## METAANALYTIC INTEGRATION OF DIAGNOSTIC ACCURACY STUDIES IN STATA



# Ben A. Dwamena, MD Department of Radiology (Nuclear Medicine)

Thursday October 25, 2007

FIRST WEST COAST STATA USERS' GROUP MEETING

### **OUTLINE**

- Evaluation of Diagnostic Tests
- □ Rationale For Meta-analysis
- □ Existing Methodology and Limitations
- □ Emerging alternative
- □ Implementation in Stata using midas

Thursday October 25, 2007

# CLINICAL EVALUATION OF A DIAGNOSTIC TEST

- □ Consider a population to be made up of two groups: those with a disease and those without it. A test aims to identify people as belonging to one of these two groups.
- It is usually assumed that a gold standard is available which can perfectly distinguish groups but cannot be used in routine practice due to problems such as invasiveness and/or cost.
- However alternative, more practical, tests are available which are imperfect.

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'
GROUP MEETING

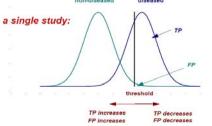
# CLINICAL EVALUATION OF A DIAGNOSTIC TEST

- Many test outcomes are measured on an explicit continuous scale (e.g. level of a chemical in the blood).
- □ Changing the threshold test level which defines a positive and a negative test will also change the performance characteristics of a test.

Thursday October 25, 2007

## THESHOLD EFFECT AND ROC

# Test threshold



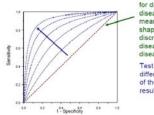
A plot of sensitivity against 1-specificity across the range of thresholds results in a receiver operating characteristic (ROC) curve.

Thursday October 25, 2007

FIRST WEST COAST STATA USERS' GROUP MEETING

### **VARIATION IN THRESHOLD VS ACCURACY**

#### ROC curve



Distributions of test results for diseased and nondiseased have the same mean, variance and shape. The test does not discriminate between the diseased and nondiseased groups.

Test accuracy increases as the difference between the means of the distributions of test results increases.

Thursday October 25, 2007

### 2 X 2 TABLE FOR SINGLE STUDY

	Present	Absent	Totals
Abnormal	a	ь	a+b
Normal	С	d	c+d
Totals	a+c	b+d	a+b+c+d
	Normal	(reference ter   Present	Abnormal a b Normal c d

Sensitivity = al/(a+c)
Specificity = dl/(b+d)
Positive Predictive value= al/(a+b)
Negative predictive value =dl/(c+d)
Likelihood ratio abnormal test= Sensitivity/(1-Specificity)
Likelihood ratio normal test= (1-Sensitivity//Specificity
Diagnostic Odds Ratio =(a×d)/(c×b)

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'
GROUP MEETING

### WHY META-ANALYSIS?

Critical Review and Statistical Combination of Results of Previous Research

- Too few patients
- Too selected a population
- No consensus regarding reproducibility, accuracy and impact of tests
- Data often scattered across many journals

Thursday October 25, 2007

### WHY META-ANALYSIS?

Using meta-analysis we can determine

- overall summary of diagnostic accuracy
- whether estimates of accuracy depend on the study design characteristics (study validity) of primary studies
- whether accuracy differs in subgroups
- areas for further research

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'
GROUP MEETING

## STATISTICAL ISSUES

- **□** Summary Estimates
- □ Threshold Variability
- Unobserved Heterogeneity
- □ Covariate effects
- □ Publication Bias
- □ Clinical Interpretation

Thursday October 25, 2007

# TRADITIONAL MODEL Summary ROC Analysis and Regression

For each study iCompute accuracy (log diagnostic odds ratio, inDOR0:  $D_i = \log \inf(y_{i'}) - \log \inf(\hat{y}_{i'})$ and implicit threshold:  $S_i = \log \inf(y_{i'}) + \log \inf(\hat{y}_{i'})$ 

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'
GROUP MEETING

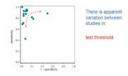
### **EXAMPLE DATA**

Peak Flow Velocity To Detect Renal Artery Stenosis



Thursday October 25, 2007

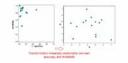
### **DISPLAY OF DATA IN ROC SPACE**



Thursday October 25, 2007

FIRST WEST COAST STATA USERS'
GROUP MEETING

## **SROC REGRESSION**



Thursday October 25, 2007

### **SROC REGRESSION**

The relationship between test accuracy and test threshold is modelled to estimate a summary ROC curve.

