Multivariate Mixed-Effects Meta-Analysis of Paired-Comparison Studies of Diagnostic Test Accuracy

Ben A. Dwamena, MD

The University of Michigan & VA Medical Centers, Ann Arbor

SNASUG - July 24, 2008

University of Michigan
Medical School

Diagnostic Test Evaluation

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

This includes:

1. Elements of medical history
2. Physical examination
3. Imaging and laboratory investigations
4. Clinical prediction rules

University of Michigan Medical School

Diagnostic Test Evaluation

1. The performance of a diagnostic test assessed by comparison of index and reference test results on a group of subjects
2. Ideally these should be patients suspected of the target condition that the test is designed to detect.
3. Most diagnostic tests have multiple or continuous outcomes
4. Categorization or application of a cutoff value is frequently applied to classify results into positive or negative
5. Such dichotomization is then represented in one 2×2 contingency table

University of Michigan
Medical School

Diagnostic Test Evaluation

Positivity threshold for dichotomization may be:

1. **Implicit**: based on interpretation/judgement/machine calibration
   e.g. radiologists classifying images as normal or abnormal
2. **Explicit**: based on a numerical threshold
   e.g. blood glucose level above patient has diabetes

Table: 2X2 Contingency Table

	Reference Test Positive	Reference Test Negative
Test Positive	a=TP	b=FP
Test Negative	c=FN	d=TN

TP: True Positive;**TN**: True Negative;
FP: False Positive;**FN**: False Negative

University of Michigan Medical School

Measures of Diagnostic Performance

Sensitivity (true positive rate)	The proportion of people with disease who are correctly identified as such by test
Specificity (true negative rate)	The proportion of people without disease who are correctly identified as such by test
Positive predictive value	The proportion of test positive people who truly have disease
Negative predictive value	The proportion of test negative people who truly do not have disease

University of Michigan
Medical School

Measures of Diagnostic 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
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.
ROC Curve	Plot of all pairs of (1-specificity, sensitivity) as positivity threshold varies

University of Michigan
Medical School

Meta-analysis of Diagnostic Performance

Utility

1. Evaluation of the quality and scope of available primary studies
2. Determination of the proper and efficacious use of diagnostic and screening tests in the clinical setting in order to guide patient treatment
3. Decision making about health care policy and financing
4. Identification of areas for further research, development, and evaluation

University of Michigan
Medical School

Meta-analysis of Diagnostic Performance

Major steps

1. Framing objectives of the review
2. Identifying the relevant literature
3. Assessment of methodological quality and applicability to the clinical problem at hand
4. Summarizing the evidence qualitatively and if appropriate, quantitatively(meta-analysis)
5. Interpretation of findings and development of recommendations

University of Michigan
Medical School

Meta-analytic Comparison of Diagnostic Tests

1. Most meta-analyses of are of single index test
2. Test comparisons are less common but of increasing importance
3. Indirect comparisons (using all studies regardless of whether included one, some or all tests under evaluation) are prone to confounding
4. Use of studies that directly comparing tests by paired design/in the same patients or by randomization are preferred
5. May be performed by adding test type as covariates in one model or applying separate models to different tests
6. Statistical models vary with summary ROC methods providing the most general approach

University of Michigan
Medical School

Summary ROC Regression

1. The true positive and false positive rates are transformed through the logarithms of their odds
2. Metaregression analysis examines the linear relationship D = **a** + **b**S : D = (logit TPR) - (logit FPR) = ln DOR; S = (logit TPR) + (logit FPR) = proxy for the threshold
3. The estimates for **a** and **b** are back-transformed and plotted in ROC space index test(s)
4. **a** and **b** may be estimated by weighted or unweighted least squares or robust regression
5. Differences between tests or subgroups may examined by adding covariates to model

Moses, Shapiro and Littenberg. Med Decis Making (1993)12:1293-1316

University of Michigan Medical School

Summary ROC Regression

The most commonly used and easy to implement method but:

