Bayesian Adaptation of the Summary ROC Curve Method for Meta-analysis of Diagnostic Test Performance

Scott W. Miller[1], Debajyoti Sinha[2], Elizabeth H. Slate[1],
Donald Garrow[1] and Joseph Romagnuolo[1]
[1]*Medical University of South Carolina and* [2]*Florida State University*

Abstract: Meta-analytic methods for diagnostic test performance, Bayesian methods in particular, have not been well developed. The most commonly used method for meta-analysis of diagnostic test performance is the Summary Receiver Operator Characteristic (SROC) curve approach of Moses, Shapiro and Littenberg. In this paper, we provide a brief summary of the SROC method, then present a case study of a Bayesian adaptation of their SROC curve method that retains the simplicity of the original model while additionally incorporating uncertainty in the parameters, and can also easily be extended to incorporate the effect of covariates. We further derive a simple transformation which facilitates prior elicitation from clinicians. The method is applied to two datasets: an assessment of computed tomography for detecting metastases in non-small-cell lung cancer, and a novel dataset to assess the diagnostic performance of endoscopic ultrasound (EUS) in the detection of biliary obstructions relative to the current gold standard of endoscopic retrograde cholangiopancreatography (ERCP).

Key words: Bayesian, biliary system, ERCP, EUS, SROC.

1. Introduction

Diagnostic tests are widely used in medicine to determine disease status, e.g. to determine whether a disease is present or absent. Commonly the test is based on an underlying (perhaps latent) continuous outcome for which values above a specified threshold are regarded as indicative of disease. As the decision threshold varies, there is a trade-off between sensitivity and specificity. To visualize the effect of threshold on the estimated sensitivity and specificity, investigators often plot a Receiver Operator Characteristic (ROC) curve, where the probability of true positives (Sensitivity) is plotted on the y-axis vs the probability of false positives (1-Specificity) on the x-axis.

One way to summarize the performance of a diagnostic test from multiple studies is by an average sensitivity and specificity. Such summaries can be misleading, however, if there is heterogeneity among the studies, and, unfortunately, tests used to detect heterogeneity lack power (Midgette *et al.*, 1993). In addition, because the sensitivity and specificity within a study are inversely related and depend on the threshold, using the average sensitivity and average specificity as summary statistics is not likely to be an adequate representation of the data (Irwig *et al.*, 1995; Vamvakas, 1998; Pepe, 2003).

This dependence on the threshold used for the test remains a concern even when the outcome is dichotomous, e.g. "present" or "absent." Different studies are likely to vary in what constitutes an abnormal reading, and failure to account for this implicit difference may bias the results of any meta-analysis (Irwig *et al.*, 1995). To correct for this potential bias, Moses *et al.* (1993) developed the technique of calculating a Summary ROC (SROC) curve, which relates the test threshold to the test accuracy via a linear regression, as described in Section 2.2.

The Bayesian framework has a number of advantages for meta-analyses: prior information is explicitly represented and included in inferences; uncertainty resulting from both the prior and sampling distributions is duly propagated through the model to posterior inferences; and the posterior predictive distribution provides a convenient summary of predictions for a new study considered exchangeable with those in the meta-analysis. Rutter and Gatsonis (2001) proposed a Bayesian Hierarchical SROC (HSROC) model that allows each study to have its own accuracy and threshold; at the cost, however, of several additional parameters, each of which requires a prior distribution.

While the Rutter and Gatsonis method may be preferable in more complex settings, such as asymmetric SROC curves, or where inter-study heterogeneity cannot be reasonably ignored, in this paper we propose a Bayesian adaptation of the SROC curve approach of Moses *et al.* which retains the simplicity of the their model for situations where it is appropriate. In Section 2, we summarize the SROC method. In Section 3 we propose a Bayesian adaptation. In Section 4, we apply the Bayesian approach to two examples and compare the results with those from a traditional analysis. The first example is a dataset assessing computed tomography in the detection of metastases in non-small-cell lung cancer. The second example compares two diagnostic procedures in the detection of biliary obstructions. We conclude with discussion in Section 5.

