

Surrogate Endpoints (Markers)

**Michael D. Hughes
Harvard School of Public Health**

Surrogate Endpoint

**"an endpoint measured in lieu of some
other so-called true endpoint"**

[Wittes et al. (1989) Stat. Med.]

Surrogate Endpoints (Markers): Examples

	Potential Surrogate Endpoint	True Endpoint
Cardiovascular	Blood pressure	Myocardial Infarctions
Cancer	Tumor response	Mortality
HIV	CD4 cell count RNA copy number	AIDS
HIV	AIDS	Mortality

Surrogate Endpoints: Why?

- **Aid understanding of drug mechanisms of action and disease pathogenesis**
- **Phase II Trials - for selecting drugs to take forward to major clinical outcome trials**
- **Accelerated Approval - preliminary regulatory approval of a drug pending clinical outcome trials**
- **Phase III Trials - to replace a rare outcome as the primary endpoint**
 - **allows for shorter follow-up and/or smaller sample size**

Prognostic Early Changes Are Not Sufficient to Validate a Surrogate Endpoint

- **Concept:**
 - **define a "response" (e.g. change from baseline)**
 - **show that "responders" have lower rate of clinical outcomes than non-responders**
- **Not a valid approach**

**Prognostic Early Changes Are Not Sufficient to
Validate a Surrogate Endpoint: Example**

**Responder
(by 8 weeks)**

**Proportion Subsequently
Developing AIDS/Dying
(median follow-up: 1 year)**

no

22/223 (9.9%)

logrank p=0.057

yes

7/151 (4.6%)

**Prognostic Early Changes Are Not Sufficient to
Validate a Surrogate Endpoint: Example**

Response variable: CD4 cell count

**Responder: Subject with increase in CD4 from
baseline of 50 cells/mm³**

**Study: ACTG 019 - asymptomatic HIV
infection**

Treatment: placebo!

Prognostic Early Changes Are Not Sufficient to Validate a Surrogate Endpoint: Morals

- **Healthier subjects may be more likely to "respond"
- the association vs. causation problem**
- **CANNOT evaluate a response variable as a surrogate
endpoint using data from a single treatment arm**
- **Need comparative studies**

Surrogate Endpoint - A More Formal Definition

"a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint" [Prentice (1989) Stat. Med.]

Operational Criteria:

- Must be prognostic for major clinical outcome**
- Treatment effect on major clinical outcome should be explained by its effect on the marker**

Surrogate Endpoint Evaluation: Concept for Comparative Studies

- **Find risk of long-term clinical outcome in each treatment group**
- **Show that risk when adjusted for marker in each treatment group is identical**

[Freedman et al. (1992) Stat. Med.]

Proportion of Treatment Effect Explained

- **Fit two proportional hazards models:**

1. $\log \text{HR}(\text{true endpt}) = \alpha + \beta(\text{trt})$

2. $\log \text{HR}(\text{true endpt}) = \alpha + \beta_a(\text{trt}) + \gamma(\text{response})$

where HR = hazard ratio

- **If response variable fully explains treatment effect, then β_a would be zero.**
- **$(\beta - \beta_a)/\beta$ is used to describe the "proportion of treatment effect (on clinical outcome) explained" by the effect on the response variable**

Example: Surrogate Endpoint Evaluation in Comparative Studies

- **Lipid Research Clinics Coronary Primary Prevention Trial**
[Freedman et al. (1992) Stat. Med.]
- **Long-term clinical outcome: death from CHD or nonfatal myocardial infarction**
(average follow-up = 7.4y)

3806 asymptomatic middle-aged men	cholestyramine	6.9%
	placebo	8.8%

Example: Surrogate Endpoint Evaluation in Comparative Studies

- **Marker = serum cholesterol at 1 year**
- **Logistic regression:**
 - without marker $b = -0.26 (0.12)$**
 - with marker $b_a = -0.13 (0.13)$**
- **Marker explains $[-0.26 - (-0.13)]/(-0.26) = 0.50$ of effect on long-term outcome**
- **95% c.i. for proportion of effect explained = 0.07 to 5.91**

Problems with the Proportion of Treatment Effect Explained (1)

