

---

# The Role of Covariable Adjustment in the Analysis of Clinical Trials

Frank E Harrell Jr

Division of Biostatistics and Epidemiology  
Department of Health Evaluation Sciences  
University of Virginia School of Medicine  
Box 800717 Charlottesville VA 22908 USA

fharrell@virginia.edu

hesweb1.med.virginia.edu/biostat

---

STATISTICS OF MULTI-CENTER TRIALS  
HENRY STEWART CONFERENCE STUDIES  
WASHINGTON, DC 14 SEP 2001

Copyright 2001 FE Harrell All Rights Reserved

## Contents

<b>1</b>	<b>Covariable Adjustment in Linear Models</b>	<b>1</b>
<b>2</b>	<b>Hidden Assumptions in <math>2 \times 2</math> Tables</b>	<b>2</b>
<b>3</b>	<b>Covariable Adjustment in Nonlinear Models</b>	<b>3</b>
3.1	Models for Binary Response . . . . .	3
3.2	Nonlinear Models, General . . . . .	9
<b>4</b>	<b>Cox / Log–Rank Test for Time to Event</b>	<b>9</b>
4.1	Sample Size Calculation Issues . . . . .	11
<b>5</b>	<b>Why are Adjusted Estimates Right?</b>	<b>12</b>
<b>6</b>	<b>How Many Covariables to Use?</b>	<b>13</b>
<b>7</b>	<b>Other Uses of Modeling</b>	<b>14</b>
7.1	Interactions . . . . .	14
7.2	Estimating Absolute Treatment Effects . .	16

<b>8</b>	<b>Cost–Effectiveness Ratios</b>	<b>20</b>
<b>9</b>	<b>Treatment Contrasts for Multi–Site Randomized Trials</b>	<b>23</b>
<b>10</b>	<b>Statistical Plan for Randomized Trials</b>	<b>25</b>
10.1	Sites vs. Covariables . . . . .	27

## 1 Covariable Adjustment in Linear Models

- Model:  $E(Y|X) = X\beta + \epsilon$
- Continuous response variable  $Y$ , normal residuals
- Statistical testing for baseline differences is scientifically incorrect (Altman & Doré 1990, Begg 1990, Senn 1994)
- If we are worried about baseline imbalance we need to search patient records for counterbalancing factors
- → imbalance is not the reason to adjust for covariables
- Adjust to gain efficiency by subtracting explained variation
- Relative efficiency of unadjusted treatment comparison is  $1 - \rho^2$
- Unadjusted analyses yields unbiased treatment effect estimate

## 2 Hidden Assumptions in $2 \times 2$ Tables

- Traditional presentation of 2–treatment clinical trial with a binary response:  $2 \times 2$  table
- Parameters:  $P_1, P_2$  for treatments 1, 2
- Test of goodness of fit:  $H_0$ : all patients in one treatment group have same probability of positive response ( $P_j$  constant)
- $\rightarrow H_0$ : no risk factors exist
- Need to account for patient heterogeneity

### 3 Covariable Adjustment in Nonlinear Models

#### 3.1 Models for Binary Response

- Model for probability of event must be non-linear in predictors unless risk range is tiny
- Useful summary of relative treatment effect is the odds ratio (OR)
- Use of binary logistic model for covariable adjustment will result in an **increase** in the S.E. of the treatment effect (log odds ratio) (Robinson & Jewell, 1991)
- But even with perfect balance, adjusted OR  $\neq$  unadjusted OR
- Adjusted OR will be greater than unadjusted OR

## Example from GUSTO-I

- Steyerberg, Bossuyt, Lee<sup>32</sup>
- Endpoint: 30-day mortality (0.07 overall)
- 10,348 patients given accelerated *t*-PA
- 20,162 patients given streptokinase (SK)
- Means and Percentages

Characteristics of 30,000 GUSTO Patients

Baseline Characteristic	<i>t</i> -PA	SK
Age	61.0	60.9
Female	25.3	25.3
Weight	79.6	79.4
Height	171.1	171.0
Hypertension	38.2	38.1
Diabetes	14.5	15.1
Never smoked	29.8	29.6
High cholesterol	34.6	34.3
Previous MI	16.9	16.5
Hypotension	8.0	8.3
Tachycardia	32.5	32.7
Anterior MI	38.9	38.9
Killip class I	85.0	85.4
ST elevation	37.3	37.8

## Unadjusted / Adj. Logistic Estimates

- With and without adjusting for 17 baseline characteristics

### Unadjusted and Adjusted GUSTO Analyses

Type of Analysis	Log OR	S.E.	$\chi^2$
Unadjusted	-0.159	0.049	10.8
Adjusted	-0.198	0.053	14.0

- Percent reduction in odds of death: 15% vs. 18%
- -0.159 (15%) is a biased estimate
- Increase in S.E. more than offset by increase in treatment effect
- Adjusted comparison based on 19% fewer patients would have given same power as unadjusted test

