



101034339 – PROMISE

Preparing for RSV Immunisation and Surveillance in Europe

WP1 – RSV epidemiology and impact of COVID-19

## D1.1 Report on novel clinical scores for RSV severity

<b>Lead contributors</b>	Harish Nair (UEDIN – University of Edinburgh) Harish.Nair@ed.ac.uk
<b>Other contributors</b>	Zakariya Sheikh (UEDIN – University of Edinburgh) Ellie Potter (UEDIN – University of Edinburgh) You Li (NJMU – Nanjing Medical University) Simon Drysdale (Saint Georges University of London) Joanne Wildenbeest (UMCU - Universitair Medisch Centrum Utrecht) Hanna Robinson (UOXF - The Chancellor, Masters and Scholars of the University of Oxford) Joseph McGinley (SERGAS - Servizo Galego de Saúde) Gu-Lung Lin (UOXF - The Chancellor, Masters and Scholars of the University of Oxford (United Kingdom)) Deniz Öner (JANSSEN - Janssen Pharmaceutica, N.V) Federico Martín-Torres (UOXF - The Chancellor, Masters and Scholars of the University of Oxford) Jeroen Aerssens (JANSSEN - Janssen Pharmaceutica, N.V) Andrew J Pollard (UOXF - The Chancellor, Masters and Scholars of the University of Oxford) Louis Bont (UMCU - Universitair Medisch Centrum Utrecht)

Reproduction of this document or part of this document without PROMISE consortium permission is forbidden. Any use of any part must acknowledge the PROMISE consortium as “This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement 101034339. This Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and EFPIA”. This document is shared within the PROMISE Consortium and is in line with the general communication guidelines described in the PROMISE Consortium Agreement.

	Rachel Cohen (GSK- GlaxoSmithKline Biologicals, S.A.)
	Gaël Dos Santos (GSK- GlaxoSmithKline Biologicals, S.A.)
<b>Reviewers</b>	Clarisse Demont (Sanofi) John Paget (Nivel)

### Document History

Version	Date	Description
V0.1	27/09/2022	First Draft
V0.2	27/10/2022	Revised to comments from reviewers
V1.0	10/11/2022	Final document
V2.0	17/08/2023	Final document with clarification regarding patient population [i.e. restricted to young infants only]

#### Note

The DoA erroneously describes the deliverable D.1.1 to “summarise all available evidence (published literature and unpublished data) and develop estimates for RSV burden in pregnant women”. However, this deliverable D.1.1 was intended to review the evidence in published literature regarding existing severity scores for RSV / Bronchiolitis in young children and subsequently develop a simplified severity score using the RESCEU multi-country infant case control dataset and then validate this score. As is evident from the DoA, the description for deliverables D1.1 and D1.3 are identical whereas the two deliverables relate to separate tasks within WP1 (D1.1 to Task 1.1.2 and D1.3 to Task 1.1.1). The typographical error in the description of deliverable in DoA is deeply regretted.

## Table of contents

Table of contents .....	3
Definitions .....	4
Abbreviations.....	5
PART I .....	7
1. Abstract .....	8
2. Introduction .....	9
3. Methods.....	10
4. Results.....	13
5. Discussion .....	42
6. References .....	43
ANNEXES .....	47
ANNEX I. Search Strategies.....	48
ANNEX II. Detailed TRIPOD Scores.....	49
ANNEX III. Detailed PROBAST Scores .....	53
PART II .....	58
1. Abstract .....	59
2. Introduction .....	60
3. Methods.....	64
4. Results.....	68
5. Discussion .....	78
6. References .....	80
ANNEXES.....	82
ANNEX I. Detailed TRIPOD score .....	83
ANNEX II. Detailed PROBAST checklist.....	86
ANNEX III. Missing data for baseline characteristics .....	90

## Definitions

- **Participants** of the PROMISE Consortium are referred to herein according to the following codes:
  1. **UEDIN.** The University of Edinburgh (United Kingdom)
  2. **UMCU.** Universitair Medisch Centrum Utrecht (Netherlands)
  3. **UA.** Universiteit Antwerpen (Belgium)
  4. **Imperial.** Imperial College of Science, Technology and Medicine (United Kingdom)
  5. **UOXF.** The Chancellor, Masters and Scholars of the University of Oxford (United Kingdom)
  6. **THL.** Terveystieteiden tutkimuskeskus (Finland)
  7. **RIVM.** Rijksinstituut voor Volksgezondheid en Milieu (Netherlands)
  8. **NIVEL.** Stichting Nedelands Instituut voor Onderzoek van de Gezondheidszorg (Netherlands)
  9. **TUCH.** Varsinais-Suomen Sairaanhoidopiirin Kuntayhtymä (Finland)
  10. **TEAMIT.** TEAM IT Research, S.L. (Spain)
  11. **ReSViNET.** Stichting Resvinet (Netherlands)
  12. **SSI.** Statens Serum Institut (Denmark)
  13. **SER GAS.** Servizo Galego de Saúde (Spain)
  14. **PENTA.** Fondazione PENTA - For the treatment and care of children with HIV and related diseases - ONLUS (Italy)
  15. **FISABIO.** Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (Spain)
  16. **MLU.** Martin-Luther-Universitaet Halle-Wittenberg (Germany)
  17. **SP.** Sanofi Pasteur, S.A. (France)
  18. **GSK.** GlaxoSmithKline Biologicals, S.A. (Belgium)
  19. **JANSSEN.** Janssen Pharmaceutica, N.V (Belgium)
  20. **Novavax.** Novavax, Inc. (United States)
  21. **Pfizer.** Pfizer Limited (United Kingdom)
  22. **AZ.** AstraZeneca AB (Sweden)
  
- **Grant Agreement.** (Including its annexes and any amendments) The agreement signed between the beneficiaries of the action and the IMI2 JU for the undertaking of the PROMISE project (Grant Agreement No. 101034339).
- **Project.** The sum of all activities carried out in the framework of the Grant Agreement.
- **Work plan.** Schedule of tasks, deliverables, efforts, dates and responsibilities corresponding to the work to be carried out, as specified in Annex I to the Grant Agreement.
- **Consortium.** The PROMISE Consortium, comprising the above-mentioned participants.
- **Consortium Agreement.** The agreement concluded amongst PROMISE participants for the implementation of the Grant Agreement. The agreement shall not affect the parties' obligations to the Community and/or to one another arising from the Grant Agreement.

## Abbreviations

Acronym / Abbreviation	Meaning
AUROC	Area under the receiver operating characteristic curve
BROSJOD	Bronchiolitis Score of Sant Joan de Déu
BSS	Bronchiolitis severity score
CAS	Clinical Asthma Score
CHWRS	Children's Hospital of Wisconsin Respiratory score
CI	Confidence interval
ED	Emergency department
ESBA	Escala de Severidad de la Bronquiolitis Aguda
GRSS	Global Respiratory Severity Score
HDU	High dependency unit
ICU	Intensive care unit
IQR	interquartile range
LIBSS	Liverpool Infant Bronchiolitis Severity Score
LOS	length of stay
mBSS	modified Bronchiolitis severity score
mRDAI	modified Respiratory Distress Assessment Instrument
mFreire	modified Freire model
mRIS	modified Respiratory Index Score
mTal	Modified Tal score
mWCAS	modified Wood's clinical asthma score
NA	not applicable
n.d.	No date
NIV	non-invasive ventilation
NPV	negative predictive value
OR	odds ratio
PCR	Polymerase chain reaction
PICU	Paediatric intensive care unit
PPV	positive predictive value
POBAST	Prediction model Risk Of Bias Assessment Tool
RDAI	Respiratory Distress Assessment Instrument
RESCEU	REspiratory Syncytial virus Consortium in EUrope
ROB	risk of bias
RR	relative risk
RSV	respiratory syncytial virus
RT-PCR	reverse transcription-polymerase chain reaction
SD	standard deviation
SE	standard error
TRIPOD	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
VUS	volume under the surface

WBSS	Wang bronchiolitis severity score
WDF	Wood- Downes-Ferrés score

## PART I

# Validity of clinical severity scores for respiratory syncytial virus: a systematic review

## 1. Abstract

### *Background*

Respiratory syncytial virus (RSV) is a widespread respiratory pathogen, and RSV-related acute lower respiratory tract infections are the most common cause of respiratory hospitalisation in children under five. Over the last two decades, a number of severity scores have been proposed to quantify disease severity for RSV in infants yet there remains no overall consensus on the most clinically useful score.

### *Methods*

We conducted a systematic review of English-language publications in peer-reviewed journals published since 2000 assessing the validity of severity scores for children ( $\leq 24$  months) with RSV and/or bronchiolitis and identified the most promising scores. For included articles, (i) validity data were extracted, (ii) quality of reporting assessed using the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis checklist, and (iii) quality assessed using the Prediction model study Risk Of Bias Assessment Tool. To guide the assessment of the validity data, standardised cut-offs were employed, and an explicit definition of what we required to determine a score was sufficiently validated.

### *Results*

Initial searches identified 7,391 results, of which 1,378 were excluded as duplicates. After title and abstract screening, 5,954 references were excluded. Following full-text screening & snowballing 26 articles, including 24 scores, were included. The most frequently assessed score was the modified Tal score; none of the scores were found to be sufficiently validated according to our definition. The best validated score was the BROSJOD score, and a number of other promising scores were identified. The reporting and/or design of all the included studies was poor.

### *Conclusions*

No scores were found to be sufficiently validated. Further work is warranted to validate the existing scores, ideally in much larger datasets collected from low-income countries.

*Keywords:* RSV, severity score, systematic review, validity

## 2. Introduction

Respiratory syncytial virus (RSV) is a common respiratory infection; it is estimated that by the age of two years most children will have experienced at least one RSV infection [1]. While the vast majority of RSV infections in infants are self-limiting and non-serious, presenting only with generic symptoms of a mild upper respiratory tract infection (e.g. cough, runny nose), a fraction of infants, will develop an acute lower respiratory tract infection, most commonly presenting as bronchiolitis or less commonly as pneumonia. The latest global burden of disease estimates suggests that in 2019, there were 33.0 million cases of RSV-related acute lower respiratory tract infections in children younger than 5, which resulted in 3.6 million hospital admissions, and 101,400 RSV-attributable overall deaths [2]. As such, RSV-related acute lower respiratory tract infections are the most common cause of respiratory hospitalisations in children aged below 5 years. Notably the vast majority of RSV-related acute lower respiratory tract infections occur in low-income countries.

Over the last two decades, a number of different scoring systems have been proposed to quantify disease severity of RSV in children to aide in clinical decision-making and serve as outcome measure/clinical endpoint for clinical trials of vaccines and therapeutics. There are many ways to assess the usefulness of these scores; this primarily consists of assessing their validity (face, discriminative, construct, criterion), reliability, responsiveness and utility [3-4].

A major review of severity scores, published more than a decade ago but still oft-cited, found all of the paediatric dyspnoea scores to be insufficiently evaluated across all domains [3]. The literature base was re-examined in a systematic review & meta-analysis published in 2017, a review published in 2018 and most recently in a rapid review published in 2020 specifically looking to identify scores for resource-limited settings [5-7]. All of these similarly found the severity scores to have been insufficiently validated.

This lack of a validated severity score is significantly impacting on clinical trials; a 2015 meeting of key academic, commercial & regulatory stakeholders in RSV vaccine development identified the lack of “clinically meaningful and reproducible indicators” as the biggest challenge to RSV vaccine development [8]. The lack of consensus was similarly expressed in a recent review of RSV vaccines [9].

Given that it has been almost three years since the last review was conducted, we sought to re-examine the literature base to identify and report on efforts to validate clinical severity scores for use in children ( $\leq 24$  months) with RSV and/or bronchiolitis and synthesise the data to report on the criterion-concurrent and construct validity of the identified severity scores, as well as the included parameters of these scores. Based on this, we identified the most promising scores.

### 3. Methods

Three online medical literature databases, MEDLINE, Embase and Global Health, were searched using the Ovid platform in June 2022 for English-language publications published in peer-reviewed journals since 2000 on the validity of severity scores for children with RSV or bronchiolitis. The search strategies for each database can be found in *ANNEX I*; they were adapted from a recent systematic review on biomarkers for disease severity in RSV [10].

A severity score was defined as a tool used to quantify disease severity over the course of the illness; as such single-purpose models, such as models designed to only predict hospital admission, were excluded.

Covidence was used to identify and automatically exclude duplicates [11]. After removing duplicates, two reviewers (EP & ZS) independently screened the titles and abstracts of the articles for relevance using pre-defined inclusion/exclusion criteria (see *Table 1*). The inclusion/exclusion criteria were similarly adapted from the aforementioned biomarkers review [10].

**Table 1: Inclusion & Exclusion criteria**

(ICU – intensive care unit; PICU – paediatric intensive care unit; RSV – respiratory syncytial virus)

Inclusion Criteria	Exclusion Criteria
Published in a peer-reviewed journal	Not published in a peer-reviewed journal
Published since 2000	Published prior to 2000
Published in the English language	Published in any language other than English
Human RSV and/or bronchiolitis studies.	Studies in animal models or cell lines, and studies of children without a RSV or bronchiolitis diagnosis
Relation explored between clinical measures and severity of RSV infection, and including a defined clinical severity score.	Studies focused on treatment, diagnostics, or epidemiology of RSV infection.
Children ( $\leq 24$ months) with RSV and/or bronchiolitis	Studies in those $>24$ months old with RSV and/or bronchiolitis.
At least 50 children with RSV and/or bronchiolitis	Less than 50 children with RSV and/or bronchiolitis
Evaluation & validation studies, case-control, cohort, placebo arm of randomised control trial and cross-sectional trials.	Literature and systematic reviews, case reports, case series, letters and protocols
Full-text available	No full-text (e.g. conference poster)
Studies in the community and primary, secondary or tertiary care.	Studies only in the ICU/PICU
	Severity scores only using treatment (e.g. ventilated) and/or hospitalisation length

For the remaining included papers, their full-text was acquired, and subsequently screened independently by two reviewers (EP & ZS) for relevance. Any uncertainty about articles at this stage was resolved through consultation with a senior researcher (HN). The reference lists of papers identified for inclusion, as well as 3 previous reviews, were examined to identify additional relevant references (i.e. snowballing) [3,6,7].

Data from the included studies were extracted independently by two reviewers (EP & ZS) into a standardised spreadsheet; this recorded the first author surname, year of publication, study design(s), whether the paper described a development or validation study, country/countries from where the data was collected, sample size(s), age range(s) (converted to months), clinical setting(s), RSV positivity rate, test used to detect RSV, severity score(s) used and the key relevant validation data reported. The World Bank's income level classification scheme was used to categorise the economies of the countries [12]. Data were simultaneously separately collected on the parameters included in each score. Additionally, score names were standardised.

Given the widely observed poor quality of publications reporting prediction models, as well as specifically for severity scores for RSV, we employed the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis checklist (TRIPOD), a 23-item checklist to quantify the quality of reporting [5, 13-15]. The related Prediction model Risk Of Bias Assessment Tool (PROBAST) was also employed to assess the risk of bias of included studies [13, 16]. For the included studies, the TRIPOD and PROBAST checklists were both assessed independently by two reviewers (EP & ZS); any uncertainty was resolved through consultation with a senior researcher (HN).

Given the heterogeneous nature of the included studies and the small amount of data on each severity score, only a narrative synthesis was made and meta-analysis not conducted. The review was registered with PROSPERO (CRD42022343781).

#### *Quality assessment of validity of identified scores*

Using the data extracted from the included studies, we assessed each of the identified scores for their face, construct (discriminative & convergent) and criterion-concurrent validity. We found, similarly to the 2014 review, a wide range of different uses of these terms and so have explicitly specified how we categorised and assessed the validity data (see *Table 2*) [3].

**Table 2: Validity types**

(LOS – length of stay; PICU – paediatric intensive care unit)

<b>Validity type</b>	<b>How assessed?</b>
Face	Subjective assessment of score components.
Criterion-concurrent	No accepted gold standard. Correlation with oxygen saturation, respiratory rate or expert classification.
<i>Construct</i>	
Convergent validity	Correlation with LOS, treatment usage (e.g. days ventilated), or other severity score.
Discriminative	Discriminative ability with hospital admission, PICU admission, treatment, or prolonged LOS.

To guide our assessment, the same cut-offs as proposed by Hakizimana et al in their rapid review were used [7]. For area under the receiver operating characteristic curve (AUROC), a score of <0.5 was classified as poor, 0.5-0.7 low, 0.7-0.9 moderate & >0.9 high, and for Spearman's correlation coefficient we took 0–0.19 as very weak, 0.2–0.39 weak, 0.4-0.59 moderate, 0.6–0.79 strong, and 0.8–1 as a very strong correlation. They didn't specify cut-offs for Pearson's correlation coefficient; we used <0.1 as negligible, 0.1-0.4 weak, 0.4-0.7 moderate, 0.7-0.9 as strong & >0.9 as very strong. For other measures we made a subjective assessment informed by the above cut-offs. We considered a p-value ≤0.01 as constituting statistical significance.

We considered a score to be sufficiently validated if at least two external validation studies with a low risk of bias rating (as assessed by PROBAST) had assessed the criterion-concurrent, convergent and/or discriminative validity for at least two separate outcomes each, and that performed at least moderately for each outcome. To identify promising scores (i.e. scores that are currently insufficiently validated), we made a subjective assessment based on the scores that were deemed that most likely could be sufficiently validated.

## 4. Results

### *Descriptive statistics*

Initial searches produced 7,391 results (see *Figure 1*) of which 59 articles were identified for full-text screening after title and abstract screening. Two additional relevant articles were identified through snowballing. As such, overall 26 articles were included, comprising 24 unique scores (see *Table 3*) [17-42]. The vast majority of the included studies used a prospective design (n=23), most commonly a cohort study (n=19), and the remaining 3 studies used either a purely retrospective design (n=2) or combination of retrospective and prospective design (n=1).

