A unification of models for meta-analysis of diagnostic accuracy studies

ROGER M. HARBORD*

*Medical Research Council Health Services Research Collaboration,
Department of Social Medicine, University of Bristol, Canynge Hall,
Whiteladies Road, Bristol BS8 2PR, UK*
roger.harbord@bristol.ac.uk

JONATHAN J. DEEKS

Centre for Statistics in Medicine, Oxford, UK

MATTHIAS EGGER

Department of Social and Preventive Medicine, University of Berne, Switzerland

PENNY WHITING, JONATHAN A. C. STERNE

*Medical Research Council Health Services Research Collaboration,
Department of Social Medicine, University of Bristol, UK*

SUMMARY

Studies of diagnostic accuracy require more sophisticated methods for their meta-analysis than studies of therapeutic interventions. A number of different, and apparently divergent, methods for meta-analysis of diagnostic studies have been proposed, including two alternative approaches that are statistically rigorous and allow for between-study variability: the hierarchical summary receiver operating characteristic (ROC) model (Rutter and Gatsonis, 2001) and bivariate random-effects meta-analysis (van Houwelingen *and others*, 1993, 2002; Reitsma *and others*, 2005). We show that these two models are very closely related, and define the circumstances in which they are identical. We discuss the different forms of summary model output suggested by the two approaches, including summary ROC curves, summary points, confidence regions, and prediction regions.

Keywords: Bivariate normal distribution; Diagnostic tests; Hierarchial models; HSROC model; Meta-analysis; ROC analysis; Sensitivity and specificity.

1. INTRODUCTION

There is increasing interest in systematic reviews and meta-analyses of data from diagnostic accuracy studies (Deeks, 2001; Deville *and others*, 2002; Bossuyt *and others*, 2003; Khan *and others*, 2003;

---

*To whom correspondence should be addressed.

Whiting *and others*, 2004; Tatsioni *and others*, 2005; Gluud and Gluud, 2005). Typically, the data from each primary study are summarized as a 2 × 2 table, based on dichotomized test result against true disease status, from which familiar measures such as sensitivity and specificity can be derived.

Several statistical methods for meta-analysis of data from diagnostic test accuracy studies have been proposed (Moses *and others*, 1993; Rutter and Gatsonis, 2001; Dukic and Gatsonis, 2003; Siadaty and Shu, 2004; Reitsma *and others*, 2005). These methods reflect two important characteristics of such data. First, a negative correlation between sensitivity and specificity is expected because of the trade-off between these measures as the test threshold varies (Moses *and others*, 1993; Deeks, 2001). Second, and in contrast to meta-analysis of data from randomized controlled trials, substantial between-study heterogeneity is to be expected and must be incorporated in the models (Lijmer *and others*, 2002). The inferential focus of these methods is also a matter of debate. Some authors propose estimating summary measures of sensitivity and specificity, or prediction regions within which we may expect the results of a future study to lie (Reitsma *and others*, 2005), while others suggest that in the presence of substantial heterogeneity, the results of meta-analyses should be presented as summary receiver operating characteristic (SROC) curves (Rutter and Gatsonis, 2001).

Littenberg and Moses (1993) (see also Moses *and others*, 1993) proposed a method of generating a SROC curve using simple linear regression that has been frequently used. However, the assumptions of simple linear regression are not met and the method is therefore approximate. There is also uncertainty as to the most appropriate weighting of the regression (Walter, 2002; Rutter and Gatsonis, 2001).

Two statistically rigorous methods for the meta-analysis of data from diagnostic test accuracy studies have been proposed (Reitsma *and others*, 2005; Rutter and Gatsonis, 2001) that overcome these problems but are necessarily more complex. In this paper, we review the characteristics of these methods. We show that although these have been discussed as alternative ways to analyze such data, they are equivalent in many circumstances and hence often lead to identical statistical inferences. Section 2 describes the bivariate model, while Section 3 describes the hierarchical summary receiver operating characteristic (HSROC) model. In Section 4, we explain the relationship between these two models. In Section 5, we discuss the different focus of inference and presentation of model estimates suggested by the two parameterizations. A worked example is presented in Section 6, and the implications of the work are discussed in Section 7.

2. The bivariate model

The bivariate model is based on an approach to meta-analysis introduced by van Houwelingen *and others* (1993) (see also van Houwelingen *and others*, 2002). It has recently been applied to meta-analysis of diagnostic accuracy studies by Reitsma *and others* (2005).

Following Reitsma *and others* (2005), we define μ_{Ai} as the logit-transformed sensitivity in study i, and μ_{Bi} as the logit-transformed specificity. We use the letter μ where Reitsma *and others* (2005) used θ to avoid a clash of notation with the HSROC model defined in Section 3. The bivariate model is a random-effects model in which the logit transforms of the true sensitivity and true specificity in each study are assumed to have a bivariate normal distribution across studies, thereby allowing for the possibility of correlation between them (Reitsma *and others*, 2005):

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \Sigma_{AB} \right) \quad \text{with } \Sigma_{AB} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}. \tag{2.1}$$

Covariates that affect either sensitivity or specificity or both can be included in a natural way by replacing one or both of the means μ_A and μ_B by linear predictors in the covariates. For example, for a single covariate Z that may affect both sensitivity and specificity, we could replace μ_A by $\mu_A + \nu_A Z_i$ and μ_B by $\mu_B + \nu_B Z_i$.