1. Assumes variability in test performance due only to threshold effect and within-study variability
2. Does not provide average estimates of sensitivity and specificity
3. Continuity correction may introduce non-negligible downward bias to the estimated SROC curve
4. Does not account for measurement error in S
5. Ignores potential correlation between D and S.
6. Confidence intervals and p-values are likely to be inaccurate

University of Michigan
Medical School

Mixed Effects Hierarchical Models

Two mathematically equivalent, flexible models for identifying underlying SROC, estimating average operating point and/or exploring heterogeneity

Hierarchical Summary ROC(HSROC) Model

1. 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 Effects Models

1. Focused on inferences about sensitivity and specificity but SROC curve(s) can be derived from the model parameters
2. Generalization of the commonly used DerSimonian and Laird random effects model

Arends et al. Med Decis Making. Published online June 30, 2008

University of Michigan Medical School

Hierarchical Summary ROC Regression

$$y_{ij} \sim Bin\left(n_{ij}, \pi_{ij}\right)$$

$$\texttt{logit}\left(\pi_{ij}\right) = \left(\theta_i + \alpha_i X_{ij}\right) \exp\left(-\beta X_{ij}\right)$$

$$\theta_i \sim N\left(\Theta, \sigma_\theta^2\right)$$

$$\alpha_i \sim N\left(\mathrm{A}, \sigma_\alpha^2\right)$$

θ_i and α_i	Study-specific threshold and accuracy parameters
y_{ij}	Number testing positive assumed to be binomially distributed
π_{ij}	Probability that a patient in study i with disease status j has a positive test result
Θ and A	Means of the normally distributed threshold and accuracy parameters
σ_θ^2 and σ_α^2	Variances of mean threshold and accuracy
X_{ij}	True disease status(coded -0.5 for those without disease and 0.5 for those with the disease)
β	Shape parameter which models any asymmetry in the SROC curve

Rutter CM, Gatsonis CA. Stat Med (2001)20:2865-2884 Macaskill et al. J. Clin Epidemiol (2004) 57:925-932

University of Michigan Medical School

Bivariate Linear Mixed Model

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim N\left(\begin{pmatrix} M_A \\ M_B \end{pmatrix}, \Sigma_{AB} + C_i\right)$$

$$\Sigma_{AB} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}$$

$$C_i = \begin{pmatrix} s_{Ai}^2 & 0 \\ 0 & s_{Bi}^2 \end{pmatrix}$$

μ_{Ai} and μ_{Bi} Logit-transforms of sensitivity and specificity of the ith study

M_A and M_B Means of the normally distributed μ_{Ai} and μ_{Bi}

Σ_{AB} Between-study variances and covariance matrix of logit transforms

C_i Matrix of within-study(sampling) variances for μ_{Ai} and μ_{Bi}

s_{Ai} and s_{Bi} Within-study variances of μ_{Ai} and μ_{Bi}

Reitsma JB et al. J. Clin Epidemiol (2005) 58:982-990

University of Michigan Medical School

Bivariate Binomial Mixed Model

$$y_{Ai} \sim Bin(n_{Ai}, p_{Ai})$$

$$y_{Bi} \sim Bin(n_{Bi}, p_{Bi})$$

$$\begin{pmatrix} \mu_{Ai} \\ \mu_{Bi} \end{pmatrix} \sim N\left(\begin{pmatrix} M_A \\ M_B \end{pmatrix}, \Sigma_{AB}\right)$$

$$\Sigma_{AB} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}$$

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 ith study
μ_{Ai} and μ_{Bi}	Logit-transforms of sensitivity and specificity of the ith study
M_A and M_B	Means of the normally distributed logit-transforms
Σ_{AB}	Between-study variances and covariance matrix

