Diagnostic Accuracy of Tests in Pediatric Gastroesophageal Reflux Disease

Rachel J. van der Pol, MD[1,*], Marije J. Smits, MSc[1,*], Leonie Venmans, PhD[3], Nicole Boluyt, MD, PhD[2], Marc A. Benninga, MD, PhD[1], and Merit M. Tabbers, MD, PhD[1]

Objective To systematically review the literature evaluating the diagnostic accuracy of commonly used diagnostic tests over conventional history taking and physical examination in children ≤18 months and >18 months suspected of gastroesophageal reflux disease (GERD).

Study design We searched Medline, Embase, and the Cochrane database for studies assessing the diagnostic accuracy of pH-metry, pH-impedance, esophagogastroscopy, barium contrast study, scintigraphy, and empirical treatment as diagnostic tools. Quality was assessed according to Quality Assessment of Studies of Diagnostic Accuracy Included in Systematic Reviews criteria.

Results Of the 2178 studies found, 6 studies were included, containing 408 participants (age 1 month-13.6 years) and 145 controls (age 1 month-16.9 years). Studies included children with GERD symptoms; 1 included an atypical presentation. In all the studies, the diagnostic accuracy of pH-metry was investigated, and in 2 studies esophagogastroscopy was investigated as well. Sensitivity and specificity were calculated in 3 studies. The range of reported sensitivity and specificity was broad and unreliable because of poor methodological quality according to Quality Assessment of Studies of Diagnostic Accuracy Included in Systematic Reviews criteria and inadequate study design.

Conclusion Diagnostic accuracy of tests in children suspected of GERD remains unclear and implications for practice are hard to give. There is an urgent need of well-designed randomized controlled trials where the effect of treatment according to specific signs and symptoms will be compared with the effect of treatment based on the results of additional diagnostic tests, for patient relevant outcomes. *(J Pediatr 2013;162:983-7)*.

Gastroesophageal reflux (GER) is a physiologic process. Regurgitation occurs in over 70% of infants multiple times a day, but it tends to disappear by the age of 12-14 months.[1,2] Gastroesophageal reflux disease (GERD) is defined and diagnosed when GER leads to troublesome symptoms and/or complications.[3] In 2009, GERD prevalence was estimated to be 12.3% in North American infants and 1% in older children.[4] Troublesome symptoms in infants may include excessive crying, back arching, regurgitation, and irritation around feedings; these could be regarded as nonspecific. In children and mainly in adolescents, heartburn is the more specific symptom occurring in GERD. Though complaints are often mild, they are troublesome and may have a significant impact on the wellbeing of the child and family life. Moreover, complications as esophagitis and hematemesis, failure to thrive, or apparent life threatening events (ALTE) have to be prevented whenever possible.[3,5,6]

Diagnosing GERD in pediatric patients is difficult because no gold standard exists, and not one combination of symptoms is conclusive. Currently, the diagnosis of GERD is based on history and physical examination. This approach might be considered as the "gold standard." However, there is a need to quantify GERD in a more objective way because the GERD diagnosis is subject to free interpretation and is probably overdiagnosed.[7] It may mimic disorders such as cow's milk allergy and eosinophilic esophagitis.[8,9]

Tests for GERD can be divided into 2 categories: tests with the ability to measure reflux events (pH-metry, pH-impedance, barium contrast studies, and scintigraphy) and tests to detect the consequences of reflux events (esophagogastroscopy).

The most widespread test used to quantify GERD is 24-hour pH-metry. A pH < 4 in the esophagus is generally considered as an acid reflux episode.[3] Acid exposure is expressed as the reflux index (RI, % of time a pH < 4 was measured), for which currently no evidence based pediatric normal values exist. The European Society for Pediatric Gastroenterology, Hepatology, and Nutrition and North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines consider a RI >7% as abnormal, a RI <3% as normal, and between 3% and 7% as indeterminate.

ALTE	Apparent life threatening event
GER	Gastroesophageal reflux
GERD	Gastroesophageal reflux disease
PPI	Proton pump inhibitor
QUADAS	Quality Assessment of Studies of Diagnostic Accuracy Included in Systematic Reviews
RI	Reflux index
SR	Systematic reviews

From the Departments of [1]Pediatric Gastroenterology and Nutrition and [2]Pediatrics, Emma Children's Hospital, Academic Medical Center, Amsterdam, The Netherlands; and the [3]Pediatric Association of The Netherlands, Utrecht, The Netherlands

*Contributed equally.

