Issues and Considerations in Composite Endpoints

Greg Soon

Division of Biometrics IV/Office of Biostatistics/OTS/CDER/FDA

BASS XIII, Savannah, Georgia, November 8, 2006

Disclaimer

- Views expressed here are of the presenter and not necessarily of the FDA

Outline

- Examples and General Considerations
- Relationship of Composite and Components
- Validation of Composite Endpoint
- Decisions Rule and Multiple Comparison
- Concluding Remarks

$E = e(E_1, E_2, \ldots, E_c)$

Composite Endpoint: Definition

- Composite or combined endpoints are defined as the combination of component (singleton) endpoints
  - Each singleton has clinical significance in its own right

- If E_1, E_2,…, E_c are clinically relevant endpoints, and together they uniquely determine

$$E = e(E_1, E_2, \ldots, E_c)$$

  Then E is a composite endpoint with component endpoints E_1, E_2,…, E_c.

- It is a single measure of effect from a combined set of different variables

Types of Composite Endpoint

I. Patient level total score or index derived from multiple item scores or counts
   - HAMD total in depression trials
   - ACR20 in rheumatoid arthritis trials
   - Quality of Life

II. Event rate after a certain period of treatment or follow-up, where the event is the occurrence of any of one event from a given set of events
   - In organ transplant patients, failure is defined as the occurrence of any of the three events: biopsy-proven acute rejection, graft loss, or death

III. Time to the first event among a set of events
   - In HIV trials, time to virologic rebound or disease progression

IV. Others
   - "Information Preserving" composite endpoints. Two success/failure type endpoint can be joined together to form a single partially ordered multi-categorical endpoint: FF<SF, FS<SS

Example 1: Histologic Responder for the Chronic Hepatitis B Trial

- Liver biopsy have 4 components (Knodell Score):
  - I. Periportal injury (0,1,3,4,5,6,10)
  - II. Lobular injury (0,1,3,4)
  - III. Portal inflammation (0,1,3,4)
  - IV. Fibrosis (0,1,3,4)

  The larger values indicate more severe conditions

- A patient is a responder if it meets two conditions:
  1) Reduction of Knodell Necroinflammatory Score (I+II+III) by 2 or more points from baseline
  2) No Worsening on fibrosis score (IV)

Example 1 Continued

- The composite endpoint is more sensitive than the key component

Number of Patients	Study 437 (HBeAg+) Adefovir 30 mg	Study 437 (HBeAg+) Adefovir 10 mg	Study 437 (HBeAg+) Placebo	Study 438 (HBeAg-) Adefovir 10 mg	Study 438 (HBeAg-) Placebo
Improvement †	**99 (60%)**	**89 (53%)**	**41 (25%)**	**78 (64%)**	**20 (35%)**
Knodell fibrosis scores					
Improved	30 (20%)	26 (17%)	15 (10%)	19 (17%)	5 (9%)
Same	112 (76%)	115 (76%)	118 (79%)	92 (81%)	43 (77%)
Worsened	5 (3%)	11 (7%)	16 (11%)	2 (2%)	8 (14%)

† Histological improvement

Example 2: Prevention of Flu

- For the prevention of influenza, the primary endpoint is the infection rate at the end of the trial, which is defined as
  - Have flu symptoms (itself a composite endpoint based on fever, coughing, etc.)
  - Confirmed by the positive viral assay
- Each Component is considered clinical meaningful, but may not be suitable as primary endpoint because
  - "Flu symptoms" can be caused by other disease
  - Viral infection without clinical manifestation is less important
- But the composite endpoint is considered clinically meaningful and is used for the primary endpoint

Example 2 continued
Composite vs. Components

		Virology Positive?		
N= 1685		Y	N	Total
Symptoms?	Y	1.4	8.7	10.0
	N	1.7	88.2	90.0
	Total	3.1	96.9	100%

