



## **Gastroesophageal reflux disease in children: how reliable is the gold standard?**

**E. van Os,<sup>1</sup> J. De Schryver,<sup>2</sup> R. H. J. Houwen,<sup>2</sup> W. E. Tjon A. Ten<sup>3</sup>**

### **Abstract**

**Objective:** To investigate the causes and degree of interobserver variability in esophageal pH monitoring for the diagnosis of gastroesophageal reflux.

**Methods:** This retrospective study included all children (n = 72) who underwent pH monitoring during 1 year at Maxima Medical Centre in Veldhoven, the Netherlands.

**Results:** An interobserver variability of 18% was found. Variability was caused by differences in opinion about the duration of registration, doubts about probe position, artifacts and drift of baseline pH.

**Conclusions:** Most of these problems can be eliminated by posttest calibration and assessment of the pH electrode position. However, a clear definition of monitoring artifacts is lacking. This study shows that mutual agreement in the interpretation of pH studies was fair (kappa coefficient of 0.70).

*J Pediatr (Rio J). 2009;85(1):84-86: Gastroesophageal reflux, gold standard, children, pH monitoring.*

### **Introduction**

Prolonged intraesophageal pH monitoring is currently considered to be the most reliable method for detection and quantification of gastroesophageal reflux (GER).<sup>1-4</sup> A standardized protocol describing the methodology of esophageal pH monitoring and interpretation of the data for the diagnosis of gastroesophageal reflux disease (GERD) is used to enable comparison with "normal values."<sup>4</sup>

The standard parameter used for the diagnosis of GERD is the percentage of time over 24 hours during which esophageal pH is less than 4, the reflux index (RI). The North American Society for Pediatric Gastroenterology, Hepatology and Nutrition (NASPGHAN) defines a RI greater than 12% in the first year and greater than 6% in older children as pathological.<sup>2</sup>

In various studies it was shown that RI calculated by an automated program may be misleading and the quality of the

pH recording should therefore be assessed manually to eliminate technical problems, such as a drift of the baseline pH during registration,<sup>5</sup> artifacts,<sup>6</sup> shifts in the pH electrode position and failures in registration.<sup>4</sup> As a manual review is somehow subjective, this may result in a different interpretation of a pH study by different observers.

In this study we identified the causes and the degree of interobserver variability in the interpretation of pH studies. Although we assume that there is a high degree of mutual agreement, to our knowledge this has not been investigated before.

### **Methods**

A retrospective study was performed of all 72 children who underwent a 24-hour pH study at Maxima Medical Centre in Veldhoven, the Netherlands, during 1 year. There were no inclusion or exclusion criteria. All pH studies performed during 1 year were included in the study. The Synectics System

1. Department of Pediatrics, University Medical Centre St. Radboud, Nijmegen, The Netherlands.
2. Pediatric gastroenterologist, Department of Pediatric Gastroenterology, Wilhelmina Hospital for Children, Utrecht, The Netherlands.
3. Pediatric gastroenterologist, Department of Pediatrics, Maxima Medical Centre, Veldhoven, The Netherlands.

No conflicts of interest declared concerning the publication of this article.

**Suggested citation:** van Os E, De Schryver J, Houwen RH, Ten WE. Gastroesophageal reflux disease in children: how reliable is the gold standard? *J Pediatr (Rio J)*. 2009;85(1):84-86.

Manuscript received Apr 07 2008, accepted for publication Jun 25 2008.

doi:10.2223/JPED.1834

**Table 1** - Causes of interobserver variability (number of patients)

| Technical problems                                 | n (%)      |
|--|------------|
| Registration too short                             | 4 (5.5)    |
| Doubts about the correct position of the electrode | 2 (2.7)    |
| No agreement about artifacts                       | 6 (8.3)    |
| Drift baseline                                     | 1 (1.4)    |
| Total variability                                  | 13/72 (18) |

with semi-disposable antimony pH electrodes (Synectics Medical AB, Stockholm, Sweden) was used. The pH electrode was calibrated before every investigation in buffers of pH = 1 and pH = 7. The pH electrode was positioned at the level of the third vertebra above the diaphragm by fluoroscopy. An external reference electrode was used. Patient's data were stored in a portable Digtrapper MK III (Synectics Medical), data were analyzed by means of PolyGram Function Testing Software of Medtronic (Medtronic Synectics, Shoreview, MN, USA) and the original data were presented for evaluation to authors W.T.A.T. and J.D.S. digitally. Both W.T.A.T. and J.D.S. are experienced pediatric gastroenterologists with a special interest in gastroesophageal reflux. J.D.S. has been working in an academic hospital for more than 20 years and W.T.A.T. has been working in a large nonacademic hospital for 10 years.

Although in real life pH monitoring is performed and interpreted while the investigator knows the indication for a pH study, both W.T.A.T. and J.D.S. were unaware of the clinical data of the patients to avoid a biased interpretation. Interpretation by W.T.A.T. and J.D.S. took place independently. A pH study was considered pathological if RI was greater than 12% in the first year and greater than 6% in older children. The causes of interobserver variability were identified and inter-rater agreement was calculated using Cohen's kappa coefficient.

## Results

The mean age of the children included in this study was 13.1 months (SD 5.6 months) and 43 of those (57%) were males. The reasons for a different interpretation of the pH study were mainly technical (Table 1). In four cases the pH tracing was assessed by J.D.S. as pathological because of very frequent reflux episodes, despite the fact that the recording was shorter than the recommended 18 hours.<sup>4</sup> In two cases there was no agreement whether there had been a shift in the position of the pH electrode. In one case there was doubt whether there had been a drift of the baseline pH. A different interpretation of artifacts was the main reason for disagreement.

