

Bayesian meta-analysis for test accuracy

Lyle D Broemeling

Broemeling & Associates Inc,
Medical Lake, WA, USA

Abstract: Determining the accuracy of a medical test is quite difficult because accuracy is an elusive parameter to estimate. A common scenario is estimating the true and false positive fractions from different studies and arriving at a common value of the accuracy of the test. The accuracy is expressed with an estimate similar to the area under the receiver operating characteristic (ROC) curve. Under the assumption that the ROC area is the same across all tests, the true and false positive fractions can be plotted on the same graph to obtain an experimental ROC curve, called the summary ROC curve (SROC) curve. The estimate of the accuracy of the curve is the ordinate of the point of intersection, where the SROC curve intersects the line with equation true positive rate + false positive rate = 1. Using a Bayesian approach, the presentation begins with summarizing information about test accuracy for tests with ordinal and continuous scores, where it is assumed the tests share a common ROC curve, but the tests may differ in the threshold used to declare a positive test. The true and false positive rates are transformed so that one may use bilogistic regression to determine the accuracy of the combined tests where the posterior distribution of the parameters of the model are determined. Bayesian inferences are based on the posterior distribution of the SROC curve and the computations are executed with the WinBUGS software package, and several examples from various areas of medicine illustrate the methodology.

Keywords: Bayesian inferences, meta-analysis, SROC curve

Introduction

Determining the accuracy of a medical test is quite difficult because accuracy is an elusive parameter to estimate. It is well known that the accuracy of a particular test will vary because of intra and inter study variation. Even though a test can be replicated under identical conditions, the accuracy will still vary between replications. The literature on the accuracy of the particular test continues to grow, each study giving some estimate of its accuracy, but of course the estimates vary because of different study conditions. The accuracy varies because the study populations of patients vary, the test itself changes somewhat, and the readers of the test scores of course also vary. Because of the various sources of variation it is important to conduct studies that summarize the accuracy of a particular medical test.

For example take the case of heart disease and the accuracy of the exercise stress test. Hundreds of studies have estimated the accuracy (the true and false positive fractions or the receiver operating characteristic [ROC] area) of this test. Suppose the exercise stress test could be replicated under identical conditions with the same patients and the same readers to interpret the test scores and the test accuracy estimated. Of course,

Correspondence: Lyle D Broemeling
Broemeling & Associates Inc,
1023 Fox Ridge Road, Medical Lake,
WA 99022, USA
Tel +1 509 299 4030
Email broemeli2@aol.com

the accuracy of the test will vary from test to test and such variation is called experimental error. It is difficult and in fact almost impossible to estimate the experimental error for most medical tests, and more common is the case of summarizing the accuracy of the exercise stress test of different studies that involve different patient populations and different sets of radiologists. There are studies in which it is possible to estimate the replication variation within readers.

The first scenario to be considered is estimating the common ROC area. The accuracy is expressed with an estimate similar to the area under the ROC curve, and the estimate of the accuracy of the so-called SROC (summary ROC curve) is the ordinate of the point of intersection, where the SROC curve intersects the line with equation true positive rate (TPR) + false positive rate (FPR) = 1.

The approach taken here is appropriate for ordinal and continuous data that employ a threshold to declare a positive test and follows the presentation of Kardaun and Kardaun¹ and Moses et al² who do not work directly with the true and false positive rates but with their logits and consequently with a regression model that allows one to work easily with the SROC curve. It should be noted that this approach is also adopted by Zhou et al³ and will be followed to some extent here, except a Bayesian model is established. Much has been accomplished using the Bayesian approach and the reader is referred to Stangl and Berry,⁴ who edited a book including many articles that address various issues in meta-analysis from a Bayesian viewpoint; however, the book focuses on summarizing various estimates of treatment efficacy in clinical trials and not on estimating various measures of test accuracy.

This presentation begins with summarizing information about test accuracy for tests with ordinal and continuous scores where it is assumed the tests share a common ROC curve, thus, the tests may differ in the threshold used to declare a positive test. The false and positive rates can be plotted to determine a common ROC curve called the SROC curve. The true and false positive rates are transformed so that one may use bilogistic regression to determine the accuracy of the combined tests where the posterior distribution of the parameters of the model are determined. The slope and intercept of the regression determine the SROC curve. The paper is continued by allowing the inclusion of study covariates that allows for inter-study variation between the various studies that comprise the meta-analysis. Two or more versions of the same tests are compared with the bilogistic regression methodology by using inter study covariates. Bayesian inference is illustrated with a well-known

example of DeVries et al⁵ which is based on two versions of ultrasonography for the diagnosis of stenosis of the peripheral arteries, and the accuracy of the two versions are compared with a regression model and by comparing the Q-statistics which measures the accuracy of the SROC. Additional examples include diagnostic studies for coronary artery disease, inflammatory bowel disease, osteomyelitis, breast cancer, and recurrent colorectal cancer. The conclusion of the paper emphasizes the summarization of tests with a common ROC area, where the posterior distribution of the ROC area and its standard deviation allow one to compute the common area, which is estimated as a weighted average of the individual ROC areas weighted by the inverse of the posterior variance.

Several of the examples are based on recent studies of meta-analysis. For example, a meta-analysis by Vanhoenacker et al⁷ summarizes the accuracy of multidetector computed tomography (CT) angiography for the diagnosis of coronary artery disease, while the Horsthuis et al⁸ meta-analysis explores the detection of inflammatory bowel disease with ultrasound, magnetic resonance (MR) scintigraphy, and CT. The latter study allows one to compare the accuracy of the three modalities. A third example by Pakos et al⁹ is a meta-analysis of a nuclear medicine procedure to diagnose osteomyelitis, and in all three examples the SROC curve is inferred by Bayesian methods. Such articles have a standard way to present their results including an introduction, a description of the methods, a reporting of the results, and a section for comments and conclusions. Such articles usually have enough detail so that others may check and replicate their results. Of paramount importance is describing just how the various studies of the meta-analysis are included in the study, and the description should include enough information to determine the heterogeneity between the various studies, including the number of readers used in each study, and the threshold value. If the SROC curve is to be determined, the homogeneity needs to be demonstrated so that one has confidence in the overall accuracy of the combined studies.

The SROC curve and bilogistic regression

The first scenario for meta-analysis is the least complicated, namely, one assumes the various studies have a common ROC curve. If the test scores are continuous or ordinal, the threshold value may vary giving different FPR and TPR values which can be plotted to give the common ROC curve referred to as the summary ROC curve or the SROC curve. Thus, there is enough homogeneity between studies

to assume a common ROC curve and this assumption needs to be checked by the information given for the various studies. The reader should remember that the inclusion and exclusion criteria may vary between studies, that of course the readers interpreting the test scores will not be the same between studies, that the threshold (the value that declares a positive test) value can vary, and lastly that the various tests, although related, may not be the same. For example when using CT to diagnose lung cancer, the CT equipment will not be the same and not operated in the same manner from study to study. On examining the recent issues of the imaging literature, the summary accuracy is usually based on a SROC analysis.

The approach taken here is to base the regression analysis on the approach of Moses et al,² who do not work with the (FPR, TPR) points directly, but with a transformation

$$B = V - U \quad (1)$$

and

$$S = V + U \quad (2)$$

where,

$$U = \text{logit}(\text{FPR}) \quad (3)$$

and

$$V = \text{logit}(\text{TPR}).$$

The B values are regressed on the S values

$$B = \text{beta}[1] + \text{beta}[2]S, \quad (4)$$

where beta[1] and beta[2] are unknown parameters.

It is assumed that U and V have logistic distributions, thus, it is reasonable to assume that B and S also have logistic distributions. The interpretation of B and S is very informative, because B is the log odds ratio where the numerator is the odds of a positive test given the disease is present and the denominator is the odds the test score is positive given the disease is not present. The variable S can be thought of as measuring the effect of the threshold value, in the sense that if S and beta[1] are zero, the true and false positive rates are equal, and also if S is positive the sensitivity (TPR) is greater than the FPR. On the other hand when the sensitivity is less than the FPR, the values of S are negative. Note the interpretation of the intercept beta[1] is the average value of B (the log odds ratio) when S = 0 and that beta[2] measures the effect of S on B, that is for each unit increase in S, B increases on the average by beta[2] units. It is probably safe to say that if beta[2] is estimated to be close to zero,

the test at hand has the same power to detect a difference in the two populations (diseased versus non-diseased) for all values of the threshold.

