# Comparing the accuracy of prediction methods

Michael W. Kattan, Ph.D.

Associate Attending Outcomes Research Scientist

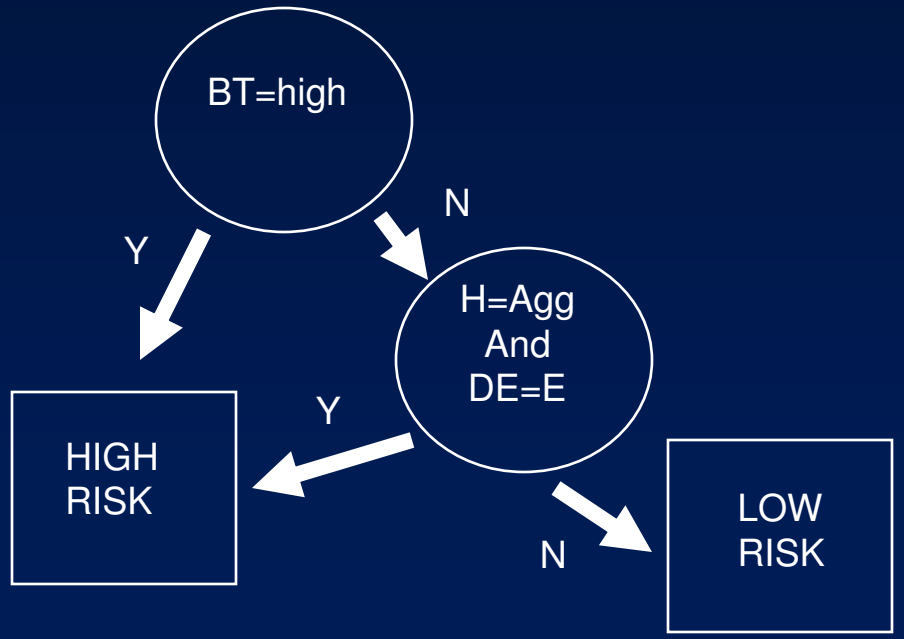
Memorial Sloan-Kettering Cancer Center

Associate Professor of Public Health

Cornell University

# How is risk typically computed?

- Based on features, we make a crude tree.
- Most cancer staging systems do this.



# The problem with crude trees

- They are very easy to use.
- But they do not predict outcome optimally.
  - » High risk groups are very heterogeneous.
    - – A single risk factor may qualify a patient as high risk.
- Other approaches, like a Cox regression model, predict more accurately.

# Some simple steps that will make a difference

1. Build the most accurate model possible.
2. Take model to bedside
   - As a nomogram,
   - In stand-alone software (desktop, handheld, web)
   - Built into the electronic medical record
- Doing this will predict patient outcome more accurately, resulting in
  - better patient counseling
  - better treatment decision making

# Desirable characteristics of an error measure

- Understandable/interpretable
- Sensitive to model improvement
- Model-free
- Unaffected by censoring

# CONCORDANCE INDEX (censored data)

- probability that, given two randomly drawn patients, the patient who fails first had a higher probability of failure.
    - assumes that the patient with the shorter follow-up fails
    - does not apply if both patients fail at the same time, or the censored patient has shorter follow-up.

$$\frac{\text{Usable patient pairs with consistent outcome}}{\text{Usable patient pairs}}$$

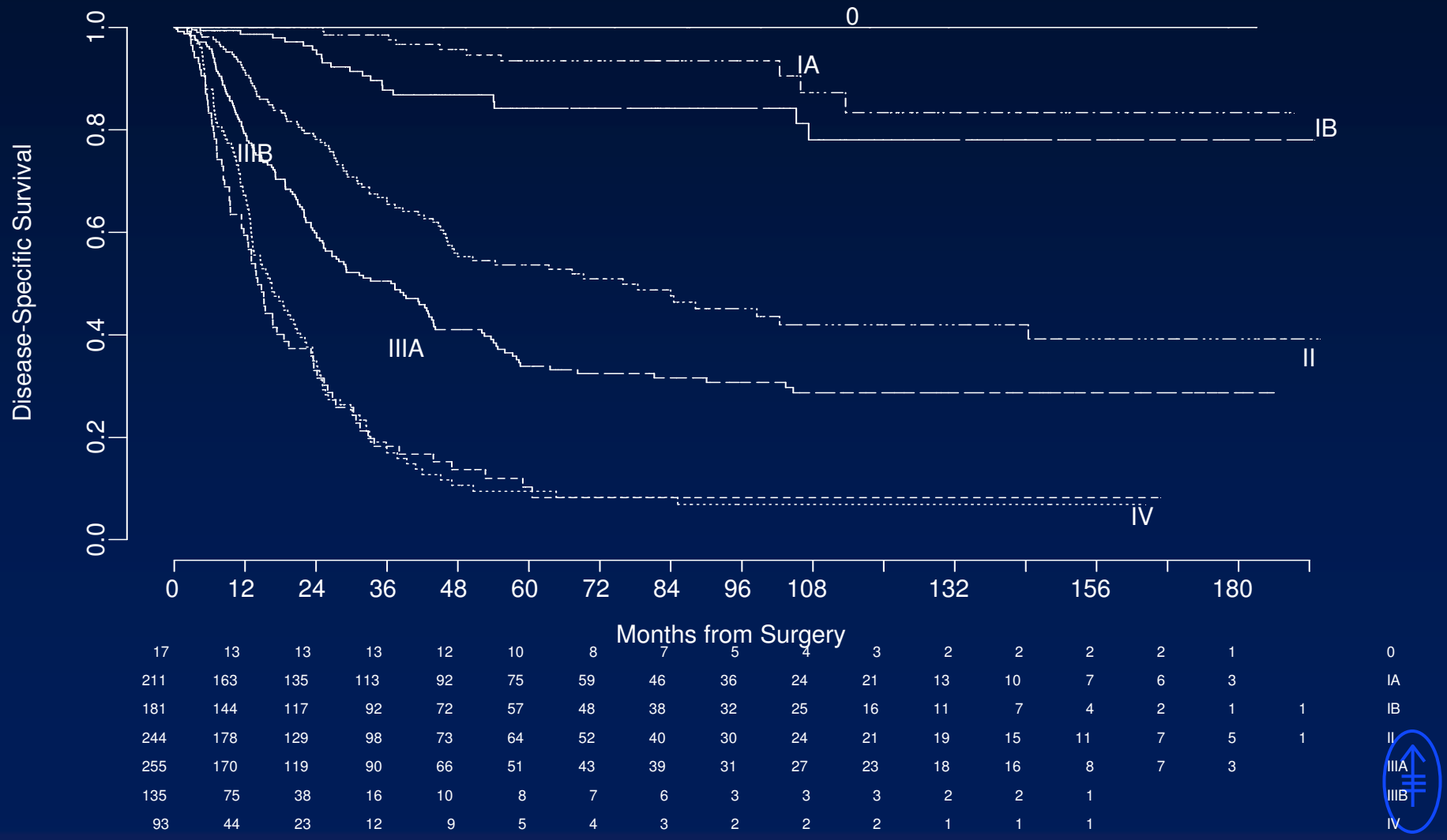
usable patient pair - patient with the shorter follow-up must fail

