Hierarchical Summary ROC Analysis: A Frequentist-Bayesian Colloquy in Stata

Ben A. Dwamena, MD

The University of Michigan Radiology & VAMC Nuclear Medicine, Ann Arbor, Michigan

Stata Conference, Chicago, Illinois - July 11-12, 2019
Outline

1. Diagnostic Test Evaluation
2. Methods for Meta-analysis of Binary Data
3. Hierarchical SROC Analysis
4. Frequentist Hierarchical SROC Analysis
5. Bayesian Hierarchical SROC Analysis
6. Concluding Remarks
Medical Diagnostic Test

Any measurement aiming to identify individuals who could potentially benefit from preventative or therapeutic intervention

This includes:

1. Elements of medical history e.g. Retrosternal chest pain
2. Physical examination e.g. Systolic blood pressure
3. Imaging procedures e.g. Chest xray
4. Laboratory investigations. e.g. Fasting blood sugar
5. Clinical prediction rules e.g. Geneva Score for Venous Thromboembolism
Diagnostic Test Types/Scales

1. **Dichotomous** using single implicit or explicit threshold
   - eg. Presence or absence of a specific DNA sequence in blood serum
   - eg. Fasting blood glucose $\geq 126$ mg/ml diagnostic of diabetes mellitus

2. **Ordered Categorical** with multiple implicit or explicit thresholds
   - eg. the BIRADS scale for mammograms: 1 ‘Benign’; 2 ‘Possibly benign’; 3 ‘Unclear’; 4 ‘Possibly malignant’; 5 ‘Malignant’
   - eg. Clinical symptoms classified as 1 ‘not present’, 2 ‘mild’, 3 ‘moderate’, or 4 ‘severe’

3. **Continuous**
   - eg. biochemical tests such as serum levels of creatinine, bilirubin or calcium
Figure: Basic Study Design

SERIES OF PATIENTS

INDEX TEST

REFERENCE TEST

CROSS-CLASSIFICATION
Diagnostic Accuracy Studies

Figure: Distributions of test result for diseased and non-diseased populations defined by threshold (DT)
**Binary Test Accuracy**

**Data Structure**

Data often reported as $2 \times 2$ matrix

<table>
<thead>
<tr>
<th></th>
<th>Reference Test (Diseased)</th>
<th>Reference Test (Healthy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Positive</td>
<td>True Positive (a)</td>
<td>False Positive (b)</td>
</tr>
<tr>
<td>Test Negative</td>
<td>False Negative (c)</td>
<td>True Negative (d)</td>
</tr>
</tbody>
</table>

1. The chosen threshold may vary between studies of the same test due to inter-laboratory or inter-observer variation

2. The higher the cut-off value, the higher the specificity and the lower the sensitivity
Binary Test Accuracy
Measures of Test Performance

Sensitivity (true positive rate)  The proportion of subjects with disease who are correctly identified as such by test \( \frac{a}{a+c} \)

Specificity (true negative rate)  The proportion of subjects without disease who are correctly identified as such by test \( \frac{d}{b+d} \)

Positive predictive value  The proportion of test positive subjects who truly have disease \( \frac{a}{a+b} \)

Negative predictive value  The proportion of test negative subjects who truly do not have disease \( \frac{d}{c+d} \)
**Binary Test Accuracy**

*Measures of Test Performance*

**Likelihood ratios (LR)** The ratio of the probability of a positive (or negative) test result in the patients with disease to the probability of the same test result in the patients without the disease

\[
LR = \frac{\text{sensitivity}}{1 - \text{specificity}} \quad \text{or} \quad \frac{1 - \text{Sensitivity}}{\text{specificity}}
\]

**Diagnostic odds ratio** The ratio of the odds of a positive test result in patients with disease compared to the odds of the same test result in patients without disease

\[
\text{OD} = \frac{\text{LR}_{+}}{\text{LR}_{-}}
\]
Diagnostic Meta-analysis
Methodological Concepts

1. Glass (1976)
Meta-analysis refers to the statistical analysis that combines the results of some collection of related studies to arrive at a single conclusion to the question at hand.

2. Meta-analysis may be based on aggregate patient data (APD meta-analysis) or individual patient data (IPD meta-analysis).
Diagnostic Meta-analysis
Methodological Concepts

1. Meta-analysis of diagnostic accuracy studies may be performed to provide summary estimates of test performance based on a collection of studies and their reported empirical or estimated smooth ROC curves.

2. Statistical methodology for meta-analysis of diagnostic accuracy studies focused on studies reporting estimates of test sensitivity and specificity or two by two data.

