

This guidance was written prior to the February 27, 1997 implementation of FDA's Good Guidance Practices, GGP's. It does not create or confer rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statute, regulations, or both. This guidance will be updated in the next revision to include the standard elements of GGP's.

**POINTS TO CONSIDER FOR COLLECTION OF DATA IN SUPPORT OF IN-VITRO  
DEVICE SUBMISSIONS FOR 510(K) CLEARANCE**

Many changes are occurring in FDA's regulation of in vitro diagnostic devices as a result of the Safe Medical Devices Act of 1990 (SMDA 90), the report of the FDA Committee for Clinical Review, and initiatives of the Office of Device Evaluation (ODE) to raise the quality of scientific review of premarket submissions in a pragmatic manner. The safety of an IVD is inversely proportional to the probability that its use will result in a misdiagnosis of the patient. SMDA 90 broadens the scope of the FDA review of IVDs and authorizes FDA to request clinical data if necessary to ensure that a new IVD is as safe and effective as a legally marketed IVD with the same intended use.

The following points represent current FDA policy regarding collection of data in support of IVD submissions for 510(k) clearance. This document is intended to supplement but not replace existing FDA guidance documents for premarket submission.

**STUDY PROTOCOL:**

Uniform protocols for all clinical sites must be established prior to and followed consistently throughout the course of data collection(51). Any changes in the study design should be clearly documented, justified, and reflected in data interpretation. A copy of the protocol should be included as part of the 510(k) submission. The protocol of clinical studies should clearly define the study population and inclusion and exclusion criteria.

**SAMPLING METHOD(S)**

1. The nature of the sampling method for selecting patient samples should be clearly addressed in the protocol and reported in the submission. Sampling may be purposive (distribution defined by the investigator) to allow for analytical method evaluation (where a wide range of normal and abnormal samples are needed to characterize performance). A study designed for delineation of expected clinical performance may require random and blind (masked) sampling.

**a. QUANTITATIVE TESTS:**

Evaluation of tests employing quantitative measurement techniques should include at a minimum an evaluation of random and systematic error (1-8,12,15,20,27,29,32,36-38,40,41,43,44,46-48,51,53,56,62-64) in comparison to a legally marketed predicate device. Comparisons may be direct between the two devices and/or indirect with the new and old devices compared to a reference method or

"gold standard". These comparison studies should include the key parameters of regression analysis with estimated slope and intercept and their 95% confidence intervals. For all tests, (and a requirement when statistical differences are noted between the new test and the predicate) decision points for device use should be identified and an error analysis (52,54) should be performed at each of these points.

Analytical specificity should be evaluated taking into consideration all known or expected confounding factors. When clinical decision making is expected to be predicated on results at the lower end of the reportable analytical range, a determination of analytical sensitivity is also required.

The statistical theory of linear regression analysis requires independence of data (i.e., only one sample pair from each patient) among various sample pairs for calculating slope, intercept, and their 95% confidence intervals. If multiple sample pairs from the same patient were used, appropriate justification and statistical procedures are needed to account for between-patient and within-patient variabilities.

b. QUALITATIVE TESTS:

Evaluation of tests employing qualitative measurement techniques should include a characterization of the test using clinical/diagnostic characteristics (9,13,14, 16,18,19,21-23,24,30,33-35, 33-35,39,42,45,49,55,56,59-62,66) and/or a description of analytical characteristics using discrimination zone and cut-off points (17,23,26).

Positive controls (calibrators) should be selected near the clinically relevant cut-off and not be limited to high titer positive controls.

c. STATISTICAL METHODS FOR EVALUATION OF DEVICE

The statistical methods used to evaluate a 510(k) submission should be appropriate for the study protocol, type of data collected and intended use of the IVD. The method should either be a well recognized referenced method (see attached bibliography for suggested protocols to follow) or should be clearly explained and justified.