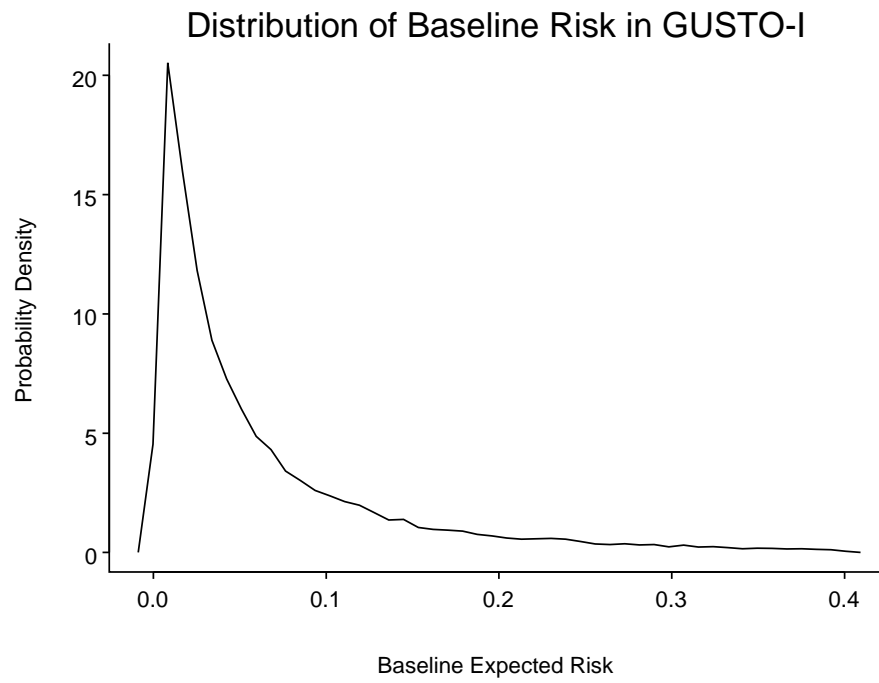


Figure 1: Kernel density estimate of risk distribution for SK treatment. Average risk is 0.07. See also [17].

- Robinson & Jewell: “It is always more efficient to adjust for predictive covariates when logistic models are used, and thus in this regard the behavior of logistic regression is the same as that of classic linear regression.”

## Simple Logistic Example – Gail 1986

$$X = -1$$

	Treatment A	Treatment B
$Y = 1$	500	100
$Y = 0$	500	900
	1000	1000

Odds Ratio: 9

$$X = 1$$

	Treatment A	Treatment B
$Y = 1$	900	500
$Y = 0$	100	500
	1000	1000

Odds Ratio: 9

$$\text{Pooled}$$

	Treatment A	Treatment B
$Y = 1$	1400	600
$Y = 0$	600	1400
	1000	1000

Odds Ratio: 5.44

### 3.2 Nonlinear Models, General

- Gail, Wieand, Piantadosi 1984 showed that for unadjusted treatment estimates to be unbiased, regression must be linear or exponential
- Gail 1986 showed that for logistic, Cox, and paired survival models unadjusted treatment effects are asymptotically biased low in absolute value
- Gail also studied normal, exponential, additive risk, Poisson

## 4 Cox / Log-Rank Test for Time to Event

- Lagakos & Schoenfeld 1984 showed that type I error is preserved if don't adjust
- If hazards are proportional conditional on covariables, they are not proportional if omit covariables

- Morgan 1986 derived asymptotic relative efficiencies (ARE) of unadjusted log-rank test if a binary covariable is omitted
- If prevalence of covariable  $X$  is 0.5:

### Efficiency of Unadjusted Log-Rank Test

$X = 1 : X = 0$ Hazard Ratio	ARE
1.0	1.00
1.5	0.95
2.0	0.88
3.0	0.72

- Ford, Norrie, Ahmadi 1995: Treatment effect does not have the same interpretation under unadjusted and adjusted models
- No reason for the two hazard ratios to have the same value
- Akazawa, Nakamura, Palesch 1997: Power of unadjusted and stratified log-rank test

Power With and Without Adjustment				
Number of Strata	Range of Log Hazards	Power		
		Unadj.	Adjusted	
1	0	.78	–	
2	0–0.5	.77	.78	
	0–1	.67	.78	
	0–2	.36	.77	
4	0–3	.35	.77	
8	0–3.5	.33	.77	

#### 4.1 Sample Size Calculation Issues

- Schoenfeld 1983 implies that covariable adjustment can only  $\uparrow$  sample size in randomized trials
- Need to recognize ill–definition of unadjusted hazard ratios

## 5 **Why are Adjusted Estimates Right?**

- Hauck, Anderson, Marcus<sup>14</sup>, who have an excellent review of covariable adjustment in nonlinear models, state:

“For use in a clinician–patient context, there is only a single person, that patient, of interest. The subject-specific measure then best reflects the risks or benefits for that patient. Gail has noted this previously [ENAR Presidential Invited Address, April 1990], arguing that one goal of a clinical trial ought to be to predict the direction and size of a treatment benefit for a patient with specific covariate values. In contrast, population–averaged estimates of treatment effect compare outcomes in groups of patients. The groups being compared are determined by whatever covariates are included in the model. The treatment effect is then a comparison of average outcomes, where the averaging is over all omitted covariates.”

