ORIGINAL ARTICLE: RESPIRATORY INFECTIONS

WILEY

Risk scoring tool to predict respiratory syncytial virus hospitalisation in premature infants

Maarten O. Blanken MD[1] | Bosco Paes FRCPC[2] | Evan J. Anderson MD[3] | Marcello Lanari MD[4] | Margaret Sheridan-Pereira FRCPI[5] | Scot Buchan BSc[6] | John R. Fullarton PhD[6] | ElizaBeth Grubb PhD[7] | Gerard Notario MD[8] | Barry S. Rodgers-Gray PhD[6] | Xavier Carbonell-Estrany MD[9]

[1] Division of Pediatric Immunology and Infectious Diseases, University Medical Center Utrecht, Utrecht, The Netherlands

[2] Neonatal Division, Department of Pediatrics, McMaster University, Hamilton, Canada

[3] Departments of Pediatrics and Medicine, Emory University School of Medicine, Atlanta, Georgia

[4] Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy

[5] Department of Paediatrics and Newborn Medicine, Coombe Women and Infants University Hospital and Department of Paediatrics, Trinity College, Dublin, Ireland

[6] Strategen Limited, Basingstoke, UK

[7] Health Economics and Outcomes Research, AbbVie Inc, North Chicago, Illinois

[8] Formerly Global Pharmaceutical Research and Development, AbbVie Inc, North Chicago, Illinois

[9] Neonatology Service, Hospital Clinic, Institut d'Investigacions Biomediques August Pi Suñer (IDIBAPS), Barcelona, Spain

Correspondence
Maarten O. Blanken, Division of Pediatric Immunology and Infectious Diseases, University Medical Center Utrecht, Utrecht, The Netherlands.
Email: m.o.blanken-2@umcutrecht.nl

Abstract

Background: The objective was to develop a risk scoring tool which predicts respiratory syncytial virus hospitalisation (RSVH) in moderate-late preterm infants (32-35 weeks' gestational age) in the Northern Hemisphere.

Methods: Risk factors for RSVH were pooled from six observational studies of infants born 32 weeks and 0 days to 35 weeks and 6 days without comorbidity from 2000 to 2014. Of 13 475 infants, 484 had RSVH in the first year of life. Logistic regression was used to identify the most predictive risk factors, based on area under the receiver operating characteristic curve (AUROC). The model was validated internally by 100-fold bootstrapping and externally with data from a seventh observational study. The model coefficients were converted into rounded multipliers, stratified into risk groups, and number needed to treat (NNT) calculated.

Results: The risk factors identified in the model included (i) proximity of birth to the RSV season; (ii) second-hand smoke exposure; and (iii) siblings and/or daycare. The AUROC was 0.773 (sensitivity: 68.9%; specificity: 73.0%). The mean AUROC from internal bootstrapping was 0.773. For external validation with data from Ireland, the AUROC was 0.707 using Irish coefficients and 0.681 using source model coefficients. Cut-off scores for RSVH were ≤19 for low- (1.0%), 20-45 for moderate- (3.3%), and 50-56 (9.5%) for high-risk infants. The high-risk group captured 62.0% of RSVHs within 23.6% of the total population (NNT 15.3).

Conclusions: This risk scoring tool has good predictive accuracy and can improve targeting for RSVH prevention in moderate-late preterm infants.

KEYWORDS
bronchiolitis, lower respiratory tract infection, prematurity, risk assessment, risk factors

1 | INTRODUCTION

Respiratory syncytial virus (RSV) is the predominant cause of lower respiratory tract infection (LRTI) in early childhood, accounting for 340 000 hospitalisations annually in children <5 years in industrialised countries.[1,2] It places a considerable strain on healthcare services, particularly during the winter months when the virus is most prevalent, with costs estimated at $545 million in the United States alone in 2009.[3] Moderate-late preterm infants (defined as 32 to 33-35 weeks' completed gestation at birth [wGA]) are at higher risk of severe RSV LRTI and greater morbidity than full-term infants.[4] Studies show that they also incur higher healthcare utilisation costs over the first 2 years of life,[5,6] and more frequent recurrent wheezing through 6 years of age compared to non-RSV hospitalised infants.[7] A pooled-analysis of seven prospective, observational studies comprising 7820 infants born at 33-35 wGA during the RSV season, reported an incidence rate of 3.4% for first confirmed RSV hospitalisation (RSVH), with 22.2% requiring intensive care and 12.7% needing mechanical ventilation.[8]

At present, palivizumab is the only licensed therapy for reducing RSVH rates,[9,10] though there are several new monoclonal antibodies on the horizon.[11,12] In order to effectively manage healthcare budgets, sub-populations of moderate-late preterms at particular risk need to be identified for intervention.[13,14] Large studies across the Northern Hemisphere have established risk factors associated with severe RSV LRTI in moderate-late preterm infants, including those related to RSV exposure (eg, daycare attendance), biological factors (eg, male sex), and social/environmental factors (eg, exposure to tobacco smoke).[15–21] Several risk scoring tools (RST) using data from these studies, identify moderate-late preterm infants at risk for RSVH in order to target RSV prophylaxis judiciously.[13,14,22–24] The models demonstrate good sensitivity (~70%) and specificity (~70%),[13,14,22,23] with the Canadian model proven to be cost-effective in clinical practice.[25,26] A model for general applicability across multiple countries has not been developed. The objective of the current study was to use a pooled dataset of studies to develop a simple and validated risk factor tool with improved performance, applicable across the Northern Hemisphere.

