Validity of Bronchiolitis Outcome Measures

Ricardo M. Fernandes, MD, Amy C. Plint, MD, MSc, Caroline B. Terwee, PhD, Cristina Sampaio, MD, PhD, Terry P. Klassen, MD, PhD, Martin Offringa, MD, PhD, Johanna H. van der Lee, MD, PhD

BACKGROUND: The Respiratory Distress Assessment Instrument (RDAI) and Respiratory Assessment Change Score (RACS) are frequently used in bronchiolitis clinical trials, but evidence is limited on their measurement properties. We investigated their validity, reliability, and responsiveness.

METHODS: We included data from up to 1765 infants with bronchiolitis enrolled in 2 studies conducted in pediatric emergency departments. We assessed RDAI construct validity by testing hypotheses of associations with physiologic measures (respiratory rate, oxygen saturation) and with constructs related to hospitalization, using correlation coefficients, and multivariable analysis. RDAI/RACS responsiveness was evaluated by using anchors of change based on these constructs; measures of responsiveness included the area under the curve. RDAI test-retest agreement and interrater reliability were evaluated by using limits of agreement and intraclass correlation coefficients.

RESULTS: Baseline RDAI scores were weakly correlated with respiratory rate ($r = 0.38, P < .001$), and scores increased in lower oxygen saturation categories ($P < .001$). Higher RDAI scores were associated with hospitalization (odds ratio: 1.36; 95% confidence interval: 1.26–1.47); scores differed between participants who were discharged, admitted, or stayed in the emergency department ($P < .001$). Our hypotheses were met, but the magnitude of associations was below our predefined thresholds. RDAI test-retest limits of agreement were $-3.80$ to $3.64$ (20% of the range), whereas interrater reliability was good (intraclass correlation coefficient = 0.93). Formulated hypotheses for responsiveness were confirmed, with moderate responsiveness (area under the curve: RDAI, 0.64–0.70; RACS, 0.72).

CONCLUSIONS: RDAI has poor to moderate construct validity, with good discriminative properties but considerable test-retest measurement error. The RDAI and RACS are responsive measures of respiratory distress in bronchiolitis but do not encompass all determinants of disease severity.

WHAT’S KNOWN ON THIS SUBJECT: The Respiratory Distress Assessment Instrument (RDAI) and the Respiratory Assessment Change Score (RACS) are the most frequently used measurement instruments in bronchiolitis clinical trials. Evidence is scarce regarding their measurement properties and their suitability for use as evaluative instruments in clinical trials.

WHAT THIS STUDY ADDS: The RDAI is an incomplete measure of respiratory distress in bronchiolitis, with poor to moderate construct validity. It has adequate discriminative properties but considerable test-retest measurement error. The RDAI and RACS were moderately responsive, but methodologic issues limit the interpretation of this finding.
Acute viral bronchiolitis is the most common lower respiratory tract infection in infants and carries substantial clinical and financial burden.\textsuperscript{1,2} There is wide practice variation in its management, with heterogeneous evidence for many therapeutic approaches.\textsuperscript{3–7} Systematic reviews have highlighted various shortcomings in randomized controlled trials (RCTs) in this field.\textsuperscript{8–10} One of the major issues is the heterogeneous choice of outcome measures. There has been inconsistency in selected measurement instruments, whose measurement properties have often not been adequately studied.\textsuperscript{8,11}

Respiratory status is an important dimension and determinant of severity in bronchiolitis. The Respiratory Distress Assessment Instrument (RDAI) and Respiratory Assessment Change Score (RACS) are often used to measure this domain in bronchiolitis.\textsuperscript{8} Lowell et al\textsuperscript{12} first described them in an RCT of epinephrine in wheezing infants. The RDAI includes items on retractions and wheezing, whereas the RACS is a change score based on the RDAI and respiratory rate. Evidence is limited regarding RDAI and RACS measurement properties and their suitability for use as evaluative instruments in clinical trials.\textsuperscript{13–15} Previous RCTs have reported some data on which reliability and validity can be assessed, whereas the first formal validation study is recent.\textsuperscript{16}

The aim of this study was to assess and compare the measurement properties of RDAI and RACS, that is, validity, reliability, and responsiveness.