## 6 **How Many Covariables to Use?**

- Try to adjust for the bulk of the variation in outcome<sup>14, 33</sup>
- Neuhaus<sup>24</sup>: “to improve the efficiency of estimated covariate effects of interest, analysts of randomized clinical trial data should adjust for covariates that are strongly associated with the outcome”

## 7 Other Uses of Modeling

### 7.1 Interactions

- Assessing differential treatment effect best done with formal interaction tests rather than subgroup analysis
- Pre-specify sensible effect modifiers
  - interactions between treatment and extent of disease
  - “learned” interventions: interaction between treatment and duration of use by physician
- Interactions with center are not necessarily sensible
- Need to use penalized estimation (e.g., interaction effects as random effects) to get sufficient precision of differential treatment effects, if # interaction d.f.  $> 4$  for example  
26, 35

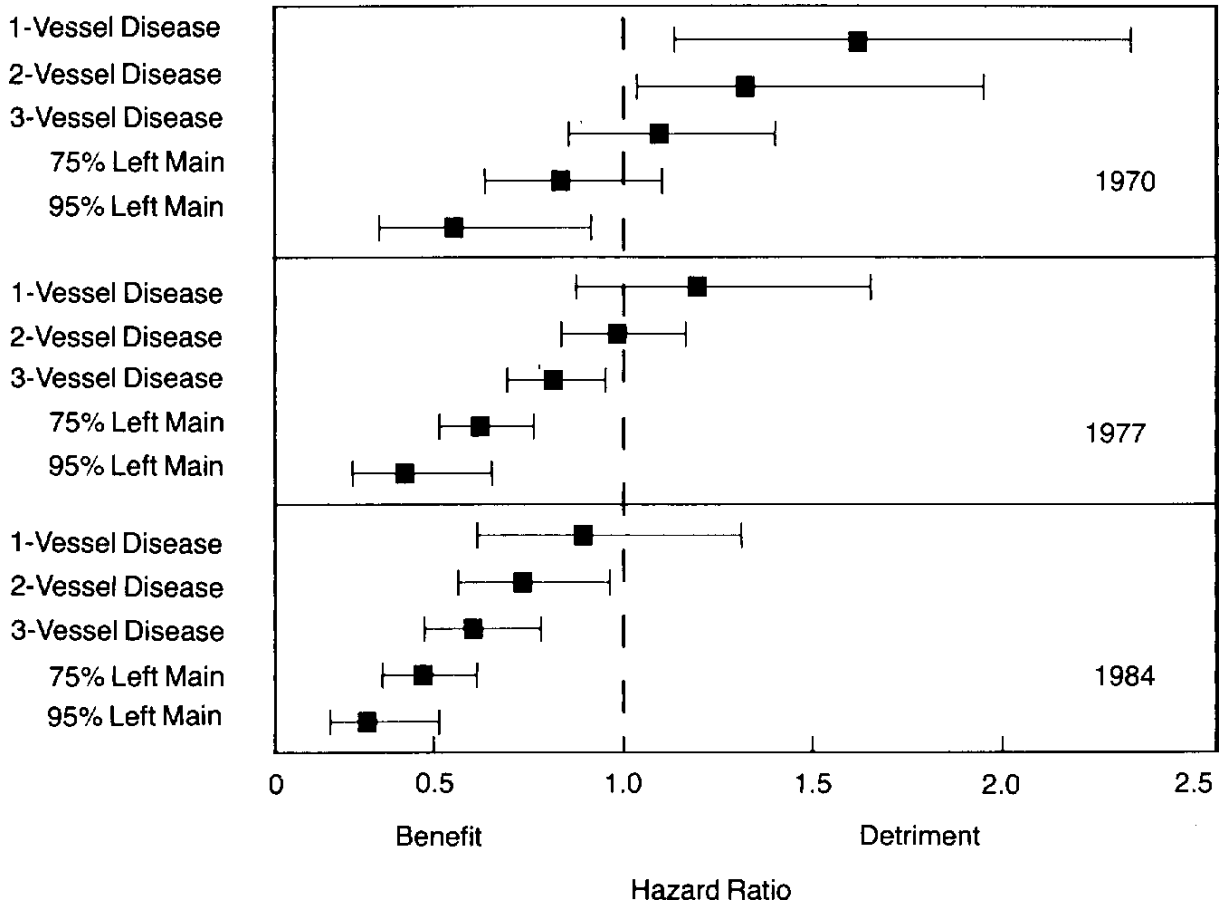


Figure 2: A display of an interaction between treatment, extent of disease, and calendar year of start of treatment<sup>5</sup>