3. Both fixed and random-effects meta-analytic models have been developed to combine information from such studies.
Methods for Dichotomized Data

1. Meta-analysis of sensitivity and specificity separately by direct pooling or modeling using fixed-effects or random-effects approaches.

2. Meta-analysis of positive and negative likelihood ratios separately using fixed-effects or random-effects approaches as applied to risk ratios in meta-analysis of therapeutic trials.

3. Meta-analysis of diagnostic odds ratios using fixed-effects or random-effects approaches as applied to meta-analysis of odds ratios in clinical treatment trials.

4. Summary ROC Meta-analysis using fixed-effects or random-effects approaches.
Summary ROC Meta-analysis

The most commonly used and easy to implement method
It is a fixed-effects model

1. Linear regression analysis of the relationship
   \[ D = a + bS \] where:
   \[ D = (\logit \ TPR) - (\logit \ FPR) = \ln \ DOR \]
   \[ S = (\logit \ TPR) + (\logit \ FPR) = \text{proxy for the threshold} \]

2. \( a \) and \( b \) may be estimated by weighted or un-weighted least squares or robust regression, back-transformed and plotted in ROC space

3. Differences between tests or subgroups may be examined by adding co-variates to model
Hierarchical/multi-level Models

Mathematically equivalent models for estimating underlying SROC and average operating point and/or exploring heterogeneity

Bivariate Mixed Effects Models

1. Generalized linear mixed model

2. Focused on inferences about sensitivity and specificity but SROC curve(s) can be derived from the model parameters

Hierarchical Summary ROC (HSROC) Model

1. Generalized non-linear mixed model

2. Focused on inferences about the SROC curve, or comparing SROC curves but summary operating point(s) can be derived from the model parameters
Bivariate Mixed Model

Level 1: Within-study variability: Approximate Normal Approach

\[
\begin{pmatrix}
\logit(p_{Ai}) \\
\logit(p_{Bi})
\end{pmatrix} \sim N\left(\begin{pmatrix}
\mu_{Ai} \\
\mu_{Bi}
\end{pmatrix}, C_i\right)
\]

\[
C_i = \begin{pmatrix}
  s^2_{Ai} & 0 \\
  0 & s^2_{Bi}
\end{pmatrix}
\]

\(p_{Ai}\) and \(p_{Bi}\)  Sensitivity and specificity of the \(i\)th study

\(\mu_{Ai}\) and \(\mu_{Bi}\)  Logit-transforms of sensitivity and specificity of the \(i\)th study

\(C_i\)  Within-study variance matrix

\(s^2_{Ai}\) and \(s^2_{Bi}\)  variances of logit-transforms of sensitivity and specificity
Bivariate Mixed Model

Level 1: Within-study variability: Exact Binomial Approach

\[ y_{Ai} \sim Bin\left(n_{Ai}, p_{Ai}\right) \]

\[ y_{Bi} \sim Bin\left(n_{Bi}, p_{Bi}\right) \]

\( n_{Ai} \) and \( n_{Bi} \)  Number of diseased and non-diseased

\( y_{Ai} \) and \( y_{Bi} \)  Number of diseased and non-diseased with true test results

\( p_{Ai} \) and \( p_{Bi} \)  Sensitivity and specificity of the \( i \)th study
Bivariate Mixed Model

Level 2: Between-study variability

\[
\begin{pmatrix}
\mu_{Ai} \\
\mu_{Bi}
\end{pmatrix}
\sim N\left(\left(\begin{array}{c}
\mu_A \\
\mu_B
\end{array}\right), \Sigma_{AB}\right)
\]

\[
\Sigma_{AB} = \begin{pmatrix}
\sigma^2_A & \sigma_{AB} \\
\sigma_{AB} & \sigma^2_B
\end{pmatrix}
\]

\(\mu_{Ai}\) and \(\mu_{Bi}\)  Logit-transforms of sensitivity and specificity of the \(i\)th study

\(\mu_A\) and \(\mu_B\)  Means of the normally distributed logit-transforms

\(\Sigma_{AB}\)  Between-study variances and covariance matrix
Hierarchical Summary ROC Regression

Level 1: Within-study variability

\[ y_{ij} \sim Bin(n_{ij}, \pi_{ij}) \]

\[ \logit(\pi_{ij}) = (\theta_i + \alpha_i X_{ij}) \exp(-\beta X_{ij}) \]

\( \theta_i \) and \( \alpha_i \) Study-specific threshold and accuracy parameters

\( y_{ij} \) Number testing positive assumed to be binomially distributed

\( \pi_{ij} \) Probability that a patient in study \( i \) with disease status \( j \) has a positive test result

\( X_{ij} \) True disease status (coded -0.5 for those without disease and 0.5 for those with the disease)
Hierarchical Summary ROC Regression

Level 2: Between-study variability

\[
\theta_i \sim N(\Theta, \sigma^2_\theta)
\]
\[
\alpha_i \sim N(A, \sigma^2_\alpha)
\]

\(\Theta\) and \(A\) Means of the normally distributed threshold and accuracy parameters

\(\sigma^2_\theta\) and \(\sigma^2_\alpha\) Variances of mean threshold and accuracy

\(\beta\) Shape parameter which models any asymmetry in the SROC curve
Motivating Data

1. Scheidler and colleagues combined information from several studies to estimate and compare the ability of LAG, CT and MR to accurately detect lymph node metastasis.

