

Special Report: Wireless Esophageal pH Monitoring*



Assessment
Program
Volume 21, No. 2
May 2006

Executive Summary

Background

Gastroesophageal reflux disease (GERD) is a disorder marked by heterogeneity in symptoms, severity, and available treatment options. Esophageal pH recording is a technique used to evaluate certain patients who may have unusual presentations or difficult to treat manifestations of GERD. Traditionally, this has been done using a pH probe placed in the esophagus and attached by wires to recording equipment, usually for a period of 24 hours. A wireless system is now available, in which the pH probe clipped to the esophagus transmits information to a pager-sized receiver worn by the patient. The wireless system may offer better patient comfort.

Evaluating the value of pH monitoring is difficult. First, there is no gold standard diagnostic test for GERD. The ability of pH monitoring to distinguish between clinically defined GERD subjects and normal controls is limited. Without a reference standard, standard measures of sensitivity and specificity cannot be calculated.

Solid evidence linking the use of any type of pH monitoring to improved patient outcomes is lacking. Current practice guidelines endorse the use of pH monitoring in selected patients, but technical reviews accompanying the guidelines acknowledge that there is modest evidence showing that pH monitoring improves patient outcomes.

Objective

This Special Report will provide information relevant to the evaluation of wireless pH monitoring. This Report will summarize information on:

- Diagnostic performance and clinical utility of traditional pH monitoring
- Feasibility of wireless pH monitoring
- Comparison of performance of traditional and wireless pH monitoring

Methods

MEDLINE was searched using the terms “wireless” and “pH monitoring” for the time period up to March 2006. Articles were identified that involved human studies of the wireless pH monitoring system for use in the diagnosis and management of patients with possible GERD. Additionally, article bibliographies and a bibliography updated to January 2006 supplied by the manufacturer of the wireless system were searched.

* This Special Report evaluates the available evidence on wireless esophageal pH monitoring. It does not address the question as to whether the TEC criteria are met, primarily because evidence supporting the clinical role of conventional esophageal pH monitoring is limited and because of the lack of a gold standard, it would be difficult to derive meaningful performance data for the use of wireless monitoring in a formal systematic review setting.

NOTICE OF PURPOSE: TEC Assessments are scientific opinions, provided solely for informational purposes. TEC Assessments should not be construed to suggest that the Blue Cross Blue Shield Association, Kaiser Permanente Medical Care Program or the TEC Program recommends, advocates, requires, encourages, or discourages any particular treatment, procedure, or service; any particular course of treatment, procedure, or service; or the payment or non-payment of the technology or technologies evaluated.



An Association
of Independent
Blue Cross and
Blue Shield Plans



Results

Several case series show that the device is successfully attached and produces pH readings for 24 to 48 hours in over 90% of attempts. Two studies compared wireless and wired monitoring in terms of patient comfort and found that in most measures, wireless monitoring is judged to be more comfortable. One study assesses test positivity in GERD subjects and normal controls and found that performance was similar to what has been found with wired monitoring. Although one study shows an 88% concordance in diagnosis of GERD when comparing wireless monitoring to simultaneous wired monitoring, it does not demonstrate that wireless monitoring is better than traditional monitoring in diagnosing or managing GERD. The study does not demonstrate that the wireless monitor is more accurate when the tests are discordant. This degree of concordance between the two methods was achieved only after adjusting test thresholds to maximize concordance. Such a post-hoc adjustment of course cannot be done with the device in use in clinical practice. More experimentation with the device on both GERD patients and normal controls will be needed to better establish appropriate test thresholds that optimize clinical utility of the test.

Studies purporting to show improved diagnostic capability of wireless monitoring when used for 48 hours versus 24 hours were fundamentally flawed because there was no gold standard for diagnosis and no control patients were enrolled in the studies.

Conclusions

Thus, the following conclusions appear to apply to wireless monitoring:

- The procedure is successfully performed and produces successful measurement of esophageal acid in a high proportion of cases. The relatively small number of cases reported does not allow for reliable estimation of rare but potentially serious adverse complications.
- Overall, it is more comfortable than traditional wired monitoring.
- Measurements correlate fairly closely to wired monitoring after adjusting test thresholds; however, because of the lack of an established gold standard, even when the two devices are discrepant, it cannot be determined which device is “correct.” Also, different studies have produced different cutoff values for a normal test. More data are needed to establish appropriate diagnostic thresholds.
- No studies establish that wireless monitoring is superior to wired monitoring in reaching a GERD diagnosis or optimally managing a patient. There is no evidence linking the use of the wireless device to an improvement in health outcome.

Contents

Objective	4	Review of Evidence	7
Background	4	Conclusions	11
Methods	7	References	12

Published in cooperation with Kaiser Foundation Health Plan and Southern California Permanente Medical Group.

TEC Staff Contributors

Lead Author—David Mark, M.D., M.P.H.; **TEC Executive Director**—Naomi Aronson, Ph.D.; **Managing Scientific Editor**—Kathleen M. Ziegler, Pharm.D.; **Research/Editorial Staff**—Claudia J. Bonnell, B.S.N., M.L.S.; Maxine A. Gere, M.S.

