



Impact in Clinical/ Translational Studies

Frank E Harrell Jr

Department of Biostatistics
Vanderbilt University School of Medicine

DESIGNING **TEAM** SCIENCE: A BIOSTATISTICIAN'S PERSPECTIVE
TRANSLATIONAL SCIENCE 2014 ACTS 11 APR 2014



- Hazards of arbitrary decisions and discarding information
- Harm from arbitrary classification of patients and avoiding probabilistic thinking
- Alternate explanations for therapeutic effects
- Limits of information: pitfalls in biomarker discovery
- Biostatistician: The Team Detective



What Does Biostatistics Offer?

Impact in
Clinical/
Translational
Studies

Overview

Thresholds
and Labeling
Outcomes

Information
Limits

Summary

References

- Experimental design—maximize power or minimize n
- Refining measurements
- Analytical design
- Avoid information loss
- Embrace probability
- Quantify information limits; demonstrate futility
- Detective work



VANDERBILT
School of Medicine
BIOSTATISTICS

Biostatistician: The Team Detective

Things are not as they seem . . .

Impact in
Clinical/
Translational
Studies

Overview

Thresholds
and Labeling
Outcomes

Information
Limits

Summary

References





Detective Work

Impact in
Clinical/
Translational
Studies

Overview

Thresholds
and Labeling

Outcomes

Information
Limits

Summary

References

- Find or rule out alternative explanations
- Show that an estimator estimates the right thing
- Check whether a result is a function of the right things
- Check whether a result is a function of the wrong things
 - improper normalization
 - change scores
 - discontinuity in disease/endpoint criteria
- Find out if an answer is arbitrary



Avoidance of Probabilistic Thinking: Arbitrary Thresholds and Labeling

Impact in
Clinical/
Translational
Studies

Overview

Thresholds
and Labeling

Outcomes

Information
Limits

Summary

References

What is the most direct cause of the enormous problem of false positives in mammography screening for breast cancer?



BI-RADS Scores: Does Category 4 Make Any Sense?

	Diagnosis	Number of Criteria
0	Incomplete	Your mammogram or ultrasound didn't give the radiologist enough information to make a clear diagnosis; follow-up imaging is necessary
1	Negative	There is nothing to comment on; routine screening recommended
2	Benign	A definite benign finding; routine screening recommended
3	Probably Benign	Findings that have a high probability of being benign ($> 98\%$); six-month short interval follow-up
4	Suspicious Abnormality	Not characteristic of breast cancer, but reasonable probability of being malignant (3 to 94%); biopsy should be considered
5	Highly Suspicious of Malignancy	Lesion that has a high probability of being malignant ($\geq 95\%$); take appropriate action
6	Known Biopsy Proven Malignancy	Lesions known to be malignant that are being imaged prior to definitive treatment; assure that treatment is completed

Breast Imaging Reporting and Data System, American College of Radiologists

<http://breastcancer.about.com/od/diagnosis/a/birads.htm>

American College of Radiology. BI-RADS US (PDF document) Copyright 2004.

Impact in
Clinical/
Translational
Studies

Overview

Thresholds
and Labeling

Outcomes

Information
Limits

Summary

References



Apparent Thresholds are Artifacts

Impact in
Clinical/
Translational
Studies

Overview

Thresholds
and Labeling

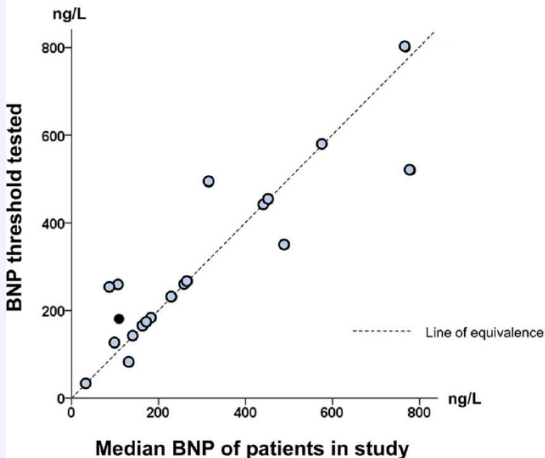
Outcomes

Information
Limits

Summary

References

BNP studies testing a single prognostic threshold, finding it to be either significant (○) or non significant (●)



In + studies:
threshold 132–800
ng/L, correlation
with study median
 $r = 0.86$.



