

Evaluating Displays of Clinical Information

David S. Pieczkiewicz, PhD
NIBIB / CIBM Postdoctoral Fellow
Biomedical Informatics Research Center
Marshfield Clinic Research Foundation
&
University of Wisconsin, Madison

What is Evaluation?

- The systematic determination of the merit, worth, or significance of an entity
- Quantitative and qualitative approaches
- Experimental and non-experimental (e.g., controlled and non-controlled)
- Focus groups, RCTs, and everything in between

Levels of Diagnostic Efficacy

| | |
|-------------------------------------|--------------------------------|
| <i>Technical efficacy</i> | physical validity? |
| <i>Diagnostic accuracy</i> | statistical performance? |
| <i>Diagnostic-thinking accuracy</i> | affects physicians' estimates? |
| <i>Therapeutic efficacy</i> | affects patient management? |
| <i>Patient-outcome efficacy</i> | affects patient health? |
| <i>Societal efficacy</i> | wider social cost/benefit? |

from Fryback and Thornbury (1991)

Evaluation for EHRs

- EHRs usually assessed in terms of efficacy
- How well do they “work”?
 - Clinical utility
 - Clinical Outcomes
 - Usability
 - User acceptance
- Many EHR evaluations stop at user acceptance

This is good, but incomplete!

Elting *et al.* (1999)

Table

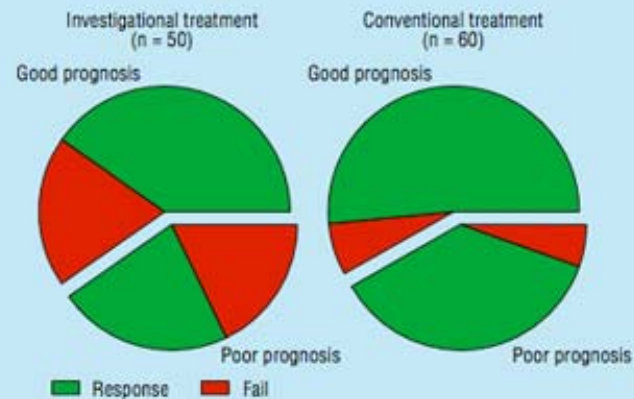
| | Conventional treatment | | Investigational treatment | |
|----------------|------------------------|--------|---------------------------|--------|
| | Total no | % Fail | Total no | % Fail |
| Good prognosis | 30 | 30 | 35 | 11 |
| Poor prognosis | 20 | 45 | 25 | 12 |
| Total | 50 | 38 | 60 | 12 |

(Negatively framed tables displayed failure rates in red.
Positively framed tables displayed response rates in green)

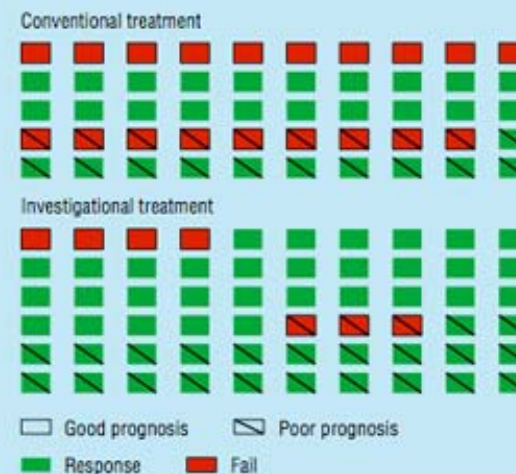
Bar graph



Pie chart



Icon



Measuring Efficacy

- *Accuracy*: How often or well the target task is completed (action, decision, etc.)
- *Latency*: How long it takes to perform the task, independent of accuracy
- *Preference*: What users feel comfortable with

