Use of Composite Endpoints in Clinical Trials

Richard Entsuah

FDA Industry Workshop 2018

Merck Research Labs

MERCK

Outline

- Background
- Definition of Composite Endpoint (CE)
- Advantages and Disadvantages
- Win Ratio with an Example
- Multiplicity Issues
- Weighting
- Conclusions

Goal of the presentation

To provide an overview of the use of Composite Endpoint by discussing the advantages, disadvantages and when to use this concept in clinical development.

Background

- ICH E9 Guidelines state: “a relevant and important treatment benefit cannot be always be achieved by evaluating a single event endpoint especially if this event type occurs with a low frequency.”

- Combining several types of events (Composite Endpoint) results in an increase of the number of expected events which tend to enlarge the overall treatment effect.

- In cardiology research, apart from ‘**Death’**, clinical events like

  a) Non-fatal myocardial infarction

  b) Cardiovascular hospitalization

  c) Non-fatal stroke

  also are of clinical interest and are included in the composite endpoint. Usually, for these studies the composite endpoint is defined as a time-to-event variable, or as a binary or continuous variable.

Definition

- Composite endpoint combines several components of interest into a single outcome variable.
- Patients who have experienced any of the events specified by the components are considered to have experienced the composite outcome.
- We shall consider examples of composite endpoints in Neuroscience, Allergy and Cardiovascular therapeutic areas.

Neuro-TA: Recent use of ADCOM

Scale	Item name
ADAS-cog	Delayed word recall Orientation Word recognition Word finding difficulty
MMSE	Orientation time Drawing
CDR-SB	Personal care Community affairs Home and hobbies Judgement and problem solving Memory Orientation
ADAS-cog, Alzheimer's Disease Assessment cognitive subscale (Neurodegeneration 2016 J, Wang)	CDR-SB; Clinical Dementia Rating sum of boxes Mini Mental State Exam -MMSE

Allergy Study

Treatment of rhino-conjunctivitis usually involves:

- Reduction in symptoms in first pollen season
- Reduction in need for pharmacologic treatment in first season

Primary endpoint is derived from two

Key Secondary Endpoints including

DSS, DMS

Total Combined Score (TCS) =

Daily Symptom Score (DSS) +

Daily Medication Score (DMS)

Primary Endpoint Allergy Studies

- Total Combined Symptom and Medication Score (TCS)

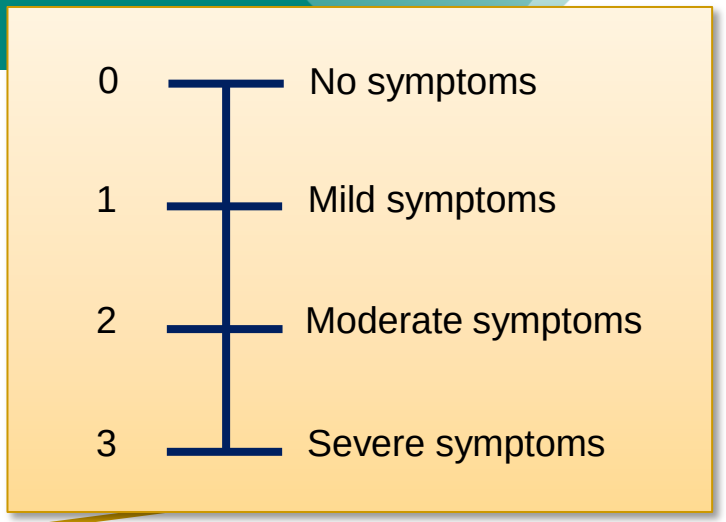
TCS=DSS + DMS

- The WAO (World Allergy Organization) recommends to adjust the reported symptom score to account for the pharmacotherapy used in immunotherapy trials[1]
- The EMA (European Medicines Agency) states that the primary endpoint for immunotherapy trials should reflect both symptoms and pharmacotherapy[2]
- The FDA (Food and Drug Administration) accepts combined symptom and medication scores as the primary endpoint[3]

[1] World Allergy Organization. Position Paper 2009.
[2] EMEA Doc.Ref. CHMP/EWP/18504/2006.
[3] R. Raben. 2011; PEI seminar.

