Hartford Hospital Research Program Research Methods Lecture Series Part IV:

Choosing the Appropriate Statistic (Part II)

Jan 4, 2010

The Effect of Attending (or Is It Presenting?) Statistics Workshops

- October: Basic concepts of research design
- November: Concepts of inferential statistics
- December: Choosing the right statistic Part I
- **January: Choosing the right statistic Part II**
- February: Meta analysis and clinical trials
- March: Grant-writing

Presenters:

- Tara McLaughlin, PhD, MPE

Senior Scientist/ Grantwriter

(tmclaug@harthosp.org)

- Jyoti Chhabra, PhD

Senior Scientist (jchhabr@harthosp.org)

- Ilene Staff, PhD

Senior Scientist (istaff@harthosp.org)

- David O'Sullivan, PhD

Senior Scientist (dosulli@harthosp.org)

Contents of the Presentation

- Multivariate analysis
- Multivariate Analysis of Variance (MANOVA), and (Multivariate) Analysis of Covariance (ANCOVA/MANCOVA)
- Multiple Regression
- Logistic Regression
- Survival Analysis
- Cox Regression
- Diagnostic Tests – ROC curves
- Other statistics

Multivariate Analysis

- We concentrate on multivariate analyses which are far more popular recently
- What is so special? – don't all analyses include more than 1 variable?
- Multivariate techniques are used when there are many outcomes ("dependent variables", DVs) or many factors ("independent variables", IVs) that are all correlated with one another to varying degrees

Why Multivariate Analysis?

Why are they more popular now?

- "The world is multivariate" - think how one outcome can be determined by many factors – genetics, nutrition, demographics, health history, access to care, etc., and how each of these might be related to the others
- Modern computers and software allow them to be done in reasonable time and effort – not a lifetime task to do one analysis!

Revisiting Differences Among Means

Analyzing Differences in Means Which Test(s) to Use

	Parametric Measures	Non-parametric Measures
Independent Groups	T-test (2 groups) One way ANOVA (>2 groups) Factorial ANOVA (2 or more independent variables)	Mann-Whitney or Wilcoxon Ranked Sum (2 groups) Kruskal-Wallis (>2 groups)
Repeated Measures	Paired t-test (2 groups) Repeated Measures ANOVA or GLM	Wilcoxon Signed Rank (2 groups) Friedman (>2 groups)

Parametric ANOVA or GLM can be expanded to include covariates (ANCOVA), multiple dependent variables (MANOVA)

MANOVA

- MANOVA = **m**ultivariate **an**alysis **o**f **va**riance
- Like t-test or ANOVA, it compares means among groups
  - T-Test for 2 groups
  - One-way ANOVA for >2 groups
  - Factorial ANOVA – for multiple groups based on more than one variable
  - MANOVA – when you have more than one dependent variable

MANOVA

- Why more than one dependent variable?
- Multiple measures of same ‘construct’
- Special case for longitudinal designs – have the same outcome at repeated intervals
- Look at the effect of factor(s) on all the dependent variables at once
- Procedure actually creates a new dependent composite that maximizes the group differences (but this is transparent to user)

MANOVA *vs.* ANOVA

- Why not just do multiple ANOVAs??
  - protects against Type I errors that might occur
- Any disadvantages/problems?
  - more complicated design - interpretation more ambiguous

ANCOVA

- ANCOVA = **An**alysis of **Cova**riance
- Extension of ANOVA looking at the effects of the factor(s) after an adjustment based on one or more other variables (covariates)
- There is only one dependent variable but can be multiple factors and multiple covariates
- Recognizes that a relationship between two variables exists in context of other relationships

ANCOVA vs ANOVA

ANCOVA facilitates:

- decrease in the "noise" or error term making it more sensitive to any differences among groups
- Adjustment of the means of the dependent measures as if they were all equal on the covariates – can simulate 'random' assignment when this is not practically or ethically possible

MANCOVA

- MANCOVA = (**M**ultivariate) **An**alysis of **Cova**riance
- A combination of both MANOVA and ANCOVA – multiple dependent variables and multiple covariates –
- Advantages and disadvantages of both
- Inclusion of covariates makes it similar to regression analysis

Revisiting Multiple Regression

Multiple Regression

- Analyzes the strength of the relationship between one dependent variable and a set of predictor variables
- Three approaches – simultaneous (all predictors at once), hierarchical (based on logic) or stepwise (based solely on statistical criteria) using individual predictors or sets (blocks) of them

Measures in Multiple Regression

- Multiple R: correlation with all predictors
- Multiple R^2 : amount of variance in DV explained by all predictors
- Change in R^2 (or R^2 cha) – for hierarchical how much additional variance explained by last predictor or block added
- Partial and Semi-partial correlations: indicators of strength of relationship for individual predictors.

Multiple Regression: Other issues

- Set up for continuous value or dichotomous predictors
- Categorical variables can be entered as predictors if expressed as series of dichotomous "dummy" variables – most software does this for you
- Sample size: $N \geq 8 * \#\ IVs + 50$ for multiple R and $N \geq \#\ IVs + 104$; increase for smaller effect sizes or DV not normal

Logistic Regression

- a special form of multiple regression
- predicts a dichotomous outcome variable, e.g., Alive/Dead, Diseased/Non-diseased
  - More complicated variations include polychotomous outcomes
- used extensively in epidemiologic research (case-control) to establish risk of disease associated with a set of exposures
- provides an odds ratio ("OR") for each of the predictor variables

Odds Ratios

- Compares likelihood of one of two outcomes given a predictor's value
- Comparison to a specific reference can be to the mean of a continuous variable or a selected value of a categorical predictor
  - Multiple categories will yield multiple ORs
  - For reference can choose first or last or each compared to one before, e.g.
- Odds ratio (OR) = 1.0, no difference; OR=1.23, 23% more likely than reference group; OR=0.65, only 65% as likely

Odds Ratios (cont'd)

- Listed as Exp (B) in output
- Should be reported with confidence intervals; CIs that cross 1.0 (e.g., 0.85 – 1.15) are not statistically significant
- Common examples:
  - Race: African-American vs. Caucasian (ref)
  - Gender:male vs. female
  - Age: 1 standard deviation above mean vs. mean

To David

Survival Analysis

Concerned with the time to the occurrence of any critical event, not just mortality; it is a "time-to-event" analysis

- Events can be positive or negative:
  - Death (e.g., disease specific death, all-cause mortality)
  - Relapse of disease; event-free survival
  - Re-admission
  - Successful conception
  - Reaching a specified level of function
- Median Survival – requires ≥50% of cases to have had an event
- Comparisons between groups generally performed with ***Kaplan-Meier***

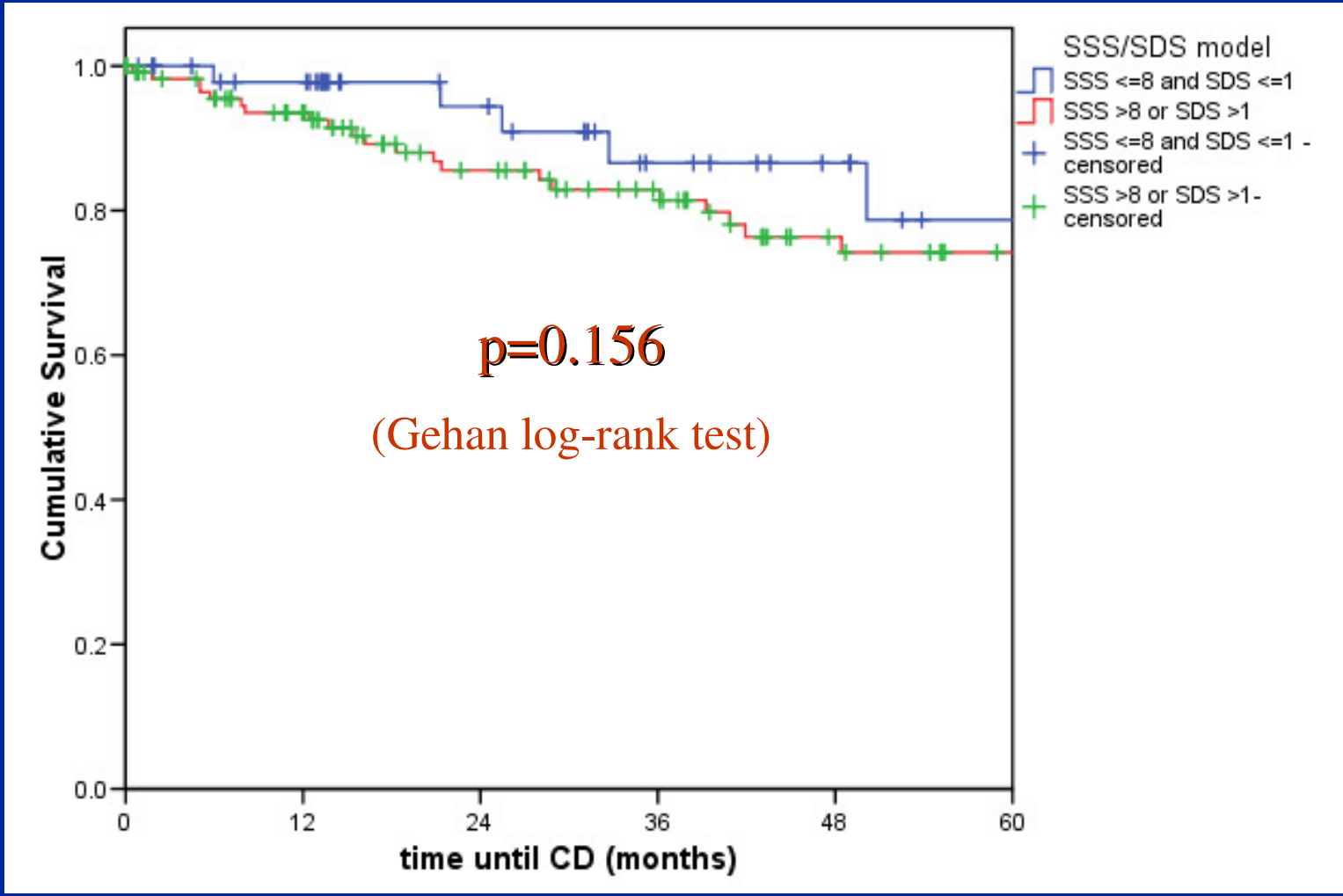
Issue in Survival Analysis: "Missing" Data

- Various reasons why information on event and time to event might be unknown:
  - Study ends before all patients reach the event
  - Patient withdraws or is excluded
  - Patient lost to follow-up (LTF)
  - Onset time is unknown (happened before study inclusion; *e.g.*, smoking history)
- These are called "censored" data; treated specially in Survival Analysis
- Information is known for some period of time - before LTF, until last day of study, etc.

Kaplan-Meier Analysis

- Used to evaluate if there is equality in the survival distribution
- Set up to handle censored data
- It is not a "simple" survival percent at a given time, nor the mean time to event
- Uses a statistical survival rate that recalculates a cumulative percent each time a subject has an event or data are lost
- Statistics available for comparisons of groups – log rank

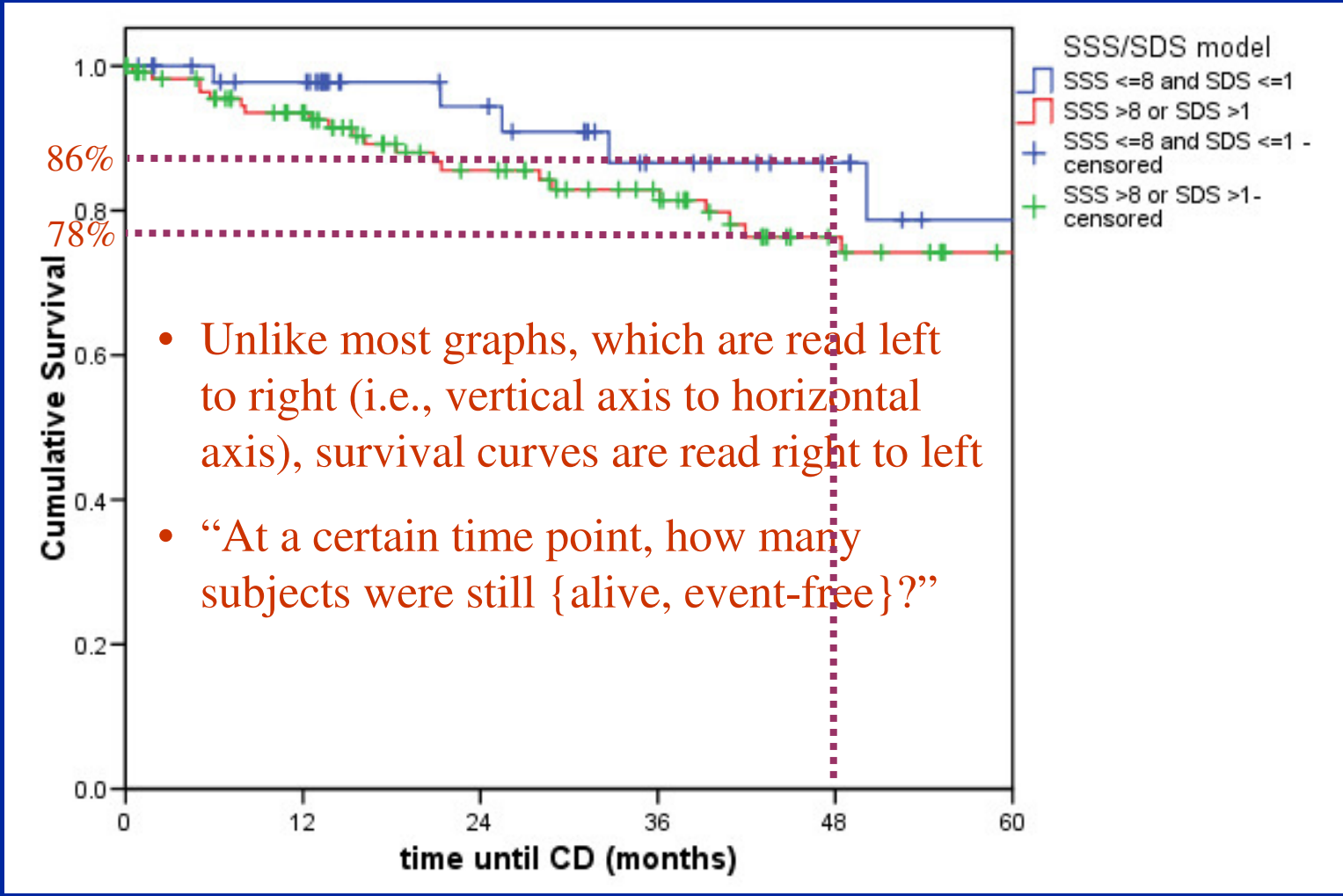
Kaplan-Meier Survival Curve ("Time-to-Event")



p=0.156

(Gehan log-rank test)

Reading/Interpreting Survival Curves



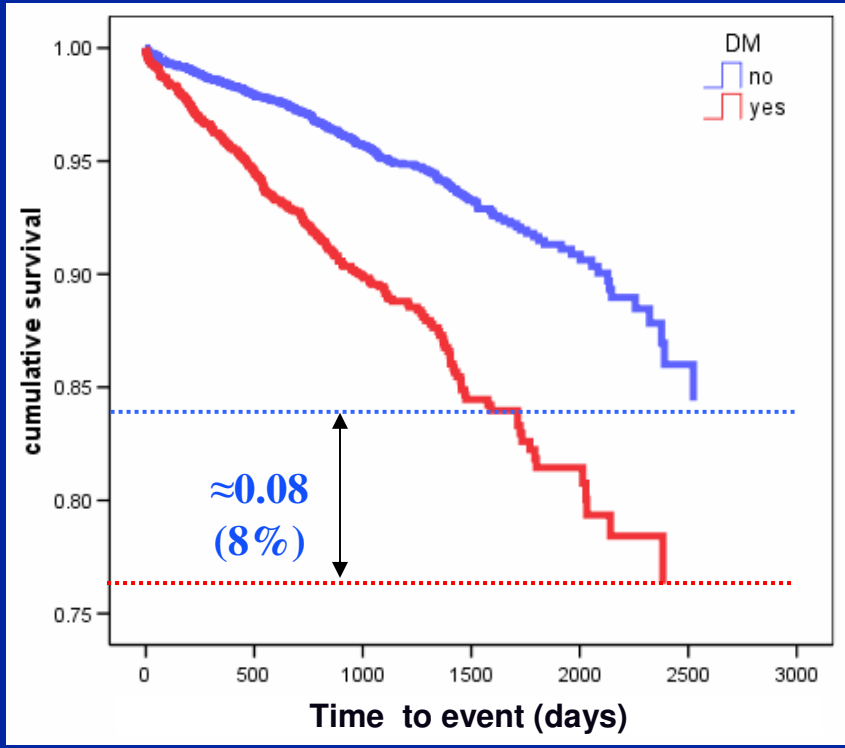
- Unlike most graphs, which are read left to right (i.e., vertical axis to horizontal axis), survival curves are read right to left
- "At a certain time point, how many subjects were still {alive, event-free}?"

Cox Regression

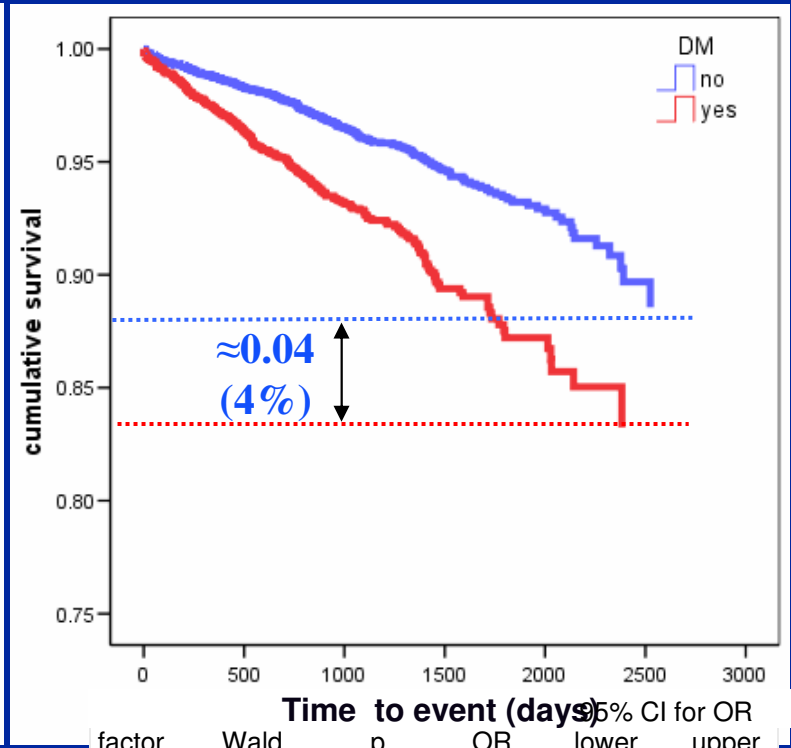
- What if you are interested in how several variables might affect “time to event”?
- Cox Regression is a method for modeling time-to-event data, including predictor variables (covariates) in the model in the presence of censored cases.
- Think of it as Kaplan-Meier meets Logistic Regression

Cox Regression - Graphically

Unadjusted (no *covariates*)



Adjusted (significant *covariates*)



factor	Wald	p	OR	95% CI for OR lower	upper
A	2.179	0.140	1.143	0.957	1.365
B	2.292	0.130	0.859	0.705	1.046
C	3.629	0.057	0.822	0.672	1.006
D	53.458	<0.001	0.408	0.321	0.519
E	24.374	<0.001	0.562	0.447	0.707
F	21.764	<0.001	1.031	1.018	1.044

Let's Change Gears a Bit

- Diagnostic tests/ROCs/AUCs
- Agreement
- Q&A

Analyzing Diagnostic Tests

Value of diagnostic test is to detect presence / absence of disease or condition

Test	Disease Disease Present	No Disease
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

Analyzing Diagnostic Tests

Identification vs. Prediction

	Disease / Condition	
Test / Risk	Present	Absent
Present	A	B
Absent	C	D

- IDENTIFICATION – focuses on the *disease* and looks *back* to see how many had the *risk*:
  - Sensitivity, SN: the probability of a positive test result when the condition is present [A / (A +C)]
  - Specificity, SP: the probability of a negative test result when the condition is not present [D / (B + D)]
- PREDICTION – focuses on the *risk* and looks *forward* to see how many acquired/manifested the *disease*:
  - Positive Predictive Value, PPV: among those that had the *risk*, what % wound up with the condition [A / (A + B)]
  - Negative Predictive Value, NPV: among those without the *risk*, what % wound up without the condition [D / (C + D)]

Analyzing Diagnostic Tests
Receiver **O**perating **C**haracteristics (ROC)

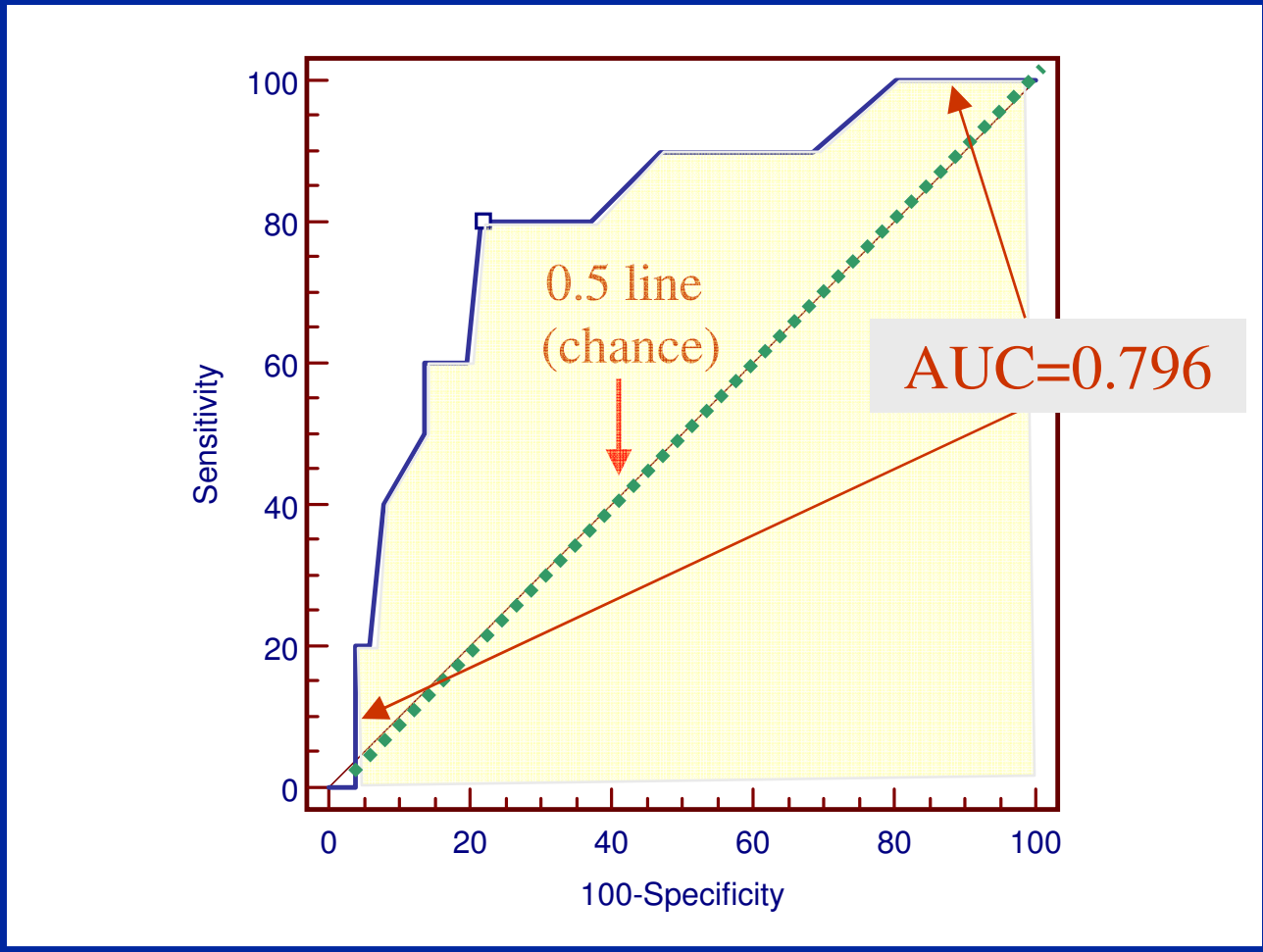
- the diagnostic test often is not dichotomous
- need to identify threshold for positive diagnosis
- ROC - procedure that compares a continuous (score) diagnostic test with a dichotomous result, *e.g.*, disease or no disease
- plots SN and SP (as 1-SP) across continuum of all available data points
- ROC analysis highlights mathematically best cut-point = highest combined sensitivity and specificity

Analyzing Diagnostic Tests
Area Under the Curve (AUC)

- AUC = summary descriptive statistic
  - *p* value can be generated (by integrating across all individual points) comparing calculated AUC to "just guessing" (=0.5)
  - Generally an investigator seeks $1.0 \geq AUC < 0.5$
  - Two or more AUCs can be compared statistically to suggest the best diagnostic or compare new test to "gold standard"
- The clinical "best" threshold may require a different SN/SP balance – implications of two possible errors

Diagnostic Tests: ROC Analyses

Area under the curve (AUC), obtained by integrating all points along the curve



But wait! There's more!

- One type of test that we couldn't seem to fit anywhere else, so…
- It gets a section all by itself:

Analyzing Agreement - Kappa

- Measures agreement between two nominal or ordinal variables; "correctness" of rating not impt.
- Usually used to assess inter-rater reliability but appropriate for two devices, methods, etc.
- *Not* "simple" (%) agreement – compares number of agreements and disagreements to that "expected by chance"
  - κ statistic = 0 if same as chance;
  - κ = 1.0 if perfect agreement;
  - κ = 0.8 "standard" for good agreement
- For ordinal data, a weighted Kappa will take into account amount of disagreement

Summary and Questions

- This session and the preceding one reviewed concepts of research design and a very quick review of most of the common statistical approaches
- Next month: meta-analysis and clinical trials
- Any questions for any of us?