3. THE HSROC MODEL

The HSROC model (Rutter and Gatsonis, 2001) was motivated by a model for ordinal regression (McCullagh, 1980) that has been used to estimate a receiver operating characteristic (ROC) curve from a single study with data available for multiple thresholds (Tosteson and Begg, 1988). The model is formulated in terms of the probability π_{ij} that a patient in study i with disease status j has a positive test result, where $j = 0$ for a patient without the disease and $j = 1$ for a patient with the disease. In the usual terminology of diagnostic accuracy studies, π_{i1} is the true-positive rate or sensitivity in study i, while π_{i0} is the false-positive rate, equal to $1 -$ specificity. The HSROC model is defined by separate equations for within-study variation (Level I) and between-study variation (Level II). (The Bayesian formulation originally presented by Rutter and Gatsonis (2001) requires an additional third level specifying the priors for the model parameters.)

3.1 *HSROC level I (within study) model*

The Level I model for study i takes the form

$$\operatorname{logit}(\pi_{ij}) = (\theta_i + \alpha_i X_{ij}) \exp(-\beta X_{ij}), \tag{3.1}$$

where X_{ij} is a dummy variable denoting the true disease status for a patient in study i with disease status j. Rutter and Gatsonis (2001) chose to code $X_{ij} = -\frac{1}{2}$ for those without disease ($j = 0$) and $+\frac{1}{2}$ for those with disease ($j = 1$). Both θ_i and α_i are allowed to vary between studies. Rutter and Gatsonis (2001) refer to the θ_i as "cutpoint parameters" or "positivity criteria," as they model the trade-off between sensitivity and specificity in each study: true-positive rate (sensitivity) and false-positive rate ($1-$specificity) both increase with increasing θ_i. The α_i are "accuracy parameters," as they measure the difference between true-positive and false-positive fractions in each study. When $\beta = 0$, the diagnostic odds ratio for each study does not depend on the cutpoint parameter θ_i, and α_i is then the log of the diagnostic odds ratio. β is a "scale parameter" or "shape parameter" which models possible asymmetry in the ROC curve by allowing true-positive and false-positive fractions to increase at different rates as θ_i increases. When $\beta \neq 0$, the diagnostic odds ratio varies with θ_i even if the accuracy parameter α_i is held fixed. β is assumed to be constant across studies, although this assumption can be relaxed somewhat, for example to allow a different value of β in each of several groups of studies (Rutter and Gatsonis, 2001).

3.2 *HSROC level II (between study) model*

Level II models the variation of the parameters θ_i and α_i between studies. In the simplest case, θ_i and α_i are assumed to have independent Normal distributions, with $\theta_i \sim \mathrm{N}(\Theta, \sigma_\theta^2)$ and $\alpha_i \sim \mathrm{N}(\Lambda, \sigma_\alpha^2)$. More generally, the means of the two distributions may be determined by linear functions of study-level covariates. For example, with a single covariate Z that affects both the cutpoint and accuracy parameters,

$$\theta_i \sim \mathrm{N}(\Theta + \gamma Z_i, \sigma_\theta^2) \tag{3.2}$$

$$\alpha_i \sim \mathrm{N}(\Lambda + \lambda Z_i, \sigma_\alpha^2), \tag{3.3}$$

where the coefficients γ and λ express the effect of the covariate Z on the cutpoint and accuracy parameters, respectively. This model may be extended to include more than one covariate, or to allow the covariates that affect the accuracy parameters to differ from those that affect the cutpoint parameters.

4. RELATIONS BETWEEN THE TWO MODELS

We now clarify the relationship between the bivariate and HSROC models. We shall start from the HSROC model. For brevity, first let $b = \exp(\beta/2)$. We can reexpress level I of the HSROC model by splitting (3.1) into separate equations for those with and without disease:

$$\text{logit}(\pi_{i1}) = b^{-1}\left(\theta_i + \frac{1}{2}\alpha_i\right) \tag{4.1}$$

$$\text{logit}(\pi_{i0}) = b\left(\theta_i - \frac{1}{2}\alpha_i\right). \tag{4.2}$$

The bivariate model is written in terms of μ_{Ai} and μ_{Bi}, the logit transforms of sensitivity and specificity in study i. In the notation introduced in Section 3, the sensitivity in study i is π_{i1} and the specificity is $1 - \pi_{i0}$, so

$$\mu_{Ai} = \text{logit}(\pi_{i1}) \tag{4.3}$$

$$\mu_{Bi} = \text{logit}(1 - \pi_{i0}) = -\text{logit}(\pi_{i0}). \tag{4.4}$$

We can therefore relate the random variables that form the basis of the two models:

$$\mu_{Ai} = b^{-1}\left(\theta_i + \frac{1}{2}\alpha_i\right) \tag{4.5}$$

$$\mu_{Bi} = -b\left(\theta_i - \frac{1}{2}\alpha_i\right). \tag{4.6}$$

This pair of equations tells us that μ_{Ai} and μ_{Bi} are linear combinations of two random variables, θ_i and α_i, which the HSROC model assumes to have independent normal distributions (conditional on any covariates). Any pair of linear combinations of random variables with normal distributions has a bivariate normal distribution (see, e.g. Dudewicz and Mishra, 1988, p. 242). Therefore, the HSROC model implies that the joint distribution of μ_{Ai} and μ_{Bi} is bivariate normal. So the HSROC model is precisely equivalent to the bivariate model. We give explicit expressions for the relationships between their parameters in the subsections that follow.