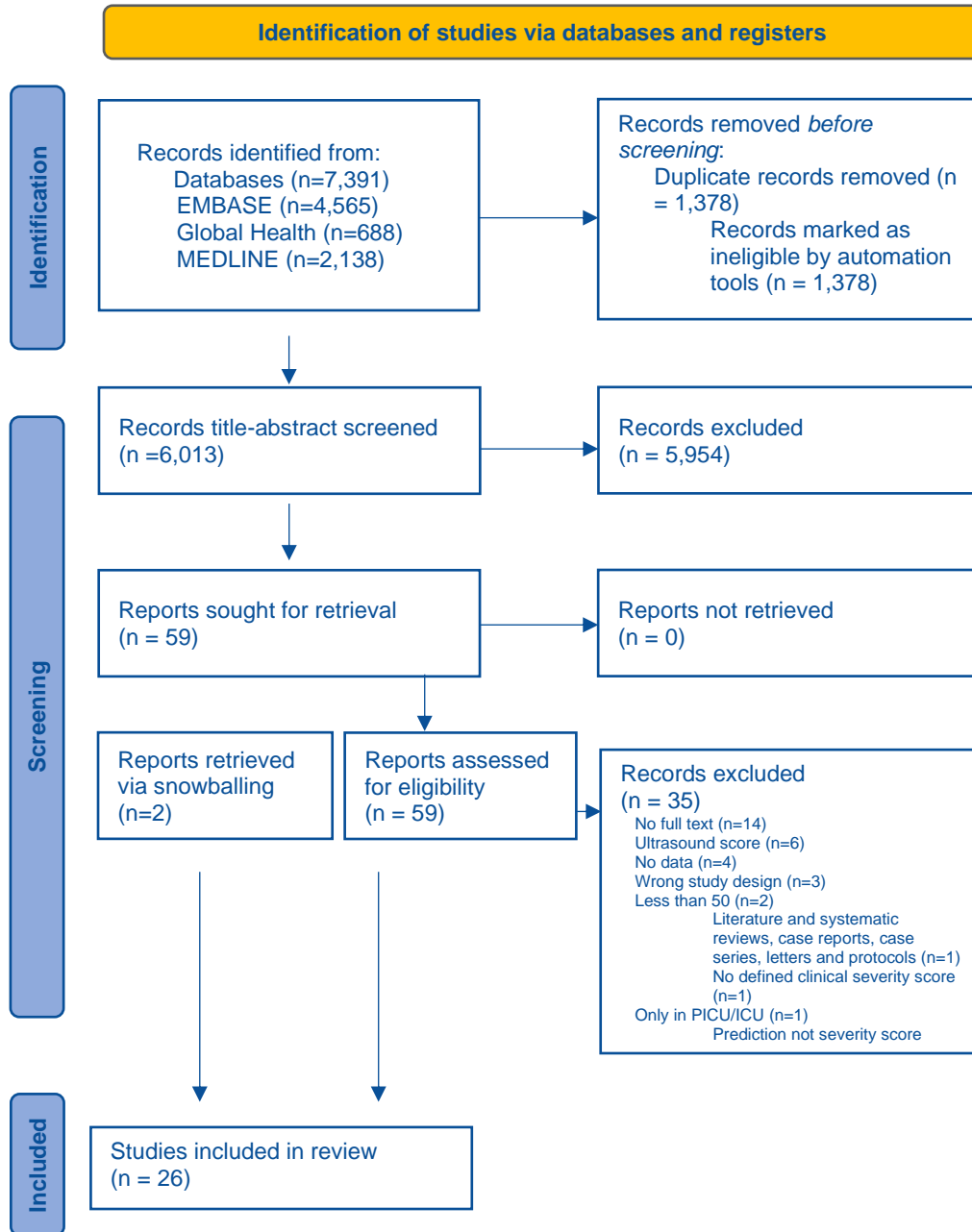
Four of the studies developed a new score, of which one included external validation in the same publication; the remaining 22 studies only validated existing scores. Seven of the studies were multi-centre studies. Twenty-two of the studies used data collected in secondary care, including two studies which also made use of data from the community; the remaining four studies used data collected in tertiary care.

The most frequently used score was the modified Tal (mTal) score which was used in four studies. Three studies used the Wang bronchiolitis severity score (WBSS); the Bronchiolitis Score of Sant Joan de Déu (BROSJOD), Escala de Severidad de la Bronquiolitis Aguda (ESBA), Global Respiratory Severity Score (GRSS), modified Respiratory Index Score (mRIS), and Wood-Downes-Ferrés score (WDF) were each used in two studies, and the remaining 16 scores were only evaluated once. Although Raita et al. [35] claimed to use the Freire model, they excluded one of the parameters included in the original Freire model, so we considered it as a separate score and referred to it as the modified Freire model (mFreire).

Most commonly discriminative validity was assessed (n=21). Fourteen studies assessed convergent validity and 3 criterion-concurrent validity.

Seven papers used data from Spain, five from the United States, four from Israel, two each from Australia, France and Turkey and one each from Canada, Colombia, Egypt, India, Ireland, Japan, New Zealand, Portugal, Singapore, and the United Kingdom. The vast majority of the included data were from high-income countries; only three studies used data from upper middle-income countries (Turkey [n=2] & Colombia), and three from a lower middle-income country (Colombia, Egypt, India). No included papers used data from any low-income country.

**Figure 1: Preferred Reporting Items for Systematic reviews and Meta-Analyses Diagram.**



**Table 3:** Included Studies. (Asterisks indicates article identified through snowballing.)

(AUROC- Area under the receiver-operator characteristic curve; BROSJOD - Bronchiolitis Score of Sant Joan de Déu; BSS-Bronchiolitis severity score; CAS – Clinical Asthma Score; CHWRS – Children’s Hospital of Wisconsin Respiratory score; CI - confidence interval; ED -Emergency department; ESBA – Escala de Severidad de la Bronquiolitis Aguda; GRSS – Global Respiratory Severity Score; IQR - interquartile range; LOS – length of stay; mBSS-modified Bronchiolitis severity score; mRDAI – modified Respiratory Distress Assessment Instrument; mFreire - modified Freire model; mRIS - modified Respiratory Index Score; mTal – Modified Tal score; mWCAS – modified Wood’s clinical asthma score; NA – not applicable; NIV – non-invasive ventilation; NPV – negative predictive value; OR - odds ratio; PCR-Polymerase chain reaction; PICU – Paediatric intensive care unit; PPV – positive predictive value; RDAI - Respiratory Distress Assessment Instrument; RR - relative risk; RSV – respiratory syncytial virus; RT-PCR -Reverse transcription polymerase chain reaction; SD - standard deviation; SE - standard error; TRIPOD - Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis checklist; VUS - volume under the surface; WBSS - Wang bronchiolitis severity score; WDF – Wood-Downes-Ferrés score)

First author surname, year	Study design(s)	Development and/or external validation	Country/countries data is collected from	Score(s)	Sample size	Age range(s) (months)	Clinical setting(s)	RSV positivity rate (%)	RSV Test	TRIPOD score (%)	Key Findings (95% CI)
Amat, 2014 [17]	Prospective case-control study	External validation	France	Wainwright severity score	222	<12	Secondary care	76%	PCR on nasopharyngeal aspirate	12/27 (44%)	<p><u>Convergent validity</u> On univariate analysis, initial severity score correlated with hospital LOS (p=0.02).</p> <p><u>Discriminative validity</u> Infants admitted had a higher severity score (p&lt;0.0001). Infants admitted to the PICU had a higher Wainwright severity score (p&lt;0.008) than those admitted to hospital but not to the PICU.</p>

											On univariate analysis severity score on admission was correlated with hospitalisation (OR 5.3 [2.7-10.5]; p=0.001), LOS (p=0.02) & need for intensive care (p<0.001).
Anil, 2017 [18]	Prospective case-control study	External validation	Turkey	WBSS	232	1.5-24	Secondary care	Not reported	NA	14/27 (52%)	<p><u>Convergent validity</u> Pearson correlation coefficient between WBSS and LOS in hospitalised patients was 0.462 (&lt;0.001).</p> <p><u>Discriminative validity</u> Significant differences between the control group, and those with a mild, moderate or severe WBSS and their mean pulse rate, respiratory rate, and oxygen saturation (&lt;0.001). Significant difference in pH and pCO<sub>2</sub> between those with a severe WBSS compared to the control, and mild &amp; moderate bronchiolitis severity. Hospitalised patients had significantly higher WBSS than those discharged (&lt;0.001) with an OR of 2.524(1.826-3.487; p&lt;0.001) on logistic regression.</p>
Balaguer, 2017 [19]	Prospective observational cohort study	External validation	Spain	BROSJOD score; Wood Downes score	112	<24	Secondary care	74.50%	Not reported	15/27 (56%)	<p><u>Criterion-concurrent validity</u> Kappa index between BROSJOD score &amp; expert opinion scale was 0.84 (0.78-0.90) at admission, 0.80 (0.68-0.92) at 24 hours, and 0.84 (0.70-0.97) at 48 hours.</p> <p><u>Convergent validity</u> Kendall rank correlation coefficient between BROSJOD &amp; Wood-Downes score was 0.66 (0.56-0.75; p &lt;0.01) at</p>

											<p>admission, 0.62 (0.53-0.70; p&lt;0.01) at 24 hours, and 0.63 (0.53-0.70; p&lt;0.01) at 48 hours.</p> <p>Significant correlation between BROSJOD score and PICU LOS (p&lt;0.001) as well as hospital LOS (p&lt;0.001).</p> <p><u>Discriminative validity</u> VUS between BROSJOD score, and expert opinion was 0.80 (0.70-0.90) at admission, 0.92 (0.85-0.99) at 24 hours, and 0.93 (0.87-0.99) at 48 hours.</p> <p>Significant association between BROSJOD score level and need for invasive mechanical ventilation. No significant correlation between BROSJOD score &amp; need for NIV (p=0.16).</p>
Bueno-Campaña, 2019 [20]	Multicentre (n=3) prospective cohort study	External validation	Spain	WDF	145	≤6	Secondary care	73.8	Antigen and/or PCR test on unstate d sample.	12/27 (44%)	<p><u>Discriminative validity</u> WDF (≥6 points) had an estimated RR 2.9 (CI 95% 1.9-4.2) for respiratory support as an outcome.</p> <p>WDF (&gt;6 points) had a predictive capacity for respiratory support with a sensitivity of 56.1% (43.3-68.2), specificity of 85.9% (76.9-91.7), PPV of 72.7 (58.2-83.7) and NPV of 74.5 (65.0-82.1).</p>
Caserta, 2017 [21]	Prospective multicent	Development	United Stat	GRSS	139	<10	Community; Seconda	100	RT-PCR on nasal	17/34 (50%)	<p><u>Convergent validity</u> Pearson correlation coefficient between GRSS and LOS for hospitalised infants</p>

	re (n=5) cohort study		es				ry care		swab		<p>was 0.586 (p&lt;0.0001).</p> <p><u>Discriminative validity</u> AUROC of 0.961 at predicting admission. AUROC was 0.962 and 0.959 for predicting admission in those ≤3 &amp; &gt;3 months, respectively.</p> <p>Mean GRSS for patients admitted to PICU was 8.5 (± 0.34 SE) compared to 5.7 (± 0.19 SE; p&lt;0.0001) for those hospitalised but not admitted to PICU.</p>
Chong, 2017 [22]	Prospective observational cohort study	Development	Singapore	mRIS	18 18	<24	Tertiary care	26.84% (n=184) for hospitalised patients (n=719)	Unstated test on nasopharyngeal aspirate	17/26 (65%)	<p><u>Discriminative validity</u> C-statistics were 0.68 for warranted admission, 0.77 for non-invasive respiratory support, 0.68 for intravenous hydration, and 0.68 for LOS.</p> <p>AUROC for composite outcome of respiratory support, intravenous hydration, and hospital stay of 2 or more days was 0.68 (0.65-0.71) with a sensitivity of 70.0%, specificity of 60.0%, PPV of 44.4%, and NPV of 81.2%.</p> <p>After adjusting for age and day of illness, the OR of a mRIS score &gt;4 in predicting for respiratory support was 4.64 (1.78-12.99; p&lt;0.001), intravenous hydration 3.56 (2.33-5.51; p&lt;0.001), LOS ≥2 days or more 3.32 (2.65-4.17; p&lt;0.001), and overall warranted admission 3.28 (2.62-4.12; p&lt;0.001).</p>

											Statistically significant adjusted OR in infants presenting with their first wheeze with a mRIS score >4 and, the need for respiratory support 3.20 (1.09-9.39), intravenous hydration, 3.40 (1.98-5.85), significant hospital stay 3.40 (2.51-4.61), and overall warranted admission 3.28 (2.62-4.11).
Chong, 2018 [23]	Prospective cross-sectional study.	External validation	Singapore	mRIS	85	<24	Tertiary care	Not reported	NA	10/26 (38%)	<p><u>Discriminative validity</u> Median (IQR) mRIS score of hospitalized and discharged patients was 3 (3-4) and 3 (2-4), respectively (p=0.2509).</p> <p>AUROC for hospitalisation was 0.5709.</p>
Destino, 2012 [24]	Prospective cohort study	External validation	United States	CHWRS; RDAI	195	<12	Secondary care	Not reported	NA	17/27 (63%)	<p><u>Convergent validity</u> Spearman's rank correlation on admission, and LOS for CHWRS and RDAI was 0.05 (p=0.61), and 0.04 (p=0.71), respectively.</p> <p><u>Discriminative validity</u> AUROC for RDAI was 0.51 at predicting admission. AUROC for CHWRS was 0.68 at predicting admission.</p> <p>CHWRS has a sensitivity of 65% and specificity of 65% (using cut-off of &gt;7.5 derived from the dataset) for predicting admission.</p>
Duarte-Dorado, 2013 [25]	Prospective cohort study	External validation	Colombia	mWCAS; Tal	54	<24	Secondary care	Not reported	NA	19/26 (73%)	<p><u>Convergent validity</u> Spearman correlation coefficient between mWCAS and Tal severity scores at admission was 0.761 (p&lt;0.001)</p>

											<p>and 0.809 (<math>p &lt; 0.001</math>) for the first and second-rater, respectively, and 0.712 (<math>p &lt; 0.001</math>) at discharge.</p> <p><u>Discriminative validity</u> Median mWCAS at admission to the PICU was 4.5 (3.6-5.2) and 4.7 (3.6-5.1; <math>p &lt; 0.001</math>), compared to 2.5 (1.5-2.5) and 2.5 (2.0-2.5; <math>p &lt; 0.001</math>) for those admitted but not taken to the PICU for the first and second-rater, respectively.</p> <p>Median mWCAS in patients at admission was 2.5 (1.9-2.5) compared to 1.0 (0.5-1.6; <math>p &lt; 0.001</math>) at discharge.</p>
El Basha, 2019 [26]	Prospective cohort	External validation	Egypt	mTal	153	<12	Tertiary care	50.98%	RT-PCR of oropharyngeal and nasopharyngeal swabs	11/26 (42%)	<p><u>Convergent validity</u> Spearman's correlation coefficient between mTal and duration of oxygen therapy was 0.653 (<math>p &lt; 0.001</math>) and 0.721 (<math>p &lt; 0.001</math>) in preterm and term infants, respectively.</p> <p>Spearman's correlation coefficient between mTal and LOS was 0.644 (<math>p &lt; 0.001</math>) and 0.803 (<math>p &lt; 0.001</math>) in preterm and term infants, respectively.</p> <p>Spearman's correlation coefficient between mTal and duration of PICU admission was 0.304 (<math>p = 0.220</math>) and 0.310 (<math>p = 0.550</math>) in preterm and term infants, respectively.</p>
Freire, 2018*	Multinational	D	Australia,	Freire	2722	<12	Secondary care	Not reported	NA	20/28 (71%)	<p><u>Discriminative validity</u> AUROC for the risk score was 0.847</p>

[27]	(n=8) multicentre (n=38) retrospective cohort chart review study.		Canada, Ireland, New Zealand, Spain, United Kingdom, United States, and Portugal.					d			(0.817-0.868) for predicting escalated care. Bootstrapping validation gave an average AUROC of 0.842 (range 0.803–0.882) and a mean overoptimism value of 0%.
Gal, 2015 [28]	Prospective cohort study	External validation	Israel	mRDAI	60	<18	Secondary care	Not reported	NA	12/27 (44%)	<p><u>Criterion-concurrent validity</u> Positive linear correlation between PtcCO<sub>2</sub> and the disease severity clinical score at study entry (p&lt;0.001).</p> <p>Disease severity clinical score remained independently associated with PtcCO<sub>2</sub> (p&lt;0.05) after controlling for PvCO<sub>2</sub> and weight.</p>
Golan-Tripto, 2018 [29]	Prospective cohort study	External validation	Israel	mTal	50	<12	Secondary care	81% (Only 48 tested)	PCR on nasal wash	19/27 (70%)	<p><u>Convergent validity</u> Spearman's correlation coefficient between first mTal score and length of oxygen support was 0.62 (p&lt;0.001), 0.50 (p=0.001), 0.48 (p=0.006) and 0.37</p>



											<p>paediatric residents, respectively.</p> <p>RR for need of oxygen support was 1.33 (1.12-1.57), 1.26 (1.1-1.46), 1.26 (1.06-1.5), and 1.21 (0.93-1.58) for paediatric pulmonologists, paediatricians, senior paediatric residents and junior paediatric residents, respectively.</p>
Jacob, 2016 [30]	Prospective single-blind cohort study	External validation	Israel	WBSS	114	<24	Secondary care	52	Not reported	11/26 (42%)	<p><u>Convergent validity</u> WBSS on arrival to the ED was the strongest predictor of hospital LOS (p=0.011).</p> <p><u>Discriminative validity</u> WBSS on arrival to ED predicted nasogastric tube feeding with an OR of 1.95 (1.12-3.39; p=0.017).</p> <p>WBSS on arrival to ED predicted desaturation days during hospitalization with an OR of 1.641 (0.977-2.75; p=0.061).</p>
Kubota, 2021 [31]	Retrospective cohort study	External validation	Japan	GRSS; WBSS	250	<10	Secondary care	100%	Rapid antigen test	16/27 (59%)	<p><u>Discriminative validity</u> AUROC of the GRSS and the WBSS were 0.875 and 0.821, respectively for predicting the need for respiratory support.</p> <p>Median GRSS of infants requiring respiratory support was 4.86 (IQR 3.67-6.10) compared to 2.79 (1.91-3.83; p &lt;0.001) for those not requiring respiratory support.</p>

