Analysis of Cost Data

Frank Harrell

Division of Biostatistics and Epidemiology

Dept. of Health Evaluation Sciences

University of Virginia School of Medicine

DIA Economic Assessment in Clinical Trials:

Design & Analysis

Hilton Head SC

6-7 Nov 97

Outline

- Nonparametric tests for comparing cost distributions
- Use of bootstrap in place of parametric confidence limits
- Multivariable modeling of cost
- Should costs for patients who die be considered complete?
- Futility of estimating a single clinical effectiveness not to mention cost-effectiveness

Nonparametric Tests

- Nonparametric tests are robust, powerful, general, transformation invariant
- If goal is to compare overall cost distributions, use Kolmogorov-Smirnov test
- Wilcoxon test tests whether costs for treatment A > costs for treatment B
- Rank-based multivariable models: proportional odds logistic, Cox proportional hazards

Bootstrap Confidence Intervals

- Mean costs may be main focus
- No value in using *t*, normal, or log-normal distribution-based confidence intervals
- Use nonparametric bootstrap confidence intervals for population mean costs or differences in mean costs between treatments
- Bootstrap makes no distributional assumptions

Multivariable Models for Cost

- Ordinary multiple linear regression has a number of shortcomings
  - Extremely non-robust
  - Lack of normality of residuals will ruin confidence intervals
  - Unlikely to predict accurately
- What about regression on log cost?
  - Residuals unlikely to be normal
  - Assumes that patient factors act multiplicatively; more reasonable to assume additivity
  - Zero costs not allowed

Cox Model for Cost

- $\Pr[\text{Cost} > y \mid X] = C(y)^{\exp(Xb)}$
- **Form** of C(y) (1 - cost CDF) estimated from data
- Assumes proportional hazards which may be checked for each X
- Estimator of mean cost | X = area under C(y | X) curve
- No relationship assumed between mean and median cost
- Does not assume that X acts additively or multiplicatively on mean costs
- Allows zero costs

Cox Model, Continued

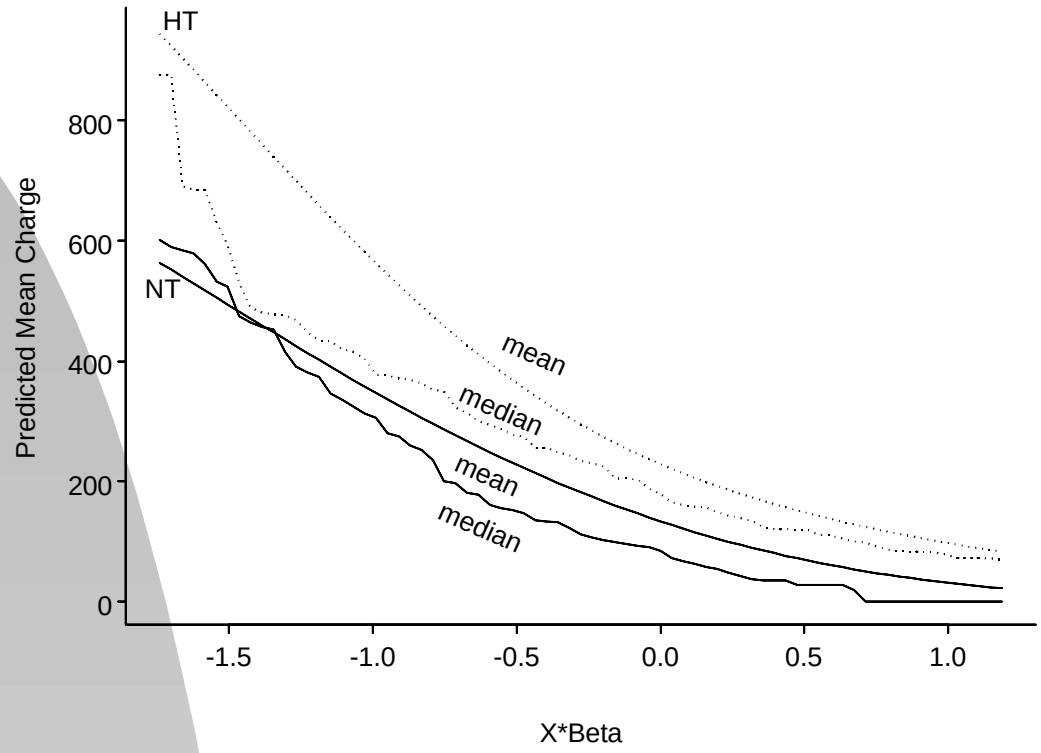
- Ref: Dudley et al, J Clin Epi 1993
- Sample size per arm for 1-sided alpha level test with power 1-beta:

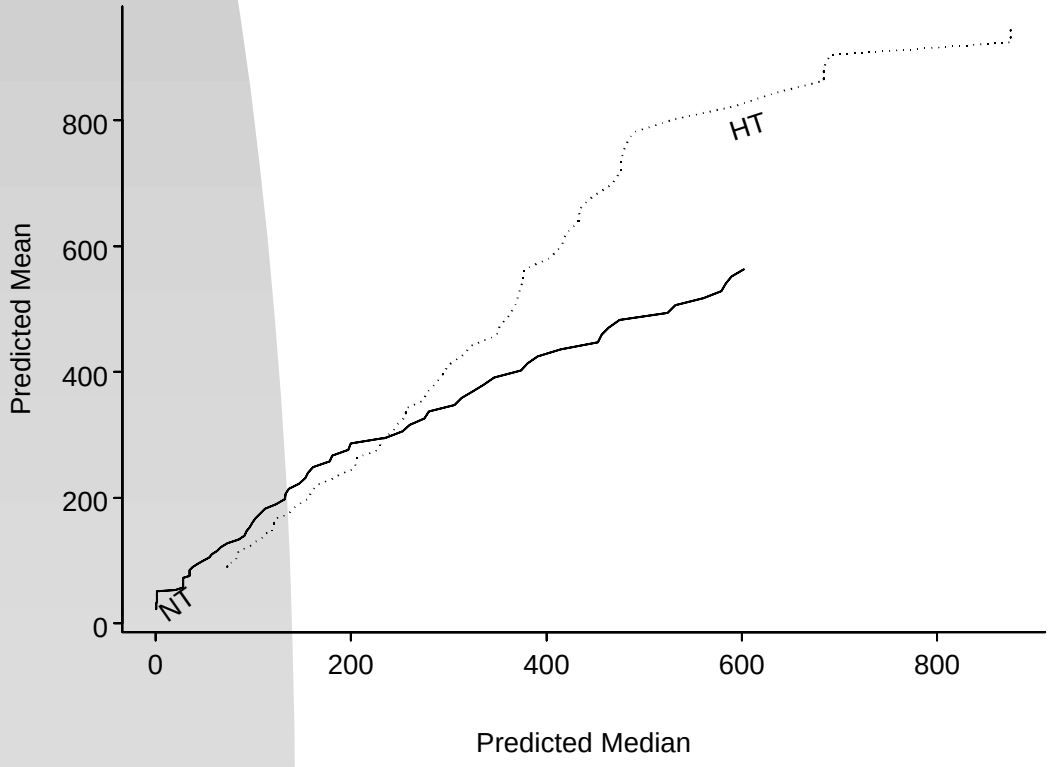
$$2(z_{1\text{-beta}} + z_{1\text{-alpha}})^2 / (\log h)^2$$

