Diagnosing clinically significant dehydration in children with acute gastroenteritis using non-invasive methods: A Meta-Analysis

Stephen Freedman MDCM, MSc;¹ Ben Vandermeer, MSc;² Andrea Milne, MLIS;² Lisa Hartling, PhD,² on behalf of the Pediatric Emergency Research Canada Gastroenteritis Study Group*  

1 Sections of Pediatric Emergency Medicine and Gastroenterology, Alberta Children’s Hospital, Alberta Children’s Hospital Research Institute, University of Calgary, Calgary, Alberta, Canada  
2 Alberta Research Centre for Health Evidence, Department of Pediatrics, University of Alberta, Edmonton, Alberta, Canada

Corresponding Author

Stephen B. Freedman MDCM, MSc  
Sections of Emergency Medicine and Gastroenterology; Department of Pediatrics  
Alberta Children’s Hospital and Research Institute; University of Calgary  
2888 Shaganappi Trail; Calgary, AB, Canada; T3B 6A8  
E-mail: stephen.freedman@albertahealthservices.ca

No reprints.

Key words: Dehydration; Gastroenteritis; Systematic Review; Clinical Score; Ultrasound

*List of members of the Pediatric Emergency Research Canada Gastroenteritis Study Group is available at www.jpeds.com (Appendix).
Supported by the Canadian Institutes of Health Research (Knowledge Synthesis grant 262955).

L.H. holds a New Investigator Salary Award from the Canadian Institutes of Health Research.

The authors declare no conflicts of interest.

**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ao</td>
<td>Aorta</td>
</tr>
<tr>
<td>ED</td>
<td>Emergency Department</td>
</tr>
<tr>
<td>CDS</td>
<td>Clinical Dehydration Scale</td>
</tr>
<tr>
<td>HSROC</td>
<td>Hierarchical Summary Receiver-Operating Curves</td>
</tr>
<tr>
<td>IVC</td>
<td>Inferior Vena Cava</td>
</tr>
<tr>
<td>LR</td>
<td>Likelihood Ratio</td>
</tr>
<tr>
<td>PRISMA</td>
<td>Preferred Reporting Items for Systematic reviews and Meta-Analyses</td>
</tr>
<tr>
<td>QUADAS-2</td>
<td>Quality Assessment of Diagnostic Accuracy Studies 2</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver-Operating Curve</td>
</tr>
</tbody>
</table>
ABSTRACT

**Objective** To determine the most accurate non-invasive method of assessing dehydration.

**Study design** The following data sources were searched: electronic databases, grey literature, scientific meetings, reference lists, and authors of unpublished studies. Eligible studies were comparative outpatient evaluations, employing an accepted reference standard, conducted in developed countries, in children aged <18 years, with gastroenteritis. Data extraction was completed independently by multiple reviewers before arriving at a consensus.

**Results** Nine studies which included 1,039 participants were identified. The 4-item Clinical Dehydration Scale, the ‘Gorelick’ score, and unstructured physician assessment were evaluated in 3, 2, and 5 studies, respectively. Bedside ultrasound, capillary digital videography, and urinary measurements were each evaluated in one study. The Clinical Dehydration Scale had a likelihood ratio (LR)+ range of 1.87–11.79 and LR- range of 0.30–0.71 to predict 6% dehydration. When combined with the 4-item Gorelick Score, the LR+ was 1.93 (95% CI 1.07, 3.49) and LR- of 0.40 (95% CI 0.24, 0.68). Unstructured dehydration assessment had a pooled LR+ of 2.13 (95% CI 1.33, 3.44) and LR- of 0.48 (95% CI 0.28, 0.82) to detect ≥5% dehydration.

**Conclusions** Clinical dehydration scales appear to have superior accuracy to unstructured clinical assessment; however, the overall accuracy is suboptimal. Current evidence does not support the routine use of ultrasound or urinalysis to determine dehydration severity.
The cornerstone of gastroenteritis management is dehydration assessment with therapy instituted based on severity.\(^1\), \(^2\) However, dehydration is difficult to determine clinically\(^3\) and change in body weight remains the “gold standard”.\(^3\) Unfortunately, recent well weights are rarely available\(^4\) and the inaccuracy of available tests limits the ability of clinicians to estimate the exact degree of dehydration.\(^3\) Consequently, research has focused on non-invasive methods of assessing dehydration (e.g. clinical scores,\(^5\)-\(^8\) bedside ultrasound\(^4\), \(^9\), \(^10\), urine ketones\(^11\)). Scores, employing combinations of examination findings, may perform better than individual signs at predicting dehydration.\(^3\) Popular examples include the “Gorelick”\(^12\) and Clinical Dehydration Scales (CDS).\(^5\) These scales have been adopted yet their ability to predict severe dehydration is suboptimal.\(^13\) For example, in a recent report the 4- and 10-point Gorelick scale had sensitivities of only 64% and 21% respectively for severe dehydration.\(^13\) Similarly the ability of bedside ultrasound to assess intravascular volume status remains a topic of debate.\(^14\) Conflicting opinions may relate to the study population and outcome measures employed.