It can be shown, see Zhou et al (eq 12.4)³ that the SROC curve is defined as

$$\text{SROC}(\text{FPR}) = [1 + \exp(-\text{beta}[1]/(1 - \text{beta}[2]))] [(1 - \text{FPR}) / \text{FPR}]^{(1+\text{beta}[2])/(1-\text{beta}[2]))-1} \quad (5)$$

and the curve is determined by plotting (FPR, SROC) where the SROC and FPR values are given by (11.5). Note that if beta[2] = 0, (5) can be modified accordingly.

There are many ways to estimate the parameters beta[1] and beta[2] of the regression of B on S, and the reader is referred to Moses et al² for some non-Bayesian estimation techniques. The approach here is Bayesian where first the posterior density of beta[1] and beta[2] are determined using (4) with a noninformative priors for the regression coefficients and scale parameter tau. The prior distribution of the beta[1] is normal (0.0001, 0.0001) and the prior distribution for tau is gamma with parameters 0.0001 and 0.0001. Note the logistic distribution has two parameters and is similar in shape to the normal distribution. The posterior distribution of the ordinates of the SROC curve (5) is induced by the posterior distribution of beta[1] and beta[2], and will be illustrated with several examples in a later section.

Bayesian analysis for summary accuracy

The obvious measure of test accuracy for a meta-analysis is the area under the SROC curve; however, since the TPR values tend to be concentrated over a relatively small range of FPR values, the entire SROC curve must be estimated by extrapolation to the entire range of FPR values over (0, 1). Because of this impediment, other measures have been devised, for example, by Moses et al,² who proposed the ordinate of the intersection between the SROC curve and the line with equation

$$\text{TPR} + \text{FPR} = 1 \quad (6)$$

which is the negative diagonal of the unit square.

It can be shown that the ordinate of the intersection is

$$Q = (1 + e^{-\text{beta}[1]/2})^{-1}, \quad (7)$$

where beta[1] is the intercept term of the bilogistic regression (4). Therefore, the posterior distribution of beta[1] induces the posterior distribution of Q. Note that 'large' values of Q close to one, indicate excellent accuracy because the

intersection point is close to the (0,1) point of the unit square, and in a similar way, values of Q close to 0.5 imply the SROC curve is close to the main diagonal of the unit square, indicating very poor accuracy. It should be noted that Q is a surrogate measure of accuracy and that generally speaking accuracy is the mean of a weighted difference between the mean transformed TPR and mean transformed FPR on where the transformation maps the ROC curve onto a straight line.

A meta-analysis for one test

The first set of examples of a meta-analysis are for studies with a common SROC curve, where the sample information consists of true and false positive ratios or equivalently the true negative (TN), true positive (TP), false positive (FP), and false negative (FN) of the 2 by 2 square for each study. Consider the DeVries et al⁵ meta-analysis that summarizes nine studies in which peripheral artery stenosis is determined with regular duplex ultrasonography. The information from the study is in the form of a 2 by 2 table of TN, TP, FP, and FN values for each study and appears in the list statement of **BUGS CODE 1** given below.

BUGS CODE 1

```
# one test
model;
{
for(i in 1:N){tpr[i]<-(tp[i] + 0.5)/(tp[i] + fn[i] + 0.05)}
for(i in 1:N){fpr[i]<-(fp[i] + 0.5)/(fp[i] + tn[i] + 0.05)}
for(i in 1:N){u[i]<-logit(fpr[i])}
for(i in 1:N){v[i]<-logit(tpr[i])}
for(i in 1:N){b[i]<-v[i]-u[i]}
for(i in 1:N){s[i]<-v[i] + u[i]}
# bilogistic regression of b on s
for(i in 1:N){b[i]~dlogis(mu[i],tau)}
mu[i]<-beta[1] + beta[2]*s[i]}
for(i in 1:2){beta[i]~dnorm(0.0000,0.0001)}
tau~dgamma(0.0001,0.0001)
P<-1 + exp(-beta[1]/2)
# accuracy of test
Q<-1/P
#sroc curve assumes slope is 0
r1<-exp(-beta[1])
for(i in 1:N){r2[i]<-(1-fpr[i])/fpr[i]}
for(i in 1:N){r3[i]<-1 + r1*r2[i]}
for(i in 1:N){sroc[i]<-1/r3[i]}
}
```

```
# data from DeVries et al5 duplex mode
list(N = 8, tn = c(516,89,235,262,488,48,156,376),
fn = c(28,8,23,20,14,7,2,31),
fp = c(20,12,5,22,9,3,14,12),
tp = c(78,59,75,89,118,48,39,121))
# data from Meijer et al6
# a 1 replaces a 0 in the data of Meijer et al6
list(N = 20,
tn = c(23,35,10,9,23,30,60,48,12,38,23,60,9,42,27,50,35,
5,37,20),
fn = c(2,1,1,1,1,1,1,1,1,1,1,1,1,1,2,1,1,1,1),
fp = c(1,2,1,1,1,6,7,4,4,1,6,7,1,4,3,5,1,1,3,3),
tp = c(36,28,29,29,16,19,20,18,88,12,35,14,25,53,38,25,
13,26,44,76))
# initial values
list(beta = c(0,0), tau = 1))
```

For example the TN, FN, FP, and TP values for the first study are 516, 28, 20, and 78 respectively and the program calculates the TP and FP rates. First, a bilogistic regression is performed which produces the posterior characteristics of the regression coefficients $\beta[1]$ and $\beta[2]$, then the posterior distribution of the Q parameter is calculated, which expressed the summary accuracy of the eight studies. Lastly, the posterior characteristics of the ordinates of the SROC curve are determined corresponding to the 8 FPR values of the eight studies. The analysis is executed with 65,000 observations generated from the joint posterior distribution, with a burn in of 5000 and a refresh of 100.

Regression analysis of B on S reveals that the slope is 'small' with a 95% confidence interval $(-0.6842, 0.5995)$ indicating that it is not unreasonable to let $\beta[2] = 0$ and implying that ultrasound duplex is discriminating between the diseased and non diseased populations in the same way for all values of the test threshold. **BUGS CODE 1** contains statements that calculate the SROC values assuming the slope is zero. The coefficient $\beta[1]$ has posterior mean 4.7 which is the average value of B when $S = 0$, indicating that the odds ratio has posterior mean 4.7, that is to say, the odds of a positive test for the diseased population (those with stenosis in the peripheral arteries) is 4.7 times more than the odds of a positive test result for the non diseased patients. Q had a posterior mean of 0.91 which implies good accuracy with ultrasound to detect stenosis.

The slope $\beta[2]$ of the regression of B on S is estimated to be close to zero and the analysis calculates the posterior characteristics of the SROC values corresponding to the FPR values which are calculated as the FPR vector. The posterior density of $\beta[2]$ is depicted in Figure 1 and shows that the posterior probability is large in the neighborhood of zero.

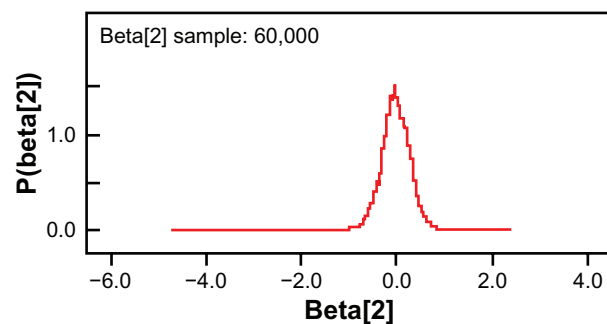


Figure 1 Posterior density of beta[2].

Simulation errors were quite small for all parameters which demonstrates that 60,000 observations are sufficient for a reliable posterior analysis. The above determination of the SROC values assumes beta[2] = 0, however, I performed the analysis assuming beta[2] is not zero and the results were quite similar to Table 1, and the analysis is executed with **BUGS CODE 2**.