Blue Cross and Blue Shield Association Medical Advisory Panel

Allan M. Korn, M.D., F.A.C.P.—Chairman, *Senior Vice President, Clinical Affairs/Medical Director, Blue Cross and Blue Shield Association*; **Alan M. Garber, M.D., Ph.D.**—Scientific Advisor, *Staff Physician, U.S. Department of Veterans Affairs*; **Henry J. Kaiser, Jr., Professor, and Professor of Medicine, Economics, and Health Research and Policy, Stanford University**; **Steven N. Goodman, M.D., M.H.S., Ph.D.**—Scientific Advisor, *Associate Professor, Johns Hopkins School of Medicine, Department of Oncology, Division of Biostatistics (joint appointments in Epidemiology, Biostatistics, and Pediatrics)*—American Academy of Pediatrics Appointee. ■ **Panel Members** **Peter C. Albertsen, M.D.**, *Professor, Chief of Urology, and Residency Program Director, University of Connecticut Health Center*; **Edgar Black, M.D.**, *Vice President, Chief Medical Officer, Blue Cross Blue Shield of the Rochester Area*; **Sarah T. Corley, M.D.**, *Physician Consultant, NexGen Healthcare Information Systems, Inc.*—American College of Physicians Appointee; **Helen Darling, M.A.**, *President, National Business Group on Health*; **Josef E. Fischer, M.D., F.A.C.S.**, *Mallinckrodt Professor of Surgery, Harvard Medical School and Chair, Department of Surgery, Beth Israel Deaconess Medical Center*—American College of Surgeons Appointee; **Willard K. Harms, M.D., M.M.**, *Medical Director, Medical Policy and Adjudication, Blue Cross Blue Shield of Illinois*; **I. Craig Henderson, M.D.**, *Adjunct Professor of Medicine, University of California, San Francisco*; **Mark A. Hlatky, M.D.**, *Professor of Health Research and Policy and of Medicine (Cardiovascular Medicine), Stanford University School of Medicine*; **Bernard L. Lo, M.D.**, *Professor of Medicine and Director, Program in Medical Ethics, University of California, San Francisco*; **Barbara J. McNeil, M.D., Ph.D.**, *Ridley Watts Professor and Head of Health Care Policy, Harvard Medical School, Professor of Radiology, Brigham and Women's Hospital*; **Brent O'Connell, M.D., M.H.S.A.**, *Vice President and Medical Director, Pennsylvania Blue Shield/Highmark, Inc.*; **William R. Phillips, M.D., M.P.H.**, *Clinical Professor of Family Medicine, University of Washington*—American Academy of Family Physicians' Appointee; **Maren T. Scheuner, M.D., M.P.H.**, *Visiting Associate Professor, Department of Health Services, UCLA School of Public Health*—American College of Medical Genetics Appointee; **J. Sanford Schwartz, M.D.**, *Professor of Medicine, Department of Medicine, University of Pennsylvania School of Medicine and Professor, Health Care Systems, Health Management & Economics, The Wharton School*; **Earl P. Steinberg, M.D., M.P.P.**, *President, Resolution Health, Inc.*; **A. Eugene Washington, M.D., M.Sc.**, *Executive Vice Chancellor, University of California, San Francisco*; **Jed Weissberg, M.D.**, *Associate Executive Director for Quality and Performance Improvement, The Permanente Federation.*

CONFIDENTIAL: This document contains proprietary information that is intended solely for Blue Cross and Blue Shield Plans and other subscribers to the TEC Program. The contents of this document are not to be provided in any manner to any other parties without the express written consent of the Blue Cross and Blue Shield Association.

Objective

Gastroesophageal reflux disease (GERD) is a disorder marked by heterogeneity in symptoms, severity, and available treatment options. Esophageal pH recording is a technique used to evaluate certain patients who may have unusual presentations or difficult to treat manifestations of GERD. Traditionally, this has been done using a pH probe placed in the esophagus and attached by wires to recording equipment, usually for a period of 24 hours. A wireless system is now available, where the pH probe attached to the esophagus transmits information to a pager-sized receiver worn by the patient. The wireless system may offer better patient comfort. The test may offer better information regarding the diagnosis of GERD due to more natural food intake and physical behavior of the patient during the test and the feasibility of expanding the time interval of testing to 48 hours.

Wireless esophageal monitoring may be considered simply a substitute for traditional wired esophageal monitoring, in which case only evidence that it produces identical information to wired esophageal monitoring is of interest. However, comparing the performance of the two systems in terms of which test is “better” in terms of improving patient outcomes is problematic. First, there is no gold standard diagnostic test for GERD. As will be described in this Report, the ability of pH monitoring to distinguish between clinically defined GERD subjects and normal controls is limited. Without a reference standard, standard measures of sensitivity and specificity cannot be calculated.