A systematic review and meta-analysis focused on developed countries can guide their integration into evidence-based practice in such countries thereby reducing the overuse of intravenous rehydration.\(^15\) Because diagnostic test characteristics (e.g. predictive values) are dependent on disease prevalence, evaluating tests of dehydration in the context of developed countries is important. With guidelines recommending that therapy be tailored to clinical scores\(^16\), \(^17\) and bedside ultrasound becoming a standard technology in pediatric emergency departments (ED) across North America,\(^18\) their roles must be defined.\(^19\), \(^20\) Thus, we conducted a systematic review of studies evaluating the diagnostic test accuracy of non-invasive methods of dehydration assessment in developed countries.
METHODS

We followed a standard protocol for the conduct and reporting of systematic reviews and meta-analyses, which was in keeping with the PRISMA guidelines.(21, 22)

An experienced medical librarian developed a search strategy in collaboration with the research team to identify studies examining the diagnostic accuracy of non-invasive methods of assessing dehydration. We (1) systematically searched MEDLINE (1946 - April 2013), EMBASE (1980 - April 2013), the Cochrane Central Register of Controlled Trials (April 2013) via the OvidSP platform, PubMed via the National Library of Medicine (last 180 days), and for grey literature; (2) hand-searched appropriate journals and major, relevant scientific meetings (Society for Pediatric Research 2010-2012, American Academy of Pediatrics 2010-2012, Canadian Pediatric Society 2010-2012, and International Conference on Emergency Medicine 2012); (3) checked reference lists of relevant studies; and (4) contacted primary authors of published and unpublished studies. The MEDLINE search strategy is appended (Table I; available at www.jpeds.com). The search was not restricted by language or publication status. We ran an updated search of the electronic databases in October 2014 to identify studies published because the first search; no eligible studies were identified. All studies contained in previous relevant systematic reviews were screened for inclusion.

Search result titles and abstracts were screened independently by two reviewers to identify potentially relevant citations. They were excluded when the title/abstract did not identify that the article addressed the accuracy of a non-invasive method of assessing dehydration. The full text
of all potentially relevant citations was obtained and assessed for inclusion by two independent reviewers using standard, predefined eligibility criteria. Disagreements were resolved by consensus. Decisions regarding inclusion and reasons for exclusion were documented. Original studies were included if they: (1) evaluated children <18 years of age suspected to have dehydration; (2) examined the diagnostic accuracy of a non-invasive method of dehydration assessment against percent change in body weight between acute presentation and stable, rehydrated, well weight (Table II) (12); (3) were conducted in a developed country as defined by the United Nations in 2011—Australia, Canada, Europe, Japan, New Zealand, and the United States(23); and (4) were conducted in an ED or similar clinical setting. We included studies where authors focused on children with acute gastroenteritis. Comparative studies meeting the above criteria were included. Review articles were excluded.

As is commonly performed, one reviewer extracted data using a structured form. Verification was performed by a second reviewer for accuracy and completeness.(24-26) The following items were extracted: study characteristics (eg, date of publication, clinical setting, country), participants (eg age, sex), dehydration scores and comparisons, outcomes (eg diagnostic accuracy), source of funding, and results. Extracted data were entered into Microsoft Excel (Microsoft, Redmond, WA) worksheets. Disagreements were resolved by consensus, or involving a third reviewer as required. The Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool(27) was used to assess the methodological quality of the relevant studies. QUADAS-2 includes four domains: patient selection, index test, reference standard, and flow of patients through the study and the timing of the index tests and reference standard (flow and timing).(28) Assessments regarding bias and applicability are made for each domain. Bias is
assessed as low, unclear, or high risk; applicability is assessed as low, unclear, or high concerns. Quality assessment was completed independently by two reviewers. Disagreements were resolved by consensus, or involving a third reviewer as required.