**METHODS**

**Population**

We used data from 2 related studies conducted simultaneously in 8 Canadian pediatric emergency departments (EDs) during 3 bronchiolitis seasons (2004–2007): a 2 × 2 factorial RCT (Canadian Epinephrine/Steroid Trial, CanBEST; \(N = 800\))\textsuperscript{17} and a prospective cohort study (\(N = 1554\) infants, 584 of whom also participated in CanBEST).\textsuperscript{18} Both studies included infants aged <12 months with acute bronchiolitis (first episode of wheezing) and excluded those with previous asthma, wheezing, or use of bronchodilators. Additional exclusion criteria in CanBEST were as follows: prematurity with corrected age <6 weeks, chronic cardiopulmonary disease or immunodeficiency, recent corticosteroid use or exposure to varicella, very mild or severe distress (pulse rate >200 beats per minute, respiratory rate >80 breaths per minute, or RDAI score <4 or >15), or lethargy.

Participants in CanBEST were randomly assigned to receive oral dexamethasone or placebo and nebulized epinephrine or placebo, both administered in the ED (Fig 1). During the first 90 minutes, only supplemental oxygen or acetaminophen was allowed. Other participants in the cohort study were given standard treatment as decided by the attending physicians. In both studies, written informed consent was obtained from the parents or guardians of the infants, and both were approved by ethics committees at each site and by Health Canada.

**Instruments and Outcome Measures**

We assessed the RDAI as described by Lowell et al\textsuperscript{12} and a modification of the RACS as reported by Schuh et al\textsuperscript{19} (Tables 1 and 2). The following measurements were performed at baseline for both studies and every 30 minutes until admission/discharge or 240 minutes for CanBEST: RDAI, respiratory rate, heart rate, oxygen saturation (\(\text{SaO}_2\)), and activity status (Fig 1). Fever was also assessed at baseline. Research nurses performed all measurements after formal training and using written instructions. \(\text{SaO}_2\) was measured by using pulse oximeters available locally. In both studies, the attending physician independently determined whether to admit or discharge the infant; RDAI was not used clinically at any site. In CanBEST, physicians and nurses were blinded to treatment interventions, and by protocol any decisions regarding admission, discharge, or continued stay in the ED were to be made only after the study interventions (ie, after 90 minutes).

![FIGURE 1](https://example.com/figure1.png)

**FIGURE 1**

Timing of intervention, measurements, and clinical decisions in the CanBEST trial. Dex, dexamethasone; Epi, epinephrine; Pla, placebo; RR, respiratory rate.
TABLE 1 Respiratory Distress Assessment Instrument (RDAI)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Score 0</th>
<th>Score 1</th>
<th>Score 2</th>
<th>Score 3</th>
<th>Score 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheezing (auscultation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expiration</td>
<td>None</td>
<td>End</td>
<td>1/2</td>
<td>3/4</td>
<td>All</td>
</tr>
<tr>
<td>Inspiration</td>
<td>None</td>
<td>Part</td>
<td>All</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>None</td>
<td>Segmental: ≤2 of 4 lung fields</td>
<td>Diffuse: ≥3 of 4 lung fields</td>
<td>0–8</td>
<td></td>
</tr>
<tr>
<td>Partial sum score</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retractions (visual assessment)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supraclavicular</td>
<td>None</td>
<td>Mild</td>
<td>Moderate</td>
<td>Marked</td>
<td>0–3</td>
</tr>
<tr>
<td>Intercostal</td>
<td>None</td>
<td>Mild</td>
<td>Moderate</td>
<td>Marked</td>
<td>0–3</td>
</tr>
<tr>
<td>Subcostal</td>
<td>None</td>
<td>Mild</td>
<td>Moderate</td>
<td>Marked</td>
<td>0–3</td>
</tr>
<tr>
<td>Partial sum score</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum score (higher scores indicate more severe disease)</td>
<td>0–17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The original RDAI as reported by Lowell et al. also included respiratory rate, ie, it did not differentiate between RDAI and RACS and only used RACS as an outcome measure.