											<p>Median WBSS of infants requiring respiratory support was 8 (IQR 7-9) compared to 6 (5-7; <math>p &lt; 0.001</math>) for those not requiring respiratory support.</p> <p>GRSS had a sensitivity of 69% (50-83%) and specificity of 90% (86-93%) for predicting the need for respiratory support using cut-offs derived from this dataset (4.52).</p> <p>WBSS had a sensitivity of 92% (76-98%) and specificity of 60% (53-66%) for predicting the need for respiratory support using cut-offs derived from this dataset (7).</p>
Marguet, 2009 [32]	Multicentre (n=4) prospective cohort study	External validation	France	CAS	209	<12	Secondary care	64.1% (including 14.3% of infants who had a dual rhinovirus & RSV infection)	Direct immunofluorescence assay from nasopharyngeal aspirate.  PCR test on nasopharyngeal aspirate	9/26 (35%)	<p><b>Convergent validity</b> Spearman's correlation coefficient between CAS and, duration of oxygen supplementation and LOS was 0.482 (<math>p &lt; 0.0001</math>) &amp; 0.393 (<math>p &lt; 0.0001</math>), respectively.</p>

								e.			
McCallum, 2013 [33]	Prospective cohort study	External validation	Australia	mTal	115	<24	Secondary care	Not reported	NA	17/28 (61%)	<p><u>Discriminative validity</u> For infants not on oxygen at enrolment (n=57), AUROC was 0.68 (0.13-1.0) and 0.75 (0.34-1.0) at predicting oxygen need at 12 and 24 hours, respectively.</p> <p>For infants on oxygen at enrolment (n=58), the AUROC was 0.60 (0.46-0.75) at predicting oxygen need at 12 or 24 hours.</p>
Özkaya, 2020 [34]	Prospective cross-sectional cohort	External validation	Turkey	mBSS	76	<24	Secondary care	Not reported	NA	15/27 (56%)	<p><u>Discriminative validity</u> Median (IQR) mBSS score was 4(3-6) and 2 (1-3) in the admitted and non-admitted group, respectively (p&lt;0.001).</p> <p>Mean mBSS score was 4.8 (±2.4) and 2.4 (±1.4) in the admitted and non-admitted group, respectively (p&lt;0.001).</p> <p>AUROC for predicting admission was 0.814 (0.71-0.91; p&lt;0.001). With a cut-off of 4, sensitivity and specificity for predicting admission was 73.81% (58-86.1) and 82.35%, respectively.</p> <p>OR of mBSS at predicting admission was 2.07 [1.43-3.0; p&lt;0.001].</p>
Raita, 2020 [35]	Multicentre (n=17) prospective cohort study	Both	United States	4 machine learning-based	1016	<12	Secondary care	80.80%	PCR on unreported sample	16/30 (53%)	<p><u>Discriminative validity</u> mFreire (the reference model) had an AUROC 0.62 (95% CI 0.53-0.70). It had a sensitivity of 0.62 (95% CI 0.49-0.75) and a specificity of 0.57 (95% CI 0.54-0.60) in predicting positive pressure</p>

				models (Lasso regression, elastic net regression, random forest, and gradient boosted decision tree); mFreire						<p>ventilation outcome. For predicting intensive treatment, mFreire had an AUC of 0.62 (95% CI 0.57-0.67), a sensitivity of 0.58 (95% CI 0.49-0.75), and specificity 0.58 (95% CI 0.50-0.66).</p> <p>The Logistic regression with Lasso regularisation had an AUC of 0.88 (0.84-0.93), sensitivity of 0.84 (0.73-0.93) and a specificity of 0.79 (0.77-0.82) in predicting positive pressure ventilation outcome. For predicting intensive treatment, the model had an AUC of 0.79 (0.76-0.83), a sensitivity of 0.75 (0.69-0.82), and specificity 0.70 (0.66-0.73).</p> <p>The Logistic regression with Elastic net regularisation had an AUC of 0.89 (0.85-0.92), sensitivity of 0.89 (0.80-0.96) and a specificity of 0.73 (0.70-0.75) in predicting positive pressure ventilation outcome. For predicting intensive treatment, the model had an AUC of 0.80 (0.76-0.83), a sensitivity of 0.72 (0.64-0.79), and specificity 0.74 (0.71-0.77).</p> <p>The Random forest model had an AUC of 0.89 (0.85-0.92), sensitivity of 0.85 (0.75-0.95) and a specificity of 0.74 (0.71-0.76) in predicting positive pressure ventilation outcome. For predicting intensive treatment, the model had an AUC of 0.79 (0.75-0.84), a sensitivity of 0.70 (0.63-0.77), and</p>
--	--	--	--	---	--	--	--	--	--	--

											specificity 0.78 (0.76-0.81).  The Gradient boosted decision tree model had an 0.88 (0.84-0.93), sensitivity of 0.89 (0.80-0.96) and a specificity of 0.77 (0.75-0.80) in predicting positive pressure ventilation outcome. For predicting intensive treatment, the model had an AUC of 0.79 (0.75-0.84), a sensitivity of 0.74 (0.67-0.80), and specificity 0.74 (0.71-0.77).
Ramos-Fernández, 2018 [36]	Prospective cohort study.	External validation	Spain	ESBA	190	<12	Secondary care	77.80%	Unstated test on nasopharyngeal aspirate	16/27 (59%)	<p><u>Discriminative validity</u> Mean (SD) initial ESBA for those admitted to the PICU was 6.41(±2,97) compared to 5.25 (±2.48) in those that weren't admitted to the PICU; insignificant.</p> <p>Mean (SD) highest ESBA for those admitted to the PICU was 10.55 (±1.12) compared to 6.35 (±2.3) in those that weren't admitted to the PICU (p&lt;0.001).</p> <p>AUROC for initial ESBA score was 0.61 (0.41-0.80; p=0.24) and for highest ESBA score was 0.94 (0.90-0.98; p&lt;0.001) for predicting admission to the PICU.</p> <p>Sensitivity and specificity of the ESBA score was 81.8% (52.3-94.9) and 91.1% (86.0-94.4) using the optimal cut-off point.</p>
Ricart,	Prospective	External	Spain	BROSJ	41	<12	Secondary	43.1	PCR on	12/27	<u>Discriminative validity</u>

2013 [37]	ve cohort study	rnal valid ation	n	OD score	0		ry care		nasoph arynge al aspirat e	(44%)	<p>Mean LOS stay in severe group (BROSJOD <math>\geq</math> 11) was 11.2 days (SD 10.4) compared to 4.3 days (SD 3.1; <math>p &lt; 0.005</math>) in the non-severe group.</p> <p>Mean days of oxygen therapy in severe group was 9.0 days (SD 8.2) compared to 2.6 days (SD 2.5; <math>p &lt; 0.005</math>) in the non-severe group.</p> <p>Maximum mean FIO<sub>2</sub> required in severe group was 43% (SD 18.4) compared to 26% (SD 7.1; <math>p &lt; 0.005</math>) in the non-severe group.</p> <p>Mean days with nasogastric feeding tube in severe group was 6.7 (range 0-58) compared to 0.7 (range 0-17; <math>p &lt; 0.005</math>) in the non-severe group.</p> <p>The percentage of the severe group admitted to the PICU was 50% compared to 3.4% (<math>p &lt; 0.005</math>) of the non-severe group.</p> <p>The percentage of the severe group who required non-invasive ventilation was 43.9% compared to 2.5% (<math>p &lt; 0.005</math>) of the non-severe group.</p> <p>The percentage of the severe group who required mechanical ventilation was 32.9% compared to 0.6 % (<math>p &lt; 0.005</math>) of the non-severe group.</p>
--------------	--------------------	------------------------	---	-------------	---	--	---------	--	--	-------	---

Rivas-Jueas, 2018 [38]	Multicentre (n=5) prospective cohort study.	External validation	Spain	ESBA; WDF	201	≤12	Secondary care	Not reported	NA	16/27 (59%)	<p><u>Convergent validity</u> Weighted kappa coefficient between ESBA &amp; WDF was -0.17 based on each scores cut-off into mild, moderate &amp; severe.</p> <p><u>Discriminative validity</u> Mean WDF score in severe group (based on presence of respiratory acidosis, admitted to ICU or, mechanical and non-invasive ventilation administered) was 7.31 (SD 1.88) compared to 5.32 (SD 1.6; p&lt;0.01) in the non-severe group.</p> <p>Mean ESBA score in severe group was 7.31 (SD 1.88) compared to 4.91 (SD 2.11; p&lt;0.01) in the non-severe group.</p> <p>WDF had a sensitivity of 46.2% (23.2-70.9) and specificity of 91.5% (86.6-94.7) at differentiating between the severe and non-severe group using each scores severe cut-off. ESBA had a sensitivity of 3.6% (0.4-26.8) and specificity of 98.1% (95.1-99.3).</p> <p>AUROC of WDF was 0.79 (0.73-0.85) at differentiating between the severe and non-severe group. AUROC of ESBA was 0.82 (0.75-0.87).</p> <p>With proposed new cut-offs derived from the same sample: WDF (&gt;5) sensitivity of 92.3% (66.7-</p>
------------------------	---	---------------------	-------	-----------	-----	-----	----------------	--------------	----	-------------	--

											98.6) and specificity of 54.8% (47.6-61.7) at differentiating between the severe and non-severe group. ESBA (>6) sensitivity of 84.6% (57.8-95.7) and specificity of 78.7% (72.3-84.0).
Rodriguez-Gonzalez, 2022 [39]	Prospective cohort study.	External validation	Spain	BROSJOD score	112	<12	Tertiary care	77	Not reported	17/27 (63%)	<p><u>Convergent validity</u> Spearman's correlation coefficient between the BROSJOD score and hospitalisation LOS, PICU LOS and duration of respiratory support was 0.55 (p&lt;0.001), -0.08 (p&gt;0.05) and 0.47 (p&lt;0.001), respectively.</p> <p><u>Discriminative validity</u> Patients requiring respiratory support presented with higher BROSJOD scores (p&lt;0.001).</p> <p>OR of a BROSJOD score &gt;10 for predicting the need for respiratory support was 28 (12-125).</p> <p>AUROC, sensitivity and specificity of a BROSJOD &gt;10 for predicting the need for respiratory support was 0.87 (0.78-0.93), 0.91 and 0.61, respectively.</p> <p>Median (IQR) BROSJOD score for those who received respiratory support (n=36) was 11 (9-13) compared to 6 (4-8) in those did not receive respiratory support (n=76) (p&lt;0.001).</p>

											BROSJOD score did not significantly correlate with PICU stay.
Shete, 2014 [40]	Prospective cross-sectional study.	External validation	India	mTal	142	1-12	Secondary care	Not reported	NA	16/27 (59%)	<p><u>Criterion-concurrent validity</u></p> <p>The mean SpO<sub>2</sub> value for children with clinical scores of 2-5 (n = 32) was 98.2 (SD 1.3); 95.2 (SD 0.9) for those with scores of 6-7 (n = 84), and 92.3 (SD 0.87) for children with scores of 8-10 (n = 26). Bonferroni's multiple comparison showed a statistically significant difference for all pairs (p&lt;0.001).</p> <p>Pearson's correlation coefficient between mTal and SpO<sub>2</sub> was -0.734 (p&lt;0.0001).</p>
Siraj, 2020* [41]	Combined retrospective and prospective, observational mixed methods study	V	United States	BSS	181	1-24 months	Secondary care	Not reported	NA	15/27 (56%)	<p><u>Convergent validity</u></p> <p>Pearson's correlation coefficient between BSS and LOS, weight-adjusted high-flow nasal canula flow rate and duration of high-flow nasal canula therapy was -0.04, 0.09 and 0.04, respectively.</p> <p><u>Discriminative validity</u></p> <p>Median BSS value before admission was 6 for children with late rescue and 4 for those without (p = 0.09). AUROC of 0.61 (0.48–0.75) for discriminating by late rescues. At the ideal threshold (≥6), sensitivity was 56% and specificity was 69%. Unadjusted logistic regression for BSS gave an OR of 1.18 (0.93–1.5; p=0.17).</p>
Somech, 2006 [42]	Multicentre (n=2) prospective	External valid	Israel	WBSS	195	<12	Community; Secondary	100	Enzyme immun	6/26 (23%)	<p><u>Discriminative validity</u></p> <p>A statistically significant difference was found in the mean (± SD) WBSS</p>

	ve observati onal cohort study	ation					ry care		oassay on unstate d mediu m.		between ambulatory (4.1±1.1), hospitalised (7.9±1.1), and PICU hospitalised (11.2±1.5) infants.
--	--	-------	--	--	--	--	---------	--	---	--	---

### *Severity score components*

For 20 of the scores, we were able to identify the parameters used; however, we were unable to identify all of the parameters used in the four machine learning models proposed by Raita et al. [35] – the authors only mention the 15 most important predictors. There was significant variation in the parameters used by each severity score model. After grouping synonymous terms (e.g. respiratory rate & respiratory frequency), 47 unique parameters were included in the scores (see *Table 4*).

The mean number of parameters in each score was 6 (range 3-10). Most commonly included was respiratory rate (n=13); the next most common parameters included oxygen saturation (n=8), retractions (n=8) and wheezing (n=8). Also commonly included was a parameter including heart rate (n=5), accessory muscle use (n=4) and cyanosis (n=4). The remaining parameters were infrequently used ( $\leq 3$  times).

**Table 4:** Parameters in included scores. Synonymous items grouped. If age/sex used within a parameter (e.g. to specify cut-offs), age/sex is not recorded.

(BROSJOD - Bronchiolitis Score of Sant Joan de Déu; BSS-Bronchiolitis severity score; CAS – Clinical Asthma Score; CHWRS – Children's Hospital of Wisconsin Respiratory score; ESBA – Escala de Severidad de la Bronquiolitis Aguda; GRSS – Global Respiratory Severity Score; mBSS - modified Bronchiolitis Severity Score; mFreire - modified Freire model; mRIS - modified Respiratory Index Score; mTal – Modified Tal score; mWCAS – modified Wood's clinical asthma score; RDAI - Respiratory Distress Assessment Instrument; WBSS - Wang bronchiolitis severity score; WDF – Wood-Downes-Ferrés score)

Parameters	BS S	BR OS JO D	CAS	CHWRS	ESBA	Freire	GRSS	mBSS	mFreire	mRDAI	mRIS	mTal	mWCAS	RDAI	Tal	Wainwright severity score	WBSS	WDF	Wood Downes score
Used by	Siraj [41]	Balaguer [19]; Rodriguez-Gonzalez [39]	Marguet [32]	Destino, [24]	Ramos-Fernández [36]; Rivas-Jueas [38]	Freire [27]	Caserta [21]; Kubota, 2021 [31]	Özkaya [34]	Raita [35]	Gal [28]	Chong [22]; Chong [23]	Duarte-Dorado [25]; El Bash a [26]; Golan - Tripto [29]; McCa llum [33]; Shete [40]	Duarte - Dorado [25]	Destino [24]	Duarte-Dorado [25]	Amat [17]	Anil [18]; Jacob [30]; Kubota [31]; Somach [42]	Bueno-Campaña [20]; Rivas - Jueas [38]	Balaguer [19]
Accessory muscles												Y	Y		Y				Y
Activity & appearance				Y															
Age						Y			Y										
Air entry		Y																Y	
Apnoea						Y			Y										
Breath sounds	Y			Y				Y											
Cerebral function													Y						Y
Chest movement synchronisation			Y																
Chest x-ray/lung sounds				Y															
Cough ability/secretions				Y															
Crackles					Y					Y									

Cyanosis							Y								Y			Y	Y
Dehydration						Y			Y										
Dyspnoea	Y			Y															
Effort				Y															
Expiratory wheezing												Y	Y						Y
Feeding			Y			Y			Y										
General appearance							Y										Y		
Heart rate		Y		Y	Y					Y								Y	
I/E ratio			Y																
Indrawing		Y																	
Inspiration/expiration ratio					Y														
Inspiratory breath sounds												Y							Y
Inspiratory wheezing													Y						
Intercostal retractions			Y											Y					
Lethargy							Y												
Location of wheeze														Y					
Lung hyperinflation chest x-ray			Y																
Mental status								Y			Y								
Nasal flaring			Y																
Nasal flaring and/or grunting						Y													
Oxygen need				Y						Y									
Oxygen saturation		Y				Y	Y	Y	Y			Y	Y			Y			
Poor air movement							Y												
Rales/rhonchi							Y												
Reactivity			Y																
Respiratory rate	Y	Y		Y	Y		Y	Y		Y	Y	Y			Y	Y	Y	Y	Y
Respiratory-effort score																Y			
Retractions	Y			Y		Y	Y		Y		Y						Y	Y	
Signs of respiratory distress (retractions &										Y									

nasal flaring)																			
Subcostal retractions														Y					
Supraclavicular retractions														Y					
Surgical status				Y															
Wheezes/rales		Y																	
Wheezing			Y		Y		Y			Y	Y	Y			Y			Y	Y
Work of breathing								Y											
Xiphoid retractions			Y																

### *Discriminative Validity*

Twenty-one of the studies assessed the discriminative validity of the scores, mostly by assessing their ability to discriminate between those discharged or admitted to the hospital, and between those admitted to the paediatric intensive care unit (PICU) and those not admitted to the PICU but hospitalised. The WBSS was assessed in four papers, the BROSJOD score in three papers, and the WDF, ESBA GRSS, mTal and mRIS in two papers; the remaining 11 scores were only evaluated once.