Confidence intervals should be included in statistical analysis whenever possible (11,28,31). ANOVA models are the preferred method because detailed ANOVA tables provide pertinent information, such as specified statistical model, type of effect (random, fixed, or

mixed), source of variation, degrees of freedom, sums of squares and mean square, F-statistic, and p-value) For ANOVA models, the variance components from ANOVA rather than confidence intervals SHOULD be presented.(32)

d. USE OF TIER III 510(k) REVIEW

The increased capability of requiring clinical data under 510(k)s may allow FDA To handle some IVDs as tier III Premarket notification 510(k) applications rather than premarket approval applications (PMA) If the following conditions apply: the analyte is well-established but there is new methodology or new matrix(es). These submissions should include a description and actual results of the clinical or diagnostic as well as the analytical performance characteristics of the new device for its intended use (9,13,14,16,18,19,21-23,24,30,33-35,39,42,45,49,55,56,59-62,66). This requires comparison of the new test against a defined laboratory or clinical standard with reporting of new test results in terms of diagnostic sensitivity and diagnostic specificity .

DEVICE USED FOR GENERATING DATA FOR SUBMISSION:

Ideally studies should be performed using a final version of the production model device (defined read head, optics, etc.) and not a prototype. Data based only on prototype devices may raise concerns about the ability of the device to perform in non-professional setting.

INSTITUTIONAL REVIEW BOARD REVIEW

All studies must be performed under the review of an Institution Review Board.

QUALITY CONTROL FOR STUDIES USED IN SUBMISSION:

Studies should be performed using accepted methods for quality control as outlined under the CLIA'88 regulations. Data collected during any run in which the assay appears to be out of control should never be used in support of a submission.

STUDY SITE REQUIREMENTS:

The types and number of study sites required for data collection and establishment of the performance characteristics described above will depend on the type of submission. If the analyte, methodology, and intended use of the IVD are well-established, less rigorous testing is required than for the clearance of an IVD where these components are not well-established.

a. For IVDs with poorly defined reference standards or

methodologies for which population and/or performance based site differences would be expected, the manufacturer should provide a minimum of three study sites. All sites should follow a uniform protocol.

- b. For well standardized devices for which population and/or performance based site differences would not be expected, a single study site may be sufficient. The source of samples used should be clearly described and documented particularly if a study site is not a clinical site.
- c. For devices to be marketed in physician office laboratories, a minimum of three study sites must be included with device operators representing the type of individuals likely to be performing the test, e.g., nurses, physicians, medical office assistants, etc.
- d. For over-the-counter (OTC) or home-use devices, a minimum of three study sites must be included. In general, several hundred untrained lay users or general consumers should participate in testing at these sites. Readability and understanding of the package insert labeling and instructions, and use of the test information by lay persons must be evaluated for all OTC devices, even those that are just collection devices. If the lay persons are supposed to actually perform the analytic procedure of the OTC IVD, test performance in the hands of lay persons must be compared to test performance in the hands of professional laboratory personnel using paired samples(50,53).

The testing should be based on the design and intended use of the IVD. Is the device limited to collecting and mailing off the specimen? Is the test intended to be performed by a lay person? Who interprets the result? What is the support and backup offered by the manufacturer to the lay user, e.g., toll-free telephone number staffed by a health care professional.

#### RESPONSIBILITY OF PRINCIPAL INVESTIGATORS OF CLINICAL STUDIES:

When studies are performed at study sites other than the manufacturer's own facility, the responsible (principal) investigator(s) must sign off on the study indicating that a study protocol was in place, was followed throughout the study course, and that the investigator has reviewed and verified the data and the manufacturer's presentation of data analysis as presented in the submission to the FDA. This is similar to the requirement of peer-reviewed medical journals.

#### PRODUCT INSERT (PACKAGE INSERT) CLAIMS ABOUT STUDY DATA:

- 7. The nature of all data should be clearly reported in both the 510(k) and the package insert. The manufacturer should indicate the number and types of study sites used to

establish the performance characteristics being described in the 510(k) and the package insert (PI). The manufacturer's claims for the IVD will be limited to the actual test data cleared by the FDA.