Placebo

		Virology Positive?		
N= 1678		Y	N	Total
Symptoms?	Y	0.2	8.7	9.0
	N	2.1	88.5	90.6
	Total	2.3	97.3	100%

Drug

Example 3: Congestive Heart Failure (CHF) trial example

- E1= CHF-related death, E2 =CHF-related hospitalization
- E = E1 “union” E2 means either the event E1 occurred or E2 occurred or both occurred
- E is called a composite event endpoint whose “component” or “singleton” endpoints are E1 and E2. One can easily generalize this definition to more than 2 endpoints
- For convenience, one usually counts first event (or the worst event) for each patient

Example 4: Treatment of Hospitalized Flu Patients

- For treatment of influenza infected patients requiring hospitalization:

1. Resolution of either fever and/or cough (and no worsening if not completely resolved), and
2. Discharge from hospital

Example 5: Rheumatoid Arthritis

- ACR20 Response
  - 20% improvement in tender joint count
  - 20% improvement in swollen joint count
  - Plus 20% improvement in 3 out of 5 of:
    - Patient pain assessment
    - Patient global assessment
    - Physician global assessment
    - Patient self-assessed disability
    - Acute phase reactant

Motivations for using a composite event endpoint as a primary endpoint

- Can reduce the size of the trial
  - If components in the composite increase the number of events (non-overlapping property)
  - If components individually have similar treatment effects or jointly increase it (homogeneity property)
- Illustrative example (CHF 2-arm trial)
  - Control 18 month mortality rate = 18%, $\alpha = 0.05$ (2-sided), power = 90%, delta = 12% reduction in mortality. These assumptions give trial size = 12,653 patients
  - Add CHF-related hospitalizations: suppose control rate increases to 36%, delta remain 12%, α and power the same. These assumptions give trial size = 5,032 patients

Motivations for using a composite event endpoint as a primary endpoint (cont'd)

- Can address to a broader aspect of a multifaceted disease. For such a disease, a result in an isolated endpoint can be misleading.
- Can change the focus of the trial from discovering a large treatment effect to a clinically meaningful small treatment effect (Aggregate small effects on components to larger effect on composite)
- Can combine "soft" components that have more frequent events with "hard" components that occur infrequently.

Some key considerations in considering a composite event endpoint as a primary

- Clinically relevant, interpretable, and regulatory acceptable
- Prospectively defined – endpoint and also its components
- Components "add" to the total treatment effect and are "sensitive" in inducing treatment effects in the same direction
- Endpoint ascertainment methods are well established for capturing accurately for both the occurrence and non-occurrence of the events
- Display of component endpoint results along with the composite endpoint result

Relationship of Composite and Components

- The characteristic of a composite endpoint will depend on its construction and the relationships among its components
- Overlapping or coincidence endpoints do not add value
- Disparate or independent endpoints are most efficient in increasing the overall event rate for Type II composite endpoints, but need to make sure it is clinically meaningful

Relationship: Two Binary Endpoints

- Assume the outcomes of Y_1 and Y_2 are 0 or 1 for each patient

$$Y = Y_1 Y_2 = \begin{cases} 1 & \text{if } Y_1 = 1 \ \& \ Y_2 = 1 \\ 0 & \text{Else} \end{cases}$$

- Let $A=\{Y_1=1\}$, $B=\{Y_2=1\}$
- Type II composite endpoint when consider its complement

$$1 - Y = (1 - Y_1) + (1 - Y_2) - (1 - Y_1)(1 - Y_2)$$

or

$$(\mathrm{A} \cap B)^C = \mathrm{A}^C \cup B^C$$

Relationship: Two Binary Endpoints

Y_1 \ Y_2	1 (B)	0	Total
1 (A)	p_{11}=p (C)	p_{10}	p_{1+}=p_A
0	p_{01}	p_{00}	p_{0+}
Total	p_{+1}=p_B	p_{+0}	1