Anil et al. [18] reported that hospitalised patients had significantly higher WBSS than those discharged, as assessed by an odds ratio (OR). There were significant differences between those classified as mild, moderate and severe (according to the WBSS) and a control group, for the pulse rate, respiratory rate and oxygen saturation. They also reported significant differences in the pH & pCO<sub>2</sub> between those with a severe WBSS score compared to the control, and mild & moderate bronchiolitis severity group. Kubota et al. [31] found that the WBSS had a moderate discriminative validity at differentiating among those hospitalised who required respiratory support. They additionally reported that the median WBSS score among those hospitalised who required respiratory support was modestly statistically significantly higher. Jacob et al. [30] reported that the WBSS was moderately associated with nasogastric tube feeding according to its OR, but this result was not statistically significant (i.e.  $p > 0.01$ ). They also reported that the WBSS did not significantly predict desaturation days during hospitalisation. Somech et al. [42] reported statistically significant differences in the mean WBSS among those ambulatory, hospitalised and admitted to the PICU.

Balaguer et al. [19] found that the BROSJOD score had a moderate validity, as assessed by its volume under the surface (VUS), at discriminating by expert classification at admission, and a high validity after 24 and 48 hours. They also found statistically significant associations between the score & hospital length of stay (LOS), PICU LOS and need for invasive mechanical ventilation; however, they found no association with need for non-invasive ventilation. Broadly consistent with these findings, Ricart et al. [37] found large statistically significant differences in the mean LOS, days of oxygen therapy, days of nasogastric tube feeding and maximum mean fraction of inspired oxygen (F<sub>I</sub>O<sub>2</sub>) among those with a more severe BROSJOD score. There were also large statistically significant differences in the percentage of those with a more severe BROSJOD score who were admitted to the PICU or required ventilation. Also, Rodriguez-Gonzalez et al. [39] reported that the BROSJOD score had a moderate ability at discriminating by need for respiratory support, but did not significantly correlate with PICU admission.

Bueno-Campaña et al. [20] found that a high WDF was moderately correlated with the need for respiratory support as assessed by its relative risk. Similarly, Rivas-Juesas et al. [38] reported that the WDF & ESBA at admission both had a moderate ability at discriminating between those classified as severe and non-severe. They also found the mean WDF & ESBA score at admission in the severe and non-severe group to be statistically significantly higher. However, Ramos-Fernández et al. [36] reported that the ESBA score at admission only had a poor ability at discriminating by admission to the PICU, but that the highest ESBA was highly discriminative.

Caserta et al. [21] reported a high discriminative validity of the GRSS, as assessed by its AUROC, at predicting admission and similar results when a sub-group analysis was conducted in those  $\leq 3$  & 3-10 months. Unfortunately however, they didn't report the CIs. They also found statistically significant difference in mean GRSS among those admitted to the PICU and those hospitalised but not admitted to the PICU. When externally validated by Kubota et al. [31], they found that the GRSS (as well as the WBSS) had a moderate discriminative validity at differentiating among those hospitalised who required respiratory support. They additionally reported that the median GRSS (and WBSS) score among those hospitalised who required respiratory support was modestly statistically significantly higher.

McCallum et al. [33] reported the mTal had a low-moderate discriminative ability as measured by the point estimate of the AUROC at predicting oxygen need at 12 hours and 24 hours; however, the confidence intervals (CIs) of the AUROCs are so wide, we ignored their results. When externally validated by Golan-Tripto et al. [29] it was found that it had overall a moderate discriminative validity at differentiating based on need for oxygen support and hospital LOS  $\geq 72$  hours. Notably, the discriminative validity for oxygen support (but not hospital LOS) was statistically significantly higher among those with greater experience.

Chong et al. [22] reported that the mRIS, a modified version of the Tal score (albeit different from the modified Tal score [mTal]) had a fair ability at discriminating between those who required non-invasive respiratory support, but a poor ability at discriminating by admission, intravenous hydration and LOS  $\geq 2$  days. Another publication [23] using a subset of the same dataset similarly reported a poor ability of the mRIS at discriminating by admission.

Freire et al. [27] reported that their model had a moderate ability at discriminating among those hospitalised who required escalated care and those who didn't; the performance was similar when internally validated using bootstrap validation. When a modified version of Freire's model was evaluated by Raita et al. [35], it was found to have a low ability at discriminating by positive pressure ventilation and intensive treatment use. Raita et al. [35] also reported validity data for the 4 machine learning models they developed; all of the models had moderate discriminative ability at discriminating by positive pressure ventilation use and intensive treatment use.

Destino et al. [24] reported a low discriminative ability, as assessed by its AUROC, for the Children's Hospital of Wisconsin Respiratory score (CHWRS) and RDAI at predicting admission. Duarte-Dorado et al. [25] reported statistically significant, albeit modest, differences in median mWCAS among patients at admission and discharge, and those hospitalised who required admission to the PICU. Özkaya et al. [34] reported that mBSS, a modified version of the WBSS, was moderately associated with admission, as assessed by the AUROC. Amat et al. [17] reported that the Wainwright severity score on admission had a moderate association with hospitalisation (assessed using an unadjusted OR) and that those admitted to the PICU had a statistically significantly higher severity score compared to those hospitalised but not admitted to the PICU. Univariate analysis also identified a correlation with need for intensive care (but the magnitude was not reported) but not with LOS.

### *Convergent Validity*

14 studies assessed convergent validity; only the mTal, BROSJOD & WBSS score were assessed more than once.

Golan-Tripto et al. [29] found the mTal to moderately correlate with duration of oxygen therapy, and hospital LOS, but also reported significant variation by clinical seniority. El Basha et al. [26] found a strong correlation, as measured by the Spearman's correlation coefficient, between the mTal & the duration of oxygen therapy; the correlation was statistically significantly stronger in term infants compared to pre-term infants. However, McCallum et al. [33] reported only a weak correlation between the mTal score and hospital LOS.

Anil et al. [18] reported that WBSS moderately correlated with hospital LOS. Jacob et al. [30] reported that the WBSS was the greatest predictor of hospital LOS however a quantitative measure of its predictive ability was not reported; regardless this finding was overall insignificant (i.e.  $p > 0.01$ ).

Balaguer et al. [19] also reported that Wood Downe's score strongly correlated with the BROSJOD score at admission, 24 hours and 48 hours. They also reported that it significantly correlated with hospital and PICU LOS although the magnitude was not reported. Rodriguez-Gonzalez et al. [39] found the BROSJOD score to be moderately correlated with hospital LOS and duration of respiratory support, but to not correlate with PICU LOS.

A single study reported on the convergent validity of the mWCAS, Tal score, Wainwright severity score, GRSS, CHWRS, RDAI, ESBA score, WDF score, Clinical Asthma Score (CAS), and BSS.

Duarte-Dorado et al. [25] found the mWCAS and Tal score to be strongly correlated at both admission and discharge. Amat et al. [17] reported that the initial Wainwright severity score was not significantly correlated with hospital LOS on univariate analysis. Caserta et al. [21] found the GRSS to be moderately correlated with hospital LOS. Destino et al. [22] found both the CHWRS & RDAI at admission to not correlate with LOS. Rivas-Juesas et al. [38] found the ESBA & WDF scores to be weakly correlated with each other. Marguet et al. [32] found the CAS to be only weakly correlated with hospital LOS. Siraj et al. [41] reported that the BSS was not correlated with hospital LOS, weight-adjusted high-flow nasal canula flow rate or duration of high-flow nasal canula therapy.

#### *Criterion-concurrent Validity*

Only 3 studies assessed criterion-concurrent validity. Shete et al. [40] reported the mTal score to be strongly correlated with oxygen saturation. Balaguer et al. [19] reported a strong correlation, unusually assessed via the Kappa index, between the BROSJOD score & expert opinion at admission, 24 hours and 48 hours. Gal et al. [28] reported that the mRDAI was correlated with PtcCO<sub>2</sub>; this correlation remained after controlling for PvCO<sub>2</sub> and weight.

#### *TRIPOD: Quality of reporting*

The quality of reporting of the included papers, as assessed by the TRIPOD score of the included articles, was poor; the mean TRIPOD score was 53% (see *Table 3* for overall TRIPOD scores, and *ANNEX II* for detailed TRIPOD scores). Particularly poor was reporting of model calibration, information around missing data, and summary characteristics of candidate predictors/score parameters. All, but one [35], of the included papers did not state use of the TRIPOD, or other similar checklists, in their reporting.

#### *PROBAST: Risk of Bias & Applicability*

The overall risk of bias & applicability classifications, as assessed using the PROBAST framework, for each included paper is listed in *Table 5* (see *ANNEX III* for detailed PROBAST scores). All of the included papers had either serious methodological issues, most commonly in their analysis, or a poor quality of reporting so that a judgement of the quality couldn't be made. The major methodological issues were small sample sizes, specifically with the datasets including few participants with the outcomes being predicted for, and as noted above, lack of sufficient reporting of calibration measures, quantity of missing data and, procedures for missing data.

**Table 5:** Summary PROBAST Results. Table *adapted* from Moons et al. [19].

(+ indicates low ROB/low concern regarding applicability; - indicates high ROB/high concern regarding applicability; and ? indicates unclear ROB/unclear concern regarding applicability)

(ROB-Risk of bias)

First author surname, year	ROB				Application			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
Amat, 2014 [17]	+	+	+	-	+	+	+	-	+
Anil, 2017 [18]	+	+	+	-	+	+	+	-	+
Balaguer, 2017 [19]	+	+	+	-	+	+	+	-	+
Bueno-Campaña, 2019 [20]	-	+	+	-	-	+	+	-	-
Caserta, 2017 [20]	+	-	-	-	+	+	+	-	+
Chong, 2017 [21]	+	+	-	-	+	+	+	-	+
Chong, 2018 [22]	+	-	+	-	+	+	+	-	+
Destino, 2012 [24]	+	?	+	-	+	+	+	-	+
Duarte-Dorado, 2013 [25]	+	+	+	-	+	+	+	-	+
El-Basha, 2019 [26]	+	+	+	-	+	+	+	-	+
Freire, 2018 [27]	-	-	-	-	+	-	+	-	-
Gal, 2015 [28]	+	?	-	-	+	+	-	-	-
Golan-Tripto, 2018 [29]	+	+	+	-	+	+	+	-	+
Jacob, 2016 [30]	+	-	+	-	+	+	+	-	+
Kubota, 2021 [31]	-	-	?	-	-	-	+	-	-
Marguet, 2009 [32]	+	-	+	-	+	-	+	-	-
McCallum, 2013 [33]	+	?	+	-	+	+	+	-	+

Özkaya, 2020 [34]	+	?	+	-	+	+	?	-	?
Raita, 2020 [35]	-	-	+	-	-	-	+	-	-
Ramos-Fernández, 2018 [36]	+	-	+	-	+	+	+	-	+
Ricart, 2013 [37]	+	+	+	-	+	+	+	-	+
Rivas-Jueas, 2018 [38]	+	+	+	-	+	+	+	-	+
Rodriguez-Gonzalez, 2022 [39]	-	-	+	-	+	+	+	-	+
Shete, 2014 [40]	+	+	+	?	+	+	+	?	+
Siraj, 2020 [41]	-	-	-	-	+	-	+	+	+
Somech, 2006 [42]	+	?	+	-	+	+	+	-	+

## 5. Discussion

We identified 24 unique scores from 26 articles and found that none of the identified scores were sufficiently validated. Across all three domains, the most promising score was the BROSJOD score, however it does require further validation. The mTal score was the next best validated score. It is relevant to note the high degree of similarity in the parameters in these two scores. The methodological quality of all the included studies and the quality of reporting, systematically assessed using the PROBAST and TRIPOD checklists, respectively, was poor. The most commonly used score, the RDAI score, had very weak discriminative ability (borderline poor) and only weak convergent-criterion validity; we do not recommend further effort being taken to validate this score or its use.

Our finding that there is no sufficiently validated score is consistent with all of the previous reviews. The most promising scores we identified, namely BROSJOD & mTal, were similarly identified by Hakizimana et al. [7]; they, however, also concluded that the Tal score and the Liverpool Infant Bronchiolitis Severity Score (LIBSS) (see below) were promising. In comparison to Bekhof, Reimink and Brand's [3], and Rodríguez-Martínez, Sossa-Briceño and Nino's [6] review we included far fewer papers (and scores); this was primarily due to our more stringent inclusion criteria and our specific focus only on validity data rather than data reporting on the responsiveness, usability or reliability of the scores. In contrast, however, we included more than triple the number of papers included by Hakizimana et al.'s rapid review [7] & Luarte-Martínez et al.'s systematic review [5]. Our findings on the geographic distribution of the data sources used to validate these scores concurs with the findings of Hakizimana et al. [7]. However, the best validated scores identified above seem feasible to implement in low-resources settings.

During the course of our searches, two additional promising scores were identified, the LIBSS and the ReSVinet score, but unfortunately no studies evaluating their validity met our inclusion criteria. The first was developed as a part of a PhD dissertation based on a comprehensive literature review, consultations with stakeholders, Delphi exercise and usability assessment, and then subsequently validated in a multicentre (n=11) prospective cohort study but no identified peer-reviewed full-text article was identified [43]. The other promising score, the ReSVinet score, has been validated in the original study that proposed and subsequently externally validated in a number of high-quality peer-reviewed studies (e.g. [45]). However, all of the identified studies using this score used a sample of infants with any acute respiratory infection, not specifically either RSV or bronchiolitis. A recently published multicentre (n=4) cross-sectional study evaluating the two scores in Rwanda in a small sample of infants with respiratory difficulty (n=100) found them to both have comparable convergent and discriminative validity to the BROSJOD score [45].

There are some limitations of this review. The major limitations of our review was the restriction of included papers to only those published in English, and not searching the grey literature; this likely means that some relevant papers may not have been included.

Further research is required to externally validate the BROSJOD, mTal, LIBSS & ReSVinet scores, ideally in low-income countries, and in primary care settings. The study designs should be guided by the PROBAST checklist or other similar tools and report their findings in accordance with the TRIPOD checklist or other similar tools to ensure the studies are both well designed and communicated. Given that there are a number of promising scores, the scientific community should initially focus on validating or improving these scores and only, if necessary, work on proposing new scores. Additionally, ideally when assessing the validity of these scores, it would be useful if analyses were also done with a threshold on the time of the outcome assessment (e.g. discriminative validity of a score at predicting ICU admission within 24 hours of taking the score), as the course of the disease is not always linear and may lead to systematic under estimation or overestimation of the actual validity of the score.

## 6. References

1. Borchers AT, Chang C, Gershwin ME, Gershwin LJ. Respiratory Syncytial Virus—A Comprehensive Review. *Clinical Reviews in Allergy & Immunology* [Internet]. **2013** [cited 2019 Sep 5]; 45(3):331–379. Available from: <https://link.springer.com/article/10.1007%2Fs12016-013-8368-9>
2. Li Y, Wang X, Blau DM, et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in children younger than 5 years in 2019: a systematic analysis. *Lancet (London, England)* [Internet]. **2022** [cited 2022 Jul 3]; 399(10340):2047–2064. Available from: <https://pubmed.ncbi.nlm.nih.gov/35598608/>
3. Bekhof J, Reimink R, Brand PLP. Systematic review: Insufficient validation of clinical scores for the assessment of acute dyspnoea in wheezing children. *Paediatric Respiratory Reviews*. 2014; 15(1):98–112.
4. Streinmer DL. A Checklist for Evaluating the Usefulness of Rating Scales. *The Canadian Journal of Psychiatry*. **1993**; 38(2):140–148.
5. Duarte-Martínez S, Rodríguez-Núñez I, Astudillo P, Manterola C. Psychometric properties of scales used for grading the severity of bronchial obstruction in pediatrics: A systematic review and meta-analysis. *Archivos Argentinos de Pediatría*. **2017**; 115(03).
6. Rodríguez-Martínez CE, Sossa-Briceño MP, Nino G. Systematic review of instruments aimed at evaluating the severity of bronchiolitis. *Paediatric respiratory reviews* [Internet]. **2018**; 25:43–57. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5557708/>
7. Hakizimana B, Saint G, Miert C van, Cartledge P. Can a Respiratory Severity Score Accurately Assess Respiratory Distress in Children with Bronchiolitis in a Resource-Limited Setting? *Journal of Tropical Pediatrics* [Internet]. **2020** [cited 2022 Feb 23]; 66(2):234–243. Available from: <https://academic.oup.com/tropej/article/66/2/234/5570307#:~:text=Though%20scores%20should%20measure%20what>
8. Roberts JN, Graham BS, Karron RA, et al. Challenges and opportunities in RSV vaccine development: Meeting report from FDA/NIH workshop. *Vaccine* [Internet]. **2016** [cited 2022 Feb 23]; 34(41):4843–4849. Available from: <https://pubmed.ncbi.nlm.nih.gov/27566900/>
9. Mazur NI, Higgins D, Nunes MC, et al. The respiratory syncytial virus vaccine landscape: lessons from the graveyard and promising candidates. *The Lancet Infectious Diseases* [Internet]. **2018**; 18(10):e295–e311. Available from: <https://www.sciencedirect.com/science/article/abs/pii/S1473309918302925>
10. Öner D, Drysdale SB, McPherson C, et al. Biomarkers for Disease Severity in Children Infected With Respiratory Syncytial Virus: A Systematic Literature Review. *The Journal of Infectious Diseases* [Internet]. **2020** [cited 2022 Feb 23]; 222(Suppl 7):S648–S657. Available from: <https://pubmed.ncbi.nlm.nih.gov/32794555/>
11. Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. Available from: [www.covidence.org](http://www.covidence.org).
12. World Bank. World Bank Country and Lending Groups – World Bank Data Help Desk [Internet]. Worldbank.org. 2022. Available from:

<https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>

13. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*. **2019**; 170(1):51.
14. Collins GS, Reitsma JB, Altman DG, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Medicine*. **2015**; 13(1):1.
15. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Annals of Internal Medicine*. **2015**; 162(1):W1.
16. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Annals of Internal Medicine*. **2019**; 170(1):W1.
17. Amat F, Henquell C, Verdán M, Roszyk L, Mulliez A, Labbé A. Predicting the severity of acute bronchiolitis in infants: Should we use a clinical score or a biomarker? *Journal of Medical Virology* [Internet]. **2014** [cited 2022 Feb 23]; 86(11):1944–1952. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7167168/>
18. Anil M, Göç Z, Avcı R, et al. B-type natriuretic peptide is a useful biomarker predicting disease severity in children with isolated bronchiolitis in the emergency department. *The Turkish Journal of Pediatrics*. **2017**; 59(5):561.
19. Balaguer M, Alexandre C, Vila D, et al. Bronchiolitis Score of Sant Joan de Déu: BROSJOD Score, validation and usefulness. *Pediatric Pulmonology*. **2016**; 52(4):533–539.
20. Bueno-Campaña M, Sainz T, Alba M, et al. Lung ultrasound for prediction of respiratory support in infants with acute bronchiolitis: A cohort study. *Pediatric Pulmonology*. **2019**; 54(6):873–880.
21. Caserta MT, Qiu X, Tesini B, et al. Development of a Global Respiratory Severity Score for Respiratory Syncytial Virus Infection in Infants. *The Journal of Infectious Diseases* [Internet]. **2017** [cited 2022 Feb 23]; 215(5):750–756. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5388274/>
22. Chong S-L, Teoh OH, Nadkarni N, et al. The modified respiratory index score (RIS) guides resource allocation in acute bronchiolitis. *Pediatric Pulmonology*. **2017**; 52(7):954–961.
23. Chong S-L, Lai OF, Castillo L, et al. Nasal high-mobility group box 1 and caspase in bronchiolitis. *Pediatric Pulmonology*. **2018**; 53(12):1627–1632.
24. Destino L, Weisgerber MC, Soung P, et al. Validity of respiratory scores in bronchiolitis. *Hospital Pediatrics* [Internet]. **2012** [cited 2022 Feb 23]; 2(4):202–209. Available from: <https://pubmed.ncbi.nlm.nih.gov/24313026/#:~:text=Conclusions%3A%20The%20CHWRS%20had%20modest>
25. Duarte-Dorado DM, Madero-Orostegui DS, Rodríguez-Martínez CE, Nino G. Validation of a scale to assess the severity of bronchiolitis in a population of hospitalized infants. *The Journal of asthma: official journal of the Association for the Care of Asthma* [Internet]. **2013** [cited 2021

- Nov 21]; 50(10):1056–1061. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4000565/>
26. El Basha NR, Marzouk H, Sherif MM, El Kholly AA. Prematurity, a significant predictor for worse outcome in viral bronchiolitis: a comparative study in infancy. *The Journal of the Egyptian Public Health Association* [Internet]. **2019** [cited 2022 Feb 23]; 94:15. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7091660/>
  27. Freire G, Kuppermann N, Zemek R, et al. Predicting Escalated Care in Infants With Bronchiolitis. *Pediatrics* [Internet]. **2018**; 142(3):e20174253. Available from: <https://pediatrics.aappublications.org/content/142/3/e20174253>
  28. Gal S, Riskin A, Chistyakov I, Shifman N, Srugo I, Kugelman A. Transcutaneous PCO<sub>2</sub> monitoring in infants hospitalized with viral bronchiolitis. *European Journal of Pediatrics*. **2014**; 174(3):319–324.
  29. Golan-Tripto I, Goldbart A, Akel K, Dizitzer Y, Novack V, Tal A. Modified Tal Score: Validated score for prediction of bronchiolitis severity. *Pediatric Pulmonology*. **2018**; 53(6):796–801.
  30. Jacob R, Bentur L, Brik R, Shavit I, Hakim F. Is capnometry helpful in children with bronchiolitis? *Respiratory Medicine*. **2016**; 113:37–41.
  31. Kubota J, Hirano D, Okabe S, et al. Utility of the Global Respiratory Severity Score for predicting the need for respiratory support in infants with respiratory syncytial virus infection. *PLoS ONE* [Internet]. **2021** [cited 2022 Feb 23]; 16(7):e0253532. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8248615/>
  32. Marguet C, Lubrano M, Gueudin M, et al. In Very Young Infants Severity of Acute Bronchiolitis Depends On Carried Viruses. *PLoS ONE* [Internet]. **2009** [cited 2022 Feb 23]; 4(2):e4596. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2644758/>
  33. McCallum GB, Morris PS, Wilson CC, et al. Severity scoring systems: are they internally valid, reliable and predictive of oxygen use in children with acute bronchiolitis? *Pediatric Pulmonology* [Internet]. **2013** [cited 2022 Feb 23]; 48(8):797–803. Available from: <https://pubmed.ncbi.nlm.nih.gov/22949369/>
  34. Özkaya AK, Yilmaz HL, Kendir ÖT, Gökay SS, Eyüboğlu İ. Lung Ultrasound Findings and Bronchiolitis Ultrasound Score for Predicting Hospital Admission in Children With Acute Bronchiolitis. *Pediatric Emergency Care*. **2018**; Publish Ahead of Print.
  35. Raita Y, Camargo CA, Macias CG, et al. Machine learning-based prediction of acute severity in infants hospitalized for bronchiolitis: a multicenter prospective study. *Scientific Reports* [Internet]. **2020** [cited 2022 Feb 23]; 10(1):10979. Available from: <https://www.nature.com/articles/s41598-020-67629-8>
  36. Ramos-Fernández JM, Piñero-Domínguez P, Abollo-López P, et al. Validation study of an acute bronchiolitis severity scale to determine admission to a paediatric intensive care unit. *Anales de Pediatría (English Edition)*. **2018**; 89(2):104–110.
  37. Ricart S, Marcos MA, Sarda M, et al. Clinical risk factors are more relevant than respiratory viruses in predicting bronchiolitis severity. *Pediatric Pulmonology*. **2012**; 48(5):456–463.

38. Rivas-Juesas C, Rius Peris JM, García AL, et al. A comparison of two clinical scores for bronchiolitis. A multicentre and prospective study conducted in hospitalised infants. *Allergologia et Immunopathologia*. **2018**; 46(1):15–23.
39. Rodriguez-Gonzalez M, Rodriguez-Campoy P, Estalella-Mendoza A, Castellano-Martinez A, Flores-Gonzalez JC. Characterization of Cardiopulmonary Interactions and Exploring Their Prognostic Value in Acute Bronchiolitis: A Prospective Cardiopulmonary Ultrasound Study. *Tomography*. **2022**; 8(1):142–157.
40. Shete S, Nagori G, Nagori P, Hamid M. Relation between pulse oximetry and clinical score in infants with acute bronchiolitis. *National Journal of Physiology, Pharmacy and Pharmacology*. **2014**; 4(2):124.
41. Siraj S, Stark W, McKinley SD, Morrison JM, Sochet AA. The bronchiolitis severity score: An assessment of face validity, construct validity, and interobserver reliability. *Pediatric Pulmonology*. **2021**; 56(6).
42. Somech R, Tal G, Gilad E, Mandelberg A, Tal A, Dalal I. Epidemiologic, socioeconomic, and clinical factors associated with severity of respiratory syncytial virus infection in previously healthy infants. *Clinical Pediatrics* [Internet]. **2006** [cited 2022 Feb 23]; 45(7):621–627. Available from: <https://pubmed.ncbi.nlm.nih.gov/16928839/>
43. Miert C van. Measuring Clinical Severity in Infants with Bronchiolitis [Internet]. **2015** [cited 2022 Aug 20]. Available from: [https://livrepository.liverpool.ac.uk/2037906/1/vanMiertCla\\_June2015\\_2037906.pdf](https://livrepository.liverpool.ac.uk/2037906/1/vanMiertCla_June2015_2037906.pdf)
44. Justicia-Grande AJ, Pardo-Seco J, Cebey-López M, et al. Development and Validation of a New Clinical Scale for Infants with Acute Respiratory Infection: The ReSVinet Scale. *PloS One* [Internet]. **2016** [cited 2022 Mar 1]; 11(6):e0157665. Available from: <https://pubmed.ncbi.nlm.nih.gov/27327497/>
45. Hakizimana B, Kalimba E, Ndatinya A, Saint G, Miert C van, Cartledge PT. Field testing two existing, standardized respiratory severity scores (LIBSS and ReSViNET) in infants presenting with acute respiratory illness to tertiary hospitals in Rwanda – a validation and inter-rater reliability study. *PLoS ONE* [Internet]. **2021** [cited 2022 Mar 4]; 16(11):e0258882. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8568200/>

## ANNEXES

## ANNEX I. Search Strategies

The search strategies employed were adapted from a recent systematic review on biomarkers for disease severity in RSV by Öner et al. [10].

Assistance in adapting them was kindly provided by Prof Harish Nair & Ruth Jenkins.

Searches were conducted on the 24<sup>th</sup> of June 2022 using the Ovid platform.

Ovid MEDLINE(R) <1946 to June Week 3 2022>

- 1 Respiratory Syncytial Viruses/ or Respiratory Syncytial Virus, Human/ or exp Respiratory Syncytial Virus Infections/ or RSV.mp.
- 2 Pneumonia/ or exp Pneumonia, Viral/
- 3 exp Bronchiolitis/
- 4 1 or 2 or 3
- 5 "severity of illness index"/ or severity.mp.
- 6 Prediction.mp.
- 7 5 or 6
- 8 4 and 7
- 9 limit 8 to (english language and humans and yr="2000 -Current" and "all infant (birth to 23 months)")

Embase <1974 to 2022 June 23>

- 1 Respiratory syncytial pneumovirus/ or Human respiratory syncytial virus/ or exp respiratory syncytial virus infection/ or RSV.mp.
- 2 exp pneumonia/
- 3 exp bronchiolitis/
- 4 1 or 2 or 3
- 5 exp disease severity/ or exp health status indicator/ or exp disease severity assessment/ or severity.mp.
- 6 exp computer prediction/ or exp prediction/ or prediction.mp.
- 7 5 or 6
- 8 4 and 7
- 9 limit 8 to (human and english and yr="2000 -Current" and infant)

Global Health <1973 to 2022 Week 25>

- 1 respiratory syncytial virus/ or exp human respiratory syncytial virus/ or RSV.mp.
- 2 exp pneumonia/
- 3 bronchiolitis/
- 4 1 or 2 or 3
- 5 exp prediction/ or prediction.mp.
- 6 severe course/ or severe infections/ or severity.mp.
- 7 5 or 6
- 8 exp infants/
- 9 4 and 7 and 8
- 10 limit 9 to (english language and english and yr="2000 -Current")

## ANNEX II. Detailed TRIPOD Scores

TRIPOD checklist: Collins et al. [14]  
 Explanation & advice: Moons et al. [15]

**Table:** TIRPOD for included studies. *Adapted* from Collins et al. [14].

(D – Development; NA – Not applicable; V – Validation). Greyed out cells indicates questions that are not included when calculating the overall score according to TRIPOD recommendations.

First author or surname, year	Type	1	2	3a	3b	4a	4b	5a	5b	5c	6a	6b	7a	7b	8	9	10a	10b	10c	10d	10e	11	12	13a	13b	13c	14a	14b	15a	15b	16	17	18	19a	19b	20	22
Amat, 2014 [17]	V	0	0	0	0	1	1	1	1	NA	1	0	1	0	1	0	NA	NA	1	0	NA	1	0	1	0	0	NA	NA	NA	NA	0	NA	1	0	1	0	0
Anil, 2017 [18]	V	0	0	1	0	1	1	1	1	NA	1	1	1	1	0	0	NA	NA	1	0	NA	1	0	1	0	0	NA	NA	NA	NA	0	NA	1	0	1	0	0
Balaguer, 2017 [19]	V	0	0	1	1	1	1	1	1	NA	1	1	1	1	0	0	NA	NA	1	0	NA		0	1	0	0	NA	NA	NA	NA	0	NA	1	0	1	1	0
Bueno-Campana, 2019 [20]	V	0	0	0	0	1	1	1	1	NA	1	0	0	0	1	0	NA	NA	1	0	NA	1	0	1	0	0	NA	NA	NA	NA	0	NA	1	0	1	1	0
Casearta,	D	0	0	1	1	1	1	1	1	NA	1	0	1	0	0	1	1	0		0		1		1	0		1	NA	1	1	0		1		1	0	







## ANNEX III. Detailed PROBAST Scores

PROBAST checklist: Wolff et al. [13]  
 Explanation & advice: Moons et al. [16]

**Table:** PROBAST for included studies. Adapted from Wolff et al. [13].

(D – Development; N – No; NA – Not applicable; NI – No information; PN – Probably No; PY – Probably Yes; V – Validation; Y – Yes)

Grey shaded box indicates inapplicability according to PROBAST recommendations.

First author surname, year	D / V	Participants				Predictors					Outcome						Analysis									Overall				
		1	1	2		2	2	2	Risk of bias introduced by predictors or their assessment	Concern that the definition, assessment or timing of predictors in the model do not match the review question	3	3	3	3	3	3	Risk of bias introduced by the outcome or its determination	Concern that the outcome, its definition, timing or determination do not match the review question	4.1	4	4	4	4	4	4	4	4	4	4.9	Risk of bias introduced by the analysis
Amat, 2014 [17]	V	Y	Y	Low	Low	PY	PN	Y	Low	Low	Y	Y	Y	Y	PN	PY	Low	Low	N	Y	Y	N		NI	NI			High	High	Low
		Rationale:				Rationale:					Rationale: Highly likely that score components affected the decision to admit (3.5).						Rationale: Only 86 included patients required oxygen therapy & 11 admitted to PICU (4.1). No procedure specified about missing data (4.4).													
Anil,	V	Y	Y	Low	Low	P	Y	Y	Low	Low	Y	Y	Y	Y	P	Y	Low	Low	PY	Y	Y	N		NI	NI			High	High	Low



2012 [24]		Rationale:	Rationale: Not specified that score assessors were blinded of patient disposition, or that scores taken prior to disposition decision (2.1).	Rationale: Highly likely that score components in the score affected the outcome decision (3.5).	Rationale: Only 6/56 were admitted to PICU (4.1). Not all enrolled patients had a score completed in the ED and so were not all included (4.3). No procedure specified about missing data (4.4).		
Duarte - Dorado, 2013 [25]	V	Y Y Low Low	Y Y Y Low Low	Y Y Y Y P N Y Low Low	N Y P N N N High	High	Low
		Rationale:	Rationale:	Rationale: Highly likely that score components in the score affected the outcome decision (3.5).	Rationale: Only 6/56 were admitted to PICU (4.1). No procedure specified about missing data (4.4).		
El-Basha, 2019 [26]	V	Y Y Low Low	Y P Y Low Low	Y Y Y Y P N Y Low Low	N Y P N N N High	High	Low
		Rationale:	Rationale:	Rationale: Highly likely that score components in the score affected the outcome decision (3.5).	Rationale: Only 24/153 admitted to PICU & 16/153 ventilated; no information on number who received oxygen therapy (4.1). No procedure specified about missing data (4.4).		
Freire, 2018 [27]	D	N Y High Low	P N Y High High	Y N Y P P P High Low	N Y N Y Y N N High High	High	High
		Rationale: Retrospective cohort chart review (1.1).	Rationale: Retrospective chart review (2.1).	Rationale: Used combi novel not clearly pre-specified outcome (3.2). Decision to give escalated care (high flow nasal canula, non-invasive ventilation and/or ICU admission) varies especially given multinational dimension (3.4). Retrospective cohort chart review (3.5).	Rationale: Only 261 patients with escalated care & 10 candidate predictors evaluated (4.1). Only Hosmer-Lemeshow test used to evaluate calibration (4.7). Interval validation only used on model not parameter selection (4.8).		
Gal, 2015 [28]	V	Y Y Low Low	Y N Y Unclear Low	N Y Y Y N N High High	N Y N N N N High	High	High
		Rationale:	Rationale:	Rationale: No clear outcome for validating clinical score (3.1).	Rationale: Only 60 participants in study (4.1). Excluded from analyses patients whose PtcCO <sub>2</sub> values were obtained under inadequate conditions (4.4).		
Golan-Tripto, 2018 [29]	V	Y Y Low Low	Y Y Y Low Low	Y Y Y Y P N Y Low Low	N Y N N N N High	High	Low
		Rationale:	Rationale:	Rationale: Highly likely that score components in the score affected the outcome decision (3.5).	Rationale: 44/50 received oxygen support ≥48 h; not specified how many had a LOS ≥72 hours (4.1). No procedure specified about missing data (4.4).		
Jacob, 2016 [30]	V	Y Y Low Low	Y N Y High Low	Y Y Y Y P N Y Low Low	N Y N N N N High	High	Low
		Rationale:	Rationale: Not blinded (2.2).	Rationale: Highly likely that score components in the score affected the outcome decision (3.5).	Rationale: 12/38 received nasogastric tube feeding (4.1). No procedure specified about missing data (4.4).		
Kubota,	V	N N High High	P N Y High High	Y Y Y Y P N Unclear Low	N Y N N N N High	High	High