The authors declare no conflicts of interest.

 http://dx.doi.org/10.1016/j.jpeds.2012.10.041

In infants, however, frequency of feeds and buffering capacities of milk may confound outcomes of pH-metry studies.[10] In addition to 24-hour pH-metry, 24-hour pH-impedance measurement (pH-impedance) was developed.[11] Equipped with a pH sensor and multiple electrode pairs, it measures the conductivity of liquid, gas, or mixed contents in the esophagus and is able to detect non-acid and alkaline reflux besides acid reflux.[12,13]

Barium contrast studies consist of a series of radiographs of the esophagus and stomach using a barium emulsion to track swallows and possible reflux, which sometimes reveal structural anatomic causes underlying GERD.[14] In gastroesophageal scintigraphy, patients consume a ^{99}technetium labeled meal prior to start of the scans, and postprandial reflux becomes visible when labeled stomach contents move upwards in the esophagus.[3]

Reflux esophagitis, one of the complications of GERD, can be measured by esophagogastroscopy. This enables both macroscopic and microscopic grading of the esophageal wall. To date, there are insufficient data to support the use of histology in diagnosing GERD. Currently, the main reason for taking biopsies is to exclude other diseases causing esophagitis such as eosinophilic esophagitis, Crohn's disease, and infections.[3,15]

Finally, a trial with an antireflux agent may be used to diagnose GERD. A proton pump inhibitor (PPI) is often the agent of choice and an empiric trial of 2-4 weeks is common.[3] Data on sensitivity and specificity is scarce, both in adults and children, amplified by the fact that GERD symptoms may improve spontaneously or respond by a placebo effect.[3]

The accuracy of the above-mentioned tests is unclear, and, therefore, it is questionable if these more invasive and expensive tools should be used. We carried out a systematic review to evaluate the accuracy of pH-metry, pH-impedance of esophagogastroscopy, barium contrast study, scintigraphy, and diagnostic treatment compared with conventional history and physical examination when diagnosing GERD.

Methods

A clinical librarian searched Medline, Embase, and the Cochrane Database of systematic reviews (SR) electronic database for SRs, and clinical studies from inception to May 2012. The key words used to describe the study population were "esophagogastroscopy," "pH-metry," "pH-impedance," "gastric emptying scintigraphy," "barium radiography," "GER," "GERD," "heartburn," "extraesophageal symptoms," "reflux esophagitis," "infant," "child," and "adolescent" (medical subject headings and all fields). No language restriction was applied. Reference lists of reviews and included studies were searched for additional studies. The full search strategy is available from the authors.

Two reviewers independently selected the abstracts of identified studies for suitability. Inclusion criteria were: (1) the study was an SR or clinical study; (2) children were aged 0-18 years presenting with signs and symptoms (through history or clinical examination) suggestive of GERD; (3) the aim of the study was to evaluate the diagnostic accuracy of esophagogastroscopy, pH-metry, pH-impedance (symptom index, symptom sensitivity index, and symptom association probability had to be given), scintigraphy, barium swallow/radiograph of esophagus/stomach, or diagnostic treatment (at least 1 week of treatment compared with history and physical examination); and (4) the study had to use a control group. Exclusion criteria were: (1) no definition of GERD; (2) patients who had a disease frequently related to GERD (eosinophilic esophagitis, malformation of the esophagus, [congenital] hernia diaphragmatica, achalasia, cystic fibrosis, gastric paresis, systemic sclerosis, children with neurologic impairment, cow's milk allergy, and rumination syndrome); (3) patients who had undergone surgical therapy; and (4) children who were treated for GERD during history and physical examination or prior to the investigated diagnostic test or vice versa (and the GERD therapy was not the investigated diagnostic tool).

All potentially relevant studies and the studies for which the abstracts did not provide sufficient information for in- or exclusion were retrieved as full articles.

Two reviewers assessed methodological quality of all identified studies by the Quality Assessment of Studies of Diagnostic Accuracy Included in Systematic Reviews (QUADAS) checklist.[16] Because the revised QUADAS checklist was published after the quality assessment process took place, we were not able to incorporate this newer checklist.[17] The revised QUADAS checklist offers additional and improved features and has improved in distinguishing between bias and applicability and is capable of rating risk of bias. From the original standardized list, we choose 11 items (scored 'yes,' 'no,' or 'unclear') that could optimally differentiate for methodological quality (**Table I**; available at www.jpeds.com). Calculations on summary scores are not provided because they ignore the importance of individual items and because cut-off values on what is a good or bad score will be arbitrarily determined; these results may be misleading.[16] In general, the more items answered with 'yes,' the higher methodological quality is presumed.

Structured data extraction was performed by 2 reviewers independently. Data derived from included articles contained items such as author and year of enrollment, diagnostic method, study setting, methods, patient characteristics, number of participants and controls, index test and execution, sensitivity, and specificity. Because symptom presentation and pathophysiology is different in infants (≤18 months) and children, we choose to extract data, if possible, for infants and children separately. If disagreement between the 2 reviewers existed, consensus was formed, or a third reviewer (M.T.) made the final judgment.

Results

The search generated 2178 studies, of which 106 met our inclusion criteria (**Figure**; available at www.jpeds.com). No valid SR was encountered. After retrieving the full-text articles, 100 articles were excluded because of the lack of a control group, comparison between 2 diagnostic tests, and, therefore, no comparison with history and physical