MERCK
Be well

Symptom Efficacy Measure



Daily Symptom Score (DSS)

Individual Symptoms	Score	Maximum Daily Score
Runny nose	0-3	3
Blocked nose	0-3	3
Sneezing	0-3	3
Itchy nose	0-3	3
Gritty feeling/red/itchy eyes	0-3	3
Watery eyes	0-3	3
Total		**18**

Rescue Medication Efficacy Measure

Daily Medication Score (DMS)

Medication Allowed in Merck Trials	Score/Dose Unit	Maximum Daily Score
Loratadine 10 mg tablet	6 points/tablet	6
Olopatadine HCl 0.1% ophthalmic solution	1.5 points/drop	6
Mometasone furoate nasal spray 50 µg	2 points/spray	8
Prednisone 5 mg tablet	1.6 points/tablet	16
Total		**36**

The maximum total score for the TCS is 18 + 36 = 54

Properties of Composite Endpoints

- Individual components of the composite should be clinically meaningful and of similar importance to the patient
- Expected effects on each component are similar, based on biological plausibility
- The clinically more important components of the composite should at least not be negatively affected

Properties of Composite Endpoints

- Avoids the need to choose a single primary endpoint when many may be of equal importance
- Improves understanding of the effect of an intervention by avoiding competing risk
- Composite endpoints have greater power to detect small but consistent differences that may occur over extended periods of time

Advantages of CEs

- Statistical Precision and Efficiency (increased Power)
- Reduces sample size requirement/Less Cost
- Results of promising treatments will be available earlier

Advantages attributed to Composite Endpoints

- In longitudinal studies, the primary advantage of a composite endpoint is that they may facilitate interpretation, as the number of pairwise comparisons is reduced.
- This advantage may be challenged by difference in directions in one or all the components of the composite endpoint or could have transient differences at specific time points.

Disadvantages of Composite Endpoints

- In early stages of clinical development, trials with more exploratory objectives of identifying which components, domain or subscales are impacted by the disease or particular therapy will not benefit from the use of composite endpoints.
- Observed treatment differences between groups from a composite endpoint may be misleading or lead to a conclusion that all the components contribute equally to that difference, when in reality only one of the components could be the driving force.

Disadvantages attributed to CEs

- Interpretation could be problematic when component endpoints are dissimilar in patients importance – ‘Qualitative Heterogeneity’
- Interpretation could be problematic if either the rates or relative risk reduction vary appreciably across components – ‘Quantitative Heterogeneity’
- Excessive influence of the more subjective (clinical-driven) component outcomes
- Alpha error must be adjusted to draw confirmatory conclusions about the components

Disadvantages of Composite Endpoints

- Calculation of power poses a challenge as the assumed effect sizes of the intervention depends on the effect sizes of the individual components and their correlations may not be known, as good historic data may not be available.
- The “net measure or effect” may not reflect the influence of the new intervention. The effect of some individual components may be very different in magnitude and point in different directions.

Win Ratio

- Consider a cardiovascular trial comparing a standard treatment to a new active treatment with composite endpoint as CV death and Hospitalization for cardiac heart failure (HF hosp.).
- Clinical Relevance: CV Death more important variable than HF Hosp.
- Hence comparing any two patients on new and standard treatment, we determine whether either one had CV death before the other.
- If this is not known, then we determine which of these two patients had HF Hosp. first.
- Risk score or risk stratification is used to select matched pairs of patients on new and standard treatment.

Matched Pair Approach:

- Form matched pairs of patients on new and standard treatments taking into account individual patient risk set.
- For each pair, study the major event (CV death) first.
- Each matched pair fits into one of the following five categories.

Classifications

Highest Importance CV death:

(a) New treatment patient had CV death first » N_a

(b) Standard treatment patient had CV death first » N_b

2nd Highest Importance HF hosp:

(c) New treatment patient had HF hosp first » N_c

(d) Standard treatment patient had HF hosp first » N_d

(e) None of the above (Ties) » N_e

Category (e) mostly comprises of pairs with neither patient having CV death or HF hosp, but for a few pairs, one may have an event but the other's follow-up time was shorter.

Win Ratio cont.