We can express the relationship more concisely using matrix notation. We may write (4.5) and (4.6) in a single matrix equation as

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} = \mathbf{S}^{-1}\begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix}, \quad \text{where } \mathbf{S}^{-1} = \begin{pmatrix} b^{-1} & \frac{1}{2}b^{-1} \\ -b & \frac{1}{2}b \end{pmatrix}. \tag{4.7}$$

Inverting this,

$$\begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix} = \mathbf{S}\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix}, \quad \text{where } \mathbf{S} = \begin{pmatrix} \frac{1}{2}b & -\frac{1}{2}b^{-1} \\ b & b^{-1} \end{pmatrix}. \tag{4.8}$$

$\mathbf{S}$ is then the transformation matrix associated with the change from the bivariate model coordinates (logit-transformed sensitivity and specificity) to the HSROC model coordinates (cutpoint and accuracy parameters). Note that $\mathbf{S}$ is not orthogonal ($\mathbf{S}^{-1} \neq \mathbf{S}^{\text{T}}$). As illustrated in Section 6, it follows that when plotted in bivariate model space (logit-ROC space), the axes corresponding to the coordinates of the HSROC model are not perpendicular to each other.

4.1   *Relation between parameters of models with no covariates*

We can then express the relationship between the parameters of the two models without covariates in terms of the transformation matrix **S** by taking the expectation and variance of both sides of (4.8):

$$\begin{pmatrix} \Theta \\ \Lambda \end{pmatrix} = \mathbf{S} \begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix} \tag{4.9}$$

$$\begin{pmatrix} \sigma_\theta^2 & 0 \\ 0 & \sigma_\alpha^2 \end{pmatrix} = \mathbf{S} \Sigma_{AB} \mathbf{S}^{\mathrm{T}}. \tag{4.10}$$

The assumption of the HSROC model that θ_i and α_i are uncorrelated, i.e. the off-diagonal elements above are zero, fixes the value of b and hence the transformation matrix **S**. So **S** is a non-orthogonal transformation that diagonalizes the variance–covariance matrix of the bivariate model. On expanding the right-hand side of (4.10), we find that these off-diagonal elements are zero if and only if $b = \sqrt{\sigma_B/\sigma_A}$ or, equivalently,

$$\beta = \log(\sigma_B/\sigma_A). \tag{4.11}$$

Thus, the shape parameter (β) of the HSROC model is determined solely by the ratio of the variances of logit sensitivity and logit specificity in the bivariate model, and, perhaps surprisingly, is unrelated to their correlation. Equations (4.9) and (4.10) then allow us to relate the other parameters of the HSROC model to those of the bivariate model:

$$\Theta = \frac{1}{2}\{(\sigma_B/\sigma_A)^{1/2}\mu_A - (\sigma_A/\sigma_B)^{1/2}\mu_B\} \tag{4.12}$$

$$\Lambda = (\sigma_B/\sigma_A)^{1/2}\mu_A + (\sigma_A/\sigma_B)^{1/2}\mu_B \tag{4.13}$$

$$\sigma_\theta^2 = \frac{1}{2}(\sigma_A\sigma_B - \sigma_{AB}) \tag{4.14}$$

$$\sigma_\alpha^2 = 2(\sigma_A\sigma_B + \sigma_{AB}). \tag{4.15}$$

We can also invert these equations to give the five parameters of the bivariate model in terms of those of the HSROC model:

$$\mu_A = b^{-1}\left(\Theta + \frac{1}{2}\Lambda\right) \tag{4.16}$$

$$\mu_B = -b\left(\Theta - \frac{1}{2}\Lambda\right) \tag{4.17}$$

$$\sigma_A^2 = b^{-2}\left(\sigma_\theta^2 + \frac{1}{4}\sigma_\alpha^2\right) \tag{4.18}$$

$$\sigma_B^2 = b^2\left(\sigma_\theta^2 + \frac{1}{4}\sigma_\alpha^2\right) \tag{4.19}$$

$$\sigma_{AB} = -\left(\sigma_\theta^2 - \frac{1}{4}\sigma_\alpha^2\right), \tag{4.20}$$

where $b = \exp(\beta/2)$, as defined above.