Statistical Analysis

We used the Consensus-based Standards for the selection of health Measurement Instruments initiative (COSMIN) definitions of measurement domains and properties regarding validity, reliability, and responsiveness.20

Construct Validity of the RDAI

There is no "gold standard" to assess bronchiolitis severity or respiratory distress. We assessed construct validity of the RDAI by formulating hypotheses about the direction and magnitude of the association between RDAI scores and other bronchiolitis severity or respiratory distress measures (respiratory rate, SaO2) and clinical decision-making constructs (decision to admit/discharge and time to admission/discharge).21 We studied both convergent and discriminative validity.

We hypothesized that baseline RDAI scores and respiratory rate would have a strong positive correlation (Pearson's r = 0.7). We used multiple linear regression analysis to explore possible confounding of this association, by activity status and fever (data from both studies) and age and weight (data from CanBEST). We further hypothesized a negative association between RDAI and SaO2, which we expected to be weaker and nonlinear (Spearman's r ≤ −0.5), and we compared RDAI scores between 3 categories of SaO2 (<92%, 92%–95%, >95%; data from both studies).

We hypothesized that a higher RDAI score would increase the risk of admission (expected odds ratio [OR] for admission: >1.5 for RDAI scores above the median). For this analysis, we used the last RDAI score assessed or registered before the time of admission/discharge (data from CanBEST). Multiple logistic regression analysis was used to evaluate whether that association was confounded by center, treatment group, age, and SaO2. Furthermore, we expected participants who stayed in the ED longer to have intermediate scores compared with those who were admitted or discharged sooner (data from CanBEST after the main trial interventions).

Reliability of the RDAI

For RDAI test-retest reliability, we considered that the group of CanBEST participants who had received both placebo interventions was stable between the 90- and 120-minute measurements. The same research nurse assessed the same child unblinded to the previous assessment. In a convenience sample of participants from each study, 2 nurses performed baseline RDAI measurements independently to obtain interrater assessments.

For both test-retest and interrater conditions, we distinguished measures of measurement error from reliability measures.22 To evaluate measurement error we calculated the standard error of measurement (SEM) and, the smallest detectable change (SDC), and we obtained a Bland-Altman plot and the 95% limits of agreement (LoA; formulas are shown in Supplemental Information). The Bland-Altman plot shows the mean differences between the test and retest scores (expressed as modified by Schuh et al. In the original RACS as reported by Lowell et al, final scores were subtracted from baseline scores (ie, positive change scores indicated improvement), and cutoffs to define respiratory rate change were defined at 10% intervals.

TABLE 2 Respiratory Assessment Change Score (RACS)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Formula</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheezing change score</td>
<td>Final partial sum score - baseline partial sum score</td>
<td>−8 to +8</td>
</tr>
<tr>
<td>Retractions change score</td>
<td>Final partial sum score - baseline partial sum score</td>
<td>−9 to +9</td>
</tr>
<tr>
<td>Respiratory rate “standardized”</td>
<td>5% change: 0 units</td>
<td>−n to +n</td>
</tr>
<tr>
<td>change score</td>
<td>6% to 15% change: −1/+1 units</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16% to 25% change: −2/+2 units</td>
<td></td>
</tr>
<tr>
<td></td>
<td>etc</td>
<td></td>
</tr>
<tr>
<td>Sum score (negative change scores indicate improvement)</td>
<td>−17 + n to +17 + n</td>
<td></td>
</tr>
</tbody>
</table>

As modified by Schuh et al. In the original RACS as reported by Lowell et al, final scores were subtracted from baseline scores (ie, positive change scores indicated improvement), and cutoffs to define respiratory rate change were defined at 10% intervals.
The LoA show the differences lie. We assessed reliability by calculating the intraclass correlation coefficient (ICC) in a 2-way random-effects model, including patient, observer, and residual variance components (formula shown in Supplemental Information).