2 | METHODS

2.1 | Pooled dataset used for modelling

Individual patient-linked data from six prospective, observational studies across the Northern Hemisphere were used to develop the predictive model underpinning the RST: 'Risk Factors Linked to Respiratory Syncytial Virus Infection Requiring Hospitalization in Premature Infants Study' (FLIP-2, Spain)[17]; 'RISK' (the Netherlands)[13]; 'Pediatric Investigators Collaborative Network on Infections in Canada' (PICNIC, Canada)[15]; 'Italian National Birth Cohort' (IBC, Italy)[19]; 'Respiratory Syncytial Virus (RSV) Respiratory Events Among Preterm Infants Outcomes and Risk Tracking Study' (REPORT, USA)[18]; and 'Predictors Associated with RSV Hospitalization in Nonprophy-laxed, Premature Infants' (PONI, multinational)[20] (Table 1). These studies had been previously identified by a systematic review of the

TABLE 1 Overview of studies included in developing and validating the risk factor model

| Demographics | | | | | Exclusions | | | | | |
Dataset	Country	Study years	Duration of follow-up	RSV season	wGA[a] range	Total N	wGA[a] >35[6]	Comorbid	Prophylaxed	Other	N included
PICNIC[15]	Canada	2000-2002	1 month post-RSV season	1 Nov-30 Apr	33⁰-35⁶	1758	0	232	0	9[b]	1517
FLIP-2[17]	Spain	2005-2007	End of May	1 Oct-30 Apr	32¹-35⁰	5441	0	193	693	0	4555
RISK[13]	The Netherlands	2008-2012	1 year of age	1 Oct-31 Mar	32¹-35⁶	2421	0	0	0	11[c]	2410
REPORT[18]	United States	2009-2011	End of May	1 Nov-31 Mar	32⁰-35⁶	1642	0	0	0	0	1642
IBC[19]	Italy	2009-2013	1 year of age	1 Nov-31 Mar	≥33⁰	2210	1184	0	0	0	1026
PONI[20]	Europe, America Asia/Middle East[d]	2013-2014	End of Apr	1 Oct-30 Apr	33⁰-35⁶	2390	0	65	0	0	2325
Total (pooled dataset)						15 862	1184	490	693	20	13 475
PREMI[21]	Ireland	2011-2014	1 year of age	1 Oct-31 Mar	32⁰-36⁶	1807	718	11	0	0	1078
Total (validation dataset)						1807	718	11	0	0	1078

[a]wGA, weeks' gestational age.
[b]Nine cases not hospital confirmed.
[c]No recorded wGA.
[d]Twenty-three countries in Western Europe (Austria, France, Norway, Portugal, Sweden, and Switzerland), Eastern Europe (Bosnia, Bulgaria, Czech Republic, Estonia, Latvia, Lithuania, Slovakia and Slovenia) and Russia, South Korea, Mexico and the Middle East (Bahrain, Egypt, Jordan, Lebanon, Oman and Saudi Arabia).

literature undertaken in 2015.[8] The key inclusion criteria for studies were: multicentre, observational, prospective design; assessed >1000 moderate-late preterm (32-35 wGA) infants at risk for severe RSV disease (defined as the need for hospitalisation); included infants with laboratory-confirmed RSV infection; and ≤15% of infants received palivizumab prophylaxis (to ensure a standardised and unbiased population). An updated search of the literature (to 18 December 2017) identified no additional studies meeting the inclusion criteria.

2.2 | Data extraction, recasting, verification and analysis

Data for infants (≤1 year) born at 32 weeks and 0 days (32^0) to 35 weeks and 6 days (35^6) gestation were extracted from each study, including information on first confirmed RSVH and corresponding risk factors. To ensure homogeneity, infants were excluded if they were born at $<32^0$ or $>35^6$ wGA, had received RSV prophylaxis, or had a relevant comorbidity (eg, congenital heart disease, bronchopulmonary dysplasia/chronic lung disease). All data were anonymised. To ensure sufficient data for analysis, the collection/recording of a risk variable in at least four studies was a requisite for inclusion in the pooled dataset. Included risk factors were recast, where necessary, into a common format across studies. To verify each study's data before inclusion, the extracted datasets were checked and approved by key study investigators and personnel (XCE, MB, BP, ML, EJA; also see Acknowledgments section). The quantity of data available for three risk factor variables from each dataset were further confirmed against the original study publication. A heterogeneity test for the dichotomous variables present in all contributory datasets was performed by comparing odds ratios (ORs) using the Breslaw-Day method. For categoric variables (>2 categories), data were converted to ranks and analysis of variance (ANOVA) conducted on the differences from mean rank in hospitalised and non-hospitalised infants. Heterogeneity for continuous variables was assessed by comparing the significance of difference between hospitalised and non-hospitalised infants using parametric t-test. Statistical significance of individual variables in the pooled dataset was assessed by two-tailed t-test (parametric data) and Mann-Whitney U-test and Mantel-Haenzel test (categoric data).

2.3 | Development of the predictive model

Logistic regression was used to develop a preliminary risk factor model that included all risk factors in the pooled dataset. RSVH was the dependent variable and the risk factors were the covariates. Where risk factor data were missing for an infant, average values for that dataset were used, or when all values for a particular risk factor were missing from a dataset, the combined data average were applied. Alternative approaches using a new category for a missing value or neutral, non-discriminatory values were also tested. The model was optimised by several mechanisms: (i) sequential removal and reinsertion of each risk factor variable from the dataset to establish its impact on predicting RSVH; (ii) using Wald test significance and exp(beta) to determine which covariates to test at each stage of removal; (iii)

assessing risk factors in combination versus use as individual predictors; and (iv) assessing different cut-off values for risk factors, where applicable. Risk factors were expressed as either dichotomous (ie, yes/no) or, if used in combination, categorical (ie, neither, one, both, etc) variables. The overall goal was to find the combination of risk factors that provided the best balance between predictive accuracy and simplicity in terms of number and type of risk factors. Predictive accuracy was assessed by a receiver operating characteristic (ROC) curve, plotting sensitivity against 1-specificity, with an area under the ROC curve (AUROC) of ≥0.75 considered 'good'.[27] The point of maximum sensitivity and specificity was also calculated for the final model using the Youden's J statistic. Lastly, for each variable in the final model, the increased adjusted risk of RSVH was expressed as an OR.

2.4 | Validation of the final model

Three main approaches were used to validate the final model. First, the model was generated in the FLIP-2,[17] PICNIC,[15] RISK[13] and PONI[20] datasets and compared to the published models for those studies (IBC[19] and REPORT[18] do not have published models).[13,22,24] Second, 100-fold bootstrapping validation was performed on the pooled dataset.[28] The pooled dataset was sampled with replacement 100 times and the model coefficients used to calculate the predictive probabilities for each case in the 100 samples. ROC curves were constructed for each sample, the AUROC values calculated, and the dispersion statistics (standard deviation and range) across the 100 samples assessed. A low level of dispersion indicates an internally consistent model. The Kolmogorov-Smirnov test was used to assess normality in the distribution of AUROCs from the samples (non-significance indicates a normal distribution) and skewness was also calculated (0.0 = absolute symmetry). Finally, the model was validated externally against data from the recently published RSV Preterm Risk Estimation Measure for RSVH in Ireland study (RSV-PREMI),[21] which was identified in the same systematic review as the studies in the pooled dataset (Table 1).[8] Data were verified by study personnel (MS-P and Acknowledgments), including three variables checked against the study publication, and heterogeneity assessed as previously described. The model was tested in two ways against the RSV-PREMI data: (i) generating a model from the RSV-PREMI data itself using the same risk factors as for the final model and (ii) the coefficients from the pooled dataset were applied to the RSV-PREMI data. For both analyses, predictive accuracy was assessed by AUROC.

2.5 | Development of the RST

To convert the final model into a RST, the logistic regression coefficient(s) for each variable was assigned a rounded multiplier with a positive value. The rounded multiplier provides a measure of the influence of a particular risk factor on the probability of RSVH relative to that of the other risk factors in the model (the higher the value, the greater the influence). The sum of the rounded multipliers, taking into consideration any categorical variables that may have more than one multiplier, represented the maximum score of the tool.

Cut-off scores for low-, moderate- and high-risk groups were determined based on RSVH rates of <2%, 2-10% and >10%, respectively, in line with the RSTs developed in Canada[22] and the Netherlands[13,14] (the FLIP-2[24] and PONI[20] models did not include cut-offs). The RSVH rate was also plotted against the risk score to determine if there were any apparent inflections in the curve from which to refine the cut-off values. A very high-risk group was defined by examining a score that would limit the RST to capturing approximately 10% of the total population. The relative risk and ORs for RSVH were compared between risk groups, positive predictive values (PPV) and negative predictive values (NPV) determined, and numbers needed to treat (NNT) calculated, assuming a palivizumab efficacy rate of 80% for 32-35 wGA infants, based on randomised controlled trials.[9,29]

All analyses were performed using SPSS for Windows version 15.0 (IBM Corporation, New York, NY), Microsoft Access 2010 SQL (Microsoft Corporation, WA) and Microsoft Access/Excel VBScript 2010 (Microsoft Corporation).