Parameter Space

$$0 \le p_A, p_B \le 1$$

$$0 \le p \le \min(p_A, p_B)$$

$$p_A + p_B - p \le 1$$

Dependence Measures

	Range	Indp
Conditional Probability: $\rho = P(A \mid B) = \frac{p}{p_B}$	$0 \sim 1$,	p_A
Pearson Correlation: $\gamma = \frac{p - p_A p_B}{\sqrt{p_A(1-p_A)p_B(1-p_B)}}$	$-1 \sim 1$,	0
Odds Ratio: $\pi = \frac{p\,(1 - p_A - p_B + p\,)}{(p_A - p\,)(p_B - p\,)}$	$0 \sim +\infty$	1

Relationship: Continued

The response rate of the composite can be expressed as

$$p = p_{11} = \rho p_B = p_A p_B + \gamma \sigma_A \sigma_B$$

$$= \sqrt{\frac{1}{1-\pi} p_A p_B + \frac{1}{4}\left(\frac{1}{1-\pi} - p_A - p_B\right)^2} - \frac{1}{2}\left(\frac{1}{1-\pi} - p_A - p_B\right)$$

Where $\sigma_A = \sqrt{p_A(1-p_A)}, \sigma_B = \sqrt{p_B(1-p_B)}, \sigma = \sqrt{p\ (1-p\ )}$

- The stronger the correlation, the larger the response rate
  - Smaller response rate in case of "either or" type definition
  - Does not mean larger or smaller effect size

Relationship: Continued

- The effects on the composite could be driven by the effects on the components, as well as by any differences in the correlation of the two component endpoints between the treatment groups.
- In general, significance effect on composite endpoint does not guarantee any effect on any of the components, unless certain assumption are made on the consistency of correlations

Relationship: Continued

Dependence measures between composite and components

$$Cov(A,B) = p - p_A p_B,$$

$$Cov(A,C) = p(1-p_A), Cov(B,C) = p(1-p_B)$$

$$corr(A,B) = \gamma$$

$$corr(A,C) = \sqrt{\frac{p}{1-p}\frac{1-p_A}{p_A}}, corr(B,C) = \sqrt{\frac{p}{1-p}\frac{1-p_B}{p_B}}$$

Composite always correlate positively with its components

Relationship: Observed Response Rate

Assuming we have n patients in a treatment arm, then the observed response rates are

$$\hat{p}_A \sim (p_A, \frac{p_A(1-p_A)}{n}), \hat{p}_B \sim (p_B, \frac{p_B(1-p_B)}{n}), \hat{p} \sim (p, \frac{p(1-p)}{n})$$

$$Cov(\hat{p}_A, \hat{p}_B) = \frac{1}{n} Cov(A, B),$$

$$Cov(\hat{p}, \hat{p}_A) = \frac{1}{n} Cov(C, A), Cov(\hat{p}, \hat{p}_B) = \frac{1}{n} Cov(C, B)$$

$$Corr(\hat{p}_A, \hat{p}_B) = Corr(A, B),$$

$$Corr(\hat{p}, \hat{p}_A) = Corr(C, A), Corr(\hat{p}, \hat{p}_B) = Corr(C, B)$$

Relationship: Testing Statistic

Assuming we have a two arm superiority trial. Let *T* and *C* stands for the testing and control arm.

Treatment effects are typically measured either by difference or odds ratio. Consider the case of using difference

$$\hat{\Delta} = \hat{p}_T - \hat{p}_C, \hat{\Delta}_A = \hat{p}_{AT} - \hat{p}_{AC}, \hat{\Delta}_B = \hat{p}_{BT} - \hat{p}_{BC}$$

$$\hat{\Delta}_k \sim (p_{kT} - p_{kC}, p_{kT}(1-p_{kT})/n_T + p_{kC}(1-p_{kC})/n_C), k = ", A, B$$