- Need statistically significant treatment effect
- Freedman *et al*:
"to make reasonably precise estimates of the proportion of the effect explained by the intermediate endpoint, we need to be explaining unadjusted [treatment] effects which are at least 4 times their standard errors"
- Given limited efficacy of treatments, will need very large trial
- Effect ≥ 4 s.e. -----> $p \leq 0.00003$ - most trials would be stopped before this point reached
-----> Early stopping means unlikely to see such significant effects

Problems with the Proportion of Treatment Effect Explained (2)

- **Estimating the proportion of treatment effect explained recognizes that there may be multiple mechanisms by which treatments could affect the true endpoint**
- **What happens to the proportion if there are multiple mechanisms?**
- **Consider simple case with two mechanisms and results from a very large trial (so that estimates of β and β_a are very precise)**

[ref: DeGruttola *et al.* (1997) J Infect Dis]

Problems with the Proportion of Treatment Effect Explained (2)

- "Net" treatment effect is being explained

i.e. $\beta = \beta_+ + \beta_-$

where β_+ is mediated through the response variable and β_- is not

- Fit model with both treatment and response variable as covariates, then

$$\beta_a = \beta_-$$

- Then proportion of treatment effect explained is

$$\begin{aligned} [\beta - \beta_a] / \beta &= [(\beta_+ + \beta_-) - \beta_-] / (\beta_+ + \beta_-) \\ &= \beta_+ / (\beta_+ + \beta_-) \end{aligned}$$

Problems with the Proportion of Treatment Effect Explained (2)

- **Proportion of effect explained = $\beta_+ / (\beta_+ + \beta_-)$**
- **If $\beta_+ = 0$, then proportion = 0**
 $\beta_- = 0$, then proportion = 1
- **If $\beta_- < 0$, then proportion > 1**
***i.e.* if there is an adverse mechanism of effect, then
proportion explained can be greater than one!!**
- **If there are multiple mechanisms of action, some good
and some bad, proportion of effect explained could take
any value (<0 , between 0 and 1, and >1)**

Caveats in Surrogate Endpoint Evaluation Based on Single Trials

	Reduction in total cholesterol	Primary endpoint rate	All-cause mortality rate
<u>Lipid Research Clinics Trial</u> (n = 3806, 7.4y average f-up)		CHD death/m.i.	
Cholestyramine	13.4%	8.1%	3.6%
	(diff = 8.5%)		
placebo	4.9%	9.8%	3.7%
<u>W.H.O. Cooperative Trial</u> (n = 10627, 5.3y average f-up)		All major IHD	
Clofibrate		3.1%	3.0%
	diff 9%		
placebo		3.9%	2.4%

D=1
d sample
T.M. 2/1

[Ref: JAMA, 1984; Br. Heart J., 1978]

Problems with Evaluating Surrogate Endpoints **Within Individual Clinical Trials**

- **Poor precision unless highly significant treatment differences on clinical outcome**
- **Does the "proportion of treatment effect explained" have any meaning when there are multiple mechanisms of action?**
- **Most likely to be useful for dismissing potential surrogate endpoints, not for validating them**

Meta-analysis for Assessing the Association between Treatment Differences on a Response Variable and on Clinical Outcome