4.2 *Relations between parameters of models with covariates*

We now move on to examine the relationship between the models when covariates are included. If the bivariate model is extended to include a single covariate Z that affects both the sensitivity and specificity, (4.10) is unchanged, while (4.9) for the expectation of (4.8) becomes

$$E \begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix} = \mathbf{S} \begin{pmatrix} \mu_A + v_A Z_i \\ \mu_B + v_B Z_i \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \{ b\mu_A - b^{-1}\mu_B + Z_i(bv_A - b^{-1}v_B) \} \\ b\mu_A + b^{-1}\mu_B + Z_i(bv_A + b^{-1}v_B) \end{pmatrix}. \tag{4.21}$$

This is of the form

$$E \begin{pmatrix} \theta_i \\ \alpha_i \end{pmatrix} = \begin{pmatrix} \Theta + \gamma Z_i \\ \Lambda + \lambda Z_i \end{pmatrix}. \tag{4.22}$$

The extension to more than one covariate, with each covariate affecting both accuracy and cutpoint parameters, is straightforward. Therefore, a bivariate model in which one or more covariates affect both sensitivity and specificity is equivalent to an HSROC model in which the same covariates are allowed to affect both accuracy and cutpoint parameters.

However, a bivariate model in which different covariates are allowed to affect sensitivity from specificity, or covariates are included for only sensitivity or only specificity, will not be equivalent to an HSROC model including covariates, unless constraints are imposed on the relationship between the coefficients of the covariates in the HSROC model. The converse is also true.

5. FOCUS OF INFERENCE AND MODEL OUTPUTS

The two approaches lead to different forms of model output appearing more natural.

5.1 *HSROC model*

The HSROC model gives rise to a SROC curve by allowing the threshold parameter θ_i to vary while holding the accuracy parameter α_i fixed at its mean Λ. For the model without covariates, the expected sensitivity for a given specificity is then given by (Rutter and Gatsonis, 2001; Macaskill, 2004)

$$\text{logit}(\text{Sensitivity}) = \Lambda\, e^{\beta/2} - e^{\beta} \text{logit}(\text{Specificity}). \tag{5.1}$$

Rutter and Gatsonis (2001) suggest that the curve is restricted to the observed range of estimated specificities of the studies to discourage extrapolation beyond the data. If $\beta = 0$, the curve is symmetric about the "sensitivity = specificity" diagonal. This SROC curve does not depict the uncertainty in any of the parameter estimates and depicts the variability in threshold but not in accuracy.

5.2 *Bivariate model*

As Reitsma *and others* (2005) suggest, confidence and prediction regions in ROC space can be constructed using the estimates from the bivariate model. As sensitivity and specificity may be highly correlated, separate confidence intervals for the mean logit sensitivity μ_A and mean logit specificity μ_B may be misleading. It is preferable to use an elliptical joint confidence region for both parameters. Such an ellipse is most easily generated using a parametric representation (Douglas, 1993):

$$\mu_A = \hat{\mu}_A + s_A c \cos t \tag{5.2}$$

$$\mu_B = \hat{\mu}_B + s_B c \cos(t + \arccos r), \tag{5.3}$$

where s_A and s_B are the estimated standard errors of $\hat{\mu}_A$ and $\hat{\mu}_B$, r is the estimate of their correlation, and varying t from 0 to 2π generates the boundary of the ellipse. The constant c has been called the boundary constant of the ellipse (Alexandersson, 2004); asymptotically, to give a $100(1 - \alpha\%)$ confidence region, $c = \sqrt{\chi^2_{2;\alpha}}$, where $\chi^2_{2;\alpha}$ is the upper $100\alpha\%$ point of the χ^2 distribution with two degrees of freedom. When the number of studies is small, it may be preferable to use a more conservative approximate confidence region given by $c = \sqrt{2 f_{2,n-2;\alpha}}$, where n is the number of studies and $f_{2,n-2;\alpha}$ is the upper $100\alpha\%$ point of the F distribution with degrees of freedom 2 and $n - 2$ (Douglas, 1993; Chew, 1966). Such an ellipse in logit-ROC space can then be back-transformed to conventional ROC space to give a confidence region for the summary operating point.

It is also possible to construct a prediction region giving the region which has a given probability (e.g. 95%) of including the "true" sensitivity and specificity of a future study. The covariance matrix for the true logit sensitivity and logit specificity in a future study is

$$\Sigma_{AB} + \mathrm{Var}\begin{pmatrix} \hat{\mu}_A \\ \hat{\mu}_B \end{pmatrix}. \tag{5.4}$$

In practice, both terms must be estimated by fitting the model to the data. The parameters s_A, s_B, and r in (5.2) and (5.3) can then be replaced by the corresponding quantities derived from this covariance matrix to give the prediction ellipse in logit-ROC space. Again, this can be back-transformed to a prediction region for the true sensitivity and specificity of a future study in conventional ROC space.

6. EXAMPLE: LYMPHANGIOGRAPHY FOR DIAGNOSIS OF LYMPH NODE METASTASIS

As an example, we shall apply both methods to data on 17 studies of lymphangiography for the diagnosis of lymph node metastasis in women with cervical cancer, one of three imaging techniques in the meta-analysis of Scheidler *and others* (1997) which has been much used as an example data set for methodological papers on diagnostic meta-analysis (Rutter and Gatsonis, 2001; Macaskill, 2004; Reitsma *and others*, 2005). A SROC plot showing the estimates of sensitivity and specificity from the individual studies is shown in Figure 1.