We developed evidence tables to describe the studies including information on design features, methodological quality, study populations, sample size, settings, dehydration scores and comparisons. For each of the included studies we extracted the raw data regarding true and false positives and negatives and constructed 2 x 2 tables to calculate sensitivity, specificity, and likelihood ratios (LR). Sensitivity and specificity are measures of test accuracy. Likelihood ratios are used to estimate the increased or decreased probability of disease (i.e. dehydration) for a patient and can be used to refine clinical judgment. The larger the positive LR, the greater the accuracy of the test and the greater the likelihood of disease after a positive test result. On the other hand, the smaller the negative LR, the lower the likelihood of disease after a negative test result. (29) Sensitivities, specificities, LRs, and predictive values are presented in a summary table that includes all dehydration assessment methods. We planned to analyze data using hierarchical summary receiver-operating curves (HSROC); however, an insufficient number of studies examining any given test were identified to enable the use of this approach. (30) Consequently, we plotted the sensitivity and specificity of the individual studies in a receiver-operating curve (ROC) space to graphically display the relative accuracy of the different measures. We pooled likelihood ratios using Maentel-Hansel methods and random effects models. We were unable to formally assess for publication bias due to the small numbers of studies examining any given test.
RESULTS

The electronic database search identified 1,454 citations; 66 were considered potentially relevant based on their title/abstract (Figure 1; available at www.jpeds.com). Of these, 4 met inclusion. Five additional studies meeting eligibility criteria were identified by reviewing the references of relevant studies (Table II). The median year of publication was 2007. The 9 eligible studies included 1,293 participants of whom 1,039 (80%) had both the diagnostic evaluation and the reference gold standard performed. Eight studies recruited patients presenting for ED care. Six studies classified dehydration employing a <5%; 5-10% and >10% scale; 3 used <3%; 3-6% and >6% as their reference standard to reflect mild/moderate/severe dehydration. Studies displayed significant heterogeneity in terms of participant age, both at the younger and older extremes.

Quality of Studies

Patient selection was rated as low risk of bias in 1 study; the remaining studies were unclear or at high risk of bias. Six studies had unclear or high concerns about applicability largely because information on patient selection was not reported. The index test domain was at low risk of bias in 5 studies. One study had high risk of bias because researchers did not use pre-specified thresholds for classifying dehydration severity. Three studies were assessed as unclear as they did not provide sufficient information to determine whether the researchers pre-specified thresholds or whether the reference standard results were known when the index test was interpreted. All studies had low concerns regarding applicability. The domain of the reference standard was at low risk of bias for 5 studies; 4 were rated as unclear because it was unknown whether the reference standard results were interpreted without knowledge of the index test result or it was unclear when the well-weights were assessed. All studies had low concerns
regarding applicability for this domain. Flow and timing was at low risk of bias for 4 studies; 5 were high risk of bias because a significant proportion of participants were lost to follow-up or had incomplete measurements (Table III and Figure 2; available at www.jpeds.com).

Non-Invasive Assessments

The CDS was evaluated in 3 studies. This 4-item scale was derived in a cohort of 137 children. The authors employed formal approaches to derive the scale. Although diagnostic accuracy was not reported in the derivation sample, subsequent work has conducted such analyses assigning scores of 0, 1 or 2 for each item which is ranked along a continuum from normal to significantly abnormal. These studies (Table IV) revealed that a score of 0 had a positive LR of 1.64 – 2.19 and negative LR of 0.79 – 0.84 to predict <3% dehydration. The ability of the score to predict ≥5% (or >6%) dehydration had positive LRs ranging between 1.87 – 11.79 and negative LR ranged between 0.30 – 0.71 (Tables IV and V and Figure 3; Table V available at www.jpeds.com).

The ‘Gorelick’ score which formalized dehydration evaluation employing clinical features uses 4 and 10 item scales. Because of the overlap in clinical features between the CDS score and the 4-item Gorelick Score these scores were combined to generate pooled estimates with positive LR = 1.93 (95% CI 1.07, 3.49) and negative LR = 0.40 (95% CI 0.24, 0.68). The overall accuracy of these scales ranged between 57 – 88% (Table IV).

Physician clinical assessment of dehydration has been evaluated employing several approaches. These include an overall hydration assessment by (1) junior doctors who were instructed to
classify children as being 5 – 10% dehydrated based on their hospital’s guideline;(32) (2) the fellow or attending physician employing a 9-item ‘standard clinical scale’;(33) (3) the attending physician, blinded to ultrasound findings employing an ordinal scale (<5%; 5% – 10%; >10%);(4) (4) the pediatric emergency physician who estimated their degree of dehydration using a 7-point Likert scale (very mild to very severe);(34) and (5) having the treating pediatric emergency medicine attending physicians record their clinical impression (1 to 10 scale) for percent dehydration.(35)

Excluding a study with 100% sensitivity because it included only children deemed to be 5 – 10% dehydrated (ie, excluded children who might have been false negatives),(32) the sensitivity of this approach to detect ≥ 5% dehydration ranged from 33% to 78%. The pooled results from four studies for positive and negative LRs were 2.13 (95% CI 1.33, 3.44) and 0.48 (95% CI 0.28, 0.82), respectively. Overall accuracy ranged between 25 – 81% (Table IV).