The second problem in evaluating the performance of these tests is that solid evidence linking the use of any type of pH monitoring to improved patient outcomes is lacking. A diagnostic test can be useful, even without a solid reference standard, if it appropriately directs treatment such that patient outcomes are better. Current practice guidelines endorse the use of pH monitoring in selected patients, but technical reviews accompanying the guidelines acknowledge that there is modest evidence showing that pH monitoring improves patient outcomes. Thus, it would be difficult or impossible to show incremental improvements in patient outcomes with an alternate test.

This Special Report will provide information relevant to the evaluation of wireless

pH monitoring. This report will summarize information on:

- Diagnostic performance and clinical utility of traditional pH monitoring
- Feasibility of wireless pH monitoring
- Comparison of performance of traditional and wireless pH monitoring

Kahrilas and Quigley (1996) summarized the available literature on the first point in their technical review accompanying the 1996 American Gastroenterological Association (AGA) guideline on use of esophageal pH recording. Their review will be extensively cited when summarizing current knowledge about traditional pH monitoring. A review of GERD treatment by Ip et al. (2005) reviewed studies that examined the association of patient factors, including esophageal pH tests, with treatment success.

Background

Gastroesophageal Reflux Disease

Gastroesophageal reflux disease (GERD) is a common disorder defined as symptoms or mucosal damage produced by the abnormal reflux of gastric contents into the esophagus. Population studies have estimated that chronic daily GERD symptoms occur in 7–10% of the population (Castell 2001) and weekly symptoms occur in up to 20% (Arora and Castell 2001). GERD is more prevalent in the elderly population, with greater than half of patients reporting chronic symptoms (Arora and Castell 2001).

The pathophysiology of GERD involves excessive exposure of the esophagus to stomach acid. However, it is clear that more than just the quantity of acid exposure is involved. GERD may also be related to the sensitivity of the esophageal mucosa, mucosal resistance to inflammation, or other contents of reflux other than acid (Kahrilas and Quigley 1996).

The clinical severity of GERD encompasses a broad spectrum. Many persons will have mild or transient symptoms that do not require treatment or are self-treated with lifestyle modifications or over-the-counter (OTC) medications. A subset of patients will have more serious disease characterized by chronic acid reflux, chronic symptoms, and esophagitis. In patients with chronic symptoms, approximately 20% develop more serious complications

(Castell 2001). Complications include erosive esophagitis, esophageal stricture, dysphagia, Barrett's esophagus and esophageal carcinoma. Pulmonary complications that can result from aspiration of stomach acid into the lungs include asthma, pulmonary fibrosis, bronchitis and bronchiectasis. Otolaryngologic complications include hoarseness, cough and chronic sore throat (DiPalma 2001).

Most GERD is diagnosed through assessment of symptoms associated with GERD along with a trial of empiric therapy. Endoscopy is indicated for those suspected of having complications of GERD. A finding of esophagitis is diagnostic for GERD. The role of additional testing, using pH monitoring or esophageal manometry, is now reserved for more complicated patients who do not respond to therapy (medical or surgical), or as part of the preoperative evaluation or postoperative monitoring of surgical treatments.

Traditional pH Monitoring

Practice Guidelines on the Use of Esophageal pH Monitoring. In 1996, the AGA published guidelines and a technical report on esophageal pH monitoring (American Gastroenterological Association 1996). Testing was indicated for 1) preoperative evaluation of endoscopy-negative surgical candidates 2) postoperative evaluation of surgical patients with continued symptoms, and 3) endoscopy-negative patients refractory to treatment. Testing was *possibly* indicated for 1) chest pain syndrome patients, 2) patients with possible otolaryngologic manifestations of GERD refractory to treatment, and 3) patients with asthma possibly linked to GERD. Testing was *not indicated* for 1) diagnosis of esophagitis and 2) evaluation of "alkaline reflux." These guidelines were essentially reiterated in brief in updated guidelines for diagnosis and treatment of GERD published in 2005, describing the indication for pH testing as evaluation of symptomatic endoscopy-negative patients refractory to treatment (DeVault and Castell 2005). The evidence supporting this recommendation was classified as Level III, which means well-designed trials but no randomized trials.

Output and Interpretation of pH Monitoring.

There are innumerable minor variations in the technique of pH recording that may affect the results, but all devices produce data regarding the pH in the esophagus measured at frequent intervals (approximately every 6 seconds) over a 24-hour period of time. Thus, over 10,000 data

points are produced, which can be analyzed in a variety of ways. Johnson and DeMeester developed a scoring system based on 6 variables derived from the data. Of these variables, the consensus is that the percent time with pH less than 4 is the most useful discriminator between normal and pathological reflux.

Another more complex set of methods to interpret the data output from pH recording is to correlate symptoms of reflux to recorded episodes of abnormal pH. Because there is not a one-to-one correspondence between reflux and symptoms, statistical methods are required to quantify the association. The AGA technical report accompanying the guidelines discusses briefly the methodologic issues of this method, but concludes "despite the evolution in sophistication, none of the proposed schemes of reflux-symptom association have been prospectively validated against an independent criterion of diagnostic accuracy...." (Kahrilas and Quigley 1996).

Diagnostic Performance of pH Monitoring.