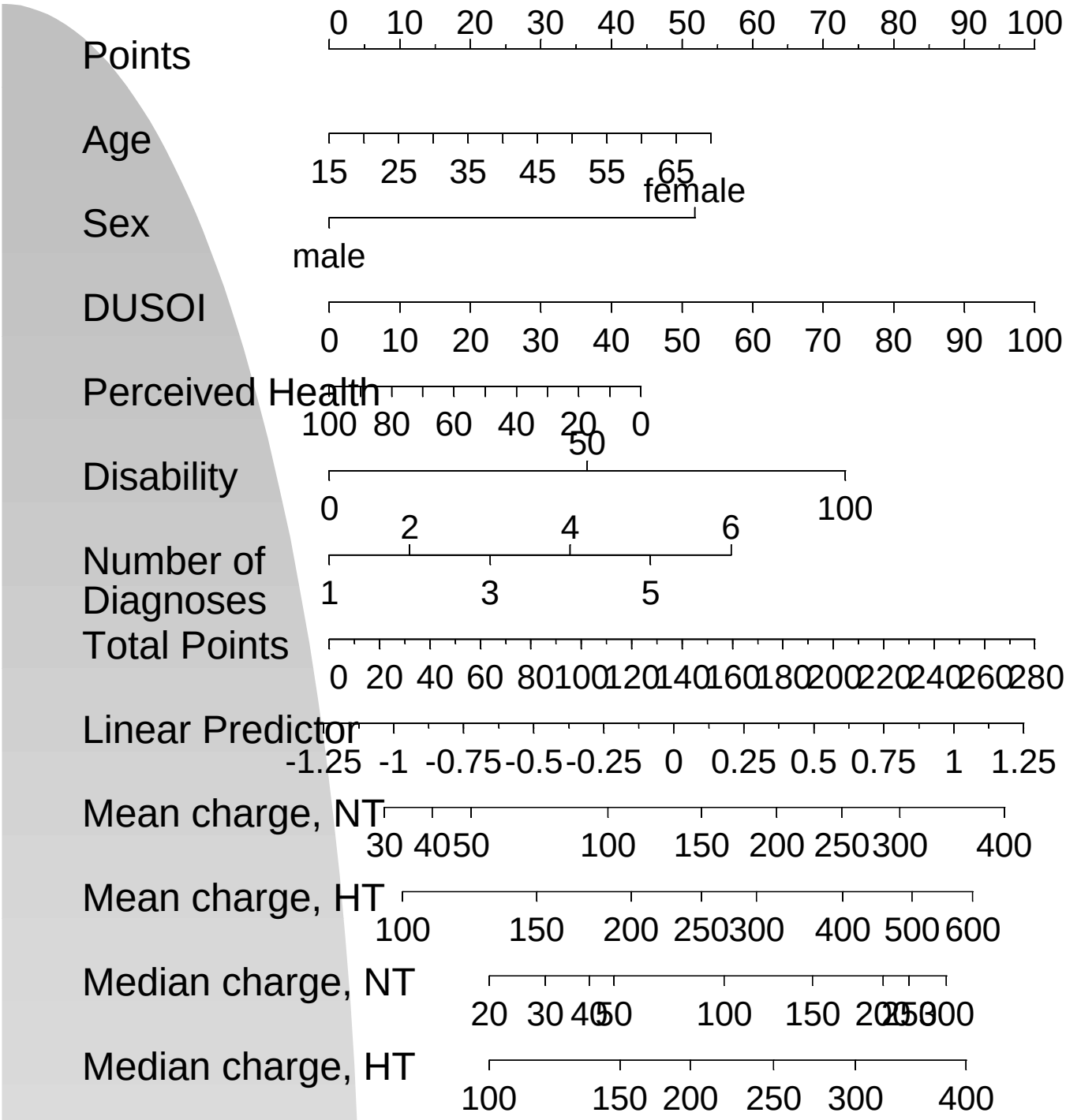
- h = hazard ratio for two treatment arms (= cost ratio for ratios near 1)
- For alpha=.025 beta=.05 sample size is $26 / (\log h)^2$
- Example: h=1.2 n per arm = 782

Parkerson Ambulatory Care Charge Data

- Model total follow-up charges for 413 primary care patients
- Many patients had no follow-ups
- Mixture of diagnoses; hypertensive vs. normotensive (HT,NT) are very important
- Non-proportional hazards for HT vs. NT; all others PH
- For HT/NT allows different shape for charge distribution, different mean-median relationship