- **Aim:**

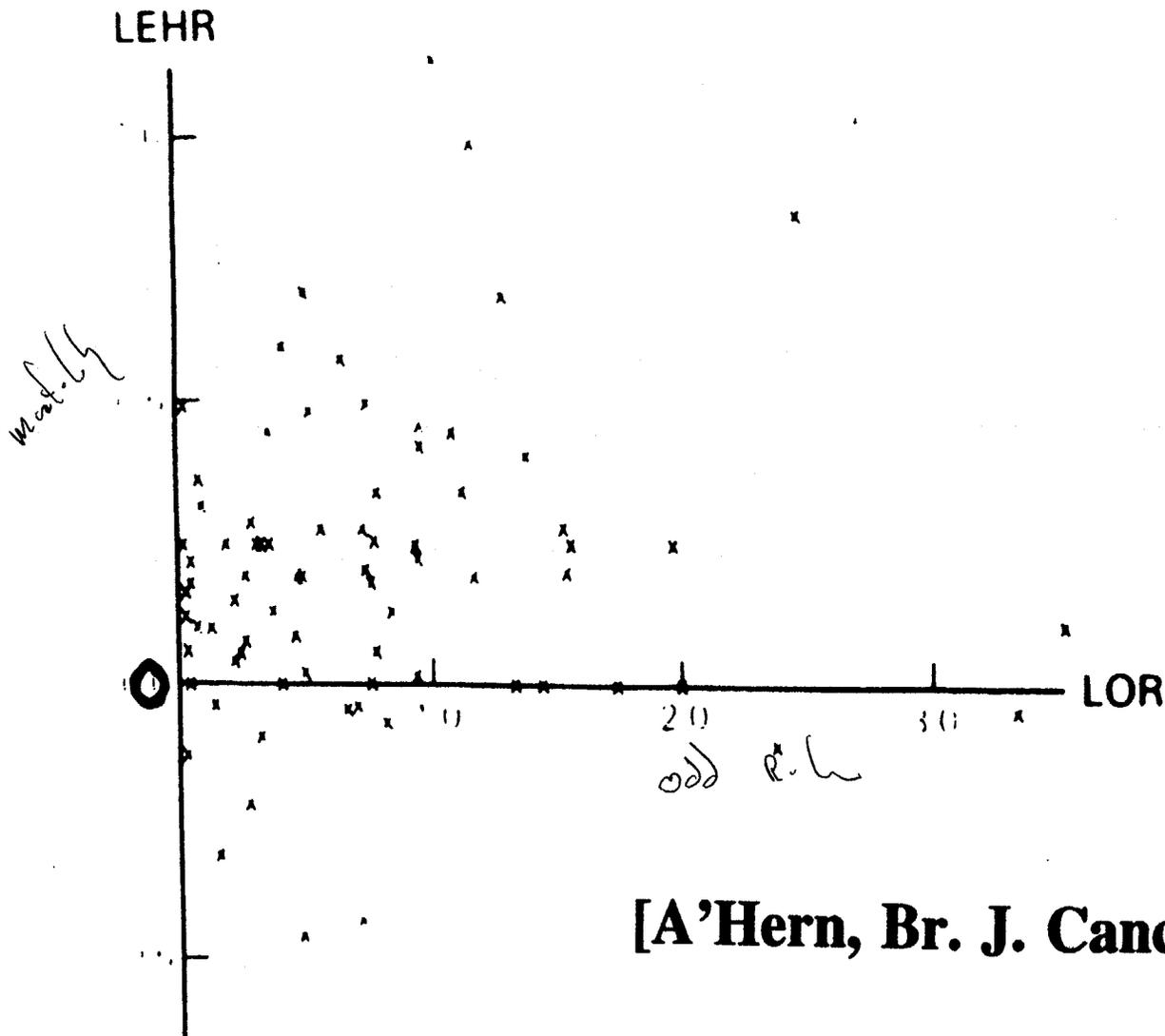
Model the association between the difference in effect on clinical outcome and the difference in effect on the response variable across clinical trials.

meta-analysis
look at diff Rx/p in many trials

- Uses information from trials with significant differences on clinical outcome and those with non-significant differences

[ref: Hughes *et al.* (1995) J. AIDS
Daniels and Hughes, (1997) Stat. Med.]

Evaluating Surrogate Endpoints: Systematic Review
Example: Tumor Response and Mortality



[A'Hern, Br. J. Cancer (1988)]