Examples:

Manufacturers who limit their testing to the manufacturing site only, will be limited to a statement in the PI such as:

**"This performance data was generated only on-site at company x" or some equivalent language.**

However, manufactures obtaining performance data at the manufacturing site and two clinical sites will be allowed to make a statement in the PI such as:

**"Performance data generated on-site at company x, at a large (university, general, etc.) (high complexity) hospital laboratory in Texas, and at a small hospital laboratory (moderate complexity) in New Jersey" or some equivalent language.**

Some IVDS, especially ones that measure analytes associated with low prevalence conditions, may require more than 3 testing sites. The manufacturer is allowed to claim in the PI the results of all testing data cleared by the FDA.

The goal for the PI is truth in labeling. PI Labeling should indicate to users the actual scope of information known about test performance for a product. Companies should be allowed to use their data to support claims based on actual testing of their product.

#### REFERENCES

1. Barnett, R.N., A scheme for the comparison of quantitative methods. Amer. J. Clin. Pathol. 43,562, 1965.
2. Barnett, R.N. and Youden, W.J. A revised scheme for the comparison of quantitative methods. Amer. J. Clin. Pathol. 54,454, 1970.
3. Bauer, S., Kennedy, J.W., Applied statistics for the clinical laboratory: III. variability over time. J. Clin. Lab. Autom. 1:35-40, 1982.
4. Bauer, S., Kennedy, J.W., Applied statistics for the clinical laboratory: VII. initial concepts for assessing accuracy. J. Clin. Lab. Autom. 2:354, 1982.
5. Bauer, S., Kennedy, J.W., Applied statistics for the clinical laboratory: VIII. measures of inaccuracy. J. Clin. Lab. Autom. 3:46, 1983.

6. Bauer, S., Kennedy, J.W., Applied statistics for the clinical laboratory: IX. distributions and normality. J. Clin. Lab. Autom. 3:183, 1983.
7. Bauer, S., Kennedy, J.W., Applied statistics for the clinical laboratory: X. the assessment of inaccuracy. J. Clin. Lab. Autom. 3:340, 1983.
8. Bauer, S., Kennedy, J.W., Applied statistics for the clinical laboratory: XI. student's t test used to assess inaccuracy. J. Clin. Lab. Autom. 4:56-63, 1984.
9. Beck, J.R., Shultz, K.: The use of relative operating characteristic (ROC) curves in test performance evaluation. Arch. Pathol. Lab. Med. 110:13, 1986.
10. Begg, C., Biases in the assessment of diagnostic tests. Statistics in Medicine 6:411, 1987.
11. Braitman, L.E., Confidence intervals extract clinically useful information from data. Ann. Intern. Med. 108:296, 1988.
12. Broughton, P.M.G., Gownlock, A.H., McCormack, J.J., Neill, D.W., A revised scheme for the evaluation of automatic instruments for use in clinical chemistry. Ann. Clin. Biochem. 11:207-208, 1974.
13. Brown, G.W., Comparing two tests. Diagnostic Medicine. pp.28-31, 1985.
14. Buck, A.A., Grant, J.J., Comparison of a screening test and a reference test in epidemiologic studies - 1. indices of agreement and their relation to prevalence. Am. J. Epidemiology. 83:586, 1966.
15. College of American Pathologists, Laboratory instrument evaluation, verification, and maintenance manual, 1989.
16. Coughlin, S.S., Trock, B., Criqui, M.H., et al., the logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test. J. Clin. Epidemiol. 45:1-7, 1992.
17. Delger, D.M., Rea, M.R., Kessler, G. et al., Criteria for evaluating nonquantitative assays: application to serum choriogonadotropin. Clin. Chem. 33:150, 1987.
18. Delong, E.R., Vernon, W.B., Bollinger, R.R., Sensitivity and specificity of a monitoring test. Biometrics. 41:947-958, 1985.
19. Draft of points to consider in the manufacture and clinical evaluation of in vitro tests to detect antibodies to the human immunodeficiency virus, type 1 (1989), Center for Biologics

Evaluation and Research.