$$D = a + bS$$

This fixed effect model is generally fitted using linear regression (unweighted or weighted by inverse variance of InDOR).

b ≠ 0 ⇒ Accuracy depends on threshold resulting in an asymmetric SROC

b = 0 ⇒ Accuracy is independent of threshold resulting in a symmetric SROC

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'
GROUP MEETING

### LIMITATIONS OF SROC MODEL

- Fixed effects method: variation due only to threshold effect and within-study variability
- No average sensitivity/specificity estimate
- · Improper weighting of study estimates
- Need for continuity correction
- Does not account for measurement error in independent variable S
- Ignores potential correlation between D and S

Thursday October 25, 2007

# BIVARIATE MIXED-EFFECTS BINOMIAL REGRESSION MODEL

Let n11, n00, n01, and n10 be the number of true positives, true negatives, false positives, and false negatives, respectively. The model may be specified as follows:

$$\begin{split} & n_{00}^{i} \big| \boldsymbol{\mu}^{i} \sim \operatorname{Binomial} \left( N_{0}^{i}, \operatorname{Sp}^{i} \right), \ \operatorname{logit} \left( \operatorname{Sp}^{i} \right) = \mathbf{X}_{i} \boldsymbol{\alpha} + \boldsymbol{\mu}^{i}, \\ & n_{11}^{i} \big| \boldsymbol{\nu}^{i} \sim \operatorname{Binomial} \left( N_{1}^{i}, \operatorname{Se}^{i} \right), \ \operatorname{logit} \left( \operatorname{Se}^{i} \right) = \mathbf{Z}_{i} \boldsymbol{\beta} + \boldsymbol{\nu}^{i}, \\ & \left( \begin{matrix} \boldsymbol{\mu}^{i} \\ \boldsymbol{\nu}^{i} \end{matrix} \right) \sim N \bigg[ \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{matrix} \right), \begin{pmatrix} \boldsymbol{\sigma}_{\mu}^{2} & \rho \boldsymbol{\sigma}_{\mu} \boldsymbol{\sigma}_{\nu} \\ \rho \boldsymbol{\sigma}_{\mu} \boldsymbol{\sigma}_{\nu} & \boldsymbol{\sigma}_{\nu}^{2} \end{matrix} \right) \bigg], \end{split}$$

where logit(z) = log(z)-log(1-z), Xi and Zi are (possibly overlapping) vectors of covariates related to Sp and Se, respectively.

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'
GROUP MEETING

# BIVARIATE MIXED-EFFECTS BINOMIAL REGRESSION MODEL

• Estimates the amount of between-study variation in sensitivity and specificity separately, in addition to the degree of correlation between sensitivity and specificity via random effects. (assessment of heterogeneity and the possibility of an implicit threshold).

Thursday October 25, 2007

# BIVARIATE MIXED-EFFECTS BINOMIAL REGRESSION MODEL

- Produces summary estimates of sensitivity and specificity and their 95% confidence interval.
- Other measures derived from sensitivity and specificity such as the diagnostic odds ratios and likelihood ratios can be calculated.

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'
GROUP MEETING

# BIVARIATE MIXED-EFFECTS BINOMIAL REGRESSION MODEL

• A confidence ellipse around the mean values of logit sensitivity and specificity or a prediction ellipse for individual values of sensitivity and specificity, taking into account the possible (negative) correlation between sensitivity and specificity, may be derived easily.

Thursday October 25, 2007

# BIVARIATE MIXED-EFFECTS BINOMIAL REGRESSION MODEL

• Several summary ROC linear regression lines based on either the regression of logit sensitivity on specificity or the regression of logit specificity on sensitivity, or a orthogonal regression line by minimizing the perpendicular distances. These lines can be transformed back to the original ROC scale to obtain a summary ROC curve

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'
GROUP MEETING

# BIVARIATE MIXED-EFFECTS BINOMIAL REGRESSION MODEL

- Covariates can be added to the bivariate model to explicitly test whether sensitivity or specificity or both are different between two diagnostic technologies.
- Does not require an ad hoc continuity correction when the number of true positives, true negatives, false positives, or false negatives is zero in a study

Thursday October 25, 2007

## BINOMIAL REGRESSION MODEL

### **XTMELOGIT CODE**

```
gen study = _n
gen ttruth1 = `tn'
gen ttruth2 = `tp'
gen num1 = `tn'+`fp'
gen num2 = `tp'+`fn'
reshape long num ttruth, i(study) j(dtruth)
tabulate dtruth, generate(disgrp)

xtmelogit ttruth disgrp1 disgrp2, noc ||study: ///
disgrp1 disgrp2, noc cov(unstr), bin(num) ///
intp(`nip') var nolr nofet noret nohead
```

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'
GROUP MEETING

### **BINOMIAL REGRESSION MODEL**

#### **GLAMM CODE**

```
gen study = _n
gen ttruth1 = `tn'
gen ttruth2 = `tp'
gen num1 = `tn'+`fp'
gen num2 = `tp'+`fn'
reshape long num ttruth, i(study) j(dtruth) string
tabulate dtruth, generate(disgrp)
eq disgrp1: disgrp1
eq disgrp2: disgrp2
gllamm ttruth disgrp1 disgrp2, nocons///
  i(study) nrf(2) eqs(disgrp1 disgrp2) ///
f(bin) l(logit) denom(num)
```

Thursday October 25, 2007

### **EXAMPLE DATASET:**

### Axillary PET Scan in Primary Breast Cancer

Author	year	tp	fp	fn	tn	pro	size	verif	tdes	rdes	sdes	rep	cons	spect	blind
Tse	1992	4	0	3	3	1	0	0	1	1	0	0	0	0	0
Adler1	1993	8	0	1	10	1	0	1	1	1	1	0	1	0	0
Hoh	1993	6	0	3	5	0	0	0	0	0	0	0	0	1	0
Crowe	1994	9	0	1	10	1	0	1	1	1	1	0	0	0	0
Avril	1996	19	1	5	26	1	1	0	1	1	1	0	0	1	0
Bassa	1996	10	0	3	3	0	0	1	1	1	1	0	0	0	0
Scheidhauer	1996	9	1	0	8	1	0	0	1	1	1	0	0	1	0
Utech	1996	44	20	0	60	1	1	1	1	1	0	0	0	1	0
Adler2	1997	19	11	0	20	1	0	0	1	1	1	0	1	1	1
Palmedo	1997	5	0	1	14	1	0	1	1	0	1	0	0	1	0
Noh	1998	12	0	1	11	0	0	1	0	0	0	0	0	1	0
Smith	1998	19	1	2	28	1	0	1	0	1	1	0	0	1	0
Rostom	1999	42	0	6	26	0	0	0	1	1	0	0	1	1	0
Yutani1	1999	8	0	2	16	1	0	0	1	1	1	0	1	1	0
Hubner	2000	6	0	0	16	0	0	0	1	0	0	1	1	1	0
Ohta	2000	14	0	5	13	0	0	1	1	1	1	0	0	1	0
Yutani2	2000	8	0	8	22	1	0	1	1	1	1	0	1	1	0
Greco	2001	68	13	4	82	1	1	1	1	1	1	1	1	1	0
Schirrmeister	2001	27	6	7	73	1	1	0	1	1	1	0	0	1	0
Yang	2001	3	0	3	12	0	0	1	1	1	1	0	0	1	0

Thursday October 25, 2007

FIRST WEST COAST STATA USERS' GROUP MEETING

## **SYNTAX FOR** midas

# midas varlist [if exp] [in range] [, id(varname) year(varname)

modeling\_options
quality\_assessment\_options
reporting\_options
exploratory\_graphics\_options
publication\_bias\_options
forest\_plot\_options
heterogeneity\_options
roc\_options
probability\_revision\_options
general\_graphing\_options \*]

Thursday October 25, 2007

## **MODELING Options**

#### estimator(g|x)

 provides a choice between estimation with xtmelogit in release 10 versus gllamm in version 9 or earlier.

#### nip()

- specifies the number of integration points used for maximum likelihood estimation based on adaptive gaussian quadrature.
- Default is set at 15 for midas (default in xtmelogit is 7).
- Higher values improve accuracy at the expense of execution times.
- model will be estimated by Laplacian approximation using nip(1)

Thursday October 25, 2007

FIRST WEST COAST STATA USERS' GROUP MEETING

## **QUALITY ASSESSMENT Options**

#### qualitab

creates, using optional varlist of study quality items (presence=1, other=0) a table showing frequency of methodologic quality items.

#### qualibar

creates, combined with optional variest of study quality items (presence=1, other=0) calculates study-specific quality scores and plots a bargraph of methodologic quality.

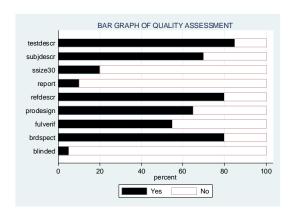
#### Qlab

may be combined with qualitab or qualibar to use variable labels for table and bargraph of methodologic items.

Thursday October 25, 2007

### **QUALITY ASSESSMENT**

midas tp fp fn tn prodesign ssize30 fulverif testdescr /// refdescr subjdescr report brdspect blinded, qualib



Thursday October 25, 2007

FIRST WEST COAST STATA USERS' GROUP MEETING

### **EXPLORATORY GRAPHICS Options**

### qqplot(dss | dlor | dlr)

plots a normal quantile plot to (a) check the normality assumption (b) investigate whether all studies come from a single population (c) search for publication bias

#### cum

produces a cumulative meta-analysis plot using year of publication as measure for temporal evolution of evidence.

#### inf

investigates the influence of each individual study on the overall meta-analysis summary estimate.

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'

## **EXPLORATORY GRAPHICS Options**

### bivbox

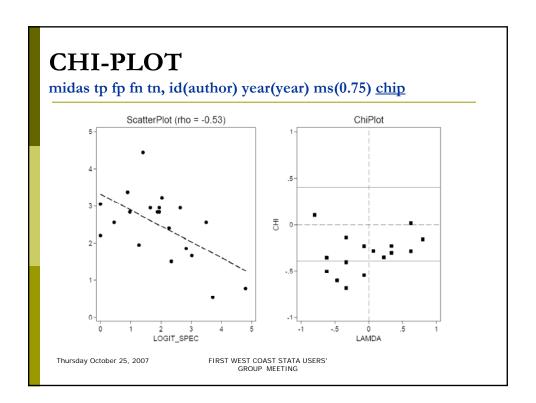
implements a two-dimensional analogue of the boxplot for univariate data

It is used to assess distributional properties of sensitivity versus specificity and for indentifying possible outliers.

### chiplot

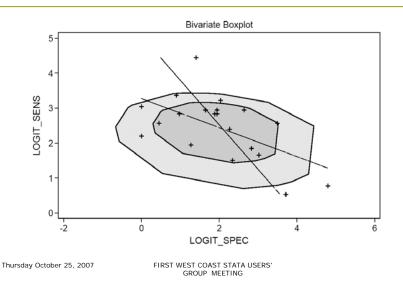
creates a chiplot for judging whether or not the paired performance indices are independent by augmenting the scatterplot with an auxiliary display.

Thursday October 25, 2007



### **BIVARIATE BOXPLOT**

midas tp fp fn tn, id(author) year(year) ms(0.75) bivbox



# **REPORTING Options**

#### table(dss | dlr | dlor)

Creates a table of study specific performance estimates

Reports measure-specific summary estimates

Provides results of homogeneity (chi\_squared) and inconsistency(I\_squared) tests.

**dss, dlr or dlor** represent the paired performance measures: sensitivity/specificity, positive/negative likelihood ratios and diagnostic score/odds ratios.

Thursday October 25, 2007

### **SUMMARY RESULTS**

midas tp fp fn tn, id(author) year(year) es(x) ms(0.75) res(all)

Bivariate Binomial Mixed Model

Number of studies = 20

Reference-positive Subjects = 393

Reference-negative Subjects = 513

Pretest Prob of Disease = 0.434

Between-study variance ( varlogitSEN) = 0.851, 95% CI = [0.290-2.499]

Between-study variance (varlogitSPE) = 1.767, 95% CI = [0.508-6.148]

Correlation (Mixed Model) = -1.000

ROC Area,  $\Lambda$ UROC = 0.97 [0.95 - 0.98]

Heterogeneity (Chi-square): LRT\_Q = 57.147, df = 2.00, LRT\_p = 0.000

Inconsistency (I-square): LRT\_I2 = 96.50, 95% CI = [93.93-99.07]

Thursday October 25, 2007

FIRST WEST COAST STATA USERS' GROUP MEETING

## **REPORTING Options**

#### results(all)

provides summary statistics, group-specific between-study variances, likelihood ratio test statistics and other global homogeneity tests.

#### results(het)

provides group-specific between-study variances, likelihood ratio test statistics and other global homogeneity tests.

#### results(sum)

provides summary statistics for all performance indices

Thursday October 25, 2007

### **SUMMARY RESULTS**

midas tp fp fn tn, id(author) year(year) ms(0.75) es(x) res(sum)

Parameter	Estimate 95% CI
Sensitivity	0.854 [0.775, 0.908]
Specificity	0.969 [ 0.912, 0.989]
Positive Likelihood Ratio	27.278 [9.811, 75.838]
Negative Likelihood Ratio	0.151 [0.098, 0.232]
Log Odds Ratio	5.198 [4.262, 6.135]
Diagnostic Odds Ratio	180.989 [70.920, 461.886]

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'
GROUP MEETING

## **PUBLICATION BIAS Options**

#### pubbias

performs linear regression of log odds ratios on inverse root of effective sample sizes as a test for funnel plot asymmetry in diagnostic metanalyses.

A non-zero slope coefficient is suggestive of significant small study bias (pvalue  $\leq$  0.10).

#### maxbias

performs Copas' worst-case sensitivity analysis for publication bias.

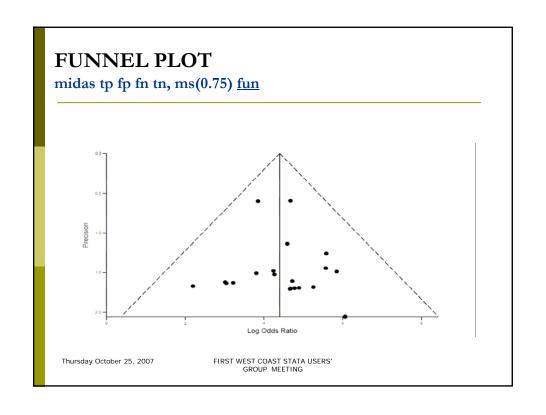
calculates the upper limit of no of missing studies that will overturn statistical significance

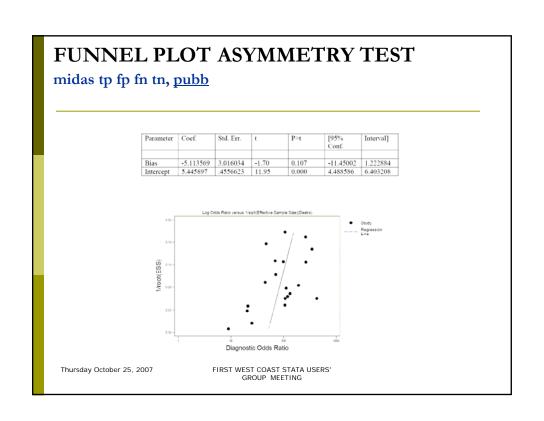
estimates the minimun likely publication probability (Copas and Jackson, 2004).

#### funnel

plots a funnel plot, a two-dimensional graph with sample size on one axis and effectsize estimate on the other axis. The funnel plot capitalizes on the well-known statistical principle that sampling error decreases as sample size increases.

Thursday October 25, 2007





# **FOREST PLOT Options**

### forest(dss | dlr | dlor)

Creates summary graphs with study-specific(box) and overall(diamond) point estimates and confidence intervals for each performance index pair

Confidence intervals lines are allowed to extend between 0 and 1000 beyond which they are truncated and marked by a leading arrow.

### fordata

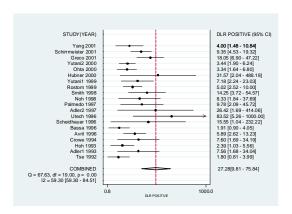
adds study-specific performance estimates and 95% CIs to right y-axis.

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'
GROUP MEETING

### **FOREST PLOT**

midas tp fp fn tn, id(author) year(year) es(x) ms(0.75) ford for(dlr)

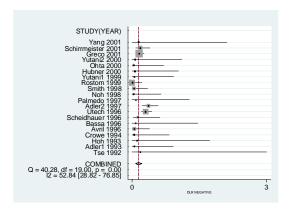


Thursday October 25, 2007

FIRST WEST COAST STATA USERS'

### **FOREST PLOT**

midas tp fp fn tn, id(author) year(year) es(x) ms(0.45) for(dlr)



Thursday October 25, 2007

FIRST WEST COAST STATA USERS' GROUP MEETING

## **HETEROGENEITY Options**

#### $galb(dss\,|\,dlr\,|\,dlor)$

The standardized effect measure (e.g. for lnDOR, lnDOR/precision) is plotted (y-axis) against the inverse of the precision(x-axis). A regression line that goes through the origin is calculated, together with 95% boundaries (starting at +2 and -2 on the y-axis). Studies outside these 95% boundaries may be considered as outliers.