Chu H, Cole SR (2006) J. Clin Epidemiol 59:1331-1332

University of Michigan Medical School

Bivariate Mixed Models

1. Easier to fit with standard mixed model software than the HSROC Model
2. Exact binomial approach preferred for sparse data and for avoiding continuity correction
3. The relation between logit-transformed sensitivity and specificity is given by: $\mu_{Ai} = \text{a} + \text{b*}\mu_{Bi}$
4. The slope b of this line equals σ_{AB}/σ_A^2, and the intercept a equals M_A - b*M_B
5. SROC may be obtained after anti-logit transformation of the regression line

University of Michigan
Medical School

Methodological Framework

1. Description of extensions of the bivariate binomial mixed model to paired design studies
2. Estimation of models with **xtmelogit**(by adaptive quadrature method(nip=7) and laplacian approximation (nip=1)
3. Assessment and comparison of fit, complexity and test performance estimates
4. Comparison of estimates and convergence times of best fit and least complex with those from the **gllamm** command

University of Michigan
Medical School

Motivating Example: PET vs CT In Lung Cancer

1. PET (uses radioisotope to evaluate tumor metabolism)
2. CT (uses xrays to evaluate tumor anatomy)
3. Both may be used to diagnose and/or examine the extent of lung cancer
4. The accuracy of these two radiological tests have been compared directly or indirectly by many researchers
5. We amalgamated data from 4 published meta-analyses to obtain 29 paired-comparison studies

- **Dwamena et al. Radiology(1999)213(2):530-6**
- **Gould et al. Ann Intern Med(2003)139(11):879-92**
- **Birim et al. Ann Thorac Surg(2005)79(1):375-82**
- **Alongi et al. Tumori(2006) 92(4):327-33**

University of Michigan
Medical School

Example Dataset: PET vs CT In Lung Cancer

Author	Year	TPpet	FPpet	FNpet	TNpet	TPct	FPct	FNct	TNct
Scott	1994	2	3	1	19	1	2	2	20
Wahl	1994	9	3	2	13	7	9	4	7
Chin	1995	7	4	2	17	5	3	4	18
Valk	1995	20	3	4	49	15	14	9	38
Scott	1996	9	0	0	18	6	3	3	15
Sazon	1996	16	0	0	16	13	7	3	9
Sasaki	1996	13	1	4	53	11	7	6	47
Vansteenskiste	1997	10	1	5	34	10	13	5	22
Steinert	1997	25	1	3	83	16	5	12	79
Guhlmann	1997	13	0	2	17	8	3	7	14
Hagberg	1997	6	0	3	9	5	0	4	9
Bury	1997	12	0	2	52	11	8	3	44
Saunders	1999	12	2	5	65	3	7	12	62
Marom	1999	40	3	4	31	26	4	18	30
Pieterman	2000	29	10	3	60	24	24	8	46
Gupta	2000	51	8	2	107	36	36	17	79
Poncelet	2001	6	8	3	44	5	17	4	36
Vansteenskiste	1998	26	6	2	28	21	5	7	26
Albes	1999	14	2	2	9	15	3	1	8
Richter	1999	9	1	0	12	5	1	4	12
Kubota	2000	3	0	3	12	4	4	2	8
Weng	2000	11	2	4	33	8	9	3	30
Luketich	2001	4	7	2	27	3	10	3	24
Kiernan	2002	22	9	3	54	16	4	9	63
VonHaag	2002	4	4	2	42	3	16	3	30
Antoch	2003	8	2	1	16	7	7	3	10
Halter	2004	72	3	10	31	63	8	9	26

Multivariate Mixed Models

1. Bivariate binomial mixed model with test type as fixed-effect covariate
2. Bivariate binomial mixed model with test type as random-effect covariate
3. Independent test-specific bivariate binomial mixed models
4. Combined test-specific bivariate binomial mixed models