The lack of a true gold standard is problematic when evaluating this test in clinically relevant populations. However, the proportion that test positive among persons with known disease can give the most optimistic estimate of sensitivity, and the proportion that test negative among normal controls gives an optimistic estimate of specificity. The degree to which these estimates apply to the clinically relevant population of problematic GERD patients is unknown. When comparing patients with endoscopically confirmed GERD to normal controls, studies show 77–100% sensitivity and 85–100% specificity of esophageal pH recording using the parameter of percent time with pH less than 4 (Kahrilas and Quigley 1996).

When comparing clinically defined but endoscopically negative GERD patients to normal controls, the diagnostic performance is much worse, with sensitivities of 0–71% and specificities of 85–100% (Kahrilas and Quigley 1996). They state, "one has to conclude that ambulatory pH studies quantify esophageal acid exposure but that this has an imperfect correlation with reflux-related symptoms, esophageal sensitivity, or response to acid suppressive therapy." With respect to capability of pH monitoring to ascertain the role of reflux in extraesophageal conditions such as asthma and laryngitis, they state, "distal pH monitoring has no proven benefit in establishing causality

between GERD and asthma but that proximal pH monitoring is potentially useful. Presently, however, there are insufficient data to state this with any certainty.”

pH Monitoring and Improvement of Patient Outcomes

Monitoring of pH might be demonstrated to be useful if it stratifies patients into different management categories that result in improved patient outcomes. Given that there is no real gold standard for the diagnosis, studies that examine this issue might be able to establish the utility of pH monitoring. The utility of pH monitoring has been evaluated in the following clinical circumstances:

Prediction of Medication Requirements.

Kahrilas and Quigley (1996) cite studies showing that there was no correlation between pH study result and eventual treatment needs. The only predictive factors found useful were a symptom profile (in the case of mild disease) and an endoscopic examination (in the case of severe disease). “For patients between these extremes, no diagnostic test is superior to an empirical trial of pharmacological agents in devising an individualized therapeutic plan.” A more recent review of medical treatment for GERD by Ip et al. (2005) cites a study in which esophageal pH findings were not associated with failure of medical therapy.

Prediction of Benefit from Antireflux Surgery.

Kahrilas and Quigley (1996) note that pH monitoring has often been used as a means of patient selection. In general, patients with abnormal pH monitoring results have high success rates after surgery. However, there is a lack of data on the outcomes of patients excluded from surgery based on pH monitoring, and thus they conclude that although “there are data supporting the view that ... pH study is a useful selection criterion ... for antireflux surgery, there are no data on the more difficult question of whether or not the test identifies patients for whom surgery is the optimal therapy.” Ip et al. (2005) in their review of studies examining factors associated with outcomes from reflux surgery, found inconsistent findings. Two studies reported better results in patients with increased preoperative acid reflux, one study reported no difference, and another study showed worse results.

Management of Treatment Failures. Kahrilas and Quigley (1996) cite several studies showing an imperfect correlation between changes in pH monitoring result and changes in symptoms after treatment. Patients can have resolution of symptoms with persistently abnormal pH tests, or persistent symptoms with normal pH results. In summarizing these studies, they conclude, “Taken together, these data suggest that ambulatory pH monitoring can be useful in documenting antireflux treatment failures, but its precise sensitivity and specificity in this application are unclear.”

In sum, traditional pH monitoring has limited capability to discriminate between GERD and non-GERD patients, and there is limited evidence that it improves patient outcomes. This makes it problematic to evaluate an alternative such as wireless monitoring.

Wireless pH Monitoring

Chotiprashidi et al. (2005) in their technology status evaluation report on behalf of the American Society for Gastrointestinal Endoscopy briefly summarize pertinent facts regarding this method of pH monitoring. The pH monitor is an oblong plastic capsule which is affixed with a metal clip to the esophageal mucosa during upper endoscopy. The capsule can be placed either orally or nasally. Data on pH are transmitted to a pager-sized receiver worn by the patient. The data produced are largely similar to those produced by traditional pH monitoring and interpreted using similar parameters. The capsule is expected to detach from the esophageal mucosa by itself and to pass through the GI tract.

Important issues regarding its use are the success rates of use, comfort and tolerability of monitoring, and of course, the utility of the results. Specific problems that have been reported in case reports have been chest pain, failure of data retrieval, premature detachment (before 24 or 48 hours), and failure to spontaneously detach within 14 days (Chotiprashidi et al. 2005).

FDA Status. The Bravo pH Monitoring System received U.S. Food and Drug Administration (FDA) 510(k) marketing clearance on September 29, 2000 (U.S. Food and Drug Administration 2000). The device is “intended

to be used for gastroesophageal pH measurement and monitoring of gastric reflux. The pH probe can be delivered and placed endoscopically or with standard manometric procedures. The pH Software Analysis Program is intended to record, store, view and analyze gastroesophageal pH data.”

Methods

Search Methods

MEDLINE was searched using the terms “wireless” and “pH monitoring” for the time period up to March 2006. Articles were identified that involved human studies of the wireless pH monitoring system for use in the diagnosis and management of patients with possible GERD. Additionally, article bibliographies and a bibliography updated to January 2006 supplied by the manufacturer of the wireless system were searched.