2. Summary ROC Model

2.1 Notation

We adopt notation similar to that of Moses *et al.* (1993), in which each of

$i = 1, 2, \ldots, m$ studies examining the same diagnostic procedure contributes a vector of data in the form $\left\{ y_{11}^{(i)}, y_{10}^{(i)}, y_{01}^{(i)}, y_{00}^{(i)} \right\}$. Here $y_{jk}^{(i)}$ is the count of subjects in the i-th study in which the diagnostic test outcome is indicated with j and the true disease status of the subject by k, with $j, k = 0$ or 1 according to whether the outcomes are negative or positive. The m studies and the subjects within each study are assumed to be independent. Study-specific estimates of sensitivity and specificity are computed from these data as $\hat{Q}_i = y_{11}^{(i)} / \left(y_{11}^{(i)} + y_{01}^{(i)} \right)$ and $1 - \hat{P}_i = y_{00}^{(i)} / \left(y_{10}^{(i)} + y_{00}^{(i)} \right)$, respectively. To avoid the potential problem of zero cells, Moses *et al.* recommend adding 0.5 to all cells in all studies prior to the calculation of $\hat{Q}_i$ and $1 - \hat{P}_i$. This empirical adjustment biases the results towards a Q^* value of 0.5; equivalent to a diagnostic test that is no better than chance. However, this bias is typically small, even in studies with a small number of subjects and high sensitivity and/or specificity (over 80%). For example, the underestimation of Q^* is approximately 2% for a study with 99% sensitivity in a simulation study with moderate sample size (Mitchell, 2003).

2.2 The Summary ROC curve

First define the transformations

$$U_i = \text{logit}\,(1 - \text{Specificity}) = \ln\left[\hat{P}_i / \left(1 - \hat{P}_i \right) \right] \quad \text{and}$$
$$V_i = \text{logit}\,(\text{Sensitivity}) = \ln\left[\hat{Q}_i / \left(1 - \hat{Q}_i \right) \right].$$

The SROC method then defines two new variables, $S_i = V_i + U_i$ and $D_i = V_i - U_i$, where S_i represents a measure of the threshold used, and D_i the diagnostic accuracy of the test, for the ith study. The SROC arises from a postulated linear regression between D_i and S_i,

$$D_i | S_i = \alpha + \beta S_i + \varepsilon_i, \tag{2.1}$$

with ε_i independently distributed normal random variables, $i = 1, 2, \ldots, m$. Estimates of the regression parameters α and β are obtained using either ordinary least squares regression, with all studies weighted equally, or using weighted regression, e.g. in which studies are assigned weights inversely proportional to the variance of the log of the diagnostic odds ratio of the study.

Once the parameter estimates $\hat{\alpha}$ and $\hat{\beta}$ are obtained, the corresponding SROC curve can be generated from the fit to equation (2.1) by evaluating the transformation

$$Q_i = \left\{ 1 + e^{-\alpha/(1-\beta)} \left[(1 - P_i) / P_i \right]^{(1+\beta)/(1-\beta)} \right\}^{-1}$$

over $P_i \in (0, 1)$ and then plotting Q_i vs $1 - P_i$

Moses *et al.* recommend that studies with false positive rates over 50% or true positive rates under 50% be excluded from the regression, as these are likely to exert undue influence on the regression coefficients, and, furthermore, these studies failed to exhibit test results sufficiently accurate for clinical use. Such exclusion has the effect of biasing the resulting SROC curve towards favorable performance (i.e. the upper left quadrant) and has therefore drawn criticism (Irwig *et al.*, 1995).

Although the area under the curve (AUC) is a common method for summarizing an ROC curve, its application to the SROC is controversial. Moses *et al.* caution against evaluating the AUC of the SROC, noting that the SROC should be evaluated only on the range supported by the underlying studies included when fitting equation (2.1). Other authors also express concern about the validity of the AUC in this context (Walter, 2002; Walter, 2005). Those who question the AUC propose another summary statistic, Q^*, which is defined as the point on the SROC curve where the sensitivity and specificity are equal, with values closer to one representing a better test. Moses *et al.* (1993) propose using a two-sample z test to compare two different tests based on their Q^* values.

The SROC model in equation (2.1) extends to allow for study-specific covariates, Z_i, that could potentially explain differences in test accuracy:

$$D_i | S_i, Z_i = \alpha + \beta S_i + \gamma Z_i + \varepsilon_i.$$

However, Vamvakas (1998) points out that, due to the small number of studies in a typical meta-analysis, generally only 1 to 2 covariates should be included to avoid over-fitting.