**Responsiveness of the RDAI/RACS**

As with validity, we studied the responsiveness of the RDAI and RACS through testing hypotheses concerning the expected associations of change scores.21 There is no clear criterion for change in bronchiolitis and none of the studies included explicit assessments of change. We based our hypotheses on physiologic and clinical constructs of change using a priori–defined criteria to identify groups of participants who improved versus those who had remained stable or deteriorated, irrespective of interventions. Criteria were based on respiratory rate and SaO2, and we used combined group data from CanBEST (Table 3).

We used different measures of responsiveness assessing statistical change or clinically important change, focusing on comparing the improved group with the stable/deteriorated group.23 These included the following: testing differences in RDAI change scores and/or RACS within and between groups, and calculating standardized/Cohen’s effect size (ES) and the responsiveness ratio (formulas shown in Supplemental Information). We hypothesized that patients who had improved would have larger change scores and ESs than patients who had not improved. We also used the area under the curve (AUC) of the receiver operating characteristic curve of improved versus stable groups (AUC >0.70 considered appropriate).

For all analyses, we excluded participants with nonvalid or missing data, with no imputation. Statistical significance was set at \( P < .05 \), and we calculated 95% confidence intervals (CIs) when applicable. We used SPSS version 19 (IBM SPSS Statistics, IBM Corporation, Armonk, NY).

**RESULTS**

Figure 2 shows data sources and participants included in the analysis of each measurement property. The baseline characteristics of participants and selected outcomes from both studies are presented in Table 4. Participants in CanBEST were older than those in the cohort study, whereas baseline severity was greater in the latter study.

**Construct Validity of the RDAI**

We found a weak positive correlation between RDAI score and respiratory rate at baseline with data from both studies (Pearson’s \( r = 0.38; 95\% \) CI: 0.35 to 0.45; \( P < .001; N = 1765 \)). Correlations for retractions and wheezing subscores were \( r = 0.41 \) and \( r = 0.17 \), respectively. By using simple linear regression, the coefficient estimate was a 1.55 (95\% CI: 1.38 to 1.73) increase in respiratory rate (breaths per minute) per increase in RDAI unit (\( P < .001 \)). The estimate was comparable when adjusting for fever and activity status (adjusted estimate: 1.52). When restricting the analysis to CanBEST data, the correlation was weaker (Pearson’s \( r = 0.22 \); unadjusted linear regression estimate: 0.98; \( n = 800 \)). The association was not confounded by age, weight, fever, or activity status (adjusted estimate: 0.92).

There was a weak negative correlation between baseline RDAI scores and SaO2 levels (Spearman’s \( r = -0.24; P < .001; n = 1761 \)). Correlations for retractions and wheezing were \( r = -0.25 \) and \( r = -0.14 \), respectively. RDAI scores increased in lower SaO2 categories (Fig 3). The median (interquartile range) RDAI scores were 10 (8–12), 8 (6–10), and 7 (5–10) for SaO2 <92%, 92%–95%, and >95%, respectively (Kruskal-Wallis test, \( P < .001 \)).

We found an association between the decision to admit or discharge and the last RDAI score of CanBEST.

**TABLE 3 Constructs of Change and Criteria Used to Assess Responsiveness of RDAI and RACS**

<table>
<thead>
<tr>
<th>Anchor of Change</th>
<th>Population</th>
<th>Timing of Measurements</th>
<th>Improved Group</th>
<th>Stable/Deteriorated Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in respiratory rate (relative change)(^a)</td>
<td>All participants</td>
<td>Inclusion</td>
<td>n = 796</td>
<td>At baseline and last measurement before admission or discharge</td>
</tr>
<tr>
<td>Change in respiratory rate (tachypnea)(^a)</td>
<td>Participants with tachypnea at baseline (respiratory rate ( &gt;50 ) breaths/minute (&lt;6 \text{ mo}) \text{ or } &gt;40 \text{ breaths/minute} \text{ (6–12 mo)} )</td>
<td>Population</td>
<td>n = 305</td>
<td>At baseline and last measurement before admission or discharge</td>
</tr>
<tr>
<td>Probability of admission at baseline (versus actual decision)</td>
<td>Participants with high baseline probability of admission (respiratory rate ( &gt;60 ) breaths/minute or SaO2 &lt;90%)</td>
<td>Population</td>
<td>n = 209</td>
<td>At baseline and last measurement before admission or discharge</td>
</tr>
</tbody>
</table>