We fitted both the bivariate and the HSROC models using the NLMIXED procedure in the statistical software package SAS (SAS Institute Inc., 2003), using code similar to that given by Macaskill (2004) and available from the authors on request. Note that our results differ slightly from those in Reitsma *and others* (2005) as they use empirical logit transforms and their standard errors followed by the MIXED procedure in SAS, where we choose to model the binomial error structure directly using the NLMIXED procedure.

Table 1 shows the parameter estimates obtained for both models, and the result of applying (4.11)–(4.20) to transform estimates from the HSROC model to the corresponding parameters of the bivariate model and vice versa. The standard errors of the transformed estimates were computed by the delta method using the ESTIMATE statement of the NLMIXED procedure. As can be seen, the results are virtually identical. (The standard errors are identical in theory due to the close relationship between the delta method and maximum likelihood; Cox, 1998; Cox and Hinkley, 1974, Exercise 4.15.) By taking the inverse logit transforms of μ_A and μ_B, respectively, and assuming their estimates have a normal distribution, the summary estimate of sensitivity is found to be 0.67 (95% CI, 0.60–0.74) and that of specificity is 0.84 (95% CI, 0.76–0.89). In this example, σ_{AB} is estimated to be positive, though with large standard error. This implies a positive correlation between sensitivity and specificity across the studies, not the negative correlation that would be expected if the between-study heterogeneity was due mainly to variation in threshold.
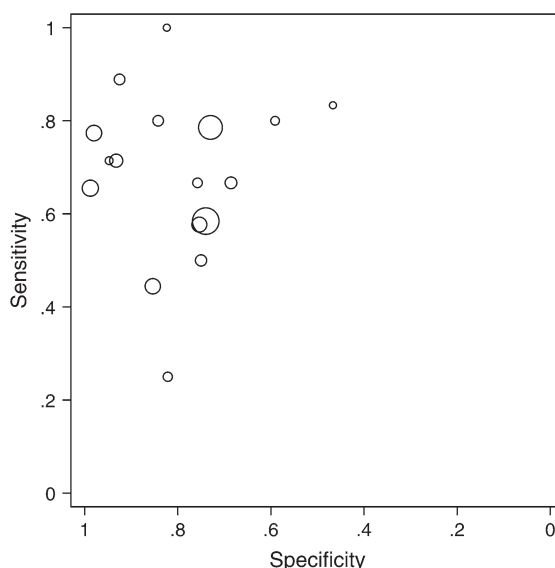
Fig. 1. SROC plot for the data of Scheidler *and others* (1997) on lymph node metastasis for diagnosis of cervical cancer. The area of the circles are in proportion to the number of patients in each study.

Table 1. *Results of fitting the bivariate and HSROC models to the lymphangiography data*

Parameter	Estimate (SE) from bivariate model	Result of applying (4.16)–(4.20) to HSROC estimates below
μ_A	0.7266 (0.1545)	0.7266 (0.1545)
μ_B	1.6390 (0.2505)	1.6390 (0.2506)
σ_A^2	0.1250 (0.1307)	0.1250 (0.1307)
σ_B^2	0.8233 (0.4056)	0.8236 (0.4058)
σ_{AB}^2	0.0766 (0.1470)	0.0765 (0.1470)
	Estimate (SE) from HSROC model	Result of applying (4.12)–(4.15) to bivariate estimates above
Θ	0.0706 (0.3271)	0.0706 (0.3271)
Λ	2.1872 (0.3087)	2.1871 (0.3087)
β	0.9427 (0.5764)	0.9427 (0.5765)
σ_α^2	0.7948 (0.5115)	0.7947 (0.5115)
σ_θ^2	0.1222 (0.1083)	0.1221 (0.1083)

SE, standard error.