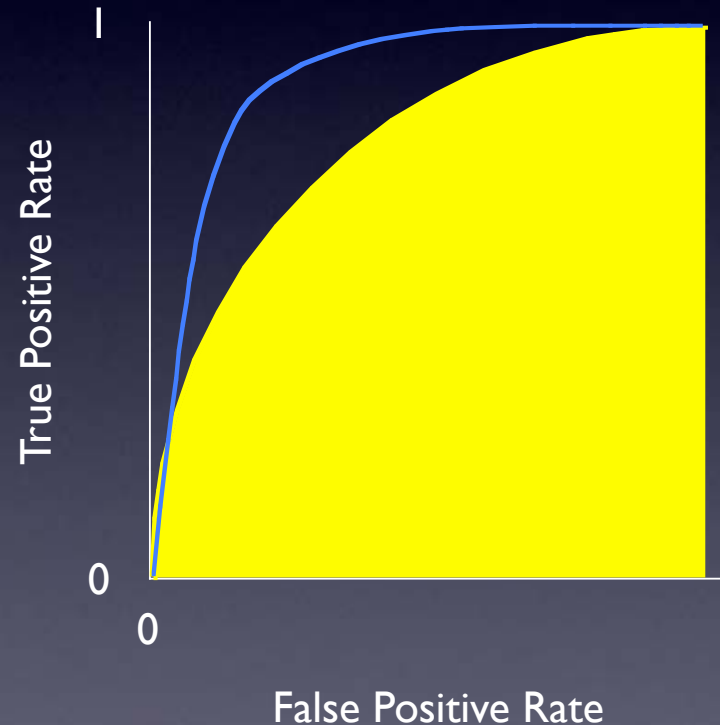
from Starren and Johnson (2000)

Decision Accuracy

- Percent correct
 - Easy to measure and report
 - Misses many decision distinctions (true and false positives and negatives, etc.)
- Sensitivity, specificity, positive predictive value, negative predictive value
 - Provides more information
 - Provides measures for particular cutoffs and prevalences

ROC Analysis

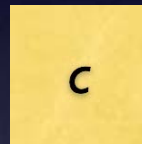
- Receiver-operating characteristic (ROC) curves describe accuracy over all cutoffs
- Area under curve describes overall accuracy of decisions
- Multiple curves can compare the performance of two or more visualizations



MRMC ROC Analysis

- Multiple-reader multiple-case (MRMC) ROC analysis developed for radiology
- Multiple readers assess multiple cases in each modality (visualization) of interest
- Decisions given on probability scale
- Decisions collated to generate ROC curve areas and variance information
- Determines if different modalities have statistically different accuracies

The MRMC Design



A case c contains the medical information needed to assess a patients' condition at a particular time

The MRMC Design

| | |
|-------|-----|
| c_1 | +/- |
| c_2 | +/- |
| ... | +/- |
| c_i | +/- |

For multiple cases c_i , some cases are positive for the feature of interest and some are negative

The MRMC Design

| | | | | |
|-------|-------|-------|-----|-------|
| c_1 | | | | |
| c_2 | | | | |
| ... | | | | |
| c_i | | | | |
| | m_1 | m_2 | ... | m_j |

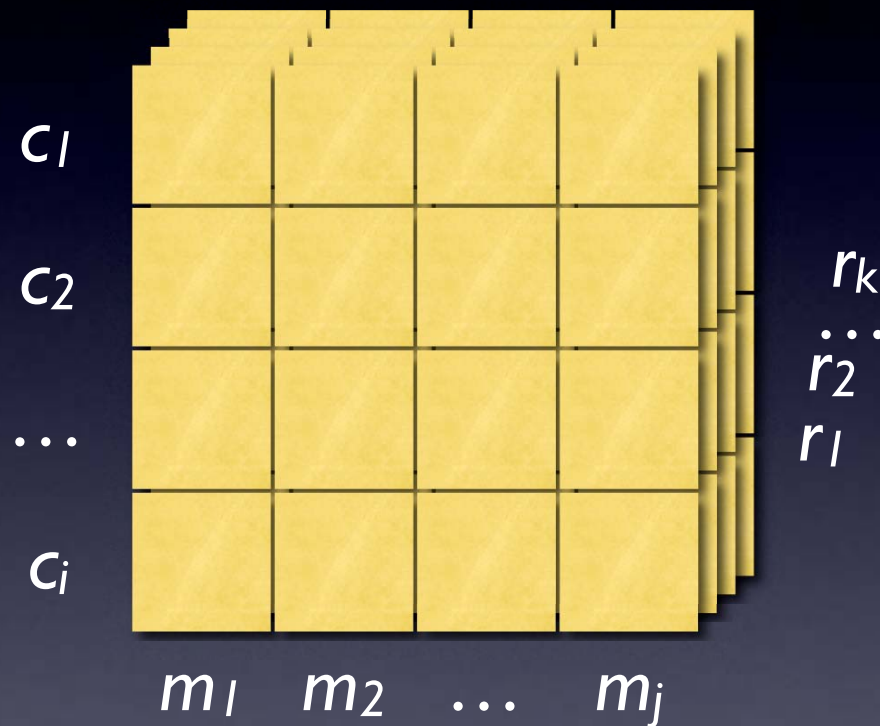
Each case c_i is viewed under each modality m_j