University of Michigan
Medical School

Model 1

- Bivariate With Fixed-effect Test Type
- Estimation by xtmelogit using Adaptive Quadrature(nip=7)

```
xi: xtmelogit(ttruth lgtse lgtsp i.testtype, ///
  noc)(study: lgtse lgtsp, noc cov(uns)), bin(num)
```

ttruth	True(positive or negative) test results
lgtse and lgtsp	Logit-transforms of sensitivity and specificity
i.testtype	Dichotomous variable(PET=0 and CT=1)

University of Michigan
Medical School

Model 2

- Bivariate With Fixed-effect Test Type
- Estimation by xtmelogit using Laplacian approximation(nip=1)

```
xi: xtmelogit(ttruth lgtse lgtsp i.testtype, ///
noc) (study: lgtse lgtsp, noc cov(uns)), ///
bin(num) laplace
```

ttruth True(positive or negative) test results

lgtse and lgtsp Logit-transforms of sensitivity and specificity

i.testtype Dichotomous variable(PET=0 and CT=1)

University of Michigan
Medical School

Model 3

- Bivariate With Random Effect Test Type
- Estimation by xtmelogit using Adaptive Quadrature(nip=7)

```
xi: xtmelogit(ttruth lgtse lgtsp  i.testtype,///
noc)(study: lgtse lgtsp , noc cov(uns)) ///
(study: i.testtype), bin(num)
```

ttruth True(positive or negative) test results

lgtse and lgtsp Logit-transforms of sensitivity and specificity

i.testtype Dichotomous variable(PET=0 and CT=1)

University of Michigan
Medical School

Model 4

- Bivariate With Random Effect Test Type
- Estimation by xtmelogit using Laplacian approximation(nip=1)

```
xi: xtmelogit(ttruth lgtse lgtsp ///
i.testtype, noc)(study: lgtse lgtsp ,///
noc cov(uns))(study: i.testtype), laplace bin(num)
```

ttruth True(positive or negative) test results

lgtse and lgtsp Logit-transforms of sensitivity and specificity

i.testtype Dichotomous variable(PET=0 and CT=1)

University of Michigan
Medical School

Model 5

- Independent Test-specific Bivariate Models
- Estimation by xtmelogit using Adaptive Quadrature(nip=7)

```
xtmelogit(ttruth lgtsePET lgtspPET lgtseCT ///
lgtspCT, noc)(study: lgtsePET lgtspPET , ///
noc cov(uns))(study: lgtseCT lgtspCT, ///
noc cov(uns)), bin(num)
```

University of Michigan
Medical School

Model 6

- Independent Test-specific Bivariate Models
- Estimation by xtmelogit using Laplacian approximation(nip=1)

```
xtmelogit(ttruth lgtsePET lgtspPET lgtseCT ///
lgtspCT, noc)(study: lgtsePET lgtspPET , ///
noc cov(uns))(study: lgtseCT lgtspCT, ///
noc cov(uns)), laplace bin(num)
```

University of Michigan
Medical School

Model 7

- Combined Test-specific Bivariate Models
- Estimation by xtmelogit using Adaptive quadrature(nip=7)
- Accounts for all between-study and between-test variability

```
xtmelogit(ttruth lgtsePET lgtspPET lgtseCT ///
lgtspCT, noc)(study: lgtsePET lgtspPET ///
lgtseCT lgtspCT, noc cov(uns)), bin(num)
```

University of Michigan
Medical School

Model 8

- Combined Test-specific Bivariate Models
- Estimation by xtmelogit using Laplacian approximation(nip=1)
- Accounts for all between-study and between-test variability

```
xtmelogit(ttruth lgtsePET lgtspPET lgtseCT ///
lgtspCT, noc)(study: lgtsePET lgtspPET ///
lgtseCT lgtspCT, noc cov(uns)), laplace bin(num)
```