Figure 2 shows the 95% confidence region for the summary operating point and a 95% prediction region for the true operating point in a single future study in both logit-transformed ROC space (left panel) and back-transformed to conventional ROC space (right panel). The prediction region covers a greater range of specificity than sensitivity, in contrast to the estimates from the separate studies shown in the SROC plot in Figure 1, which exhibit more variation in estimated sensitivity than specificity. This is due to the fact that most of the studies had a considerably larger number of patients with negative results on the reference test than positive results, leading to greater sampling variability in the estimates
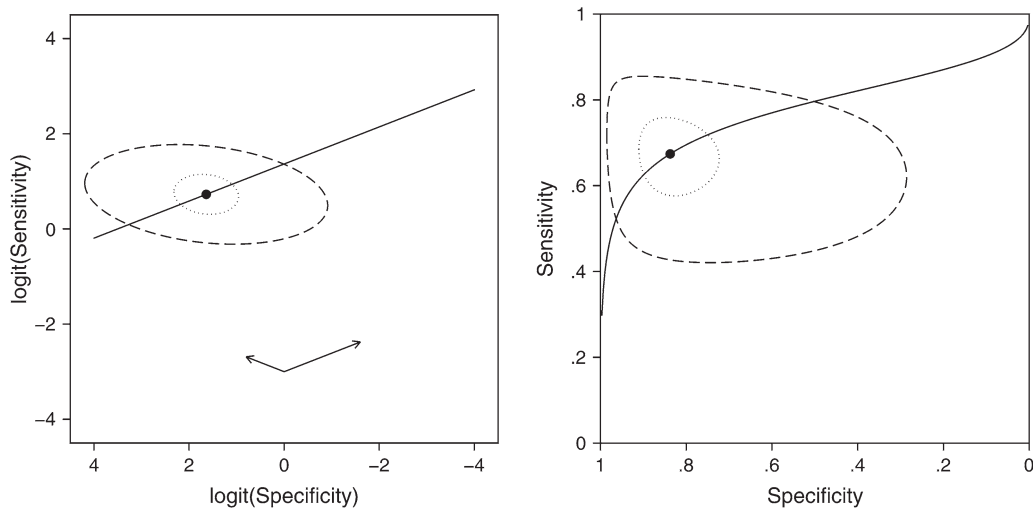
Fig. 2. Summary points, lines, and regions in logit-transformed ROC space (left) and conventional ROC space (right). Filled circle: summary point. Solid line: SROC curve. Dotted line: boundary of the confidence region for the summary point. Dashed line: boundary of prediction region. The left-hand panel also shows the HSROC model coordinate axes in logit-transformed ROC space. Note that these axes do not align with the major or minor axes of the ellipse.

of sensitivity than specificity. The prediction region is for the true sensitivity and specificity in a future study, not the estimated values.

Also shown in Figure 2 is the SROC curve (a straight line in logit-transformed ROC space). Note that the SROC curve takes a conventional shape despite the positive estimate of the correlation between sensitivity and specificity. The left-hand panel also shows the HSROC coordinate axes in logit-transformed ROC space. Note that these axes do not align with the major or minor axes of the ellipse. The θ axis is parallel to the ROC curve, while its horizontal reflection is parallel to the α axis. The method of Littenberg and Moses (1993) (using unweighted linear regression) gives a curve similar to, but slightly above, the HSROC curve, as shown in Macaskill (2004).

7. DISCUSSION

We have shown that the HSROC model and the bivariate random-effects model for meta-analysis of diagnostic accuracy studies are very closely related, and in common situations identical. In the absence of study-level covariates, they are different parameterizations of the same model. The bivariate model allows inclusion of covariates that affect sensitivity or specificity or both, while the HSROC model allows covariates that affect accuracy or threshold parameters or both. An HSROC model that allows one or more covariates to affect both accuracy and threshold parameters is equivalent to a bivariate model that allows the same covariates to affect both sensitivity and specificity. However, the HSROC model can be more easily extended to include a covariate to affect the degree of asymmetry of the SROC curve.

The models may differ in the options for introducing greater model parsimony by dropping or combining parameters: The HSROC framework allows the analyst to drop the random effect for the accuracy parameter and assume this is fixed across all studies, and hence that only the threshold parameter varies between studies. This corresponds to perfect negative correlation between the logit transforms of sensitivity and specificity in the bivariate model ($\sigma_{AB} = -\sigma_A \sigma_B$). The confidence and prediction regions then collapse to lie along the SROC curve. The HSROC framework also allows the assumption of a symmetric

SROC curve with constant diagnostic odds ratio by setting $\beta = 0$, which in the bivariate model corresponds to equal variances of logit sensitivity and logit specificity ($\sigma_A^2 = \sigma_B^2$). The ability to enforce such constraints on bivariate model parameters may vary between software packages. By contrast, it does not appear natural to set any of the parameters of the bivariate model to zero. One practical advantage of the bivariate model is that it can be fitted in a wider range of software, for example MLwiN, SAS, or the Stata package "gllamm" (Rabe-Hesketh *and others*, 2004), whereas the HSROC model is at present only estimable using WinBUGS or the NLMIXED procedure in SAS.

As we have seen, the different parameterizations of the HSROC and bivariate models arise from different ideas of the most appropriate meta-analytic summaries of the results of diagnostic test accuracy studies, and have primarily been used to produce these chosen summaries. The HSROC parameterization naturally leads to a SROC curve when the threshold parameter θ is allowed to vary between studies but the accuracy parameter α is fixed at its mean. This may be reasonable when there is little or no detectable heterogeneity in the accuracy parameter, i.e. σ_α^2 is estimated to be close to zero, or when there is considerably greater variability in threshold than in accuracy. The bivariate model parameterization naturally leads to a summary operating point, i.e. a summary sensitivity and specificity, together with confidence intervals for each or a joint confidence region for both together. When there is a considerable degree of between-study heterogeneity, as is common in meta-analysis of diagnostic accuracy studies, a prediction region may be preferable to a confidence region.