Note that in equation (2.1), if the choice of threshold does not impact sensitivity and specificity, then $\beta = 0$ and α is the common diagnostic odds ratio (DOR, i.e. the log odds ratio of the accuracy) of the test. In this situation, there are other ways to estimate this common DOR, e.g. the Mantel-Haenszel method, but these generally do not account for between-study variability and hence yield narrower confidence intervals than is warranted (Moses *et al.*, 1993).

3. Bayesian Formulation

The parameters in equation (2.1) are assumed to be normally distributed, so a bivariate normal prior can be assigned to $\begin{pmatrix} \alpha & \beta \end{pmatrix}$ to implement the model in a Bayesian setting.

$$D_i | S_i, \sigma_i^2 \sim N\left(\alpha + \beta S_i, \sigma_i^2\right) \tag{3.1}$$

$$1/\sigma_i^2 \sim Ga\left(0.001, 0.001\right)$$

with the regression parameters assigned a bivariate normal distribution

$$\left(\begin{array}{cc} \alpha & \beta \end{array}\right)^T \sim N_2\left(\mu, \Sigma\right)$$

where $\mu = \left(\begin{array}{cc} \mu_\alpha & \mu_\beta \end{array}\right)^T$ and $\Sigma = \left[\begin{array}{cc} \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_\beta^2 \end{array}\right]$.

Clinical information could be incorporated into the prior for the intercept, with the prior for the slope of the threshold effect centered at zero.

3.1 Covariate extension

The model can be easily extended to incorporate clinically relevant covariates which could potentially improve the predictive ability of the test. For example, it may be that there are differences between how the different studies incorporated into the meta-analysis were designed, and some of these design differences may affect the observed accuracy. Accounting for these covariates could result in an improved estimate of the underlying accuracy of the diagnostic test.

Modifying equation (3.1) by adding the parameter(s) γ to represent the effect of covariates yields

$$D_i|S_i, Z_i, \sigma_i^2 \sim N\left(\alpha + \beta S_i + \gamma Z_i, \sigma_i^2\right)$$

The priors can then be modified by changing the bivariate normal distribution used above by increasing the number of dimensions to the appropriate multivariate normal distribution.

3.2 Prior elicitation

Bayesian implementation of this model requires prior distributions to be placed on both the intercept and the slope of the SROC curve. For the intercept, this requires the clinician to express a point estimate and some level of certainty in terms of the log diagnostic odds ratio, which may not be an intuitive scale for clinical investigators. A more intuitive method is to take advantage of the fact that Q^* is a 1:1 transformation of the intercept in the SROC method, calculated via $Q^* = \left[1 + e^{-\hat{\alpha}/2}\right]^{-1}$.

Rearranging to solve for α yields

$$\hat{\alpha} = w\left(Q^*\right) = -2\ln\left(Q^{*-1} - 1\right) \tag{3.2}$$

The distribution of Q^* can be shown to be

$$g\left(Q^*\right) = \begin{cases} \frac{2}{Q^*(1-Q^*)\sigma_\alpha\sqrt{2\pi}} \exp\left\{\frac{-1}{2\sigma_\alpha^2}\left[-2\ln\left(Q^{*-1}-1\right)-\mu_\alpha\right]^2\right\} & 0 < Q^* < 1 \\ 0 & \text{otherwise} \end{cases}$$
$$\tag{3.3}$$

Having elicited the most likely value for Q^* from clinicians, say Q^{**}, we set the prior mean for α as $\mu_\alpha = w(Q^{**})$ from equation (3.2). We then further elicit a series of upper percentiles for Q^*, e.g. Q_p for $p = 0.90, 0.95$, and then set priorthen set the prior standard deviation for α as approximately satisfying $\sigma_\alpha = (w(Q_p) - \mu_\alpha)/z_p$, where z_p is the pth percentile of the standard normal distribution.

As the SROC curve is *a priori* expected to be symmetric, β is *a priori* expected to be close to zero, so μ_β would be set to zero with σ_β^2 selected such that only values close to zero are likely. The regression parameters α and β are assumed *a priori* to be independent. Covariates are easily accommodated by incorporating them into the mean in equation (3.1) and adopting a multivariate normal prior for the corresponding regression parameters. Again, prior clinical information can be incorporated here.

