What aspect of reader performance are we interested in?

Stephen L. Hillis, PhD, Departments of Radiology and Biostatistics

Rationale

There seems to be no end to the debate over which reader performance estimator should be used for comparing modalities in a diagnostic radiologic imaging study. Should one use a receiver-operating-characteristic (ROC), free-response ROC (FROC), region-of-interest (ROI), or location ROC (LROC) method? If one uses ROC, should one use ROC AUC, pAUC,..., etc.?

Methods

I propose that the reason that this debate never ends is because there is not agreement on what information is being sought. More specifically, if we cannot agree on what parameter (known in statistical jargon as an estimand) we want to estimate, then there can never be agreement on which estimator to use, since the estimators mentioned above estimate different parameters, i.e., different aspects of reader performance. Astonishingly, there is little discussion at this meeting or other radiologic meetings that I've attended as to what information is being sought. A primary purpose of this talk is to encourage such discussion.

Results

To illustrate the importance of knowing what performance information is being sought, I consider three different situations – screening mammography, diagnostic mammography, and CAD – and discuss, from my understanding, what aspect of reader performance is most relevant for each situation. I also discuss which of the estimators mentioned above appears to best provide the desired reader performance information. I suggest a new method that is an extension of LROC that allows the research flexibility in specifying the reader performance aspect of interest and demonstrate this method with an example.

Conclusions

The purpose of this talk is primarily to point out a large deficiency in the field of diagnostic radiology – the absence of discussion as to what makes one reader's performance better than another reader's performance for different situations. Because currently available reader-performance estimators (such as the ROC AUC and the JAFROC statistic) estimate different aspects of reader performance, there cannot be agreement as to which is most appropriate unless

there is agreement on what we want to know. Hopefully, as a result of this talk there will be more discussion regarding what reader performance information is sought for various situations. Once it is clear what information is being sought, then it will be possible determine which of the currently available estimators is most appropriate, or whether there is a need for developing a new estimator that provides more relevant information.