2021 [31]		Rationale: Retrospective study (1.1). Excluding patients with missing score-items (1.2).	Rationale: Retrospective chart review (2.1, 2.2).	Rationale: Highly likely that score components in the score affected the outcome decision (3.5).	Rationale: 26/250 received respiratory support (4.1). Excluded patients for which GRSS score components not in chart (4.4)		
Marguet, 2009 [32]	V	Y Y Low Low	Y N N High High	Y Y Y Y P N Y Low Low	N Y N N N N High	High	High
		Rationale:	Rationale: Highest score was taken during hospitalisation, and so admission status would be known (2,2). Chest x-ray used (2.3).	Rationale: Highly likely that score components in the score affected the outcome decision (3.5).	Rationale: Only 96 patients (46.9%) required oxygen supplementation (4.1). No procedure specified about missing data (4.4).		
McCallum, 2013 [33]	D	Y Y Low Low	Y N Y Uncle ar Low	Y Y Y Y P N Y Low Low	NI Y P N Y N N N NA High	High	Low
		Rationale:	Rationale:	Rationale: Highly likely that score components in the score affected the outcome decision (3.5).	Rationale: 4.1 Unclear how many infants required oxygen at each time point. No procedure specified about missing data (4.4). Neither univariate not multivariate analysis used in predictor selection (4.9).		
Özkaya, 2020 [34]	V	Y Y Low Low	Y N Y Uncle ar Low	Y Y Y Y P N Low Unclear	N Y Y N N N High	High	Unclear
		Rationale:	Rationale: 2.2 Unclear when mBSS was assessed.	Rationale: Highly likely that score components in the score affected the outcome decision (3.5).	Rationale: 42/76 patients were admitted (4.1). No procedure specified about missing data (4.4).		
Raita, 2020 [35]	D	N Y High High	P N P High High	Y Y Y Y P N Y Low Low	PN Y N Y P N N N NA High	High	High
		Rationale: Retrospective chart review & interview (1.1).	Rationale: Prospective chart review & interview (2.1, 2.2).	Rationale: Highly likely that score components in the score affected the outcome decision (3.5).	Rationale: Only 55/1,016 patients required PPV, however 163/1,016 patients required intensive treatment (4.1). Internal validation not used in predictor selection (4.8). Neither univariate not multivariate analysis used in predictor selection (4.9).		
Ramos-Fernández, 2018 [36]	V	Y Y Low Low	Y N Y High Low	Y Y Y Y P N Y Low Low	PN Y Y N N N N High	High	Low
		Rationale:	Rationale:	Rationale: Highly likely that score components in the score affected the outcome decision (3.5).	Rationale: 11/190 patients were admitted to the PICU (4.1). No procedure specified about missing data (4.4).		
Ricart, 2013 [37]	V	Y Y Low Low	Y Y Y Low Low	Y Y Y Y P N Y Low Low	PN Y N N N N N High	High	Low
		Rationale:	Rationale:	Rationale: Highly likely that score components in the score affected the outcome decision (3.5).	Rationale: 82/420 in severe group, 58/420 admitted to PICU; number of patients receiving mechanical & non-mechanical ventilation not stated (4.1). Abstract states only 410/420 included in analysis (4.3). No procedure specified about missing data (4.4).		
Rivas-Juessa,	V	Y Y Low Low	Y P Y Low Low	Y N Y Y P N Y Low Low	N Y N N N N High	High	Low
		Rationale:	Rationale:	Rationale: Used novel combi outcome of	Rationale: Only 13/201 patients with novel outcome		

2018 [38]											severity (3.2). Highly likely that score components in the score affected the outcome decision (3.5).		of severity (4.1). No procedure specified about missing data (4.4).																	
Rodriguez-Gonzalez, 2022 [39]	V	Y	N	High	Low	Y	N	Y	High	Low	Y	Y	Y	Y	P	Y	Low	Low	N	Y	Y	N		N	N			High	High	Low
		Rationale: Excluded patients with incomplete research intervention (1.2).				Rationale: Done by attending physicians				Rationale: Highly likely that score components in the score affected the outcome decision (3.5).				Rationale: 36/112 required respiratory support (4.1). No procedure specified about missing data (4.4).																
Shete, 2014 [40]	V	Y	Y	Low	Low	Y	Y	Y	Low	Low	Y	Y	Y	Y	Y	Y	Low	Low	Y	Y	Y	N		N	N			Unclear	Unclear	Low
		Rationale:				Rationale:				Rationale:				Rationale: Continuous variable outcome (SaO <sub>2</sub> ) on all 142 participants.																
Siraj, 2021 [41]	V	N	P	High	Low	P	N	Y	High	High	Y	P	Y	P	P	P	High	Low	N	Y	N	N		N	N			High	High	High
		Rationale: For validity retrospective design was used (1.1). Excluded those without BSS score (1.2)				Rationale: Retrospective design (2.2, 2.3).				Rationale: Highly likely that score parameters affected decision to admit to PICU (3.5).				Rationale: Only 18/181 had a late rescue (4.1). Patients with missing score data excluded (4.4)																
Somech, 2006 [42]	V	Y	Y	Low	Low	Y	N	Y	Unclear	Low	Y	Y	Y	Y	N	Y	Low	Low	PN	Y	N	N		N	N			High	High	Low
		Rationale:				Rationale:				Rationale:				Rationale: 100/195 infants were hospitalised, but only 13/195 infants were admitted to PICU. (4.1)																

## PART II

# **External validation of the discriminative validity of the ReSVinet score & development of simplified ReSVinet scores in secondary care**

## 1. Abstract

### *Background*

There is no consensus on how to best quantify disease severity in infants with respiratory syncytial virus (RSV) and/or bronchiolitis; this lack of a sufficiently validated score complicates the provision of clinical care and, the evaluation of trials of therapeutics and vaccines. The ReSVinet score appears to be one of the most promising; however it is too time-consuming to be incorporated into routine clinical care. We aimed to develop and externally validate simplified versions of this score.

### *Methods*

Data were used from a multinational (Netherlands, Spain & United Kingdom) multicentre case-control observational study of infants with RSV to develop simplified versions of the ReSVinet by conducting a grid search to determine the best combination of equally weighted parameters to maximise for the discriminative ability of the scores across a range of outcomes (hospitalisation, intensive care unit admission, ventilation requirement). Subsequently discriminative validity of the score for a range of secondary care outcomes was externally validated by conducting a secondary analysis of data collected in infants with respiratory infection from tertiary hospitals in Rwanda and Colombia. We compared the discriminative validity of the simplified scores to the original ReSVinet score.

### *Results*

Three candidate simplified scores were identified using the development dataset; in the development dataset they had an area under the receiver operating characteristic curve (AUROC)  $>0.9$  at discriminating for a range of outcomes, and their performance was not statistically significantly different to the original ReSVinet score despite having fewer parameters. In the external validation datasets, the simplified scores were moderate-excellent (AUROC 0.7-1) across a range of outcomes. In all outcomes, except for in a single dataset (Rwanda) at predicting admission to the high dependency unit, they performed at least as well as the original ReSVinet score.

### *Conclusions*

Three promising candidate simplified scores were developed; however further external validation work in larger datasets, ideally from resource-limited settings given this is where the greatest disease burden lies, needs to be conducted before any recommendation regarding their use.

## 2. Introduction

There remains no consensus on how best to quantify the severity of bronchiolitis and/or respiratory syncytial virus (RSV) associated acute lower respiratory tract infections in infants; this is an important challenge as such a score can be useful indicator for clinicians when providing care and, serve as clinical endpoints in clinical trials for therapeutics and, increasingly importantly, vaccines. Of the scores proposed, one of the most promising is the ReSVinet score [1-3]

The ReSVinet score was proposed almost a decade ago to quantify disease severity in infants (<24 months) with acute respiratory infections, including RSV [1]. It was designed based on an unpublished literature review of previous related scores and subsequently evaluated by 90 paediatricians ensuring high face validity. The devised score consists of 7-parameters (feeding intolerance, medical intervention, respiratory difficulty, respiratory frequency, apnoea, general condition, fever) and has an overall total score of 20. The original developers of the score later proposed thresholds (based on the findings from two studies) to indicate grades of severity: 0-6 as signifying a mild affection, 7-13 for moderate distress, and  $\geq 14$  for a severe episode [3]. Notably, this score doesn't include either oxygen saturation or heart rate, objective indicators often included in other RSV/bronchiolitis severity scores, making it more widely usable, especially in settings with limited equipment (e.g. outpatient and inpatient settings in low-income countries) [2]. A parental version of this score has also been developed, although the focus of this paper will be on the version for health professionals.

The original study that proposed the score retrospectively validated it in a small sample of 170 infants (<24 months) hospitalised in 3 centres in Spain with an acute respiratory infection; in this study they demonstrated that the score had good internal consistency, strong inter-rater reliability (between investigators and, between investigators and parents) and moderate construct validity [1]. Subsequently, a number of prospective external validation studies have been conducted showing similar results; studies assessing the validity of the clinician version of the ReSVinet score, that we are aware of, are summarised in *Table 1* [1, 4-7]. As such it currently appears to be one of the most promising and best validated severity scores for acute respiratory infections in infants.

Anecdotally, the major reported weakness of the score is that it is too time-consuming to use, making it unfeasible for integrating it into routine clinical care (although still suitable for clinical trials). Therefore, we aimed to propose and externally validate a simplified version of the ReSVinet score. We additionally externally validated the original ReSVinet score in two new datasets

**Table 1: Published studies validating the ReSVinet score.**

(AUROC- Area under the receiver operating characteristic curve; CI – confidence interval; HDU - high dependency unit; LIBSS - Liverpool Infant Bronchiolitis Severity Score; LOS – length of stay; n.d. – no date; NA - not applicable; OR – odds ratio; PICU – paediatric intensive care unit; RSV - respiratory syncytial virus; SD - standard deviation)

First author surname, year	Study design(s)	Country	Sample size	Age range (months)	Clinical setting	RSV positivity rate (%)	RSV Test	Key Findings (95% CI)
<i>Development of the ReSVinet score</i>								
Justicia-Grande, 2016 [1]	Multicentre (n=3) retrospective cohort study	Spain	170	<24	Secondary care	77.1% (only 166 tested)	Direct immunofluorescence of respiratory secretions	<p><u>Discriminative Validity</u> Patients admitted to PICU had statistically significantly (<math>p&lt;0.001</math>) higher mean score; mean score (SD) was 15.7 (<math>\pm 2.6</math>) vs 10.2 (<math>\pm 2.5</math>), 15.4 (<math>\pm 2.7</math>) vs 9.2 (<math>\pm 2.4</math>), and 14 (<math>\pm 2.6</math>) vs 9.4 (<math>\pm 2.4</math>) among investigators, and 15.9 (<math>\pm 2.5</math>) vs 10.6 (<math>\pm 2.7</math>) for parents.</p> <p><u>Convergent validity</u> ReSVinet score correlated with Wood-Downes score (<math>p&lt;0.001</math>).</p> <p>Spearman's correlation coefficients between hospital LOS &amp; ReSVinet score was 0.48–0.60 for investigators (<math>p&lt;0.001</math>) and 0.33–0.35 for parents (<math>p=0.027</math>).</p>
<i>External validation of the ReSVinet score</i>								
Hakizimana, 2021 [4]	Multicentre (n=4) prospective cross-sectional study	Rwanda	100	1-12	Tertiary care	Not reported	NA	<p><u>Discriminative Validity</u> AUROC for hospital admission was 0.973 (0.94–1.0; <math>p&lt;0.001</math>) and 0.956 (0.92–0.99; <math>p&lt;0.001</math>) for nurse and resident respectively.</p> <p>AUROC for admission to PICU or HDU was 0.880 (0.80–0.96; <math>p&lt;0.001</math>) &amp; 0.872 (0.787–0.957; <math>p&lt;0.001</math>) for nurse and resident respectively.</p> <p>AUROC for mortality was 0.974 (0.944–1.0; <math>p&lt;0.001</math>) &amp;</p>

								<p>0.980 (0.954–1.0; <math>p &lt; 0.001</math>) for nurse and resident respectively.</p> <p>AUROC for patients having a longer hospital admission than the median LOS was 0.637 (0.531–0.747; <math>p &lt; 0.001</math>) &amp; 0.639 (0.531–0.747; <math>p &lt; 0.001</math>) for nurse and resident respectively.</p> <p><u>Convergent Validity</u> Pearson's correlation coefficient between ReSVinet and LIBSS was 0.815 (0.70–0.93; <math>p &lt; 0.001</math>) &amp; 0.836 (0.73–0.95; <math>p &lt; 0.001</math>) for resident and nurse respectively.</p>
Justicia Grande, n.d. [5]	Unclear	Unclear	245	1-168 (1 month - 14 years)	Primary care	Not reported	NA	<p><u>Convergent validity</u> Pearson's correlation coefficient between ReSVinet Score and the Wood-Downes Score was 0.57 (<math>p &lt; 0.0001</math>).</p>
Tungsupreechameth, 2021 [6]	Retrospective case-control study	Thailand	269	1-60	Secondary care	100%	Immunoassay of nasopharyngeal swabs	<p><u>Discriminant validity</u> Those with a ReSVinet score <math>\geq 14</math> had significantly lower mean weight, attended day-care centre less frequently, had a significantly longer hospital stay and were more likely to require ICU and mechanical ventilation.</p> <p>Adjusted OR of 19.56 (1.81-212.05; <math>p = 0.014</math>) for hospital LOS <math>&gt; 5</math> days for those with a ReSVinet score <math>\geq 14</math> compared to those with a ReSVinet score <math>&lt; 14</math>.</p>
Vandendijck, 2022 [7]	Multicentre (n=10) prospective cohort study	Argentina, Chile, and the Unit	124	$\leq 36$	Secondary care	100%	Rapid antigen assay or PCR	<p><u>Convergent validity</u> Pearson correlation coefficient between the PRESORSv6 Clinician-Reported Outcome questionnaire overall RSV severity score and ReSVinet score was 0.45; (0.30–0.58).</p>

		ed Stat es						
--	--	------------------	--	--	--	--	--	--

### 3. Methods

#### Data Sources

For this study, we made use of three existing datasets. One recently published dataset was used to develop candidate simplified scores [8,9], and these scores were subsequently externally validated in three additional datasets, all of which have already been used to validate the original ReSVinet score and their findings published [4, 10]. Further detail on each dataset can be found below; the inclusion and exclusion criteria of the datasets are listed in *Table 2*.

**Table 2:** Eligibility Criteria for included participants in the analysis. *Adapted* from [4, 8-10].

(PCR-Polymerase chain reaction; RSV - Respiratory syncytial virus; RESCEU- REspiratory Syncytial virus Consortium in Europe)

Name	Development dataset	External validation dataset	
	RESCEU [8,9]	Camacho-Cruz [10]	Hakizimana [4]
<b>Inclusion criteria</b>	Parent/carer of infant is willing and able to give informed consent for participation in the study.	Parents were invited to participate by filling out/completing the scale and those who completed it were included.	Parents could consent.
	Male or female, aged <12 months at enrolment.	Children aged < 2 years	Infants 1 to <12 months of age
	Positive RSV PCR test of nasopharyngeal swab.	Patients presenting with signs and symptoms of any type of acute respiratory infection (rhinopharyngitis, croup, bronchiolitis, wheezing episodes or pneumonia)	Presenting with respiratory distress due to any respiratory illness. Case definition of symptoms and signs indicative of respiratory distress were: apnoea, subcostal or intercostal recession, tracheal tug, nasal flaring, head bobbing, grunting, cyanosis, oxygen desaturation, tachypnoea, wheezing, stridor, oxygen requirement, or reduced air-entry
	Hospitalised for <48 hours at enrolment.		
	Live near enough to a participating study centre for the 6-8 week home visit/hospital appointment to be feasible.		
	Parent has a telephone.		
		Parents spoke Spanish	
<b>Exclusion</b>	History of receipt		

<b>criteria</b>	of medication to treat RSV infection (e.g. ribavirin). Prior exposure to an investigational RSV vaccine or medication		
		History of significant comorbidities, for example, heart disease, chronic lung disease, anatomical lung malformation, neurological diseases, severe malnutrition, any immune deficiency or cancer.	Infants with known chronic lung disease, or infants who presented with a non-respiratory cause of respiratory distress (e.g. cardiac disease).

#### Development dataset (RESCEU)

The dataset used to develop simplified ReSVinet scores was collected as part of a recently published multinational (Netherlands, Spain & United Kingdom) multicentre case-control observational study run by the REspiratory Syncytial virus Consortium in EUrope (RESCEU) to investigate biomarkers of RSV associated acute lower respiratory tract infection severity in infants (<12 months) [8,9]. The original study aimed to recruit 500 previously healthy infants with RSV, 50 infants with comorbidities and with RSV and 80 healthy controls without RSV. Patients were recruited between October 2017 to April 2021.

For the purpose of this study, we only included the cases (i.e. infants with RSV), including those with comorbidities. The eligibility criteria for this group can be found in *Table 2*. An anonymised version of this dataset was provided by the RESCEU investigators. For each patient, basic demographic data (e.g. sex, gestational age), clinical data (e.g. heart rate, oxygen saturation) outcome data (e.g. hospitalisation, need for respiratory support) and the ReSVinet score parameters and total score at the time of recruitment were supplied. Score assessors were not blinded for outcomes if the outcomes had occurred by the time of score assessment. Outcome data were retrieved from patients' medical notes; outcomes used in our analysis were hospitalisation, admission to the intensive care unit (ICU) and requirement for synchronised intermittent mandatory ventilation (SIMV).

#### External validation datasets (Camacho-Cruz, Hakizimana and Justicia-Grande)

Subsequently we externally validated the simplified scores in two datasets, henceforth referred to as the Camacho-Cruz and Hakizimana datasets respectively. In this section, we provide a high-level summary of each dataset; however more detailed information on the methods can be found in the original study publications [4, 10].

The Camacho-Cruz dataset was collected in a prospective observational cohort study of children (<24 months) presenting to the emergency department at a single general hospital in Colombia with an acute respiratory infection with the aim of externally validating the reliability of the ReSVinet score [10]. The ReSVinet score was taken simultaneously by parents, paediatric doctor-professors (faculty) and residents in the emergency room. They were not blinded to the infant's outcome or to the other predictors. Information was gathered on enrolled infants via their electronic health records and phone

consultation 10-days post-discharge. An anonymised version of the dataset was provided by the investigators. Outcomes analysed were hospitalisation and paediatric intensive care unit (PICU) admission.