$$Cov(\hat{\Delta}_A, \hat{\Delta}_B) = Cov(\hat{p}_{AT}, \hat{p}_{BT}) + Cov(\hat{p}_{AC}, \hat{p}_{BC}) = Cov(AT, BT)/n_T + Cov(AC, BC)/n_C$$

$$Cov(\hat{\Delta}, \hat{\Delta}_A) = Cov(\hat{p}_T, \hat{p}_{AT}) + Cov(\hat{p}_C, \hat{p}_{AC}) = Cov(CT, AT)/n_T + Cov(CC, AC)/n_C$$

$$Cov(\hat{\Delta}, \hat{\Delta}_B) = Cov(\hat{p}_T, \hat{p}_{BT}) + Cov(\hat{p}_C, \hat{p}_{BC}) = Cov(CT, BT)/n_T + Cov(CC, BC)/n_C$$

Relationship: Testing Statistic

- If sample size is large then the three effects sizes are jointly normally distributed.
- One can obtain the conditional distribution of the effect sizes on the components given the effect size on the composite endpoint
- If one regards standard errors of the observed effect sizes as known, then it is possible to derive conditional probability for statistical significance for the components given observed significance on the composite
- In general it could be analytically difficult to derive the formulas, computation intensive methods may be used for calculations

Relationship Among the Testing Statistics

In general, assume the testing statistic for the composite is $\hat{T}$, the testing statistics are $\hat{T}_A$ and $\hat{T}_B$ for the two components, respectively. Assume positive numbers indicate better response, and 0 indicate no effect. Assume the critical values for rejection are C_α, A_α and B_α.

Define

$$A(c,a) = \Pr(\hat{T}_A > a \mid \hat{T} > c),$$

$$B(c,b) = \Pr(\hat{T}_B > b \mid \hat{T} > c),$$

$$AB(c,a,b) = \Pr(\hat{T}_A > a, \hat{T}_B > b \mid \hat{T} > c),$$

Relationship Among the Testing Statistics

Let $A(\alpha) = A(C_\alpha, 0),\ B(\alpha) = B(C_\alpha, 0),\ AB(\alpha) = A(C_\alpha, 0, 0),$

These quantities describes how likely the component endpoints will be in the right direction given the significance on the composite

Let $A(\alpha, \alpha_A) = A(C_\alpha, A_{\alpha_A}), B(\alpha, \alpha_B) = A(C_\alpha, B_{\alpha_B}),$