Article Selection

After review of the articles, we found that the information coalesced into certain categories that may provide evidence on the utility of wireless monitoring. We classified findings in the articles into these categories, and selected articles for presentation.

Technical Feasibility and Success. A representative sample of studies which presented quantitative data on the success rate of monitor placement and successful monitoring on a consecutive series of patients were selected.

Comfort and Satisfaction with Wireless Monitoring. Only studies comparing traditional and wireless monitoring were selected. Case series only evaluating wireless monitoring were not reviewed.

Sensitivity and Specificity of Wireless Monitoring. Studies which calculated sensitivity and specificity using confirmed GERD patients and normal controls were selected.

Comparison of Traditional and Wireless Monitoring. Studies which evaluated the concordance of traditional and wireless monitoring in the same patient were selected.

Utility of 48-Hour Versus 24-Hour pH Monitoring. These studies were selected for review because of the claim that 48 hours is superior to 24 hours of monitoring.

Evidence Linking the Use of Wireless Monitoring to Improvement in Health Outcomes. Although this type of information would be valuable in establishing the utility of wireless pH monitoring in patients with GERD, no studies directly linking the use of the device to an improvement in health outcome were found.

We did not select case reports, abstracts, or studies reporting unusual applications or rare populations for whom wireless monitoring might be used.

Medical Advisory Panel Review

The Blue Cross and Blue Shield Association Medical Advisory Panel (MAP) reviewed this Special Report on February 23, 2006. To maintain the timeliness of scientific information in the Report, literature search updates were conducted subsequent to the Panel’s review (see “Search Methods”). If the search updates identified any additional studies that met the criteria for detailed review, the results of these studies were included in the text where appropriate. There were no studies that would change the Report’s conclusions.

Review of Evidence

Studies Describing the Technical Feasibility of Wireless Monitoring

Several studies only include information relating to technical aspects of inserting the device and achieving successful monitoring over a period of time. Selected results of these studies are summarized in Table 1. Also included are studies that assessed other aspects of the wireless monitor, but included such results in the paper. Technical failure can occur when attempt to attach the device to esophagus fails. Subsequently, monitoring can fail if the device detaches from the esophagus prematurely or if there are failures in transmitting the signal to the receiver. Most studies show a high rate (>95%) of attempts result in successful placement of the pH monitor in the esophagus, and a high rate of success in achieving 48 hours of monitoring. Although these studies provide information on overall success and failure rates, the number of cases reported is probably too small to understand the risk of serious, probably rare adverse consequences that may occur when foreign bodies are placed in the human body.

Studies Evaluating Patient Comfort and Tolerability of Wireless pH Monitoring

Several case series studies have reported rates of symptoms like chest discomfort and difficulty swallowing. It is difficult to evaluate such information because the methods of measuring and evaluating these symptoms are not reported in the studies. Two studies were found that provide more rigorous data on the issue of patient comfort with a formal comparison to traditional pH monitoring (Table 2).

Pandolfino et al. (2005) surveyed 29 patients who had had wireless monitoring and 30 age- and sex-matched patients who underwent traditional monitoring regarding comfort,

satisfaction and interference with living activities. About half of the patients who had wireless monitoring were asymptomatic volunteer controls, whereas all 30 patients with wired monitoring were symptomatic patients. Thus, the groups were not really well matched. On most measures, wireless monitoring had fewer adverse effects than traditional monitoring. However, patients with wireless monitoring reported higher frequency of esophageal discomfort (34% vs. 17%, $p < 0.05$).

Wong et al. (2005) enrolled 50 patients in a randomized trial comparing patient experiences with the wireless and the traditional monitor. Adverse effects, daily activity, sleep, overall

Table 1. Studies Reporting Placement and Monitoring Success Rates of Wireless Monitoring

Author	n	Successful Placement	Successful 24-hour Monitor	Successful 48-hour Monitor	Overall Success
Belafsky 2004	46	96%	—	—	85%
Tu 2004	30	100%	100%	97%	
Pandolfino 2003	85	100%	96%	89%	
Ward 2004	60	98%		97%	
Remes-Troche 2005	84	95%	92%	92%	
Wenner 2005	50	100%	100%	96%	

Table 2. Comparison of Wireless and Traditional Monitoring Regarding Comfort and Satisfaction During Test, Comparative Studies

Study	Outcome (Question Format)	Wireless	Traditional	p-value
Observational Study				
Pandolfino 2003	overall satisfaction (0–5, 0=best)	0.8	1.9	0.001
	% throat discomfort (y/n)	73	14	0.001
	% esophageal discomfort	34	17	0.05
	% diet affected	3	47	0.001
	% activity affected	0	60	0.001
	% disrupted sleep	9	30	0.07
Randomized Study				
Wong et al. 2005	% nose pain	32	60	0.05
	% runny nose	52	96	0.001
	% throat discomfort	48	92	0.001
	% headache	20	56	0.009
	% chest discomfort	36	8	0.037
	% work during test	58	11	0.049
	Interference daily life (0–4, 0=none)	0.32	1.3	0.001
	Overall experience (ordinal scale)	better	24	0.004
	Overall satisfaction (ordinal scale)	better		0.023
	SF-36 overall score	18		0.191

satisfaction, and quality of life as measured by SF-36 were compared between the groups. On most measures of adverse symptoms, the patients with the traditional monitor reported more problems. However, chest discomfort was greater with the wireless monitor. Patients with wireless monitors reported less interference with activities of daily life, a better overall experience, and greater overall satisfaction.