3.3 Priors

Before examining the data, three priors (diffuse, skeptical and enthusiastic) were selected in order to represent a range of clinically reasonable beliefs. All three are normal distributions on the intercept (α) of the model, with mean and variance (0,12.96), (0,0.5625) and (2.5,2.25), respectively. Using the distribution of Q^* in equation (3.3), the prior belief in terms of α can be transformed into the corresponding clinical beliefs in terms of Q^*; these are presented graphically in Figure 1.
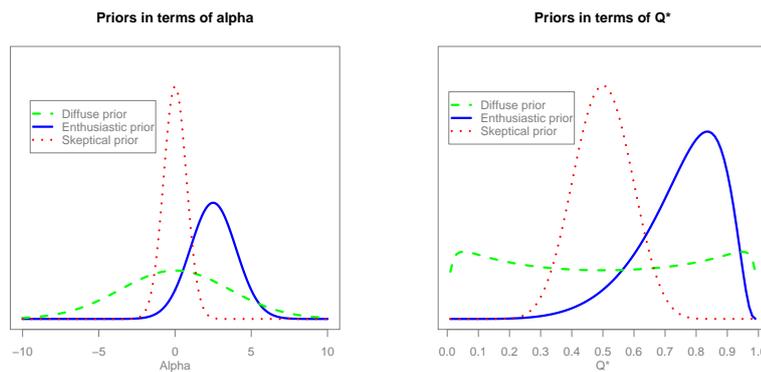
Figure 1: Diffuse, skeptical and optimistic priors used in this analysis: Left panel: in terms of α; right panel: corresponding priors in terms of Q^*.

Clearly, the distribution of Q^* is skewed when not centered at 0.5, as would be expected given the confinement to the unit interval. Interestingly, the distribution of Q^* is sensitive to the variance used. Using a standard deviation of

3.6 with a mean at 0.5 results in a fairly flat distribution over the whole range; increasing the standard deviation beyond this value causes the distribution to become increasingly bi-modal with most of the density at the extremes and little in the center.

3.4 Implementation

The Bayesian analysis was implemented using WinBUGS version 1.4. Briefly, a bivariate normal prior was assigned to the intercept and slope of the regression model. Three over-dispersed chains were run for 5,000 iterations each. After discarding the first 2,000 iterations from each chain, convergence was assessed via examination of $\hat{R}$, trace, history and quantiles, and the resulting data used for inference. This process was sufficient to reach convergence for all three sets of priors used in both of the examples examined.

The traditional approach (henceforth referred to as the frequentist approach) was implemented using SAS version 9.1 using a macro written for this purpose. Graphs were generated using R. Both the SAS macro as well as the WinBUGS code are available from the first author upon request.

Table 1: Example 1: data from Inouye and Sox on computed tomography in non-small-cell lung cancer. TP: true positives, FP: false positives, FN: false negatives, TN: true negatives. *: Study 79 was omitted by Moses et al

Study	TP	FN	FP	TN
78	11	7	6	26
79*	2	5	15	13
80	15	2	2	32
81	4	1	4	13
82	22	21	1	6
83	17	1	9	15
84	29	10	1	54
85	8	6	4	23
86	7	6	12	25
87	20	1	10	18
88	19	1	9	19
89	18	6	8	65
90	18	1	9	13
91	17	3	6	49

4. Applications

We illustrate the proposed model via two examples: an analysis of 13 studies examining computed tomography in the detection of metastases in non-small-cell

lung cancer and an analysis of 35 studies examining the performance of endoscopic ultrasound relative to endoscopic retrograde cholangiopancreatography in the context of detecting biliary obstructions. The first example is provided to facilitate comparison of the proposed model with the results obtained in Moses et al; the second example is the motivating dataset, and demonstrates the extension of this method to include a covariate.