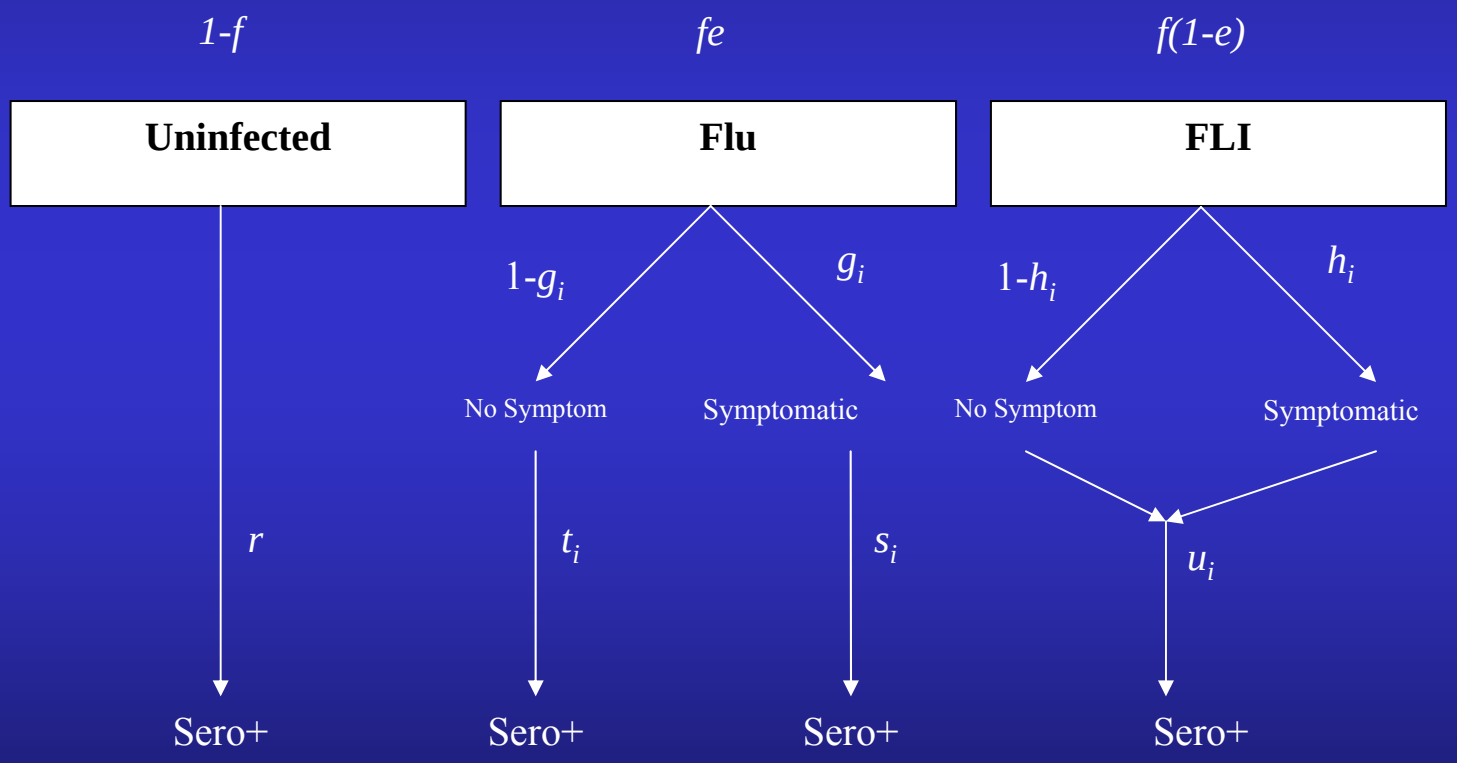
$AB(\alpha, \alpha_A, \alpha_B) = AB(C_\alpha, A_{\alpha_A}, B_{\alpha_B})$

These quantity describes how likely the component endpoints will also be statistically significant given the significance on the composite

