Reviewer's report

Title: Exploring the Uncertainties of Early Detection Results: Model-Based Interpretation of Mayo Lung Project

Version: 1 Date: 21 July 2010

Reviewer: Deborah L Goldwasser

Reviewer's report:

Review of “Exploring the Uncertainties of Early Detection Results: Model-Based Interpretation of Mayo Lung Project”

Summary:
The motivation of this paper is to examine alternative explanations for the Mayo Lung Project (MLP) results that may better explain the MLP trial results than the predominantly cited explanation of over-diagnosis. The Mayo Lung Project had the unusual feature that while an aggressive regimen of screening by chest x-ray increased both lung cancer detection during the duration of the trial and demonstrated a survival benefit of screen-detected cancers, there was no evidence of a mortality reduction attributable to screening. In fact, after long-term follow-up of a median of 20.5 years, there was an excess of 34 lung cancer deaths in the screening arm relative to the control arm. If screening for lung cancer is to be implemented on a broader scale in the future, it is very important to understand the underlying causes of screening trial failures. Mathematical modeling is an important tool in understanding clinical trial results and designing optimal future trials.

Major Criticisms:
The Marcus paper infers that the majority of excess screen-detected cancers represent overdiagnosis on the basis of long-term follow-up data results. This paper incorporates only the initial six years of data as a contribution to the deviance measure. It is unclear whether the initial trial data is alone sufficient to distinguish between the hypotheses under question.

The authors fail to mention the possibility that a risk difference between the two trial arms was accrued during the trial itself as a result of screening. Therefore, the list of alternative hypotheses to over-diagnosis given is not an exhaustive list of possibilities. A recent paper by Goldwasser et al. (2010) in the journal Cancer proposes that a risk difference may be attributable to the excess radiation risk received during the trial. This hypothesis was consistent with individual screen histories and the long-term mortality follow-up data. This paper addresses many of the same issues but is not in the list of citations.

Incorporating over-diagnosis by the addition of a parameter for the detection of indolent cancers should improve the fit to the overall data-set when mortality is considered in addition to incidence. The fact that it did not do so in the model suggests a failure in the methodology and in particular, the choice of the
deviance measure given. The baseline/simple model assumes perfect sensitivity. Given that the sensitivity of chest x-ray is far less than 100%, in order to achieve the goodness of fit demonstrated in table 2, the underlying natural history model must be unrealistic. Therefore, it may not be reasonable to expect that the simple model serves as a valid reference point against which to examine a distinct natural history model that incorporates over-diagnosis.

In the methods section, it would be useful to have a mathematical definition for the deviance in the text as well as a measure of statistical significance for the deviance reduction, in particular for the risk difference model compared to the simple model.

The authors conclude that the trial results are best explained by an initial risk difference between the study arms, in other words, a failure in randomization at the onset of the trial. The probability of randomization failure resulting in a mortality difference as large as that observed can itself be quantified in the context of simulation.

The authors report that individuals in the screening arm smoked on average half a cigarette more per day than individuals in the control arm. This finding is supportive of the authors' suggestion that a risk difference at the onset of the trial may explain the trial results. It would be useful to look at the distribution of cigarettes per day, in particular as a cross-section with age in order to examine whether this difference was sufficient to explain a risk difference as large as the one observed in the MLP.

Minor Criticisms:

There are several typos in the tables and some grammatical errors. In the introduction, it is implied that there is three years of data, when the initial trial lasted for six years with one additional year of follow-up.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests.