSV twice in one year would be limited, and would not impact our findings.

There are certain variables, such as a history of preterm birth, that have been found to be associated with prolonged length of stay with RSV[18] that are not captured in the NIS database. To mitigate this limitation, we used the ICD-10 code for chronic respiratory disease originating in the perinatal period (classified as chronic lung disease) to capture infants born preterm who are experiencing lasting respiratory effects. Additionally, we were unable to categorize age into smaller groups, which limits the analysis because the risk of severe infection is not uniform across the first year of age.[38] Due to these limitations, we feel that the development of a clinical decision-making tool would be misleading at this time. The next steps should involve capturing clinically relevant variables and collaboration with clinicians to develop probability thresholds that could alert the healthcare team about the children most likely to have a prolonged length of stay. This could benefit patients and families by adequately preparing them for a longer hospital stay, as well as ensuring the hospital units are appropriately staffed.

CONCLUSIONS

The analysis suggests that the variables that can predict a prolonged length of stay for RSV include younger age, transport from an acute care hospital, intubation, increased number of comorbidities (or specifically one or more of congenital cardiac disease, chronic lung disease, neuromuscular disease, trisomy 21, or immunosuppression), urban hospitals, and teaching hospitals. These results are consistent with previous literature that demonstrated these variables were associated with increased severity of disease or increased length of stay with RSV.[1,10–12,26] Future studies could explore the clinical presentations of children with RSV that predict the prolonged length of stay, perhaps increasing the predictive ability of this model. With electronic medical records being used in most acute care hospitals,[39] accessing clinical data such as vital signs, respiratory assessments, oxygen requirements, and interventions may facilitate the development of a more clinically usable prediction model.