The MRMC Design

| | | | | |
|-------|----------|----------|----------|----------|
| c_1 | d_7 | d_2 | d_9 | d_{10} |
| c_2 | d_{12} | d_5 | d_{14} | d_3 |
| ... | d_{11} | d_{16} | d_1 | d_6 |
| c_i | d_4 | d_{15} | d_8 | d_{13} |
| | m_1 | m_2 | ... | m_j |

Decisions d_{ij} and other data are collected in random order to wash out viewing-order influences

The MRMC Design



Process is repeated for each reader r_k , with a different random case ordering for each

MRMC ROC Software

- DBM MRMC—University of Iowa
 - Windows application, ready-to-run
 - SAS program for sample size estimation
- OBUMRM—Cleveland Clinic Foundation
 - FORTRAN program
 - Must be compiled to use
- Both packages freely available

Decision Latency

- t -tests and ANOVAs most accessible
- Repeated measures ANOVA takes correlation patterns into account
- Also provides better accounting for sources of variance
- Does not handle missing data very well

Mixed Models

- Type of generalized linear model which can encompass repeated measures ANOVAs
- Also takes correlations into account
- Factors can be “fixed” or “random”
- More efficient use of experimental data
- Much more robust to missing data

Mixed Models

- MRMC design translates into fully-crossed mixed model
- Latency modeled by fixed modality factor and random reader and case factors
- P -values of modality slopes are tests of whether modalities differ by latency
- Can more easily investigate other factors
- MRMC ROC analysis actually a form of mixed modeling

Mixed Model Commands

R and S-Plus

`lme ()`

SAS

`proc mixed`

SPSS

`mixed`

Stata

`xtmixed`

Lung Transplant Home Monitoring Program

- Created by the University of Minnesota and Fairview-University Transplant Center
- Patients use a portable electronic spirometer to record pulmonary and symptom information
- Data uploaded and triaged weekly