Cutpoints are Disasters

Impact in
Clinical/
Translational
Studies

Overview

Thresholds
and Labeling

Outcomes

Information
Limits

Summary

References

Cutpoints may be found that result in both increasing and decreasing relationships with **any** dataset with zero correlation

Range of Delay	Mean Score	Range of Delay	Mean Score
0-11	210	0-3.8	220
11-20	215	3.8-8	219
21-30	217	8-113	217
31-40	218	113-170	215
41-	220	170-	210



Data from Wainer [2006]

Impact in
Clinical/
Translational
Studies

Overview

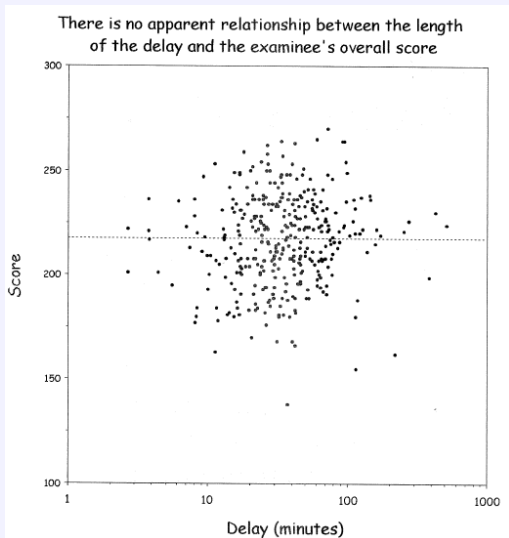
Thresholds
and Labeling

Outcomes

Information
Limits

Summary

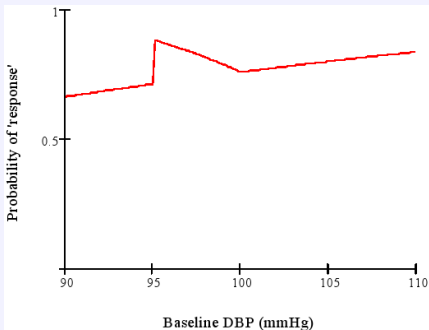
References





Arbitrary Categorization of Outcomes

- Arbitrary, low power, can be difficult to interpret
- Example: “The treatment is called successful if either the patient has gone down from a baseline diastolic blood pressure of ≥ 95 mmHg to ≤ 90 mmHg or has achieved a 10% reduction in blood pressure from baseline.”



Is a mean difference of 5.4mmHg more difficult to interpret than A:17% vs. B:22% hit clinical target?



Damage Caused by Improper Accuracy Score

- Predicting probability of an event, e.g., Prob[disease]
- $N = 400$, 0.57 of subjects have disease
- Classify as diseased if prob. > 0.5

Model	c	χ^2	Proportion Correct
age	.592	10.5	.622
sex	.589	12.4	.588
age+sex	.639	22.8	.600
constant	.500	0.0	.573

Adjusted Odds Ratios:

age (IQR 58y:42y) 1.6 (0.95CL 1.2-2.0)

sex (f:m) 0.5 (0.95CL 0.3-0.7)

Test of sex effect adjusted for age (22.8 – 10.5):

$P = 0.0005$



Alternate Explanation for Heterogeneous Tx Effect

Understanding Personalized Medicine

Impact in
Clinical/
Translational
Studies

Overview

Thresholds
and Labeling
Outcomes

Information
Limits

Summary

References

- Failure to measure treatment response on the right scale
- Failure to properly account for all main effects before examining interaction effects
- Patients with the same disease are not necessarily different
 - Randomness can dominate
- Unlike \times -over designs, parallel group designs are limited in ability to distinguish
 - small benefit to all patients vs.
 - large benefit for a few



Difficulties of Picking “Winners” : Limits of Information

Impact in
Clinical/
Translational
Studies

Overview

Thresholds
and Labeling
Outcomes

Information
Limits

Summary

References

- Multiple comparison problems
- Extremely low power; high false negative rate
- Potential markers may be correlated with each other
- Small changes in the data can change the winner
- Significance testing can be irrelevant; is a ranking and selection problem