#### hetfor

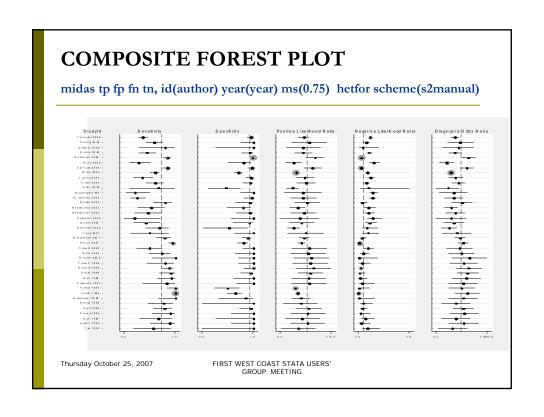
creates composite forest plot of all performance indices to provide a general view of variability. Confidence intervals lines are allowed to extend between 0 and 1000 beyond which they are truncated and marked by a leading arrow.

#### covars

combined with an optional varlist permits univariable metaregression analysis of one or multiple covariables.

Thursday October 25, 2007

z   0.93  0.91  0.05
0.93
0.93
0.91
0.08
1.00



# **ROC Options**

#### sroc1

plots observed datapoints, summary operating sensitivity and specificity in SROC space.

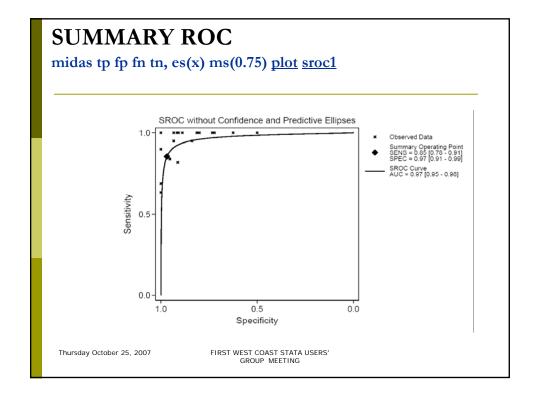
#### sroc2

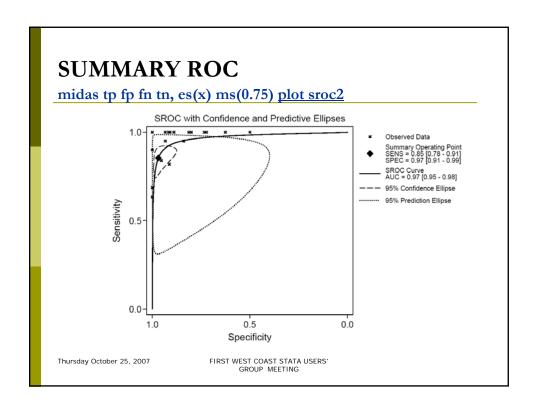
adds confidence and prediction contours.

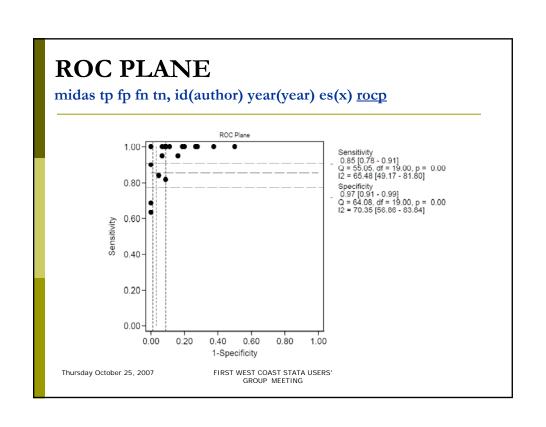
#### rocplane

plots observed data in receiver operating characteristic space (ROC Plane) for visual assessment of threshold effect.