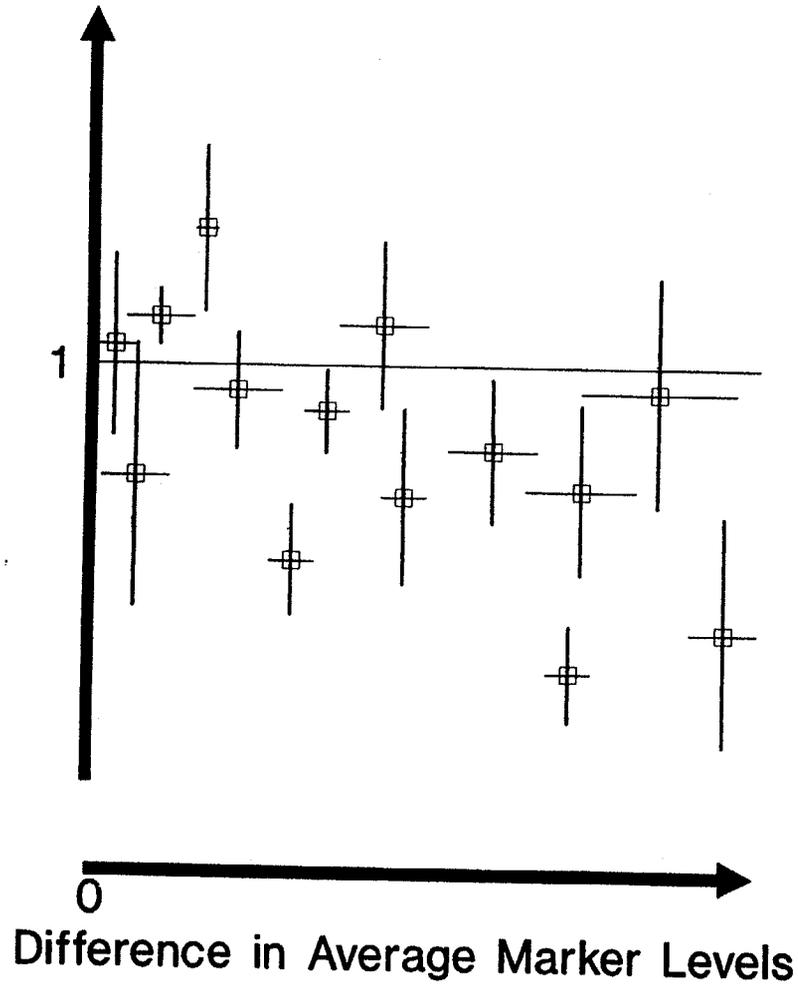
**Meta-analysis for Assessing the Association between
Treatment Differences on a Response Variable
and on Clinical Outcome**

- **Enables an answer to the question:**

Given an observed difference in effect on a response variable in a new trial, based on past experience (*i.e.* the model), what difference in clinical outcome would be predicted?

Evaluating Surrogate Endpoints: Systematic Review A Weak Surrogate Marker for Prediction