REFERENCES

1. Smith, D. K., Seales, S. & Budzik, C. Respiratory syncytial virus bronchiolitis in children. *Am. Fam. Physician* **95**, 94–99 (2017).
2. Shi, T. et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. *Lancet* **390**, 946–958 (2017).
3. Schweitzer, J. W. & Justice, N. A. *Respiratory Syncytial Virus Infection (RSV)* (Stat-Pearls Publishing, 2019).
4. Pierce, H. C. et al. Variability of intensive care management for children with bronchiolitis. *Hosp. Pediatr.* **5**, 175–184 (2015).
5. Masarweh, K. et al. Factors predicting length of stay in bronchiolitis. *Respir. Med.* **161**, 105824 (2020).
6. Sruamsiri, R., Kubo, H. & Mahlich, J. Hospitalization costs and length of stay of Japanese children with respiratory syncytial virus: a structural equation modeling approach. *Medicine* **97**, e11491 (2018).
7. Stockwell, D. C. et al. Using length of stay to understand patient flow for pediatric inpatients. *Pediatr. Qual. Saf.* **3**, e050 (2018).
8. Gold, J. M. et al. Long length of hospital stay in children with medical complexity: LOS in children with medical complexity. *J. Hosp. Med.* **11**, 750–756 (2016).
9. Geoghegan, S. et al. The experience of long-stay parents in the ICU: a qualitative study of parent and staff perspectives. *Pediatr. Crit. Care Med.* **17**, e496–e501 (2016).
10. Pham, H., Thompson, J., Wurzel, D. & Duke, T. Ten years of severe respiratory syncytial virus infections in a tertiary paediatric intensive care unit. *J. Paediatr. Child Health* https://doi.org/10.1111/jpc.14491 (2019).
11. Tam, J. et al. Pediatric Investigators Collaborative Network on Infections in Canada Study of Respiratory Syncytial Virus-associated Deaths in Pediatric Patients in Canada, 2003-2013. *Clin. Infect. Dis.* **68**, 113–119 (2019).
12. Hall, C. B. et al. The burden of respiratory syncytial virus infection in young children. *N. Engl. J. Med.* **360**, 588–598 (2009).
13. McMillan, J. A. et al. Prediction of the duration of hospitalization in patients with respiratory syncytial virus infection: use of clinical parameters. *Pediatrics* **81**, 22–26 (1988).
14. Beckhaus, A. A. & Castro-Rodriguez, J. A. Down Syndrome and the risk of severe RSV infection: a meta-analysis. *Pediatrics* **142**, e20180225 (2018).
15. Heys, M., Rajan, M. & Blair, M. Length of paediatric inpatient stay, socio-economic status and hospital configuration: a retrospective cohort study. *BMC Health Serv. Res.* **17**, 274 (2017).
16. Steyerberg, E. W. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating* (Springer, 2019).
17. Ramos-Fernández, J. M. et al. [Prediction of severe course in infants with RSV bronchiolitis under 6 months. Spain]. *Rev. Esp. Salud Publica* **91**, e201701006 (2017).
18. Simões, E. A. et al. A predictive model for respiratory syncytial virus (RSV) hospitalisation of premature infants born at 33–35 weeks of gestational age, based on data from the Spanish FLIP study. *Respir. Res.* **9**, 78 (2008).
19. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann. Intern. Med.* **162**, 55 (2015).
20. HCUP. NIS overview. https://www.hcup-us.ahrq.gov/nisoverview.jsp (2021).
21. Cai, W. et al. Evaluation of using ICD-10 code data for respiratory syncytial virus surveillance. *Influenza Other Respir. Viruses* **14**, 630–637 (2020).
22. Bont, L. et al. Defining the epidemiology and burden of severe respiratory syncytial virus infection among infants and children in western countries. *Infect. Dis. Ther.* **5**, 271–298 (2016).
23. Boyce, T. G., Mellen, B. G., Mitchel, E. F., Wright, P. F. & Griffin, M. R. Rates of hospitalization for respiratory syncytial virus infection among children in Medicaid. *J. Pediatr.* **137**, 865–870 (2000).
24. Holberg, C. J. et al. Risk factors for respiratory syncytial virus-associated lower respiratory illnesses in the first year of life. *Am. J. Epidemiol.* **133**, 1135–1151 (1991).
25. Simoes, E. A. F. Environmental and demographic risk factors for respiratory syncytial virus lower respiratory tract disease. *J. Pediatr.* **143**, 118–126 (2003).
26. Reeve, C. A. et al. Predicting respiratory syncytial virus hospitalisation in Australian children. *J. Paediatr. Child Health* **42**, 248–252 (2006).
27. Kaiser, S. V. et al. Risk factors for prolonged length of stay or complications during pediatric respiratory hospitalizations. *Hosp. Pediatr.* **5**, 461–473 (2015).
28. Mansfield, E. R. & Helms, B. P. Detecting multicollinearity. *Am. Stat.* **36**, 158–160 (1982).
29. Steyerberg, E. W. et al. Internal validation of predictive models. *J. Clin. Epidemiol.* **54**, 774–781 (2001).
30. Steyerberg, E. W. in *Clinical Prediction Models* 83–100 (Springer New York, 2009).
31. Rodríguez-Martínez, C. E., Sossa-Briceño, M. P. & Nino, G. Predictors of prolonged length of hospital stay for infants with bronchiolitis. *J. Investig. Med.* **66**, 986 (2018).
32. Moler, F. W. et al. Respiratory syncytial virus morbidity and mortality estimates in congenital heart disease patients: a recent experience. *Crit. Care Med.* **20**, 1406–1413 (1992).
33. Doyle, C. M., Lix, L. M., Hemmelgarn, B. R., Paterson, J. M. & Renoux, C. Data variability across Canadian administrative health databases: differences in content, coding, and completeness. *Pharmacoepidemiol. Drug Saf.* **29**, 68–77 (2020).
34. Metcalfe, A., Sibbald, B., Lowry, R. B., Tough, S. & Bernier, F. P. Validation of congenital anomaly coding in Canada's administrative databases compared with a congenital anomaly registry. *Birth Defects Res. A Clin. Mol. Teratol.* **100**, 59–66 (2014).

35. Debray, T. P. A. et al. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J. Clin. Epidemiol.* **68**, 279–289 (2015).
36. Houchens, R. Missing data methods for the NIS and the SID. http://www.hcup-us.ahrq.gov/reports/methods/methods.jsp (2015).
37. Nakamura, M. M. et al. Pediatric readmissions after hospitalizations for lower respiratory infections. *Pediatrics* **140**, e20160938 (2017).
38. Vizcarra-Ugalde, S. et al. Intensive care unit admission and death rates of infants admitted with respiratory syncytial virus lower respiratory tract infection in Mexico. *Pediatr. Infect. Dis. J.* **35**, 1199–1203 (2016).
39. Adler-Milstein, J. et al. Electronic health record adoption in US hospitals: the emergence of a digital "advanced use" divide. *J. Am. Med. Inform. Assoc.* **24**, 1142–1148 (2017).

AUTHOR CONTRIBUTIONS

K.W. conceptualized and designed the study, conducted the analysis, drafted the initial manuscript, and reviewed and revised the manuscript. T.P. contributed to the analysis and reviewed the manuscript. D.B. participated in the design of the study and reviewed the manuscript. A.M. participated in the design of the study, obtained the data, contributed to the analysis, and reviewed the manuscript. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work. The work is original, not previously published, and has not been submitted for publication or consideration elsewhere.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Patient consent was not required.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41390-022-02008-9.

Correspondence and requests for materials should be addressed to Krista Wollny.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.