Define

$N_b + N_d = N_w$ as the number of 'winners' for the new treatment (those matched pairs where the standard treatment performed worse)

$N_a + N_c = N_L$ as the of number of 'losers' for the new treatment

The Win Ratio is defined as

$$\mathbf{R_w = N_w / N_L}$$

Win Ratio Test Statistics

Note that $p_w = N_w / (N_w + N_L)$ is the proportion of winners with

95% CI of $p_w \pm 1.96 \{ p_w(1-p_w) / (N_w + N_L) \}^{1/2}$

(p_L, p_U) where p_L is the lower CI and p_U is the upper CI

Thus $\mathbf{R_w = p_w / (1- p_w)}$

with 95% CI as $\{ p_L / (1-p_L) , p_U/(1- p_U) \}$.

For significance test

$$\mathbf{Z = (p_w - 0.5) / [[p_w (1 - p_w) / (N_w + N_L)]^{1/2}}$$

$$\mathbf{\sim N (0,1)}$$

The Unmatched Approach

We compare every patient on a new treatment with every patient on standard treatment, each time noting who 'won'.

Let N_n and N_s be the number of patients on new and standard treatments.

There are N_n x N_s paired comparisons.

Each pair is classified into one of the following categories

a), b), c), d), or e) based on CV death and HF hospitalization.

b) and d) are winners for the new treatment while

a) and c) are "losers"

In this set up $\mathbf{N_a + N_b + N_c + N_d + N_e = N_n \times N_s}$

$\mathbf{N_b + N_d = N_w}$

$\mathbf{N_a + N_c = N_L}$

The **Win Ratio** is define as $\mathbf{R_w = N_w / N_L}$

Unmatched Approach

Note that N_n x N_s unmatched pairs are not independent comparisons. Since each patient on a new and standard treatment are used N_n and N_s times, respectively.

Pocock et. al.(2012) recommend the matched pair approach provided a pre-defined basis for matching exit.

EMPHASIS –HF Trial

Eplerenone vs Placebo in patients with NYHA class II heart failure

- Ejection fraction ≤ 35%, N=2737,
- Median follow-up per outcome is = 21 months
- There were 1364 patients on eplerenone and 1373 on placebo.
- Groups are equal-sized by randomly removing 9 patients from the placebo group.
- The 1364 in each group are risked matched using their **risk scores**. Patients on each treatment are ranked by their risk scores.
- From top-rank to bottom-rank, each eplerenone patient is paired with the same –ranked placebo patient.

Zannad F et.al. & the EMPHASIS Study Group -2011

Results

Composite primary outcome CV death or HF Hospitalization for heart failure

Eplerenone	18.3%
Placebo	25.9%

Hazard Ratio = 0.63 with

95% CI of (0.54, 0.74), p < 0.0001

Hospitalizations tend to occur first, hence any impact of eplerenone on CV mortality gets lost in the Composite Endpoint.

Components Results for the HF Trial

	Components	Matched Pairs	Matched Pairs Time Stratified	Unmatched Pairs
(a)	**CV death on eplerenone first**	90	105	124825
(b)	**CV death on placebo first**	118	148	163129
(c)	**HF hosp on eplerenone first**	61	61	86127
(d)	**HF hosp on placebo first**	131	137	175606
(e)	**None of the above**	964	913	1323085
	Win Ratio for Composite	**1.65**	**1.72**	**1.61**
	95% CI	1.35, 2.03	1.42, 2.09	1.37, 1.89
	Z-Score	5.05	5.81	5.45
	Win ratio for CV death only	1.31	1.41	1.31

Multiple Testing for Composite Endpoints

- In addition to analysis of composite endpoint an evaluation of the individual components is required (CPMP, 2002). This assessment has to be done with care as the components of the composite endpoint often define ‘competing risks’.

Multiple Comparisons

- A possible confirmatory claim would be to show superiority for the composite and non-inferiority for all components. This constitutes an **intersection-union** test problem that does not require an adjustment of the significance level for the individual hypothesis (Berger, 1982).
- The simultaneous rejection of all individual null hypotheses is a strong claim resulting in a considerable loss of power compared to the single hypothesis testing for the composite alone.

