Evaluation of Prepackaged "Kits"*

REX B. CONN, JR., M.D.

Professor of Laboratory Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland

The proliferation of commercially-available "kits" for the clinical laboratory has resulted in such a technological surfeit that one or more kits are available for virtually every type of commonly performed test. Such kits offer prepackaged convenience and, under some circumstances, economies in the laboratory. A few offer technical procedures which are superior to standard methods.

In spite of their apparent simplicity, commercial availability of a reagent kit does not provide any assurance that it will perform satisfactorily or that the resulting data will be accurate. In contrast to therapeutic agents, federal regulation of the manufacture of diagnostic kits has only recently been instituted, and it remains for the user to determine whether a particular kit does in fact meet the specifications stated in promotional material. The selection and the continuing evaluation of diagnostic kits present problems to every laboratory, whether the laboratory is in the physician's office or in a large hospital.

Perhaps the simplest type of kit is a prepared reagent for a certain determination. This category would include specific antisera for blood grouping or other purposes, as well as standard chemical reagents. Usually, however, the word "kit" is used to describe a prepackaged set of multiple reagents for carrying out a certain test in the laboratory. Many such multiple-reagent kits are based upon standard, accepted methods; those from reliable manufacturers offer the advantages of standard methodology and elimination of reagent preparation. Some kits are based upon manufacturer-developed methods which are usually patented or kept as proprietary secrets. Some of these manufacturer-developed methods are acceptable; some, no doubt, have been developed primarily to permit a wider profit margin. Lastly, several well-known firms are marketing kits which are suitable only for their own analytical instruments. This presents a double dilemma for the purchaser, since it is necessary not only to evaluate the kit but the instrument as well.

Selection of Kits. Most physicians are deluged with flyers and advertisements which suggest that the purchase of a few kits is an efficient and entirely satisfactory method of installing an instant clinical laboratory. These blandishments frequently lead to an illogical and expensive system for providing laboratory data. The first question to be asked is not which kits to buy but rather which determinations should be done in the laboratory. If a test is definitely needed, purchase of a kit is one of the alternative methods for making it available. Particularly in the physician's office, the convenience, the elimination of reagent preparation and savings in personnel time are advantageous. Use of a kit usually results in a higher per-test cost than having the test done in a large, automated laboratory, but the advantages of using kits sometimes outweigh these higher costs. Kits from different manufacturers will offer different analytical methods, differing numbers of tests per kit, different instrument requirements and, of course, different prices.

An important principle in selecting any type of laboratory kit is to require the manufacturer to provide relevant experimental data which substantiate any claims regarding performance of the kit. Descriptions such as accurate, precise, simple, inexpensive and reliable are all relative. Unless the manufacturer can produce data, preferably substantiated by an independent investigator, regarding these parameters of performance, further consideration of using the kit should not be entertained. Most reputable manufacturers will supply reprints of articles describing such evaluations.

Evaluation of Kits. If a laboratory is to produce

* Presented by Dr. Conn at the 44th Annual McGuire Lecture Series, March 22, 1973, at the Medical College of Virginia, Richmond.

MCV QUARTERLY 9(4): 350–354, 1973

reliable data, a critical evaluation of each method must be carried out prior to its introduction as a routine procedure. It is immaterial whether or not the procedure utilizes a prepackaged kit. Such an evaluation should make full use of any data collected by other laboratories, but an experimental evaluation by the laboratory which is to use the kit is essential.

Although kits come in various forms, there are two major categories of laboratory tests for which they will be used—qualitative tests and quantitative tests. Qualitative tests are those for which the results can be expressed as a yes or no, positive or negative or present or absent report. Pregnancy tests and tests for urinary glucose are examples of qualitative tests. Quantitative tests are those which are used for measurement of a specific constituent and results are expressed in numerical terms. The evaluation of a kit will differ depending on whether it is used for a qualitative or quantitative procedure.

Comparison with Reference Method. Evaluation, no matter what is being evaluated, is a comparative process. If a kit merely supplies reagents necessary to do a standard laboratory test, data on the standard test are readily available in published form. More often, the kit will be a modification of a standard method or, occasionally, a new approach to measuring the same constituent. In either situation, the manufacturer should provide experimental data which compare the kit procedure to an accepted, established method.

For quantitative tests, the parameters which should be examined are accuracy, precision and range of linearity over which measurements can be made. Accuracy is an elusive parameter to evaluate, and a comparison of results using the kit with those obtained by a standard method is acceptable. Precision can be evaluated by replicate determinations and linearity by measuring a series of samples with varying concentrations of the constituent being measured. An excellent example of such a comparison is the study by Barnett, Cash and Junghans (2) in which cholesterol measurements using 12 different kits were compared to those using the Abell-Kendall method. They concluded that only two of the 12 kits being marketed at the time of the study were acceptable for clinical use. Other published evaluations should be equally disconcerting to any laboratory which uses a kit without first subjecting it to rigorous performance trials. Kim, Waddell and Logan (6) measured sodium and potassium with chemical kits manufactured by the Stanbio Laboratory, San Antonio, Texas and by Medi-Chem, Santa Monica, California. Results were compared with those obtained by standard flame photometric techniques, and the authors concluded that both kits gave "diagnostically unsatisfactory results." A study by Dietz, Rubenstein and Lubrano (5) which involved serum cholinesterase measurements using the Acholest® kit manufactured by E. Fougera and Co., Hicksville, New York, provides a shocking indictment of the lax standards set by some commercial firms. The Acholest® method "failed to detect 12 of 20 cases at high risk of prolonged apnea after succinylcholine." Similar comparative studies of kits for less complicated procedures such as glucose and urea indicate that some are entirely suitable for these measurements (7, 9, 11). It is noteworthy that Logan, Waddell and Krynski (9) found that the kits which were the most expensive and which revealed the least information concerning their constitution gave the poorest performances.

