Innovative Non- or Minimally-Invasive Technologies for Monitoring Health and Nutritional Status in Mothers and Young Children

13C Breath Tests: Visions and Realities

Peter D. Klein

Research and Development, Meretek Diagnostics, Inc. and Departments of Pediatrics and Medicine, Baylor College of Medicine, Houston, TX 77030

ABSTRACT Breath tests have been used in research laboratories for over 25 y. Originally, the tests were based on the use of 14C, rather than on the nonradioactive isotope, 13C. When 13C became widely available at a reasonable cost, research groups in the United States and Europe developed methodologies to measure 13C abundance in samples of CO2. The tests used a variety of substrates and measured pancreatic function, fat absorption, bacterial overgrowth and \( P_{450} \) mixed-function oxidase. Thus far, the only test to be approved by the Food and Drug Administration is the 13C-urea breath test. This manuscript describes the process by which approval is gained, and indicates the steps necessary for other tests to receive Food and Drug Administration approval. J. Nutr. 131: 1637S-1642S, 2001.

KEY WORDS: \( \bullet \) H. pylori \( \bullet \) carbon-13 \( \bullet \) \( 13\mathrm{CO}_2 \) \( \bullet \) diagnostic test \( \bullet \) noninvasive

The search for noninvasive diagnostic and assessment methodology in nutrition is driven by three considerations: 1) target populations including infants, children, and pregnant and lactating women; 2) normal subjects; and 3) methods applicable to field use and epidemiological surveys. In other words, the methods should be safe, simple and effective.

Since the advent of tracer methodology, initially using 14C- and later, 13C-labeled isotopic compounds, interest in the use of the compounds in breath tests has been ongoing. The feasibility of administering a labeled compound and obtaining metabolic or diagnostic information from its metabolism and conversion to CO2 has been perennially attractive.

The underlying concept is simple: 13C is introduced into one or more functional groups in a substrate. The functional groups are linked to the rest of the molecule through bonds that are cleaved by specific enzymes. Once cleavage occurs, the functional group is further oxidized until CO2 is produced and excreted in breath. The appearance of excess 13CO2 in respiratory CO2 provides three types of information: 1) it indicates the presence of enzymatic activity, 2) the rate and extent of label appearance over time can often be correlated with the level of enzyme present in the whole body, and 3) it may reflect the rate of a physiological process or indicate the presence of a foreign, e.g., bacterial, enzyme. Examples of substrates that have been proposed and the function to be monitored are shown in Table 1.

The simplicity of the breath test method is seductive: if one could distinguish between two groups or two populations on the basis of a difference in the recovery of labeled CO2, valuable diagnostic information should have been generated. Unfortunately, this belief is often unsupported by external comparisons. A putative breath test, to have utility, must be compared with an established or predicate method of diagnosis, and the efficacy of the breath test must be established.

Establishing breath test efficacy

The efficacy of the test versus the predicate device is determined by simultaneous application of both measures in two groups: affected and nonaffected subjects. Preferably, these two groups are equal in size. If the predicate method involves an invasive procedure (e.g., liver biopsy or endoscopy), there may be recruitment problems, especially if there are no alternative indications for using the invasive procedure on otherwise asymptomatic subjects. Good statistical design considerations point to a study size of at least 60 individuals in each category or a total of 120 to 150 comparisons to the standard.

Four outcomes are possible from these comparisons: the breath test may correctly identify the true-positive (TP) and true-negative (TN) individuals, e.g., those in whom the condition is present and those in which it is absent. In addition,
the breath test may falsely identify individuals as false-positive (FP) or as false-negative (FN). From the outcomes from a specific clinical trial, it is possible to calculate the following parameters:

- Sensitivity = TP/(TP + FN).
- Specificity = TN/(TN + FP).
- Positive predictive value = TP/(TP + FP).
- Negative predictive value = TN/(TN + FN).
- Accuracy = (TP + TN)/total.