University of Michigan
Medical School

Model 9

- Bivariate Model with Fixed-effect Test Type
- Estimation by gllamm using Adaptive quadrature(nip=7)

```
xi: gllamm ttruth lgtse lgtsp  i.testtype, ///
i(study) nocons f(bin) l(logit) ip(m) nip(7)///
nrf(2) eqs(lgtse lgtsp) denom(num) adapt
```

ttruth	True(positive or negative) test results
lgtse and lgtsp	Logit-transforms of sensitivity and specificity
i.testtype	Dichotomous variable(PET=0 and CT=1)

University of Michigan
Medical School

Model 10

- Combined Test-specific Bivariate Models
- Estimation by gllamm with adaptive quadrature(nip=7)
- Accounts for all between-study and between-test variability

```
gllamm ttruth lgtsePET lgtspPET lgtseCT ///
lgtspCT, i(study) nocons f(bin) l(logit)///
eqs(lgtsePET lgtspPET lgtseCT ///
lgtspCT) nip(7) nrf(4) denom(num) ip(m) adapt
```

University of Michigan
Medical School

Results

Fit and Complexity Measures

Model	nparm	Deviance	BIC
Bivariate-FECov(Model 1)	6	488	516
Laplace-Bivariate-FECov(Model 2)	6	489	517
Bivariate-RECov(Model 3)	7	487	520
Laplace-Bivariate-RECov(Model 4)	7	487	520
Bivariate-Bivariate(Model 5)	10	484	531
Laplace-Bivariate-Bivariate(Model 6)	10	484	531
Quadrivariate(Model 7)	14	479	545
Laplace-Quadrivariate(Model 8)	14	479	545
Gllamm-Bivariate-FECov(Model 9)	6	488	516
Gllamm-Quadrivariate(Model 10)	14	479	545

University of Michigan
Medical School

Results

Fixed-effects Estimates

Parameter	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
lgtse	1.79(0.14)	1.79(0.14)	1.79(0.14)	1.79(0.13)	-	-	-	-
lgtsp	2.49(0.15)	2.49(0.15)	2.49(0.15)	2.49(0.15)	-	-	-	-
testtype	-1.68(0.10)	-1.68(0.10)	-0.16(0.11)	-0.16(0.11)	-	-	-	-
lgtsePET	-	-	-	-	1.77(0.16)	1.78(0.16)	1.74(0.16)	1.74(0.16)
lgtseCT	-	-	-	-	2.59(0.20)	2.59(0.20)	2.57(0.20)	2.57(0.20)
lgtspPET	-	-	-	-	0.69(0.14)	0.69(0.14)	0.66(0.14)	0.66(0.14)
lgtspCT	-	-	-	-	1.33(0.16)	1.33(0.16)	1.33(0.16)	1.33(0.15)

lgtse and lgtsp Logit-transforms of sensitivity and specificity
testtype Dichotomous variable(PET=0 and CT=1)

University of Michigan
Medical School

Results

Random-effects Estimates

Parameter	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
var(lgtse)	0.18(0.10)	0.18(0.10)	0.17(0.10)	0.16(0.10)	-	-	-	-
var(lgtsp)	0.32(0.13)	0.32(0.13)	0.31(0.13)	0.31(0.13)	-	-	-	-
var(testtype)	-	-	0.04(0.49)	0.04(0.48)	-	-	-	-
var(lgtsePET)	-	-	-	-	0.17(0.18)	0.17(0.18)	0.20(0.19)	0.20(0.18)
var(lgtseCT)	-	-	-	-	0.22(0.13)	0.21(0.12)	0.24(0.13)	0.23(0.13)
var(lgtspPET)	-	-	-	-	0.47(0.28)	0.46(0.27)	0.47(0.28)	0.46(0.27)
var(lgtspCT)	-	-	-	-	0.43(0.18)	0.42(0.18)	0.42(0.18)	0.42(0.18)
cov(lgtse,lgtsp)	-0.10(0.08)	-0.10(0.08)	-0.11(0.08)	-0.11(0.08)	-	-	-	-
cov(lgtsePET,lgtspPET)	-	-	-	-	-0.06(0.16)	-0.07(0.16)	-0.03(0.16)	-0.03(0.16)
cov(lgtseCT,lgtspCT)	-	-	-	-	-0.19(0.12)	-0.19(0.12)	-0.18(0.12)	-0.18(0.12)