2. They combined data from 36 studies, of which 17 examined LAG, 19 examined CT and 10 examined MR.

3. Nine of the 36 studies examined more than one test. In particular, two studies examined CT and LAG, four studies examined CT and MR, and two studies examined CT twice.

4. The two studies that examined CT twice reported data separately for para-aortic and pelvic nodes.

5. This dataset of 46 estimates of test sensitivity and specificity was reanalyzed by Rutter and Gatsonis using bayesian HSROC (BUGS) and by Macaskill using adaptive quadrature (proc nlmixed in SAS)
The NLMIXED procedure for nonlinear mixed models in SAS can fit the HSROC model.

NLMIXED allows for a nonlinear function of model parameters and non-normal error distributions, including the binomial distribution.

Random effects are restricted to be normally distributed.

The syntax closely follows the model specification.
HSROC Using NLMIXED

1. NLMIXED uses maximum likelihood estimation to fit the model.
2. NLMIXED provides empirical Bayes estimates of the random effects.
3. The marginal likelihood is maximized using adaptive Gaussian quadrature.
4. Starting values are estimated by first fitting the model in NLMIXED with no random effects.
1. The ESTIMATE facility in NLMIXED allows a function of the model parameters to be estimated.

2. The delta method is used to estimate the asymptotic standard error of the function of parameter estimates based on the covariance matrix of the parameter estimates.

3. This approach allows the summary estimates of sensitivity, specificity, and likelihood ratios and their asymptotic confidence intervals to be computed.
HSROC using PROC NLMIXED: MACASKILL’S CODE

data scheid;

input study test pos n dis;

T1=0; T2=0; /* create dummy variables for test type */

if test eq 1 then t1=1; /* using LAG as the referent test */

if test eq 2 then t2=1;

datalines;

1 0 19 29 0.5
1 0 1 82 0.5
46 2 16 18 0.5
46 2 2 24 0.5

;
HSROCs using PROC NLMIXED

proc nlmixed data=scheid;

parms theta=0 tc=0 tm=0 alpha=2 ac=0 am=0 beta=0 bc=0 bm=0 s2ut=1 s2ua=1; /* starting values */

logitp = (theta + ut + tc*t1 + tm*t2 + (alpha + ua + ac*t1 + am*t2)*dis)*exp(-(beta + bc*t1 + bm*t2)*dis);

p = exp(logitp)/(1+exp(logitp));

model pos ~ binomial(n,p);

random ut ua ~ normal([0, 0],[s2ut,0,s2ua]) subject=study;

run;
cap prog drop hsroclike

program define hsroclike

args todo b lnf g

tempvar Theta Alpha Beta lnsTheta lnsAlpha

mleval 'Theta' = 'b', eq(1)
mleval 'Alpha' = 'b', eq(2)
mleval 'Beta' = 'b', eq(3)
mleval 'lnsTheta' = 'b', eq(4) scalar
mleval 'lnsAlpha' = 'b', eq(5) scalar

tempname varTheta varAlpha

scalar 'varTheta'(exp('lnsTheta'*2))
scalar 'varAlpha'(exp('lnsAlpha'*2))
tempvar lnpi sum L last
gen double `lnpi'=0
gen double `sum'=0
gen double `L'=0
by study: gen byte `last'=_n==_N

tempname x1 x2
gen double `x1' = 0
gen double `x2' = 0

forvalues r=1/ ${draws} {replace `x1' = (((`Theta' + avar1'r'*sqrt(`varTheta')) + ///
0.5*(`Alpha' + avar2'r'*sqrt(`varAlpha')))/exp((`Beta'/2))
replace `x2' = (((`Theta' + avar1'r'*sqrt(`varTheta')) - ///
0.5*(`Alpha' + avar2'r'*sqrt(`varAlpha'))))exp((`Beta'/2))
replace `lnpi' = cond(dtruth==1, ///
(y*ln(invlogit(`x1'))) + ((1-y)*ln(invlogit(-`x1'))), ///
(y*ln(invlogit(-`x2'))) + ((1-y)*ln(invlogit( `x2'))))
by study: replace `sum' = sum(`lnpi')
by study: replace `L' = `L' + exp(`sum')*wvar'r' if `last'
}