Point estimates and 95% intervals
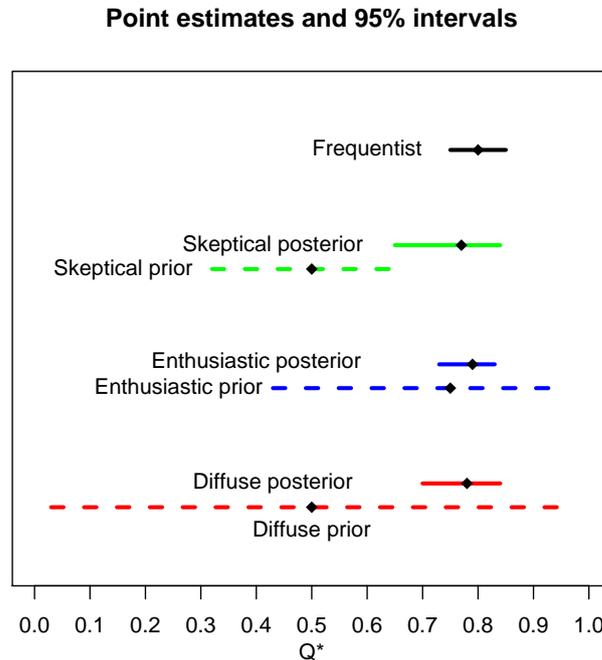
Figure 2: Graphical representation of the point estimates and 95% intervals for each of the analyses conducted (Bayesian with diffuse, enthusiastic or skeptical priors and frequentist) for detection of metastases in non-small-cell lung cancer using computed tomography from Example 1

4.1 Example 1: Non-small-cell lung cancer

Moses *et al.* (1993) use as an example a dataset, first presented by Inouye and Sox (1986), assessing computed tomography (CT) in the detection of metastases in non-small-cell lung cancer. The data are presented in Table 1.

As can be seen in Figure 2, regardless of which prior is used, all three of the posterior point estimates for Q^* are pulled slightly toward 0.5 relative to the frequentist intervals. However, despite the disparity in initial beliefs reflected in

the three priors used, all three posteriors have relatively strong agreement on the range of reasonable values for Q^*, indicating that the data are reasonably robust to a broad range of possible clinical beliefs.

Despite the fact that the skeptical prior had the narrowest credible interval of the three priors used, the corresponding posterior credible interval is the widest of the three. This, combined with the observation that the skeptical posterior point estimate for Q^* is closer to the frequentist Q^* than to the skeptical prior point estimate, suggests that the skeptical prior is inconsistent with the observed data.

4.2 Example 2: Biliary obstructions

The second example examines the detection of biliary obstructions with endoscopic ultrasound. 35 studies examining this question were identified using the search criteria set forth by Irwig *et al.* (1994). Briefly, a MEDLINE search was conducted to locate studies relevant to the detection of biliary obstructions using EUS. The data were then extracted from these studies, and 2×2 tables generated for each; for details, see Garrow *et al.* (2007).

The biliary system is the set of tubes connecting the gallbladder to the liver, pancreas and the rest of the digestive system. Under certain circumstances, obstructions can block the small tubes that compose the biliary system. These obstructions can cause problems, such as pancreatitis, cholangitis or cholecystitis (inflammation of the pancreas, bile ducts or gallbladder, respectively.)

Three main methods are commonly used to image the biliary system: endoscopic retrograde cholangiopancreatography (ERCP), considered the gold standard; magnetic resonance cholangiopancreatography (MRCP), a less invasive method; and endoscopic ultrasound (EUS).

Although ERCP is currently considered the gold standard for biliary visualization, and the rate of complications is low, around 5% (Lahmann *et al.*, 2004), several potentially serious and possibly life-threatening events are possible. As a result, it would be preferable for certain subgroups of patients to undergo the safer procedure of EUS and perform an ERCP only if needed therapeutically.

The purpose of this meta-analysis was to compare the diagnostic performance of EUS to ERCP (MRCP has been compared to ERCP in a previous meta-analysis.)

Table 2 displays the studies used in this analysis, as well as whether those interpreting the second test were blinded as to the results from the first test.

Table 2: Example 2 data for the meta-analysis of EUS performance in diagnosing the cause of biliary obstruction. TP: true positives, FP: false positives, FN: false negatives and TN: true negatives. Blinded is an indicator set to 1 if the reader of the second test was blinded to the results from the first test.