2.6 | Transparency of reporting

The *Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis* (TRIPOD) statement was followed for this manuscript (Supplementary Table S1).[30] The TRIPOD statement provides a framework for the full and clear reporting of a prediction model study, such that risk of bias and potential usefulness can be adequately assessed.[30]

3 | RESULTS

3.1 | Pooled dataset

The six studies (FLIP-2,[17] RISK,[13] PICNIC,[15] IBC,[19] REPORT,[18] PONI[20]) contained individual patient-linked data collected from 2000 to 2014 for a total of 15 862 infants, of whom 13 475 were born between 32^0 and 35^6 wGA and met the inclusion criteria for the pooled dataset (Table 1). The primary reasons for exclusion were birth ≥36 wGA (*n* = 1184), receiving RSV prophylaxis (*n* = 693), and having an exclusionary comorbidity (*n* = 490). Each study contributed at least 1000 infants to the pooled dataset, with all providing data for infants born at 33-35 wGA and three studies contributing data as well for 32 wGA infants (FLIP-2,[17] RISK,[13] REPORT[18]). The overall distribution by wGA was 32 wGA (6.9%), 33 wGA (24.4%), 34 wGA (38.1%) and 35 wGA (30.7%).

Of the 13 475 infants in the pooled dataset, 484 (3.6%) had a confirmed RSVH within the first year of life. A total of 18 possible risk factors for RSVH were present in four of the six studies and were recast to a common format (Supplementary Table S2). Prior to inclusion in the pooled dataset, the extracted data for each study were confirmed and verified against the published data with no apparent discrepancies (Supplementary Table S3). Heterogeneity tests revealed no significant differences for 11 of the 12 risk factor variables present in all six datasets; smokers in the household differed significantly (*P* = 0.04) between studies, with rates varying between 4 and 67% across studies (Supplementary Table S4).

3.2 | Risk factor model

The final logistic regression model comprised three variables, combining a total of five risk factors: birth between 3 months before and 2 months after season start date; smokers in the household and/ or maternal smoking whilst pregnant; and siblings (excluding multiple births) and/or daycare attendance (recorded as 'planned', reflecting how the RST would be used in practice). Treating all risk factors as categorical covariates (ie, assigning into groups and treating as non-linear scales), the derived model had an AUROC of 0.773 (95% confidence interval [CI] 0.753-0.792) and a maximum sensitivity and specificity of 0.689 and 0.730, respectively (Figure 1). The most predictive variable was the combination of siblings and daycare, though age relative to the start of the RSV season was the single most powerful risk factor (Table 2). Refining the siblings variable to pre-school age (<6 years), which is a highly significant risk factor for RSVH,[18,20] increased the AUROC minimally to 0.775. It was considered more practical to exclude a sibling age criterion, particularly when 'pre-school age' is defined differently across countries. Substituting (any) siblings for a broader 'crowding' variable of >4 in the household including infant, >4 being the most predictive cut-off, or adding this variable to the model did not increase overall predictive accuracy (AUROC 0.764 for both substitution and addition). Unlike the other five datasets, PONI recorded only month (not day) of birth.[20] The age variable birth between 3 months before and 2 months after season start date was intended to simplify the calculated 13 weeks before to 8.5 weeks after the start of the RSV season. The use of a new category or imputation of neutral, non-discriminatory values for missing data resulted in models with similar discrimination (new category, AUROC 0.773; non-discriminatory, AUROC 0.770), confirming the absence of unrecognised bias associated with using average values.
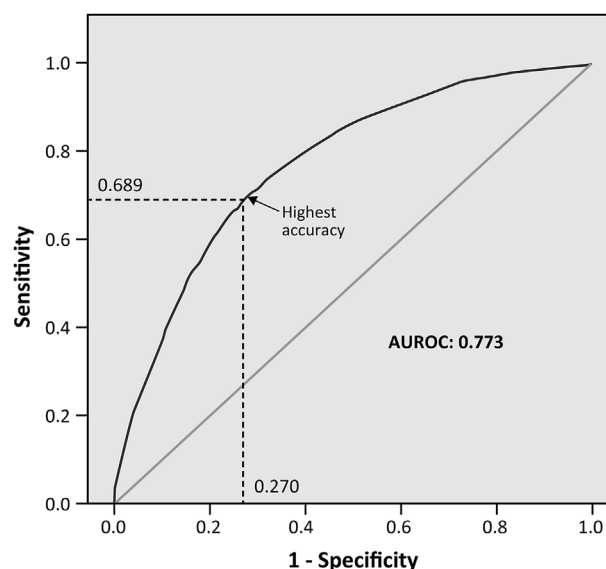

FIGURE 1 Receiver operating characteristic (ROC) curve for the final three-variable model derived from the pooled dataset

TABLE 2 Variables in the final logistic regression model for the risk scoring tool derived from the pooled dataset

Variable	Odds ratio (95%CI), P-value[a]	Logistic regression coefficient	Score (rounded integer)
Birth between 3 months before and 2 months after season start date [yes or no]	2.0 (1.7-2.5), P < 0.001	0.338	6
Smokers in household and/or while pregnant [neither, either or both]	Household: 1.4 (1.2-1.7), P = 0.001 Pregnant: 1.7 (1.3-2.1); P < 0.001	Either: 0.209 Both: 0.479	Either: 5 Both: 11
Siblings (excluding multiple birth siblings) and/or (planned) day care [neither, either or both]	Siblings: 1.6 (1.4-2.0), P < 0.001 Daycare: 1.6 (1.3-1.9), P < 0.001	Either: 0.740 Both: 1.639	Either: 14 Both: 39

[a]Increased adjusted risk of respiratory syncytial virus hospitalisation for individual variables.

