DADE BEHRING
Dade Behring, Inc
P.O. Box 6101
Newark, DE 19714

2923 '03 JUN 11 A9:25

June 10, 2003

Dockets Management Branch (HFA-305)
Food and Drug Administration
5630 Fishers Lane Room1061
Rockville, MD 20852

To Whom It May Concern:

We are sending responses for Docket Number 03D-0044 electronically for your review and consideration. This is the draft guidance titled: Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic tests; Draft guidance for Industry and FDA Reviewers.

We wish to thank you in advance for considering these comments. If you have questions on these responses, please contact me at 302-631-9798 or send a Fax to 302-631-6299.

Sincerely yours,

George M. Plummer
Regulatory Affairs and
Compliance Manager

03D-0044 C 6

Comments on Docket Number 03D-0044: Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests: Draft Guidance for Industry an FDA Reviewers

[illegible]	[illegible]	Suggested Change/Comments/Questions	[illegible]
1	Scope, page 1	It is recommended to clarify within the scope section whether this guidance applies to Add-To-File (ATF) submissions. This guidance appears to pertain to new diagnostic tests and therefore not apply to ATF's. Statements such as "This guidance is not intended for Add-to-File submissions. Refer to FDA's guidance "Data for Commercialization of Original Equipment Manufacturer, Secondary and Generic Reagents for Automated Analyzers, June 10, 1996" are suggested.	This document provides guidance for the submission of premarket notification (510k) and premarket approval (PMA) applications for diagnostic tests.
2	General question (not applicable)	If this guidance is adopted, will it be required to know prevalence and/or true status of the patient (diseased/nondiseased) in addition to the test results with the perfect standard?	Not applicable
3	2nd paragraph, page 4	The FDA should provide a list of suggested perfect standards for the various device branches and products within those branches. Where there are known difficulties with current "perfect standards" in use, such as the NCCLS frozen reference panel for Microbiology (performance varies by broth used), information should be provided on acceptability of use or suggest alternate standards.	From a purely statistical perspective, the best approach is to compare the new test to the patients' clinical status or to a perfect standard using specimens from patients who are representative of the intended use population.
4	3rd paragraph, page 4	Additional guidance or examples should be provided for what is meant by "impractical" in the second bullet point on page 4 ("if a perfect standard is available but impractical, use it to the extent possible") to avoid confusion on FDA expectations.	If a perfect standard is not available but impractical, use it to the extent possible

5	5th paragraph, page 4	This could be less accurate than the imperfect standard. Most likely, it will be redundant with the leading imperfect standard. If the physicians know the multivariate models, then they would probably be using them currently. Additional clarity is needed to understand FDA expectations when constructing such a standard with respect to requirements to show the new test may serve as a "perfect standard". Development of such a standard could take considerable time, require in depth data collection and analysis, and result in significant delays to introduction of new diagnostic products. Further, in a paper in the Journal of the American Statistical Association in 1975, Goldberg provides formulas that allow one to remove the bias in their estimates when sufficient information is available regarding the diagnostic performance of the imperfect standard. These results should be presented in the Appendix. (The effects of misclassification on the bias in the difference between two proportions and the relative odds in the fourfold table. Journal of the American Statistical Association, volume 70, Issue 351, page 561-567	If a perfect standard is not available consider constructing one, page 4
6	4th paragraph, page 5	A list is provided for information that should be included for comparative results. Is this information required in the submissions for both 510k and PMA submissions? Which data is required in the submission versus recommended as supporting documentation in the company technical files? Some information in this list may not be possible for samples obtained from reference laboratories – what is acceptable in this case?	However, all descriptions of comparative results should include a clear description of all methods used, and how and what data were collected. This includes: ----------
7	2nd Paragraph, page 6	What is the purpose of showing the fractions if the 2x2 table is included in customer supplied literature?	All statistical measures such as sensitivity, specificity, and agreement should be reported both as fractions (e.g., 490/500) *and* as percentages (e.g., 98%).

8	Last paragraph, page 6 and 1st paragraph, page 7	Phrasing of the two bolded bullet points in the section titled "Common reporting practices that are statistically inappropriate" is awkward and somewhat difficult to understand.	1. You should not use the terms "sensitivity" and "specificity" to describe the comparison of a new test to an imperfect standard is inappropriate. 2. You should not use results from discrepant resolution alone to estimate the sensitivity and specificity of a new test or agreement between a new test and a comparative method is inappropriate
9	Table 2/paragragh 2, page 10	For the example in Table 2, the estimated specificity is 168/169 = 99.4%. The corresponding 95% confidence interval is reported on page 10 as (96.8%, 100%). From statistical tables, the actual 97.5% upper confidence limit is 99.99%, which was apparently rounded to an even 100%. A true specificity of 100% means that it would be impossible to misdiagnose any of the 169 cases. If this example is retained in the guidance document, it is suggested to use at least one significant figure for p or 1-p. In this case, the upper limit as 99.99% rather than 100%.	Exact 95% confidence intervals (based on the binomial distribution) for sensitivity and specificity are (73.7%, 94.3%) and (96.8%, 100%), respectively
10	Calculating an Estimate of Agreement, page 10 -13	Different methods of statistical calculations and data presentation are suggested here when imperfect standards are used. This may be new information that physicians and laboratory personnel do not have the statistical background to understand. Does the FDA have plans for determining the impact of these changes to these users or for how these users will be provided with appropriate training? If product literature provides information that they do not know how to use or they use it incorrectly, this could have significant negative impact on proper use of products.	See section titled "Calculating and Estimate of Agreement" in the Appendix

11	Last paragraph, page 16	Are additional acceptable options available for resolving discrepancies e.g., adding a non-biased repeat day (where all tests are repeated regardless of their agreement to the reference) to the clinical trial protocol or can specific literature references on this topic be suggested?	In summary, it is not appropriate to revise the original 2x2 table of results based on discrepant resolution because the revision is based on assumptions that aern't verified and usually aren't correct. As a result, it is inappropriate to make sensitivity and specificity type calculations or agreement calculations using the revised table. Instead, FDA recommends reporting the original 2.2 table of results (Table 4), a description of the imperfect standard, an agreement measure and its confidence interval.