In our example in Section 6, fitting both models to the same data gave near-identical results in agreement with the formulae derived in Section 4.1, when both models were fitted using the NLMIXED procedure in SAS. However, such close agreement may not always be found in practice, particularly if the models are fitted using different approaches in different software. Rutter and Gatsonis (2001) originally proposed fitting the HSROC model using a Bayesian Markov chain Monte-Carlo method. Unlike maximum likelihood estimates, Bayesian posterior means or medians are not invariant under nonlinear transformations such as those in Section 4.1. Reitsma *and others* (2005) fit the bivariate model using the MIXED procedure in SAS, which, unlike the NLMIXED procedure, requires first calculating empirical estimates of the logit transforms of sensitivity and specificity and their standard errors, treating the latter as fixed and approximating the within-study variability of the logits by a normal distribution. This approach is less computationally demanding but involves some degree of approximation when the study sizes are small. In addition, regardless of the method of estimation, there is typically little information on the covariance parameter σ_{AB} of the bivariate model unless there are many studies of reasonable size with considerable variation in sensitivity and specificity between them. Its estimation may therefore prove troublesome (R. Riley *and others*, in preparation).

Another reason for apparent discrepancies between results in previous publications is that when fitting models to the data of Scheidler *and others* (1997), authors have made different assumptions about the equality of parameters between the three imaging techniques assessed, of which for simplicity we have only considered one, lymphangiography, in the example here. Rutter and Gatsonis (2001) allowed all five parameters of the HSROC model to differ between the three imaging techniques. Macaskill (2004) assumed that the two variance parameters σ_α^2 and σ_θ^2 were the same while the other three parameters Λ, Θ, and β differed. Reitsma *and others* (2005) estimated a bivariate model in which the three variance–covariance parameters σ_A, σ_B, and σ_{AB} are the same for the three imaging techniques and the two location parameters μ_A and μ_B differ, thereby constraining the three SROC curves to have the same degree of asymmetry.

It may initially seem surprising that (4.11) for β, the shape parameter of the HSROC model, does not involve the covariance σ_{AB} of the bivariate model but only the ratio of the variances. In fact, σ_{AB} only enters (4.14) and (4.15) for the variances of the HSROC parameters. It follows that the equation for the SROC curve given by the HSROC model does not require this covariance. It is therefore possible to use (4.11), (4.13), and (5.1) to estimate the equation of the SROC curve from separate conventional

"univariate" random-effects meta-analyses of logit-transformed sensitivity and logit-transformed specificity. These could be performed using any of the widely available packages for random-effects meta-analysis. The estimates obtained by such an approach will not be identical to those from the bivariate model as joint marginal normality of two random variables does not imply they have a bivariate normal distribution. However, if the bivariate normal model does hold, separate univariate analyses should give consistent estimates of the means and variances, with only a slight loss in efficiency (Riley *and others*, 2006). Separate univariate analyses may therefore provide an alternative to the method of Littenberg and Moses (1993) as a way of generating a SROC curve using widely available algorithms. Separate univariate analyses may also be useful in providing starting values for the iterative procedures required to fit either the bivariate or the HSROC models, which may aid convergence.

There is empirical evidence that aspects of the design and conduct of diagnostic accuracy studies can lead to bias or increased variation in their results. Exploration of potential sources of heterogeneity is therefore a crucial component of systematic reviews of such studies. Sources of between-study heterogeneity may include differences in patient selection and clinical setting, disease severity, specifics of the index and reference tests, and interobserver variability (Lijmer *and others*, 1999; Whiting *and others*, 2004). The expected effect of a covariate on test performance may lead to a preference for one of the two parameterizations implied by the HSROC and bivariate models. For example, "spectrum bias," in which the subjects included in a study are not representative of the patients who will receive the test in practice (Whiting *and others*, 2003), might be expected to affect test accuracy rather than threshold, and might therefore be most appropriately investigated using the HSROC approach. Conversely, between-study variation in disease severity will affect sensitivity but not specificity, leading to a preference for the bivariate approach. For most study characteristics, however, there are few *a priori* reasons to prefer one approach over the other; further empirical research on this issue is needed.

The methods explored in this paper assume that only summary data from each study are available in the form of a 2 × 2 table. Meta-analysis of individual patient data may offer particular advantages for diagnostic research (Khan *and others*, 2003). It would allow differences in patient spectra to be properly accounted for, and enable assessment of the additional information provided by a test above that already known from patient history and clinical examination. For test results that are originally numerical or ordered categorical, it would also capture within-study information about the ROC curve that is lost when a particular threshold is chosen and the results collapsed into a summary 2 × 2 table.

In summary, we have demonstrated that the HSROC and bivariate models are very closely related and often identical. The parameter estimates from either model can be used to produce a summary operating point, an SROC curve, confidence regions, or prediction regions. The choice between these parameterizations depends partly on the degrees of and reasons for between-study heterogeneity. Empirical evidence about this would be useful in guiding analysts.

REFERENCES

ALEXANDERSSON, A. (2004). Graphing confidence ellipses: an update of ellip for Stata 8. *Stata Journal* **4**, 242–56.