consistent outcome - failure more likely for the shorter follow-up patient

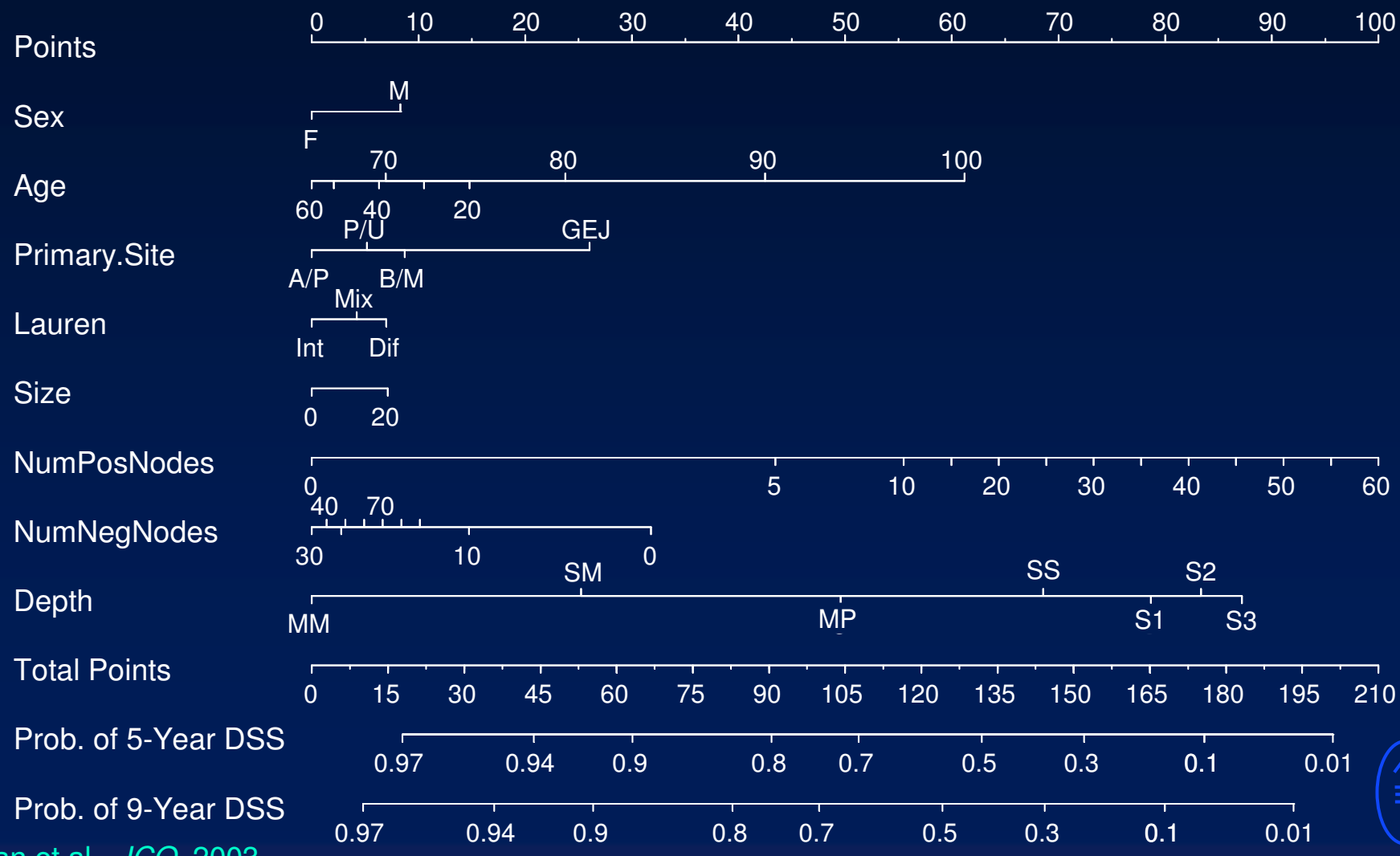
tied predicted probabilities get 1/2

(Harrell, 1982)

# Gastric Cancer Disease-Specific Survival by AJCC Stage



# Gastric Cancer Disease-Specific Survival Nomogram




Kattan et al., *JCO*, 2003

# How to tell if we are doing any better than existing models?

Compare jackknife predicted probabilities of new model to existing model predictions:

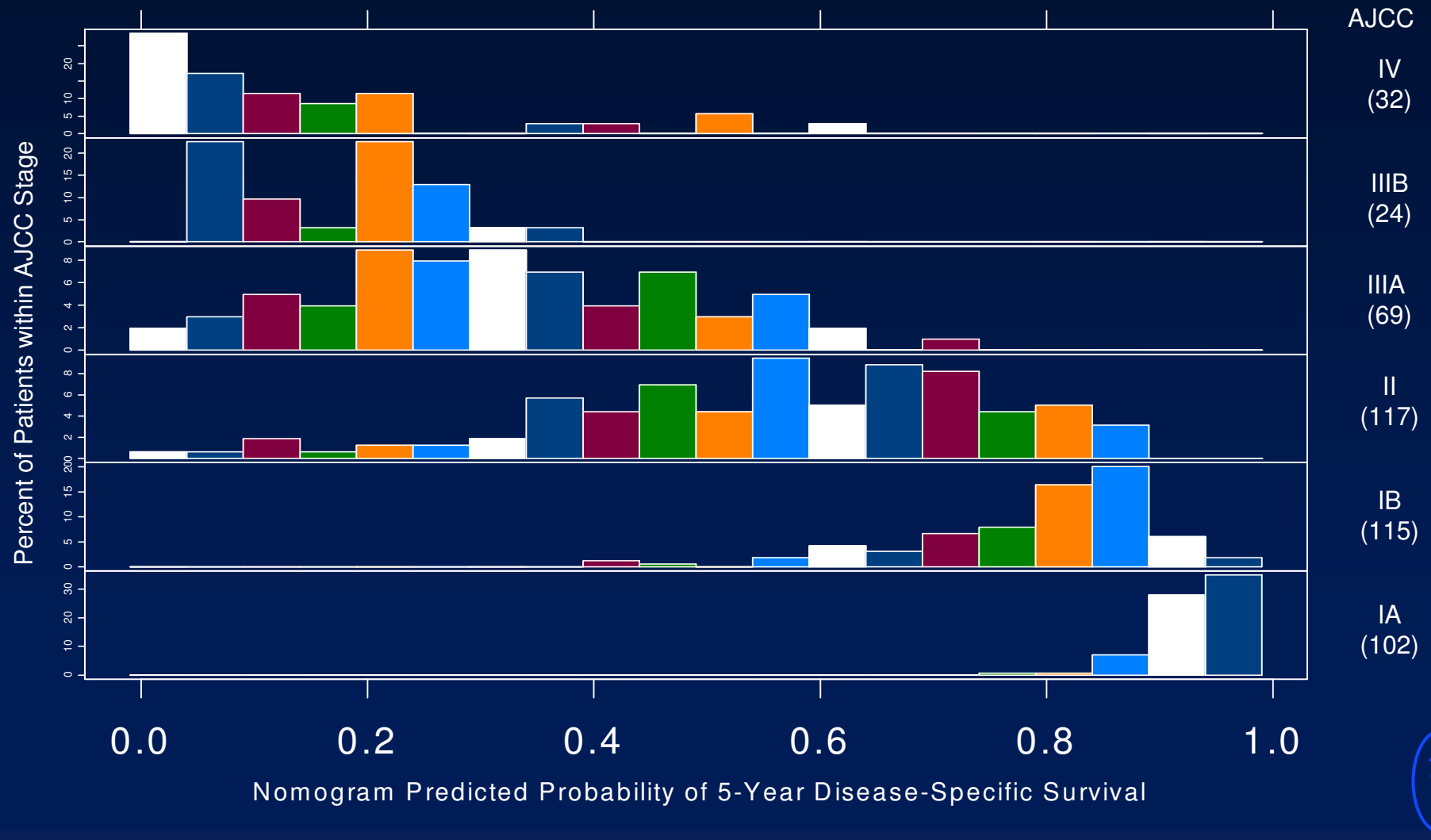
| Method | Concordance Index |
|---|---|
| AJCC Stage | 0.77 |
| Nomogram (jackknife) | 0.80 |