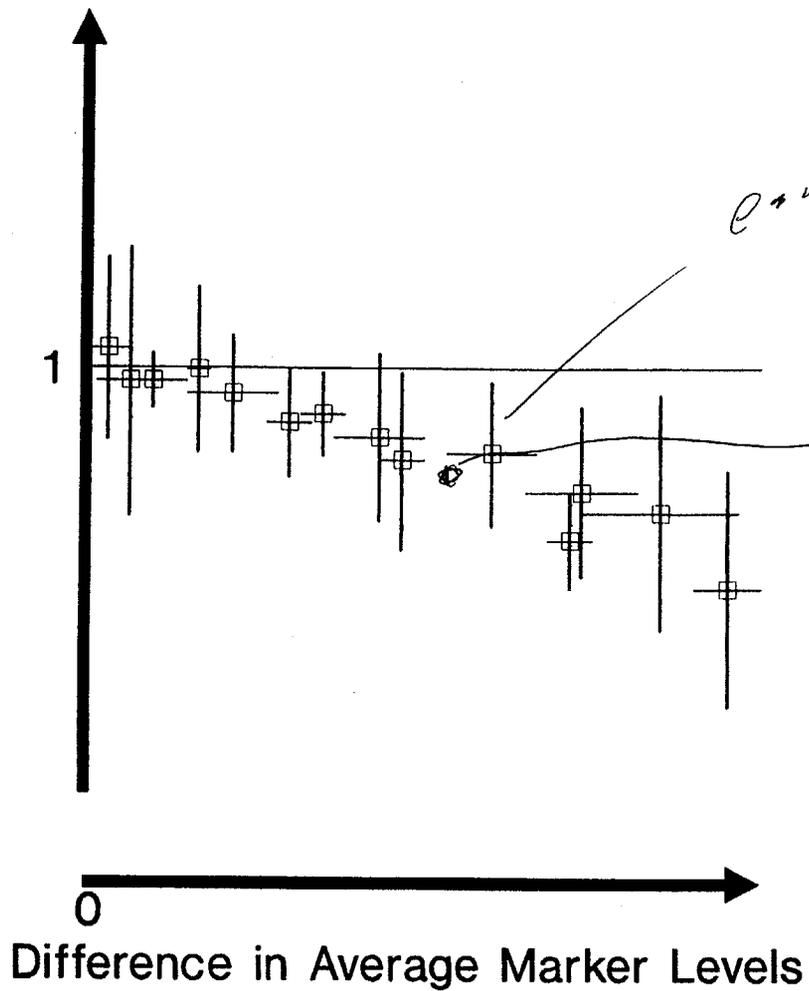
Relative Risk of Clinical Outcome



no trend
but for trial
Result would be
more precise
uncertainty

Evaluating Surrogate Endpoints: Systematic Review A Good Surrogate Marker for Prediction

Relative Risk of Clinical Outcome



each point = clinical trial

so. now that we can judge about a clinical trial

Evaluating Surrogate Endpoints: Meta-analysis
Changes in CD4 Count as an Intermediate Endpoint

- **All randomized trials of the AIDS Clinical Trials Group**
- **Estimates of treatment differences:**
 - **AIDS/death during two years**
 - **change in CD4 cell count from baseline to week 24**
- **Plotted AIDS/death log hazard ratio against difference in mean change in CD4 count**

Table I. Treatment differences for the log hazard ratio for the development of AIDS or death over 2 years (θ_i) and the difference in mean change in CD4 cell count between baseline and 6 months ($\hat{\gamma}_i$) for studies of the AIDS Clinical Trial Group ($\hat{\sigma}_i$, $\hat{\delta}_i$, $\hat{\rho}_i$ are estimates of the standard error of θ_i , the standard error of $\hat{\gamma}_i$, and the correlation between θ_i and $\hat{\gamma}_i$)