In sum, except for a consistently higher proportion of subjects reporting chest discomfort, wireless monitoring shows improved comfort, fewer reported effects on activities of daily living, and greater satisfaction when compared to traditional monitoring.

Studies Evaluating Sensitivity and Specificity of pH Monitoring

One study by Pandolfino et al. (2003) evaluated wireless pH monitoring in both confirmed GERD subjects (both endoscopy positive and endoscopy negative) and normal controls, and calculated sensitivity and specificity values (Table 3). Note that this study does not enroll patients for whom pH monitoring is clinically indicated, and that both groups are likely to be convenience samples. However, the design is similar to the many studies cited by Kahrilas and Quigley (1996) that attempt to calculate diagnostic performance of traditional pH monitoring.

Forty-four healthy subjects and 41 GERD patients were enrolled. Percent time that pH was less than 4 was used as the measure to assess GERD, and the 95th percentile of that measure in normal controls was used as the diagnostic threshold (5.3%). Sensitivity and specificity were calculated based on the whole

2-day period of observation, the first day of observation, or the worst of 2 days of observation. Given that the cutoff was calculated based on 2 days of observation of the control group, the specificity of the test at 2 days observation is “fixed” at 95%.

The sensitivities and specificities calculated are typical of the numbers found in similar studies of traditional pH monitoring cited by Kahrilas and Quigley (1996). A rigorous quantitative comparison with other studies is not warranted, given unknown differences between these patients and other studies. Of note is the extremely poor sensitivity of 35.7% for endoscopy-negative subjects, when using data from the entire 2-day monitoring period. When other definitions of positivity are used, such as the worst day value, sensitivity always increases and specificity always decreases.

Studies Comparing Simultaneous Wireless and Traditional pH Monitoring

There are 2 studies in which subjects had both wireless and traditional pH monitors inserted and results compared. Such experiments could assess the degree of concordance between the 2 techniques as a measure of the comparability. If the 2 tests are highly concordant, then the tests are measuring the same thing, and are substitutes for each other. If the 2 tests are discordant, then without an independent reference standard or a measure of clinical outcome, then it is impossible to know which test is superior.

A study by Pandolfino et al. (2005) studied 25 normal controls with both traditional and wireless pH monitors. Surprisingly, there is no data analysis presented that shows the correlation

Table 3. Sensitivity and Specificity of Wireless Monitoring, from Pandolfino et al. (2003)

Groups Compared	Observation Period	Sensitivity (%)	Specificity (%)
	for Calculation of Test Characteristics		
Entire GERD vs. control (n's, 37, 39)	2 days	64.9	94.8
	first day	67.5	84.5
	worst day	83.8	84.5
endoscopy (+) GERD vs. control (n's, 23, 39)	2 days	78.3	94.8
	first day	73.9	84.5
	worst day	100	84.5
endoscopy (-) GERD vs. control (n's, 14, 39)	2 days	35.7	94.8
	first day	57.1	84.5
	worst day	57.1	84.5

or concordance of the 2 methods. However, they did find differences between the 2 systems regarding the calibration of pH values, which resulted in traditional monitoring showing higher mean acid exposure times than wireless monitoring. When the devices were appropriately adjusted using samples of known pH, the two devices produced similar mean pH values. However, they did not analyze concordance of individual measurements.

The other study by Bruley des Varannes et al. (2005) studied 40 patients with GERD symptoms with both types of monitors. They found a highly significant correlation ($r=0.87$) between the two systems using the first day's data. They also found that the traditional monitor showed higher acid exposure times than wireless monitoring. After lowering the test threshold of wireless monitoring to 2.9% time with pH <5 in order to maximize the concordance of diagnosis between the 2 devices, the concordance in the diagnosis of GERD was 88%, corresponding to a kappa of 0.76. Out of 33 patients with successful monitoring over the first 24 hours, 11 had GERD with both tests, 3 were positive only with traditional monitor, 2 were positive only with wireless monitoring, and 17 were negative with both tests. Other analyses in this study included comparison of first day monitoring with second day monitoring, and comparison of symptom association probability between the 2 techniques, with no significant findings noted. Thus, the study shows reasonably good concordance between the 2 methods of monitoring. Note that this study design maximizes the concordance between the 2 tests by using simultaneous measurements. pH monitoring is known to have only fair reproducibility when using identical techniques in the same individual on different occasions.