var(lgtse) and var(lgtsp) variances of logit-transforms of sensitivity and specificity
var(testtype) variance of dichotomous variable(PET=0 and CT=1)
cov(lgtse,lgtsp) covariance of logit-transforms of sensitivity and specificity

University of Michigan
Medical School

Results

Test Performance Estimates

Index	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
PETsen	0.86(0.02)	0.86(0.02)	0.86(0.02)	0.86(0.02)	0.85(0.02)	0.85(0.02)	0.85(0.02)	0.85(0.02)
PETspe	0.92(0.01)	0.92(0.01)	0.92(0.01)	0.92(0.01)	0.93(0.01)	0.93(0.01)	0.93(0.01)	0.93(0.01)
PETlrp	11.2(1.6)	11.2(1.6)	11.2(1.6)	11.2(1.6)	12.2(2.3)	12.2(2.3)	12.0(2.2)	12.0(2.2)
PETlrn	0.15(0.02)	0.15(0.02)	0.15(0.02)	0.15(0.02)	0.16(0.02)	0.16(0.02)	0.16(0.02)	0.16(0.02)
PETldor	4.28(0.21)	4.28(0.21)	4.28(0.21)	4.28(0.20)	4.36(0.25)	4.36(0.24)	4.31(0.25)	4.31(0.25)
PETdor	72(15)	72(15)	72(15)	72(15)	78(19)	78(19)	74(19)	75(19)
CTsen	0.65(0.03)	0.65(0.03)	0.65(0.03)	0.65(0.03)	0.67(0.03)	0.67(0.03)	0.66(0.03)	0.66(0.03)
CTspe	0.79(0.02)	0.79(0.02)	0.79(0.02)	0.79(0.02)	0.79(0.03)	0.79(0.03)	0.79(0.03)	0.79(0.03)
CTlrp	3.09(0.34)	3.09(0.34)	3.12(0.37)	3.12(0.37)	3.20(0.37)	3.20(0.37)	3.15(0.37)	3.15(0.37)
CTlrn	0.44(0.034)	0.44(0.04)	0.44(0.04)	0.44(0.04)	0.42(0.04)	0.42(0.04)	0.43(0.04)	0.43(0.04)
CTldor	1.94(0.17)	1.94(0.17)	1.96(0.19)	1.96(0.18)	2.03(0.17)	2.03(0.17)	1.99(0.18)	1.99(0.18)
CTdor	6.99(1.20)	6.99(1.20)	7.08(1.32)	7.09(1.31)	7.61(1.30)	7.62(1.29)	7.30(1.29)	7.30(1.28)

University of Michigan
Medical School

Conclusions

1. The preferred model accounting for all between-study and between-test variability is the **Combined bivariate(quadrivariate) model**
2. The **Combined bivariate(quadrivariate) model** is, however, the most computationally complex and expensive
3. If interest is in diagnostic performance only, then the **Bivariate model with fixed-effect test type** may be preferred for simplicity

University of Michigan
Medical School

Conclusions

1. For the purpose of meta-analysis, one may use **Adaptive Gaussian Quadrature** or **Laplacian Approximation** without much loss of accuracy in coefficient and variance estimation
2. For each comparative model, estimates from **gllamm** were similar to **xtmelogit** in fit, complexity and test performance estimates
3. Convergence times appeared equivalent between **gllamm** and **xtmelogit** for estimation of either the more complex **Combined bivariate(quadrivariate) model** or the simpler **Bivariate model with fixed-effect test type**

University of Michigan
Medical School