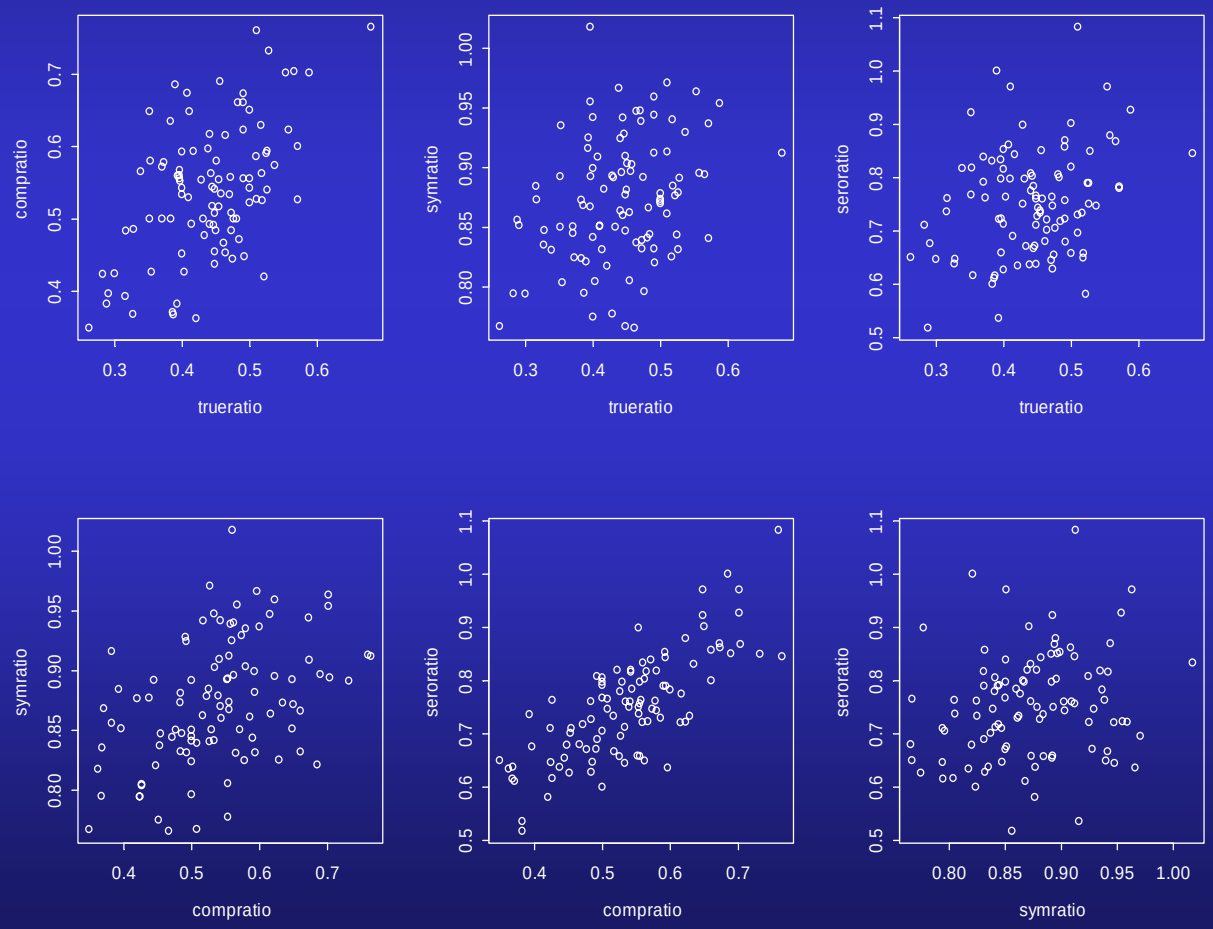
A Simulation for Example 2

- Primary endpoint is infection, defined by symptoms of flu, with confirmation by serology to rule out any flu-like illnesses.
- Simulation based on an assumed mechanism:
  - Only a fraction of infecting virus that cause flu-like symptoms are influenza
  - A fraction of subjects are expected to be invaded by the virus
  - Treatments will impact the emergence of flu-like symptoms
  - Viral assay have reasonably good but not ability of identifying inflenza virus

A Simulation for Example 2, (Continued)



A Simulation for Example 2, (Continued)



Simulation Continued

- The response rates on the composite could be smaller than both components, yet it is more powerful
- The noise in the trial can make the relationship between the composite and components weak

Validation of Composite Endpoint

- Ideally, a significant results on the composite should provide confidence to the significance of the key components or key sub-composite endpoints
- The quantities defined in the last few slides provide such measures within each trial. These represents the probability of a positive or significant effect sizes on the components given the significance on the composite, if the same trial were repeated infinitely many times
- One should also examine the probability of falsely claiming significance on components when the composite is significant, and is mainly driven either by the other component or difference in correlations

Validation: Meta Analysis

- Similar to the validation of surrogate endpoint, the validity of the composite can be assessed across many trials
- The observed effects and responses on the composite and components should be consistent

Considerations in defining "Win"

- Is the composite a measure of the disease (individual components do not fully measure the disease) or is it for convenience of analysis?
  - Sparse events
  - Competing risk
  - Multiplicity
- Are the events surrogates for other events or surrogates for something else?
  - CV events are an outcome of underlying disease
  - Liver fibrosis may lead to carcinoma

