Margaret S. Pepe, Ph.D.
Associate Member, Modeling and Methods, Fred Hutchinson Cancer Research Center
Professor, Biostatistics Department, University of Washington
F-672 Health Sciences Building., Box 357232
Seattle, Washington 98195
Phone: (206)543-1044; Email: mspepe@u.washington.edu

May 13, 2003

Dockets Management Branch
Division of Management Systems and Policy
Office of Human Resources and Management Services
Food and Drug Administration
5630 Fishers Lane, Room 1061, (HFA-305)
Rockville, MD 20852

To Whom it May Concern,

Enclosed are some comments regarding the FDA draft guidance document #1428. I hope that they will be helpful to the FDA in finalizing the document

Should you have any questions, feel free to call me at (206)543-1044 or email me at mspepe@u.washington.edu.

Sincerely,

Margaret S. Pepe, Ph.D.

encl: Comments on FOD #1428

03D-0044

C4

Comments on FDA Draft Guidance Document
(FOD #1428)

Overall I found that this was a well written document that provides useful information. However, I think that there are some major omissions and the bibliography needs to be updated. Here are my suggestions for revising the document.

1. Add a technical discussion of verification biased sampling to the document. At several points the document alludes to the fact that estimation is possible when the perfect standard is obtained only for a subset of subjects. It should be clearly stated that the adjusted estimates of sensitivity and specificity require specialized techniques and references to the seminal paper by Begg and Greens (1983) should be provided. A thorough discussion of verification biased sampling, its ramifications and procedures for dealing with it are provided in Pepe (2003, Section 7.1).

2. In the section on General Reporting Recommendations it might be useful to include the following table that summarizes the key classes of bias that commonly occur in studies of diagnostic tests. Manufacturers will want to avoid bias, and this checklist would help them critique their plans for study design.

Type of Bias	Description
Verification bias	Non-random selection for definitive assessment for disease with the gold standard reference test
Errors in the reference	True disease status is subject to misclassification because the gold standard is imperfect
Spectrum bias	Types of cases and controls included are not representative of the population
Test interpretation bias	Information is available that can distort the diagnostic test
Unsatisfactory tests	Tests that are uninterpretable or incomplete do not yield a test result
Extrapolation bias	The conditions or characteristics of populations in the study are different from those in which the test will be applied

From *The Statistical Evaluation of Medical Tests for Classification and Prediction* by Margaret S. Pepe, Oxford University Press, United Kingdom (2003)

3. Add to the last paragraph of the section on 'General Reporting Recommendations' that efforts should be made to identify factors or circumstances that affect the sensitivity and specificity of the diagnostics. Statistical comparisons of sensitivity (or specificity) under 2 or more circumstances are straightforward.

4. There is a major emphasis on the inappropriateness of discrepant resolutions for dealing with the imperfect gold standard problem. Another technique that has been heavily promoted by some statisticians is 'Latent Class Analysis.' This is a technically sophisticated but scientifically flawed approach. See Pepe and Alonzo (2001) for a short illustrative discussion or Pepe (2003, Section 7.3.5) for a thorough discussion of this important topic.

5. Numerical illustrations tend to be more convincing than text in may opinion. Can an illustration be added to demonstrate the 'unscientific and potentially misleading' results that can be obtained when the performance of a new test is established 'by comparing it to a procedure that uses the same new test' (page 8)?

6. I gave a shortcourse to the FDA December 2002. The slides from that course are available to the public via Internet at `www.fhcrc.org/labs/pepe/fdacourse`. Reference to this resource may be helpful to industry in planning and evaluating studies.

7. A glossary of terms would be most helpful.

References

[1] Pepe MS. (2003) *The Statistical Evaluation of Medical Tests for Classification and Prediction* Oxford University Press, United Kingdom.

[2] Pepe MS and Alonzo T (2001) Author's reply: Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Statistics in Medicine* **20**:658–660.

[3] Begg CB and Greenes RA (1983) Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics* **39**:207–215.

Submitted by Margaret S. Pepe (mspepe@u.washington.edu)