Each of these measures has an associated uncertainty, which is expressed as a range above and below the calculated value. For example, a study with a small number of unaffected subjects may have a specificity value of 0.95, but a confidence value. For example, a study with a small number of unaffected subjects may have a specificity value of 0.95, but a confidence range from 0.50 to 0.99.

Another means of assessing the data is to generate a receiver operator characteristic curve in which the sensitivity is plotted against (1-specificity). In such plots, a rectilinear curve rising steeply to a maximum and maintaining a plateau is expected. For example, a study with a small number of unaffected subjects may have a specificity value of 0.95, but a confidence range from 0.50 to 0.99.

### Are there effective breath tests?

The establishment of the efficacy of a breath test marks the boundary between its use in basic research and its potential use in patient diagnosis and management. To date, only the 13C-breath test for Helicobacter pylori has actually made such a transition by comparison with endoscopy, biopsy, culture and histologic examination. Tests using some of the other substrates can also be compared with predicate methods: for example, gastric emptying breath tests can be compared with the use of 99mTc-labeled meals and scintigraphic observation of stomach emptying, oxidation of substrates impaired by inborn errors of metabolism can be ranked against genetic phenotyping, and fat absorption breath tests can be compared with fecal fat measurements. In early studies on aminopyrine, work with animal models showed that removal of 70% of the liver reduced the response in breath by a corresponding amount, whereas inducing liver hypertrophy with phenobarbital increased the response in proportion to the increase in liver mass. Direct manipulation of the active hepatocyte mass is not feasible in humans; thus, a direct validation has not been possible.

Nonetheless, these substrates continue to be used and are the subject of many abstracts and numerous articles. This circumstance is possible because the research protocols under which they are used require only local institutional review and minimal precautions in substrate preparation and administration. The execution of the protocols is usually funded by research grants, and the outcome is of intellectual interest to the investigator. The benefit to the subject, if any, is coincidental.

When the focus of the test shifts to the subject or patient, an entirely different paradigm emerges. When the test contributes materially to the diagnosis or management of a patient, the costs of administering the test are more likely to be covered by health insurance plans or Medicare. To qualify, the test must have an assigned a cost reimbursement code from the American Medical Association and the Health Care Finance Administration. In turn, the test must have a demonstrated safety and efficacy as represented by a Food and Drug Administration (FDA)-approved new drug application and/or 510 K application. This application may include a Drug Master File detailing the preparation and packaging of the substrate.

### Barriers to the introduction of breath tests in clinical nutrition

Until the mid-1980s, 13C-breath tests were used exclusively in research. This circumstance was dictated by two factors: first, the lack of clinical access to 13CO2/12CO2 isotope ratio measurements, and second, the lack of reimbursement for test use. The introduction of the EUROPA company’s Automatic Breath Carbon Analyzer (Crewe, Cheshire, UK), a gas-isotope-ratio mass spectrometer, greatly simplified and broadened the isotope ratio measurement process. However, it was not until 1986 that breath tests began to play a significant role in the diagnosis and treatment of patients.

The discovery in 1983 by Warren (59) of a “campylobacter-like organism” in biopsies of gastric and duodenal ulcers, and the championship of an infectious basis of ulcers by Marshall et al. (60), focused interest on the organism now designated H. pylori. This spiral organism inhabiting the mucus layer in the crypts of the gastric mucosa induces an inflammatory response with leukocyte invasion of the mucosal cells. The gastritis, which is always present in H. pylori infection, gives rise to vacuolization and, in susceptible individuals, progresses to ulcer formation. Alternatively, the mucosal inflammation can progress to atrophic gastritis and eventually to gastric carcinoma.

In a 1994 Consensus Statement, the National Institutes of Health declared ulcers to be an infectious process associated with H. pylori and the World Health Organization has designated H. pylori as a class I carcinogen (61). H. pylori is characterized by a high content of urease, which serves to protect it against the low pH of the stomach. Urea that passes from the blood stream through the tight junctions in the crypt cells is hydrolyzed to form CO2 and ammonia, and the ammoniacal plume surrounding the H. pylori raises the ambient pH to 5.