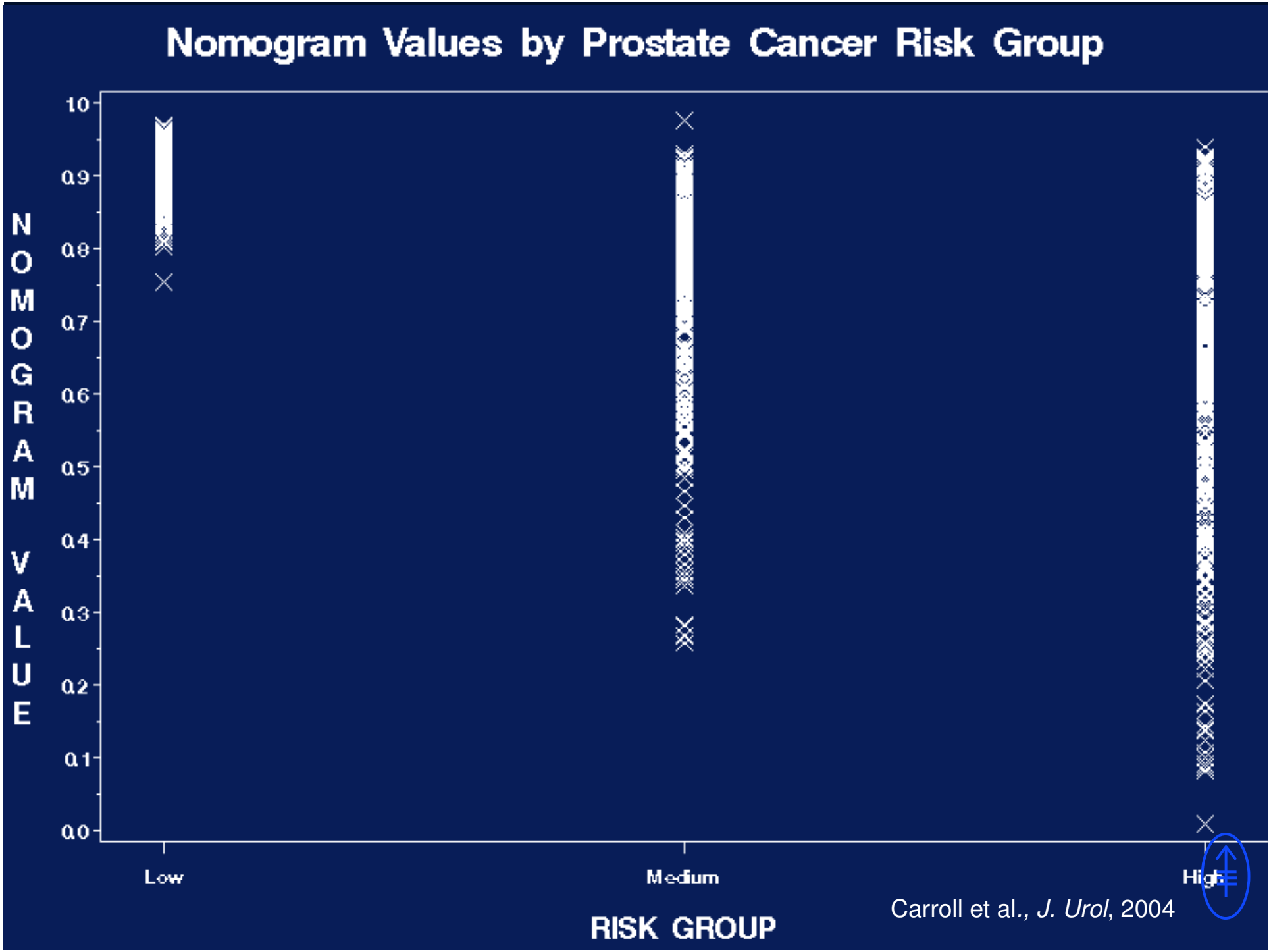
($p<0.001$).

# How to tell if we are doing any better than existing models? Validation dataset

| | Concordance Index | |
|---|---|---|
| Method | Original | Dutch Trial (n=459) |
| AJCC Stage | 0.77 | 0.75 |
| Nomogram | 0.80 | 0.77 |
| | (p<0.001) | (p<0.001) |