Tabular Modality

| Lastname, Firstname Male, Age XX (as of 24 Jun 2000) Single-lung Tx for COPD | | | | | MID: YYY Transplanted XX Jan XXXX XXX days since transplant (as of 24 Jun 2000) | | | | | | |
|--|-----------------|------------------|----------------------|--------------------------|---|------------------------|----------|----------|------------------|-----------------|--|
| Date | FVC (Liters) | FEV1 (Liters) | PEFR (Liters/sec) | FEF25_75 (Liters/sec) | Temperature (Degrees F) | Shortness of Breath | Coughing | Wheezing | Sputum Amount | Sputum Color | |
| 06/24/00 | 4.29 | 1.96 | 4.21 | 0.70 | 97.2 | None | 2-3x/hr | 1-2x/day | Moderate | White | |
| 06/23/00 | | | | | | | | | | | |
| 06/22/00 | 5.18 | 2.02 | 4.38 | 0.75 | 97.1 | Mild | 2-3x/hr | 1-2x/day | Moderate | White | |
| 06/21/00 | 5.11 | | | | | | | | | | |
| 06/20/00 | 4.89 | | | | | | | | | | |
| 06/19/00 | 5.41 | | | | | | | | | | |
| 06/18/00 | 4.78 | | | | | | | | | | |
| 06/17/00 | 5.04 | | | | | | | | | | |
| 06/16/00 | 5.22 | | | | | | | | | | |
| 06/15/00 | 4.96 | | | | | | | | | | |
| 06/14/00 | 5.00 | | | | | | | | | | |
| 06/13/00 | 4.63 | | | | | | | | | | |
| 06/12/00 | 5.41 | | | | | | | | | | |
| 06/11/00 | 5.41 | | | | | | | | | | |
| 06/10/00 | | | | | | | | | | | |
| 06/09/00 | | | | | | | | | | | |
| 06/08/00 | 4.89 | | | | | | | | | | |
| 06/07/00 | 4.89 | | | | | | | | | | |
| 06/06/00 | 5.11 | | | | | | | | | | |
| 06/05/00 | 5.22 | | | | | | | | | | |
| 06/04/00 | 5.22 | | | | | | | | | | |
| 06/03/00 | 5.51 | | | | | | | | | | |
| 06/02/00 | 5.22 | | | | | | | | | | |
| 06/01/00 | 5.91 | | | | | | | | | | |
| 05/30/00 | 4.71 | | | | | | | | | | |
| 05/29/00 | | | | | | | | | | | |
| 05/28/00 | 5.31 | | | | | | | | | | |
| 05/27/00 | 5.01 | | | | | | | | | | |
| 05/26/00 | 4.41 | | | | | | | | | | |
| 05/25/00 | | | | | | | | | | | |
| 05/24/00 | | | | | | | | | | | |
| 05/23/00 | | | | | | | | | | | |
| 05/22/00 | 4.63 | | | | | | | | | | |
| 05/21/00 | 4.71 | | | | | | | | | | |
| 05/20/00 | 5.21 | | | | | | | | | | |
| 05/19/00 | 5.21 | | | | | | | | | | |
| 05/18/00 | 5.61 | | | | | | | | | | |
| 05/17/00 | 5.21 | | | | | | | | | | |
| 05/16/00 | 5.21 | | | | | | | | | | |
| 05/15/00 | 5.31 | | | | | | | | | | |
| 05/14/00 | 5.07 | 2.87 | 5.14 | 1.13 | 96.7 | None | EFH | 1-2x/day | Small | Watery | |
| 05/13/00 | 5.28 | 2.90 | 5.25 | 1.01 | 96.8 | None | EFH | EFH | Small | Watery | |
| 05/12/00 | 5.15 | 2.91 | 5.53 | 0.99 | 96.6 | Mild | EFH | EFH | Small | Watery | |
| 05/11/00 | 4.69 | 2.56 | 4.96 | 0.79 | 96.6 | None | EFH | EFH | Small | Watery | |
| 05/10/00 | 4.83 | 2.65 | 5.01 | 0.88 | 96.6 | Mild | EFH | EFH | Small | Watery | |
| 05/09/00 | 4.68 | 2.48 | 4.71 | 0.69 | 96.5 | Mild | EFH | EFH | Small | Watery | |
| 05/08/00 | 5.13 | 2.85 | 5.17 | 1.04 | 96.9 | None | EFH | EFH | Small | Watery | |
| 05/07/00 | 4.88 | 2.74 | 5.07 | 1.05 | 96.6 | Mild | EFH | EFH | Small | Watery | |
| 05/06/00 | 4.99 | 2.79 | 5.08 | 0.96 | 96.8 | Mild | EFH | EFH | Small | Watery | |
| 05/05/00 | 5.01 | 2.78 | 5.29 | 0.91 | 97.9 | Mild | EFH | EFH | Small | Watery | |
| 05/04/00 | | | | | | | | | | | |
| 05/03/00 | 5.17 | 2.88 | 5.46 | 0.99 | 96.4 | Mild | EFH | EFH | Small | Watery | |
| 05/02/00 | 5.01 | 2.84 | 5.15 | 0.95 | 96.4 | Mild | EFH | EFH | Small | Watery | |
| 05/01/00 | 4.80 | 2.74 | 5.22 | 0.98 | 96.5 | Mild | EFH | EFH | Small | Watery | |
| 04/29/00 | 5.01 | 2.76 | 5.15 | 0.88 | 96.1 | Mild | EFH | EFH | Small | Watery | |

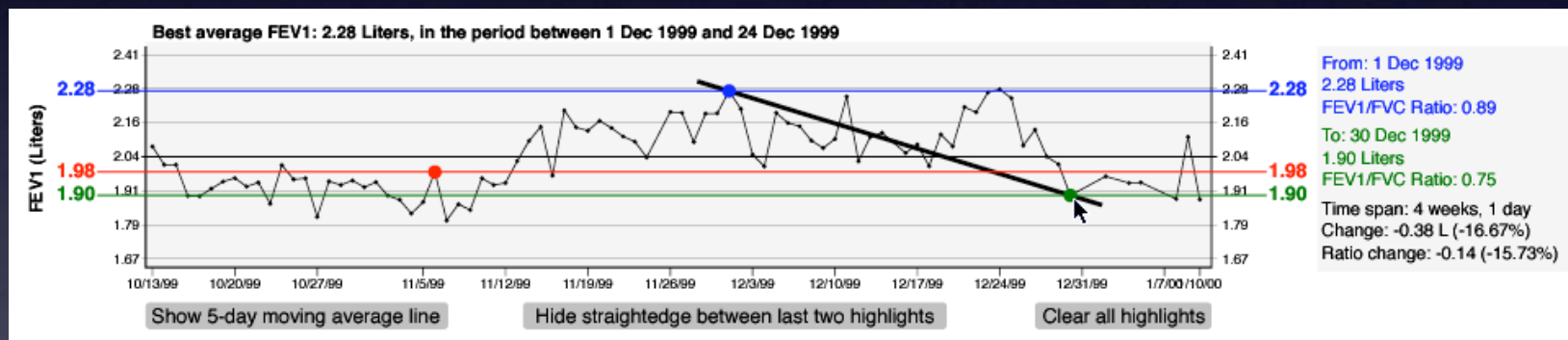
Lastname, Firstname
Male, Age XX (as of 24 Jun 2000)
Single-lung Tx for COPD