### TABLE 1

<table>
<thead>
<tr>
<th>Organ</th>
<th>Function</th>
<th>Substrate</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver</td>
<td>P450 mixed-function oxidase</td>
<td>Aminopyrine</td>
<td>1–14</td>
</tr>
<tr>
<td>Liver</td>
<td>P445 mixed-function oxidase</td>
<td>Caffeine</td>
<td>15–26</td>
</tr>
<tr>
<td>Liver</td>
<td>Mitochondrial function</td>
<td>Ketaisocaproic acid</td>
<td>27</td>
</tr>
<tr>
<td>Liver</td>
<td>Bile salt production, oxidation</td>
<td>Palmitic acid</td>
<td>28–31</td>
</tr>
<tr>
<td>Pancreas</td>
<td>Lipid, fat absorption</td>
<td>Triglycerides</td>
<td>32–42</td>
</tr>
<tr>
<td>Stomach</td>
<td>Gastric emptying</td>
<td>Acetate, octanolic acid</td>
<td>43–51</td>
</tr>
<tr>
<td>Intestine</td>
<td>Bacterial overgrowth</td>
<td>Cholylglycine, xylose</td>
<td>52–54</td>
</tr>
<tr>
<td>Intestine</td>
<td>Transit time</td>
<td>Lactosylureide</td>
<td>55</td>
</tr>
<tr>
<td>Inborn error</td>
<td>Establish genotype</td>
<td>Phenylalanine, leucine, galactose</td>
<td>56–58</td>
</tr>
</tbody>
</table>
The first use of $^{13}$C-urea in a breath test to detect the presence of H. pylori was described by Graham et al. in 1987 (62) and was widely reproduced by others in subsequent publications (63–94). In this test, a baseline breath sample is collected before the subject consumes a 1000-kJ meal to inhibit gastric emptying. A solution containing 125 mg of $^{13}$C urea is then consumed and one or more postdose breath samples are collected. Today, one sample is collected 30 min after substrate ingestion. If H. pylori are present in the stomach, the organisms will hydrolyze the urea with consequent liberation of $^{13}$CO$_2$ that will be detected in expired air as an increase over the baseline abundance. A change as small as 2.4‰, or ~26 parts per million of $^{13}$CO$_2$, is evidence of active H. pylori infection. The test is easy to perform and because use of the test reduced endoscopy costs in clinical trials, a commercial breath test diagnostic kit for the diagnosis of H. pylori appeared on the market in January 1997.

By this time, a new environment for $^{13}$C-breath testing in the United States had been developed and tested. Kits to administer the test could be shipped to the physician from a central warehouse. The kits included simplified directions for performance of the test in the physician’s office, the instructions for the return of the breath samples by overnight express mail to the analysis facility and the prompt transmission of test results by fax within 24–48 h. This ensemble of components eliminated the need for local isotope ratio measurements and gave all physicians access to the test.

The use of the $^{13}$C-urea breath test in Europe has relied much more heavily on local measurements, chiefly because of the added revenue received by the physician for performing the analysis in his office. In addition to the EUROPA-type of gas-isotope-ratio mass spectrometer, a variety of infrared devices have been introduced (95,96). These instruments have a lower analytical precision than mass spectrometry, but in most instances have the advantages of lower cost, simpler operation and immediacy of results. Their use creates a gray zone in which the lowest levels of H. pylori infection cannot be resolved from results of uninfected subjects. The role that this technology will play in other breath test applications remains to be demonstrated. Nevertheless, in Europe as well as in the United States, the arrival of the $^{13}$C-urea breath test has introduced a whole new generation of physicians to the concepts and benefits of this noninvasive diagnostic procedure.