# Heterogeneity within stages

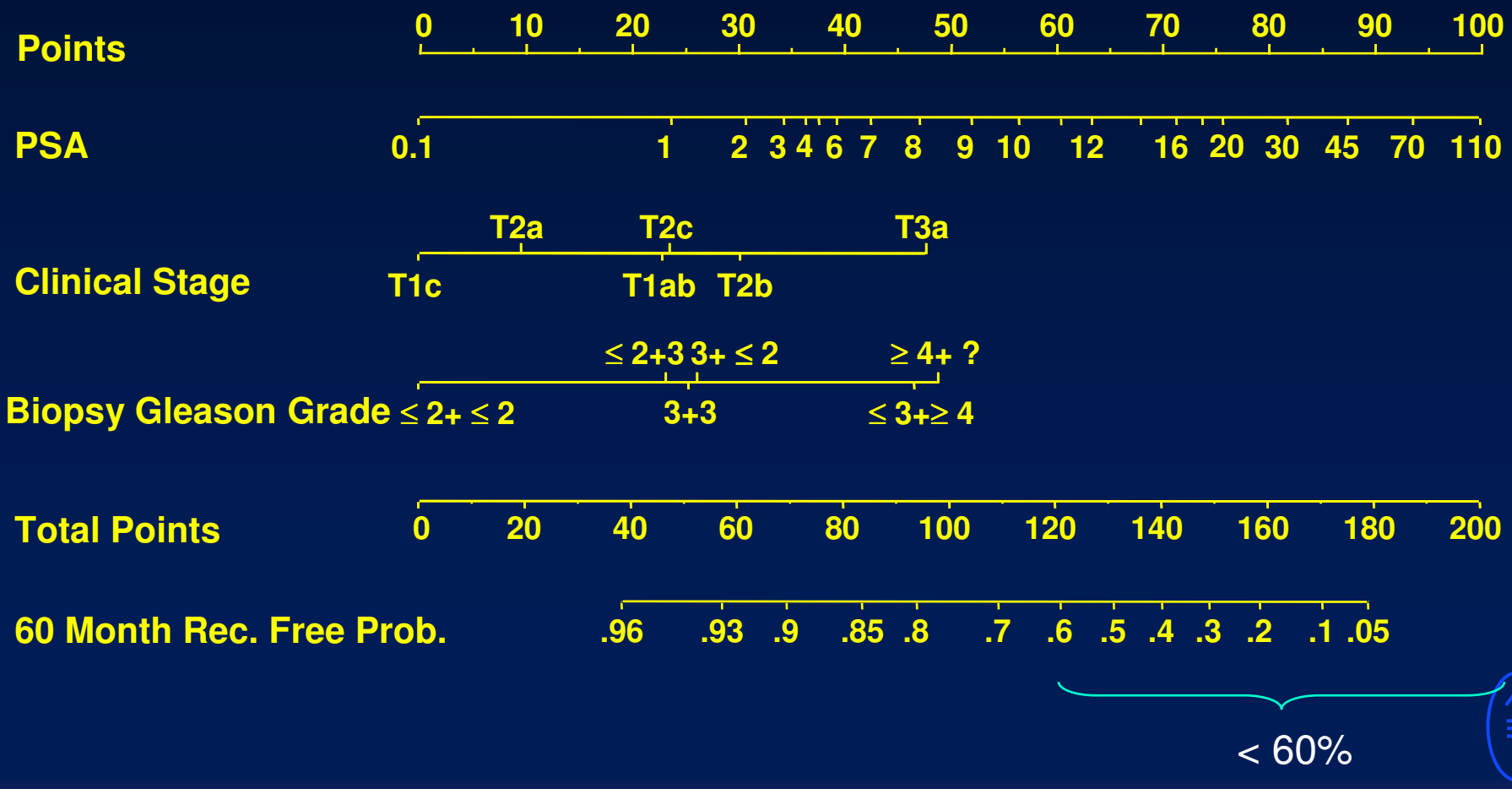





Carroll et al., *J. Urol*, 2004

# Nomograms for clinical trial design

- Example: CALGB 90203, preoperative therapy for patients at high risk of failure following surgery for prostate cancer

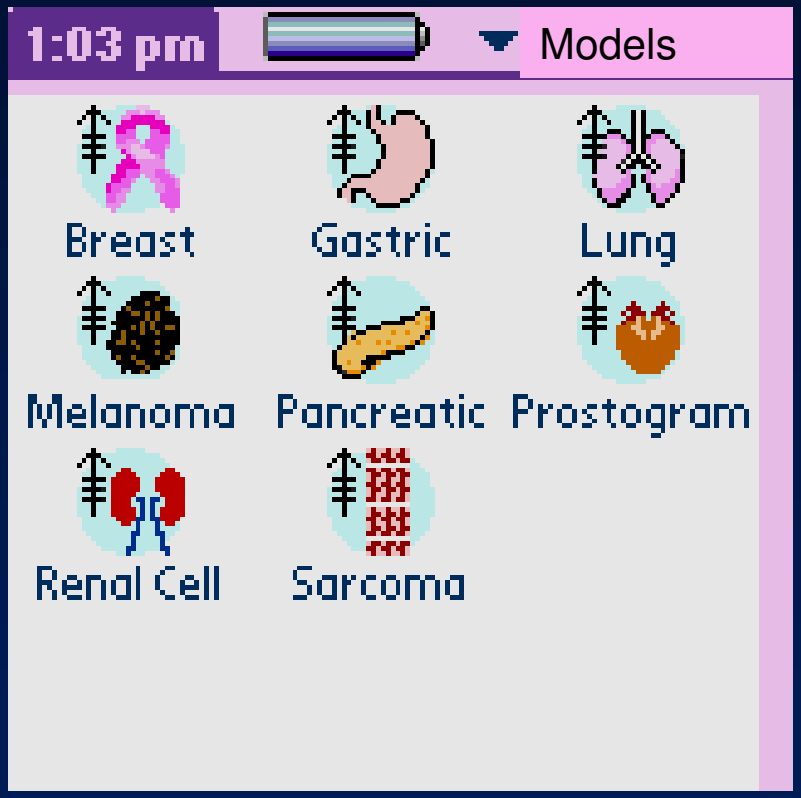


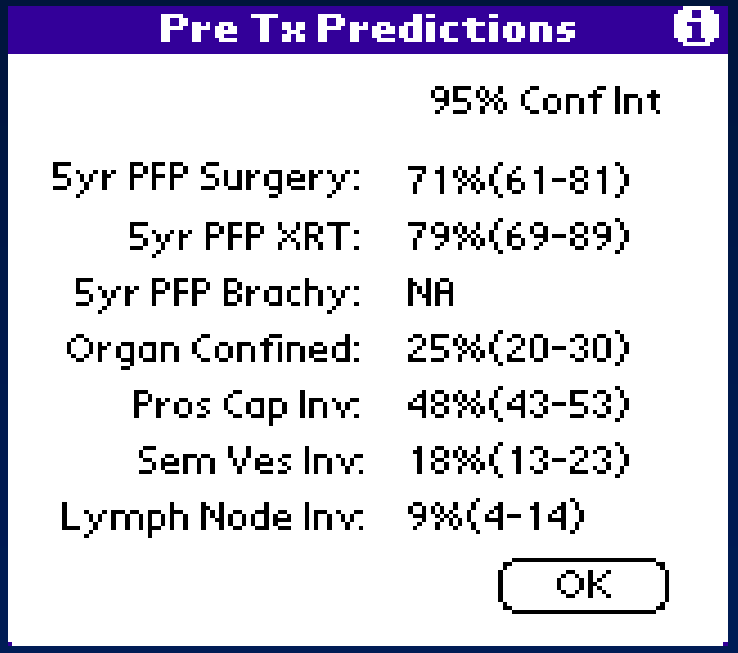
# Continuous Models vs. Staging/Grouping Systems

| Model | Comparator | CI (M vs C) |
|---|---|---|
| Preop | L/I/H Risk Groups | 0.67 vs. 0.64 |
| Preop + IL6/TGFβ1 | L/H Risk Groups | 0.84 vs. 0.73 |
| Pre XRT | L/I/H Risk Groups | 0.76 vs. 0.69 |
| Melanoma SLN+ | AJCC Stage | 0.69 vs. 0.66 |
| Pancreatic Ca | AJCC Stage | 0.64 vs. 0.56 |
| Gastric Ca | AJCC Stage | 0.77 vs. 0.75 |
| Breast Ca | NPI Groups | 0.69 vs. 0.64 |
| Sarcoma | CART Groups | 0.77 vs. 0.74 |