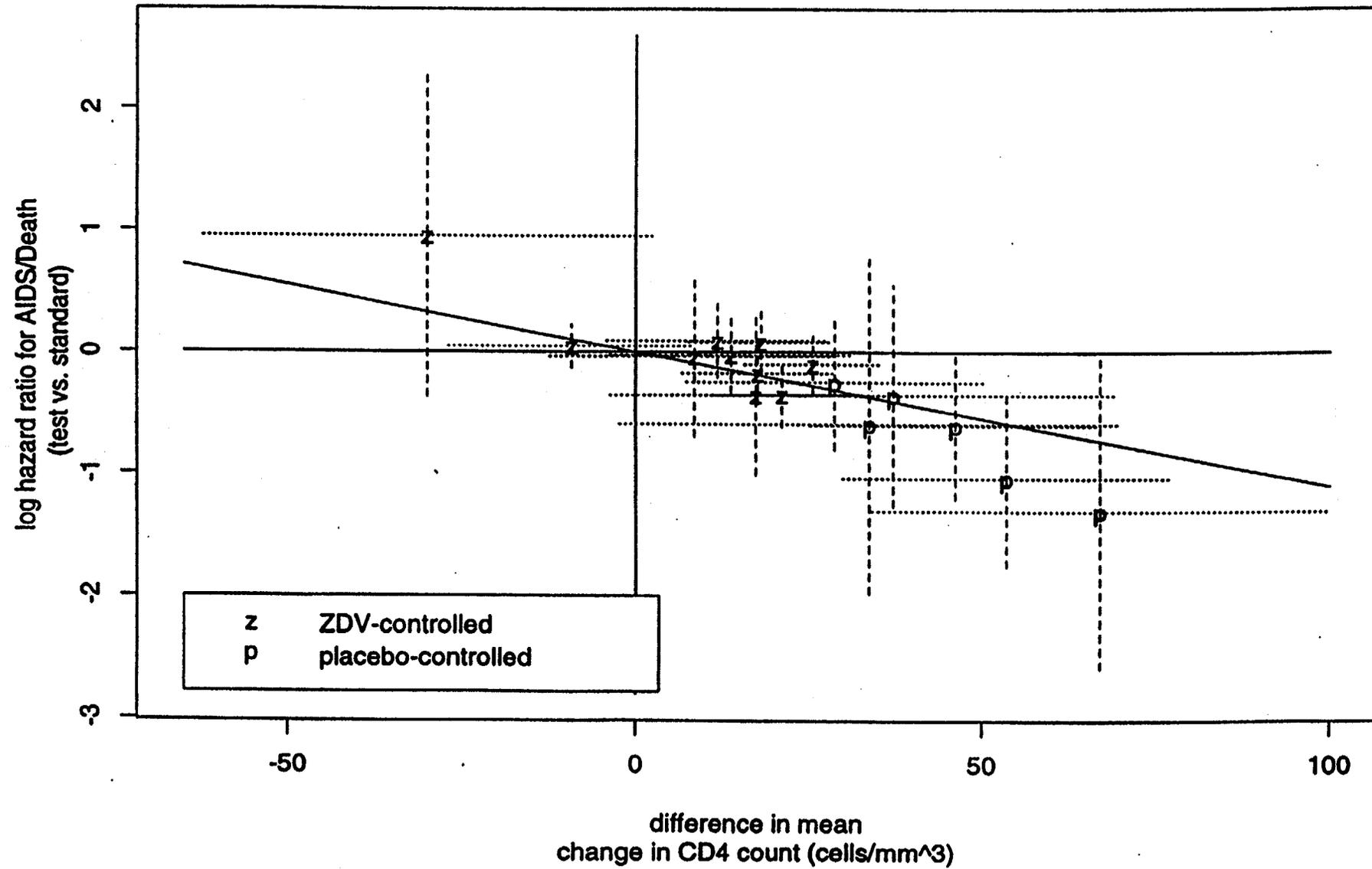
Study	Reference	Test treatment*	Standard treatment*	$\theta_i(\hat{\sigma}_i)$	$\hat{\gamma}_i(\hat{\delta}_i)$	$\hat{\rho}_i$
002	31	ZDV[600]	ZDV[1500]	0.048 (0.092)	-9.2 (9.0)	-0.14
016	32	ZDV[1200]	placebo	-1.035 (0.370)	56.0 (11.8)	-0.02
019a	33	ZDV[1500]	placebo	-0.235 (0.282)	28.8 (11.0)	-0.13
		ZDV[500]	placebo	-0.594 (0.307)	46.1 (10.7)	-0.15
019b	34	ZDV[1500]	placebo	-1.313 (0.651)	67.1 (16.8)	0.01
		ZDV[500]	placebo	-0.359 (0.465)	37.2 (16.3)	-0.00
036	35	ZDV[1500]	placebo	-0.598 (0.707)	32.2 (18.0)	-0.06
112	†	ddC[‡]	ZDV[‡]	-0.447 (0.732)	-4.7 (6.1)	0.17
114	†	ddC[2.25]	ZDV[600]	0.267 (0.121)	-9.1 (5.6)	-0.22
116a	36	ddI[750]	ZDV[600]	0.096 (0.156)	11.8 (8.4)	-0.15
		ddI[500]	ZDV[600]	-0.022 (0.161)	12.8 (8.6)	-0.19
116b	37	ddI[750]	ZDV[600]	0.180 (0.130)	15.9 (5.3)	-0.07
		ddI[500]	ZDV[600]	-0.355 (0.137)	22.2 (5.4)	-0.11
118	38	ddI[200]	ddI[750]	0.112 (0.121)	-8.9 (5.8)	-0.06
		ddI[500]	ddI[750]	0.166 (0.120)	-5.5 (5.8)	-0.05
119	39	ddC[2.25]	ZDV[600]	-0.035 (0.340)	12.8 (9.5)	-0.08
155	40	ZDV/ddC[600/2.25]	ZDV[600]	-0.102 (0.121)	27.5 (4.2)	-0.09
		ddC[2.25]	ZDV[600]	0.083 (0.129)	17.1 (4.5)	-0.10
175	41	ZDV/ddC[600/2.25]	ZDV[600]	-0.348 (0.202)	36.1 (6.5)	-0.13
		ZDV/ddI[600/400]	ZDV[600]	-0.467 (0.207)	71.2 (6.4)	-0.17
		ddI[400]	ZDV[600]	-0.487 (0.207)	40.9 (6.4)	-0.19
229	42	ZDV/SQV[600/1800]	ZDV/ddC[600/2.25]	0.148 (0.518)	7.3 (10.2)	-0.13
		ZDV/ddC/SQV[600/2.25/1800]	ZDV/ddC[600/2.25]	-0.841 (0.680)	15.9 (10.2)	-0.16
241	43	ZDV/ddI/NVP[600/400/400]	ZDV/ddI[600/400]	0.211 (0.258)	25.8 (7.3)	-0.17