**How do $^{13}$C breath tests progress from research probe to clinical tool?**

The $^{13}$C-urea breath test provides a paradigm or map for the progressive transformation of a breath test from a research probe to an established clinical test. Four different areas make up this map: 1) establish the medical efficacy of the test, 2) obtain regulatory approval of the test, 3) assess how the test will affect the cost of health care delivery, and 4) provide the economic resources required to market the test.

**Establishment of medical efficacy.** The first and most serious resource hurdle appears at this stage of a breath test’s development. Clinical trials are both complex and expensive to organize and execute. To provide meaningful results, the trial must be conducted in accordance with the ground rules of good clinical practice. These rules govern the manner in which records of patient selection, assignment, testing and comparisons are maintained. Personnel with the specialized skills to accomplish these tasks are usually available only to pharmaceutical companies or to academic units that specialize in drug evaluation studies. Moreover, experience has shown that without adequate financial support for their execution, it is extremely difficult to conduct clinical trials on a collaborative basis that meet rigorously defined standards. Ultimately, the quality of the trials will be determined and/or limited by the funds available for their execution. One figure of merit is that real costs per subject are seldom < $1000 and can easily reach $5000. For even an efficiently organized trial, the cost is likely to be $500,000 or more. Under certain fortunate circumstances, it may be possible to find a planned or ongoing clinical drug trial in which the predicate device is being used. Addition of the breath test as a piggyback procedure may entail minimal additional costs while deriving the benefits of the investments supporting the main protocol. Under the most sanguine circumstances, the trial sponsor may anticipate enough benefit from the breath test results to subsidize the test costs.

**Regulatory approval of the test.** In the United States, approval of a noninvasive device, such as a breath test, hinges primarily on its clinical efficacy and its safety. However, in addition, the substrate must be produced in accordance with chemical, manufacturing and control processes that follow good manufacturing practices. This means that the manufacturer must have demonstrated experience in the synthesis or preparation of the substrate and must have in place the quality control procedures that identify the source and batch number of each ingredient used in substrate production. The manufacturer is required to prepare a certificate of analysis, which documents the chemical and isotopic purity of the product using previously validated analytical methods, and he is required to document the stability of the bulk substance under conventional and accelerated aging conditions.

The bulk substrate product must then undergo confirmatory analysis by an outside reference laboratory before it can be received by the facility in which doses are packaged. When the bulk material has been dispensed into individual units, regression time and accelerated stability studies of the packaged doses must be conducted and verified by the outside reference laboratory. The packaged substrate is then incorporated into the final kit together with the means of breath collection, storage and sample return. Included in the kit are a sample box, individual Vacutainers, a patient form with bar coded identification and a return-shipping label. The kit contains test performance instructions and the label of the kit must include the indications for which its use is intended.

All suppliers and subcontractors in the kit production must be site-visited and must pass good manufacturing practices inspections before approval is issued by the agency. Theoretically, the entire approval process is completed within one year after all required information has been supplied and accepted. Recent statistics show a mean approval time of 105 d and the $^{13}$C urea breath test required > 400 d.

**Financial assessment of test use impact on health care delivery costs.** FDA approval of a diagnostic procedure does not ensure its adoption by the medical community. It is necessary to construct an algorithm for the disease process, its presentation and differential diagnosis. In this algorithm, the costs and outcomes of alternatives to the test are compared with those in which the test is used. The construction of these decision trees takes into account the course followed by a prudent physician and identifies the costs savings and benefits from use of the test to the patient as well as to the physician. In today’s market, both must benefit if the test is to be commercially successful.

Ultimately, reimbursement decisions stem from the deliberations of the technology assessment committees of the Health Care Finance Administration, the Blue Cross/Blue Shield organizations and other major health care providers and...
insurers. Gaining access to these committees and presenting justification of the test use requires extensive experience and detailed knowledge of the health care industry.