HT
NT
mean
median
mean
median
Predicted Mean Charge
0
200
400
600
800
-1.5
-1.0
-0.5
0.0
0.5
1.0
X*Beta


HT
NT
Predicted Mean
0
200
400
600
800
0
200
400
600
800
Predicted Median

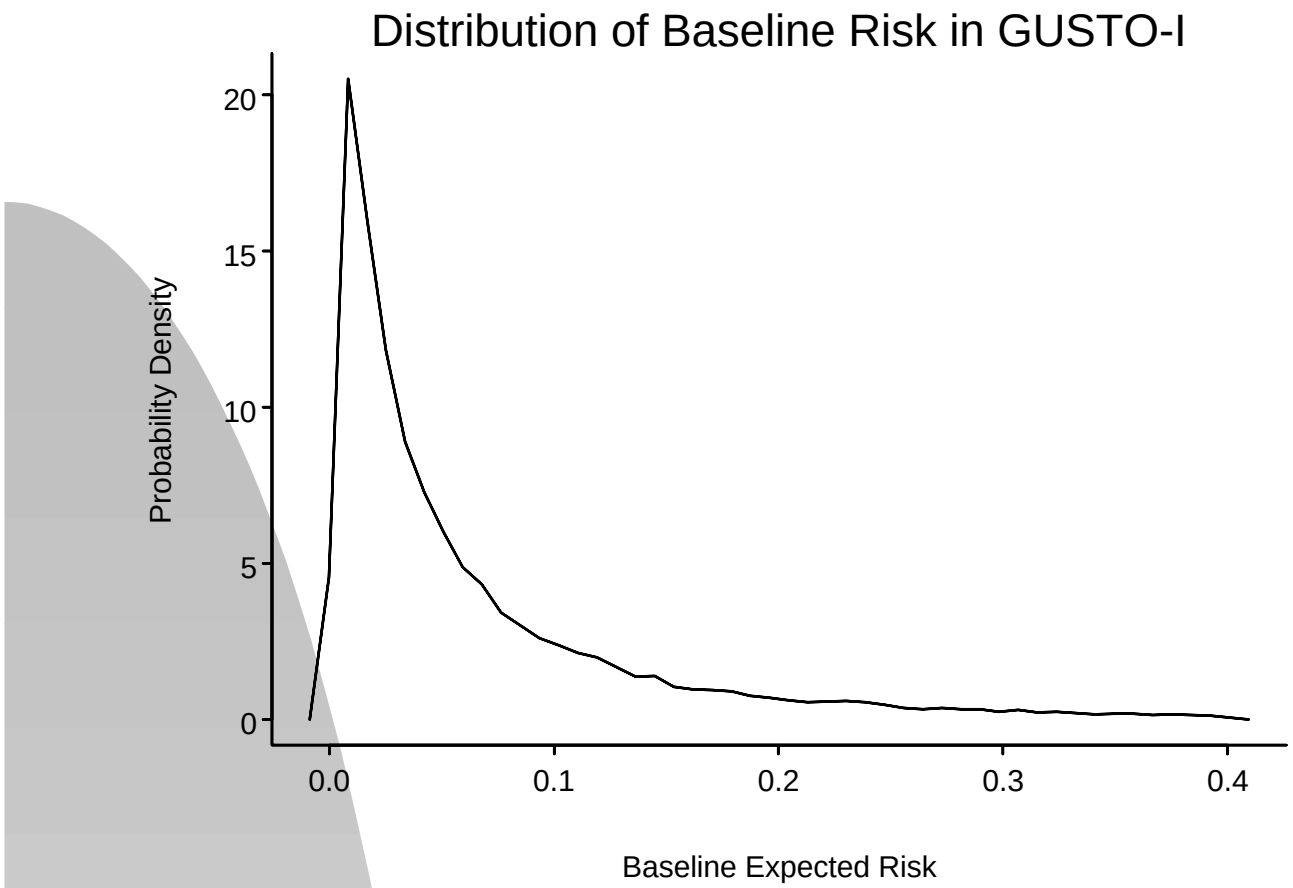

Points
0 10 20 30 40 50 60 70 80 90 100
Age
15 25 35 45 55 65
Sex
female
male
DUSOI
0 10 20 30 40 50 60 70 80 90 100
Perceived Health
100 80 60 40 20 0
Disability
50
0
100
Number of Diagnoses