Considerations in defining "Win"

- How to interpret components?
  - Significant in one and weak or worse in others
  - None significant, but all in right direction
  - Should you analyze components individually?
  - The composite did not show significance but a key component showed high statistical significance
- Need to think these issues before trial starts and define the "win" scenarios properly

Statistical testing strategies for the composite event endpoint and its components

Testing Strategy 1:

- Test the composite endpoint at level α (e.g., $\alpha = 0.05$, 2-sided)
- Summarize the results of the components using descriptive statistics. Pre-declared intention of not claiming any benefit for a component

Disadvantage:

Composite endpoint = mortality + morbidity

P-value for the composite = 0.085

P-value of the mortality endpoint = 0.028

✓ No claim for the composite or the mortality endpoint. Why?

Statistical testing strategies for the composite event endpoint and its components (cont'd)

Testing Strategy 2:

- Test the composite endpoint at level α (e.g., $\alpha = 0.05$, 2-sided)
- If the null hypothesis is rejected then test the components in a pre-specified fixed sequence at the same level α

Example:

Composite endpoint = mortality + morbidity

P-value for the composite = 0.049

P-value of the mortality endpoint = 0.028

✓ Result for the composite and also for the mortality endpoint if the mortality endpoint was tested first in the sequence

Statistical testing strategies for the composite event endpoint and its components (cont'd)

Testing Strategy 3:

- Test the composite endpoint at level α (e.g., $\alpha = 0.05$, 2-sided)
- If the null hypothesis is rejected then test the components using Bonferroni method controlling FWER at level α

Example:

Composite endpoint = mortality + morbidity

P-value for the composite = 0.049

P-value of the mortality endpoint = 0.028

✓ Result for the composite but not for the mortality endpoint

Statistical testing strategies for the composite event endpoint and its components (cont'd)

Testing Strategy 4 (Fallback method):

Composite endpoint = mortality + morbidity

- Test the composite at a slightly reduced significance level of $\alpha = 0.04$ (save 0.01)
- If significant, then test the mortality endpoint at the full significance level of $\alpha = 0.05$
- If not significant then test the mortality endpoint at the reduced significance level of $\alpha = 0.01$
- Example: p (composite) = 0.50, p (mortality) = 0.009

✓ Result for the mortality endpoint

Issue of "soft" components in a composite event endpoint

Example (Cardiovascular Trial):

- Composite event = (fatal/non-fatal MI, fatal/non-fatal stroke, revascularization, unstable angina)
- Hard components: fatal/non-fatal MI, and fatal/non-fatal stroke - clinically convincing, unambiguous ascertainment of events)
- Soft components: revascularization and unstable angina – clinically less convincing and subject to differences in clinical practice

Statistical testing strategies for a composite with "hard" and "soft" components

- Sub-composite approach: sub-composite includes only hard components. Example:
  - Test the full composite at level α_1 less than α (e.g., $\alpha_1 = 0.01$)
  - If significant, then test the sub-composite at level $\alpha = 0.05$, else, test it at level $\alpha - \alpha_1 = 0.04$. (Fallback method)
- The weighting method: pre-determine "weights" with sum of the weights equal to one (e.g., harder components can get 3 times larger weights than the softer components)
  - Advantage: avoids multiplicity
  - Disadvantage: clinicians may have difficulty in assigning weights

Improving the Power by taking advantage of Correlations

- Composite, sub-composite, and component endpoints are usually positively correlated by construction.
- Alpha allocation can be chosen to improve power while controlling Type-I error

Summary

- Multiple ways of constructing composite endpoints
- Relationship of composite and its components can be complicated
- How effects on the composite influence the components should be studied both within the trial as well as through meta-analysis
- Proper decision rule need to be pre-specified and proper statistical adjustment planned to control the type-I error rate.