\(^a\) Only for RDAI, not used to measure RACS responsiveness because respiratory rate is included in the RACS formula.
participants. The preceding RDAI score was higher in admitted patients than in those who were discharged (mean difference: 2.28; 95% CI: 1.75 to 2.81; \( t \) test, \( P < .001; n = 798 \)). A higher RDAI score was associated with higher risk of admission (OR: 1.36; 95% CI: 1.26 to 1.47) per increase in RDAI unit; and 2.54 (95% CI: 1.65 to 3.92) when the RDAI was > 8. Adjusted analyses for center, treatment group, age, and \( \text{SaO}_2 \) revealed no relevant changes in these associations.

In addition, we found that RDAI scores measured after CanBEST interventions differed between the groups of participants who were discharged (median [interquartile range]: 5 [2–6]), hospitalized (8 [5–10]), or who stayed in the ED (6 [4–8]) (\( n = 695 \); Kruskal-Wallis test, \( P < .001 \) (Fig 4). Differences between discharged participants and the 2 latter groups were statistically significant (\( P = .01 \) and \( P < .01 \), respectively; Bonferroni post hoc test). Patients with a higher RDAI score at 90 minutes had higher risk of ED stay > 240 minutes (OR: 1.33; 95% CI: 1.24 to 1.43 per increase in RDAI unit).

Overall, although results were in accordance with our validity hypotheses, the magnitude of the associations was mostly below our predefined thresholds.

Reliability of the RDAI

Test-retest assessments were available from 79 CanBEST participants. The mean difference between the 2 repeated assessments was 0.08 (95% CI: –0.35 to 0.5; \( P = .72 \)). The SEM was 1.34 and the SDC was 3.72 RDAI units. The test-retest 95% LoA were 2.3 to 3.8 RDAI units (95% CI: –2.8 to –1.3, to 2.91 to 4.37). This finding means that if a child is assessed twice, the second score could be between 2.3 points lower and 2.8 points higher than the first score, just because of measurement error. The magnitudes of differences between repeated measurements remained the same over the whole range of mean values as shown in the Bland-Altman plot (Fig 5). The test-retest ICC was 0.80 (95% CI: 0.70 to 0.87).

Inter-rater assessments were performed in 107 participants. There was no significant difference between 2 repeated assessments (mean difference: –0.06; 95% CI: –0.28 to 0.15; \( P = .54 \)). The SEM was 0.78, and the inter-rater LoA were –2.1 to 2.22 RDAI units (95% CI: –2.46 to –1.74, to 1.86 to 2.58). The ICC was 0.93 (95% CI: 0.9 to 0.94).

Responsiveness of the RDAI and RACS

Measures of responsiveness for RDAI and RACS based on the different constructs of change are presented in Table 5. By using both anchors, the mean RDAI scores decreased in both improved and stable groups (paired \( t \) test, \( P < .001 \); for all within-group comparisons), with larger mean changes in scores of the improved group (unpaired \( t \) test, \( P < .001 \); for all between-group comparisons).

These results were in accordance with our predefined hypotheses. Between-group differences in mean RDAI change scores ranged from –1.31 (95% CI: –1.85 to –0.77) for the 25% respiratory rate reduction criterion to –2.03 (95% CI: –2.9 to –1.16) for the probability of admission criterion. Standardized ESs for the improved group ranged from 1.43 to 1.71, whereas responsiveness ratios ranged from 1.54 to 1.61, and AUCs from 0.64 to 0.7. The RACS was larger in the improved group (between-group difference: –2.81; 95% CI: –3.92 to –1.7), with a responsiveness ratio of 1.96 and an AUC of 0.72.