BOSSUYT, P. M., REITSMA, J. B., BRUNS, D. E., GATSONIS, C. A., GLASZIOU, P. P., IRWIG, L. M., LIJMER, J. G., MOHER, D., RENNIE, D. AND DE VET, H. C. W. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *British Medical Journal* **326**, 41–4.

CHEW, V. (1966). Confidence, prediction, and tolerance regions for the multivariate normal distribution. *Journal of the American Statistical Association* **61**, 605–17.

COX, C. (1998). Delta method. In: Armitage, P. and Colton, T. (editors), *Encyclopedia of Biostatistics*, 1st edition. Chichester, UK: Wiley. pp 1125–1127.

COX, D. R. AND HINKLEY, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

DEEKS, J. J. (2001). Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *British Medical Journal* **323**, 157–62.

DEVILLE, W., BUNTINX, F., BOUTER, L., MONTORI, V., DE VET, H., VAN DER WINDT, D. AND BEZEMER, P. (2002). Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Medical Research Methodology* **2**, 9.

DOUGLAS, J. B. (1993). Confidence regions for parameter pairs. *American Statistician* **47**, 43–5.

DUDEWICZ, E. J. AND MISHRA, S. N. (1988). *Modern Mathematical Statistics*. New York: Wiley.

DUKIC, V. AND GATSONIS, C. (2003). Meta-analysis of diagnostic test accuracy assessment studies with varying number of thresholds. *Biometrics* **59**, 936–46.

GLUUD, C. AND GLUUD, L. L. (2005). Evidence based diagnostics. *British Medical Journal* **330**, 724–6.

KHAN, K. S., BACHMANN, L. M. AND TER RIET, G. (2003). Systematic reviews with individual patient data meta-analysis to evaluate diagnostic tests. *European Journal of Obstetrics & Gynecology and Reproductive Biology* **108**, 121–5.

LIJMER, J. G., BOSSUYT, P. M. M. AND HEISTERKAMP, S. H. (2002). Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Statistics in Medicine* **21**, 1525–37.

LIJMER, J. G., MOL, B. W., HEISTERKAMP, S., BONSEL, G. J., PRINS, M. H., VAN DER MEULEN, J. H. AND BOSSUYT, P. M. (1999). Empirical evidence of design-related bias in studies of diagnostic tests. *Journal of the American Medical Association* **282**, 1061–6.

LITTENBERG, B. AND MOSES, L. E. (1993). Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. *Medical Decision Making* **13**, 313–21.

MACASKILL, P. (2004). Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *Journal of Clinical Epidemiology* **57**, 925–32.

MCCULLAGH, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B, Methodological* **42**, 109–42.

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**, 1293–316.

RABE-HESKETH, S., PICKLES, A. AND SKRONDAL, A. (2004). GLLAMM manual. *U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 160*. http://www.bepress.com/ucbbiostat/paper160.

REITSMA, J. B., GLAS, A. S., RUTJES, A. W. S., SCHOLTEN, R. J. P. M., BOSSUYT, P. M. AND ZWINDERMAN, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* **58**, 982–90.

RILEY, R. D., ABRAMS, K. R., LAMBERT, P., SUTTON, A. J. AND THOMPSON, J. R. (2006). An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine*. doi:10.1002/sim.2524.

RUTTER, C. M. AND GATSONIS, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine* **20**, 2865–84.

SAS Institute Inc. (2003). *The SAS System for Windows. Version 9.1*. Cary, NC: SAS Institute Inc.

SCHEIDLER, J., HRICAK, H., YU, K. K., SUBAK, L. AND SEGAL, M. R. (1997). Radiological evaluation of lymph node metastases in patients with cervical cancer. a meta-analysis. *Journal of the American Medical Association* **278**, 1096–101.

SIADATY, M. AND SHU, J. (2004). Proportional odds ratio model for comparison of diagnostic tests in meta-analysis. *BMC Medical Research Methodology* **4**, 27.

TATSIONI, A., ZARIN, D. A., ARONSON, N., SAMSON, D. J., FLAMM, C. R., SCHMID, C. AND LAU, J. (2005). Challenges in systematic reviews of diagnostic technologies. *Annals of Internal Medicine* **142**, 1048–55.

TOSTESON, A. N. AND BEGG, C. B. (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making* **8**, 204–15.

VAN HOUWELINGEN, H., ARENDS, L. R. AND STIJNEN, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* **21**, 589–624.

VAN HOUWELINGEN, H. C., ZWINDERMAN, K. H. AND STIJNEN, T. (1993). A bivariate approach to meta-analysis. *Statistics in Medicine* **12**, 2273–84.

WALTER, S. D. (2002). Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Statistics in Medicine* **21**, 1237–56.

WHITING, P., RUTJES, A. W., REITSMA, J. B., BOSSUYT, P. M. AND KLEIJNEN, J. (2003). The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* **3**, 25.

WHITING, P., RUTJES, A. W. S., REITSMA, J. B., GLAS, A. S., BOSSUYT, P. M. M. AND KLEIJNEN, J. (2004). Sources of variation and bias in studies of diagnostic accuracy—a systematic review. *Annals of Internal Medicine* **140**, 189–202.