Study	TP	FP	FN	TN	blinded
1	33	2	1	14	1
2	42	0	1	23	1
3	88	1	11	63	1
4	88	1	11	63	1
5	16	6	4	34	0
6	6	1	2	19	1
7	32	1	1	83	1
8	33	1	1	65	1
9	49	0	2	91	1
10	74	1	4	44	0
11	11	0	1	19	1
12	60	14	11	227	0
13	85	2	6	41	1
14	13	1	0	54	0
15	237	2	5	215	0
16	32	0	1	29	0
17	74	1	5	39	1
18	21	1	3	25	1
19	37	10	7	161	0
20	32	4	1	13	1
21	15	1	3	45	1
22	10	1	0	21	0
23	25	0	3	104	1
24	15	0	1	25	1
25	24	2	0	2	1
26	4	0	1	50	1
27	31	0	3	16	1
28	12	0	16	22	1
29	16	4	4	16	1
30	11	0	12	12	0
31	28	2	9	6	1
32	10	1	1	28	1
33	45	4	2	11	0
34	42	0	0	8	0
35	101	0	1	30	0

The results from each of the three priors used (diffuse, skeptical and enthusiastic) as well as the corresponding results from the traditional approach, are displayed graphically in Figure 3. Despite an initial wide range of clinical opinions reflected in the priors used, all three posteriors are in near total agreement.

Point estimates and 95% intervals
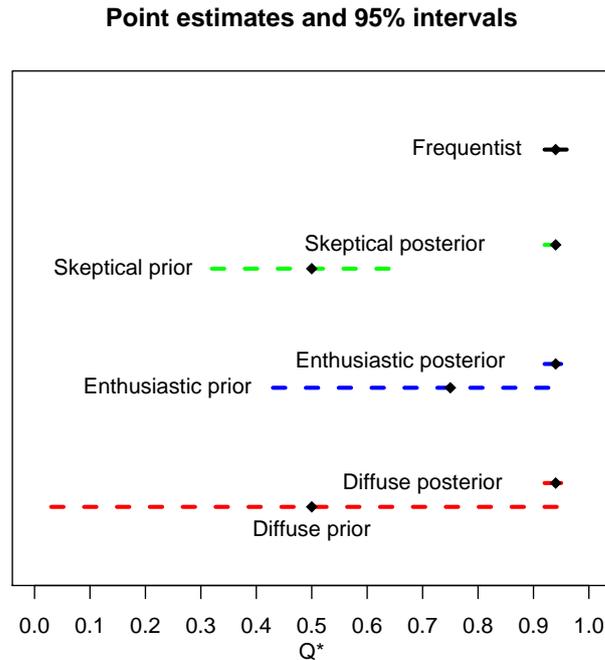
Figure 3: Graphical representation of the point estimates and 95% intervals for each of the analyses conducted (Bayesian with diffuse, enthusiastic or skeptical priors and frequentist) for detection of obstruction with EUS relative to ERCP from Example 2.

5. Discussion

For both datasets examined, all three posterior intervals are in high agreement with one another. The priors appear to have more influence on the results from the CT data in example 1, presumably due to the smaller number of studies included (11 vs 35 in the EUS example). The similarity between the posteriors for both studies suggest that both datasets are sufficiently robust to a range of clinically reasonable priors, and are pulling all observers into agreement. Clinically, these results suggest that CT is moderately effective in detecting metastases in non-small-cell lung cancer, and that EUS compares favorably with ERCP in the detection of biliary obstruction.

We have presented a Bayesian adaptation of the SROC method of Moses *et al.* (1993). The Bayesian adaptation converges quickly and yields results similar to the frequentist method. A simple transformation of Q^* was presented which

Table 3: Results of 3 different priors for detection of obstruction data using blinding of the second reader to the results of the first test result as a covariate. Shown are the mean of the prior and corresponding posterior, as well as the 95% credible interval width.

Analysis	Blinded	Mean	LB	UB	Width
Prior (diffuse)	0	0.483	0.030	0.975	0.945
	1	0.505	0.006	0.994	0.988
Posterior (diffuse)	0	0.932	0.900	0.956	0.056
	1	0.932	0.900	0.956	0.056
Prior (skeptical)	0	0.502	0.336	0.673	0.337
	1	0.508	0.017	0.984	0.967
Posterior (skeptical)	0	0.932	0.900	0.957	0.057
	1	0.938	0.915	0.957	0.042
Prior (enthusiastic)	0	0.756	0.449	0.945	0.496
	1	0.676	0.052	0.996	0.944
Posterior (enthusiastic)	0	0.929	0.893	0.954	0.061
	1	0.938	0.914	0.957	0.043
Frequentist	0	0.931	0.897	0.965	0.068
	1	0.916	0.888	0.944	0.056

facilitates prior elicitation. Some suggestions on example priors for use in prior sensitivity analysis were also presented.