DISCUSSION

This study of measurement properties of RDAI and RACS in acute bronchiolitis identifies strengths and...
limitations of their use as outcome measures. The RDAI was evaluated in 3 systematic reviews of measurement properties of asthma or wheezing severity scales in children.\textsuperscript{13–15} Limited data on its reliability and responsiveness were provided in the original description of the scale and in later reports of RCTs.\textsuperscript{12,19,24} However, none of these were adequately designed measurement studies, and no formal assessment of validity was found. Destino et al\textsuperscript{16} recently reported the first validation study on RDAI in bronchiolitis, showing poor construct validity, interrater reliability, and responsiveness. Findings on validity were fairly consistent with our results; differences in setting, raters, and methods may explain why results on reliability and responsiveness were distinct.

Our results show that the RDAI has poor to moderate construct validity. The RDAI was developed ad hoc with no elaboration on the underlying conceptual model, item selection, scoring, or weighting. Although in the original report only the RACS was used as an outcome measure, later trials used RDAI scores separately for single or repeated assessments.\textsuperscript{12} In our conceptual framework, respiratory distress was putatively reflected by RDAI items (ie, reflective model) and contributed to the multidimensional construct of bronchiolitis.\textsuperscript{21} We found poor convergent validity with respiratory rate, but RDAI scores discriminated well between clinically meaningful \textit{SaO}_2 subgroups. Measurement properties from other respiratory scales or their individual items, which often include respiratory rate or \textit{SaO}_2, are seldom available.\textsuperscript{13–15} When they are reported, there is substantial heterogeneity in correlations with \textit{SaO}_2, ranging from poor to moderate. Thus, our predefined cutoffs may have been too strict. Most, but not all, studies are consistent with our findings of weaker correlations.

\begin{table}
\centering
\caption{Baseline Characteristics of Participants and Selected Outcomes From the CanBEST Trial and the Cohort Study}  
\begin{tabular}{llll}
\textbf{Baseline Characteristics} & \textbf{CanBEST Trial (N = 800)} & \textbf{Cohort Study (N = 1554)}
\hline
Demographic characteristics & & \\
Age, median (IQR), mo & 5 (3–7) & 4 (2–7) \\
Male gender, n (%) & 483 (62) & 948 (61) \\
White, n (%) & 654 (82) & 1243 (80) \\
History & & \\
Personal history of atopy, n (%) & 88 (11) & 157 (10) \\
Prematurity, n (%) & 83 (10) & 202 (13) \\
Household smoking, n (%) & 305 (38) & 575 (37) \\
Symptom length, median (IQR), d & 4 (2–5) & 4 (2–5) \\
Clinical characteristics & & \\
Respiratory rate & & \\
Breaths/minute, median (IQR) & 48 (42–58) & 48 (42–60) \\
\geq60 Breaths/minute, n (%) & 196 (25) & 441 (28) \\
Oxygen saturation & & \\
\%, median (IQR) & 97 (95–98) & 97 (95–98) \\
\textless 90%, n (%) & 24 (3) & 121 (8) \\
Heart rate & & \\
Beats/minute, median (IQR) & 150 (139–160) & 152 (140–164) \\
\textgreater 180 Beats/minute, n (%) & 33 (4) & 143 (9) \\
RDAI score & & \\
Median (IQR) & 8 (6–10) & 8 (6–10) \\
\textgreater 12, n (%) & 76 (10) & 211 (14) \\
Patient outcomes & & \\
Admission during enrollment visit, n (%) & 119 (15) & 370 (24) \\
Time to admission/discharge, n (%) & & \\
\textless 90 minutes & 103 (13) & NA \\
90–120 minutes & 261 (33) & \\
120–240 minutes & 248 (31) & \\
\textgreater 240 minutes & 188 (23) & \\
\hline
\end{tabular}
\end{table}

