Statistical criteria for components of a primary composite endpoint to support benefit-risk assessment at an interim analysis using blinded data

Jochen Brumm, Roche

Acknowledgements: Markus Abt

Motivating Problem

- Stopping a trial for positive benefit/risk at an interim analysis requires **supportive evidence on secondary endpoints** (besides the criterion on the primary endpoint)
- These analyses are often not adequately powered, so the **criteria tend to be qualitative**, using language like "should not be in contradiction" or "directionally consistent"
- In the case of a **composite endpoint** (= minimum of a number of components), **the components will be key secondary endpoints**
- It is intuitively clear that the different components lead to different weighting of the evidence (e.g. a rarely occurring component will have little information on the efficacy of the drug)
- We attempt to provide **quantitative guidance for decision makers** allowing to distinguish informative analyses from hopelessly underpowered analyses

Useful (?) methodological results for trials with a composite endpoint

- An alternative title for this talk could be: "*Multivariate Modeling of the Hazard Ratios for a Composite Endpoint and its Components*"
- we also used this distribution to assess the operating characteristic of a hierarchical testing procedure
- we don't use any specific properties of cardiovascular endpoints -> some of our results may extend beyond the metabolism area
- so even if you are not dealing with an interim analysis there may be some interesting methodological aspects for you … ☺

Outline

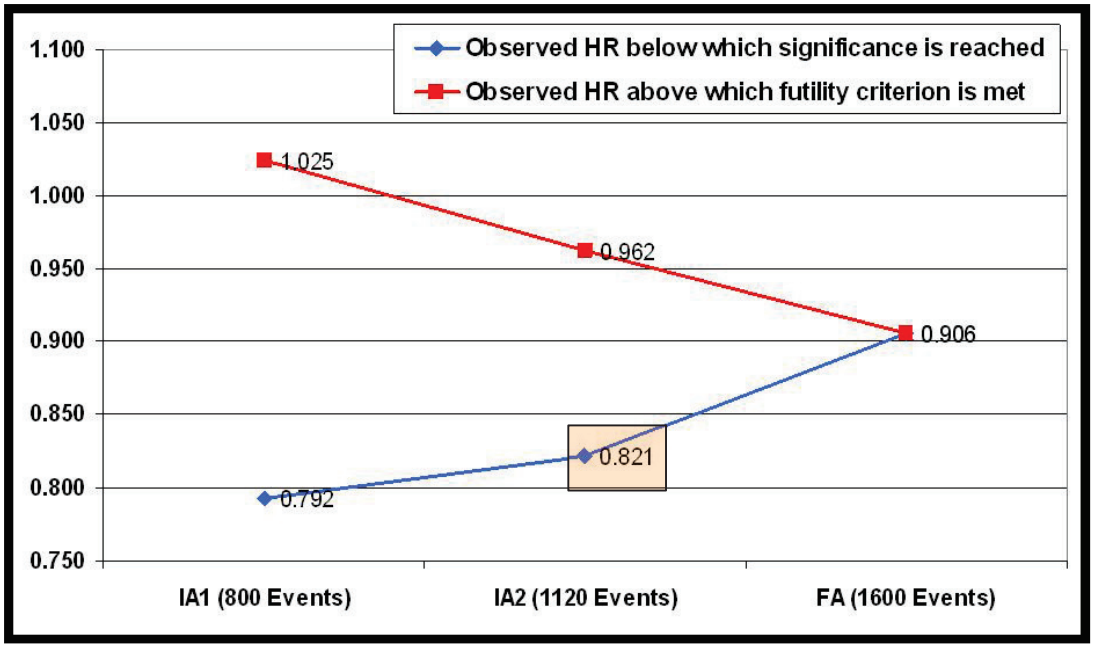
- our motivating problem in detail
  - introduce dal-Outcomes 1 study
  - benefit/risk considerations for the interim analysis
  - decision to stop the study at interim: what are the concerns ?
  - a proposal for quantitative guidance for decision makers
- methods and results
  - model the joint distribution of the log-hazard ratio of the composite and its components

dal-Outcomes 1

- dal-Outcomes 1 was a study designed to show CV risk reduction in patients with a recent acute coronary syndrome (ACS). Powered for a 15% risk reduction (hazard ratio 0.85) it was designed to get 1600 events and enrolled ~15900 patients in two arms (placebo and active drug)
- The study was stopped for futility at the second interim analysis; this talk is based on the situation prior to the interim analysis
- The **primary endpoint is the composite (=minimum) of several components**: CHD death, non-fatal MI, hospitalization for ACS, resuscitated cardiac arrest, stroke (all endpoints adjudicated by a clinical endpoints committee)
- target for interim 1120 primary events, to snapshot date there were (reference primary composite) 63% non-fatal MI, 14% CHD death, 13% hosp for ACS, 9% strokes, 2% cardiac arrests

Interim analyses

- two planned IAs, note in particular the 0.821 threshold for the observed HR to meet the 'efficacy' criterion (we will use that later).



Observed Hazard Ratios for Which Statistical Criteria for Futility and Efficacy are Met

Benefit / Risk Considerations (from the analysis plan)

- "For all cause mortality as well as the individual components of the primary endpoint the results *should not be in contradiction* to those of the primary analysis. […] Even if the true risk reduction for any component was in the range of 15% or more, the observed hazard ratio could still be indicating an increase in risk by chance if only very few events are observed. […]."
- there are also considerations of subgroups and prognostic factors as well as robustness considerations which we will ignore here (although the subgroup analyses may have similar issues)

What is the concern ?

- In our example, if we pass the efficacy threshold at the interim analysis and the observed HR for CHD death is ~1, is that a concern ?
- concerns are both internally (should we really stop or continue to collect more information ?) as well as externally (can we justify a treatment effect on each component to the regulators ?)
- we will take the following position here: **"assuming the drug is (equally) effective for all components, what is the biggest (=least favorable) HR that would still be plausible ?"**
- in other words: can the lack of power explain a high HR (so is it still plausible to assume a favorable treatment effect ?)

Can we put numbers on this ?

- For notational simplicity, (wlog) consider only two components
- We model the multivariate distribution of the estimated log hazard ratios (HR). From this, we can compute the tail probability

$$P(\log \widehat{\mathrm{HR}}_2 > \log(c_2) | \log \widehat{\mathrm{HR}}_{\min} < \log(c_{\min}), \lambda_1, \lambda_2)$$

for arbitrary cutpoints c and true effect sizes lambda. We take cmin to be the HR needed to pass the interim analysis, and consider various true effect sizes (but we always assume

$$\lambda_1 = \lambda_2$$

Comparison of tail probabilities

- Table of P(EP > c | …), the probability to see something as bad as c or worse given the treatment effect of 0.85, in %
- after unblinding, plug in the observed HR as c and worry if the probability is low
- Most interesting rows: CHD death (14%), hosp for ACS (13%) and stroke (9%)

	c = 0.9	c = 0.95	c = 1
stroke	24	16	10
hosp for ACS	20	12	6
CHD death	17	8	4
cardiac arrest	38	33	29
non fatal MI	1	0	0
all cause mortality	16	7	2

Comparison to univariate analysis

- We could ignore the information that we passed the interim analysis and simply compute a tail probability with the number of events for each components
- As the table shows, this is inefficient

tail for c=1; true HR=0.85	bivariate tail probability	univariate tail probability
stroke	10	26
hosp for ACS	6	22
CHD death	4	22
cardiac arrest	29	39
non fatal MI	0	5

Methods

- model the multivariate distribution of the estimated log-hazard ratios Ymin and Y2
- we know that approximately $\begin{pmatrix} Y_{min} \\ Y_2 \end{pmatrix} \sim N(\lambda, \Sigma)$

where the mean is the true log HR, the variance for each Y is simply 4/(total number of events) (correlation/covariance will be discussed below)

- using this distribution, we get:

$$P(Y_2 > \log(c_2) | Y_{min} < \log(c_{min})) = \frac{P(Y_2 > \log(c_2), Y_{min} < \log(c_{min}))}{P(Y_{min} < \log(c_{min}))}$$

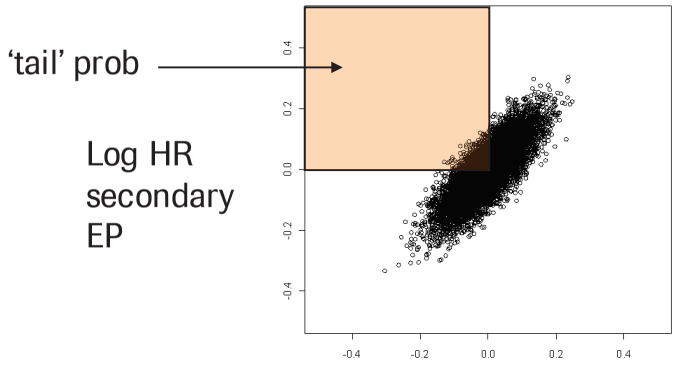
- In R, computations are easily done using the package “mvtnorm” (a question of seconds)

Estimating the correlation

- we can get the *correlation between the log HR of the primary and the log HR of the secondary endpoint* under H0 by permuting the treatment allocations on the blinded data
- Is this correlation the same as under assumed treatment effects ? Not 100% sure, but at least in exponential model a simulation confirmed it (-> backup)

Component which equals primary in 62%

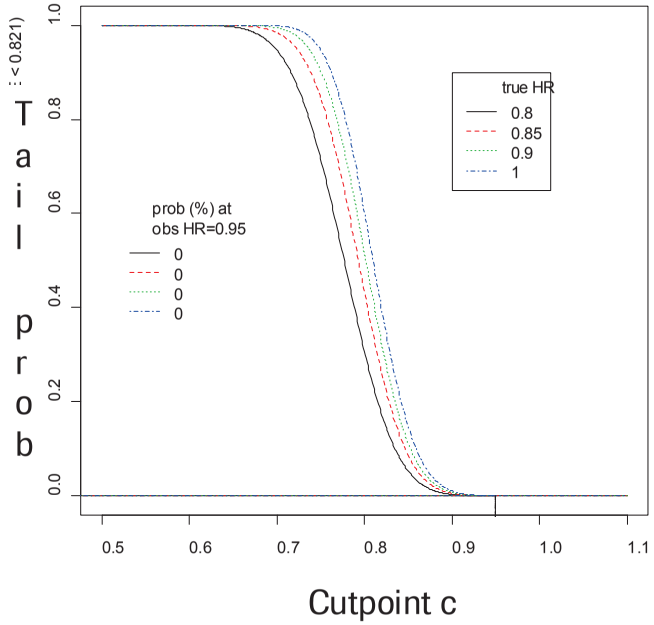
cor = 0.80



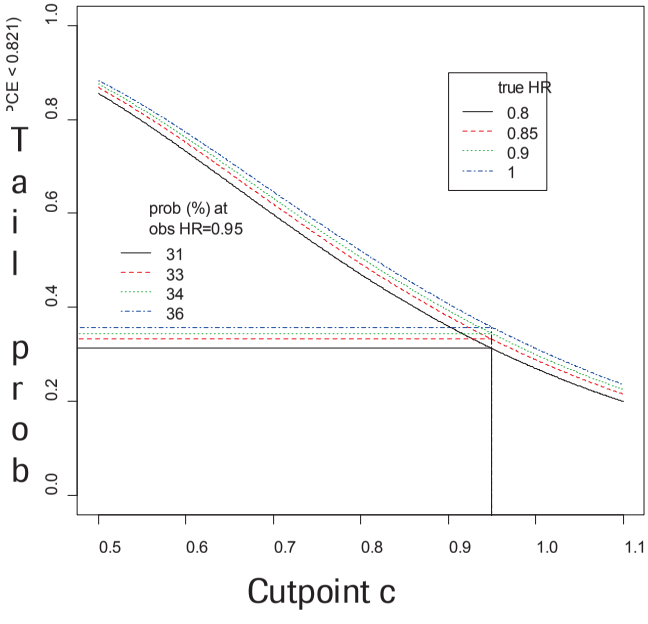
Operating Characteristics of the Tail Probabilities

- Impact of assumed treatment effect: high for high-percentage components

Non-fatal MIs (62% of composite)



Cardiac Arrest (2% of composite)



Correlations and percentages

- one can see that the correlation between the hazard ratio of a component to the composite endpoint is related to the percentage the composite equals the component

	non fatal MI (62%)	CHD death (14%)	hosp for ACS (13%)	stroke (9%)	cardiac arrest (2%)
PCE	**80%**	**41%**	**36%**	**31%**	**15%**
non fatal MI		11%	5%	4%	5%
CHD death			3%	3%	7%
hosp for ACS				0%	1%
stroke					0%

log HR cor in %

Mathematical insight: can we estimate correlation prior to the trial ?

- It turns out that in the exponential model we can prove that approximately (without censoring; maybe possible to extend to censored case)

$$cor(\log HR_{min}, \log HR_2) = \frac{\lambda_2}{\sum_i \lambda_i}$$

- In the exponential model, the contribution of the k-th component to the composite is given by

$$P(X_k = \min\{X_1, \ldots, X_n\}) = \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_n}$$

- Hence if we have a guess about the event rates or about the relative contributions of the components, we can estimate the correlation between the hazard ratios

Conclusions

- the quantitative assessment sets some expectations about directional consistency:
  - for non-fatal MI (62% of the events) we should expect consistency
  - for cardiac arrests (2%) we should be ready to observe a HR > 1
  - for CHD death (14%), expect directional consistency with a 96% probability
  - for stroke (9%), expect directional consistency with a 90% probability
- a priori not clear what % of events would be sufficient to expect consistency, so the statistical method provides additional insight beyond common sense
- We don't present the probabilities for multiple components simultaneously, but this is feasible
- Other issues like subgroup analyses should also be considered when considering to stop for efficacy

Thank you !
Questions ?

Roche

Doing now what patients need next

Impact of Treatment on Correlation: Exponential model

- no impact of treatment effect on correlations between the HR for the minimum and a component in a small simulation study (exponential model, no censoring, 3000 patients per arm, 200 trials simulated)
- setup: proportions: Var 1: 70%, Var 2: 20% (simulations are set up in a way that these proportions are preserved for the different treatment effects)

Correlation between HRmin and HRi

Variable	True HR	Min	Q1	Med	Q3	Max
Hrmin,HR1	0.8	0.69	0.69	0.72	0.74	0.75
	0.9	0.67	0.69	0.70	0.71	0.75
	1	0.67	0.71	0.72	0.73	0.74
Hrmin,HR2	0.8	0.09	0.16	0.17	0.21	0.33
	0.9	0.10	0.15	0.21	0.23	0.28
	1	0.06	0.12	0.19	0.21	0.27

Sketch of the proof for exponential model

1. Use that if the variance is “low” (as in our case), the correlation between two log-normal random variables is approximately equal to the correlation of the corresponding normal random variables (!)
2. Use that in the exponential model, $$\hat{\lambda}_P = \frac{\text{no of deaths}}{\sum_i t_{i,P}}$$
3. We hence have
$$cor(\widehat{HR}_{min}, \widehat{HR}_2) = cor(\frac{\sum t_{min,i,P}}{\sum t_{min,i,A}}, \frac{\sum t_{2,i,P}}{\sum t_{2,i,A}})$$
4. Use an approximation for the correlation of ratios in terms of the correlations of the survival times

Key Result: Correlation of Ratios

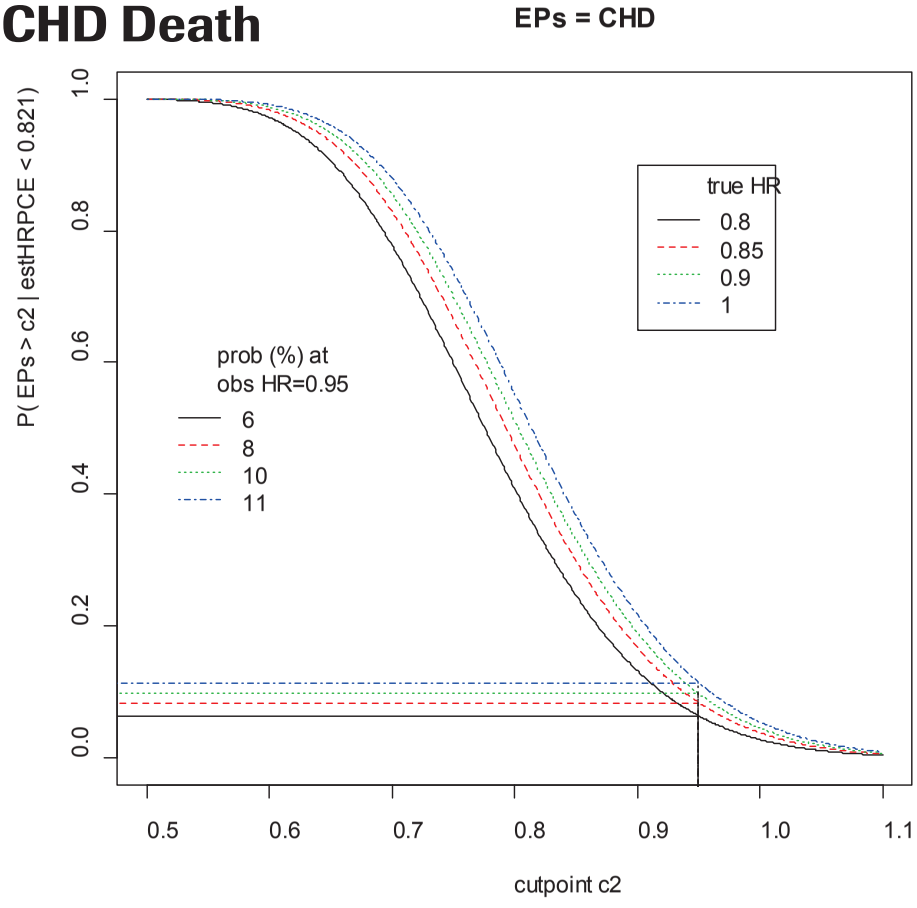
In 1897 Karl Pearson presented an approximate formula for a correlation of ratios expressed in terms of the correlations and coefficients of variation of its individual elements:

$$r_{\frac{Y}{Z}\frac{X}{W}} = \frac{r_{YX} V_Y V_X - r_{YW} V_Y V_W - r_{XZ} V_X V_Z + r_{ZW} V_Z V_W}{\sqrt{V_Y^2 + V_Z^2 - r_{YZ} V_Y V_Z}\ \sqrt{V_X^2 + V_W^2 - r_{XW} V_X V_W}}$$

from “The Correlation of Ratios or Difference Scores Having Common Terms”, G.V. Guguitt and S. Lieberson, 1972

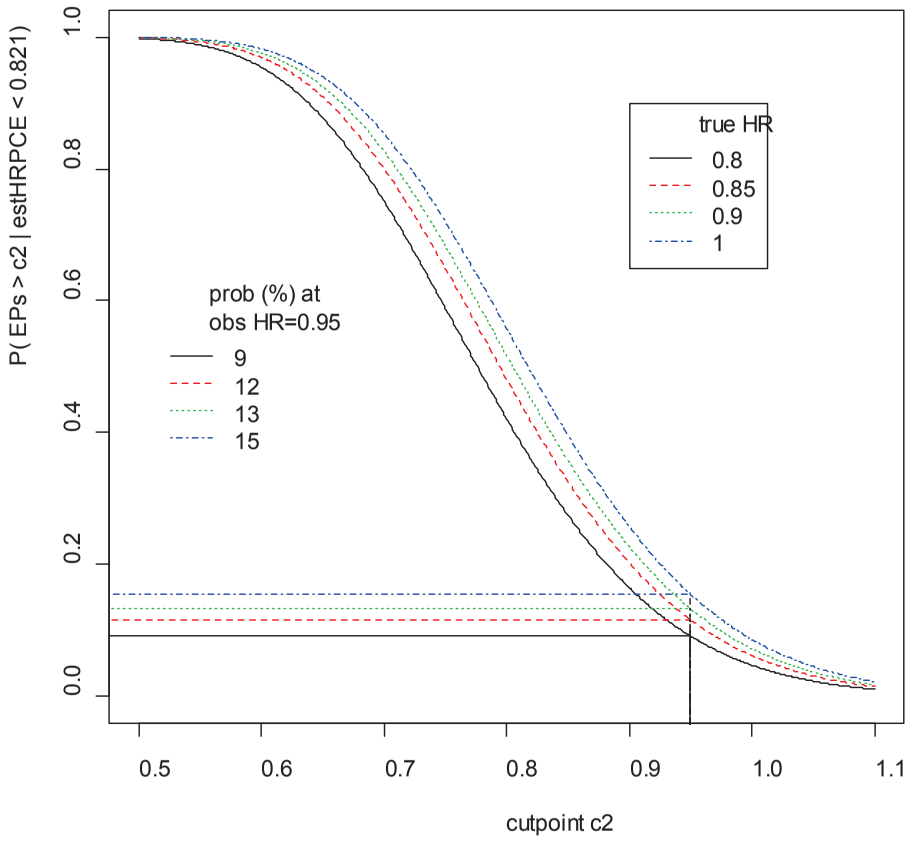
does anyone have the Pearson reference (or a book with the derivation) ?

Results: CHD Death



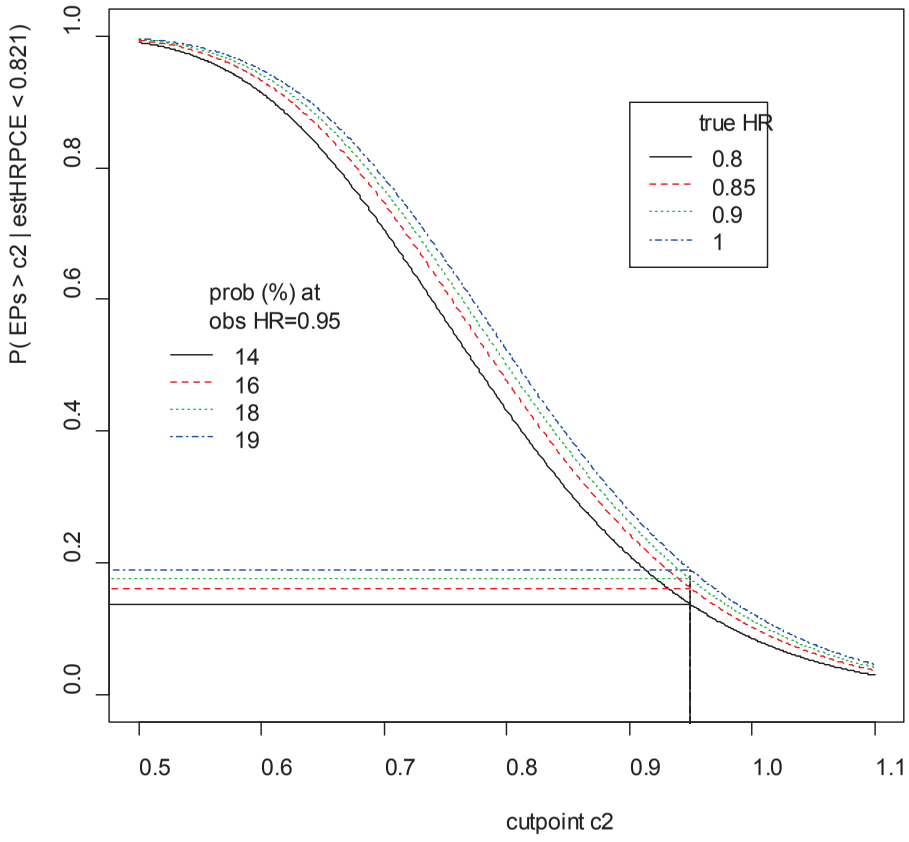
Results: hosp for ACS

EPs = MCH

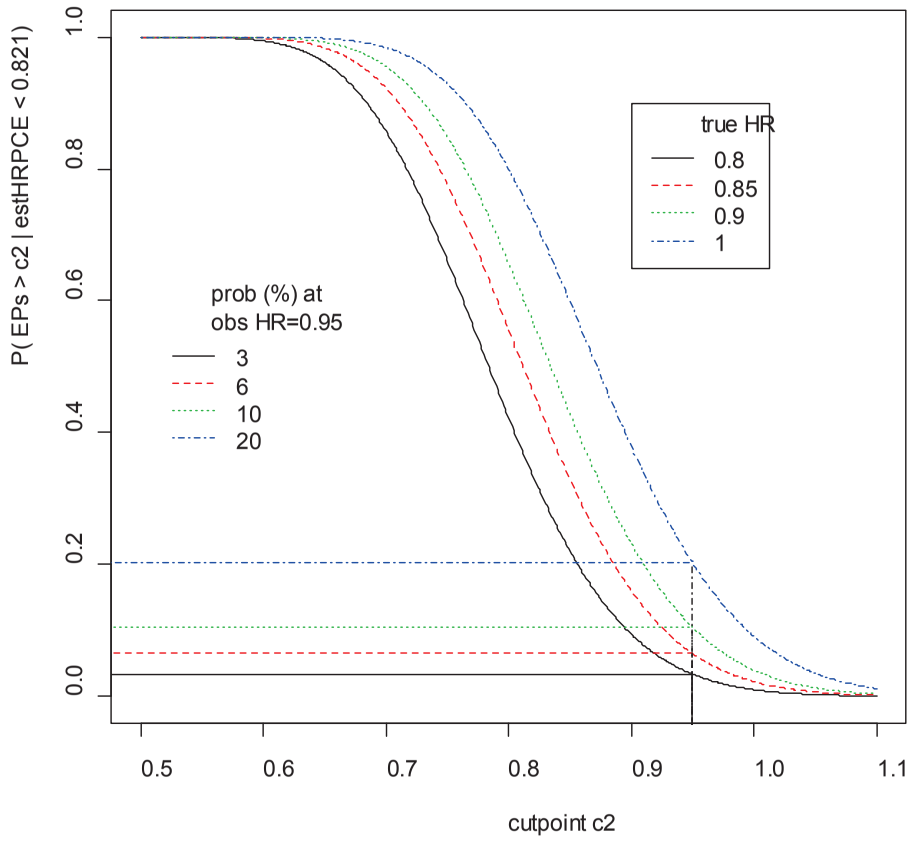


Results: Strokes

EPs = STR



Results: ACM


EPs = ACM
P(EPs > c2 | estHRPCE < 0.821)
0.0
0.2
0.4
0.6
0.8
1.0
0.5
0.6
0.7
0.8
0.9
1.0
1.1
cutpoint c2
true HR
0.8
0.85
0.9
1
prob (%) at
obs HR=0.95
3
6
10
20