# Software to facilitate real-time predictions

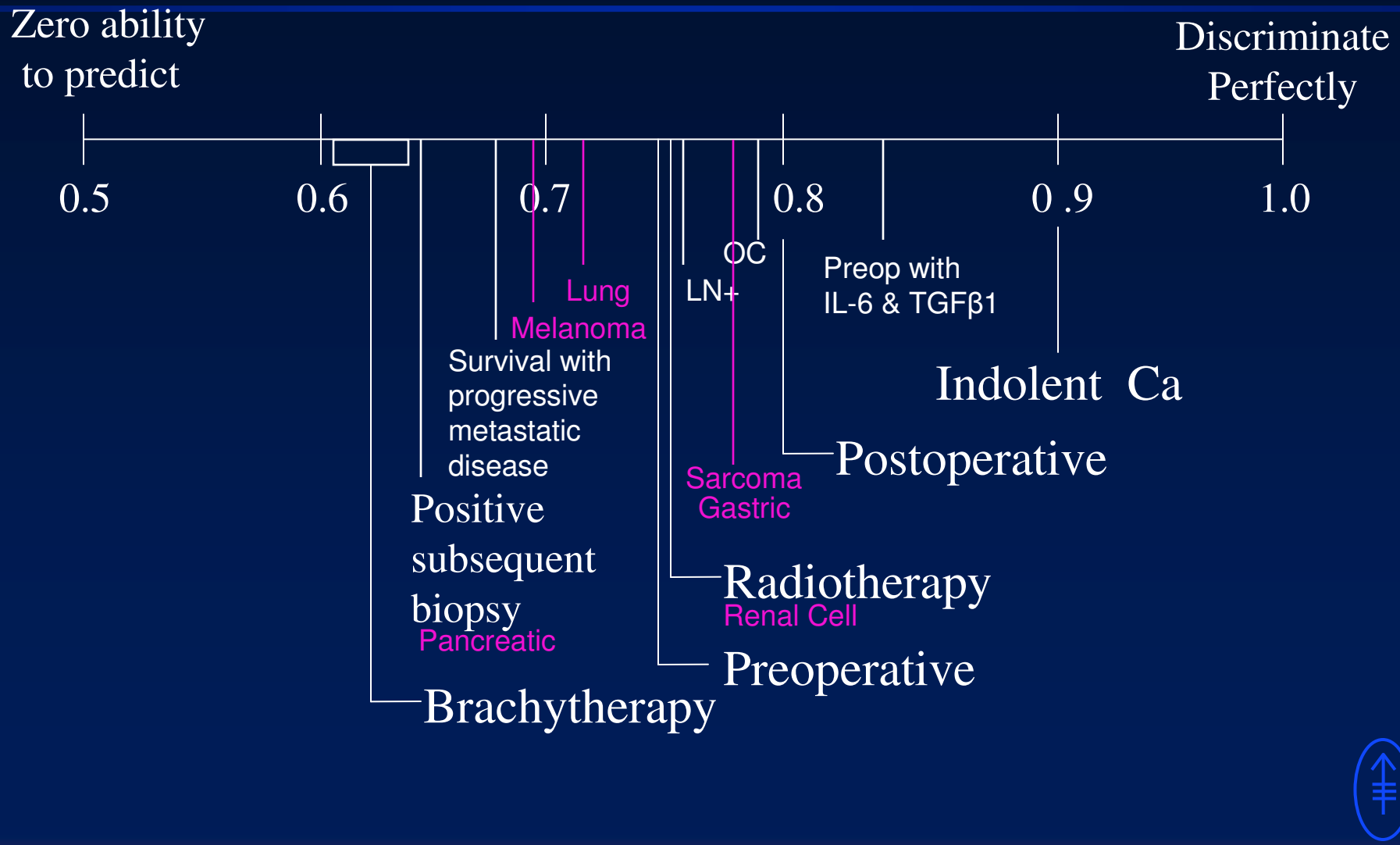
## Software for the Palm Pilot, PocketPC, and Windows Desktop Computers




- Software is free from http://www.mskcc.org/predictiontools
- Prostate, renal cell, gastric, sarcoma, breast, lung available now.
- Pancreatic, melanoma available soon.