The Hakizimana dataset was collected in a prospective cross-sectional study conducted in four tertiary hospitals (n=4) in Rwanda where infants (1-12 months) presenting with signs of respiratory distress were opportunistically enrolled to externally validate the ReSVinet score and the Liverpool Infant Bronchiolitis Severity Score (LIBSS) [4]. An anonymised version of this dataset was publicly available on the Harvard Dataverse [11]. The ReSVinet score of each infant was assessed independently by a nurse and resident within an hour of the infant presenting to hospital. Further information on enrolled infants was gathered through standardised questionnaires filled in by health-care providers involved in the delivery of care to the infant; as such they were not blinded to the infant's outcomes or to the other predictors. The dataset contained no missing values. Outcomes analysed were hospitalisations, admission to the high dependency unit (HDU), admission to the PICU, intubation or ventilation and death.

### *Study design & reporting*

The design and reporting of this study was guided, as far as possible within the pre-existing constraints of the datasets, by the Prediction model Risk Of Bias ASsessment Tool (PROBAST) and the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) checklist [12,13]; for the detailed populated checklists see ANNEX I & II.

### *Simplified score*

To develop the simplified scores, we only made use of the development dataset (RESCEU). Specifically, we performed a grid search to explore all the possible combinations of parameters, and identified the equally-weighted combination of parameters giving the highest point estimate of area under the receiver operating characteristic curve (AUROC), for requirement of admission to hospital, admission to ICU, and need for SIMV.

After developing the candidate simplified scores, we assessed the AUROC for a range of outcomes in the external validation datasets (Hakizimana & Camacho-Cruz datasets).

### *Validation of the original ReSVinet score*

We additionally used the RESCEU and Camacho-Cruz datasets to externally validate the discriminative validity of the original ReSVinet score by assessing the AUROC for the outcomes mentioned above. This was not done for the Hakizimana dataset as this has already been done and reported in the original publication [4].

### *Statistical Analysis*

For each of the above measures we calculated confidence intervals (CIs) of the AUROCs and created p-value matrices. We considered a p-value <0.005 as constituting statistical significance and used DeLong's method to calculate both CIs and paired p-values between AUROCs [14]. We used R version 4.1.2 and Python version 3.6.2 to conduct the analysis.

### *Missing data*

In the development dataset, we excluded all infants with missing items for the ReSVinet score. The Hakizimana dataset had no missing data, as the participants with any missing data (n=7) had already been excluded in the dataset available publicly.

For outcome data, we excluded participants from specific analyses if the outcome was not available; in all cases we explicitly stated the number of participants included in each analysis.

## 4. Results

### Included participants

The RESCEU study enrolled 325 cases of which 14 had missing data for the ReSVinet score parameters and so were excluded; 11 had no data for any of the score parameters and the remaining 3 had no data for two parameters each. As such, overall 311 patients were included. The Camacho-Cruz study enrolled 191 patients of which three were excluded as two did not meet the inclusion criteria and one was lost to follow up; as such, overall data on 188 patients were included. The Hakizimana study recruited 107 patients, of which for 7 data collection was incomplete and so were excluded in the publicly available dataset; overall data on 100 infants were included.

### Baseline characteristics

Baseline characteristic for included participants by dataset are summarised in *Table 3*.

**Table 3:** Baseline Characteristics (Greyed out cell indicates data not available in dataset or not applicable. See ANNEX III for the number of missing values for each baseline characteristic by dataset) (%)

(SD – Standard deviation)

Characteristics	Development dataset	External validation datasets	
	RESCEU	Hakizimana	Camacho-Cruz
<b>Sex</b>			
Male	179 (58%)	63 (63%)	110 (59%)
Female	132 (42%)	37 (37%)	78 (46%)
<b>Country of site</b>			
Netherlands	135 (43%)		
Spain	28 (9%)		
United Kingdom	148 (48%)		
Rwanda		100 (100%)	
Colombia			188 (100%)
<b>Birth Body Weight (kg)</b>			
Mean (SD)	3.32 (6.39)		
Range	1.04-5.03		
<b>Current Body Weight (kg)</b>			
Mean (SD)	6.12 (2.03)	6.56 (2.43)	
Range	2-12	2-13.3	
<b>RSV Subgroup</b>			
A (%)	149 (52%)		
A/B (%)	1 (0%)		
B (%)	137 (48%)		
<b>Season</b>			
2017-18 (%)	57 (20%)		
2018-19 (%)	113 (39%)		
2019-20 (%)	141 (49%)		
<b>Gestational age (weeks)</b>			
Very preterm (28-31) (%)	8 (3%)		
Moderate preterm (32-33) (%)	5 (2%)		
Late preterm (34-	13 (4%)		

36) (%)			
Early term (37-38) (%)	87 (28%)		
Late term (39-41) (%)	191 (62%)		
Post term ( $\geq 42$ ) (%)	5 (2%)		
Premature			
<i>Age at enrolment (months)</i>			
Median (SD)	3 (3.31)	7 (3.77)	10 (7.09)
Range	0.3-12	0.9-11.9	0.4-24
$\leq 2$	112 (37%)	16 (16%)	32 (17%)
>2-4	64 (21%)	17 (17%)	15 (8%)
>4-6	39 (13%)	12 (12%)	21 (11%)
>6-8	36 (12)	16 (16%)	14 (7%)
>8-10	32 (10%)	11 (11%)	19 (10%)
>10-12	23 (8%)	28 (28%)	14 (7%)
>12-14			16 (9%)
>14-16			13 (7%)
>16-18			8 (4%)
>18-20			15 (8%)
>20-22			9 (5%)
>22-24			12 (6%)
<i>Diagnosis (%)</i>			
None	24 (8%)		
Rhinopharyngitis			76 (40%)
Bronchiolitis	128 (41%)	36 (36%)	54 (29%)
Upper respiratory tract infection	141 (46%)	1 (1%)	
Croup	1 (0%)		32 (17%)
Wheezing	1 (0%)	6 (6%)	23 (12%)
Pneumonia	13 (4%)	51 (51%)	3 (2%)
Other	1 (0%)	6 (6%)	
<i>Viral tests</i>			
RSV	311 (100%)		17 (9%)
Adenovirus			3 (2%)
Parainfluenza 1			2 (1%)
Adenovirus and RSV			1 (1%)
<i>Pre-existing condition</i>			
No	270 (87%)		
Yes	41 (13%)		
Prematurity	26 (8%)	5 (5%)	
Malnutrition		12 (12%)	
Asthmatic		1 (1%)	
Human immunodeficiency virus		2 (2%)	
Other	15 (5%)		
<i>Ubedehe Social Group</i>			
High		47 (47%)	

Low		53 (53%)	
<i>Residence</i>			
Urban		32 (32%)	
Rural		68 (68%)	
<i>Living sibling</i>			
No		20 (20%)	
Yes		80 (80%)	
<i>Vaccinations complete</i>			
No		0 (0%)	
Yes		100 (100%)	
<i>Maternal marital status</i>			
Single		15 (15%)	
Married		84 (84%)	
Divorced		1 (1%)	
Widowed		0 (0%)	
<i>Maternal age (years)</i>			
15-19		11 (11%)	
20-24		21 (21%)	
25-29		26 (26%)	
30-34		19 (19%)	
35-39		17 (17%)	
40-44		5 (5%)	
45-49		1 (1%)	
<i>Paternal age (years)</i>			
15-19		3 (3%)	
20-24		11 (11%)	
25-29		24 (24%)	
30-34		24 (24%)	
35-39		19 (19%)	
40-44		12 (12%)	
45-49		6 (6%)	
50-55		1 (1%)	
<i>Maternal occupation</i>			
Unemployed		19 (19%)	
Manual labourer		64 (64%)	
Professional		17 (17%)	
<i>Paternal occupation</i>			
Unemployed		5 (5%)	
Manual labourer		67 (67%)	
Professional		28 (28%)	
<i>Maternal educational level</i>			
None		13 (13%)	
Primary		38 (38%)	
Secondary		28 (28%)	
Higher		21 (21%)	
<i>Paternal educational level</i>			
None		9 (9%)	
Primary		42 (42%)	
Secondary		30 (30%)	
Higher		18 (18%)	

### Outcome data

The outcomes for included participants by dataset are summarised in *Table 4*.

**Table 4:** Outcomes. (First bracketed figure is the percentage, and the second figure, separated by a semicolon, is the number of patients for which data are missing. Greyed out cells indicate data not available in respective dataset.)

(HDU - high dependency unit; ICU - intensive care unit; PICU – Paediatric intensive care unit; RESCEU- REspiratory Syncytial virus Consortium in Europe; SIMV - synchronised intermittent mandatory ventilation)

Outcome	Development dataset	External Validation datasets	
	RESCEU (n=311)	Hakizimana (n=100)	Camacho-Cruz (n=188)
<b>Hospitalised (%; missing)</b>	232 (75%; 0)	93 (93%; 0)	
<b>HDU Admission (%; missing)</b>		22 (22%; 0)	
<b>ICU/PICU Admission (%; missing)</b>	86 (28%; 0)	7 (7%; 0)	5 (3%; 0)
<b>SIMV (%; missing)</b>	70 (25%; 31)		
<b>Intubation or ventilation (%; missing)</b>		5 (5%; 0)	
<b>Death</b>		6 (6%; 0)	0 (0%; 7)

### Development of simplified ReSVinet scores

Using all of the development dataset (i.e. the RESCEU dataset), the best combination of equally weighted parameters to maximise for the point-estimate of AUROC for discriminating by hospitalisation, ICU admission and SIMV requirement were determined; this produced three unique candidate simplified scores (ReSVinet-3, ReSVinet-4 and ReSVinet-6). These scores, as assessed by their AUROC, were excellent in the RESCEU dataset at discriminating by the patients' hospitalisation status, ICU admission and requirement for SIMV (See PART I of this report for cut-offs) (see *Table 5*). For ICU admission and requirement for SIMV, there were no statistically significant differences in the performance of the original and candidate scores (see *Table 6* for p-value matrices). For hospitalisations although there were no statistically significant differences between each candidate scores and the original score, there were significant differences between the candidate scores; specifically, ReSVinet-6 was marginally better than both ReSVinet-3 & ReSVinet-4, and ReSVinet-4 marginally better than ReSVinet-3.

We additionally externally validated the discriminative validity of the original ReSVinet score in the RESCEU dataset (see *Table 5*); in this dataset it was excellent at discriminating on hospitalisation, and borderline-excellent for ICU and need for SIMV.

**Table 5:** AUROC for original ReSVinet & candidate simplified scores in development (RESCEU) dataset (n=311)

(AUROC- Area under the receiver-operating characteristic curve; CI - confidence interval; ICU - intensive care unit; RESCEU- REspiratory Syncytial virus Consortium in Europe; SIMV - synchronised intermittent mandatory ventilation)

Maximising AUROC for	Name	Parameters included	Parameters excluded	AUROC (95% CI)		
				Hospitalisation	ICU Admission	SIMV requirement (n=280)
	ReSVinet	Feeding intolerance Medical intervention Respiratory difficulty Respiratory frequency Apnoea General condition Fever		0.9437 (0.9191-0.9683)	0.8995 (0.8629-0.9361)	0.8958 (0.8551-0.9364)
<b>Hospitalisation</b>	ReSVinet-6	Feeding intolerance Medical intervention Respiratory difficulty Respiratory frequency Apnoea General condition	Fever	0.9484 (0.9257-0.9711)	0.9080 (0.8734-0.9427)	0.9069 (0.8684-0.9454)
<b>ICU Admission</b>	ReSVinet-3	Respiratory difficulty Apnoea General condition	Feeding intolerance Medical intervention Respiratory frequency Fever	0.9165 (0.8867-0.9464)	0.9156 (0.8836-0.9477)	0.9164 (0.8819-0.951)
<b>SIMV requirement (n=280)</b>	ReSVinet-4	Medical intervention Respiratory difficulty Apnoea General condition	Feeding intolerance Respiratory frequency Fever	0.9378 (0.9135-0.9621)	0.9150 (0.8818-0.9483)	0.9234 (0.8892-0.9576)

**Table 6:** P-value matrices score in development (RESCEU) dataset. (Cells where  $p < 0.005$  [i.e. cut-off adopted for statistically significant difference] are highlighted in green)  
(ICU - intensive care unit; SIMV - synchronised intermittent mandatory ventilation)

Hospitalisation (n=311)					ICU Admission (n=311)					SIMV requirement (n=280)				
	ReSVin et	ReSVinet-6	ReSVine t-3	ReSVine t-4		ReSVin et	ReSVine t-6	ReSVine t-3	ReSVine t-4		ReSVin et	ReSVine t-6	ReSVine t-3	ReSVine t-4
ReSVine t					ReSVine t					ReSVine t				
ReSVine t-6	1				ReSVine t-6	0.02776				ReSVine t-6	0.02169			
ReSVine t-3	0.012	0.0007127			ReSVine t-3	0.09679	0.3828			ReSVine t-3	0.05608	0.2942		
ReSVine t-4	0.4268	0.00092684 27	0.00259 1		ReSVine t-4	0.03306	0.2698	0.913		ReSVine t-4	0.00592 2	0.05091	0.1783	

### *External validation*

Subsequently we externally validated the three candidate simplified scores using the external validation datasets (Camacho-Cruz & Hakizimana) (see *Table 7*). In these datasets, they generally performed similarly to the original score. There were only two instances in which there were statistically significant differences (see *Table 8* for the p-value matrices). The first occurred in the Hakizimana dataset where both ReSVinet-3 & ReSVinet-4 performed extremely poorly at discriminating for HDU admission but both the original ReSVinet score, and ReSVinet-6 performed moderately well. The second occurred in the Camacho-Cruz dataset where all of the candidate scores outperformed the original score, and ReSVinet-4 outperformed both ReSVinet-6 & ReSVinet-3.

We additionally externally validated the discriminative validity of the original ReSVinet score in the Camacho-Cruz dataset (see *Table 7*); in this dataset it was moderate at discriminating for both hospitalization and PICU admission.

**Table 7: AUROC for original ReSVinet & candidate simplified scores in external validation datasets**

(AUROC- Area under the receiver operating characteristic curve; CI - confidence interval; HDU - high dependency unit; ICU - intensive care unit; PICU – Paediatric intensive care unit)

Score	AUROC (95% CI)													
	Hakizimana (n=100)										Camacho-Cruz (n=188)			
	Nurse					Resident					Faculty		Resident	
	Hospitalisation	HDU Admission	PICU Admission	Intubation or ventilation	Death	Hospitalisation	HDU admission	PICU Admission	Intubation or ventilation	Death	Hospitalisation	PICU Admission	Hospitalisation	PICU Admission
ReSVinet	0.9731 (0.9442-1)	0.7955 (0.6909-0.9)	0.9178 (0.8284-1)	0.9053 (0.7779-1)	0.9743 (0.9454-1)	0.9562 (0.9167-0.9957)	0.7838 (0.6767-0.8909)	0.8994 (0.7659-1)	0.8726 (0.6835-1)	0.9796 (0.9543-1)	0.8031 (0.7408-0.8654)	0.8219 (0.6541-0.9896)	0.8084 (0.7455-0.8714)	0.8213 (0.6563-0.9864)
ReSVinet-6	0.9747 (0.9471-1)	0.7777 (0.6742-0.8811)	0.9209 (0.8274-1)	0.9021 (0.7688-1)	0.9707 (0.9386-1)	0.9547 (0.9157-0.9937)	0.7751 (0.669-0.8812)	0.9278 (0.8353-1)	0.9032 (0.7718-1)	0.9796 (0.9514-1)	0.8843 (0.8376-0.9311)	0.8754 (0.7417-1)	0.8861 (0.8387-0.9335)	0.8716 (0.7361-1)
ReSVinet-3	0.937 (0.8808-0.9932)	0.2745 (0.1782-0.3708)	0.9224 (0.811-1)	0.8989 (0.7392-1)	0.9867 (0.9666-1)	0.9416 (0.8931-0.9901)	0.2678 (0.1712-0.3644)	0.9194 (0.8052-1)	0.8958 (0.7322-1)	0.9858 (0.9648-1)	0.8787 (0.8322-0.9253)	0.9213 (0.8553-0.9873)	0.8779 (0.8301-0.9256)	0.9082 (0.832-0.9843)
ReSVinet-4	0.9616 (0.912-1)	0.2666 (0.1714-0.3618)	0.9194 (0.81-1)	0.9 (0.7425-1)	0.9814 (0.9579-1)	0.9539 (0.9155-0.9924)	0.2637 (0.1675-0.3599)	0.9194 (0.8053-1)	0.8979 (0.734-1)	0.9832 (0.9609-1)	0.9255 (0.8879-0.9631)	0.9377 (0.8848-0.9906)	0.9318 (0.892-0.9716)	0.9311 (0.8731-0.9892)

**Table 8:** P-values matrices for external validation datasets. (Cells where  $p < 0.005$  [i.e. cut-off adopted for statistically significant difference] are highlighted in green)  
 (HDU - high dependency unit; PICU – Paediatric intensive care unit)