Chen et al evaluated the accuracy of an inferior vena cava (IVC) to aorta ratio of < 0.8 to predict ≥ 5% weight loss.(4) Although ultrasound had a LR negative value of 0.27, the LR positive of 1.95, was significantly lower than of both the Clinical Dehydration and the Gorelick Scales for identifying > 6% dehydration. A single study also evaluated the ability of digital capillary refill to predict ≥ 5% body weight change. The authors reported an LR positive of 11.67 and LR negative of 0.00 and the highest overall accuracy of any measure evaluated (93%).(34)

Inter-observer Assessments (ie, reproducibility)
These were performed in 4 of the 9 studies and varied in size related to the entire cohort. Kappa values were lowest for the clinical measures – 0.65 for the CDS and 0.69 for the Gorelick Scale.(7, 12) The values were higher for ultrasound (Pearson correlation coefficient = 0.76)(4) and digital capillary refill (intraclass correlation = 0.99).(34)

**DISCUSSION**

The practice of clinical medicine is founded on the reliability of history and physical examination to guide management. However, the evidence regarding the diagnostic accuracy of clinical examination is often lacking; pediatric dehydration is no exception. The use of the CDS or Gorelick Scale both consistently reported accuracy values of > 80% in identifying significant dehydration, although lower values were noted by Jauregui et al.(35). Overall, the clinical scales evaluated provide some improved diagnostic accuracy. However, test characteristics indicate that a substantial gap remains in how well the scores predict the outcome of interest. Similarly, bedside ultrasound, which may aid in ruling out dehydration (LR negative of 0.27), should not be used to confirm the presence of dehydration (LR positive of 1.95). The most promising intervention we identified was the use of digital videography capillary refill time; given the paucity of data however, further research is required to confirm the preliminary findings.

Our findings are congruent with a previous review that included studies evaluating dehydration assessment in developed and developing countries.(3) This review concluded that the quantity and quality of research on dehydration assessment is extremely limited with most studies being of only moderate sample size and at significant risk of patient selection bias. Many were conducted at pediatric institutions and enrolled exclusively children deemed to be dehydrated, and those needing rehydration. This highlights how study populations may not reflect the real
world situation, where clinicians must identify the small number of children with significant dehydration from those who are not dehydrated.

When considering our findings it is important to consider what the most important measure of test performance is – overall test accuracy (true positive + true negative) versus sensitivity or negative likelihood ratio. The decision reached is likely based on one’s perspective on the number to over-diagnose (ie, unnecessary intravenous fluids) to identify one case of severe dehydration. Clinicians must consider the limited importance of distinguishing small increments of dehydration. This issue is highlighted in guidelines published by the Centers for Disease Control which group children into none/minimal (< 3-5%), mild – moderate (3 – 9%) and severe (≥9%) dehydration and tie management to these categories.(2)

Ultrasound has been touted as having the potential to serve as a useful adjunct for this clinical dilemma.(36) In the single study identified (4), a pair of investigators obtained transverse images of the IVC and the aorta and then calculated their ratio (IVC/aorta). The authors reported that an IVC/Ao cut-off of 0.8 produced a sensitivity of 86%, a specificity of 56% and a positive predictive value of 62% for the diagnosis of significant dehydration. These characteristics are not superior to clinical dehydration scales, and less than two-thirds of test positive patients will have significant dehydration. Thus, there is insufficient evidence to support the routine use of ultrasound due to the potential to result in “over-treatment” if test characteristics are not considered. More recent work has also concluded that the IVC and aorta measurements by bedside ultrasound are not reliable indicators of intravascular volume as determined by central venous pressure.(14) Digital capillary refill time has only been described in a single study(34).
The present study has several potential limitations. The greatest challenge in quantifying dehydration severity is the gold standard. Although this study required, in keeping with expert opinion and previous studies,(3, 5, 12, 32, 33) that post-illness weight be used as the gold standard, the included studies differed significantly in the precise day used as the reference measure of stable post-illness weight (Table I).(4, 12, 33, 37) An additional concern relates to the timing within a given day, of rehydration weights which are influenced by stooling, voiding, and feeding, especially in infants.(3) Nonetheless, this study adhered to the use of this gold standard as the criteria for defining true percent dehydration. Although this study followed accepted methodological standards for the conduct of systematic reviews of diagnostic test accuracy including a comprehensive search of the literature, there remains the possibility that some unpublished studies may not have been identified. Although formally evaluating publication bias through the use of funnel plots and statistical tests would be ideal, due to insufficient numbers of studies included in this review, a formal analysis of publication bias was not possible. Clinical dehydration scales appear to have superior accuracy to unstructured clinical assessment however, the overall accuracy is suboptimal.

ACKNOWLEDGMENTS

We thank the following for their contributions to this project: Melanie Muise and Dion Pasichnyk for their assistance with screening, study selection, data extraction and verification, and Donna Dryden, MD, for assistance with quality assessment.