BUGS CODE 2

```
# one test with beta[2] not zero
model;
{
  for(i in 1:N){tpr[i]<-(tp[i] + 0.5)/(tp[i] + fn[i] + 0.05)}
  for(i in 1:N){fpr[i]<-(fp[i] + 0.5)/(fp[i] + tn[i] + 0.05)}
  for(i in 1:N){u[i]<-logit(fpr[i])}
  for(i in 1:N){v[i]<-logit(tpr[i])}
  for(i in 1:N){b[i]<-v[i]-u[i]}
  for(i in 1:N){s[i]<-v[i] + u[i]}
  # bilogistic regression of b on s
  for(i in 1:N){b[i]~dlogis(mu[i],tau)}
  mu[i]<-beta[1] + beta[2]*s[i]}
  for(i in 1:2){beta[i]~dnorm(0.0000,0.0001)}
  tau~dgamma(0.0001,0.0001)
```

```
P<-1 + exp(-beta[1]/2)
# accuracy of test
Q<-1/P
# sroc curve, does not assume slope is zero
r1<-exp(-beta[1]/(1-beta[2]))
for(i in 1:N){r2[i]<-(1-fpr[i])/fpr[i]}
r3<-(1 + beta[2])/(1-beta[2])
for(i in 1:N){r4[i]<-pow(r2[i],r3)}
for(i in 1:N){r5[i]<-1 + r1*r4[i]}
for(i in 1:N){sroc[i]<-1/r5[i]}
}
# data from DeVries et al5 duplex mode
list(N = 8, tn = c(516,89,235,262,488,48,156,376),
  fn = c(28,8,23,20,14,7,2,31),
  fp = c(20,12,5,22,9,3,14,12),
  tp = c(78,59,75,89,118,48,39,121))
# initial values
list(beta = c(0,0), tau = 1))
```

Summary accuracy for scintigraphy

Our second example of a meta-analysis involves the diagnosis of osteomyelitis using antigranulocyte scintigraphy with ^{99m} radiolabeled monoclonal antibodies, and the analysis consists of 19 non-overlapping studies with reference standards of cell culture, histologic examination, and clinical follow-up. The study was conducted by Pakos et al⁹ with MEDLINE and Embase searches and a summary ROC curve was constructed, and the reader is referred to the article for additional important information as to the heterogeneity of the meta-analysis. Our approach is Bayesian and will be made up of determining the overall accuracy as expressed by the Q parameter, performing the regression analysis of the B scores on the S scores, and finally computing the posterior characteristics of the SROC values. Note, for this example, the input values are the true

Table 1 Posterior analysis for peripheral artery stenosis

Parameter	Mean	SD	Error	2 I/2	Median	97 I/2
Q	0.9104	0.0270	<0.0001	0.861	0.9132	0.946
beta[1]	4.7	0.551	0.0098	3.647	4.707	5.728
beta[2]	-0.0275	0.3372	0.0060	-0.6842	-0.0231	0.5995
sroc[1]	0.802	0.0857	0.0015	0.604	0.8149	0.9244
sroc[2]	0.9319	0.0455	<0.0001	0.8411	0.9399	0.9775
sroc[3]	0.7104	0.1046	0.0018	0.4735	0.7219	0.8781
sroc[4]	0.8946	0.0587	0.0011	0.7674	0.905	0.9636
sroc[5]	0.673	0.1102	0.0019	0.4277	0.62	0.8569
sroc[6]	0.8796	0.0636	0.00125	0.7384	0.8907	0.9577
sroc[7]	0.9017	0.0563	0.0011	0.7814	0.9116	0.9663
sroc[8]	0.7737	0.9023	0.0016	0.5607	0.7865	0.9109
tau	2.063	0.6882	0.0051	0.9117	1.995	3.602

and false positive rates of the 19 studies contained in the list statement of **BUGS CODE 3** below.

BUGS CODE 3

```
# one test
model;
{
for(i in 1:N){ u[i]<-logit(fpr[i])}
for(i in 1:N){ v[i]<-logit(tpr[i])}
for(i in 1:N){ b[i]<-v[i]-u[i]}
for(i in 1:N){ s[i]<-v[i] + u[i]}
# bilogistic regression of b on s
for(i in 1:N){ b[i]~dlogis(mu[i],tau)
mu[i]<-beta[1] + beta[2]*s[i]}
for(i in 1:2){beta[i]~dnorm(0.0000,0.0001)}
tau~dgamma(0.0001,0.0001)
P<-1 + exp(-beta[1]/2)
# Q value
Q<-1/P
# sroc curve assumes slope is 0
r1<-exp(-beta[1])
for(i in 1:N){r2[i]<-(1-fpr[i])/fpr[i]}
for(i in 1:N){r3[i]<-1 + r1*r2[i]}
for(i in 1:N){sroc[i]<-1/r3[i]}
}
```

data from Pakos et al⁹

```
list(N=19, tpr=c(0.67,0.75,0.85, 35,0.99,0.95,0.90,0.99,0.8
0,0.78,0.95,0.61,0.90,0.89,0.88,0.93,0.93,0.43,0.38), fpr=c
(0.15,0.05,0.11,0.17,0.22,0.33,0.33,0.01,0.33,0.60,0.43,0.04,
0.17,0.01,.10,0.70,0.01,0.01,0.57))
```

initial values

```
list( beta = c(0,0), tau = 1))
```

A Bayesian analysis is performed with 65,000 observations generated for the Monte Carlo Markov Chain (MCMC) simulation, with a burn in of 5000 and a refresh of 100 and the results appear in Table 2.

It should be noted the 19 SROC values correspond to the following (in that order) FPR ratios

```
fpr = (0.15,0.05,0.11,0.17,0.22,0.33,0.33,0.01,0.33,0.60,
0.43,0.04,0.17,0.01,0.10,0.70,0.01,0.01,0.57)
```

where the SROC values are computed according to formula (5), assuming $\beta[2] = 0$. The 95% confidence interval for $\beta[2]$ is $(-0.7051, 0.2755)$ and contains zero, thus, I let $\beta[2] = 0$ in the code for SROC of **BUGS CODE 3**. A regression of B on S values give a posterior mean (sd) for the intercept of 3.559 (0.5512) implying that the odds ratio for a positive test is 3.6 times larger for those patients with osteomyelitis than for those without the disease (when $S = 0$). Figure 2 is a plot of the SROC curve for the Pakos

Table 2 Bayesian analysis for the osteomyelitis meta-analysis

Parameter	Mean	SD	Error	2 I/2	Median	97 I/2
Q	0.8524	0.0346	<0.0001	0.7756	0.8555	0.9113
beta[1]	3.559	0.5512	0.0048	2.481	3.558	4.658
beta[2]	-0.2063	0.2454	0.0020	-0.7051	-0.2032	0.2755
sroc[1]	0.8486	0.0702	<0.0001	0.6783	0.8609	0.949
sroc[2]	0.6899	0.1189	0.0010	0.3861	0.6487	0.8473
sroc[3]	0.7994	0.08594	<0.0001	0.5963	0.8126	0.9287
sroc[4]	0.8661	0.064	<0.0001	0.7099	0.8778	0.9557
sroc[5]	0.8981	0.0511	<0.0001	0.7712	0.9082	0.9675
sroc[6]	0.9382	0.0335	<0.0001	0.8548	0.9453	0.9811
sroc[7]	0.9382	0.0335	<0.0001	0.8548	0.9453	0.9811
sroc[8]	0.2745	0.1052	<0.0001	0.1077	0.2616	0.5158
sroc[9]	0.9382	0.0335	<0.0001	0.8548	0.9453	0.9811
sroc[10]	0.9785	0.0127	<0.0001	0.9472	0.9813	0.9937
sroc[11]	0.9585	0.0235	<0.0001	0.9001	0.9636	0.9876
sroc[12]	0.5881	0.1243	<0.0001	0.3324	0.5938	0.8146
sroc[13]	0.8661	0.064	<0.0001	0.7099	0.8778	0.9557
sroc[14]	0.2745	0.1052	<0.0001	0.1077	0.2616	0.5158
sroc[15]	0.7826	0.0906	<0.0001	0.5704	0.7958	0.9214
sroc[16]	0.9861	0.0084	<0.0001	0.9654	0.9879	0.996
sroc[17]	0.2745	0.1052	<0.0001	0.1077	0.2616	0.5185
sroc[18]	0.2745	0.1052	<0.0001	0.1077	0.2616	0.5185
sroc[19]	0.9758	0.0142	<0.0001	0.9406	0.9789	0.9929
tau	0.7552	0.1549	<0.0001	0.4822	0.7445	1.087

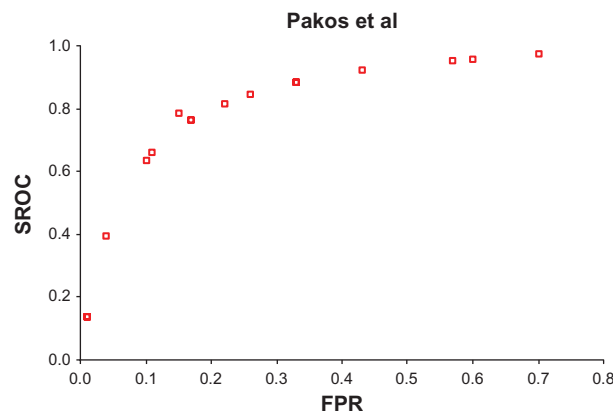


Figure 2 Summary receiver operating characteristic curve (Pakos et al).⁹

et al⁹ meta-analysis whose ordinates are the posterior means portrayed in Table 2.