## 7.2 Estimating Absolute Treatment Effects

- Absolute efficacy measures:
  - Risk difference ( $\delta$ )
  - number needed to treat (reciprocal of risk difference)
  - Years of life saved
  - Quality-adjusted life years saved
- Binary response, no interactions with treatment, risk for control patient  $P$ :
$$\delta = P - \frac{P}{P + (1 - P) / OR}$$
- $\delta$  is dominated by  $P$

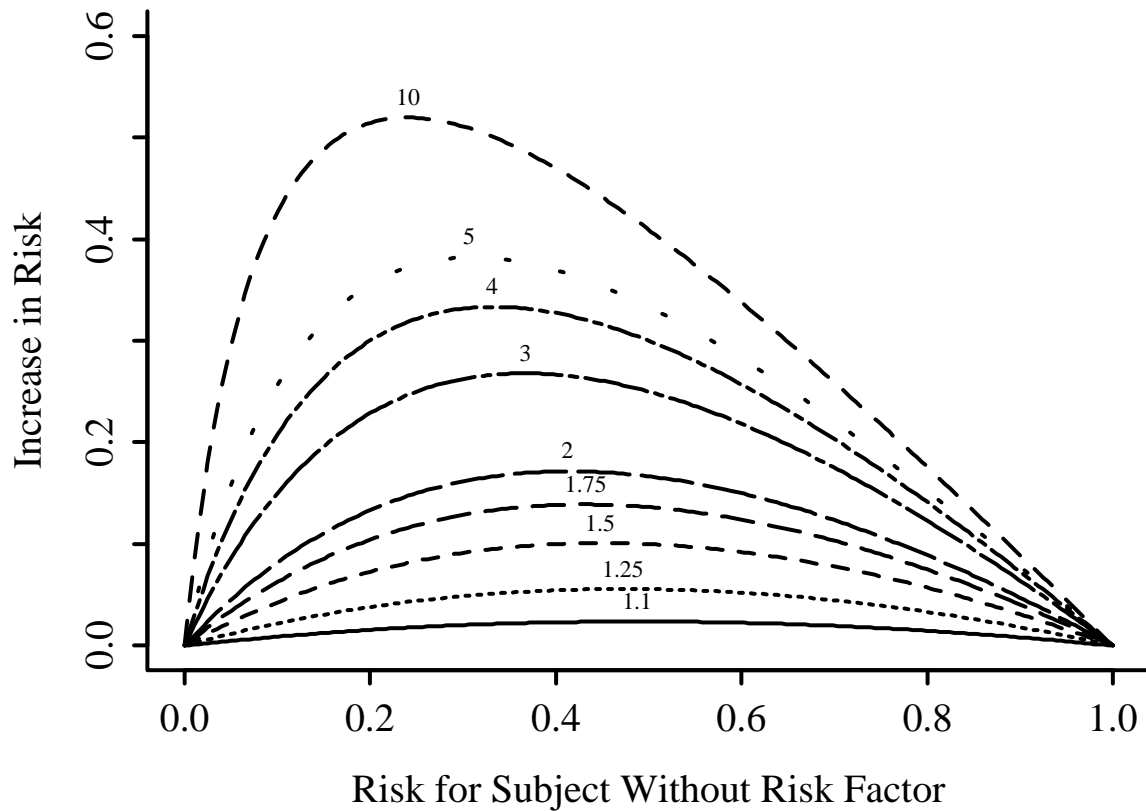


Figure 3: *Absolute risk reduction as a function of risk for control subject. Numbers on curves are treatment:control odds ratios.*