\begin{figure}
\centering
\includegraphics{figure3}
\caption{Box plot displaying baseline RDAI scores by categories of \textit{SaO}_2. The box spans the interquartile range (IQR), the solid horizontal line through the box is the median value, and the whiskers denote values within 1.5 IQRs lower than the first quartile and 1.5 IQRs higher than the third quartile.}
\end{figure}
between \( \text{SaO}_2 \) and auscultatory items when compared with work of breathing.\(^{13-15,25}\) These results reflect the pathophysiology and clinical correlates of respiratory distress in bronchiolitis. It is known that as disease progresses and severity increases, so do the disturbances in ventilation and ventilation-perfusion matching.\(^{26,27}\) Many patients have effective compensatory mechanisms for these disturbances, although others do not. However, clinical signs of respiratory distress may not capture hypoxemia/hypercapnia balance equally. Furthermore, the correlation between \( \text{SaO}_2 \) (reflecting oxygenation), and respiratory rate (also dependent on respiratory drive and ventilation) varies across conditions.\(^{25}\) Therefore, the RDAI likely does not represent all dimensions of respiratory distress in bronchiolitis, and a combination of parameters may be more relevant for the measurement of respiratory distress, as seen in formally developed scales.\(^{28,29}\) However, most other scales were not developed specifically for bronchiolitis, and their measurement properties cannot be transferred between different respiratory conditions without further validation.

We found that the RDAI had reasonable predictive validity based on its association with hospitalization and length of stay in the ED. Our findings are consistent with those of Corneli et al,\(^{30}\) who identified RDAI score, \( \text{SaO}_2 \), and respiratory rate as predictors of hospitalization in bronchiolitis. On the contrary, Destino et al\(^{16}\) found that RDAI sum scores did not discriminate well between admitted and discharged patients, but the item on retractions did. Two large prognostic studies have also identified retractions as predictors of severe disease in ED and hospitalized patients.\(^{31,32}\) Decisions regarding hospitalization and length of stay in the ED are multifactorial. Nonrespiratory severity parameters (eg, feeding), prognostic factors (eg, age), social issues, clinical judgment, available resources, and local practices influence decision-making.\(^{30}\) Furthermore, there are limits to the validity of static
measurements of respiratory distress in a highly dynamic condition. From an outcome measure perspective, the RDAI does not encompass all determinants of bronchiolitis severity. Interrater reliability measured by the ICC was good, both at the group and individual level, as was interrater measurement error. These findings mean that RDAI scores can adequately discriminate participants assessed by different raters at the same time point in both clinical and research settings. Data from previous RCT reports also showed good interrater reliability, but Destino et al.\(^\text{16}\) found a strikingly low ICC. Differences may relate to training, familiarity with the instrument, raters, and population heterogeneity. On the other hand, we found considerable test-retest measurement error at the individual level, because a patient should change at least close to 4 points (approximately one-fifth of the scale) before a change is detectable beyond measurement error. Thus, in clinical practice, changes in individual patients should be interpreted with caution. For the RACS, we must also consider measurement error for respiratory rate.\(^\text{33,34}\) The SDC is paramount to interpretability parameters such as the minimal important change (MIC), because a large SDC relative to the MIC means that observed change may be caused by measurement error rather than change per se.\(^\text{35}\) At the individual level, taking repeated measurements and averaging the value would reduce the measurement error with a factor $\sqrt{k}$ ($k$ is the number of measurements). Although reassessment is a key component when evaluating children with respiratory distress, many repeated measurements might not be practical in clinical practice. At the group level, the SDC of a mean change is equal to $\text{SDC}/\sqrt{n}$, which reduces its impact.\(^\text{21}\) Because the ICC was high, the RDAI is reliable for use in studies. Overall, these results suggest that the RDAI has adequate discriminative properties, but test-retest measurement error should be minimized.