3.3 | Validation of the risk factor model

3.3.1 | Generation of the model in individual datasets

Generating the final model in the individual datasets resulted in functions that were more powerful in FLIP-2: AUROC 0.762 versus 0.687,[24] respectively, and in the other cases was within 3-12% of the predictive power of the published models (PICNIC: 0.673 vs 0.762[22]; RISK: 0.680 vs 0.703;[13] PONI: 0.701 vs 0.755[20]) (Supplementary Table S5).

3.3.2 | Internal validation

The bootstrap validation resulted in a tight distribution of results for the 100 samples (total of ~1.35 million infants), with the median AUROC being 0.773 (range 0.753-0.805; interquartile range 0.01) (Supplementary Figure S1). The Kolmogorov-Smirnov test indicated that the distribution of AUROCs from the samples was normal (0.059, degrees of freedom 100; P = 0.200), whilst the Skewness statistic showed a symmetrical distribution containing a slightly greater number of larger values (0.322 ± 0.241).

3.3.3 | External validation

RSV-PREMI[21] included 1078 infants born 32^0-35^6 wGA of whom 46 (4.3%) were hospitalised with RSV LRTI in the first year of life (Table 1). All risk factors comprising the final model were available in RSV-PREMI and were recast in exactly the same format as the pooled dataset. Analysis revealed no apparent discrepancies between the extracted and published data for RSV-PREMI (Supplementary Table S3). The risk factors in the final model were shown to behave similarly within RSV-PREMI and the pooled dataset (Supplementary Table S4).

Generating a model in the RSV-PREMI[21] data comprised of the risk factors included in the final model produced an AUROC of 0.707 (95%CI 0.637-0.778) (Supplementary Figure S2A). Applying the coefficients from the final model from the pooled dataset to the RSV-PREMI data resulted in an AUROC of 0.681 (95%CI 0.588-0.773) (Supplementary Figure S2B).

3.4 | RST

Converting the logistic regression coefficients for each variable in the final model into rounded multipliers resulted in a maximum risk score of 56 (Table 2 and Figures 2A and 2B). The RST was created as a nomogram with a score ≤19 representing a low-risk of RSVH (average risk 1.0%), 20-45 representing a moderate-risk (average risk 3.3%) and ≥50 representing a high-risk (average risk 9.5%). Plotting the RSVH rate against the risk score resulted in a curve with a natural inflection at a score of ~45 (Supplementary Figure S3). This was set as the medium/high-risk boundary. The high-risk group identified 62.0% of all RSVHs whilst selecting 23.6% of the total study population. The corresponding figures for the moderate- and low-risk groups were 23.2%/25.1% and 14.8%/51.3%, respectively (Figure 3). The high- and moderate-risk groups both had a significantly higher RSVH risk than the low risk group (OR 10.1, 95%CI 7.9-12.9, P < 0.001; and OR 3.3, 95%CI 2.5-4.4, P < 0.001, respectively; combined high- and moderate-risk: OR 6.4, 95%CI 5.1-8.2, P < 0.001). The NNT for the high-risk group was 15.3, while the combined high- and moderate-risk group had a NNT of 33.3. A very high-risk group was defined as a score of 56, which captured 39.3% of RSVHs whilst selecting 11.9% of the total population, with a corresponding NNT of 10.8.

4 | DISCUSSION

A simple RST was developed for predicting the risk of RSVH in moderate-late (32^0-35^6 wGA) preterm infants in the Northern Hemisphere, from six large datasets and validated in a seventh large dataset. Three risk factor variables—birth between 3 months before and 2 months after season start date, smokers in the household and/or maternal smoking whilst pregnant, and siblings (excluding multiples) and/or (planned) daycare attendance—were shown to accurately and reliably predict RSVH. The RST is practical and can facilitate decision making for clinicians, parents and policy makers regarding RSV prophylaxis. Importantly, two out of the five identified risk factors in our model—smoking in the household and daycare—are modifiable and the tool could be used accordingly to educate parents.

The model underpinning the RST compares favourably in terms of simplicity and predictive accuracy with other published models in moderate-late preterm infants, including those contained within the pooled dataset: AUROC of 0.773 with three variables versus 0.791 with seven variables (Spanish [FLIP][23]); 0.762 with seven variables (Canadian [PICNIC][22]); 0.755 with six variables (PONI[20]); 0.72 with five variables (Dutch [RISK-II][14]); 0.703 with four variables (Dutch
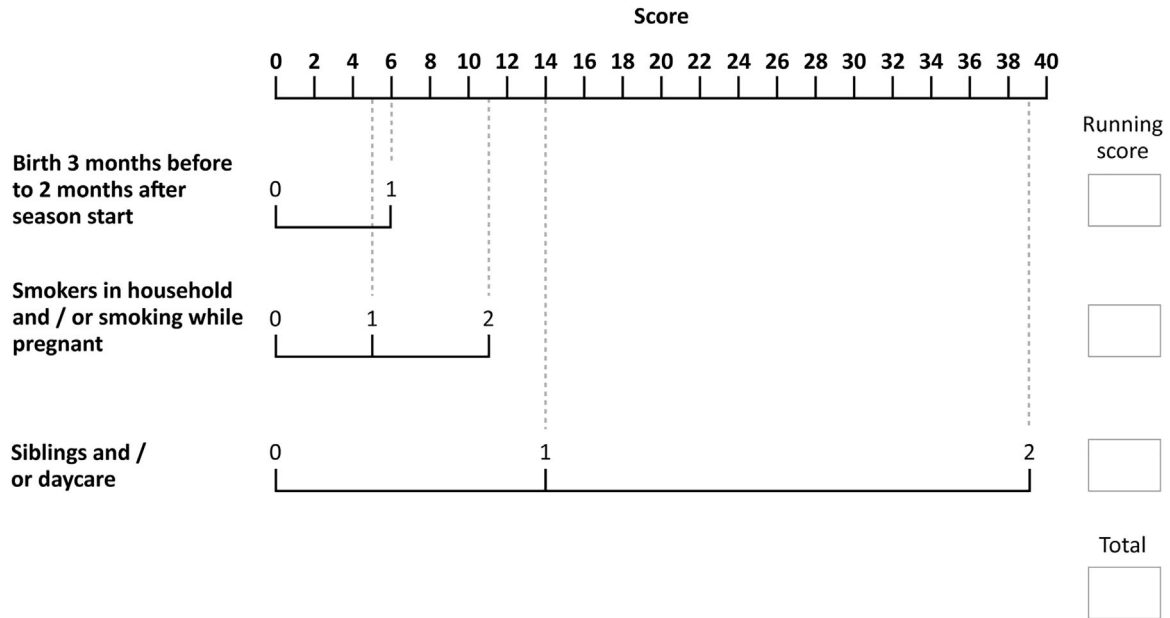
Score

FIGURE 2 Risk factor scoring tool. Key: 0 = no/not present; 1 = yes/present for one risk factor; 2 = yes/present for both risk factors

[RISK][13]) and 0.687 with four variables (Spanish [FLIP-2][24]). All of the models included variables associated with age relative to the RSV season and siblings/daycare, highlighting the importance of these risk factors in determining RSVH risk. The combination of siblings and daycare is particularly powerful and non-linear (individual score: 14 vs combined score: 39), suggesting that these risk factors reinforce each other in terms of exposure to RSV and in combination, increase discrimination in the model. Smoking, the other risk factor included in the pooled model, was also part of previously published models (FLIP-2,[24] PICNIC,[22] and PONI[20]). The combined smoking variable is approximately linear and less powerful (individual score: 5; combined score: 11) than siblings/daycare, despite similar ORs (1.4-1.7 vs 1.6,

respectively). This may partly be due to greater overlap in the variance explained by the two smoking risk factors within the model, since average values were imputed for smoking whilst pregnant in PICNIC[15] and REPORT,[18] which only captured smokers in the household. Combined with the validation against the RSV-PREMI dataset and the homogeneity of risk factor data across all studies, this reinforces the universal applicability of the RST across the Northern Hemisphere.

The key strength of this RST was the development from a pooled dataset of six independent, multicentre, observational, prospective studies involving >14 500 infants with both internal and external validation. However, certain limitations should be addressed. The
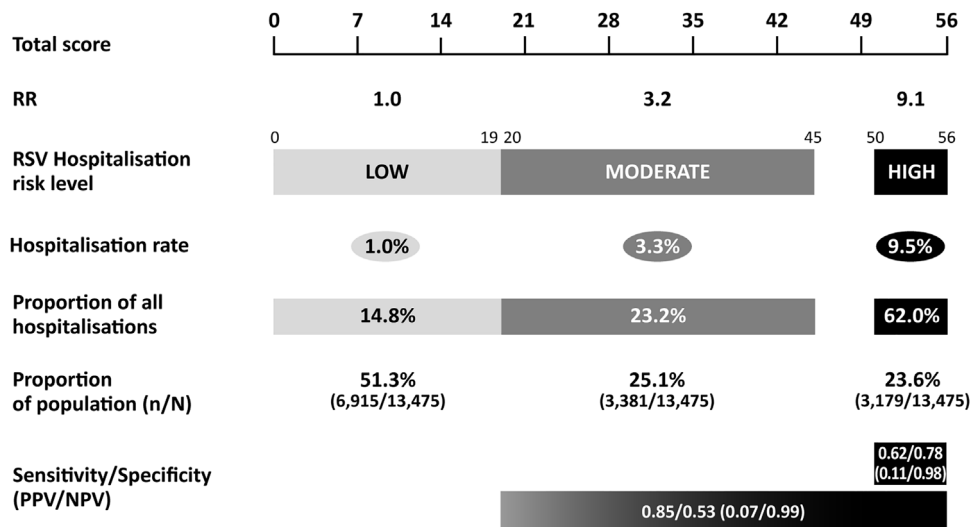