An advantage of the proposed model is that it allows investigators to explicitly model how strongly they believe that the SROC curve is symmetric. As the use of the SROC method strongly implies this belief, the ability to quantify the strength of this belief allows investigators to avoid moving from this method to a more complicated one based solely on a single p-value from the hypothesis test of the significance of the slope parameter β.

As the proposed model is a Bayesian adaptation of the SROC method, it retains several of the limitations of SROC curves. Firstly, SROC curves of this type implicitly assume that the best summary curve of the data will be symmetric; there are cases where this assumption is invalid. However, the simplicity of the model makes it preferable to more complex models.

A further issue is the use of the Q^* as a summary statistic. As Q^* represents the point on the summary curve where the sensitivity and specificity are equal, its use implies that the risks of a true and false positive test are of equal importance. This is perhaps less of a concern than it would first appear, as this can be

expanded as follows.

Table 3 shows how the results of EUS detection of obstruction are affected by including blinding as a covariate (studies were coded as 1 if the reader of the second test result in the ith study was blinded to the results of the first test, and 0 if not.) Of the 35 studies, 23 (66%) were blinded and 12 were not.

The posterior distributions for all three priors for blinding as a covariate are in agreement that blinded assessment of the second reader does not appear to significantly impact the diagnostic accuracy of EUS in these studies.

Instead of writing Q^*, express this summary statistic as Q_p, where the subscript p represents an angle from the bottom right corner of the SROC graph to the upper left corner. In this notation, Q^* would be Q_{45}, which represents the setting where both sensitivity and specificity are of equal importance. In settings where one is more relevant, e.g. for a screening test, in which a lower specificity might be tolerated in exchange for a higher sensitivity, this number could be adjusted to reflect this change.

Future research on this model would include adapting it to allow for an Empirical Bayes analysis approach. In addition, some tests are not categorized into two groups, but three: e.g. "normal," "abnormal" and "indeterminate". Accounting for this could also be of clinical interest. Finally, existing methods assume a perfect gold standard; accounting for an imperfect gold standard could also be of interest.

Appendix: Derivations

The following derivations of the equation for the SROC curve and Q^* are reproduced here from the paper by Moses *et al.* (1993). The formula for the SROC curve is derived by Moses *et al.* in their Appendix, p1312-1313; it is reproduced here for ease of reference.

Derivation of the SROC curve equation

Recall that the model defines the sum and difference of the logit-transformed true and false positive rates;

$$U = \ln\left[P/\left(1 - P\right)\right] \text{ and } V = \ln\left[Q/\left(1 - Q\right)\right]$$

$$D = V - U \text{ and } S = V + U$$

Then fit the simple linear model

$$D = \alpha + \beta S$$

Their derivation of the SROC formula is then:

$$V - U = \alpha + \beta \left(V + U \right)$$

rearranging to get like terms on the same side yields

$$V \left(1 - \beta \right) = \alpha + U \left(1 + \beta \right)$$

$$V = \alpha / \left(1 - \beta \right) + U \left(1 + \beta \right) / \left(1 - \beta \right)$$

$$\ln \left[Q / \left(1 - Q \right) \right] = \alpha / \left(1 - \beta \right) + \ln \left[P / \left(1 - P \right) \right] \left(1 + \beta \right) / \left(1 - \beta \right)$$

$$\left[Q / \left(1 - Q \right) \right] = \exp \left[\alpha / \left(1 - \beta \right) \right] \left[P / \left(1 - P \right) \right]^{(1+\beta)/(1-\beta)}$$

$$Q = \left[1 + \exp \left[-\alpha / \left(1 - \beta \right) \right] \left[\left(1 - P \right) / P \right]^{(1+\beta)/(1-\beta)} \right]^{-1}$$

Derivation of Q^*

Start by noting that Q^* is the location on the SROC curve which intersects the diagonal $P + Q = 1$. Thus $P = 1 - Q$.

$$\text{logit} \left(P \right) = \text{logit} \left(1 - Q \right) = -\text{logit} \left(Q \right)$$

Recall

$$U = \ln \left[P / \left(1 - P \right) \right] \ \text{and} \ V = \ln \left[Q / \left(1 - Q \right) \right]$$

Therefore, $U = -V$, so $S = U + V = 0$.