20. Emancipator, K., Kroll, M.H., A quantitative measure of nonlinearity, Clin. Chem. 39:766, 1993.
21. Gail, M.H., Evaluating serial cancer marker studies in patients at risk of recurrent disease. Biometrics. 37:67-78, 1981.
22. Galen, R.S., Gambino, S.R. Beyond Normality, the Predictive Value and Efficiency of Medical Diagnosis. New York: Wiley and Sons, 1975.
23. Garrett, P.E., It looks okay to me or should we evaluate qualitative immunoassays? J. Clin. Immunol. 10:147, 1987.
24. Goodman, S.N., Berline, J.A., The use of predicted confidence intervals when planning experiments and the missues of power when interpreting results. Ann. Intern. Med. 121:200-206, 1994.
25. Greenberg, R.A., Jekel, J.F., Some problems in the determination of the false positive and false negative rates of tuberculin tests. Amer. Rev. Resp. Dis. 100:645, 1969.
26. Guidelines for the evaluation of diagnostic kits - part 2: general principles and outline procedures for the evaluation of kits for qualitative tests. European Committee for Clinical Laboratory Standards, 1987.
27. Hackney, J.R., Cembrowsky, G.S., Need for improved instrument and kit evaluations. A.J.C.P. 86:391, 1986.
28. Harris, E.K., On P values and confidence intervals (why can't we P with more confidence?) Clin. Chem. 39:927, 1993.
29. Harris, E.K., Regression, Least Squares, and Correlation. In CRC Handbook of Clinical Chemistry - Volume III. Boca Raton, Florida, CRC Press, pp.193-213, 1982.
30. Henderson, A.R., Assessing test accuracy and its clinical consequences: a primer for receiver operating characteristic curve analysis. Ann. Clin. Biochem. 30:521, 1993.
31. Henderson, A.R., Chemistry with confidence: should Clinical Chemistry require confidence intervals for analytical and other data? Clin. Chem. 39:929, 1993.
32. Henry, J.B., Beeler, M.F., Copeland, B.E. and Wert, E.B. A format for description of methods in clinical pathology. Amer. J. Clin. Pathol. 52:296, 1969.
33. Hsu, S.L., Walter, S.D., Estimating the error rates of diagnostic tests. Biometrics. 36:167, 1980.



34. Ilstrup, D.M., Statistical methods in microbiology. Clin. Micro. Rev. 3:219, 1990.
35. Kent, D.L., Decision analysis and assessment of the clinical impact of diagnostic tests. Lab. Med. 22:718, 1991.
36. Krouwer, J.S., Steward, W.N., Schlain, B., A multi-factor experimental design for evaluating random-access analyzers. Clin. Chem. 33:1894-6, 1988.
37. Krouwer, J.S., Multi-factor designs IV. how multi-factor designs improve the estimate of total error by accounting for protocol-specific biases. Clin. Chem. 37:26-29,, 1991.
38. Krouwer, J.S., Estimating total analytical error and its sources - techniques to improve method evaluation. Arch. Pathol. Lab. Med. 116:726, 1992.
39. Lachs, M.S., Hachamkin, I., Edelstein, P.H., et al., Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid dipstick test for urinary tract infection. Ann. Intern. Med. 118:230, 1993.
40. Lasky, F.D., Achieving accuracy for routine clinical chemistry methods by using patient specimen correlations to assign calibrator values. Arch. Pathol. Lab. Med. 117:412, 1993.
41. Lawton, W.H. Sylvestre, E.A., Young-Ferraro, B.J., Statistical comparison of multiple analytic procedures: application to clinical chemistry. Technometrics. 21:397, 1979.
42. Metz, C.E., Basic principles of ROC analysis. Sem. Nucl. Med. 8:283, 1978.
43. National Committee for Clinical laboratory Standards. Evaluation of the linearity of quantitative analytical methods; proposed guideline, 1986.
44. Lasky, F.D., Achieving accuracy for routine clinical chemistry methods by using patient specimen correlations to assign calibrator values - a means for managing matrix effects. Arch. Pathol. Lab. Med. 117:412-419, 1993.
45. National Committee for Clinical Laboratory Standards. Assessment of the clinical accuracy of laboratory tests using receiver operating characteristic (ROC) plots; tentative guideline, 1993.
46. National Committee for Clinical Laboratory Standards. Method comparison and bias estimation using patient samples; tentative guideline, 1993.
47. National Committee for Clinical Laboratory Standards.