The evaluation of comparative data for qualitative tests frequently is more difficult than for quantitative tests. The two parameters corresponding to accuracy and precision are validity and reproducibility. Reproducibility can be studied through replicate tests on the same group of samples; validity, however, like accuracy, may have to be evaluated by comparison with a reference method. The objective for all qualitative tests is a positive result when the constituent or disease is present and a negative result when it is not. For example, serological tests for syphilis are usually compared to the fluorescent treponemal antibody-absorbed (FTA-ABS) test. Since the serological test is used primarily as a screening procedure, acceptable performance would result in no false negatives and as few false positives as possible. Frequently, the evaluation must include consideration of the clinical context in which the test will be used. Tests for pregnancy are usually evaluated by testing large numbers of pregnant and nonpregnant women. A test which gave positive results in 97% of women in the second trimester of pregnancy clearly would not be as useful as one which gave similar accuracy during the first three weeks. The undesirability of false positive pregnancy tests is readily apparent.

In addition to comparisons with reference methods, information provided by the manufacturer should include predisposing test conditions or patient abnormalities which will affect the test and give inaccurate or undependable results. Such interfering

conditions are particularly troublesome with newly developed tests such as radioimmunoassays. Tests for digoxin, for example, may measure not only digoxin but also its metabolites (10). More worrisome, however, are reports that therapeutically administered compounds such as cortisol and spironolactone may react with the digoxin-binding antibody to give erroneous results.

Laboratory Trials. Although few laboratories will carry out an evaluation of a kit as elaborate as those which are described in scientific journals, it is falacious to assume that results similar to those in published articles can be obtained automatically by any kit purchaser. The procedures for experimental evaluation may vary from kit to kit, but critical testing of every type of kit by the laboratory in which it is to be used is essential.

Meticulous examination of the instructions which accompany the kit will frequently eliminate unnecessary work. The instructions should present a logical and detailed outline of each step in the procedure, with a clear indication of where errors might occur, what types of instruments are suitable, in which steps timing is critical and how results are to be calculated from instrument readings. The instructions should be followed compulsively under all circumstances. Suitable standards should be included with all kits for quantitative measurements and the standards should have concentrations which span the range of clinically useful measurements. Controls should be run with each batch of any procedure, whether it is a quantitative or a qualitative one and, if possible, control solutions should be obtained from a manufacturer other than the supplier of the kit. Qualitative tests generally should have both a positive and a negative control; these usually accompany the kit, however, since they may be the only suitable controls available.

Replicate determinations on different days of one, or preferably several, control solutions provides an indication of the precision which might be expected. If reproducibility is unsatisfactory, the problem may reside either in the kit or in the technique. If errors can be traced to faulty technique which is corrected, the experiments should be repeated; if not, the kit should not be accepted for routine use in the laboratory. A second useful step in evaluating a kit is separation of patient samples into two aliquots, one to be run by the kit method and the other to be submitted to a reference laboratory. A minimum of a dozen, and preferably several times that number, split samples should be analyzed before acceptance of the kit for routine use. Many manufacturers will supply free samples of kits for preliminary evaluation; this practice, however, should have no influence on the laboratory in regard to which kits are tested and which are finally selected for routine use.

Evaluation of an instrument which is designed for use with kits for a variety of procedures may be quite time-consuming, since each procedure for which a kit is available must be evaluated separately. Logan and Sunderland (8) evaluated the Unitest System® marketed by Bio-Dynamics, Inc., Indianapolis, Indiana, and Diagnotest® marketed by Dow Chemical Company, Diagnostic Products Division, Indianapolis, Indiana. For each instrument, some of the kits performed unsatisfactorily, and the authors concluded that personnel without technical training could not obtain reliable data with either system.

Continuing Evaluation. Initial evaluation and acceptance of a kit for routine use in the laboratory does not assure continuing satisfactory performance. Most important, the use of kits does not obviate the need for strict quality control measures. Suitable quality control solutions should be run with each batch of procedures, strict limits for variation of the control solution should be set and all data should be rejected if control readings are out of the predetermined range. Lot numbers of each reagent or kit should be entered into the laboratory log book and when a new lot number is used, samples should be run in duplicate with the old and the new reagents to permit comparison between lots.