mlsum `lnf' = ln(`L') if `last'

if (`todo'==0|`lnf'>.) exit
use "e:\rghsrocsmsle.dta", clear

gen y1=tp

gen y2=tn

gen num1=tp+fn

gen num2=tn+fp

gen study=_n

reshape long num y, i(study) j(dtruth)

gen _dfreq=1

_binomial2bernoulli y, fw(_dfreq) binomial(num)

expand _dfreq
STATA: Pseudo-random Monte Carlo

mata: ndraws=1000

mata: rseed(12345)

mata: hsrocdraws=rnormal(2,ndraws,0,1)

mata: hsrocdraws=hsrocdraws\J(1,cols(hsrocdraws), 1/cols(hsrocdraws))

mata: st_matrix("r(hsrocdraws)",hsrocdraws)

matrix hsrocdraw=r(hsrocdraws)

global draws= colsof(hsrocdraw)
mata: burn=100

mata: ndraws=1000

mata: hsrocdraws =halton(ndraws,2,(1+burn+ndraws),.)'

mata: hsrocdraws =hsrocdraws\J(1,cols(hsrocdraws), 1/cols(hsrocdraws))

mata: st_matrix("r(hsrocdraws)",hsrocdraws)

matrix hsrocdraw=r(hsrocdraws)

global draws= colsof(hsrocdraw)
mata: ndraws=35

mata: hsrocdraws=_gauss_hermite_nodes(ndraws)

mata: hsrocdraws =hsrocdraws\J(1,cols(hsrocdraws), 1/cols(hsrocdraws))

mata: st_matrix("r(hsrocdraws)",hsrocdraws)

matrix hsrocdraw=r(hsrocdraws)

global draws= colsof(hsrocdraw)
mata: ndraws=25
mata: hsrocdraws=nwspgr("KPN", 2, ndraws)
mata: hsrocdraws = hsrocdraws\J(1,cols(hsrocdraws), 1/cols(hsrocdraws))
mata: hsrocdraws=hsrocdraws’
mata: st_matrix("r(hsrocdraws)",hsrocdraws)

matrix hsrocdraw=r(hsrocdraws)

global draws= colsof(hsrocdraw)
forvalues r = 1/$draws {
  bysort study: gen avar1‘r’=hsrocdraw[1,’r’]
  bysort study: gen avar2‘r’=hsrocdraw[2,’r’]
  bysort study: gen wvar‘r’=hsrocdraw[3,’r’]
}

ml model d1 hsroclike (Theta:i.test) ///
  (Alpha:i.test)(Beta:i.test) /lnsTheta /lnsAlpha, technique(nr) ///
nopreserve group(study) maximize search(on) skip ///
difficult tol(1e-2) ltol(1e-2) nooutput

ml display, noheader cformat(%7.2f) pformat(%4.3f) sformat(%4.3f) ///
diparm(lnsTheta, function(exp(@)) deriv(exp(@)) prob label("sdTheta")) ///
diparm(lnsAlpha, function(exp(@)) deriv(exp(@)) prob label("sdAlpha"))
nois nlcom (sen_lag: invlogit(_b[Theta:_cons] + _b[Alpha:_cons]*0.5)*exp(-_b[Beta:_cons])*0.5) (spe_lag: 1 - invlogit(_b[Theta:_cons] - _b[Alpha:_cons]*0.5)*exp(_b[Beta:_cons])*0.5) (sen_ct: invlogit((_b[Theta:_cons] + _b[Theta:2.test]) + (_b[Alpha:_cons] + _b[Alpha:2.test])*0.5)*exp(-(b[Beta:_cons] + _b[Beta:2.test])*0.5)) (spe_ct: 1 - invlogit((_b[Theta:_cons] + _b[Theta:2.test]) - (_b[Alpha:_cons] + _b[Alpha:2.test])*0.5)*exp((b[Beta:_cons] + _b[Beta:2.test])*0.5)) (sen_mr: invlogit((_b[Theta:_cons] + _b[Theta:3.test]) + (_b[Alpha:_cons] + _b[Alpha:3.test])*0.5)*exp(-(b[Beta:_cons] + _b[Beta:3.test])*0.5)) (spe_mr: 1 - invlogit((_b[Theta:_cons] + _b[Theta:3.test]) - (_b[Alpha:_cons] + _b[Alpha:3.test])*0.5)*exp((b[Beta:_cons] + _b[Beta:3.test])*0.5))

noheader cformat(%7.2f) pformat(%4.3f) sformat(%4.3f)
### STATA: Summary Test Performance