2 4 6
1 3 5
Total Points
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280
Linear Predictor
-1.25 -1 -0.75 -0.5 -0.25 0 0.25 0.5 0.75 1 1.25
Mean charge, NT
30 40 50 100 150 200 250 300 400
Mean charge, HT
100 150 200 250 300 400 500 600
Median charge, NT
20 30 40 50 100 150 200 250 300
Median charge, HT
100 150 200 250 300 400

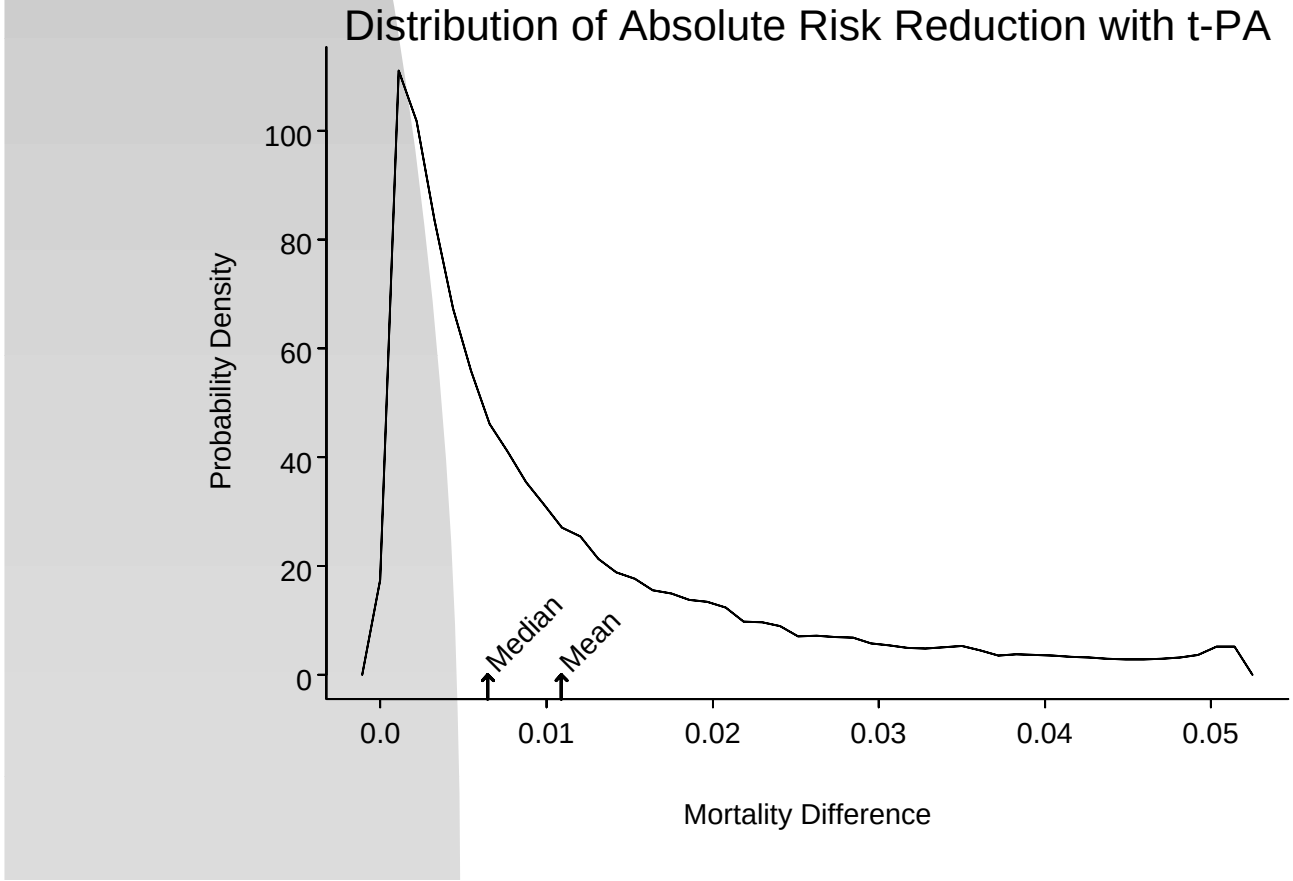
How to Analyze Deaths?