Multiple Comparisons

- A more realistic approach could be to prove non-inferiority for as many components as possible once superiority for the composite has been demonstrated.
  - This situation corresponds to a hierarchical test procedure where the null hypothesis for the components are only tested if H_o^{CE} can be rejected at one-sided level α.
  - The same overall α can then also be used in the second step for the assessment of the components (Wiens and Dmitrienko, 2010).
- A test procedure like the Bonferroni-correction (for small # of components) which controls the experiment-wise α is applied.

Weighted Combined Effect Measure

- In general we can define a composite endpoint for the i^{th} subject as a summary of the data over J components.

$$S_i = \sum_{i=0}^{J} w_j \; f(Yij)$$

This assigns higher weights w_j to the more important components with the intention that an opposite effect in a relevant component (e.g. “death”) is less likely to be masked by a large effect in a component of secondary importance.

Summarizing across Domain or Subscale:

Some weightings options are the following:

- Weights proportional to the number of questions for each component
- Equal weights for each component
- Factor analytic weights
- Weights derived from patient preference measure
- Statistically derived weights (e.g. correlation among the components)

There is implicit weighting of components that is dependent on the number of items.

Concluding Remarks

- The composite endpoint should be associated with the primary objective of the trial.
- Correct (a prior) identification of the primary composite endpoint can increase the statistical power of the trial.
- Components of composite outcomes should be defined as secondary outcomes and reported alongside the results of the primary analysis.
- Appropriate multiplicity adjustments need to be performed.
- Assignments of weights of similar importance to each component need to be done cautiously, as there is heterogeneity in patient's preferences due to age, race and household income. (Joshua M Strolker et. al.; 2014).
- The Win Ratio prioritizes the major components of the composites.

References

1. Nick Freemantel, Melanie Calvert, John Wood, Joanne Eastaugh, Carl Griffin. Composite Outcomes in Randomized Trials Greater Precision But With Greater Uncertainty? JAMA, May 21, 2003 – Vol 289, No 19
2. Geradine Rauch, Bernhard Rauch, Svenja Schuler, Meinhard Kieser. Opportunities and Challenges of clinical trials in cardiology using composite primary endpoints. World J Cardiol 2015 January 26; 7(1):1-5
3. Gloria Cordoba, Lisa Schwartz, Steven Woloshin, Harold Bae, Peter C Gotzsche. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review: BMJ 2010;341:c3920 doi:10.1136/bjm.c3920
4. Stuart J. Pocock, Cono A. Ariti, Timothy J. Collier, and Duolao Wang. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities:; European Heart Journal (2012) 33, 176-182
5. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH harmonized tripartite guideline: statistical principles for clinical trials. Stat Med. 1999; 1905-1942

References

7. Ignacio Ferrera-Gonzalez, Gaieta Permanyer-Miralda, Jason W. Busse, Dianne M. Bryant, Victor M. Montori, Pablo Alonso-Coello, Stephen D. Walter, Gordon H. Guyatt Methodological discussions for using and interpreting composite endpoints are limited, but still major concerns: Journal of Clinical Epidemiology 60 (2007) 651-657

8. Dawson JD, Lagakos SW. (1991) Analyzing laboratory marker changes in AIDS clinical trials. Journal of Acquired Immune Deficiency Syndromes, 4;: 667-676

9. Dawson JD, Lagakos SW. (1993) Size and power of two-sample test of repeated measures data. Biometrics, 49: 1022-1032

10. Dawson JD. (1994) Stratification of summary statistics tests according to missing data patterns. Statistics in Medicine, 13: 1853-1863

11. Omar PZ, Wright EM, Turner RM, Thompson SG. (1999). Analyzing repeated measurements data. A practical comparison of methods. Statistics in Medicine, 18: 1587-1608

12. O'Brien PC.(1984) Procedures for comparing samples with multiple endpoints. Biometrics, 40: 1079-1087

References

13. Cox DR, Fitzpatrick R, Fletcher AI, Gore SM, Spiegelhalter DJ, Jones DH. (1992) Quality-of-life assessment: Can we keep it simple? Journal of the Royal Statistical Society A, 155: 353-393

14. Ware JE, Borrk RH, Davies AR, Lohr KN. (1981) Choosing measures of health status for individuals in general populations. American Journal of Public Health, 71: 620-625

15. Farivar SS, Cunningham WE, Hayes RD, (2007) Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey,V.1.Health and Quality of Life Outcomes, 5: 54-61