MID: YYY
Transplanted XX Jan XXXX
XXX days since transplant (as of 24 Jun 2000)

from Pieczkiewicz et al.

from Pieczkiewicz et al. (2007)

Graphical Modalities



from Pieczkiewicz *et al.* (2007)

ROC Analysis Options

Curve-Fitting

- ☒ PROPROC
- ☐ Contaminated Binormal
- ☐ Decontaminated Binormal
- ☐ One-parameter RSCORE
- ☐ Trapezoidal/Wilcoxon
- ☐ RSCORE

Analysis

- ☒ Area
- ☐ Partial area
 - ☒ bounds on FP
 - lower bound
 - ☐ bounds on TP
 - upper bound
- ☐ Sensitivity at specificity of
- ☐ Specificity at sensitivity of
- ☐ Cutpoints/Thresholds

ANOVA

The following ANOVA routines are available. You may select multiple options. You must select at least one option.

- ☒ Analysis treating both readers and cases as random samples
- ☐ Analysis treating only cases as a random sample
- ☐ Analysis treating only readers as a random sample

Other ANOVA options are available

- ☒ Display the Obuchowski-Rockette components of variance

OK Cancel

```
=====
*****      Analysis 1: Random Readers and Random Cases      *****
=====
```

(Results apply to the population of readers and cases)

a) Test for H_0 : Treatments have the same AUC

| Source | DF | Mean Square | F value | Pr > F |
|-----------|------|-------------|---------|--------|
| Treatment | 1 | 0.47140141 | 6.39 | 0.0526 |
| Error | 5.00 | 0.07372649 | | |

Error term: $MS(TR) + \max[MS(TC) - MS(TRC), 0]$

Conclusion: The treatment AUCs are not significantly different, $F(1,5) = 6.39$, $p = .0526$.

b) 95% confidence intervals for treatment differences

| Treatment | Estimate | StdErr | DF | t | Pr > t | 95% CI |
|-----------|----------|---------|------|-------|--------|--------------------|
| 1 - 2 | -0.06268 | 0.02479 | 5.00 | -2.53 | 0.0526 | -0.12639 , 0.00104 |

H_0 : the two treatments are equal.
Error term: $MS(TR) + \max[MS(TC) - MS(TRC), 0]$