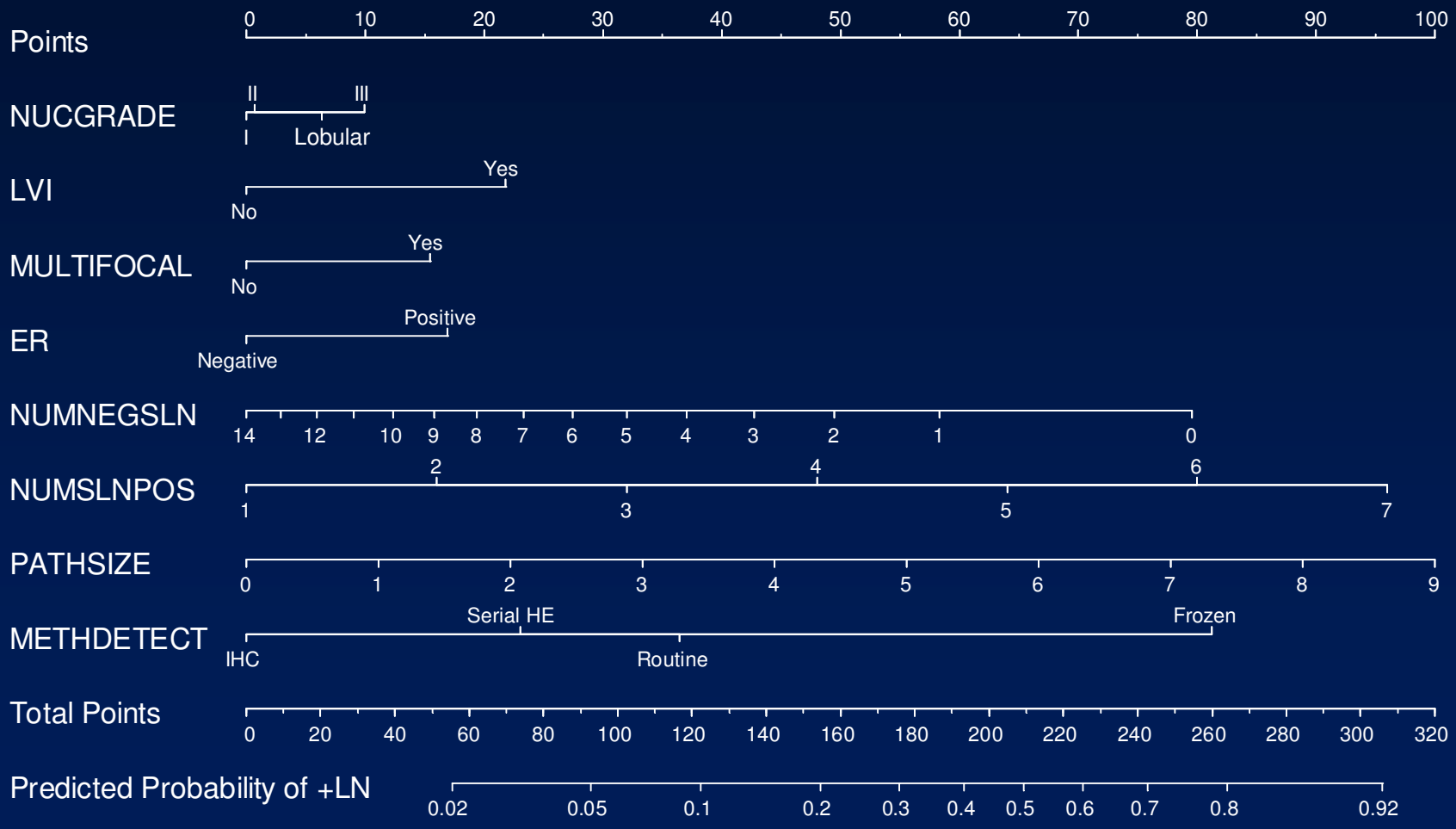
# Levels of discrimination for some prediction tools



# When The Patient Wants A Prediction, What Options Does The Clinician Have?

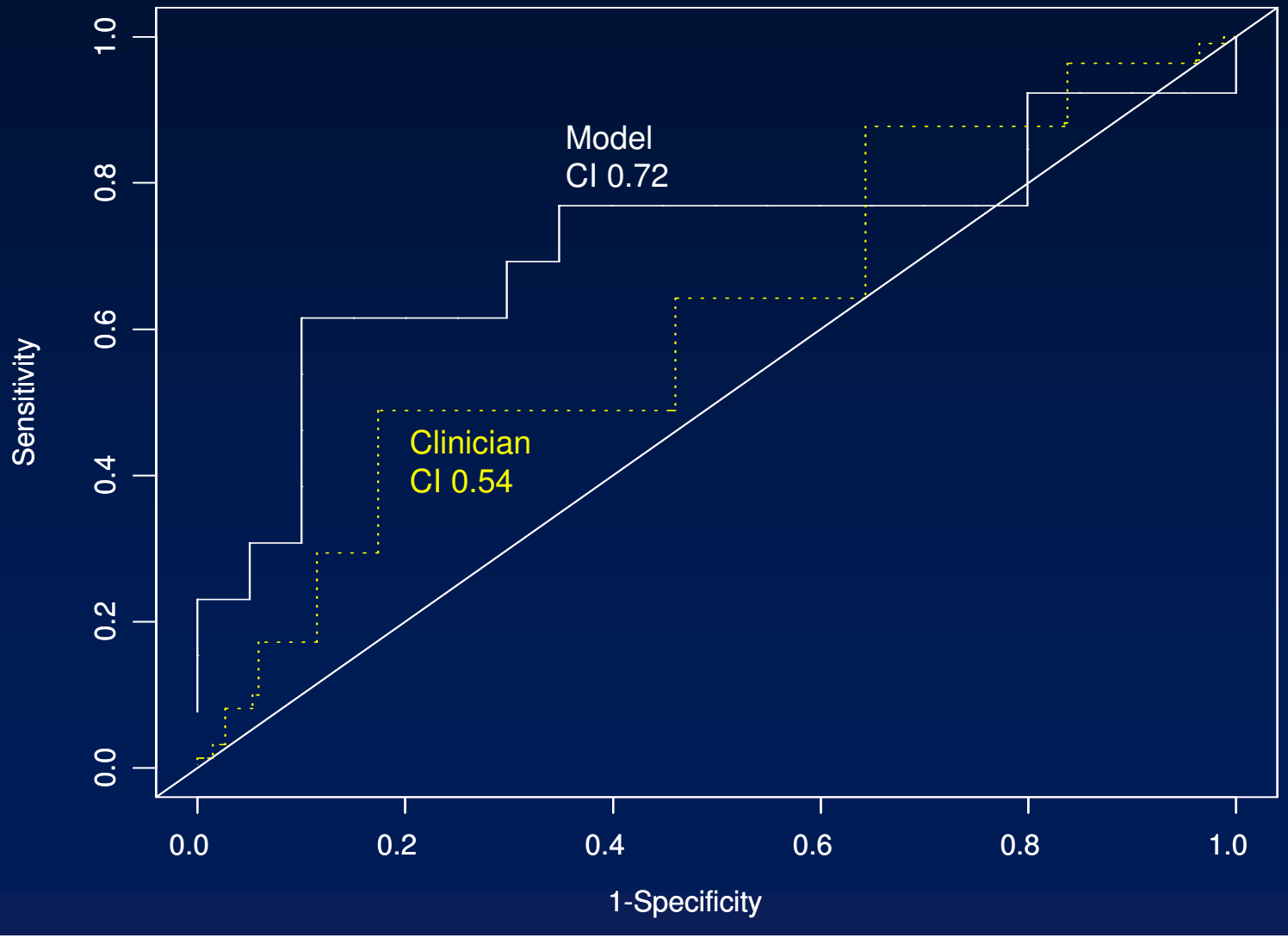
- Predict based on knowledge and experience
- Deny ability to predict at the individual patient level
- Quote an overall average to all patients
- Assign the patient to a risk group, i.e. high, intermediate, or low
- Apply a model

