Diagnosing Clinically Significant Dehydration in Children with Acute Gastroenteritis Using Noninvasive Methods: A Meta-Analysis

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

Objective To determine the most accurate, noninvasive method of assessing dehydration.

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data sources were searched: electronic databases, gray literature, scientific meetings, reference lists, and authors of unpublished studies.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

Results Nine studies that included 1039 participants were identified. The 4-item Clinical Dehydration Scale (CDS), 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 CDS had a positive likelihood ratio (LR) range of 1.87-11.79 and a negative LR range of 0.30-0.71 to predict 6% dehydration. When combined with the 4-item Gorelick Score, the positive LR was 1.93 (95% CI 1.07-3.49) and negative LR was of 0.40 (95% CI 0.24-0.68). Unstructured dehydration assessment had a pooled positive LR of 2.13 (95% CI 1.33-3.44) and negative LR of 0.48 (95% CI 0.28-0.82) to detect ≥5% dehydration.

Conclusions Overall, the clinical scales evaluated provide some improved diagnostic accuracy. However, test characteristics indicate that their ability to identify children both with and without dehydration is suboptimal. Current evidence does not support the routine use of ultrasound or urinalysis to determine dehydration severity. *(J Pediatr 2015;166:908-16)*.

The cornerstone of gastroenteritis management is the assessment of dehydration, 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 noninvasive methods of assessing dehydration (eg, clinical scores,[5-8] bedside ultrasound,[4,9,10] urine ketones[11]). Scores, by using combinations of examination findings, may perform better than individual signs at predicting dehydration.[3] Popular examples include the "Gorelick"[12] and Clinical Dehydration Scales (CDSs).[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 used.

A systematic review and meta-analysis focused on research in developed countries can enhance the integration of evidence into clinical care in such countries. A key focus of quality improvement efforts in children with gastroenteritis in developed countries is to promote the use of oral rehydration therapy when appropriate in order to minimize the unnecessary administration of intravenous fluids.[15] Because diagnostic test characteristics (eg, predictive values) are dependent on the prevalence of disease, 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 (EDs) across North America,[18] their roles must be defined.[19,20] Thus, we conducted a systematic review of studies in which investigators evaluated the diagnostic test accuracy of noninvasive methods of dehydration assessment in developed countries.

Ao	Aorta
CDS	Clinical Dehydration Scale
ED	Emergency department
IVC	Inferior vena cava
LR	Likelihood ratio

From the [1]Sections of Pediatric Emergency Medicine and Gastroenterology, Alberta Children's Hospital, Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Alberta, Canada; and [2]Alberta Research Centre for Health Evidence, Department of Pediatrics, University of Alberta, Edmonton, Alberta, Canada

*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.

http://dx.doi.org/10.1016/j.jpeds.2014.12.029

Methods

We followed a standard protocol for the conduct and reporting of systematic reviews and meta-analyses, which was in keeping with the Preferred Reporting Items for Systematic reviews and Meta-Analyses 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 noninvasive methods of assessing dehydration. We (1) systematically searched MEDLINE (1946 to April 2013), EMBASE (1980 to 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 gray literature; (2) handsearched 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 after 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 2 reviewers to identify potentially relevant citations. They were excluded when the title/abstract did not identify that the article addressed the accuracy of a noninvasive method of assessing dehydration. The full text of all potentially relevant citations was obtained and assessed for inclusion by 2 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 noninvasive 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 US[23]; and (4) were conducted in an ED or similar clinical setting. We included studies in which authors focused on children with acute gastroenteritis. Comparative studies meeting the aforementioned criteria were included. Review articles were excluded.

As is commonly performed, 1 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, Washington) worksheets. Disagreements were resolved by consensus, or involving a third reviewer as required. The Quality Assessment of Diagnostic Accuracy Studies 2 tool[27] was used to assess the methodologic quality of the relevant studies. The Quality Assessment of Diagnostic Accuracy Studies 2 includes 4 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 2 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, methodologic 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 2×2 tables to calculate sensitivity, specificity, and likelihood ratios (LRs). Sensitivity and specificity are measures of test accuracy. LRs are used to estimate the increased or decreased probability of disease (ie, 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. In contrast, 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; 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 space to graphically display the relative accuracy of the different measures. We pooled LRs using Maentel-Hansel methods and random effects models. We were unable to formally assess for publication bias because of the small numbers of studies examining any given test.

Results

The electronic database search identified 1454 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 1293 participants, of whom 1039 (80%) had both the diagnostic evaluation and the reference gold standard performed. Eight studies