What is the area under the curve? This is left as an exercise! It appears the accuracy implied by the SROC curve is quite good, which is confirmed by the value of the Q parameter which has a posterior mean 0.8524 and a 95% confidence interval of (0.7756, 0.9113).

Meta-analysis with two tests

When two tests are used to diagnose the same disease, the main emphasis is on comparing the accuracy of the two tests. There are many examples of a meta-analysis that summarize the accuracy of two sets of studies. The first set is a series of studies that use one test to diagnose disease, and the other set focuses on a series of tests with another test for diagnosing the same disease. For such a situation, two regressions of B on S are performed, two values of Q are computed and the SROC curve for each test is determined.

Thus, let

$$B_1 = \text{beta1}[1] + \text{beta1}[2]S_1 \quad (8)$$

be the first regression, and

$$B_2 = \text{beta2}[1] + \text{beta2}[2]S_2 \quad (9)$$

be the second, where

$$B_i = V_i - U_i$$

and the U_i and V_i are the logits of the TPR and FPR for the i -th test $i = 1, 2$.

See (1) and (2) for the formal definitions.

In a similar way, the accuracy of the i -th test is based on the Q parameters

$$Q_i = [1 + e^{-\text{beta}[2]/2}]^{-1} \quad (10)$$

for $i = 1, 2$.

Finally, the summary ROC curve is defined by its ordinates for the i -th test as

$$\begin{aligned} \text{SROC}_i(\text{FPR}) &= [1 + \exp(-\text{beta}[1]/(1-\text{beta}[2])) \\ &\times [(1-\text{FPR})/\text{FPR}]^{(1+\text{beta}[2])/\text{beta}[2]})]^{-1} \end{aligned} \quad (11)$$

corresponding to the FPR of the i -th test, where $i = 1, 2$. Note if the regression analysis implies $\text{beta}[2] = 0$, then one should modify (1) accordingly.

For the first example, consider the meta-analysis of Horsthuis et al⁸ which is a study of the use of ultrasound and MRI to diagnose inflammatory bowel disease, and nine studies are based on ultrasound and seven on MRI. On the basis of the meta-analysis, our purpose is to compare the two modalities. The authors found 1406 articles in a MEDLINE search and reduced the number to 16 for the present analysis and much information was extracted about the number of readers, the threshold values for a positive test, and other study covariates. A later section will be devoted to including study covariates into the determination of test accuracy. The Bayesian analysis will consist of performing regression analyses for the two modalities and determining if the slope coefficient is zero. Depending on the value of $\text{beta}[2]$, the appropriate formula for the SROC values for each test are used to determine the summary curve, and last the Q values are computed to compare the two modalities for accuracy.

Consider **BUGS CODE 4** below.

BUGS CODE 4

```
# two tests
model;
{
  # for test 1
  for(i in 1:N1){u1[i]<-logit(fpr1[i])}
  for(i in 1:N1){v1[i]<-logit(tpr1[i])}
  for(i in 1:N1){b1[i]<-v1[i]-u1[i]}
  for(i in 1:N1){s1[i]<-v1[i] + u1[i]}
  # bilogistic regression test 1
  for(i in 1:N1){b1[i]~dlogis(mu1[i],tau1)}
  mu1[i]<-beta1[1] + beta1[2]*s1[i]}
  for(i in 1:2){beta1[i]~dnorm(0.0000,0.0001)}
  tau1~dgamma(0.0001,0.0001)
  P1<-1 + exp(-beta1[1]/2)
  # accuracy of test 1
  Q1<-1/P1
  #sroc test 1, assumes slope is 0
  r11<-exp(-beta1[1])
```

```

for(i in 1:N1){r12[i]<-(1-fpr1[i])/fpr1[i]}
for(i in 1:N1){r13[i]<-1 + r11*r12[i]}
for(i in 1:N1){sroc1[i]<-1/r13[i]}
# for test 2
for(i in 1:N2){u2[i]<-logit(fpr2[i])}
for(i in 1:N2){v2[i]<-logit(tpr2[i])}
for(i in 1:N2){b2[i]<-v2[i]-u2[i]}
for(i in 1:N2){s2[i]<-v2[i] + u2[i]}
# bilogistic regression test 2
for(i in 1:N2){b2[i]~dlogis(mu2[i],tau2)}
mu2[i]<-beta2[1] + beta2[2]*s2[i]}
for(i in 1:2){beta2[i]~dnorm(0.0000,0.0001)}
tau2~dgamma(0.0001,0.0001)
#sroc test 2, assumes slope is zero
r21<-exp(-beta2[1])
for(i in 1:N2){r22[i]<-(1-fpr2[i])/fpr2[i]}
for(i in 1:N2){r23[i]<-1 + r21*r22[i]}
for(i in 1:N2){sroc2[i]<-1/r23[i]}
P2<-1 + exp(-beta2[1]/2)
# accuracy of test 2
Q2<-1/P2
# difference in accuracy of two tests
d<-Q1-Q2
}
# below are data from DeVries et al5 for duplex and color
# duplex is test 1
# color is test 2
list(N1 = 8, fpr1 = c(0.04,.12,0.02,0.08,0.02,0.06,
0.08, 0.03),
tpr1 = c(0.74,0.88,0.77,0.82,0.89,0.87,0.95,0.80),
N2 = 6, fpr2 = c(0.01,0.01,0.02,0.06,0.05,0.05),
tpr2 = c(0.90,0.99,0.88,0.89,0.99,0.96))
# Horsthuis et al8 study
# test 1 is ultrasound (US) and test 2 is MR
list (N1 = 9, fpr1 = c(0.10,0.01,0.06,0.12,0.07,0.07,0.01,
0.33,0.04),
tpr1 = c(0.78,0.93,0.90,0.81,0.93,0.88,0.87,0.96,0.92),
N2 = 7, fpr2 = c(0.01,0.01,0.39,0.39,0.01,0.08,0.15),
tpr2 = c(0.99,0.99,0.91,0.87,0.82,0.95,0.89))
# the initial values
list(beta1 = c(0,0), tau1 = 1, beta2 = c(0,0), tau2 = 1).

```

The second list statement is the information for the Horsthuis et al study, where the first test is ultrasound and the second is MRI. The third list statement is for the initial values of the parameters. A Bayesian analysis is executed with 65,000 observations for the simulations, with a burn in of 5000 and a refresh of 100.

The two Q values are quite close, where for ultrasound, the posterior mean is 0.9098 and that for MRI is 0.9324 and the difference d between the two has a 95% confidence interval of (-0.0956, 0.0991) implying that the accuracy of US and MRI are about the same, which is confirmed to some extent by a plot of the two SROC curves portrayed by Figure 3.

The Bayesian regression analysis indicates that the slope of each test is zero, thus, the SROC values for both tests are computed assuming $\beta_1[2]$ and $\beta_2[2]$ are zero. Also, the intercept term $\beta_1[1]$ has a posterior mean of 4.668, implying that the odds (when $S_1 = 0$) of a positive MRI test result for those with inflammatory bowel disease are 4.6 times that of the odds of a positive result for those without the disease. Also since the two slopes are close to zero, the implication is that both tests are differentiating equally between disease and non-disease for all values of the threshold. See Horsthuis et al⁸ for additional information about the choice of threshold for the various studies for each modality.