#### Pseudo-random Monte Carlo

|       | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-------|-------|-----------|-------|-----|---------------------|
| sen_lag | 0.68  | 0.04      | 17.986| 0.000| 0.60 - 0.75         |
| spe_lag | 0.84  | 0.03      | 28.349| 0.000| 0.78 - 0.90         |
| sen_ct  | 0.48  | 0.07      | 6.694 | 0.000| 0.34 - 0.63         |
| spe_ct  | 0.93  | 0.01      | 67.486| 0.000| 0.90 - 0.96         |
| sen_mr  | 0.54  | 0.09      | 5.690 | 0.000| 0.35 - 0.72         |
| spe_mr  | 0.95  | 0.01      | 72.557| 0.000| 0.93 - 0.98         |

#### Quasi-random Monte Carlo

|       | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-------|-------|-----------|-------|-----|---------------------|
| sen_lag | 0.69  | 0.04      | 18.170| 0.000| 0.61 - 0.76         |
| spe_lag | 0.85  | 0.03      | 28.912| 0.000| 0.79 - 0.90         |
| sen_ct  | 0.49  | 0.07      | 7.314 | 0.000| 0.36 - 0.63         |
| spe_ct  | 0.93  | 0.01      | 67.886| 0.000| 0.90 - 0.96         |
| sen_mr  | 0.55  | 0.09      | 6.403 | 0.000| 0.38 - 0.72         |
| spe_mr  | 0.95  | 0.01      | 70.574| 0.000| 0.92 - 0.98         |

#### Sparse Grids Quadrature

|       | Coef. | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-------|-------|-----------|-------|-----|---------------------|
| sen_lag | 0.68  | 0.04      | 18.711| 0.000| 0.61 - 0.75         |
| spe_lag | 0.86  | 0.02      | 35.973| 0.000| 0.82 - 0.91         |
| sen_ct  | 0.53  | 0.07      | 7.858 | 0.000| 0.40 - 0.66         |
| spe_ct  | 0.93  | 0.01      | 70.201| 0.000| 0.90 - 0.95         |
| sen_mr  | 0.54  | 0.10      | 5.564 | 0.000| 0.35 - 0.73         |
| spe_mr  | 0.95  | 0.01      | 74.814| 0.000| 0.93 - 0.98         |
1. The HSROC model as discussed previously is defined by separate equations for within-study (Level I) and between-study (Level II) variation.

2. The Bayesian formulation requires an additional third level specifying priors for model parameters.

3. The priors for accuracy, threshold and shape parameters were chosen to reflect all plausible ranges.
1. Rutter and Gatsonis used BUGS, a publicly available software for Markov Chain Monte Carlo sampling.

2. BUGS uses derivative-free adaptive rejection sampling to draw from log-concave distributions and the Griddy-Gibbs method to estimate draws from non-log-concave distributions.

3. WinBUGS, a windows version of BUGS, is also publicly available and more user-friendly (has GUI).
MCMC USING BUGS (RUTTER-GATSONIS): DATA PREPARATION

model dxmeta;

const

N = 46;

var

CT[N], MR[N], fp[N], neg[N], tp[N], pos[N],
theta[N], alpha[N], pi[2,N], t[N], a[N], b[N],
THETA, LAMBDA, beta, gamma[2], lambda[2], bcov[2],
prec[2,3], sigmasq[2,3];

data CT, MR, tp, pos, fp, neg in "dxmeta.dat";
inits in "dxmeta.ini";
Bayesian Hierarchical SROC Analysis

MCMC USING BUGS (RUTTER-GATSONIS): PRIORS

{
  THETA~dunif(-10,10);
  LAMBDA~dunif(-2,20);
  beta~dunif(-5,5);
  for(i in 1:2) {
    gamma[i]~dunif(-10,10);
    lambda[i]~dunif(-10,10);
    bcov[i]~dunif(-5,5);
    for(j in 1:3) {
      prec[i,j] ~ dgamma(2.1,2); sigmasq[i,j] = 1.0/prec[i,j];
    }
  }
}
for(i in 1:N){
    t[i] <- THETA+CT[i]*gamma[1]+MR[i]*gamma[2];
    l[i] <- LAMBDA+CT[i]*lambda[1]+MR[i]*lambda[2];
    theta[i]~dnorm(t[i],prec[1,test[i]]);
    alpha[i]~dnorm(l[i],prec[2,test[i]]);
    b[i] <- exp((beta+CT[i]*bcov[1]+MR[i]*bcov[2])/2);
    logit(pi[1,i]) <- (theta[i] + 0.5*alpha[i])/b[i];
    logit(pi[2,i]) <- (theta[i] - 0.5*alpha[i])*b[i];
    tp[i] ~ dbin(pi[1,i],pos[i]);
    fp[i] ~ dbin(pi[2,i],neg[i]);
}
BAYESIAN ESTIMATION IN STATA: bayesmh