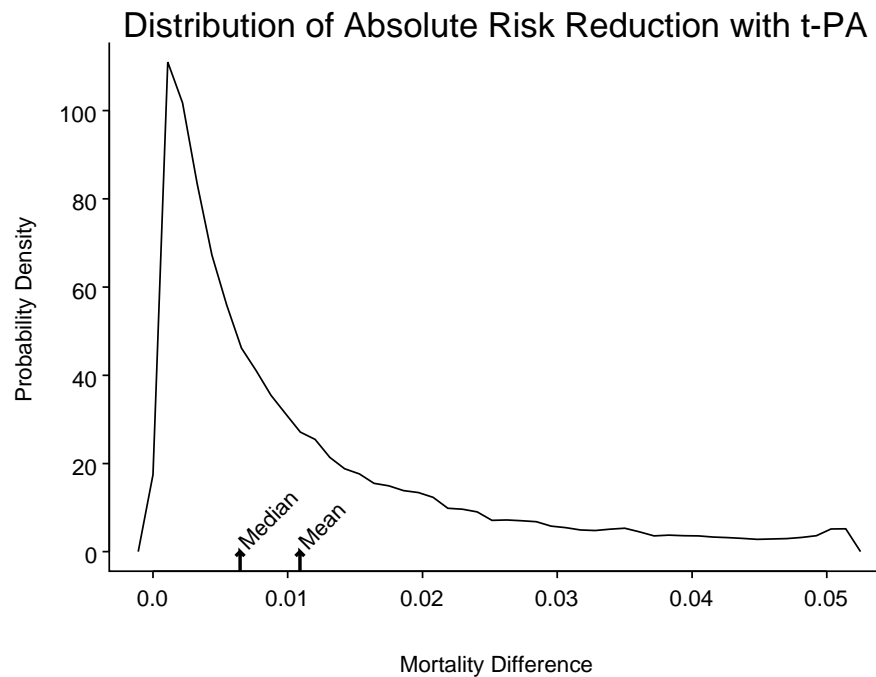


Figure 4: Distribution of absolute benefit of *t*-PA vs. SK

## Absolute Treatment Effects for GUSTO-I

- No evidence for interactions with treatment
- Misleading subgroup analysis showed that elderly patients not benefit from *t*-PA; result of strong age  $\times$  Killip class interaction
- Wide variation in absolute benefit of *t*-PA
- Overall mortality difference of 0.011 dominated by high-risk patients

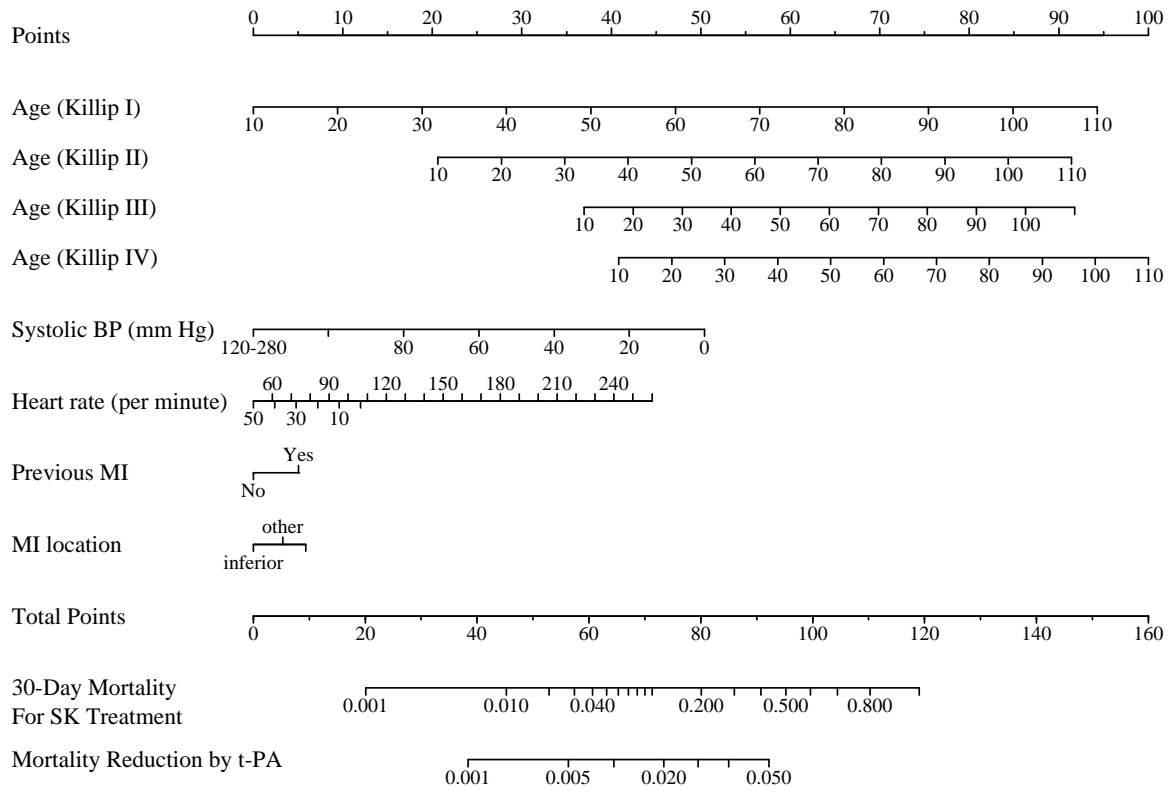


Figure 5: Nomogram to predict SK - t-PA mortality difference, based on the difference between two binary logistic models.

