Author’s response to reviews

Title: Estimating the Cumulative Risk of False Positive Cancer Screenings

Authors:

Dr Stuart G Baker (sb16i@nih.gov)
Diane Erwin (d.erwin@btinternet.com)
Barnett S Kramer (KramerB@OD.NIH.GOV)

Version: 2 Date: 5 Jun 2003

PDF covering letter
Dr. Yu Shen
We are delighted that Dr. Shen found the work interesting and thank her for the helpful comments.

1. We changed the reference.
2. We left the notation for R precisely to illustrate the point that Dr. Shen makes, namely that if dropout depends only on past events, the essential components are the probabilities of Y.
3. No longer applicable; we dropped the paragraph.
4. No longer applicable; we changed the notation.
5. No longer applicable; we changed the notation.
6. The parameters in the two logistic functions have different meanings, so we cannot estimate them from the single likelihood.
7. We changed the wording.
8. We now describe the unnecessary work-up in HIP.
9. We agree that the underlying trend should be monotonic. However because the standard errors are large, it is not surprising that the point estimates are not monotonic.

Martin C. Mahoney

We thank Dr. Mahoney for the helpful comments.

1. We added age and year to the abstract. Because there are numerous patient characteristics, we recommend that the interested reader look in the references. We now provide the data we used for the analysis, which is stratified by the covariates we used. Screening number refers to the number of the visit after randomization in which the women received mammography and physical examination.

2. A biopsy is unnecessary in the sense that if there were no biopsy, the women's chance of getting cancer would be unchanged. The term "unnecessary biopsy" is also used elsewhere. Therefore we have left the term in the paper.

3. We added a short discussion of the results of fitting a logistic regression to dropout in terms of the covariate of primary interest, previous FP. (The logistic regression also included predictors for age, screening number, and time since last screen; there were no data on race). We think a more detailed discussion would detract from the flow of the paper. In addition, this sort of analysis was reported in reference [5].

4. We have rewritten this section.

5. We thought this was an excellent suggestion! We moved the likelihood formulation to the Appendix.
6. We mention that the expected number of false positives is useful for cost calculations.

7. We restated the proportions in terms of the questions.

8. It is difficult to gauge the impact of new technology or additional experience in reading mammograms. The reason is that the false positive rate could either increase or decrease because of the tradeoff between true and false positives.

9. We mention that one can construct a curve for "surviving" FP to n screens by computing one minus the result in equation (6). The model does not allow extrapolation from data with one-year screening interval to data with a two-year screening interval.

10. We think that for public health and clinical decisions making it is important to accurately estimate the absolute risk of a health outcome or the expected number of health outcomes.

Hans Reitsma

We thank Dr. Reitsma for the helpful comments. Regarding the general remarks, we have moved many technical details in the discussion. With this modification, we think that the paper should appeal to both non-technical and technical readers.

1. (a) We have rewritten the part comparing our approach to that of Gelfand and Wang (GW). We write "We bolster and extend the methodology in GW. First, we show that an unrealistic assumption of GW is unnecessary. GW thought that they needed to assume that dropping out (either by loss-to-follow-up or refusing additional screenings) is independent of the prior history of false positives. This assumption is unrealistic [5] and makes their approach untenable. We show that by reformulating the problem with different random variables, one can obtain essentially the same result without the unrealistic assumption. Second we estimate an additional quantity to that estimated by GW. GW only estimated the probability of at least one false positive in n screenings. To better quantify the cumulative burden of false positives (particularly for economic analysis), we also estimated the expected number of false positives in n screenings. Third, to simplify computations for some data sets, we introduced a logistic regression model."

(b) We did not have the software to implement the Gelfand and Wang model on our data.

2. (a) We think that both the probability of at least one FP and the expected number of FP's are useful to estimate, and the relevance could vary with individuals. For an economic analysis the latter would be more useful.

(b) For the risk of at least one false positive, the main difference is that we use a logistic regression model in a frequentist framework and Gelfand and Wang use a proportional hazards model in a Bayesian framework. We think that our approach is simpler to
implement. But, more importantly, we obviate the unrealistic drop assumption in Gelfand and Wang. We have tried to make this more clear.

3. Rather than provide summary statistics, we present the counts used in the analysis in a separate file (which is easy with BioMed Central) The interested reader can then replicate the analysis or compute a desired quantity.

4. We agree that the assumption would be violated in the proposed scenario. If this scenario were likely, it would be necessary to collect additional data to avoid an incorrect conclusion. One could ask women on screening and women who dropped out if they had detected a lump on self examination. We would then include a variable in the logistic regression for whether or not a lump was detected. In this way dropout is dependent on past history and the logistic regression model would still be valid.