The finding of a systematic difference in pH values between wireless monitoring and traditional monitoring raises an important issue of what should be considered the threshold value for a "normal" test. Different investigators have used different methods to define this value, and it is critical to understand how this value might affect the utility of the test. Pandolfino et al. (2003) has defined the cutoff at 5.3% time with pH less than 4 based on the 95% percentile of an asymptomatic control sample ($n=59$), thus "fixing" specificity to be 95%. Bruley des Varannes et al. (2005) as described above set the cutoff at 2.9%, in order to achieve the

highest diagnostic concordance with the cutoff value of simultaneously measured traditional pH monitoring. There is one other study by Wenner et al. (2005) that replicates the method of Pandolfino and co-workers in establishing a cutoff value. In this sample of 50 control patients, the 95th percentile of normal resulted in a cutoff value of 4.4%. Thus, different studies using either different samples or different methodology have established different cutoff values for a normal pH monitoring test result. Different cutoffs would affect the sensitivity and specificity of the test, and potentially alter the interpretation of the test.

Studies Comparing the Yield of 24- Versus 48-Hour Monitoring

A proposed advantage of wireless monitoring is that 48-hour monitoring is feasible due to patient tolerability and comfort. Forty-eight hours of monitoring is purported to provide improved diagnostic accuracy for GERD. Two studies (Prakash and Clouse 2005 and Tseng et al. 2005) report case series of patients with possible GERD undergoing wireless monitoring for 48 hours (Table 4). In lieu of an independent 24-hour monitoring period, the first 24 hours was used as a comparison period. Both studies report increased numbers of positive tests when using various methods of incorporating 48 hours of data.

The increase in positive yield from 48 hours is interpreted by the authors as an improvement in test performance. However, this argument is flawed. An increase in positive tests is a foregone conclusion when using a test with limited reproducibility and the criteria for a positive test is a positive test on either day. The study by Tseng et al. assumes that all patients actually have GERD, and that the increase in yield reflects all true-positive patients. The study has no control group to examine any loss of specificity due to using 48 hours of data. The study by Prakash et al. incorrectly uses the 48-hour data as a gold standard to calculate sensitivities and specificities for 24-hour monitoring. The sensitivities and specificities from that article are not presented here because 48-hour monitoring is not a "gold standard." Thus, both studies ignore the possibility of false-positive tests, and thus, do not provide meaningful information on whether 48-hour monitoring better discriminates between subjects with and without GERD.

Table 4. Results of Studies Comparing 24 versus 48 Hours of Monitoring

Study	n	Measure of GERD	% Positive		
			Test from 24-hour Data	Method of Using 48-hour Data	
Prakash and Clouse (2005)					
All subjects	121	% time pH>4, cutoff 5.33%	50%	worst of 2 days	62%
Typical symptoms	67	% time pH>4, cutoff 5.33%	65%	worst of 2 days	80%
Atypical symptoms	54	% time pH>4, cutoff 5.33%	35%	worst of 2 days	50%
Tseng et al. (2005)					
	190	Johnson-DeMeester >14.7	49%	48 hours of data	61%
	190	% time pH>4, cutoff 5.33%	44%	worst of 2 days	54%

Conclusions

Although one study shows an 88% concordance in diagnosis of GERD when comparing wireless monitoring to simultaneous wired monitoring, there is little evidence that wireless monitoring is better than traditional monitoring in diagnosing or managing GERD. The concordance between the two methods was achieved only after adjusting an apparent calibration problem between the two types of devices. Such a post-hoc adjustment, of course, cannot be done with the device in use in clinical practice. More experimentation with the device on both GERD patients and normal controls will be needed to better establish appropriate test thresholds that optimize clinical utility of the test.

Studies purporting to show improved diagnostic capability of wireless monitoring when used for 48 hours are fundamentally flawed. No control patients were enrolled in the studies, and the investigators assumed that there were no false positives with wireless monitoring. Proving that one diagnostic test is better than another is problematic in this case because of lack of a that gold standard. Another method to show a diagnostic test is superior to another is to demonstrate that it better stratifies patients into different treatments, producing overall better patient outcomes. No such studies exist for wireless monitoring, or even for wired monitoring. Lack of strong evidence showing that

wired monitoring is effective, even though it is a generally accepted diagnostic test for certain patients, makes it difficult to evaluate an alternative such as wireless monitoring.

Thus, the following conclusions appear to apply to wireless monitoring:

- The procedure is successfully performed and produces successful measurement of esophageal acid in a high proportion of cases. The relatively small number of cases reported does not allow for reliable estimation of rare but potentially serious adverse complications.
- Overall, it is more comfortable than traditional wired monitoring.
- Measurements correlate fairly closely to wired monitoring after adjusting for calibration measures; however, because of the lack of an established gold standard, even when the two devices are discrepant, it cannot be determined which device is “correct.” Also, different studies have produced different cutoff values for a normal test. More data are needed to establish appropriate diagnostic thresholds.
- No studies establish that wireless monitoring is superior to wired monitoring in reaching a GERD diagnosis or optimally managing a patient. There is no evidence linking the use of the wireless device to an improvement in health outcome.