Overall a different interpretation was present in 13 of the 72 pH tracings (18%) (kappa 0.70 – the kappa coefficient captures the degree of agreement after chance-adjusted inter-observer agreement; values between 0.40 and 0.75 represent fair to good agreement).

## Discussion

Prolonged intraesophageal pH monitoring is currently considered to be the most reliable method for diagnosing GERD.<sup>1-4</sup> Although the current study shows a kappa coefficient of 0.70, this coefficient together with a reproducibility of 24-hour pH monitoring on 2 consecutive days of 70 to 80%<sup>7,8</sup> may result in a misclassification of a considerable number of patients with GERD. A registration shorter than 18 hours, although recommended,<sup>1,4</sup> may not cause problems in clinical practice, since many clinicians only tend to assess such a pH study as pathological when there are frequent reflux episodes and the patient history fits GERD. When in doubt intraesophageal pH monitoring will be repeated.

Doubts about the position of the electrode will not occur if one uses a standardized protocol, which includes posttest registration of the electrodes position. A drift of the baseline pH during 24-hour pH studies using antimony probes is common. Although the drift is generally small, it may result in a change in interpretation after posttest calibration drift adjustment for pH threshold is applied.<sup>5</sup> In this study posttest calibration was not applied; therefore, whether drift of baseline pH affected the results is not known.

In adults with suspected laryngopharyngeal reflux, Harrell<sup>6</sup> defined four potential pH monitoring artifacts for pH drops less than 4 that may apply to distal pH studies: 1) meal periods; 2) liquid swallows outside of meals; 3) pH out of range (pH = 0 or pH > 8); and 4) short pH drop (lasting less than 5 seconds). In infants the ingestion of weaning foods with a pH less than 4, such as fruit juices, results in a pH drop.

Mothers often fail to report all the child's feeds, especially liquids taken between meals. To detect these artifacts, simultaneous proximal and distal pH monitoring is necessary.<sup>9</sup> In infants the pH of the esophagus varies between 4 and 7 and of the stomach between 1 and 7.<sup>10</sup> It would therefore be safe to assume that pH values below 1 or above 8 are artifacts. Whether pH drops less than 5 seconds are always artifacts was not investigated in children. A clear definition of pH monitoring artifacts for pH drops less than 4 is needed, as it may aid in the interpretation of pH studies.

Although this was a retrospective study, this study shows that mutual agreement in the interpretation of a pH study is

fair, with a kappa coefficient of 0.70. It can further improve with a standardized protocol which includes posttest calibration and assessment of the pH electrode position and a clear definition of pH monitoring artifacts for pH drops less than 4.

## References

1. Vandenplas Y, Blecker U, Heymans HS. [Gastroesophageal reflux in infants; recommendations for diagnosis and treatment](#). Ned Tijdschr Geneesk. 1995;139:366-70.
2. Rudolph CD, Mazur LJ, Liptak GS, Baker RD, Boyle JT, Colletti RB, et al.; North American Society for Pediatric Gastroenterology and Nutrition. [Guidelines for evaluation and treatment of gastroesophageal reflux in infants and children: recommendations of the North American Society for Pediatric Gastroenterology and Nutrition](#). J Pediatr Gastroenterol Nutr. 2001;32 Suppl 2:S1-31.
3. Vandenplas Y, Goyvaerts H, Helven R, Sacre L. [Gastroesophageal reflux, as measured by 24-hour pH monitoring, in 509 healthy infants screened for risk of sudden infant death syndrome](#). Pediatrics. 1991;88:834-40.
4. Working Group of the European Society of Pediatric Gastroenterology and Nutrition. [A standardized protocol for the methodology of esophageal pH monitoring and interpretation of the data for the diagnosis of gastroesophageal reflux](#). J Pediatr Gastroenterol Nutr. 1992;14:467-71.
5. Wise JL, Kammer PK, Murray JA. [Post-test calibration of single-use, antimony, 24-hour ambulatory esophageal pH probes is necessary](#). Dig Dis Sci. 2004;49:688-92.
6. Harrell SP, Koopman J, Woosley S, Wo JM. [Exclusion of pH artifacts is essential for hypopharyngeal pH monitoring](#). Laryngoscope. 2007;117:470-4.
7. Mahajan L, Wyllie R, Oliva L, Balsells F, Steffen R, Kay M. [Reproducibility of 24-hour intraesophageal pH monitoring in pediatric patients](#). Pediatrics. 1998;101:260-3.
8. Nielsen RG, Kruse-Andersen S, Husby S. [Low reproducibility of 2 x 24-hour continuous esophageal pH monitoring in infants and children: a limiting factor for interventional studies](#). Dig Dis Sci. 2003;48:1495-502.
9. Maldonado A, Diederich L, Castell DO, Gideon RM, Katz PO. [Laryngopharyngeal reflux identified using a new catheter design: defining normal values and excluding artifacts](#). Laryngoscope. 2003;113:349-55.
10. Omari TI, Davidson GP. [Multipoint measurement of intragastric pH in healthy preterm infants](#). Arch Dis Child Fetal Neonatal Ed. 2003;88:F517-20.

## Correspondence:

W. E. Tjon A. Ten  
 Department of Pediatrics, Maxima Medical Centre  
 Veldhoven, PO Box 7777  
 5500 MB - Veldhoven - The Netherlands  
 Tel.: +31 (40) 888.8270  
 Fax: +31 (40) 888.8273  
 E-mail: w.tjonaten@mmc.nl