- Interference testing in clinical chemistry; proposed guideline, 1986.
48. National Committee for Clinical Laboratory Standards. Evaluation of precision performance of clinical chemistry devices; tentative guideline, 1992.
  49. National Committee for Clinical Laboratory Standards. Specifications for Immunological Testing for Infectious Diseases; proposed guidelines, 1991.
  50. National Committee for Clinical Laboratory Standards. Labeling of home-use in vitro testing products., proposed guidelines,
  51. Mills, J.L., Data torturing, N.E.J.M., 329:1196-1199, 1993.
  52. Peters, T., Westgard, J.O., Evaluation of methods, Chapter 7 in: Tietz, N.W., ed., Fundamentals of Clinical Chemistry, Third Edition: Philadelphia, Saunders, pp.225-237, 1987.
  53. Points to consider regarding labeling and premarket submissions for home-use in vitro diagnostic devices, 1987, Center for Devices and Radiological Health.
  54. Powers, D.M. Establishing and maintaining performance claims -a manufacturer's viewpoint. Arch. Pathol. Lab. Med. 116:718, 1992.
  55. Poynard, T., Chaput, J., Etienne, J., Relations between effectiveness of a diagnostic test, prevalence of the disease, and percentages of uninterpretable results. Med. Dec. Making. 2:25, 1982.
  56. Ransohoff, D.F., Feinstein, A.R., Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N.E.J.M. 299:926, 1978.
  57. Ross, J.W., Evaluation of Precision. In CRC Handbook of Clinical Chemistry - Volume I. Boca Raton, Florida, CRC Press, pp.391-422, 1982.
  58. Schatzkin, A. Connor, R.J., Taylor, P.R., Bunnag, B., Comparing new and old screening tests when a reference procedure cannot be performed on all screenees. Am. J. Epidemiology. 125:672, 1987.
  59. Staquet, M., Rozenzweig, Lee, Y.J., Muggia, F.M., Methodology for the assessment of new dichotomous diagnostic tests. J. Chron. Dis. 34:599, 1981.
  60. Thibodeau, L.A., Evaluating diagnostic tests. Biometrics. 37:801, 1981.

61. Valenstein, P.V., Evaluating diagnostic tests with imperfect standards. Am. J. Clin. Pathol. 93:252, 1990.
62. Walter, S.D., Irwig, L.M., Estimation of test error rates, disease prevalence, and relative risk from misclassified data: a review. J. Clin. Epidemiol. 41:923-937, 1988.
63. Westgard, J.O., Precision and accuracy: concepts and assessment by method evaluation testing. CRC Crit. Rev. Clin. Lab. Sci. 13:283, 1981.
64. Westgard, J.O., Carey, R.N., and Wold, S., Criteria for judging precision and accuracy in method development and evaluation. Clin. Chem. 20:825, 1974.
65. Westgard, J.O., Hunt, M.R., Use and interpretation of common statistical tests in method-comparison studies. Clin. Chem. 19:49-57, 1973.
66. Zweig, M.H., Campbell, G., Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. 29:561, 1993.