FIGURE 3 Interpretation of risk score and risk group characteristics (Please note that it is not possible to achieve a score of 46-49 based on the individual variable scores)

individual studies varied in objectives and design, which influenced the included gestational age range of infants and how and what risk factors were collected. Of the six studies, only three included data on 32 wGA infants, but these represented Europe (FLIP-2,[17] RISK[13]) and North America (REPORT[18]). In total, >900 32 wGA infants were included in the pooled dataset and, importantly, the RSV-PREMI[21] validation dataset involved 32 wGA infants. Whilst the FLIP-2[17] dataset provided around one-third of infants in the pooled dataset, each study contributed >1000 infants. Recasting risk factors to a simpler, common format results in loss of some statistical power; however, this was justified by the objective to create a user-friendly tool. All of the risk factors in the final model were available in all the datasets, except for smoking whilst pregnant. The PONI[20] dataset captured only month not day of birth, which could have weakened the birth between 3 months before and 2 months after season start date variable, although rounding to whole months helped to mitigate this effect. The studies spanned 15 years (2000-2014), with likely variations in hospital practice and RSV testing. Our ability to develop a robust predictive model suggests intrinsic compatibility amongst the datasets and supports the high predictive value of these risk factors. The internal and external validations demonstrated that the model is internally consistent, not overly optimistic (ie, there is little or no over-fitting), and can be applied effectively across the Northern Hemisphere.

The RST has a scale of 0-56 with defined cut-off scores for low- (≤19), moderate- (20-45) and high-risk (≥50) infants. The cumulative RSVH risk was 3.6% (484/13 475) in the pooled dataset, with the combined moderate- and high-risk groups being 6.3%, the high-risk group 9.5% and the very high-risk group (score of 56) 11.9%. The NNT for the combined high- and moderate-risk groups was 33.3, which falls to 15.3 in the high-risk group and 10.8 for very high-risk infants. A balance must be struck between the cost-effectiveness of palivizumab versus potential therapeutic benefits, with the very high-risk group having a compelling NNT, but missing 60% of predicted RSVHs. Ultimately, the final decision regarding appropriate cut-offs should be made locally, taking into consideration the overall risk-cost-benefit relative to each clinical setting.

The validated RST described herein is simple and has good predictive accuracy to assess RSVH risk in moderate-late preterm infants. Developing the tool from six datasets confirms its predictive capabilities, generalisability and applicability across the Northern Hemisphere. The RST is a powerful instrument to determine RSVH risk and direct RSV therapies cost-effectively to the most vulnerable moderate-late preterm infants.

ORCID

Maarten O. Blanken [ID] http://orcid.org/0000-0003-4065-7368

REFERENCES

1. Shi T, McAllister DA, O'Brien KL, 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*. 2017;390:946–958.
2. Jain S, Williams DJ, Arnold SR, et al. Community-acquired pneumonia requiring hospitalization among U.S. children. *N Engl J Med*. 2015;372: 835–845.
3. Hasegawa K, Tsugawa Y, Brown DF, Mansbach JM, Camargo CA, Jr. Trends in bronchiolitis hospitalizations in the United States, 2000–2009. *Pediatrics*. 2013;132:28–36.
4. Figueras-Aloy J, Manzoni P, Paes B, et al. Defining the risk and associated morbidity and mortality of severe respiratory syncytial virus infection among preterm infants without chronic lung disease or congenital heart disease. *Infect Dis Ther*. 2016;5:417–452.
5. Palmer L, Hall CB, Katkin JP, et al. Respiratory outcomes, utilization and cost 12 months following a respiratory syncytial virus diagnosis among commercially insured late-preterm infants. *Curr Med Res Opin*. 2011;27:403–412.
6. Shefali-Patel D, Paris MA, Watson F, Peacock JL, Campbell M, Greenough A. RSV hospitalisation and healthcare utilisation in moderately prematurely born infants. *Eur J Pediatr*. 2012;171: 1055–1061.
7. Carbonell-Estrany X, Pérez-Yarza EG, García LS, et al. Long-term burden and respiratory effects of respiratory syncytial virus hospitalization in preterm infants—the SPRING study. *PLoS ONE*. 2015;10: e0125422.
8. Anderson EJ, Carbonell-Estrany X, Blanken M, et al. Burden of severe respiratory syncytial virus disease among 33-35 weeks gestational age infants born during multiple respiratory syncytial virus seasons. *Pediatr Infect Dis J*. 2017;36:160–167.
9. The IMpact-RSV Study Group. Palivizumab, a humanized respiratory syncytial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high-risk infants. *Pediatrics*. 1998;102:531–537.