1. Fits a variety of Bayesian models using an adaptive Metropolis-Hastings (MH) algorithm

2. Provides various likelihood models including univariate normal linear and nonlinear regressions, multivariate normal linear and nonlinear regressions, generalized linear models such as logit and Poisson regressions, and multiple-equations linear models

3. Provides various prior distributions including continuous distributions such as uniform, Jeffreys, normal, gamma, multivariate normal, and Wishart and discrete distributions such as Bernoulli and Poisson

4. For a not-supported or nonstandard likelihood, you can use the llf() option within likelihood() to specify a generic expression for the observation-level likelihood function
The `bayesmh` command for Bayesian analysis includes three functional components:

1. Setting up a posterior model which includes a likelihood model that specifies the conditional distribution of the data given model parameters and prior distributions for all model parameters. The prior distribution of a parameter can itself be specified conditional on other parameters, also referred to as hyperparameters.

2. Performing MCMC simulation

3. Summarizing and reporting results
use "i:\multitest.dta", clear

gen y0 = fp

gen y1 = tp

gen num0 = tn+fp

gen num1 = tp+fn

gen study = _n

reshape long num y, i(study) j(dtruth)

replace dtruth=-0.5 if dtruth ==0

replace dtruth=0.5 if dtruth ==1

fvset base none study testcat
bayesmh y, likelihood(dbinomial(invlogit((({theta:}+{xbtheta:i.testcat, noconstant})+ ///
({alpha:}+{xalpha:i.testcat, noconstant})*dtruth)*exp(-({beta} + ///
{xbbeta:i.testcat, noconstant})*dtruth)), num)) ///

redefine(theta:i.study) ///
redefine(alpha:i.study) ///

prior({theta:i.study}, normal({mutheta}, {vartheta})) ///
prior({alpha:i.study}, normal({mualpha}, {varalpha})) ///
prior({mutheta}, uniform(-10,10)) prior({xbbeta:}, uniform(-5,5)) ///
prior({mualpha}, uniform(-2,20)) prior({beta}, uniform(-5,5)) ///
prior({xbtheta:} {xbeta:}, uniform(-10,10)) ///
prior({vartheta} varalpha), igamma(2.1,2.0)) ///

block({vartheta} {varalpha} {mutheta} {mualpha}, split) ///
block({xbtheta:} {xbalpha:}{xbbeta:}, split) ///

noshow({theta:i.study} {alpha:i.study}) ///
nomodelsummary rseed(13456677) burnin(50000) thin(2) dots(1000) ///
mcmcsize(50000) saving("i:\hsroctests", replace)

estimates store hsroctests
## BAYESMH: ESTIMATES

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>xbalphat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.96529</td>
<td>1.554068</td>
<td>.446378</td>
<td>2.21716</td>
<td>-1.587397 4.282242</td>
</tr>
<tr>
<td>2</td>
<td>3.066355</td>
<td>1.764614</td>
<td>.475079</td>
<td>3.332689</td>
<td>-.6274569 5.87028</td>
</tr>
<tr>
<td>3</td>
<td>3.811619</td>
<td>2.095353</td>
<td>.554116</td>
<td>3.802728</td>
<td>-.5743477 7.647241</td>
</tr>
<tr>
<td>xbbetat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.78733</td>
<td>1.866076</td>
<td>.53337</td>
<td>1.428655</td>
<td>-1.351259 4.838397</td>
</tr>
<tr>
<td>2</td>
<td>.7260084</td>
<td>1.893</td>
<td>.544684</td>
<td>.3239997</td>
<td>-2.388184 3.94731</td>
</tr>
<tr>
<td>3</td>
<td>.7577294</td>
<td>1.854283</td>
<td>.530049</td>
<td>.3592742</td>
<td>-2.372203 4.052735</td>
</tr>
<tr>
<td>xbthetat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-1.868499</td>
<td>1.122751</td>
<td>.325802</td>
<td>-1.567034</td>
<td>-4.311378 -.234388</td>
</tr>
<tr>
<td>2</td>
<td>-3.442445</td>
<td>1.114605</td>
<td>.309764</td>
<td>-3.218207</td>
<td>-5.976668 -1.688818</td>
</tr>
<tr>
<td>3</td>
<td>-3.640552</td>
<td>1.092858</td>
<td>.286307</td>
<td>-3.438157</td>
<td>-6.191774 -1.942339</td>
</tr>
<tr>
<td>beta</td>
<td>-1.233781</td>
<td>1.84388</td>
<td>.535988</td>
<td>-.7840052</td>
<td>-4.289635 1.750596</td>
</tr>
<tr>
<td>mutheta</td>
<td>1.728991</td>
<td>1.096002</td>
<td>.319087</td>
<td>1.444648</td>
<td>.2239442 4.21629</td>
</tr>
<tr>
<td>varthetata</td>
<td>.5992724</td>
<td>.1848735</td>
<td>.007505</td>
<td>.5735596</td>
<td>.3153643 1.037242</td>
</tr>
<tr>
<td>mualphata</td>
<td>.330738</td>
<td>1.566817</td>
<td>.450673</td>
<td>.0471519</td>
<td>-1.862259 3.893121</td>
</tr>
<tr>
<td>varalphata</td>
<td>.7970011</td>
<td>.2870122</td>
<td>.013143</td>
<td>.7565332</td>
<td>.3772263 1.486745</td>
</tr>
</tbody>
</table>
Bayesian Hierarchical SROC Analysis

