Accuracy in Clinical Chemistry—Does Anybody Care?

Norbert W. Tietz

The role of clinical chemistry in the diagnostic process has been enhanced by the evolution of better instrumentation and analytical techniques. The quality of some laboratory tests, however, has not kept pace with these advances. I present three examples—serum iron, serum lipase, and nonisotopic immunoassays—assays where some currently used methodologies are highly flawed. Causes for the less than optimal performance of some clinical laboratories are discussed.

Indexing Terms: iron/lipase/immunoassays/proficiency testing

Better instrumentation and the emergence of highly accurate analytical techniques provide unequalled opportunities to raise the quality of diagnostic laboratory services to levels never seen before. Despite these significant developments, the quality of some laboratory tests has not kept pace with these advances. Some instruments and methodologies offered by industry and selected for use in clinical laboratories lack analytical accuracy, reliability, or clinical specificity (see later examples). A recent article, “Excelling in Mediocrity,” confirms the contemporary trend and points out the “increasing acceptance of mediocrity and decreasing quest for excellence” (1).

Serum Iron Determinations

Twenty years ago, several manual methods with clinically acceptable accuracy were available for the determination of serum iron. However, some recently introduced automated and “simplified” manual methods have been shown to lack necessary accuracy, as has been documented recently (2). Among the methods used in the proficiency test offered by the College of American Pathologists (CAP) (specimens C01-05, 1993), only a few had a peer group mean within 10% of the respective values obtained with the reference method of the International Committee for Standardization in Haematology (ICSH). Proficiency testing programs sanction such performance because most laboratories submitting unsatisfactory values are protected by the “peer group” concept (3). No matter how deficient a method is, the results are certified as acceptable as long as a group of peers using the same analytical system obtains results within a certain—often wide—range of “acceptability.”

In the case of iron, the acceptable range is ±10% of the peer group mean.

The matrix effect observed with some proficiency testing materials has frequently been blamed for the discrepancies in results observed in such assays. This statement does not take into consideration that methods subject to the matrix effect are frequently not robust methods. In this context, Rej (4) writes: “by current conventional wisdom, anomalies encountered with quality-control serum specimens are considered to be matrix effects. Unfortunately this designation is often accompanied by an implicit acceptance of such aberrant results, and the contribution of the analytic method to this deviation is frequently overlooked.” Matrix effect is an error introduced by any or all other components in the specimen and must be considered to be an interference. This observation is also shared by Büttner (5). According to the IFCC definition (6), specificity is “the ability of an analytical method to determine solely the component(s) it purports to measure;” thus, the matrix effect is a result of an inherent lack of specificity of the analytic technique. Traditional analytic techniques of separation of the analyte from interferences in a matrix have largely been abandoned in favor of the speed and convenience of direct measurements on serum specimens (4).

On patients’ specimens, certain methods for serum iron perform even worse than in proficiency tests. When comparing results obtained with the ICSH reference method with results obtained with five different routine analyzers, my colleagues and I documented errors in measuring low iron values (i.e., <750 μg/L) of >600% (2). Several of these methods may give iron values that suggest iron deficiency, when in fact iron values are within normal limits when measured with the ICSH method. Such erratic results may complicate the detection of iron deficiency and may even lead to a wrong diagnosis.

As judged by the results of the CAP survey for iron and the widespread use of unsatisfactory methods, seemingly few individuals care about (or are aware of) this degree of inaccuracy of iron measurements, despite the established clinical need for accurate iron measurements. We should remember that: 6% of Americans are in significant negative iron balance, ~1% of the population have iron overload (7), iron has recently been implicated as a risk factor for myocardial infarction (8), and iron may be a factor in the pathogenesis of rheumatoid arthritis (9). Accurate measurement of this element is essential.
Lipase in Serum

Lipase is determined with increased frequency, because it is more specific than amylase measurements for diagnosing pancreatitis (10). Although optimal reaction conditions for this enzyme are now well defined, procedures for lipase with suboptimal reaction systems are still widely used (10). The substrate in a popular dry-slide technique is more likely to be hydrolyzed by intestinal lipase than by pancreatic lipases (11). Clinical evaluations of this method have shown a low clinical specificity; i.e., the method reports a disproportionately high incidence of increased lipase values in patients with nonpancreatic abdominal diseases (12). Nevertheless, this procedure was used in the US in 1992 by 670 laboratories, 43% of all laboratories reporting lipase values (13). Another procedure, evidently popular (i.e., used by 44% of all laboratories) because of its conveniently pre-packaged reagents, produced in our hands highly variable results, exhibited unsatisfactory linearity, and had a high positive intercept in comparison with a method that meets established optimal reaction conditions.

Immunaoassays With and Without Isotopic Labels

The advent of radioimmunooassays (RIAs) enhanced the diagnosis of endocrine and other disorders, but the reluctance to handle isotopes, the desire to automate immunoassays, and the quest for increased sensitivity motivated the use of nonisotopic labels—enzyme, fluorescent, and chemiluminescent labels—in these immunoassays. We investigated 10 of these procedures on five different instruments and found good correlation with results of established manual RIA techniques for only two analytes, folitropin (follicle-stimulating hormone) and thyrotropin (thyroid-stimulating hormone). We saw great discrepancies between values obtained with these five automated immunoassays and those from manual RIAs, yet all 10 methods are in routine use. One example, assay of total triiodothyronine, is shown in Fig. 1.

The four most discrepant values (ng/L) for the two methods measuring this analyte and the associated differences were as follows: 10 vs 380, +3700%; 70 vs 870, +1143%; 1870 vs 810, −57%; and 1180 vs 5120, +80%. These values are marked in Fig. 1 by a square. Reference ranges were similar for both methods.

As methods with new labels are refined, we can expect improvements, but it is wrong to use new immunoassays before thorough documentation that their accuracy equals or exceeds that of existing methods. It is disturbing to see that some manufacturers have introduced simpler, rather than more accurate methods, and that some laboratorians have yielded to the temptation to accept these methods before extensive evaluation as to whether or not the candidate method gives analytically and clinically superior results.

How Did We Reach This Situation?

Reasons for the developments and concerns expressed above are multifold, and include the following:

Evidently, accuracy as a criterion for the selection of instruments and procedures is now overshadowed by considerations of simplicity and expediency. Moreover, concern for the financial “bottom line” overshadows professional standards; financial issues can be given too much importance at the expense of the patient. For example, the laboratory’s use of a lower-priced analytical system that generates results in disagreement with another instrument in the same laboratory may confuse the physician and possibly lead to errors in treatment. A system with low reliability can generate inaccurate results, leading to repeat testing, increased hospital stay, a delayed diagnosis, or even a wrong diagnosis—situations that increase cost and ultimately negate any initial savings in the purchase of a flawed instrument.

Many laboratories do not receive sufficient financial support to purchase reliable equipment and hire well-qualified personnel; instead, laboratories are all too frequently used as a source for revenue rather than a place to improve and speed up the diagnostic process where timely and accurate information of significant benefit to the physician and patient can be provided. Such diversion of laboratory funds forces the adaptation of inferior methodology and the purchase of suboptimal instrumentation. Healthcare administrators overlook the fact that accurate and efficient laboratory service lowers hospital costs by facilitating speedy diagnosis, treatment, and follow-up of patients.

Regulations imposed by the US Food and Drug Administration (FDA) are very elaborate, time-consuming, and expensive for manufacturers of instruments and reagent systems. Nonetheless, FDA approval does not guarantee adequate performance of the diagnostic device. Furthermore, FDA regulations restrict laboratorians so severely that development and evaluation of new and more accurate procedures are not only discouraged but in some situations made impossible. For example, for our laboratory to evaluate tumor markers that are currently approved “for investigational use only,” we need approval by the institutional research committees and certification from the physician that the test will
not be used for diagnosis (positions with which we agree). However, recently, manufacturers, as a result of FDA directives, have refused to ship the reagents for such studies, unless the company is given access to the patients’ files (which would violate confidentiality requirements). Also, it is not well known that assays of tumor markers, such as TdT (terminal deoxribonucleotidyld transferase, for the differential diagnosis and treatment selection of leukemia), SCC antigen (for the diagnosis of squamous cell carcinoma), β-subunit of chorionic gonadotropin (for the diagnosis of choriocarcinoma), estrogen-receptor assays (for selection of treatment of breast cancer), and cancer antigen CA 19-9 (for the diagnosis of pancreatic tumors) are still not FDA-approved, although they have been thoroughly evaluated, have been widely used for many years, and are today part of the accepted standard diagnostic procedure for the respective disorders.

Regulations of the 1988 Clinical Laboratory Improvement Act (CLIA) have lowered the laboratory personnel standards to such a degree that some laboratories no longer have the expertise to evaluate instruments and analytical techniques. This development gives rise to the “push-button” mentality; i.e., instruments are selected that simply require the push of a button. Administrators and poorly qualified laboratory workers fail to realize that these instruments may not be the best available and that skilled laboratorians might still be required to recognize faulty operation when it occurs.

Only few medical schools incorporate courses into the curriculum that adequately deal with the use and evaluation of laboratory methods and data. Concepts of clinical sensitivity and clinical specificity, predictive value, receiver-operator characteristic (ROC) curves, use of parallel and sequential testing, and other information useful to a practicing physician for evaluating and selecting the most reliable laboratory methods are rarely well understood. Therefore, physicians may not readily recognize clinically or analytically deficient methods. Despite the intense competition for curriculum time, these topics must receive more attention if we are to enhance the diagnostic process and use the laboratory resources most efficiently.

Exchange of information between the clinician and the laboratorian is too often lacking. Without feedback from the clinician, the laboratory is handicapped in the evaluation of an untried technique. Likewise, the clinician, without feedback from the laboratorian, lacks an understanding of the clinical advantages and analytical limitations of specific procedures. Frequent, direct contact between physicians and laboratorians would help resolve this problem. Physicians should be encouraged to visit the laboratory and be made welcome there.

Recommendations

Being greatly concerned about the lack of accuracy of some of the diagnostic services, I urge all those concerned to take full advantage of the methodological and instrumental advances and to reject systems that do not meet required clinical and analytical standards. Current knowledge and technological capabilities allow for provision of better care without increasing healthcare cost. What it takes is a commitment to excellence and a demonstration that we in the clinical laboratory care by:

1. Intensifying studies of the accuracy and clinical utility of laboratory tests, and incorporating into medical school curricula instructions for proper selection and clinical evaluation of laboratory tests;
2. Insisting that industry improves the design of instruments and reagent systems;
3. Returning to the concept that accuracy is an important criterion for the selection of analytical systems, with cost, speed, convenience, and expediency being additional considerations;
4. Requiring government agencies to generate guidelines that will encourage and not suppress initiatives for improving diagnostic systems;
5. Instituting standards for laboratory personnel that assures competence;
6. Improving proficiency testing criteria by setting acceptable limits in accordance with clinical need and requiring that comparison of results obtained by routine methods with those obtained by reference methods are within those limits.

My sincere thanks go to Dankwart Stamm, Elizabeth L. Pruden, A. Ralph Henderson, J. Stanton King, Robert Rej, Denis O. Rodger, and Ronald J. Whitley, as well as Alan D. Rinker and Marilyn R. Retzlaff for their review of the manuscript and for helpful suggestions. I also thank Terry A. Weeks for the generation of statistical data.

References