# Nomogram for predicting the likelihood of additional nodal metastases in breast cancer patients with a positive sentinel node biopsy




Vanzee K, et al., *Ann Surg Oncol.*, 2003.

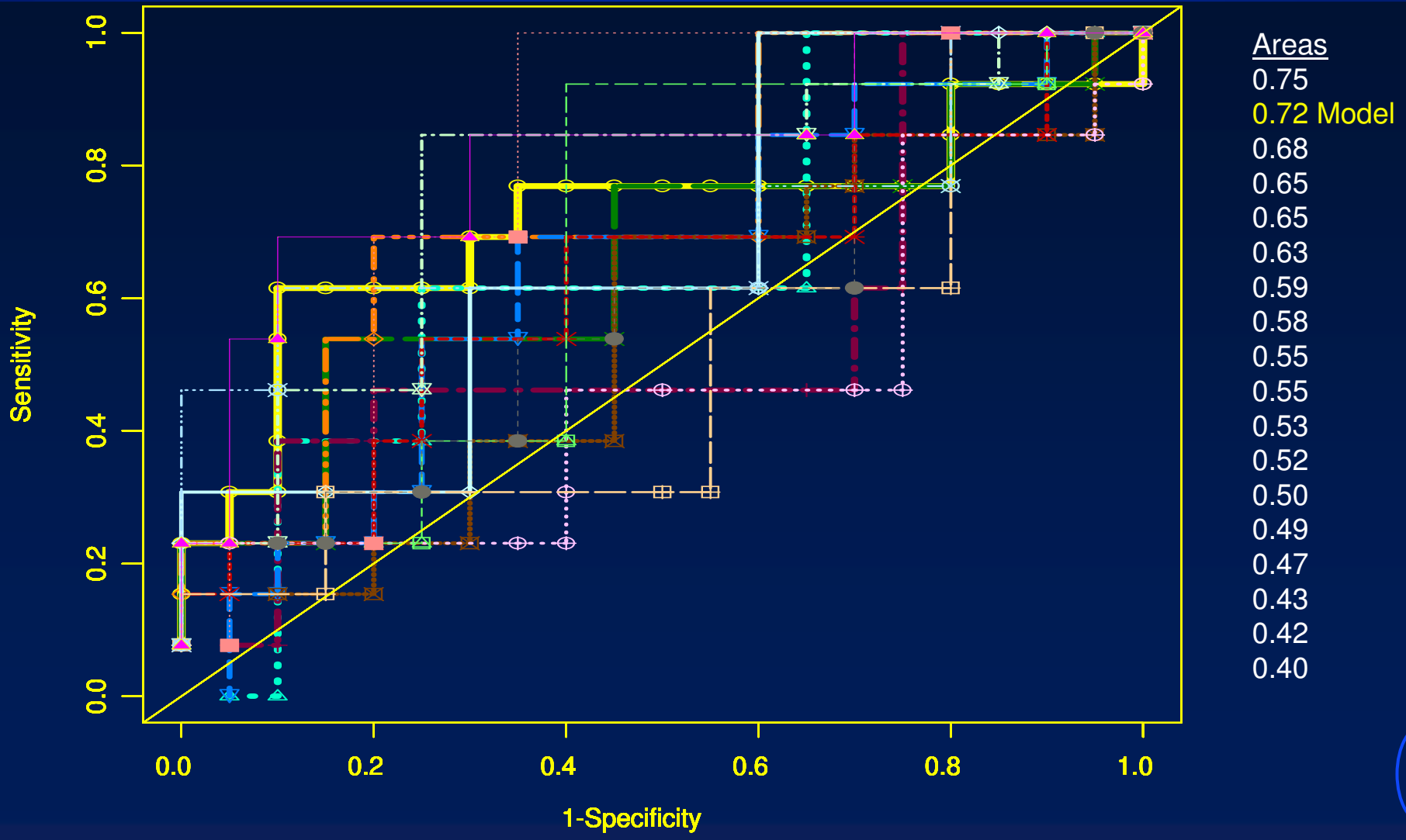
# Breast Cancer Prediction: 17 Clinicians vs. Model on 33 Patients




Sensitivity: Proportion of women with positive nodes predicted to have positive nodes

Specificity: Proportion of women with negative nodes predicted to have negative nodes

# ROC Curves
# Individual Clinicians and Model



# Conclusions

- Concordance index is a useful metric by which to compare rival prediction models.
- The decision whether to use any model vs. assume homogeneous risk is context dependent.

# Collaborators

- Methods
  - Biostatistics
    - Mithat Gonen
    - Glenn Heller
    - Peter Bach
    - Colin Begg
    - Frank Harrell
  - Informatics
    - Paul Fearn
    - David Ladanyi
    - John Davey
    - Pat Turi
    - Jacob Rockowitz
    - Drumbeat Digital
- Applications
  - Peter Scardino
  - Murray Brennan
  - Marty Karpeh
  - Kim VanZee
  - Dan Coit