\textbf{bayesmh: Summary Test Performance}

\begin{verbatim}
bayesstats summary ///
(sen_lag:invlogit(((xbtheta:1bn.testcat)+ mutheta)) + ///
({mualpha} + {xbalpha:1bn.testcat}*0.5)*exp(-({beta} + ///
{xbbeta:1bn.testcat})*0.5))) ///

(spe_lag:1-invlogit(((xbtheta:1bn.testcat)+ mutheta)) - ///
({mualpha} + {xbalpha:1bn.testcat}*0.5)*exp(({beta} + ///
{xbbeta:1bn.testcat})*0.5))) ///

(sen_ct:invlogit(((xbtheta:2.testcat)+ mutheta))) + ///
({mualpha} + ({xbalpha:2.testcat})*0.5)*exp(-({beta} + ///
{xbbeta:2.testcat})*0.5))) ///

(spe_ct:1-invlogit(((xbtheta:2.testcat)+ mutheta)) - ///
({mualpha} + ({xbalpha:2.testcat})*0.5)*exp(({beta} + ///
{xbbeta:2.testcat})*0.5))) ///

(sen_mr:invlogit(((xbtheta:3.testcat)+ mutheta)) + ///
({mualpha} + ({xbalpha:3.testcat})*0.5)*exp(-({beta} + ///
{xbbeta:3.testcat})*0.5))) ///

(spe_mr:1-invlogit(((xbtheta:3.testcat)+ mutheta)) - ///
({mualpha} + ({xbalpha:3.testcat})*0.5)*exp(({beta} + ///
{xbbeta:3.testcat})*0.5))), noleg hpd
\end{verbatim}
<table>
<thead>
<tr>
<th>HPD</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>MCSE</th>
<th>Median</th>
<th>[95% Cred. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>sen_lag</td>
<td>.6785673</td>
<td>.0467938</td>
<td>.002559</td>
<td>.6796891</td>
<td>.5830788 .765461</td>
</tr>
<tr>
<td>spe_lag</td>
<td>.8382703</td>
<td>.0431682</td>
<td>.002212</td>
<td>.8414481</td>
<td>.7508609 .9183696</td>
</tr>
<tr>
<td>sen_ct</td>
<td>.494813</td>
<td>.0766115</td>
<td>.002932</td>
<td>.4949665</td>
<td>.3457088 .6456879</td>
</tr>
<tr>
<td>spe_ct</td>
<td>.9291934</td>
<td>.015615</td>
<td>.000653</td>
<td>.9305332</td>
<td>.8982031 .9581979</td>
</tr>
<tr>
<td>sen_mr</td>
<td>.5458161</td>
<td>.1009772</td>
<td>.004455</td>
<td>.549018</td>
<td>.3440935 .7370919</td>
</tr>
<tr>
<td>spe_mr</td>
<td>.9524953</td>
<td>.0146328</td>
<td>.000547</td>
<td>.9543193</td>
<td>.9224045 .9777359</td>
</tr>
</tbody>
</table>
## COMPARATIVE SUMMARY TEST PERFORMANCE