Thursday October 25, 2007







### **PROBABILITY REVISION Options**

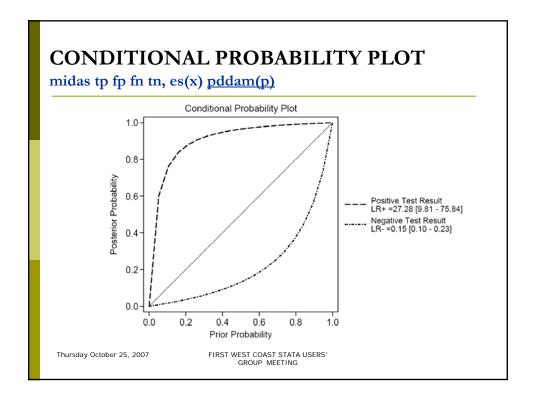
### pddam(p|r)

produces a line graph of post-test probabilities versus prior probabilities between 0 and 1 using summary likelihood ratios

#### **Irmatrix**

creates a scatter plot of positive and negative likelihood ratios with combined summary point. Plot is divided into quadrants based on strength-of-evidence thresholds to determine informativeness of measured test.

Thursday October 25, 2007



#### UNCONDITIONAL PREDICTIVE VALUES

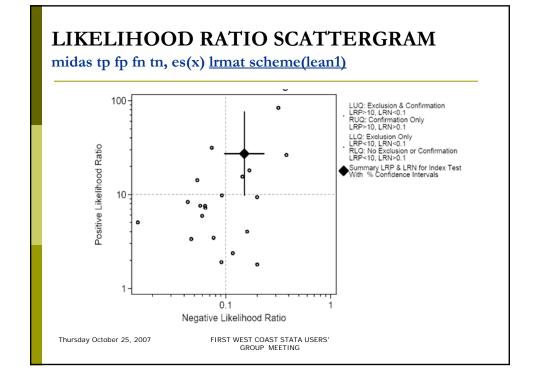
midas tp fp fn tn, ms(0.75) es(x) pddam(r)

Prior Prevalence = Uniform [0.00,1.00]

Unconditional Positive Predictive Value= 0.91, 95% CI = [0.87-0.95]

Unconditional Negative Predictive Value = 0.78, 95% CI = [0.74-0.82]

Thursday October 25, 2007



### **PROBABILITY REVISION Options**

### fagan

creates a plot showing the relationship between the prior probability, the likelihood ratio(combination of sensitivity and specificity), and posterior test probability.

### prior()

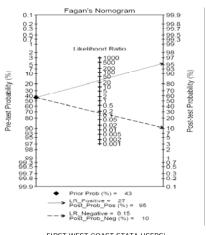
combined with fagan allows user to specify a pretest probability overriding the default of using disease prevalence calculated from data when fagan is invoked alone.

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'

### FAGAN'S NOMOGRAM I

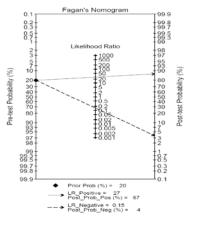
midas tp fp fn tn, es(x) fagan scheme(lean1)



Thursday October 25, 2007

### **FAGAN'S NOMOGRAM II**

midas tp fp fn tn, es(x) fagan prior(.20) scheme(lean1)



Thursday October 25, 2007

FIRST WEST COAST STATA USERS'
GROUP MEETING

### STANDING ON SHOULDERS OF GIANTS

Thanks to

- -Richard Sylvester and Ruth Carlos for encouragement, suggestions and testing of midas
- -Richard Riley for trusting me with pre-prints of his work on bivariate meta-analysis.
- -Joseph Coveney for posting syntax for the bivariate model using gllamm on Statalist.
- -Sophia-Rabe-Hesketh and other authors of **gllamm** for their work,
- -Derek Wenger for assistance with coding Fagan's plot.
- -Nick Cox for his **polarsm** which was adapted for **bivbox** option in **midas** and for **mylabels**.
- -Roger Harbord for his metareg, metafunnel and metamodbias programs which provided very useful ideas for {help midas}.

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'

### THANK YOU FOR YOUR ATTENTION

- QUESTIONS?
- SUGGESTIONS
- COMMENTS

Thursday October 25, 2007

FIRST WEST COAST STATA USERS'
GROUP MEETING

# **COMING SOON**

New and Improved MIDAS package with routines for metaanalysis and meta-regression of roc curve area and continuous test result data.

Thursday October 25, 2007