* Figures in brackets are the total daily dose (mg), ZDV = zidovudine, ddI = didanosine, ddC = zalcitabine, NVP = nevirapine, SQV = saquinavir

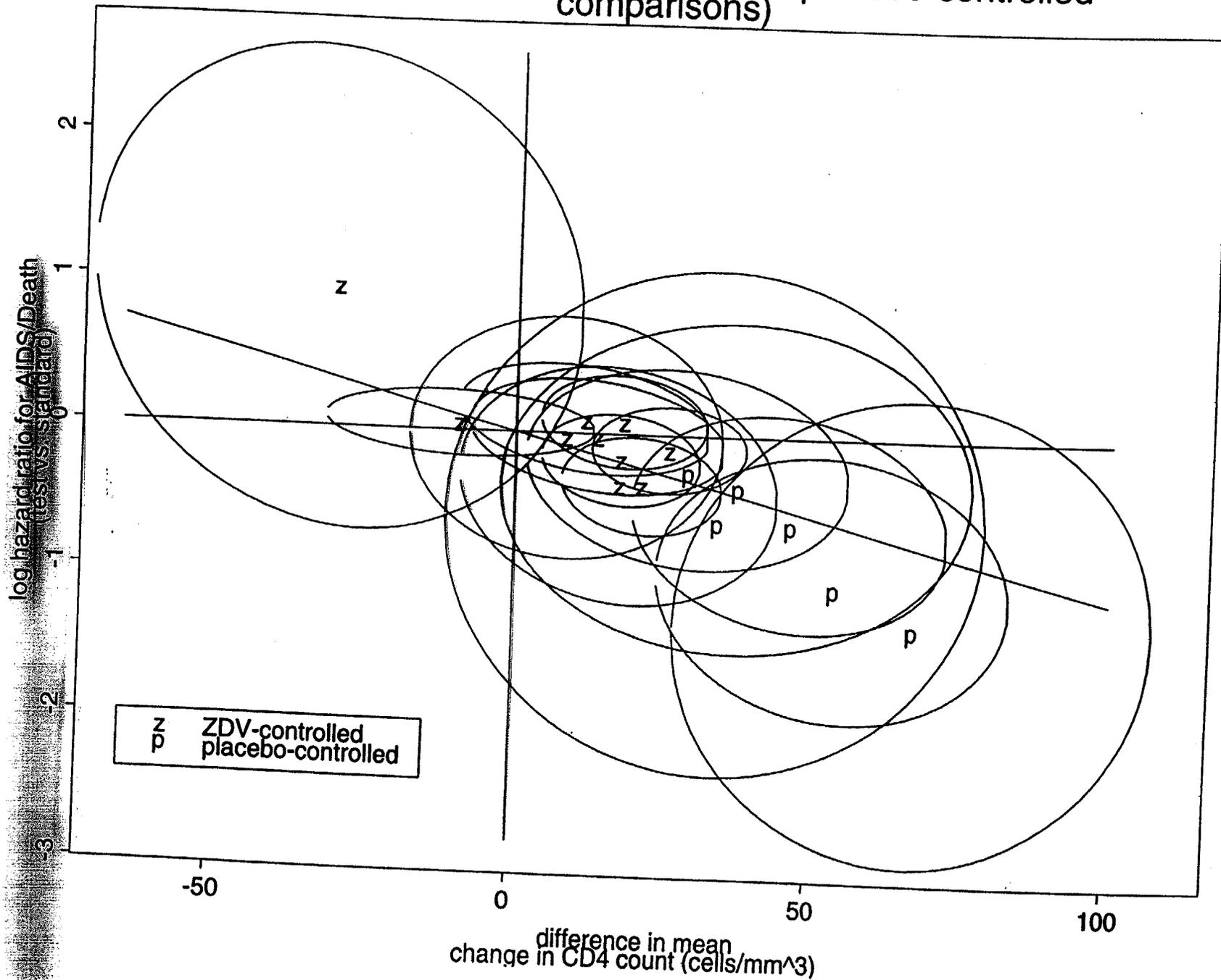
† Not published

‡ The ddC dose was weight dependent and the ZDV dose depended on the dose of ZDV being received prior to study entry