<table>
<thead>
<tr>
<th></th>
<th>bayesmh</th>
<th>winBUGS</th>
<th>NLMIXED</th>
<th>ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>sen_lag</td>
<td>0.68(0.58-0.76)</td>
<td>0.68(0.58-0.77)</td>
<td>0.68(0.60-0.76)</td>
<td>0.69(0.61-0.76)</td>
</tr>
<tr>
<td>spe_lag</td>
<td>0.84(0.75-0.92)</td>
<td>0.84(0.74-0.91)</td>
<td>0.84(0.74-0.90)</td>
<td>0.85(0.79-0.90)</td>
</tr>
<tr>
<td>sen_ct</td>
<td>0.49(0.35-0.65)</td>
<td>0.48(0.31-0.66)</td>
<td>0.49(0.35-0.63)</td>
<td>0.49(0.36-0.63)</td>
</tr>
<tr>
<td>spe_ct</td>
<td>0.93(0.90-0.96)</td>
<td>0.93(0.89-0.96)</td>
<td>0.93(0.90-0.95)</td>
<td>0.93(0.90-0.96)</td>
</tr>
<tr>
<td>sen_mr</td>
<td>0.55(0.34-0.74)</td>
<td>0.54(0.29-0.77)</td>
<td>0.55(0.37-0.71)</td>
<td>0.55(0.38-0.72)</td>
</tr>
<tr>
<td>spe_mr</td>
<td>0.95(0.92-0.98)</td>
<td>0.95(0.91-0.98)</td>
<td>0.95(0.92-0.97)</td>
<td>0.95(0.92-0.98)</td>
</tr>
</tbody>
</table>
Summary

1. Recent availability of `bayesmh` and the myriad of post-estimation commands allows comprehensive bayesian hierarchical summary ROC analysis in Stata.

2. Although there is no Stata-native generalized non-linear mixed modeling command, frequentist hierarchical summary ROC analysis is possible by means of `ml` programming.

3. Frequentist estimation approximates likelihood by either quadrature or simulation-based numerical integration techniques.

4. The results obtained using Stata are comparable with those obtained with other software in both frequentist and bayesian frameworks.
Aertgeerts B., Buntinx F., and Kester A.
The value of the CAGE in screening for alcohol abuse and alcohol dependence in general clinical populations: a diagnostic meta-analysis.
J clin Epidemiol 2004;57:30-39

Bivariate Random Effects Meta-Analysis of ROC Curves.
Med Decis Making 2008;28:621-628

Begg C.B. and Mazumdar M.
Operating characteristics of a rank correlation test for publication bias.
Biometrics 1994;50:1088-1101

Chu H. and Cole S.R.
Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach.
J Clin Epidemiol 2006;59:1331-1332

Dendukuri N., Chui K. and Brophy J.M.
Validity of EBCT for coronary artery disease: a systematic review and meta-analysis.
BMC Medicine 2007;5:35
Dukic V. and Gatsonis C.
Meta-analysis of diagnostic test accuracy studies with varying number of thresholds.
Biometrics 2003;59:936-946

Dwamena, B.
midas: Module for Meta-Analytical Integration of Diagnostic Accuracy Studies
Boston College Department of Economics, Statistical Software Components 2007;
s456880: http://ideas.repec.org/c/boc/bocode/s456880.html.

Ewing J.A.
Detecting Alcoholism: The CAGE questionnaire.
JAMA 1984;252:1905-1907

Unification of models for meta-analysis of diagnostic accuracy studies.
Biostatistics 2007;8:239-251

An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary
Journal of Clinical Epidemiology 2008;61;1095-1103
Harbord R.M., and Whitting P.
metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression
Stata Journal 2009;2:211-229

Irwig L., Macaskill P., Glasziou P. and Fahey M.
Meta-analytic methods for diagnostic test accuracy.

Kester A.D.M., and Buntinx F.
Meta-Analysis of ROC Curves.

Littenberg B. and Moses L. E.

Macaskill P.
Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis.
J Clin Epidemiol 2004;57:925-932
References IV

Moses L.E., Shapiro D. and Littenberg B.
Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations.

Pepe M.S.
Receiver Operating Characteristic Methodology.

Pepe M.S.
The Statistical Evaluation of Medical Tests for Classification and Prediction.
2003; Oxford: Oxford University Press

Reitsma J.B., Glas A.S., Rutjes A.W.S., Scholten R.J.P.M., Bossuyt P.M. and Zwinderman A.H.
Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews.
J Clin Epidemiol 2005;58:982-990

Rutter C.M., and Gatsonis C.A.
A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations
Stat Med 2001;20:2865-2884
Toledano A. and Gatsonis C.A.
Regression analysis of correlated receiver operating characteristic data.
Academic Radiology 1995;2:S30-S36

Tosteson A.A. and Begg C.B.
A general regression methodology for ROC curve estimation.
Medical Decision Making 1988;8:204-215

White I.R.
Multivariate Random-effects Meta-analysis.
Stata Journal 2009;1:40-56