## Absolute Benefit on Survival Prob.

- Cox PH model
- Modeling can uncover time course of treatment effect
- $X_1 = \text{treatment}$ ,  $A = X_2, \dots, X_p$  adjustment variables

- Survival difference is

$$\begin{aligned} & S(t|X_1 = 1, A) - S(t|X_1 = 0, A) \\ &= S(t|X_1 = 0, A)^{HR} - S(t|X_1 = 0, A) \end{aligned}$$

## 8 Cost-Effectiveness Ratios

- Effectiveness E (denominator of C-E ratio) is always absolute
- Absolute treatment effectiveness varies greatly with patient characteristics
- $\rightarrow$  C-E ratio varies greatly
- A C-E ratio based on average E and average C may not apply to any existing patient!

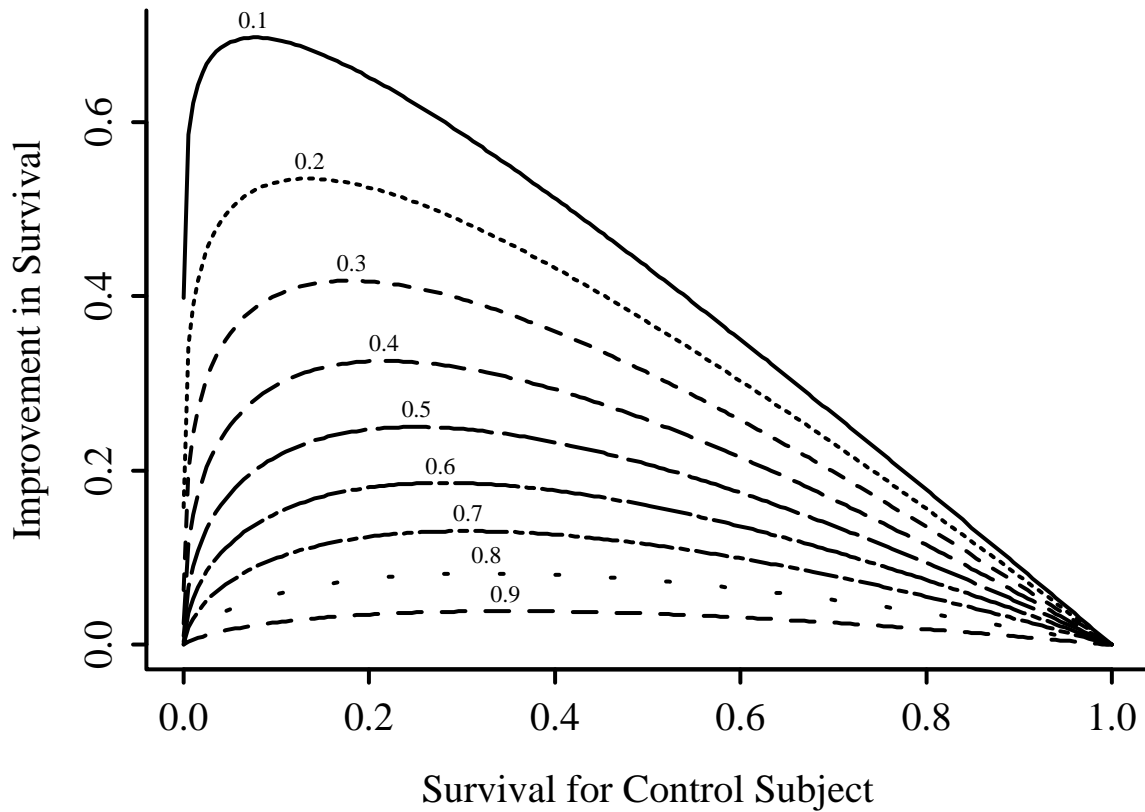


Figure 6: *Relationship between baseline risk, relative treatment effect (hazard ratio — numbers above curves) and absolute treatment effect.*