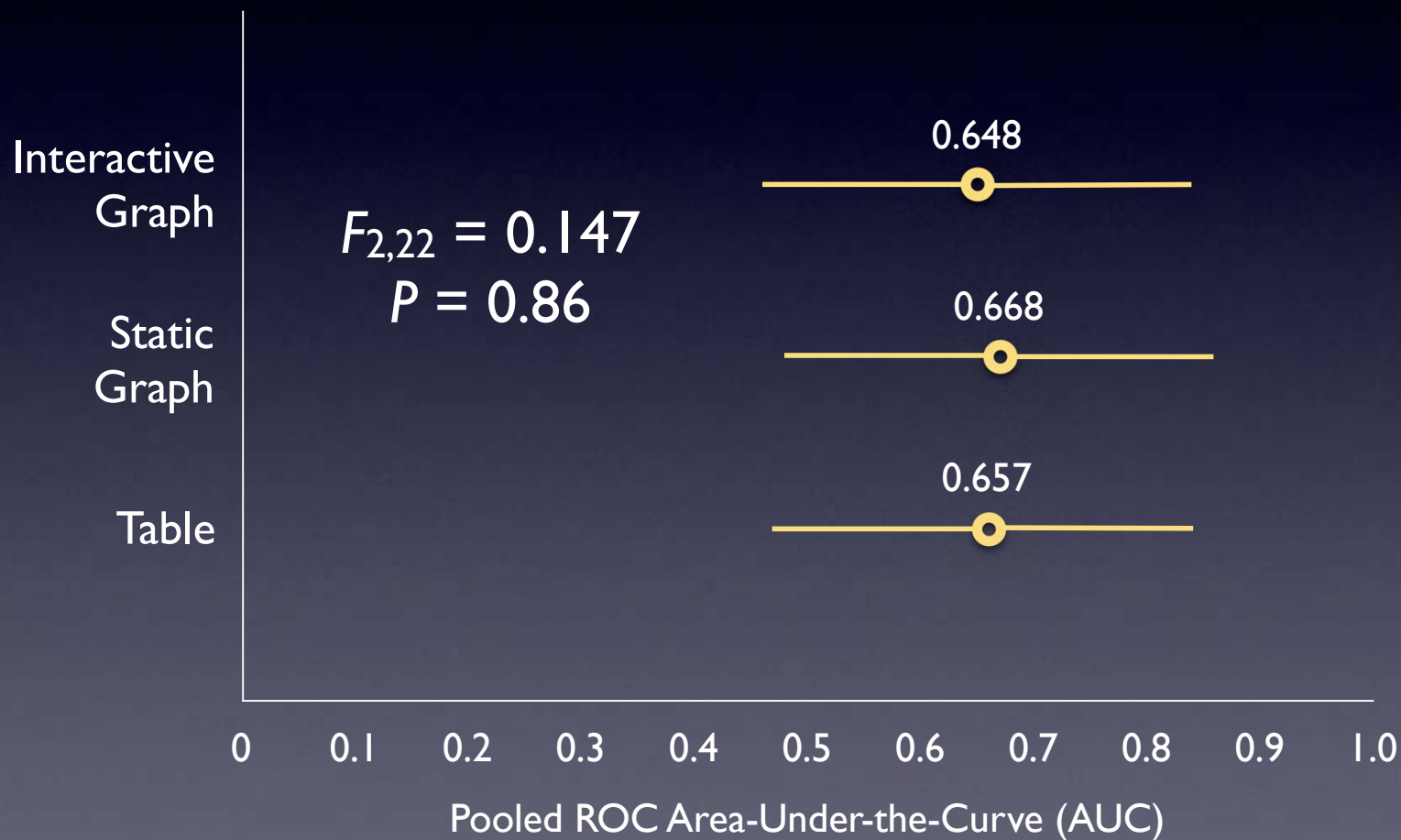
c) 95% treatment confidence intervals based on reader x case ANOVAs for each treatment (each analysis is based only on data for the specified treatment)

| Treatment | Area | Std Error | DF | 95% Confidence Interval |
|-----------|------------|------------|-------|---------------------------|
| 1 | 0.78356094 | 0.02755194 | 16.12 | (0.72518772 , 0.84193415) |
| 2 | 0.84623745 | 0.03697621 | 12.60 | (0.76609538 , 0.92637952) |

Error term: $MS(R) + \max[MS(C) - MS(RC), 0]$

Accuracy Results

$C = 20$ (10^+ / 10^-), $M = 3$, $R = 12$



```
. xi: xtmixed lntime i.modality || _all:R.case || _all:R.reader
i.modality      _Imodality_1-7      (naturally coded; _Imodality_1 omitted)
```

Performing EM optimization:

Performing gradient-based optimization:

```
Iteration 0:  log restricted-likelihood = -526.85469
Iteration 1:  log restricted-likelihood = -526.85469
```

Computing standard errors:

```
Mixed-effects REML regression          Number of obs    =      720
Group variable: _all                   Number of groups   =        1

Obs per group: min =      720
                  avg =    720.0
                  max =      720
```

```
Log restricted-likelihood = -526.85469      Wald chi2(2)      =    48.91
                                           Prob > chi2       =    0.0000
```

| | lntime | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] |
|--------------|--------|-----------|-----------|-------|-------|----------------------|
| _Imodality_6 | | -.1332807 | .0433225 | -3.08 | 0.002 | -.2181913 - .0483702 |
| _Imodality_7 | | .1689817 | .0433225 | 3.90 | 0.000 | .0840711 .2538923 |
| _cons | | 3.813324 | .153672 | 24.81 | 0.000 | 3.512132 4.114516 |

| Random-effects Parameters | Estimate | Std. Err. | [95% Conf. Interval] |
|---------------------------|----------|-----------|----------------------|
| _all: Identity | | | |
| sd(R.case) | .1280731 | .0287307 | .0825102 .1987962 |
| _all: Identity | | | |
| sd(R.reader) | .5121313 | .1107496 | .3352023 .7824484 |
| sd(Residual) | .4745745 | .012803 | .450133 .5003431 |