Meta-analysis with study covariates and one test

Often the information in a meta-analysis contains information about the various studies and should be included in the meta-analysis. Of course, patient covariates and individual study information do indeed affect the accuracy of the medical test being assessed, and need to be taken into account when estimating test accuracy. This will be done using the following bilogistic regression model for the meta-analysis of one test:

$$B = \beta[1] + \beta[2]S + \sum_{i=1}^{i=k} \eta_i X_i, \quad (12)$$

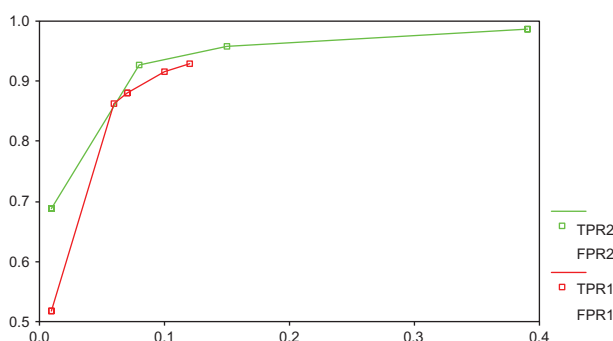


Figure 3 Summary receiver operating characteristic curve (Horsthuis et al⁸ meta-analysis).

Note: The green curve corresponds to US and the red to MRI.

where $B = V - U$, $S = V + U$, and the X_i are k study covariates, while the parameters are unknown. Thus, for a given study of the analysis there are k covariates such as age, the number of readers, the percentage of males to females, and other information. Of course, it depends on the meta-analysis just what information is available and often the information is not available for some studies. The information for a particular meta-analysis will consist of a column for the false positive rate, the true positive rate, and a separate column for each study covariate. Once the regression is performed, the SROC ordinates can be computed along with the accuracy parameter Q .

The meta-analysis of Horsthuis et al⁸ provides an excellent example of using covariates to estimate the summary accuracy of ultrasound in order to diagnose inflammatory bowel disease where the fraction of patients with Crohn's disease and the fraction of males per study serve as covariates. Recall in the previous example with the ultrasound information of Horsthuis et al,⁸ the accuracy of ultrasound was 0.9098 as measured by the posterior mean of the Q parameter. Will the accuracy change when study age and fraction of males is included in the analysis? In order to answer that question the Bayesian regression (12) is executed with 65,000 observations for the simulation, with a burn in of 5000 and a refresh of 100. Note the first list statement of **BUGS CODE 5** contains the necessary information to execute the Bayesian analysis, where the emphasis will be on assessing the effects of the covariates on B and using Q and the SROC values to assess the accuracy of ultrasound to detect inflammatory bowel disease.

BUGS CODE 5

```
# one test with covariates
model;
{
  for(i in 1:N){u[i]<-logit(fpr[i])}
  for(i in 1:N){v[i]<-logit(tpr[i])}
  for(i in 1:N){b[i]<-v[i]-u[i]}
  for(i in 1:N){s[i]<-v[i] + u[i]}
  # bilogistic regression of b on s
  for(i in 1:N){b[i]~dlogis(mu[i],tau)}
  mu[i]<-beta[1] + beta[2]*s[i] + neta[1]*x1[i] + neta[2]*
  x2[i]}
  for(i in 1:2){beta[i]~dnorm(0.0000,0.0001)}
  for(i in 1:2){neta[i]~dnorm(0.0000,0.0001)}
  tau~dgamma(0.0001,0.0001)
  P<-1 + exp(-beta[1]/2)
  # accuracy of test Q<-1/P
  # sroc curve assumes slope is 0
```

```
r1<-exp(-beta[1])
for(i in 1:N){r2[i]<-(1-fpr[i])/fpr[i]}
for(i in 1:N){r3[i]<-1 + r1*r2[i]}
for(i in 1:N){sroc[i]<-1/r3[i]}
}
# data from Horsthuis et al8
# x1 is fraction with Crohn's disease
# x2 is fraction of males
# true and false positive ratios are on per patient basis
list(N = 9, tpr = c(0.78,0.935,0.90,0.812,0.931,0.884,
0.87,0.96,0.92),
fpr = c(0.10,0.01,0.05,0.12,0.07,0.07,0.01,0.33,0.03),
# the first component of x2 is the average of the
remaining 8
x1 = c(0.322,0.093,0.322,1,0.406,0.643,1,1,0.642),
x2 = c(0.4088,0.097,0.457,0.571,0.423,0.465,0.366,
0.571,0.321))
# Pakos et al
list(N = 19,
tpr = c(0.67,0.75,0.85,0.35,0.99,0.95,0.90,0.99,0.80,0.78,0.
.95,0.61,0.90,0.89,0.88,0.93,0.93,0.43,0.38),
fpr = c(0.15,0.26,0.11,0.17,0.22,0.33,0.33,0.01,0.33,0.6,
0.43,0.04,0.17,0.01,0.10,0.70,0.01,0.01,0.57),
# x1 is the average age
# x2 is the percentage of males
# components 2,3,4,14, and 17 of x1 were given the age 54,
the average of remaining
# components 2,3,4,14, and 17 of x2 were given the value
0.70, the average of remaining
x1 = c(59,54,54,54,48,57,61,56,66,45,58,48,47,54,60,58,5
4,48,46),
x2 = c(83,70,70,70,81,53,68,67,86,72,82,76,71,70,58,68,7
0,59,67))
# initial values
list(beta = c(0,0), neta = c(0,0), tau = 1))
```

The surprising result is that the adjusted accuracy of ultrasound is 0.9748 with covariates compared with 0.9098 without covariates, even though the effect of the covariates is somewhat negligible. For example, the 95% confidence interval for $neta[1]$, which is the effect for the proportion of patients with ulcerated colitis, is $(-1.916, 3.858)$ which implies a small effect of that covariate on the average value of B . Also, it should be noted that the 95% credible interval for $beta[2]$ contains zero, thus, the ordinates of the points on the SROC curve are computed assuming $beta[2] = 0$ (see formula (5)).

The effect of the covariates on the SROC curve is evident from Figure 4, because the SROC curve corresponding to the

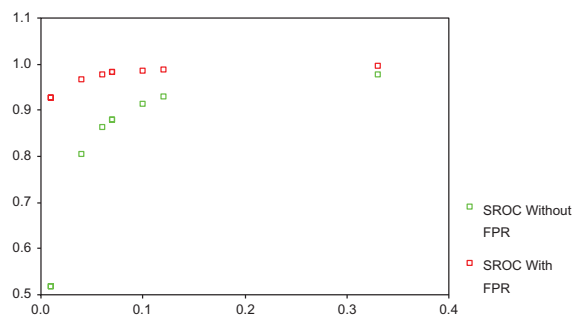


Figure 4 Effect of covariates on Summary receiver operating characteristic (Horsthuis et al).⁸

inclusion of the two covariates dominates the SROC curve when the covariates are not included over the range where $0 < \text{fpr} < 0.35$.

Meta-analysis with covariates for several tests

A Bayesian analysis provides an incisive approach to comparing two medical tests with covariates when the information is provided by a meta-analysis, and the Horsthuis et al⁸ article presents such an example. It consists of two modalities, ultrasound and MRI, to diagnose inflammatory bowel disease, where the two covariates are the fraction of patients with Crohn's disease and the fraction of male patients. The disease presents as Crohn's disease or as ulcerative colitis, while the disease attacks both genders with about the same frequency. The analysis is based on **BUGS CODE 6** listed below.