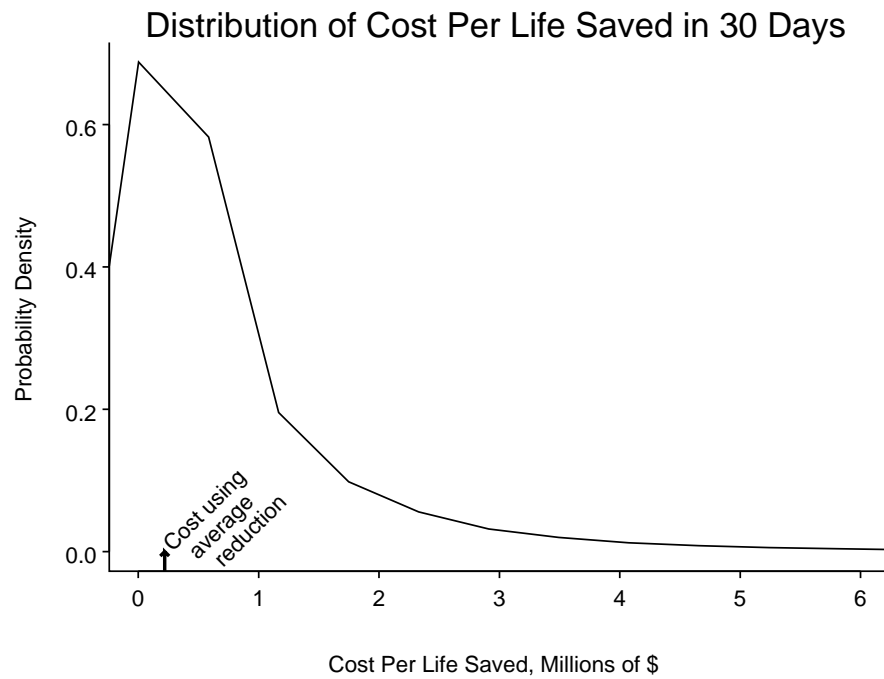


Figure 7: *Distribution of cost per life saved in GUSTO-I*

- Need a model to estimate E
- C may also depend on patient characteristics

## 9 Treatment Contrasts for Multi-Site Randomized Trials

- Primary model: covariables, treatment, site main effects
- Planned secondary model to assess consistency of treatment effects over sites (add site  $\times$  treatment interactions)
- Advantages for considering sites as random effects (or use penalized MLE to shrink site effects, especially for small sites). See [3] for a random effects Cox model and a demonstration that treatment effects may be inconsistent when non-zero site main effects are ignored in the Cox model. See also [35].
- Types of tests / contrasts when interactions are included<sup>28</sup>:
  - Type I: not adjusted for center
  - Type II: average treatment effect, weighted by size of sitesS-PLUS or R Design library command:

```
sites ← levels(site)
contrast(fit, list(treat='b', site=sites),
         list(treat='a', site=sites),
         type='average', weights=table(site))
```

## – Type III: average treatment effect, unweighted

```
contrast(fit, list(treat='b', site=sites),
         list(treat='a', site=sites), type='average')
# Built-in to S-Plus: anova(fit from lm, ssType=3)
```

Low precision; studies are not powered for Type III tests.

- Another interesting test: combined effects of treatment and site  $\times$  treatment interaction; tests whether treatment was effective at *any* site.

## 10 Statistical Plan for Randomized Trials

- When a relevant dataset is available before the trial begins, develop the model from the dataset and use the predicted value as a single adjustment covariable in the trial (Knaus et al. 1993)
- Otherwise: CPMP Working Party: Finalize choice of model, transformations, interactions before merging treatment assignment into analysis dataset.  
Edwards<sup>7</sup>: Pre-specify family of models that will be used, along with the strategy for selecting the particular model.  
Masked model derivation does not bias treatment effect.

“In confirmatory trials, a model is pre-specified, and it is necessary to pretend that it is true. In most other statistical applications, the choice of model is data-driven, but it is necessary to pretend that it is not.”<sup>7</sup>

See also Siqueira and Taylor<sup>31</sup>.

- Choose predictors based on expert opinion
- Impute missing values rather than discarding observations
- Keep all pre-specified predictors in model, regardless of  $P$ -value
- Use shrinkage (penalized maximum likelihood estimation) to avoid over-adjustment
- Detailed strategy in REGRESSION MODELING STRATEGIES, Springer, 2001.

**10.1 Sites vs. Covariables**

- Site effects (main or interaction) are almost always trivial in comparison with patient-specific covariable effects
- It is not sensible to include site in the model when important covariables are omitted
- The most logical and usually the most strong interaction with treatment is not site but is the severity of disease being treated