**Economic resources required.** From the description thus far, one can see that demonstration of clinical efficacy and obtaining regulatory approval are not simple, rapid or inexpensive. Although not explicitly stated previously, development of a test to this point soon exceeds the resources of an individual academic investigator. Supporting the use of the test after approval also requires an extensive infrastructure. This includes capabilities for production and packaging, distribution and analysis, the means to accomplish tracking and reporting requirements, and finally, if appropriate, reimbursement mechanisms. For a successful first test, these capabilities must be created from scratch. Fortunately, with the infrastructure in place, subsequent tests face a much lower threshold for their introduction.

An additional question arises as to where and how breath test samples should be analyzed. Two models exist for this process. The first is possible through the advent of overnight express delivery and centralized facilities. It is no longer necessary to own a mass spectrometer to be able to carry out breath tests. All that is required is to collect the samples in the designated tubes and expedite their delivery to a centralized analytical facility. Preference for this model is reinforced by the emerging economy of scale in gas-isotope-ratio mass spectrometer systems, which are now capable of performing on the order of 300,000 analyses per instrument on an annual basis. The second model uses one of several alternative nonmass spectrometric methods for analyzing $^{13}$CO$_2$ based on infrared or laser assisted isotope ratio spectrometry.

Given time and cost constraints and the regulatory requirements of the Clinical Laboratories Improvement Act, most physicians in the United States prefer to send their samples to outside laboratories for analysis. In contrast, in European countries, where analysis of the breath samples in the physician's office may be a significant source of income, onsite analyses are more likely to be performed. A major source of concern, however, is that quality controls in the production, delivery and conduct of the test may be abandoned, if reimbursement does not fully cover costs. The absence of regulatory oversight in this process may lead to deterioration in the test quality.

**The future of $^{13}$C breath tests and their application to infants and children**

One may ask whether other breath tests are likely to reach the same general clinical application that now exists in the use of the urea breath test for *H. pylori*. Inspection of the list in Table 1 shows few candidates likely to find the type of widespread application that would justify the investment required for their commercial development.

For example, the attempt to determine the proportion of dietary fat malabsorbed and excreted from the amount consumed cannot be determined from the amount that is absorbed and oxidized. Correlations between breath tests with labeled triglycerides and fecal fat excretion have not shown a stoichiometric equivalence, and most studies have required breath collections over a 12- to 24-h period. The attendant personnel costs further reduce any perceived advantage over fecal fat measurements.

A practical maximum duration of ~120 min for the clinical conduct of a breath test seems reasonable if the test is to be widely accepted. Within this breath collection interval, there are three diagnostic test candidates that may eventually become available in commercial form. These are assessments of liver function with aminopyrine; solid-phase gastric emptying, using a prepackaged meal that requires no cooking; and the genotyping of individuals at risk for inborn errors of metabolism.

A $^{13}$C liver function breath test has demonstrated the ability to discriminate between at least three levels of liver injury as documented by liver biopsy. Moreover, its assessment of the active hepatocyte mass of the liver is not provided by any other device or method. This test also has potential application in longitudinal monitoring of liver transplantation candidates to forecast the urgency of the procedure. Despite its recognized efficacy, the aminopyrine breath test has never been the subject of an organized clinical trial with an established protocol that could be translated into an FDA application.

Solid-phase gastric emptying studies carried out with $^{13}$C have two inherent advantages over conventional radionuclide methods. Not only is there a reduction in radioactive exposure, but also there is no reliance on a nuclear medicine imaging facility. The option of test performance in a doctor’s office makes these studies particularly attractive for development. One obstacle remaining in the existing $^{13}$C solid-phase gastric emptying tests with octanoic acid is the requirement that the meal be cooked on site before the test can be administered. Despite this constraint, at least one standardized meal that can be shipped to the physician and stored until required has been developed.

The next 10 y will tell us whether the $^{13}$C-urea breath test for *H. pylori* was a fluke of circumstances or the forerunner of a flood of applications. At this time, some 25 y after its inception, the $^{13}$C-breath test remains a singularity in medical practice.

**LITERATURE CITED**


13C BREATH TESTS