In (S, D) space, the relationship is linear, with $D = \alpha + \beta S$. Plugging in $S = 0$ and solving yields $D = \alpha$, implying that $D = V - U = \alpha$.

From above, we know that $U = -V$, so $2V = \alpha$, or, equivalently, $\ln \left[Q / \left(1 - Q \right) \right] = \alpha/2$.

Rearranging yields $Q / \left(1 - Q \right) = e^{\alpha/2}$, leading to the formula for Q^*; $Q^* = \left[1 + e^{-\alpha/2} \right]^{-1}$.

References

Garrow, D., Miller, S., Sinha, D., Hoffman, B., Hawes, R., Conway, J. and Romagnuolo, J. (2007). Endoscopic ultrasound: a meta-analysis of test performance in suspected biliary disease. *Clinical Gastroenterology and Hepatology* **5**, 616-623.

Inouye, S. and Sox, H. (1986). Standard and computed tomography in the evaluation of neoplasms of the chest: a comparative efficacy assessment. *Annals of Internal Medicine* **105**, 906-924.

Irwig, L., Macaskill, P., Glasziou, P. and Fahey, M. (1995). Meta-analytic methods for diagnostic test accuracy. *Journal of Clinical Epidemiology* **48**, 119-30; discussion 131-2.

Irwig, L., Tosteson, A. N., Gatsonis, C., Lau, J., Colditz, G., Chalmers, T. C. and Mosteller, F. (1994). Guidelines for meta-analyses evaluating diagnostic tests. *Annals of Internal Medicine* **120**, 667-676.

Lahmann, B. E., Adrales, G. and Schwartz, R. W. (2004). Choledocholithiasis-principles of diagnosis and management. *Current Surgery* **61**, 290-293.

Midgette, A., Stukel, T. and Littenberg, B. (1993). A meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. *Medical Decision Making* **13**, 253-257.

Mitchell, M. D. (2003). Validation of the summary ROC for diagnostic test meta-analysis: a Monte Carlo simulation. *Academic Radiology* **10**, 25-31.

Moses, L. E., Shapiro, D. and Littenberg, B. (1993). Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Statistics in Medicine* **12**, 1293-1316.

Pepe, M. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*, first edition, Oxford University Press.

Rutter, C. M. and Gatsonis, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine* **20**, 2865-2884.

Vamvakas, E. C. (1998). Meta-analyses of studies of the diagnostic accuracy of laboratory tests: a review of the concepts and methods. *Archives of pathology and laboratory medicine* **122**, 675-686.

Walter, S. D. (2002). Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Statistics in Medicine* **21**, 1237-1256.

Walter, S. D. (2005). The partial area under the summary ROC curve. *Statistics in Medicine* **24**, 2025-2040.

Scott Miller
Division of Biostatistics, CDRH
Food & Drug Administration
10903 New Hampshire Ave
Building 66, Room 2321
Silver Spring, MD 20993-0002
Scott.Miller@fda.hhs.gov

Debajyoti Sinha
Florida State University
Department of Statistics
214 OSB
117 N. Woodward Avenue
PO Box 3064330
Tallahassee, FL USA 32306-4330
sinhad@stat.fsu.edu

Elizabeth H. Slate
Medical University of South Carolina
Department of Biostatistics, Bioinformatics and Epidemiology
135 Cannon Street, suite 303
Charleston,SC USA 29425
slateeh@musc.edu

Don Garrow
Medical University of South Carolina
Department of Gastroenterology and Hepatology
96 Jonathan Lucas Street, room 210
Clinical Science Building
Charleston,SC USA 29425
garrowd@musc.edu

Joseph Romagnuolo
Medical University of South Carolina
Department of Gastroenterology and Hepatology
96 Jonathan Lucas Street, room 210
Clinical Science Building
Charleston,SC USA 29425
romagnuo@musc.edu