## References

- [1] K. Akazawa, T. Nakamura, and Y. Palesch. Power of logrank test and Cox regression model in clinical trials with heterogeneous samples. *Statistics in Medicine*, 16:583–597, 1997. []
- [2] D. G. Altman and C. J. Doré. Randomization and baseline comparison in clinical trials. *Lancet*, 335:149–153, 1990. []
- [3] P. K. Andersen, J. P. Klein, and M. Zhang. Testing for centre effects in multi-centre survival studies: A monte carlo comparison of fixed and random effects tests. *Statistics in Medicine*, 18:1489–1500, 1999. [23]
- [4] C. B. Begg. Significance tests of covariate imbalance in clinical trials. *Controlled Clinical Trials*, 11:223–225, 1990. []
- [5] R. M. Califf, F. E. Harrell, K. L. Lee, J. S. Rankin, et al. The evolution of medical and surgical therapy for coronary artery disease. *Journal of the American Medical Association*, 261:2077–2086, 1989. [15]
- [6] CPMP Working Party. Biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products. *Statistics in Medicine*, 14:1659–1682, 1995. []
- [7] D. Edwards. On model pre-specification in confirmatory randomized studies. *Statistics in Medicine*, 18:771–785, 1999. [25, 26]
- [8] I. Ford, J. Norrie, and S. Ahmadi. Model inconsistency, illustrated by the Cox proportional hazards model. *Statistics in Medicine*, 14:735–746, 1995. []
- [9] M. Gail, S. Wieand, and S. Piantadosi. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*, 71:431–444, 1984. []
- [10] M. H. Gail. Adjusting for covariates that have the same distribution in exposed and unexposed cohorts. In S. H. Moolgavkar and R. L. Prentice, editors, *Modern Statistical Methods in Chronic Disease Epidemiology*, pages 3–18. Wiley, New York, 1986. []
- [11] M. H. Gail, W. Y. Tan, and S. Piantadosi. Tests for no treatment effect in randomized clinical trials. *Biometrika*, 75:57–64, 1988. []
- [12] F. E. Harrell. *Regression Modeling Strategies, with Applications to Linear Models, Survival Analysis and Logistic Regression*. Springer, New York, 2001. []
- [13] F. E. Harrell, K. L. Lee, and D. B. Mark. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15:361–387, 1996. []
- [14] W. W. Hauck, S. Anderson, and S. M. Marcus. Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Controlled Clinical Trials*, 19:249–256, 1998. [12, 13]
- [15] W. W. Hauck, J. M. Neuhaus, J. D. Kalbfleisch, and S. Anderson. A consequence of omitted covariates when estimating odds ratios. *Journal of Clinical Epidemiology*, 44:77–81, 1991. []
- [16] International Conference on Harmonisation. Guideline E-9: Statistical Principles for Clinical Trials, 1998. []
- [17] J. P. A. Ioannidis and J. Lau. The impact of high-risk patients on the results of clinical trials. *Journal of Clinical Epidemiology*, 50:1089–1098, 1997. [6]
- [18] A. Källén. Treatment-by-center interaction: What is the issue? *Drug Information Journal*, 31:927–936, 1997. []
- [19] W. A. Knaus, F. E. Harrell, C. J. Fisher, D. P. Wagner, S. M. Opan, J. C. Sadoff, E. A. Draper, C. A. Walawander, K. Conboy, and T. H. Grasela. The clinical evaluation of new drugs for sepsis: A prospective study design based on survival analysis. *Journal of the American Medical Association*, 270:1233–1241, 1993. []

- [20] S. W. Lagakos and D. A. Schoenfeld. Properties of proportional hazards score tests under misspecified regression models. *Biometrics*, 40:1037–1048, 1984. []
- [21] J. A. Lewis. Statistical issues in the regulation of medicines. *Statistics in Medicine*, 14:127–136, 1995. []
- [22] D. B. Macnaughton. Which sums of squares are best in unbalanced analysis of variance? <http://www.matstat.com/ss/>, 1998. []
- [23] T. M. Morgan. Omitting covariates from the proportional hazards model. *Biometrics*, 42:993–995, 1986. []
- [24] J. M. Neuhaus. Estimation efficiency with omitted covariates in generalized linear models. *Journal of the American Statistical Association*, 93:1124–1129, 1998. [13]
- [25] L. D. Robinson and N. P. Jewell. Some surprising results about covariate adjustment in logistic regression models. *International Statistical Review*, 59:227–240, 1991. []
- [26] D. J. Sargent and J. S. Hodges. A hierarchical model method for subgroup analysis of time-to-event data in the Cox regression setting. Presented at the Joint Statistical Meetings, Chicago, 1996. [14]
- [27] D. A. Schoenfeld. Sample size formulae for the proportional hazards regression model. *Biometrics*, 39:499–503, 1983. []
- [28] G. Schwemer. General linear models for multicenter clinical trials. *Controlled Clinical Trials*, 21:21–29, 2000. [23]
- [29] S. Senn. Testing for baseline balance in clinical trials. *Statistics in Medicine*, 13:1715–1726, 1994. []
- [30] S. Senn. *Statistical Issues in Drug Development*. Wiley, Chichester, England, 1997. []
- [31] A. L. Siqueira and J. M. G. Taylor. Treatment effects in a logistic model involving the Box-Cox transformation. *Journal of the American Statistical Association*, 94:240–246, 1999. [26]
- [32] E. W. Steyerberg, P. M. M. Bossuyt, and K. L. Lee. Clinical trials in acute myocardial infarction: Should we adjust for baseline characteristics? *American Heart Journal*, pages 745–751, 2000. Editorial, pp. 761-763. [4]
- [33] J. W. Tukey. Tightening the clinical trial. *Controlled Clinical Trials*, 14:266–285, 1993. [13]
- [34] P. Verweij and H. C. van Houwelingen. Penalized likelihood in Cox regression. *Statistics in Medicine*, 13:2427–2436, 1994. []
- [35] T. Yamaguchi and Y. Ohashi. Investigating centre effects in a multi-centre clinical trial of superficial bladder cancer. *Statistics in Medicine*, 18:1961–1971, 1999. [14, 23]