

The Dark Side of Composite End Points

Jim Neaton
Minneapolis, Minnesota

Heart Failure 2006
Helsinki, Finland
June 2006



Outline of Talk

- **Background on choosing end points and on composites**
- **Design issues with composites**
- **Reporting issues with composites**
- **Accounting for severity and multiple events**
- **Summary (not all dark and scary)**

Characteristics Desired for Endpoint

- **Clinically relevant and easy to interpret**
- **Easy and reliably diagnosed**
- **Ascertained and classified in an unbiased manner**
- **Sensitive to changes induced by treatment**

General Considerations - 1

More commonly occurring endpoints (high incidence) will result in smaller sample sizes than less frequent events (low incidence) as long as expected relative difference between treatment groups is similar

- CHD death + non-fatal MI vs. CHD death

Continuous response variables usually result in smaller sample sizes than binary or time to event

- BP change vs. % hypertensive

General Considerations - 2

More serious events should be considered along with less serious ones

- **Count CHD deaths along with non-fatal MIs**

Related to this, some events may have to be included as part of the outcome to avoid mis-interpretation

Composite Event (def.)

“An event that is considered to have occurred if any one of several different events or outcomes are observed.”

Meinert CL. *Clinical Trials Dictionary*, 1996.

Combined Endpoint = Composite Event

Usual Rationale for a Composite Endpoint

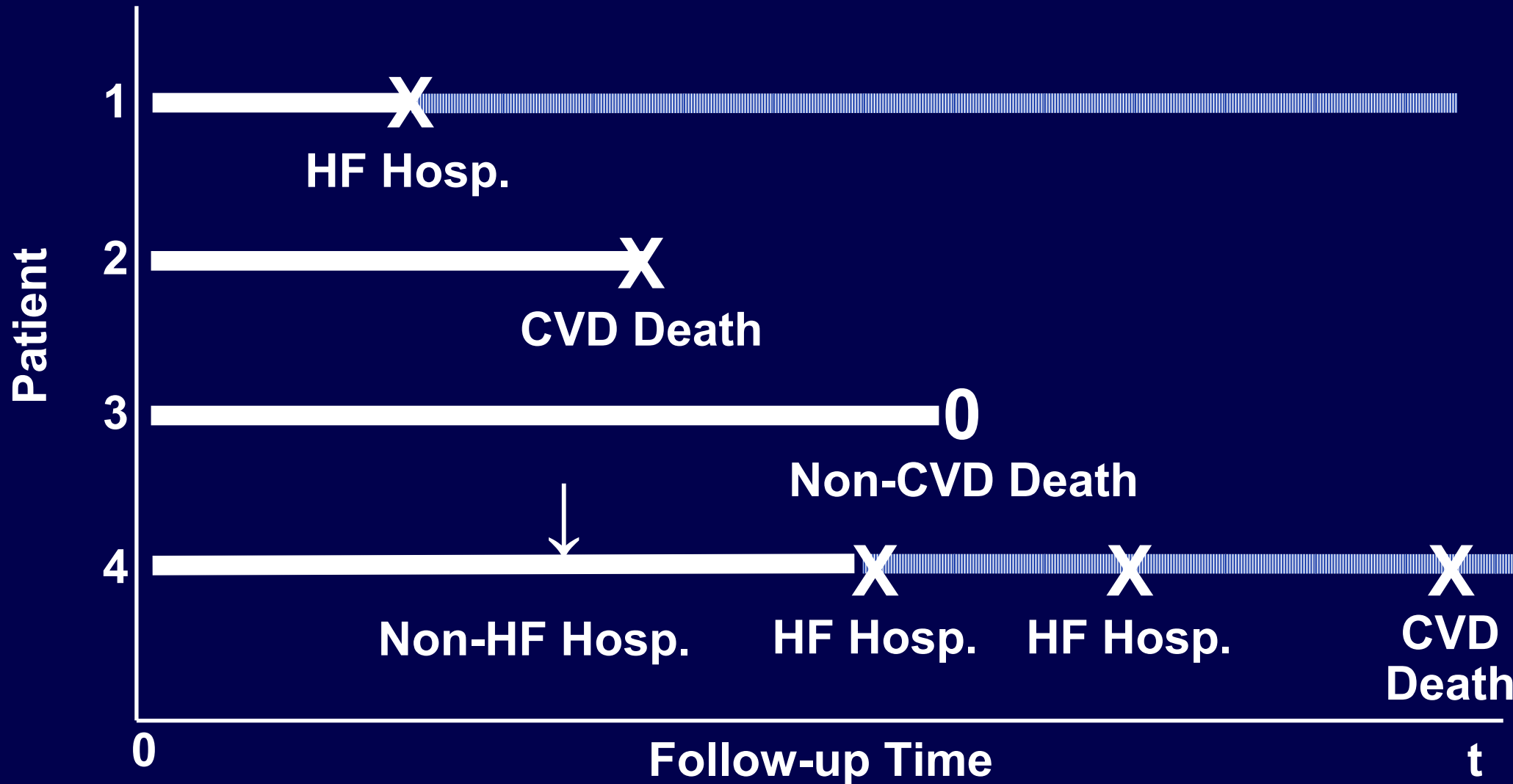
- **More events = greater power (or smaller sample size or shorter trial duration)**
- **Inclusion of some components may reduce/eliminate bias due to informative censoring**
- **A solution to handling disagreement over which outcome should be primary and for dealing with multiplicity issues**

Composite Examples in HF Trials

Time to Event Analysis

- Time to the 1st occurrence of any of the outcomes that are part of the combined endpoint:
 - Death or hospitalization
 - Death or CVD hospitalization
 - CVD death or CVD hospitalization
 - **CVD death or hospitalization for HF**

Composite Example: CVD Death or HF Hospitalization



Composites Are Also Common in Safety Analyses

- **Examples:**
 - Any adverse event
 - Any serious adverse event
 - Any serious, “drug-related” adverse event

Composite Endpoint Cautions

Loss of power if:

- Treatment has little or no effect on some components
- Early events are less likely to represent “treatment failures” compared to later events (Yusuf and Negassa referred to this as “masking” of events)

Unclear interpretation if:

- Components show a different pattern for treatments
- Less serious or more subjectively assessed events are accounting for treatment difference

Neaton JD et al, *Stat Med* 1994 and Yusuf S and Negassa A, *Amer Heart J* 2002.

Design Considerations – Journey to Dark Side Begins

- More events \neq more power
- Excluding some events may bias findings



Number of Events Required to Detect Specified Reduction in Hazard

| <u>Reduction in Hazard</u> | <u>No. Events</u> |
|----------------------------|-------------------|
| 50% | 70 |
| 40% | 120 |
| 30% | 250 |
| 20% | 630 |

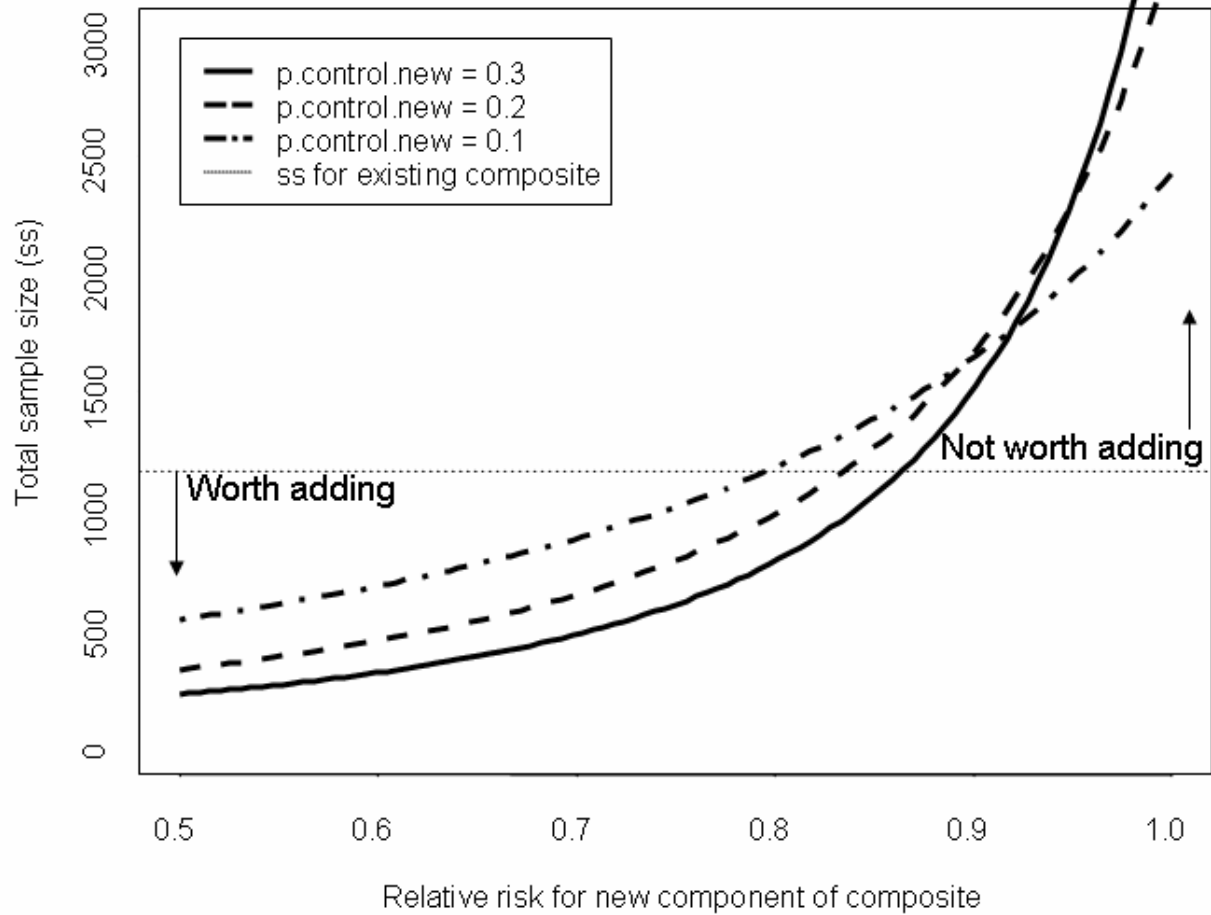
Composite = more events = more power – Not Always

Adding a Component to a Composite Does Not Always Have a Favorable Effect on Sample Size

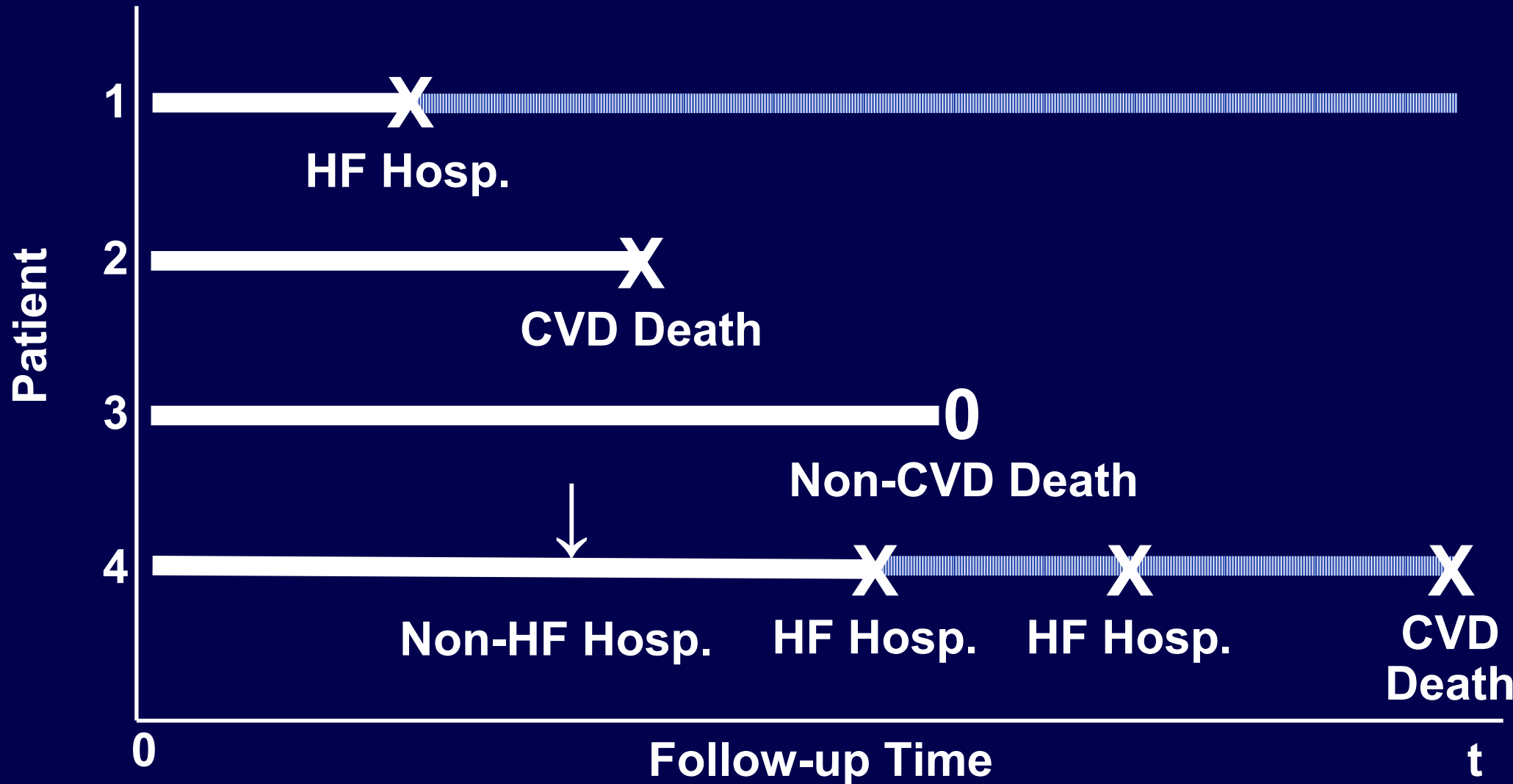
- 10% versus 5% event rate – 1,170 patients total
- Add a new component
 - 30% versus 15% event rate – 330 patients
 - 30% versus 22.5% event rate – 1,450 patients

alpha = 0.05 (2-sided) and power = 0.90

Effect of adding a new component to a composite (existing composite: p.control=0.1, RR = 0.5)



Censoring Non-CVD Death



Informative Censoring

- **If a patient dying from a non-CVD cause would have had a different risk of HF hospitalization (had they survived) than survivors, the censoring is “informative”.**
- **Bias could result if risk of non-CVD death varied by treatment group.**

PICO Trial: Ranked Clinical Outcome at 24 Weeks

| | Assigned Treatment | |
|-----------------------------------|-----------------------|--------------------|
| | Pimobendan (N=209) | Placebo (N=108) |
| Test same/higher than baseline | 132 (63%) | 64 (59%) |
| Test lower duration than baseline | 48 (23%) | 34 (31%) |
| Too sick to undergo exercise test | 5 (2%) | 4 (4%) |
| Died before 24 weeks | 24 (12%) | 6 (6%) |

$P=0.5$ for 63% versus 59%; $P < 0.05$ for difference in exercise duration.

The Price for Avoiding Informative Censoring

Loss of power due to deaths unrelated to treatment

| <u>% 1st Event Deaths</u> | <u>% Deaths Unrelated</u> | <u>Power To Detect HR Adj. for Dilution</u> |
|--------------------------------------|---------------------------|---|
| 66.7 | 50.0 | 0.58 |
| 50.0 | 50.0 | 0.69 |
| 50.0 | 20.0 | 0.85 |
| 30.0 | 50.0 | 0.80 |
| NA | 0.0 | 0.92 |

910 events provide 92% power to detect at 20% reduction in hazard

Recommendations

- **Include more severe events (e.g., death) in composite outcome and adjust power and sample size accordingly.**
- **Include non-CVD mortality as part of mortality outcome unless one is very sure that censoring will not be informative.**
- **If a large fraction of deaths are expected to be CVD, use all-cause mortality as outcome.**

Reporting: Key Points

- All components of the composite must be collected for the duration of the trial in order to do a proper analysis
- All components of the composite should be reported.

Otherwise, the journey to the dark side continues.



Freemantle Guidelines for Reporting

1. **Components of composite outcomes should always be defined as secondary outcomes and reported alongside the results of the primary analysis, preferably in a table.**
2. **Ensure that the reporting of composite outcomes is clear and avoids the suggestion that individual components of the composite have been demonstrated to be effective.**
3. **Systematic overviews and quantitative meta-analysis should be used to identify the effects of treatments on rare but important endpoints that may be included as part of composite outcomes in individual trials.**

Freemantle N, et al. JAMA 2003.

Guide to Interpreting Composite End Points

- 1. Are the component end points of similar importance to patients?**
- 2. Did the more and less important end points occur with similar frequency?**
- 3. Is the underlying biology of the component end points similar?**
- 4. Are the point estimates of the relative risk reduction similar and the confidence intervals sufficiently narrow?**

Montori VM et al, BMJ 2005.

Reporting of Composite Outcomes

- How often did each component contribute to composite outcome (descriptive)?
- What is the relative hazard for each component of the composite - the separate number of events and rate for each component (“Consumer Reports approach”)?

Reporting a Composite Outcome – EPHESUS Trial (N Engl J Med 2003)

No. of Patients with Event

| <u>Primary Outcomes and Components</u> | <u>Eplerenone (N=3,319)</u> | <u>Placebo (N=3,313)</u> | <u>HR</u> |
|--|---------------------------------|------------------------------|-------------|
| CVD Death or Hosp. | 885 | 993 | 0.87 |
| - CVD Death | 407 | 483 | 0.83 |
| - CVD Hosp. | 606 | 649 | 0.91 |
| - MI | 224 | 229 | 0.97 |
| - Heart failure | 345 | 391 | 0.85 |
| - Stroke | 70 | 51 | 1.34 |
| - Ventricular Arrhythmia | 52 | 54 | 0.95 |

Non-CVD Mortality in EPHEBUS

| | No. of Patients with Event | | HR |
|----------------------------------|----------------------------|----------------------|------|
| | Eplerenone (N=3,319) | Placebo (N=3,313) | |
| All cause ⁺ mortality | 478 | 554 | 0.85 |
| CVD mortality | 407 | 483 | |
| Non-CVD mortality | 71 | 71 | |

⁺ Co-primary endpoint

The Dark Side of Reporting

- We used a five-point scale...to grade adverse events occurring while the patient was taking study drugs and during the eight weeks after their permanent discontinuation”.
- “All analyses were performed according to intention to treat”.

NEJM 1996;335:1099-1106.



Pitfalls of Usual Approach

- Usual analysis of composites focuses on 1st event
- Components of composite usually vary in severity and in impact on quality of life
- Many patients experience multiple events

Approaches for considering multiple events of varying severity need to be studied.

Weighting the Components of Composite Outcomes

- **Risk of death associated with different components**
- **Rank-ordering of outcomes in terms of severity and quality of life by clinicians and patients**
- **Rating the entire event profile**

General Approaches for Accounting for Severity of Events and Event Histories

- **Ranking of entire event histories (Follmann et. al., 1992)**
- **Marginal models with ranking of events according to risk of death or subjective ranking by clinicians and/or patients (Neaton et.al.,1994)**
- **Rule based ranking (Bjorling and Hodges, 1997)**
 - **Severity, timing, number**

Severity, Timing, Number Rule

- **Death > non-fatal MI**
- **MI on day 30 > MI on day 100**
- **Stroke and MI > MI alone**

Considerations in Analysis of All Events

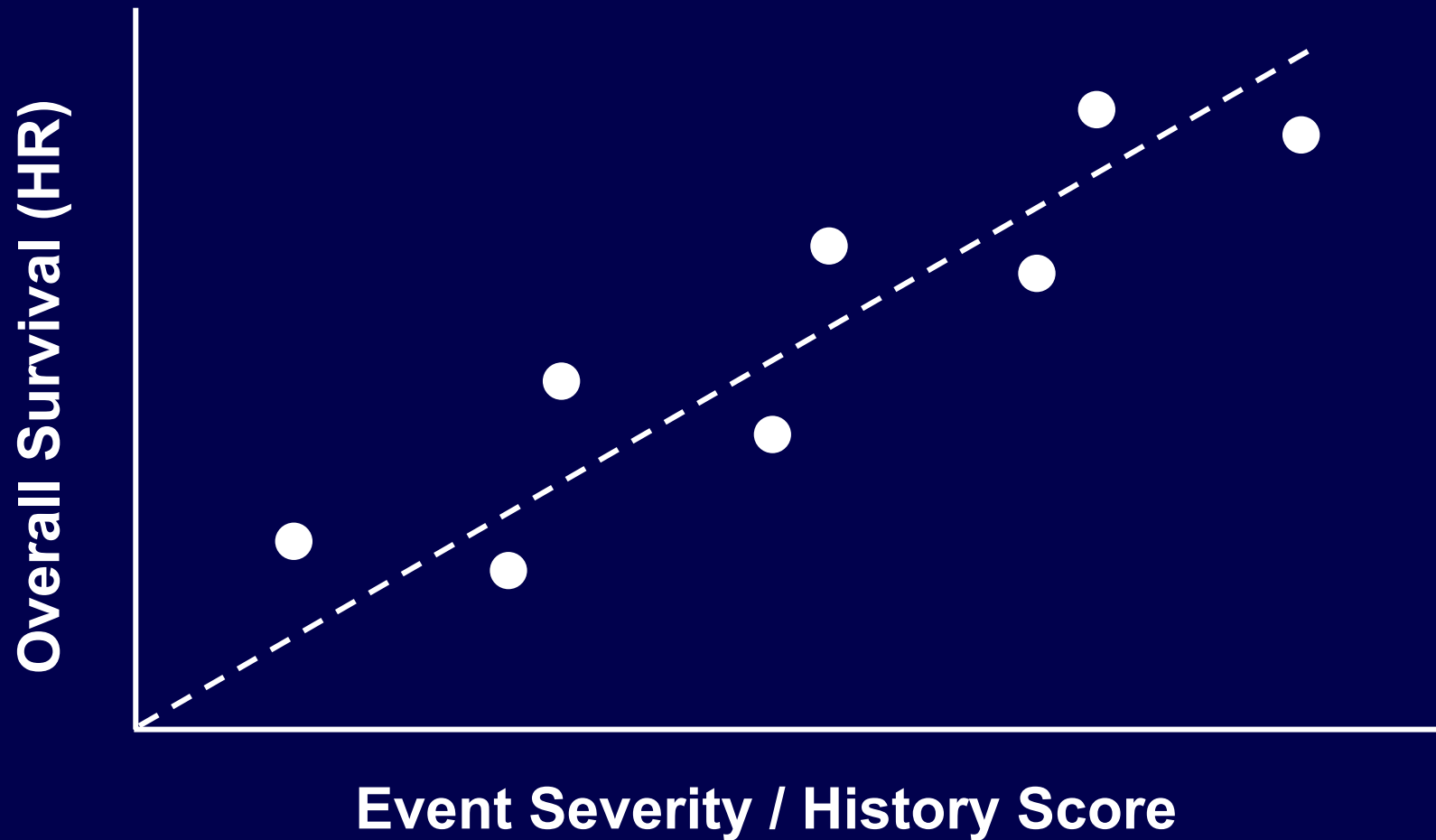
- Events are not independent – SE's have to be adjusted
- 2nd, 3rd ... events may not add much to signal from 1st event
- A loss of power could result if treatment was modified after 1st event

Composite Scoring System for A-HeFT: Possible Score -6 to +2

| <u>Variable Scored</u> | <u>Criteria</u> | <u>Score</u> |
|--|--|--------------|
| Death | Death from any cause during trial | -3 |
| | Alive at end of trial | 0 |
| Hospitalization | At first hospitalization for heart failure | -1 |
| | No hospitalization for heart failure | 0 |
| Change in quality of life at 6 months (MLHF) | Reduction by -10 = markedly improved | +2 |
| | Reduction by -5 to -9 = improved | +1 |
| | Changed by -4 to +4 = no change | 0 |
| | Increased by +5 to +9 = worsened | -1 |
| | Increased by +10 = markedly worsened | -2 |

Other Examples

- Braunwald's "weighted unsatisfactory-outcome" – hierarchical with patient receiving score for event with highest score (*Circulation*, 1992 and *Amer J Cardiol*, 1993)
- Packer's ordinal composite score (improved, unchanged, or worse) (*J Cardiac Failure*, 2001)
- Cleland's "patient journey" score that takes account of symptoms and survival, e.g., symptom adjusted days alive (*Euro J HF*, 2002).



Recommendation: use completed trials to study relationships between survival and novel scoring systems

Composites Versus Other Approaches

Advantage

Disadvantage

Single
outcome

Simple

Sample size; multiple
endpts are a reality

Composite

Sample size

Interpretation not easy
if components show
different patterns

Co-primary
outcomes

Eggs not all in
one basket

Sample size and power

Global index

Power

Not easily interpretable

Hierarchical
scoring

Power; clinical
relevance

Clinical relevance

Summary

Multiple outcomes are a reality, and combining them often makes sense.

Composites do not have to be scary!



Some References

Yusuf S, *American Heart Journal*, 2002

Lubsen J, *Stat Med*, 2002

Freemanthe N, *JAMA*, 2003

Montori VM, *BMJ*, 2005

Neaton JD, *Stat Med*, 1994

Neaton JD, *J Card Fail*, 2005

Collaborators

Gerry Gray

Bram Zuckerman

Marvin Konstam

Daniel Kussow (Grandson as Darth Vader)