Hakizimana: Nurse, Hospitalisations (n=100)					Hakizimana: Nurse, HDU (n=100)					Hakizimana: Nurse, PICU (n=100)					Hakizimana: Nurse, Intubation or ventilation (n=100)					Hakizimana: Nurse, Death (n=100)					
	Re SVi net	ReS Vin et-6	ReS Vin et-3	ReS Vin et-4		Re SVi net	ReS Vin et-6	ReS Vin et-3	ReS Vin et-4		Re SVi net	ReS Vin et-6	ReS Vin et-3	ReS Vin et-4		Re SVi net	ReS Vin et-6	ReS Vin et-3	ReS Vin et-4		Re SVi net	ReS Vin et-6	ReS Vin et-3	ReS Vin et-4	
ReS Vin et					ReS Vin et					ReS Vin et					ReS Vin et					ReS Vin et					
ReS Vin et-6	0.8 192				ReS Vin et-6	0.2 716				ReS Vin et-6	0.6 844				ReS Vin et-6	0.7 002				ReS Vin et-6	0.5 909				
ReS Vin et-3	0.1 326	0.06 958			ReS Vin et-3	4.6 9e-08	3.71 5e-07			ReS Vin et-3	0.7 932	0.91 8			ReS Vin et-3	0.7 686	0.87 85			ReS Vin et-3	0.1 782	0.19 77			
ReS Vin et-4	0.5 983	0.48 18	0.12 88		ReS Vin et-4	2.6 61e-08	6.22 4e-08	0.06 422		ReS Vin et-4	0.9 225	0.90 49	0.41 64		ReS Vin et-4	0.7 979	0.91 45	0.69 54		ReS Vin et-4	0.3 408	0.28 05	0.27 07		
Hakizimana: Resident, Hospitalisations (n=100)					Hakizimana: Resident, HDU (n=100)					Hakizimana: Resident, PICU (n=100)					Hakizimana: Resident, Intubation or ventilation (n=100)					Hakizimana: Resident, Death (n=100)					
	Re SVi net	ReS Vin et-6	ReS Vin et-3	ReS Vin et-4		Re SVi net	ReS Vin et-6	ReS Vin et-3	ReS Vin et-4		Re SVi net	ReS Vin et-6	ReS Vin et-3	ReS Vin et-4		Re SVi net	ReS Vin et-6	ReS Vin et-3	ReS Vin et-4		Re SVi net	ReS Vin et-6	ReS Vin et-3	ReS Vin et-4	
ReS Vin et					ReS Vin et					ReS Vin et					ReS Vin et					ReS Vin et					
ReS Vin et-6	0.8 497				ReS Vin et-6	0.6 532				ReS Vin et-6	0.2 021				ReS Vin et-6	0.3 337				ReS Vin et-6	0.7 395				
ReS Vin et-3	0.3 827	0.37 02			ReS Vin et-3	7.3 56e-08	2.11 8e-07			ReS Vin et-3	0.1 109	0.57 29			ReS Vin et-3	0.1 774	0.75 1			ReS Vin et-3	0.3 826	0.39 48			
ReS Vin et-4	0.8 721	0.94 41	0.36 71		ReS Vin et-4	7.7 14e-08	1.50 6e-07	0.12 4		ReS Vin et-4	0.1 058	0.55 38	1		ReS Vin et-4	0.1 264	0.81 85	0.61 15		ReS Vin et-4	0.5 892	0.51 71	0.58 62		

Camacho-Cruz: Faculty, Hospitalisations (n=188)					Camacho-Cruz: Faculty, PICU (n=188)				
	Re SVi net	ReS Vin et-6	ReS Vin et-3	ReS Vin et-4		Re SVi net	ReS Vin et-6	ReS Vin et-3	ReS Vin et-4
ReS Vin et					ReS Vin et				
ReS Vin et-6	5.13e-07				ReS Vin et-6	0.1072			
ReS Vin et-3	0.001204	0.7172			ReS Vin et-3	0.09639	0.3396		
ReS Vin et-4	3.789e-07	0.001689	4.335e-05		ReS Vin et-4	0.0732	0.2223	0.05062	
Camacho-Cruz: Resident, Hospitalisations (n=188)					Camacho-Cruz: Resident, PICU (n=188)				
	Re SVi net	ReS Vin et-6	ReS Vin et-3	ReS Vin et-4		Re SVi net	ReS Vin et-6	ReS Vin et-3	ReS Vin et-4
ReS Vin et					ReS Vin et				
ReS Vin et-6	1.275e-06				ReS Vin et-6	0.122			
ReS Vin et-3	0.002475	0.5334			ReS Vin et-3	0.1068	0.4266		
ReS Vin et-4	6.03e-07	0.000636	8.355e-06		ReS Vin et-4	0.07007	0.2336	0.03829	

## 5. Discussion

Overall, the candidate simplified scores were moderate-excellent at discriminating, in both the development and external validation datasets, for a range of outcomes spanning the entire range of secondary-care disease severity. They performed equally well or better than the original ReSVinet score at discriminating by all of the outcomes evaluated; the only exception to this was in one dataset (Hakizimana) where two (ReSVinet-3 & ReSVinet-4) out of three of the candidate scores performed extremely poorly at differentiating by admission to the HDU. Additionally, the ReSVinet score was externally validated in two new datasets (Camacho-Cruz & Hakizimana) in which it performed moderate to excellent.

Significantly, both of the external validation datasets (but not the development dataset) were collected in non-European countries, and they were collected in a low-income and middle-income country; this is relatively unusual as vast majority of previous validation efforts have occurred in high-income countries [PART I of this report, 2].

In the three datasets used there were differences in when the ReSVinet score was taken. In the RESCEU dataset, the score was taken at the time of recruitment and so there is potentially a large degree of variability at the point in the disease course at which the scores were taken, whereas in the Hakizimana & Camacho-Cruz datasets the scores were taken within an hour of presentation and in the emergency room respectively. This greater variability reduces the usefulness of the validation in the RESCEU dataset.

There were also some differences in the inclusion criteria of the datasets, specifically the age ranges and, presence of RSV and comorbidities. It is not clear, on the basis of our analysis, if these differences affected the discriminative ability of the scores.

To systematically assess bias, we used a modified version of the PROBAST tool to conduct an internal risk of bias assessment (see *ANNEX II*) Based on this we deemed overall that there was a high risk that our estimates of the discriminative validity of these scores are biased.

The primary reason for this is due the small number of patients with positive outcomes owing to the small size of the datasets. Only in the RESCEU dataset for 1 outcome (hospitalisations) were the PROBAST thresholds ( $\geq 20$  events per variable for validation [RESCEU];  $\geq 100$  for validation [Hakizimana & Camacho-Cruz]) met. In 5 instances, the number of participants with the outcome of interests is a single-digit figure. This is similar to what is observed in published literature, presumably due to difficulties in collecting large datasets of these sorts. For illustration purposes, to be confident in reporting on the discriminative validity for death (assuming the same approximate ratio of deaths), the Hakizimana dataset would need to be approximately 17-times larger. Additionally, ideally, we would have used multiple imputation for missing data instead of excluding patients with missing score-items and additionally assessed calibration; this was not done due to time constraints. Regardless, neither of these factors would have changed our overall classification of this study as having an associated high risk of bias.

An additional limitation, not captured in the PROBAST checklist above, was that this was a secondary evaluation of (prospectively collected) datasets.

To guide, as well as assess the transparency of reporting, the TRIPOD checklist was employed (see *ANNEX I*). The overall score was 26/32 (81%). This is considerably better than the average score, 53%, found in our systematic review [PART I of this report], although there still is room for

improvement. The issues with reporting were that not all required information was reported in the title (due to length constraints), the number of positive outcomes were not reported in the abstract (due to word limit constraints), the number of centres enrolment took place at for the RESCEU study was not reported, that interaction terms were not tested for and, that calibration was not reported (as not assessed).

#### *Implications on research, policy & practice*

The findings indicate that fever as a criterion is redundant and may be excluded from the original ReSVinet score (i.e. ReSVinet-6 should be adopted). However, given the small number of other validation studies (see *Table 1*) and more generally studies evaluating the ReSVinet score (e.g. reliability, responsiveness, utility), we cannot at this time recommend use of either the original ReSVinet score or ReSVinet-6 in routine clinical care or as a primary outcome for clinical trials.

Further research on much-larger datasets is required to validate the usefulness of the original ReSVinet score, as well as the simplified scores we have developed. Given the relatively small number of patients with positive outcomes in the development dataset, it would also be appropriate to repeat this exercise, using similar methods, in larger datasets to examine if different combinations of parameters (i.e. scores) are selected. Ideally the datasets employed for these exercise as well as being larger, would be gathered in resource-limited settings where the burden is disproportionately high [2,3].

## 6. References

1. Justicia-Grande AJ, Pardo-Seco J, Cebey-López M, et al. Development and Validation of a New Clinical Scale for Infants with Acute Respiratory Infection: The ReSVinet Scale. *PloS One* [Internet]. **2016** [cited 2022 Mar 1]; 11(6):e0157665. Available from: <https://pubmed.ncbi.nlm.nih.gov/27327497/>
2. Hakizimana B, Saint G, Miert C van, Cartledge P. Can a Respiratory Severity Score Accurately Assess Respiratory Distress in Children with Bronchiolitis in a Resource-Limited Setting? *Journal of Tropical Pediatrics* [Internet]. **2020**; 66(2):234–243. Available from: <https://academic.oup.com/tropej/article/66/2/234/5570307#:~:text=Though%20scores%20should%20measure%20what>
3. Justicia-Grande AJ, Martínón-Torres F. The ReSVinet Score for Bronchiolitis: A Scale for All Seasons. *American Journal of Perinatology* [Internet]. **2019** [cited 2022 Mar 13]; 36(S 02):S48–S53. Available from: <https://www.thieme-connect.com/products/ejournals/pdf/10.1055/s-0039-1691800.pdf>
4. Hakizimana B, Kalimba E, Ndatinya A, Saint G, Miert C van, Cartledge PT. Field testing two existing, standardized respiratory severity scores (LIBSS and ReSVINET) in infants presenting with acute respiratory illness to tertiary hospitals in Rwanda – a validation and inter-rater reliability study. *PLoS ONE* [Internet]. **2021** [cited 2022 Mar 4]; 16(11):e0258882. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8568200/>
5. Justicia-Grande, A.J., Pardo Seco, J., Alvarez Garnelo, M.E., Pena Nieto, J., Sánchez Lastres, J.M., Del Río Garma, M., Valls Duran, M.T., San José González, M.A., Regueiro Martínez, J., Alonso Alonso, C., Fernández Bustillo, J.M., García Sánchez, N., Castro Paz, L., López-Silvarrey Varela, Á., López Del Olmo, M.C., Rivero Calle, I., Obando Pacheco, P., Porto Silva, M., Grela Beiroa, A. and Serén, S. Validation of a Clinical Respiratory Score (the ReSVinet Scale) in Pediatric Patients Suffering from Respiratory Infection in a Primary Care Setting [Internet]. **2017**. Available from: <https://espido2017.kenes.com/Documents/ESPID17%20abstracts.pdf>
6. Tungsupreechameth A, Srisingh K. Factors Associated with Severe Lower Respiratory Tract Infection from Respiratory Syncytial Virus (RSV) in Thai Children. *Siriraj Medical Journal*. **2021**; 73(12):808–814.
7. Vandendijck Y, Scott J, Heerwegh D, et al. Measuring respiratory syncytial virus infection severity in hospitalized children using the Pediatric Respiratory Syncytial Virus Electronic Severity and Outcome Rating System (PRESORS). *Influenza and Other Respiratory Viruses*. **2022**;
8. Jefferies K, Drysdale SB, Robinson H, et al. Presumed Risk Factors and Biomarkers for Severe Respiratory Syncytial Virus Disease and Related Sequelae: Protocol for an Observational Multicenter, Case-Control Study From the Respiratory Syncytial Virus Consortium in Europe (RESCEU). *The Journal of Infectious Diseases*. **2020**; 222(Supplement\_7):S658–S665.
9. McGinley JP, Lin GL, Öner D, et al. Clinical and Viral Factors Associated With Disease Severity and Subsequent Wheezing in Infants With Respiratory Syncytial Virus Infection. *The Journal of Infectious Diseases*. **2022**; 226(Supplement\_1):S45–S54.

10. Camacho-Cruz J, Briñez S, Alvarez J, Leal V, Gómez LV, Vasquez-Hoyos P. Use of the ReSVinet Scale for parents and healthcare workers in a paediatric emergency service: a prospective study. *BMJ Paediatrics Open* [Internet]. **2021** [cited 2022 Apr 22]; 5(1):e000966. Available from: <https://bmjpaedsopen.bmj.com/content/5/1/e000966>
11. Cartledge P. LIBBS and ReSVinet Rwanda - Data (V9 for DOI).csv, Field testing two existing standardised respiratory severity scores (LIBSS and ReSVinet) in infants presenting with acute respiratory illness to tertiary hospitals in Rwanda [Internet]. Harvard Dataverse. **2020** [cited 2022 Aug 26]. Available from: <https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/N4O05G/AKSNOZ>
12. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Annals of Internal Medicine*. **2019**; 170(1):W1.
13. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Annals of Internal Medicine*. **2015**; 162(1):W1.
14. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*. **1988**; 44(3):837.

## ANNEXES

## ANNEX I. Detailed TRIPOD score

TRIPOD Explanation & advice: Moons et al. [13]

**Table:** TRIPOD for data project.

(D – Development; V – Validation).

A green cell indicates adherence, a red cell lack of adherence and a grey cell lack of applicability.

Section /Topic		Validation /Development	Checklist Item	
<b>Title and abstract</b>				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	
<b>Introduction</b>				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	
<b>Methods</b>				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	
	5b	D;V	Describe eligibility criteria for participants.	
	5c	D;V	Give details of treatments received, if relevant.	
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	
Sample size	8	D;V	Explain how the study size was arrived at.	

Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	
	10c	V	For validation, describe how the predictions were calculated.	
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	
<b>Results</b>				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	
	15b	D	Explain how to use the prediction model.	
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	
<b>Discussion</b>				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	

Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	
<b>Other information</b>				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	

\*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

## ANNEX II. Detailed PROBAST checklist

To systematically assess the risk of bias, a modified version of the PROBAST checklist was employed. Given that this was not as part of a scoping review, the questions regarded applicability were ignored. For the ease of the readers, the filled in PROBAST checklist is presented separately for each dataset.

PROBAST Explanation & advice: Moons et al., 2019 [12]

**Table:** Modified PROBAST for data project.

(N – No; NA – Not applicable; PN – Probably No; PY – Probably Yes; Y – Yes)

Grey shaded box indicates inapplicability according to PROBAST recommendations.

	<b>RESCEU</b>	<b>Hakizimana</b>	<b>Camacho-Cruz</b>
Type of prediction study	Development	Validation	
Models of interest	ReSVinet ReSVinet-6 ReSVinet-3 ReSVinet-4		
Outcome of interest	Hospitalisation ICU Admission SIMV requirement	Hospitalisation HDU Admission PICU Admission Intubation or ventilation Death	Hospitalisation PICU Admission
<b>DOMAIN 1: Participants</b>			
1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	Y	Y	Y
1.2 Were all inclusions and exclusions of participants appropriate?	Y	Y	Y
<b>Risk of bias introduced by selection of participants (low/ high/ unclear)</b>	Low	Low	Low
<i>Rationale:</i>			
<b>DOMAIN 2: Predictors</b>			
2.1 Were predictors defined and assessed in a similar way for all participants?	Y	Y	Y
2.2 Were predictor assessments made	PN	PN	PN

without knowledge of outcome data?			
2.3 Are all predictors available at the time the model is intended to be used?	Y	Y	Y
<b>Risk of bias introduced by predictors or their assessment (low/ high/ unclear)</b>	Low	Low	Low
<i>Rationale:</i>			
<b>DOMAIN 3: Outcome</b>			
3.1 Was the outcome determined appropriately?	Y	Y	Y
3.2 Was a pre-specified or standard outcome definition used?	Y	Y	Y
3.3 Were predictors excluded from the outcome definition?	Y	Y	Y
3.4 Was the outcome defined and determined in a similar way for all participants?	Y	Y	Y
3.5 Was the outcome determined without knowledge of predictor information?	PN	PN	PN
3.6 Was the time interval between predictor assessment and outcome determination appropriate?	Y	Y	Y
<b>Risk of bias introduced by the outcome or its determination (low/ high/ unclear)</b>	Low	Low	Low
<i>Rationale:</i>			
<b>DOMAIN 4: Analysis</b>			
4.1 Were there a reasonable number of participants with the outcome?	Y for hospitalisations. N for ICU Admission or SIMV requirement.	N for all.	N for all.
4.2 Were continuous and categorical predictors handled appropriately?	Y	Y	Y

4.3 Were all enrolled participants included in the analysis?	N	N	Y
4.4 Were participants with missing data handled appropriately?	N	N	Y
4.5 Was selection of predictors based on univariable analysis avoided?	NA		
4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls) accounted for appropriately?	Y	Y	Y
4.7 Were relevant model performance measures evaluated appropriately?	N	N	N
4.8 Were model overfitting and optimism in model performance accounted for?	Y		
4.9 Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	NA		
<b>Risk of bias introduced by the analysis (low/ high/ unclear)</b>	High	High	High
<i>Rationale</i>	Patients were excluded with missing predictors (4.3) and excluded from analyses with missing outcome data (4.4). Calibration was not assessed (4.7)	Patients with incomplete data were excluded (4.3, 4.4). Calibration was not assessed (4.7)	Not all enrolled patients were included, but the reason for exclusion was reasonable (4.3). Calibration was not assessed (4.7)
<b>Overall judgement of risk of bias</b>			
<b>Overall judgement of risk of bias (low/ high/ unclear)</b>	<b>High</b>	<b>High</b>	<b>High</b>
<i>Summary of sources of potential bias:</i>	Small number of participants with	Small number of participants	Small number of participants with positive outcomes.

	<p>positive outcome data for ICU admission and SIMV requirement. Exclusion of participants with missing data. Calibration not assessed.</p>	<p>with positive outcomes. Exclusion of participants with missing data. Calibration not assessed.</p>	<p>Calibration not assessed.</p>
--	---	---	----------------------------------

## ANNEX III. Missing data for baseline characteristics

**Table:** Missing data for baseline characteristics by dataset (Greyed out cell indicates data not available in dataset or inapplicability)

Characteristics	Development dataset	External Validation datasets	
	RESCEU	Hakizimana	Camacho-Cruz
Sex	0	0	0
Country of site	0	0	0
Birth Body Weight	8		
Current Body Weight	7	0	
RSV Subgroup	24		
Season	0		
Gestational age	2		
Premature			
Age at enrolment	5	0	0
Diagnosis	2	0	0
Viral tests	0		?
Pre-existing condition	0		
Prematurity	0	0	
Malnutrition		0	
Asthmatic		0	
Human immunodeficiency virus		0	
Ubedehe Social Group		0	
Residence			
Living sibling		0	
Vaccinations complete		0	
Maternal marital status		0	
Maternal age		0	
Paternal age		0	
Maternal occupation		0	
Paternal occupation		0	
Maternal educational level		0	
Paternal educational level		1	