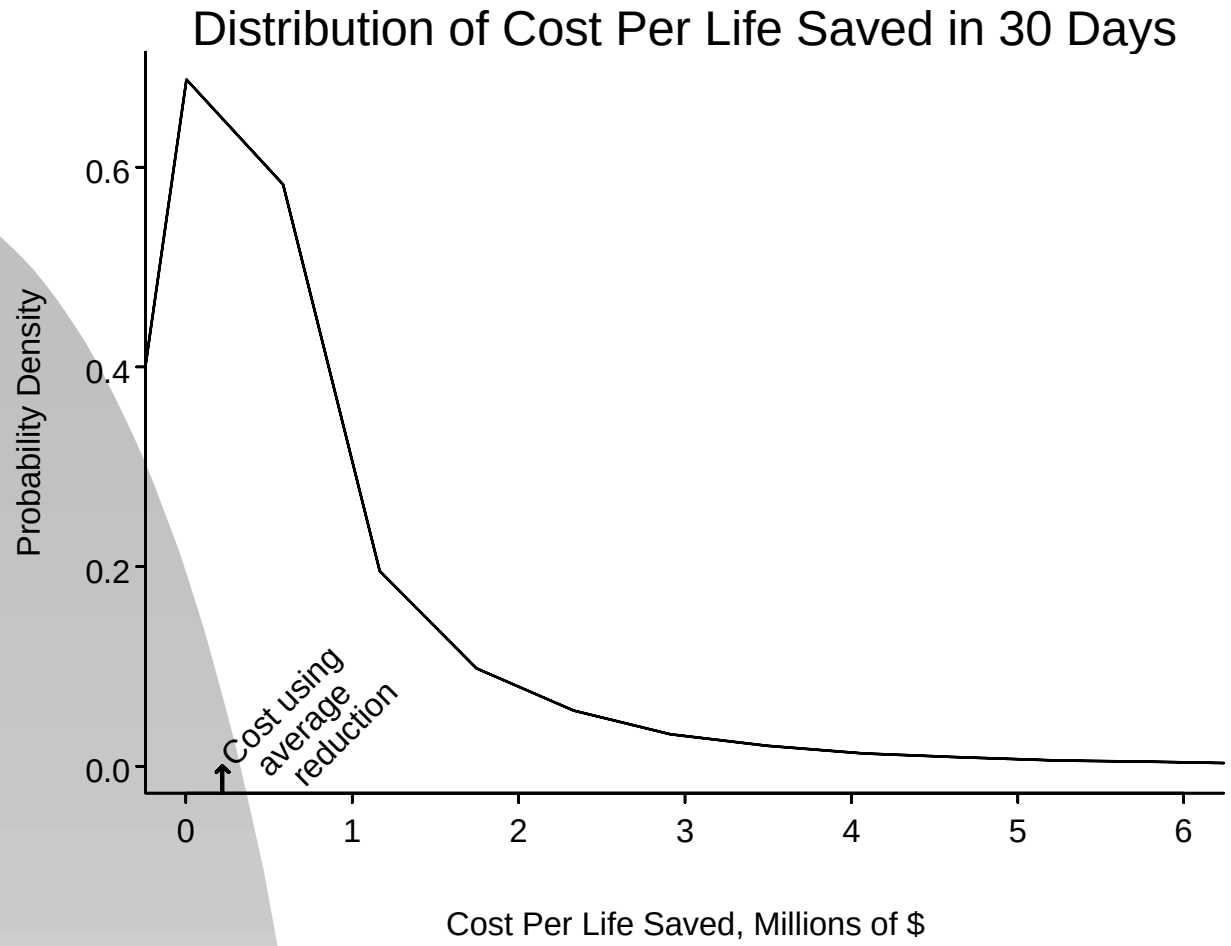
- Consider studies where deaths are common
- Costs would have been greater had patient lived longer
- Don’t want to reward treatment with higher mortality
- May want to censor costs at time of death but account for informative censoring
- Lancaster and Intrator (1995): Joint model for survival and cost