BUGS CODE 6

```
# two tests with covariates
model;
{
  for(i in 1:N1){u1[i]<-logit(fpr1[i])}
  for(i in 1:N1){v1[i]<-logit(tpr1[i])}
  for(i in 1:N1){b1[i]<-v1[i]-u1[i]}
  for(i in 1:N1){s1[i]<-v1[i] + u1[i]}
  # bilogistic regression of b on s for test 1
  for(i in 1:N1){ b1[i]~dlogis(mu1[i],tau1)}
  mu1[i]<-beta1[1] + beta1[2]*s1[i] + neta1[1]*x11[i] + neta1[2]*x12[i]}
  for(i in 1:2){ beta1[i]~dnorm(0.0000,0.0001)}
  for(i in 1:2){ neta1[i]~dnorm(0.0000,0.0001)}
  tau1~dgamma(0.0001,0.0001)
  P1<-1 + exp(-beta1[1]/2)
  # accuracy of test 1
  Q1<-1/P1
  # sroc curve assumes slope is 0
```

```
# sroc values for test 1
r11<-exp(-beta1[1])
for(i in 1:N1){r12[i]<-(1-fpr1[i])/fpr1[i]}
for(i in 1:N1){r13[i]<-1 + r11*r12[i]}
for(i in 1:N1){sroc1[i]<-1/r13[i]}
# for test 2
for(i in 1:N2){u2[i]<-logit(fpr2[i])}
for(i in 1:N2){v2[i]<-logit(tpr2[i])}
for(i in 1:N2){b2[i]<-v2[i]-u2[i]}
for(i in 1:N2){s2[i]<-v2[i] + u2[i]}
# bilogistic regression of b2 on s2 for test 2
for(i in 1:N2){b2[i]~dlogis(mu2[i],tau2)}
mu2[i]<-beta2[1] + beta2[2]*s2[i] + neta2[1]*x21[i] + neta2[2]*x22[i]}
for(i in 1:2){beta2[i]~dnorm(0.0000,0.0001)}
for(i in 1:2){neta2[i]~dnorm(0.0000,0.0001)}
tau2~dgamma(0.0001,0.0001)
P2<-1 + exp(-beta2[1]/2)
# accuracy of test 2
Q2<-1/P2
# sroc curve assumes slope is 0
r21<-exp(-beta2[1])
for(i in 1:N2){r22[i]<-(1-fpr2[i])/fpr2[i]}
for(i in 1:N2){r23[i]<-1 + r21*r22[i]}
for(i in 1:N2){sroc2[i]<-1/r23[i]}
# difference of accuracy of two tests
d<-Q1-Q2
}
# data from Horsthuis et al8 for US and MRI
# US is test 1 and MRI is test 2
# x11 is fraction with Crohn's disease
# x12 is fraction of males
# tpr and fpr are on a per patient basis
list(N1 = 9, tpr1 = c(0.78,0.935,0.90,0.812,0.931,0.884,0.87,0.96,0.92),
fpr1 = c(0.10,0.01,0.05,0.12,0.07,0.07,0.01,0.33,0.03),
# the first component of x12 is the average of the remaining 8
x11 = c(0.322,0.093,0.322,1,0.406,0.643,1,1,0.642),
x12 = c(0.4088,0.097,0.457,0.571,0.423,0.465,0.366,0.571,0.321),
# the following is for test 2 or MRI
N2 = 7,
x21 = c(0.6,0.54,1,1,0.34,0.36,1),
# below 0.49 is the average of the other six components of x22
x22 = c(0.6,0.42,0.46,0.36,0.49,0.56,0.52),
# tpr and fpr are on a per patient basis
```

```

tpr2 = c(0.99,0.99,0.913,0.87,0.818,0.956,0.889),
fpr2 = c(0.18,0.01,0.29,0.29,0.01,0.08,0.143))
# initial values
list( beta1 = c(0,0),neta1 = c(0,0), tau1 = 1,beta2 = c(0,0),
neta2 = c(0,0), tau2 = 1))

```

The analysis consists of two bilogistic regressions corresponding to the two tests.

$$B_1 = \beta_1[1] + \beta_1[2]S_1 + \sum_{i=1}^{i=k} \eta_{1i} X_{1i} \quad (13)$$

for the first test and

$$B_2 = \beta_2[1] + \beta_2[2]S_2 + \sum_{i=1}^{i=k} \eta_{2i} X_{2i} \quad (14)$$

for the second, where the same k covariates apply to both tests. The B and S values is defined the usual way in terms of U and V values, which are the logits of the false and true positive fractions respectively. See (1)–(3) for the definitions of the U , V , B , and S values used in the above regression.

Once the regressions are performed, the posterior distribution of the regression parameters induce a posterior distribution for the accuracy parameters, namely

$$Q_1 = (1 + e^{-\beta_1[1]/2})^{-1} \quad (15)$$

for the first test and

$$Q_2 = (1 + e^{-\beta_2[1]/2})^{-1} \quad (16)$$

for the second.

For the last part of the analysis, the posterior distribution of the SROC values is generated which allows one to compare SROC curves and hence to compare graphically the accuracy of the two tests. Formula (11) should be modified in order to compute the SROC values for the two tests.

The analysis is executed with 325,000 observations with a burn in of 5000 and a refresh of 100. Comparing US and MRI is based on Q_1 and Q_2 and indicates that (based on the posterior means) US has more accuracy, but this comparison should be made with caution because the posterior distribution of Q_2 is highly skewed with a median for Q_2 of 0.998 and a posterior mean of 0.866, thus if the comparison is made with the means there appears to be a difference, but if based on the medians there appears to be no difference. When the two SROC curves are plotted for the two tests implying that the US test has more accuracy than MRI.

Based on the 95% confidence interval, the slope of the two regressions is zero, that is, the effect of the S values on the B value is negligible, but note that the intercept for the regression of US is 8.106 which implies that the odds of a positive test with US for the diseased population is 8 times the odds of a positive US for those without inflammatory disease, whereas, for MRI the intercept posterior mean is 12.15 implying the odds of a positive test with MRI for the diseased group is about 12 times the odds of a positive test for the non diseased group. Also, based on the 95% confidence interval, for all values of the threshold the both tests discriminate equally between those with inflammatory bowel disease and those without.

Do the covariates have any effect on the accuracy of the two tests? The 95% credible interval for $neta1[2]$ is (−17.69, −1.522) which implies that the percentage of male patients has an effect on the B score for US, however, additional analysis is needed in order to determine if the effect is non-negligible by calculating the Q value for US when covariates are not used in the regression.

One note of caution in using regression for the meta-analysis to assess the accuracy is that the relatively small sample size (the number of studies) might not produce a model with a good fit to the data, which in turn implies that one may not have high confidence in one's assessment of the accuracy of the medical test.

Other meta-analyses

Up to this point, the emphasis has been on summarizing the accuracy of various medical tests, but now the focus will be on estimating the complications of various medical tests. As is well-known, many tests are accompanied by various complications: (a) Coronary angiography can result in stroke and damage to the coronary arteries, and (b) Certain contrast media used in diagnostic imaging can result in damage to the kidneys. It is for the latter scenario that a Bayesian approach will be taken and is based on the meta-analysis of Heinrich et al¹⁰ who compare the nephrotoxicity of iso-osmolar iodixanol with nonionic low-osmolar contrast media.

The meta-analysis included a thorough search of MEDLINE, Embase, and BIOSIS databases, trial registries, conference proceedings and requests from companies. Also, randomized clinical trials assessing the serum creatinine levels before and after the intravascular administration of iodixanol were included. The main endpoint was the incidence of contrast media-induced nephropathy (kidney disease), which is measured by a change in serum creatinine

values. Their main conclusion was based on 25 trials and on conventional statistical procedures that showed that iodixanol is not associated with significantly reduced risk of contrast induced nephropathy.

Our approach is Bayesian, where the main endpoint is the incidence of kidney complications (defined as a certain percent increase in serum creatinine, which is measured before and after the administration of the procedure). There are two groups, the one administered contrast media, labeled LOCM, and the other group of patients to which contrast media was not administered, referred to as the iodixanol group. Some of the trials were randomized, and there are various types of contrast media, as well as many other study covariates that are of interest in how they affect the complication rate.

Suppose the analysis is based on the logistic model, where for the control group (patients not receiving the contrast media)

$$X_i \sim \text{binomial}(n_i, p_i) \quad (17)$$

for $i = 1, 2, \dots, n$,

and for the patients receiving the contrast agent

$$Y_i \sim \text{binomial}(m_i, q_i) \quad (18)$$

with $i = 1, 2, \dots, m$.

In addition, let

$$\log(p_i/(1 - p_i)) = \theta_i \quad (19)$$

for $i = 1, 2, \dots, n$,

and

$$\log(q_i/(1 - q_i)) = \phi_i + \theta_i \quad (20)$$

with $i = 1, 2, \dots, m$.

Prior information for the parameters is specified as

$$\theta_i \sim \text{nid}(\mu_\theta, \tau_\theta) \quad (21)$$

and

$$\phi_i \sim \text{nid}(\mu_\phi, \tau_\phi). \quad (22)$$

At this stage there are several choices: One could let the θ_i and ϕ_i have noninformative normal distributions (0.0000, 0.00001), or let the μ_θ and μ_ϕ have normal distributions (informative or non-informative) and the precisions τ_θ and τ_ϕ have gamma distributions (informative or non informative). It depends on the prior information available to the analyst. For additional information about the model, see Berry.¹¹ The first part of the analysis will be based on the model defined by (17)–(22), then the analysis will be expanded to include the study covariates.