The RDAI was responsive according to our predefined hypotheses based on 2 distinct constructs of change. Previous data on RDAI responsiveness are scarce.\(^\text{13–15}\) Hardly any intervention can be considered clearly effective in bronchiolitis in the ED setting, and thus none is a reasonable gold standard to assess change. Destino et al.\(^\text{16}\) reported a mild correlation between the change in the RACS and the Children’s Hospital of Wisconsin Respiratory Score, but data on responsiveness of this latter scale are also missing. We anchored our constructs of change on physiologic change and change in clinical status likely to be relevant for decisions regarding patient disposition at the ED. Measures of responsiveness that took into account both improved and stable groups (responsiveness ratio and AUC) were comparable between anchors for the RDAI. The AUC value was close to the frequently used cutoff of acceptability (0.7) for both the RDAI and RACS, with the RACS being slightly more responsive. These data suggest that the RDAI and RACS are moderately responsive, but any comparison with other respiratory scales is limited.

Our study has limitations related to design constraints of both included studies. First, less heterogeneity of RDAI scores in the selected sample of CanBEST participants may explain why we found a weaker correlation with respiratory rate and lower test-retest ICC scores. Further validation is needed when considering children with very mild or severe disease, who were excluded in CanBEST. Second, our results are applicable to infants with a first episode of wheezing and no relevant comorbidities and should be interpreted with caution when defining bronchiolitis differently in other populations.\(^\text{36}\) Third, concurrent factors that affect decisions of hospitalization were not collected, and the exact timing of this changes in individual patients should be accounted for in future research.

### TABLE 5 Measures of Responsiveness for RDAI and RACS Using Different Anchors of Change

<table>
<thead>
<tr>
<th>Responsiveness Measure</th>
<th>Anchor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RDAI</td>
</tr>
<tr>
<td>Within-group mean change (RDAI) or change score (RACS) ± SD</td>
<td></td>
</tr>
<tr>
<td>Improved group</td>
<td>$-4.17 \pm 2.86$</td>
</tr>
<tr>
<td>Stable group</td>
<td>$-2.86 \pm 2.71$</td>
</tr>
<tr>
<td>Between-group difference in mean change (RDAI) or change score (RACS) (95% CI)</td>
<td></td>
</tr>
<tr>
<td>Standardized ES</td>
<td></td>
</tr>
<tr>
<td>Improved group</td>
<td>1.54</td>
</tr>
<tr>
<td>Stable group</td>
<td>1.22</td>
</tr>
<tr>
<td>Responsiveness ratio</td>
<td>1.54</td>
</tr>
<tr>
<td>AUC (95% CI)</td>
<td>0.64 (0.59–0.68)</td>
</tr>
</tbody>
</table>

Measurements were made at baseline and before admission or discharge. NA, not applicable.
decision was not known. Although, in ideal conditions, managing physicians would be blinded to RDAI/RACS scores, blinding to their individual items is not expected. Finally, defining stability and change can be problematic and time-dependent due to the dynamic nature of bronchiolitis. When assessing responsiveness by using data collected at different time points (mostly between 90 and 240 minutes), we observed significant improvements in RDAI scores in groups that we considered a priori to be stable. This finding is likely a limitation of our anchors and may also reflect the effect of supportive measures and the nebulized “placebo.” These limitations should be considered when calculating the MIC of the RDAI, which will be the focus of future work.

In conclusion, we found the RDAI to be an incomplete measure of respiratory distress in bronchiolitis, with poor to moderate construct validity and adequate interrater reliability. The RDAI had considerable test-retest measurement error, and although both the RDAI and RACS were moderately responsive, methodologic issues may limit the interpretation of this finding. Finally, the RDAI does not encompass all determinants of bronchiolitis severity.

ACKNOWLEDGMENTS

We thank the children and families who participated in CanBEST and the prognostic cohort study, the site research coordinators, the coordinating site staff, the data manager, all of the research nurses and participating site physicians and nurses, and Pediatric Emergency Research Canada members.

REFERENCES

13. van der Windt DA, Nagelkerke AF, Bouter LM, Dankert-Roelse JE, Veerman AJ.


Validity of Bronchiolitis Outcome Measures
Ricardo M. Fernandes, Amy C. Plint, Caroline B. Terwee, Cristina Sampaio, Terry P. Klassen, Martin Offringa and Johanna H. van der Lee

*Pediatrics* 2015;135:e1399; originally published online May 18, 2015;
DOI: 10.1542/peds.2014-3557

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/135/6/e1399.full.html