NOTICE OF PURPOSE: TEC Assessments are scientific opinions, provided solely for informational purposes. TEC Assessments should not be construed to suggest that the Blue Cross Blue Shield Association, Kaiser Permanente Medical Care Program or the TEC Program recommends, advocates, requires, encourages, or discourages any particular treatment, procedure, or service; any particular course of treatment, procedure, or service; or the payment or non-payment of the technology or technologies evaluated.

CONFIDENTIAL: This document contains proprietary information that is intended solely for Blue Cross and Blue Shield Plans and other subscribers to the TEC Program. The contents of this document are not to be provided in any manner to any other parties without the express written consent of the Blue Cross and Blue Shield Association.

References

- American Gastroenterological Association. (1996).** American Gastroenterological Association medical position statement: guidelines on the use of esophageal pH recording. *Gastroenterology*, 110:1981.
- Arora AS, Castell DO. (2001).** Medical therapy for gastroesophageal reflux disease. *Mayo Clin Proc*. 76:102-106.
- Belafsky PC, Allen K, Castro-Del Rosario L et al. (2004).** Wireless pH testing as an adjunct to unsedated transnasal esophagoscopy: the safety and efficacy of transnasal telemetry capsule placement. *Otolaryngol Head Neck Surg*, 131:26-8.
- Bruley des Varannes S, Mion F, Ducrotte P et al. (2005).** Simultaneous recordings of oesophageal acid exposure with conventional pH monitoring and a wireless system (Bravo). *Gut*, 54:1682-6.
- Castell DO. (2001).** Medical, surgical and endoscopic treatment of gastroesophageal reflux disease and Barrett's esophagus. *J Clin Gastroenterol*, 35:262-266.
- Chotiprashidi P, Liu J, Carpenter S et al. for the Technology Assessment Committee, American Society for Gastrointestinal Endoscopy. (2005).** ASGE Technology Status Evaluation Report: wireless esophageal pH monitoring system. *Gastrointest Endosc*, 62:485-7.
- DeVault KR, Castell DO for the American College of Gastroenterology. (2005).** Updated guidelines for the diagnosis and treatment of gastroesophageal reflux disease. *Am J Gastroenterol*, 100:190-200.
- DiPalma JA. (2001).** Management of severe gastroesophageal reflux disease. *J Clin Gastroenterol*, 32:19-26.
- Ip S, Bonis P, Tatsioni A et al. (2005).** Comparative effectiveness of management strategies for gastroesophageal reflux disease. Evidence Report/Technology Assessment No. 1. Rockville, MD: Agency for Healthcare Research and Quality. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.
- Kahrilas PJ, Quigley EM. (1996).** Clinical esophageal pH recording: a technical review for practice guideline development. *Gastroenterology*, 110:1982-96.
- Pandolfino JE, Richter JE, Ours T et al. (2005).** Ambulatory esophageal pH monitoring using a wireless system. *Am J Gastroenterol*, 98:740-9.
- Pandolfino JE, Schreiner MA, Lee TJ et al. (2005).** Comparison of the Bravo wireless and Digitrapper catheter-based pH monitoring systems for measuring esophageal acid exposure. *Am J Gastroenterol*, 100:1466-76.
- Prakash C, Clouse RE. (2005).** Value of extended recording time with wireless pH monitoring in evaluating gastroesophageal reflux disease. *Clin Gastroenterol Hepatol*, 5:529-54.
- Remes-Troche JM, Ibarra-Palomino J, Carmona-Sanchez RI et al. (2005).** Performance, tolerability, and symptoms related to prolonged pH monitoring using the Bravo system in Mexico. *Am J Gastroenterol*, 100:2582-6.
- Tseng D, Rizvi AZ, Fennerty MB et al. (2005).** Forty-eight-hour pH monitoring increases sensitivity in detecting abnormal esophageal acid exposure. *J Gastrointest Surg*, 9:1045-51.
- Tu CH, Lee YC, Wang HP et al. (2004).** Ambulatory esophageal pH monitoring by using a wireless system: a pilot study in Taiwan. *Hepatogastroenterology*, 51:1586-9.
- U.S. Food and Drug Administration. (2000).** 510(k) summary for the Bravo pH Monitoring System™. Available online at <http://www.fda.gov/cdrh/pdf/k002028.pdf>. Last accessed February 2006.
- Ward EM, DeVault KR, Bouras EP et al. (2004).** Successful oesophageal pH monitoring with a catheter-free system. *Aliment Pharmacol Ther*, 19:449-54.
- Wenner J, Johnsson F, Johansson J et al. (2005).** Wireless oesophageal pH monitoring: feasibility, safety and normal values in healthy subjects. *Scand J Gastroenterol*, 40:768-74.
- Wong WM, Bautista J, Dekel R et al. (2005).** Feasibility and tolerability of transnasal/per-oral placement of the wireless pH capsule vs. traditional 24-h oesophageal pH monitoring--a randomized trial. *Aliment Pharmacol Ther*, 21:155-65.



**Technology
Evaluation
Center**

**Blue Cross and
Blue Shield Association**
225 North Michigan Avenue
Chicago, Illinois 60601-7680
www.bcbs.com/tec