Our analysis is executed with **BUGS CODE 7** with 65,000 observations for the simulation, a burn in of 5000, and a refresh of 100.

BUGS CODE 7

model;

```
{
# binomial distributions for two tests
for(i in 1:17){x[i]~dbin(p[i],m[i])}
for(i in 1:17){y[i]~dbin(q[i],n[i])}
# logit models for the two complication rates
for(i in 1:17){logit(p[i])<-theta[i]}
for(i in 1:17){logit(q[i])<-phi[i]}
# prior distributions
for(i in 1:17){theta[i]~dnorm(muth,tauth)}
for(i in 1:17){phi[i]~dnorm(muphi,tauphi)}
muth~dnorm(0.000,0.0001)
muphi~dnorm(0.000,0.0001)
tauth~dgamma(0.00001,0.00001)
tauphi~dgamma(0.00001,0.00001)
# mean of complication rates
pee<-mean(p[])
qee<-mean(q[])
d<-pee-qee
}
# Heinrich et al10 meta analysis
list(m = c(123,72,54,210,58,105,76,134,35,25,64,54,
32,100,20,59,101),
n = c(125,76,48,204,56,116,77,125,35,25,65,49,32,50,19,
60,99),
x = c(6,5,1,14,0,4,2,12,2,4,2,5,1,7,1,0,0),
y = c(7,0,5,9,1,1,0,17,9,4,17,14,0,3,0,0,2))
# initial values
# must initiate other chain with specification tool
list(muth = 0,muphi = 0,tauth = 1,tauphi = 1)
```

The above code is labeled with the appropriate identification for the Bayesian analysis, such as pee and qee are the means of the complication rates for the two groups of patients, and the data for the Heinrich et al¹⁰ meta-analysis where the components of the x vector are the number of complications for the control or iodixanol patients and the y vector contains the number of complications for the treatment for the patients receiving the contrast media. In addition the m vector is the number of patients in the control group and the n vector contains the number of patients in the treatment group. Note that only 17 of the 22 studies are used for the analysis because they are the ones with complete information on the complication rates. Table 3 portrays the Bayesian analysis for the Heinrich et al¹⁰ meta-analysis.

Table 3 Posterior analysis for the inflammatory bowel disease meta-analysis

Parameter	Mean	SD	Error	2 I/2	Median	97 I/2
Q1	0.9098	0.0203	<0.0001	0.8629	0.9123	0.942
Q2	0.9324	0.0493	<0.0001	0.8148	0.9426	0.9863
beta1[1]	4.668	0.4757	0.0048	3.679	4.683	5.575
beta1[2]	-0.5686	0.3238	0.0034	-1.242	-0.5631	0.0664
beta2[1]	5.647	1.418	0.0129	2.963	5.597	8.557
beta2[2]	-0.901	1.003	0.0112	-2.929	-0.8595	0.9871
d	-0.02263	0.05346	<0.0001	-0.0956	-0.03004	0.09919
sroc1[1]	0.9151	0.0409	<0.0001	0.8148	0.9232	0.967
sroc1[2]	0.5176	0.1107	0.0011	0.2857	0.522	0.7271
sroc1[3]	0.8626	0.0594	<0.0001	0.7165	0.8734	0.9439
sroc1[4]	0.9294	0.03521	<0.0001	0.8438	0.9365	0.9729
sroc1[5]	0.8805	0.0535	<0.0001	0.7488	0.8906	0.952
sroc1[6]	0.8805	0.0535	<0.0001	0.7488	0.8906	0.952
sroc1[7]	0.5176	0.1107	0.0011	0.2857	0.522	0.7271
sroc1[8]	0.9791	0.0120	<0.0001	0.9512	0.9816	0.9924
sroc1[9]	0.806	0.0754	<0.0001	0.6227	0.8184	0.9166
sroc2[1]	0.6875	0.2199	0.0019	0.1636	0.7314	0.9813
sroc2[2]	0.6875	0.2199	0.0019	0.1636	0.7314	0.9813
sroc2[3]	0.986	0.0410	<0.0001	0.9253	0.9942	0.9997
sroc2[4]	0.986	0.0410	<0.0001	0.9253	0.9942	0.9997
sroc2[5]	0.6875	0.2199	0.0019	0.1636	0.7314	0.9813
sroc2[6]	0.926	0.1041	<0.0001	0.6274	0.9591	0.9978
sroc2[7]	0.9581	0.0748	<0.0001	0.7736	0.9794	0.9989
tau1	1.538	0.4736	0.0030	0.735	1.498	2.588
tau2	0.5946	0.2134	0.0015	0.2411	0.5732	1.072

Key parameters are d , the difference in the two averages of the complication rates, and μ_{phi} , which is the difference in the logits of the complication rates, and both imply there is a difference in the complication rates. The average complication rate for the group not receiving the contrast agent is 0.0489, compared with 0.0826 for the one receiving the contrast agent, and their difference d has a 95% confidence interval of $(-0.0556, -0.0125)$, while the 95% credible interval for μ_{phi}

is $(-4.532, -2.365)$. Note the large posterior mean for τ_{ϕ} ; however, the median is 5.058, which should be used as the estimate of the location of the precision parameter. This example is easily generalized to include study covariates in the exercises.

Comments and conclusions

This paper introduces the reader to the Bayesian approach to meta-analysis with emphasis on the summary ROC curve,

Table 4 Bayesian analysis for meta-analysis of ultrasound for diagnosis of inflammatory bowel disease

Parameter	Mean	SD	Error	2 I/2	Median	97 I/2
Q	0.9748	0.0394	0.0015	0.9034	0.9826	0.996
beta[1]	7.978	1.591	0.0827	4.472	8.006	11.04
beta[2]	-0.0813	0.3484	0.0141	-0.8162	-0.0714	0.5889
neta[1]	0.9398	1.443	0.0598	-1.916	0.988	3.858
neta[2]	-8.634	4.293	0.2369	-16.65	-8.773	0.4042
sroc[1]	0.9868	0.0528	0.0021	0.9067	0.9972	0.9999
sroc[2]	0.9262	0.1329	0.0064	0.4691	0.9698	0.9984
sroc[3]	0.9773	0.0749	0.0030	0.8216	0.9941	0.9997
sroc[4]	0.9887	0.0546	0.0019	0.9227	0.9977	0.9999
sroc[5]	0.9825	0.0662	0.0026	0.8682	0.9958	0.9998
sroc[6]	0.9825	0.0662	0.0026	0.8682	0.9958	0.9998
sroc[7]	0.9262	0.1329	0.0064	0.4691	0.9698	0.9984
sroc[8]	0.9954	0.0384	0.0012	0.9773	0.9994	1.0
sroc[9]	0.9665	0.0903	0.0039	0.7302	0.9899	0.9995
Tau	2.167	0.7886	0.0190	0.8561	2.089	3.919

Table 5 Bayesian analysis for comparing US and MRI for inflammatory bowel disease

Parameter	Mean	SD	Error	2 I/2	Median	97 I/2
Q1	0.9774	0.0335	<0.0001	0.9318	0.9828	0.9962
Q2	0.866	0.2919	0.0109	0.00021	0.998	1
beta1[1]	8.106	1.53	0.0442	5.231	8.093	11.13
beta1[2]	-0.058	0.3414	0.0075	-0.7215	-0.0595	0.6222
beta2[1]	12.15	12.58	0.4865	-16.93	12.45	35.1
beta2[2]	0.2252	0.8867	0.0199	-1.656	0.2409	1.931
d	0.1114	0.2933	0.019	-0.05664	-0.0007	0.9798
neta1[1]	1.195	1.43	0.0313	-1.568	1.172	4.005
neta1[2]	-9.288	4.252	0.1304	-17.69	-9.193	-1.522
neta2[1]	-6.287	5.884	0.1805	-17.61	-6.324	6.22
neta2[2]	-5.574	20.13	0.7688	-42.63	-6.059	40.85
sroc1[1]	0.9903	0.0489	0.00125	0.9541	0.997	0.9999
sroc1[2]	0.9392	0.1068	0.0028	0.6538	0.9706	0.9985
sroc1[3]	0.9828	0.0609	0.0015	0.9077	0.9942	0.9997
sroc1[4]	0.9916	0.0462	0.0011	0.9623	0.9978	0.9999
sroc1[5]	0.9869	0.0547	0.0014	0.9336	0.996	0.9998
sroc1[6]	0.9869	0.0547	0.0014	0.9336	0.996	0.9998
sroc1[7]	0.9392	0.1068	0.0028	0.6538	0.9706	0.9985
sroc1[8]	0.9966	0.0338	0.000821	0.9893	0.9994	1
sroc1[9]	0.9742	0.0723	0.0019	0.8525	0.9902	0.9995
sroc2[1]	0.852	0.3289	0.01212	0	1	1
sroc2[2]	0.7841	0.377	0.0140	0	0.9996	1
sroc2[3]	0.8627	0.3192	0.0117	0	1	1
sroc2[4]	0.8627	0.3192	0.0117	0	1	1
sroc2[5]	0.7841	0.377	0.0140	0	0.9996	1
sroc2[6]	0.8344	0.3436	0.0127	0	1	1
sroc2[7]	0.847	0.3332	0.0123	0	1	1
tau1	2.174	0.7841	0.0056	0.8769	2.094	3.93
tau2	0.8198	0.3824	0.0066	0.2321	0.7694	1.699