AIDS vs. CD4 (ZDV-controlled and placebo-controlled comparisons)



AIDS vs. CD4 (ZDV-controlled and placebo-controlled comparisons)



Statistical Model

- **Diff between treatments in clinical endpoint = θ ;
Diff between treatments in marker = γ**
- **Simple model: $\theta = \alpha + \beta\gamma + \epsilon$, with $\epsilon \sim N(0, \tau^2)$**
- **Have a surrogate endpoint if $\beta \neq 0$ so that marker differences are predictive of clinical endpoint differences**
- **Also desirable to have τ^2 as small as possible so that predictions of clinical endpoint difference more precise**
- **And good to have $\alpha = 0$ as then zero difference in marker is consistent with zero difference in clinical endpoint**
- **Model fitting is complicated by the fact that true θ and γ are not known
..... only have estimates from each trial.**

Evaluating Surrogate Endpoints: Meta-analysis Changes in CD4 Count as an Intermediate Endpoint

- **With α not set to zero:**

$$\hat{\alpha} = 0.07 \text{ (95\% pred. int: -0.04 to 0.19)}$$

---> not significant

- **With α set to zero:**

$$\hat{\beta} = -0.009 \text{ (95\% pred. int: -0.012 to -0.005)}$$

$$\hat{\tau}^2 = 0.0047 \text{ (95\% pred. int: 0.0001 to 0.0266)}$$

Evaluating Surrogate Endpoints: Meta-analysis
Changes in CD4 Count as an Intermediate Endpoint

- What are the implications of $\tau^2 = 0.0047$?
- given TRUE difference in effect on CD4, can predict true hazard ratio for AIDS/death:

CD4 Diff.	Hazard Ratio	95% pred. int.
0	1.00	0.84, 1.20
10	0.92	0.76, 1.11
20	0.84	0.70, 1.04
30	0.78	0.62, 0.97
40	0.71	0.56, 0.91
50	0.65	0.50, 0.85

Evaluating Surrogate Endpoints: Meta-analysis Changes in CD4 Count as an Intermediate Endpoint

- But don't know TRUE difference on response var.
- Given estimated difference on response variable, can predict difference in AIDS/death ████████████████████

eg $\hat{\gamma} = 30$ cells/mm³, s.e.($\hat{\gamma}$) = 15 cells/mm³

---> predicted hazard ratio = 0.78

95% prediction interval is 0.58, 1.01

Evaluating Surrogate Endpoints: Meta-analysis
Changes in CD4 Count as an Intermediate Endpoint

CD4 Diff.	Hazard Ratio	95% int. (s.e.=0)	95% int. (s.e.=10)	95% int. (s.e.=20)
0	1.00	0.84, 1.20	0.79, 1.28	0.69, 1.48
10	0.92	0.76, 1.11	0.72, 1.17	0.62, 1.35
20	0.84	0.70, 1.04	0.65, 1.08	0.56, 1.24
30	0.78	0.62, 0.97	0.58, 1.01	0.51, 1.14
40	0.71	0.56, 0.91	0.52, 0.94	0.46, 1.06
50	0.65	0.50, 0.85	0.47, 0.88	0.41, 0.98

Meta-analysis of CD4 Count as an Intermediate Endpoint: Conclusions

- **Strong association but possibility of nonzero τ^2**
- **For nucleoside analogues studied, need observed difference in change in CD4 ≥ 40 cells/mm³ to indicate clinical difference**
- **Value of CD4 count as a surrogate endpoint is limited by measurement error/biologic variation and limited effect of treatments on CD4 cell count**
- **??? other classes of drugs**

Evaluating Markers: Conclusions

- Never likely to obtain a perfect surrogate endpoint
- Markers must be evaluated in comparative studies
- Need large studies to obtain sufficient clinical events
- Only real way forward is formal meta-analyses of results for markers/clinical outcome
- Needs rapid collaboration and sharing of data between trial groups/companies