Best of 213 Protein Biomarkers

Impact in
Clinical/
Translational
Studies

Overview

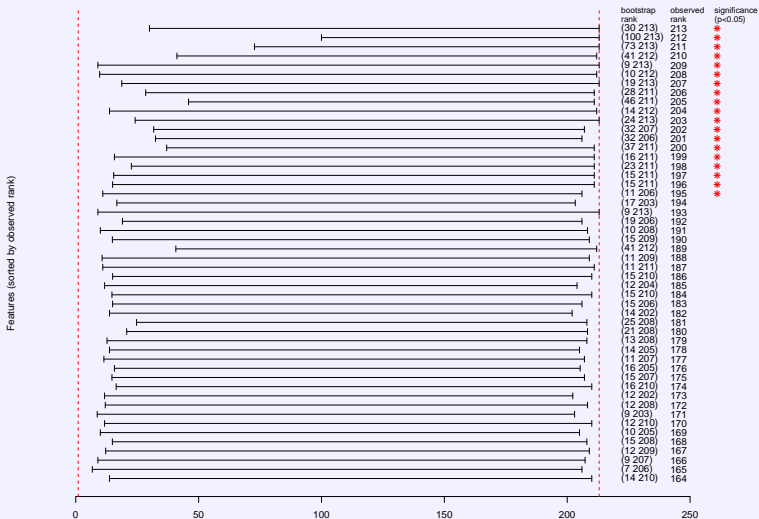
Thresholds
and Labeling

Outcomes

Information
Limits

Summary

References





Worst of 213 Biomarkers

Impact in
Clinical/
Translational
Studies

Overview

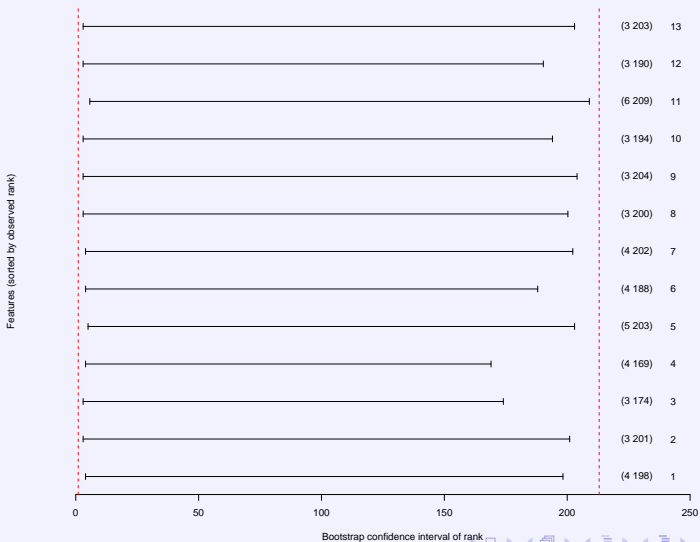
Thresholds
and Labeling

Outcomes

Information
Limits

Summary

References





Summary

Impact in
Clinical/
Translational
Studies

Overview

Thresholds
and Labeling
Outcomes

Information
Limits

Summary

References

Things are often not as they seem

Better to have a biostatistician as your detective rather than referees/study section

Statistics is not a toolbox, but a way of thinking



Statisticians have spent the past 200 years figuring out what traps lie in wait when we try to understand the world through data. The data are bigger, faster and cheaper these days—but we must not pretend that the traps have all been made safe. They have not. . . “Big data” has arrived, but big insights have not.

Tim Harford
Financial Times March 28, 2014

This work used only free software

L^AT_EX





References

- A. Giannoni, R. Baruah, T. Leong, M. B. Rehman, L. E. Pastormerlo, F. E. Harrell, A. J. Coats, and D. P. Francis. Do optimal prognostic thresholds in continuous physiological variables really exist? Analysis of origin of apparent thresholds, with systematic review for peak oxygen consumption, ejection fraction and BNP. *PLoS ONE*, 9(1):e81699, 2014.
- P. Royston, D. G. Altman, and W. Sauerbrei. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*, 25:127–141, 2006.
- S. J. Senn. Dichotomania: an obsessive compulsive disorder that is badly affecting the quality of analysis of pharmaceutical trials. In *Proceedings of the International Statistical Institute, 55th Session*, Sydney, 2005.
- H. Wainer. Finding what is not there through the unfortunate binning of results: The Mendel effect. *Chance*, 19(1):49–56, 2006.