that is, where one assumes that the separate studies have a common ROC curve. The Bayesian approach is based on a regression analysis and consists of computing the posterior distribution of the SROC values and the accuracy Q of the combined studies. It should be noted that Miller et al¹² gave one of the first accounts of the Bayesian approach to meta-analysis, including the use of the SROC curve, and this presentation is very similar, but goes beyond what they presented to include more complex study designs. The author presents an approach to meta-analysis based on the Miller method; however, there are other models that could have been employed. For example, the approach of Paul et al¹³ is a valuable contribution and an alternative to the Miller

method. The reader should also refer to the following studies: Harbord et al,¹⁴ Leeflang et al,¹⁵ Reitsma et al,¹⁶ and Rutter and Gatsonis.¹⁷ For the practitioner, the Leeflang et al study is quite informative.

This paper began by considering a meta-analysis that involves only one test and via a MCMC simulation determines the posterior distribution of the relevant parameters. A regression of B on S is the basis of the analysis, where the intercept and slope give us valuable information about the worth of the test in differentiating between diseased and non-diseased patients. $B = V - U$ and $S = V + U$, where U and V are the logits of the FPF and TPF, respectively. It can be shown that the intercept is the odds ratio of a positive test result for those with and

Table 6 Bayesian analysis for the safety of contrast media study

Parameter	Mean	SD	Error	2 I/2	Median	97 I/2
d	-0.0337	0.01098	<0.0001	-0.0556	-0.0336	-0.0125
muphi μ_ϕ	-3.319	0.547	0.0033	-4.532	-3.274	-2.365
muth μ_θ	-3.085	0.2284	0.0050	-3.617	-3.057	-2.722
pee	0.0489	0.0063	<0.0001	0.0374	0.0486	0.0622
qee	0.0826	0.0089	<0.00001	0.0660	0.0823	0.1009
tauphi τ_ϕ	0.3825	0.2161	0.0017	0.1003	0.3378	0.9219
tauth τ_θ	2596	14400	429	0.6974	5.058	28040

without the disease (when $S = 0$), and that the value of the slope is very informative about the accuracy. For example, if the slope is close to zero the implication is that the test's ability to discriminate between the two groups of patients is independent of the threshold values used in the separate studies. Once the posterior distribution of the slope and intercept are determined, their joint posterior distribution induces a posterior distribution for the SROC values and the accuracy parameter Q .

Various interesting examples exemplify the Bayesian analysis and are taken from the medical research literature. For example, the Horsthuis et al⁸ study is especially interesting because it is well documented and involves several tests, namely US, MR, scintigraphy, and CT, to diagnose inflammatory bowel disease, and is one of the few meta-analyses to study several modalities simultaneously. If covariates are involved, the posterior distribution of the relevant regression coefficients tells one if they contribute to the analysis, and another positive feature of the Horsthuis et al⁸ analysis is the inclusion of many covariates to assess the inter-study variation.

The Horsthuis et al⁸ example is examined in three stages: (a) US was only considered without any covariates, (b) US and MR were compared, without using any covariates, and (c) US was used along with two covariates, the fraction of patients with Crohn's disease, and the fraction of male patients. Thus, this example allows one to display the full complement of Bayesian methods in order to conduct the analysis.

As mentioned earlier, the foundation of the Bayesian approach is the bilogistic regression model of B on S ; however, Zhou et al (p 403)³ describe another method based on a binary regression model. This was not considered, but instead the bilogistic regression model was employed. Also, very little has appeared in the literature about the meta-analysis of tests with continuous scores from a Bayesian viewpoint, although Zhou et al (p 409)³ do present such an approach, based on a weighted average of estimated ROC areas of the various studies. Note, it would be very difficult to do a Bayesian analysis, because most if not all such meta-analyses are not done in a Bayesian fashion, but instead are analyzed with non-Bayesian methodology, that is, each individual ROC area is estimated by conventional methods along with an associated estimated standard error (standard deviation).

Disclosure

The author declares no conflicts of interest in this work.

References

1. Kardaun JW, Kardaun OJ. Comparative diagnostic performance of three radiological procedures for the detection of lumbar disk herniation. *Methods Inf Med.* 1990;29:12–22.
2. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med.* 1993;12:1293–1316.
3. Zhou Xiao-Hua, Obuchowski N, McClish DK. *Statistical Methods in Diagnostic Medicine*. New York: John Wiley and Sons Inc; 2002.
4. Stangl DK, Berry DA. *Meta-Analysis in Medicine and Health Policy*. New York: Marcel Dekker Inc; 2000.
5. De Vries SO, Hunink MG, Polak JF. Summary receiver operating characteristic curves as a technique for meta-analysis for the diagnostic performance of duplex ultrasonography in peripheral artery disease. *Acad Radiol.* 1996;3:361–369.
6. Meijer AB, O YL, Geleijns J, Kroft LJ. Meta-analysis of 40- and 64-MDTC angiography for assessing coronary artery stenosis. *AJR Am J Roentgenol.* 2008;191:1667–1675.
7. Vanhoenacker PK, Heijenbroek-Kal MH, Van Heste R, et al. Diagnostic performance of multidetector CT angiography for assessment of coronary artery disease. *Radiology.* 2007;244:419–428.
8. Horsthuis K, Bipat S, Bennisink RJ, Stoker J. Inflammatory bowel disease diagnosed with US, MR, scintigraphy, and CT: Meta-analysis of prospective studies. *Radiology.* 2008;247:64–79.
9. Pakos EE, Koumoulis HD, Fotopoulos AD, Ioannidis JP. Osteomyelitis: antigranulocyte scintigraphy with ^{99m}Tc radiolabeled monoclonal antibodies for diagnosis-meta-analysis. *Radiology.* 2007;245:732–741.
10. Heinrich MC, Haberle L, Müller V, Bautz W, Uder M. Nephrotoxicity of iso-osmolar iodixanol compared with nonionic low-osmolar contrast media: meta-analysis of randomized controlled trials. *Radiology.* 2009; 250:68–85.
11. Berry SM. Meta-analysis versus large trials: resolving the controversy: In Stangl DK, Berry DA, editors. *Meta-Analysis in Medicine and Health Policy*. New York: Marcel-Dekker: 2000:65–81.
12. Miller SW, Sinha D, Slate EH, Garrow D, Romagnuolo J. Bayesian adaptation of the summary ROC curve method for meta-analysis of diagnostic test performance. *J Data Sci.* 2009;7:349–364.
13. Paul M, Riebler A, Bachman LM, Rue H, Held L. Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. *Stat Med.* 2010;29:1325–1339.
14. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic test accuracy studies. *Biostatistics.* 2007;8:239–251.
15. Leeflang MMG, Deeks JJ, Gatsonis C, Bossuyt PMM; on behalf of the Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Ann Int Med.* 2008;149:889–897.
16. Reitsma JB, Glas AS, Rutjes AW, Scholten RJP, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol.* 2005;58:982–990.
17. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med.* 2001;20:2865–2884.

Open Access Medical Statistics

Publish your work in this journal

Open Access Medical Statistics is an international, peer-reviewed, open access journal publishing original research, reports, reviews and commentaries on all areas of medical statistics. The manuscript management system is completely online and includes a very quick and fair

Submit your manuscript here: <http://www.dovepress.com/open-access-medical-statistics-journal>

Dovepress

peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.