FDA Regulations Regarding Kits. It perhaps should be stressed that most of the laboratory kits on the market today were developed during a period when there were no federal regulations setting minimum performance standards. As early as 1966 the American Association of Clinical Chemists (1) published policies regarding reagent sets and kits which, had they been followed by all manufacturers, might have greatly reduced the number of subsequent articles devoted to documenting the inadequacies of many kits. Manufacturers also could voluntarily submit kits to the College of American Pathologists for evaluation; however, compliance with the recommendations of the college in the case of inadequate kits was also voluntary. Published evaluations of all types of kits clearly indicate that some are unsatisfactory, some are satisfactory and some are outstanding in meeting performance standards.

In January 1972, the Food and Drug Administration announced that existing legislation gave them authority to exercise regulatory control over diagnostic kits to ensure that they deliver a consistently high level of quality and performance (3). A statement of procedures for developing policy and interpretive regulations was published in August 1972 (4). Briefly, these regulations require that diagnostic kits be accurate and reliable, that manufacturers test and evaluate kits prior to marketing them and verify results against a generally accepted procedure, that premarket testing is done to find if predisposing patient abnormalities will affect the test and that the labeling of all kits contains adequate directions for use. The labeling directions must include complete information on accuracy, reproducibility and sensitivity performance. The FDA intends to establish standards of performance for each type of laboratory kit and require manufacturers to meet these standards. Establishing pertinent standards will be time consuming (the FDA intends to start with glucose and hemoglobin), but merely requiring manufacturers to provide evidence that kits will perform as claimed will be helpful for anyone who must decide which kit to purchase. The FDA is already enforcing these regulations and has required two manufacturers of pregnancy test kits to recall their products.

Cost Evaluation. The cost of performing laboratory tests is causing increasing concern because of the dramatic increase in the use of laboratory data in patient care and because many of the newer tests are more complicated and thus more expensive. Evaluation of a laboratory procedure should include the cost of doing it. In a physician's office or a small laboratory, such cost accounting can be quite simple, since it is easy to calculate the cost per test done by a kit method and to measure personnel time involved. Since most kits have expiration dates, the cost per test of the kit should take into account the necessity for discarding outdated reagents. Generally speaking the cost of a quantitative measurement carried out by a kit method will be considerably higher than the same test carried out on automated laboratory instruments. On the other hand, even large laboratories use prepared reagents and kits for performing some of the simpler tests and these tests can be performed in the office laboratory at the same or perhaps lower cost than in a large laboratory. Higher costs for performing tests in physicians' offices may be offset by convenience to the patient or the necessity for having data immediately available.

Summary. Prepackaged laboratory kits for performing diagnostic procedures are frequently the most suitable alternative in the selection of laboratory methods, especially in physicians' offices and small laboratories. Because of the previous lack of governmental regulations covering the manufacture of kits, many kits now on the market do not perform adequately and may produce misleading results. Each laboratory must evaluate each type of kit before it is put into routine use. This evaluation should include a review of published experimental data, comparison of results using the kit to results using a reference method and an experimental evaluation of the kit in the laboratory in which it is to be used.

REFERENCES

1. American Association of Clinical Chemists Committee on Standards and Controls. AACC policy regarding reagent sets and kits. *Clin. Chem.* 12:43, 1966.

2. Barnett, R. N., Cash, A. D. and Junghans, S. P. Performance of "kits" used for clinical chemical analysis of cholesterol. *N. Engl. J. Med.* 279:974, 1968.

3. Department of Health, Education and Welfare. *In vitro* diagnostic products for human use: Notice to manufacturers, packers and distributors. *Federal Register,* January 19, 1972.

4. Department of Health, Education and Welfare. *In vitro* diagnostic products for human use: Proposed establishment of procedures for developing statements of policy or interpretive regulations. *Federal Register,* August 17, 1972.

5. Dietz, A. A., Rubenstein, H. and Lubrano, T. Detection of patients with low serum cholinesterase activity: Inadequacy of "Acholest" method. *Clin. Chem.* 18:565, 1972.

6. Kim, E. K., Waddell, L. D. and Logan, J. E. Evaluation of four reagent kits and two flame photometers used to determine sodium and potassium in serum. *Clin. Chem.* 18:124, 1972.

7. Logan, J. E., Renton, H. M. and Eby, P. W. Observations on diagnostic kits for the determination of urea nitrogen. *Clin. Biochem.* 2:189, 1969.

8. Logan, J. E. and Sunderland, M. L. E. Assessment of two blood-analyzer systems intended for diagnostic

testing in the physician's office. *Clin. Chem.* 16:990, 1970.

9. LOGAN, J. E., WADDELL, L. D. AND KRYNSKI, I. A. Observations on diagnostic kits for the determinations of glucose. *Clin. Biochem.* 3:129, 1970.

10. STOLL, R. G., CHRISTENSEN, M. S., SAKMAR, E. AND WAGNER, J. G. The specificity of the digoxin radioimmunoassay procedure. *Res. Comm. Chem. Pathol. Pharmacol.* 4:503, 1972.

11. WRIGHT, W. R., RAINWATER, J. C. AND TOLLE, L. D. Glucose assay systems: Evaluation of a colorimeter hexokinase procedure. *Clin. Chem.* 17:1010, 1971.