10. Feltes TF, Cabalka AK, Meissner HC, et al. Palivizumab prophylaxis reduces hospitalization due to respiratory syncytial virus in young children with hemodynamically significant congenital heart disease. *J Pediatr*. 2003;143:532–540.

11. Simões EA, DeVincenzo JP, Boeckh M, et al. Challenges and opportunities in developing respiratory syncytial virus therapeutics. *J Infect Dis*. 2015;211:S1–S20.

12. Rezaee F, Linfield DT, Harford TJ, Piedimonte G. Ongoing developments in RSV prophylaxis: a clinician's analysis. *Curr Opin Virol*. 2017;24:70–78.

13. Blanken MO, Koffijberg H, Nibbelke EE, Rovers MM, Bont L, Dutch RSV Neonatal Network. Prospective validation of a prognostic model for respiratory syncytial virus bronchiolitis in late preterm infants: a multicenter birth cohort study. *PLoS ONE*. 2013;8:e59161.

14. Korsten K, Blanken MO, Nibbelke EE, Moons KG, Bont L, Dutch RSV Neonatal Network. Prediction model of RSV-hospitalization in late preterm infants: an update and validation study. *Early Hum Dev*. 2016;95:35–40.

15. Law BJ, Langley JM, Allen U, et al. The Pediatric Investigators Collaborative Network on Infections in Canada study of predictors of hospitalization for respiratory syncytial virus infection for infants born at 33 through 35 completed weeks of gestation. *Pediatr Infect Dis J*. 2004;23:806–814.

16. Figueras-Aloy J, Carbonell-Estrany X, Quero J, IRIS Study Group. Case-control study of the risk factors linked to respiratory syncytial virus infection requiring hospitalization in premature infants born at a gestational age of 33-35 weeks in Spain. *Pediatr Infect Dis J*. 2004;23:815–820.

17. Figueras-Aloy J, Carbonell-Estrany X, Quero-Jimenez J, et al. FLIP-2 Study: risk factors linked to respiratory syncytial virus infection requiring hospitalization in premature infants born in Spain at a gestational age of 32 to 35 weeks. *Pediatr Infect Dis J*. 2008;27:788–793.

18. Ambrose CS, Anderson EJ, Simões EA, et al. Respiratory syncytial virus disease in preterm infants in the U.S. born at 32-35 weeks gestation not receiving immunoprophylaxis. *Pediatr Infect Dis J*. 2014;33:576–582.

19. Lanari M, Prinelli F, Adorni F, et al. Risk factors for bronchiolitis hospitalization during the first year of life in a multicentre Italian birth cohort. *Ital J Pediatr*. 2015;41:40.

20. Straňák Z, Saliba E, Kosma P, et al. Predictors of RSV LRTI hospitalization in infants born at 33 to 35 weeks gestational age: a large multinational study (PONI). *PLoS ONE*. 2016;11:e0157446.

21. Sheridan-Pereira M, Murphy J, Sloan J, et al. Respiratory syncytial virus preterm (32-36 completed weeks' gestation) risk estimation measure for RSV hospitalization in Ireland: a prospective study. *Pediatr Infect Dis J*. 2016;35:19–24.

22. Sampalis JS, Langley J, Carbonell-Estrany X, et al. Development and validation of a risk scoring tool to predict respiratory syncytial virus hospitalization in premature infants born at 33 through 35 completed weeks of gestation. *Med Decis Making*. 2008;28:471–480.

23. Simões EA, Carbonell-Estrany X, Fullarton JR, 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*. 2008;9:78.

24. Figueras-Aloy J, Quero-Jiménez J, Fernández-Colomer B, et al. Usefulness of different risk factor associations in predicting admissions due to respiratory syncytial virus in premature newborns of 32 to 35 weeks gestation in Spain. *An Pediatr (Barc)*. 2009;71:47–53.

25. Paes B, Steele S, Janes M, Pinelli J. Risk-scoring tool for respiratory syncytial virus prophylaxis in premature infants born at 33-35 completed weeks' gestational age in Canada. *Curr Med Res Opin*. 2009;25:1585–1591.

26. Lanctôt KL, Masoud ST, Paes BA, et al. The cost-effectiveness of palivizumab for respiratory syncytial virus prophylaxis in premature infants with a gestational age of 32-35 weeks: a Canadian-based analysis. *Curr Med Res Opin*. 2008;24:3223–3237.

27. Brubaker PH. Do not be statistically cenophobic: time to roc and roll! *J Cardiopulm Rehabil Prev*. 2008;28:420–421.

28. Efron B, Tibshirani RJ. *An introduction to the bootstrap*. London: Chapman and Hall; 1993.

29. Blanken MO, Rovers MM, Molenaar JM, et al. Respiratory syncytial virus and recurrent wheeze in healthy preterm infants. *N Engl J Med*. 2013;368:1791–1799.

30. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.