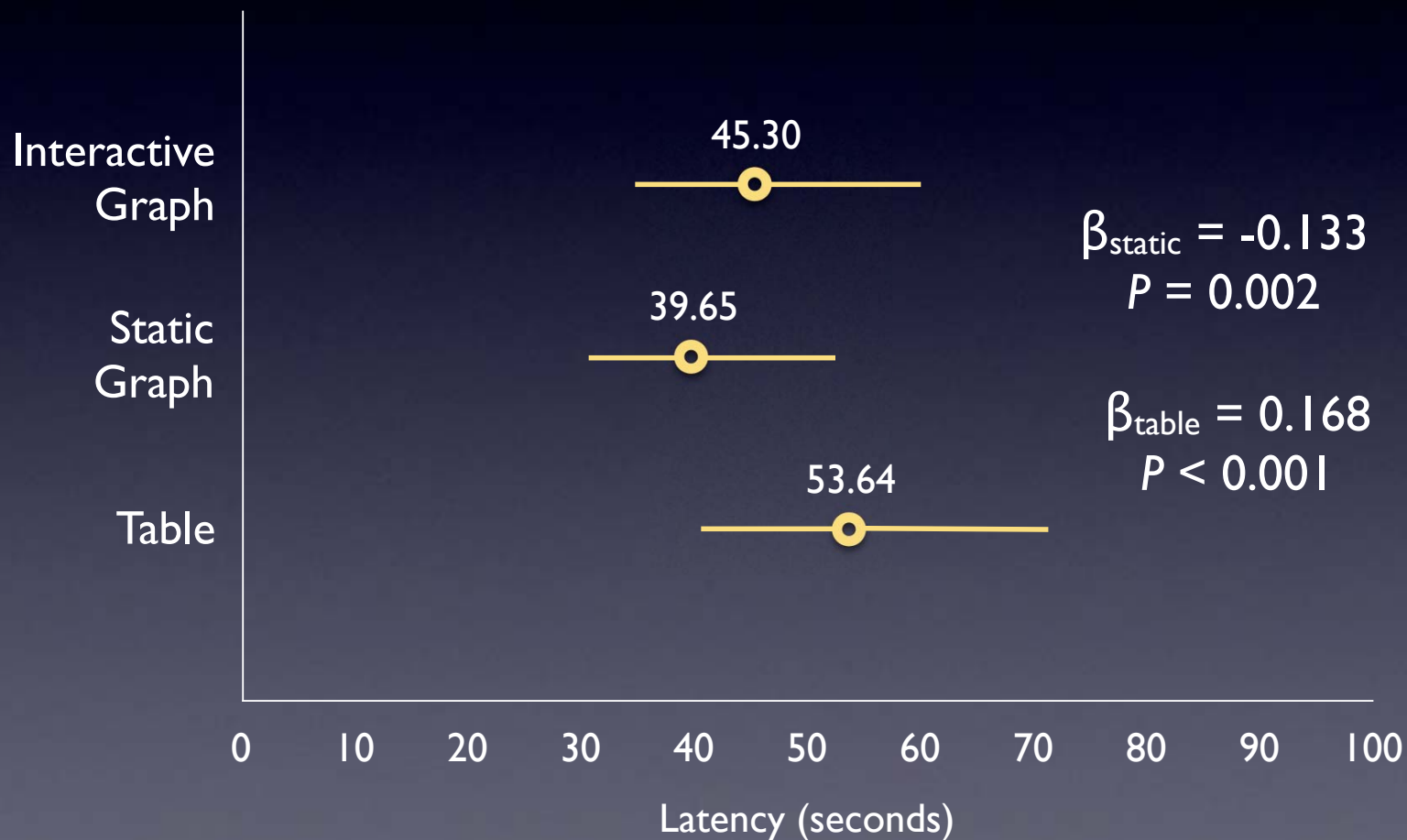
```
LR test vs. linear regression:      chi2(2) =   474.66   Prob > chi2 = 0.0000
```

Note: LR test is conservative and provided only for reference.