Clinical Effectiveness is Not a Number

- Studies designed to detect relative effectiveness (odds ratio, hazard ratio)
- High-risk patients can dominate the trial (Ioannidis and Lau, J Clin Epi 1997)
- For binary outcome, logistic model dictates that risk difference = P - P/[P + (1-P)/OR]
- Example: GUSTO-I; *t*-PA vs. streptokinase in acute MI
- Endpoint: 30d mortality
- No treatment interactions $\rightarrow$ benefit as given above (Califf et al. Am Heart J 1997)


Distribution of Baseline Risk in GUSTO-I
Probability Density
20
15
10
5
0
0.0
0.1
0.2
0.3
0.4
Baseline Expected Risk


Distribution of Absolute Risk Reduction with t-PA
Probability Density
100
80
60
40
20
0
Median
Mean
0.0
0.01
0.02
0.03
0.04
0.05
Mortality Difference



- Cost per life-year saved (Mark et al. NEJM 1995)
- Low risk patient: $203K
- Medium risk: $ 33K
- High risk: $ 13K
- More variation in CER: cost varies

Summary

- Consider nonparametric tests
- Parametric C.L. are obsolete
- Cox model has greatest potential for cost outcomes
- Special consideration needed for deaths
- Clinical effectiveness of therapy has enormous variability
- Statistical modeling is needed even for simple RCTs
- CE Ratios need to be estimated by the ratio of differences in two multivariable models

Abstract

Typical analyses of cost data utilize tests of differences in mean cost or testing differences after taking logarithms. Nonparametric tests are not often considered, but they have several advantages: (1) some of them, such as the Wilcoxon test, test a very general hypothesis: whether group A tends to have higher costs than group B; (2) they are not affected by transformations on the dependent variable; (3) they are powerful; and (4) nonparametric tests give the proper weight to "outliers." The most general nonparametric test is the Kolmogorov- Smirnov test, which for larger sample sizes will detect any difference in the cost distribution. One can make a case that such general tests should be considered when comparing costs between treatments. On the other hand, when one really does want to restrict the comparison to one of means, the bootstrap should be used for getting confidence limits on the mean difference, as it makes no distributional assumption about costs within treatment groups.

For multivariable modeling of costs, the most common approach is to use multiple linear regression on the log of cost. This approach has two major drawbacks: (1) logarithms typically do not yield Gaussian residual distributions for cost data, and (2) this approach assumes that patient symptoms, severity of disease, and other risk factors act multiplicatively on costs, which is unlikely to be true. The Cox model on the other hand has been found to provide an excellent fit in several datasets (see Dudley et al., J Clin Epi 46:261-71; 1993). The Cox model does not assume either that risk factors act additively or multiplicatively on costs, and it does not assume a mathematical relationship between mean and median costs.

Finally, in estimating cost-effectiveness ratios, most researchers assume that effectiveness is constant across patients, which is far from true. This talk will argue for multivariable modeling of effectiveness and perhaps for cost, before C-E ratios are considered.

For related technical reports, software, and bibliographic database visit
http://www.med.virginia.edu/medicine/clinical/hes/biostat.htm