Stata 10.0

Latency Results

$C = 20$ (10^+ / 10^-), $M = 3$, $R = 12$

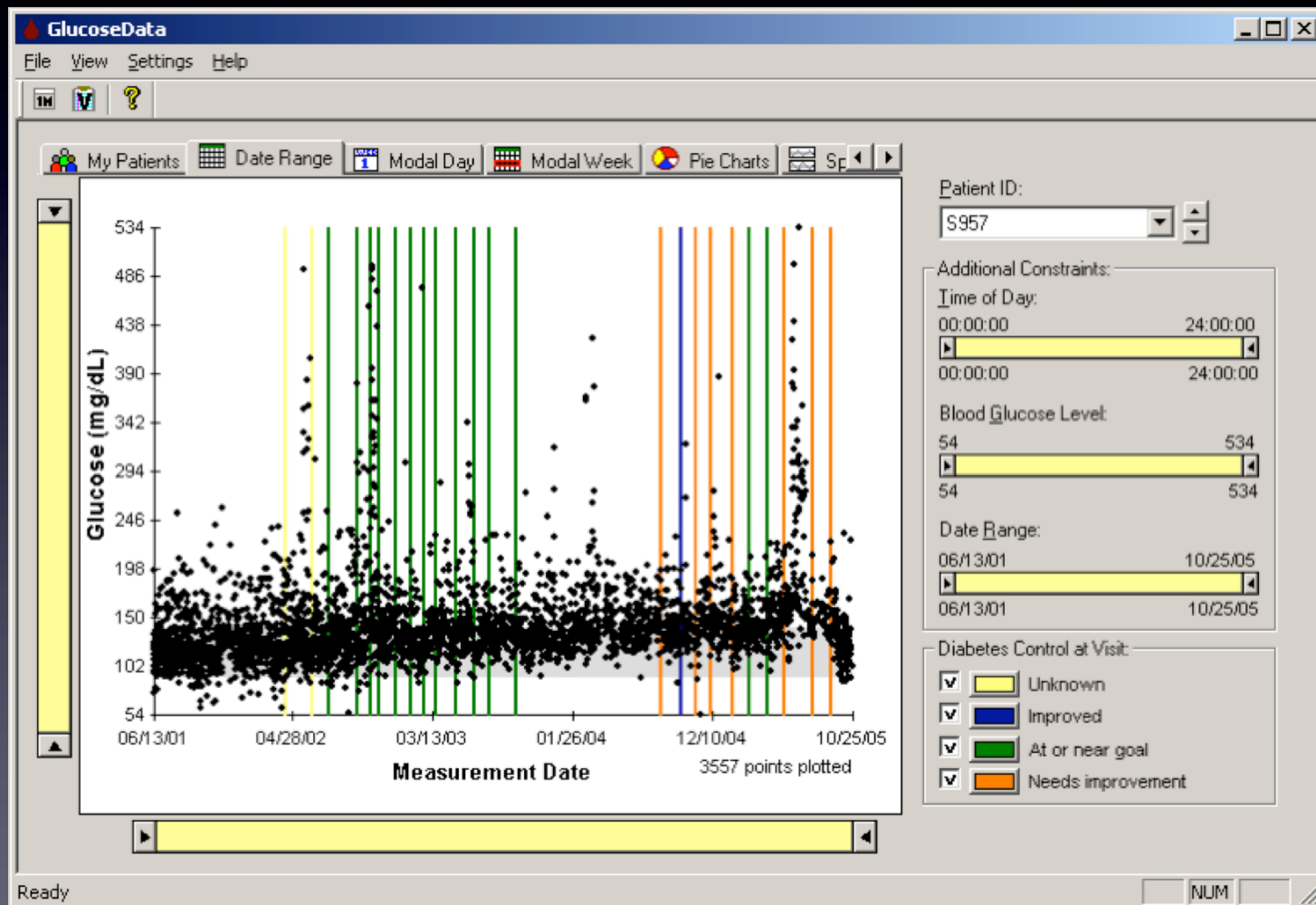


Preference Results

| Modality | Average Rank |
|-------------------|--------------|
| Interactive Graph | 1.1 |
| Static Graph | 2.2 |
| Table | 2.8 |

($R = 12$ readers)

Glucose Data Viewer



Disadvantages

- Methods not as “easy” as traditional ones
- Sample size requirements can be unclear
- MRMC ROC software takes skill to use
- Mixed models more computationally-intensive, and possibly nonconvergent
- May not apply to some aspects of EHR evaluation and research

Conclusions

- Efficacy studies usually stop at user satisfaction and/or user preference
- Accuracy and latency can be useful, objective measures of EHR efficacy
- ROC methodologies can be applied to measure decision accuracy in EHRs
- Mixed models can be used to assess latency
- Software now readily available for these purposes

Acknowledgments

- Stan Finkelstein, PhD
- Marshall Hertz, MD
- Justin Starren, MD, PhD
- Luke Rasmussen
- Kevin Berbaum, PhD
- Kevin Schartz, PhD
- Nancy Obuchowski, PhD
- Computation and Informatics in Biology and Medicine Program, University of Wisconsin
- National Institute of Biomedical Imaging and Bioengineering, NIH
- National Library of Medicine