Chapter 7

Introduction to Composite Endpoints

This chapter is the first of seven consecutive chapters that applies the multiple analysis methodology that we have developed thus far to specific, complex circumstances that commonly occur within modern clinical trials. Both this and the next chapter focus on the use of the composite or combined endpoint as a primary analysis variable. Composite or combined endpoints are defined as the combination of component (singleton) endpoints, each of which has clinical significance in its own right. In this chapter, the complications involved in the construction of the composite endpoint are discussed, and the issue of homogeneity vs. heterogeneity of treatment effect is addressed.

7.1 Introduction

Composite or combined* endpoints have been incorporated into the design of many clinical trials over the past thirty years. The use of these complicated endpoints has expanded both in scope and in complexity as investigators have become more accustomed to their features. A well-designed clinical trial that prospectively embeds a composite endpoint into its primary analysis plan is empowered to measure small effects. The use of the combined endpoint improves the resolving

* The terms “combined endpoint” and “composite endpoint” are synonymous and will be used interchangeably.
ability of the clinical trial, strengthening its capacity to pick out weaker signals of effect from the background noise of sampling error. If larger effect sizes are of interest, then a trial using a composite endpoint can gauge the effect of therapy using a smaller sample size (everything else being equal).

However, the entry of a composite endpoint into a clinical trial introduces that trial to complications in both endpoint construction and endpoint interpretation, complexities that can weaken the trial’s ability to reach reliable conclusions. In some circumstances, the combined endpoint can be exceedingly difficult to analyze in a straightforward, comprehensible manner. The components of the endpoint, if not carefully chosen, may produce a conglomerate endpoint that is off balance. The medical community’s resultant difficulty in understanding the meaning of this unequilibrated endpoint can cast a shadow over the effect of the clinical trial’s intervention. This can reduce what appeared as a stunningly successful demonstration of clinical and statistical efficacy to merely the demonstration of tepid and ultimately irrelevant effectiveness against an endpoint that in the end was seen to be of dubious clinical value.

7.2 Definitions and motivations
While our ultimate focus in this chapter will be on the complications that composite endpoints add to the problem of multiple statistical analyses in clinical trials, we will begin our discussion with an introductory overview of the combined endpoint, assessing its strengths and weaknesses.

A combined or composite endpoint in a clinical trial is a clinically relevant endpoint that is constructed from combinations of other clinically relevant endpoints, termed component endpoints or singleton endpoints. Two examples of singleton endpoints are 1) the cumulative incidence of total mortality and 2) the cumulative incidence of total hospitalization. A patient experiences a combined endpoint based on these two singleton endpoints if they either die or are hospitalized. If a patient experiences either or both of these component events during the course of the clinical trial, then that patient is considered to have experienced the composite endpoint, commonly referred to as total mortality/total hospitalization or total mortality + total hospitalization.

7.3 Notation
We can introduce some elementary notation to further clarify the constitution of the composite endpoint. Consider a composite endpoint that is composed of the two singleton endpoints A and B. Then as stated earlier, the composite endpoint occurs if either the event A or the event B has happened. We can denote this occurrence as \( A \cup B \) where \( \cup \) is called “union”. \( A \cup B \) (said as “A union B”) denotes the occurrence of either the event A, the event B, or both events. Simply put, \( A \cup B \) means that at least A or B has occurred. If A denotes the occurrence of a death during the course of a trial and B denotes the occurrence of a hospitalization during the course of the trial, then \( A \cup B \) accurately describes the combined event of at least a death or a hospitalization. Thus, the union event is precisely the combined endpoint.
7.4 Theoretical motivations for the use of combined endpoint

There are theoretical and practical motivations that guide investigators as they consider the use of a composite endpoint as a primary analysis variable in a clinical trial. Each of these motivations must be considered in turn so that we might gain some insight into how to prospectively construct a functional combined endpoint for such a study.

It is a truism that disease, and certainly chronic disease, manifests itself in different ways. As an example, consider congestive heart failure (CHF). CHF can produce death – CHF also increases the likelihood of hospitalization, as well as prolonging it. Congestive heart failure impairs the patient’s ability to exercise, and reduces that patient’s quality of life. In addition, CHF is associated with chronic effects on measures of cardiac function including but not limited to left ventricular ejection fraction, end systolic volume, end diastolic volume, stroke volume, cardiac output, and blood pressure. If the investigator wishes to attempt to measure the effect of an intervention in alleviating the signs and symptoms of congestive heart failure, which of these measures should she use?

As we saw in chapter four, the investigators who design a clinical trial attempt to choose a measure of disease that the intervention will positively affect. However, considering the many possible signs and symptoms of congestive heart failure, this choice can seem to be an impossible one for investigators to make, even after following the analysis triage tactic reviewed in chapter four. Alternatively, by building a combined endpoint from several of the signs and symptoms of congestive heart failure outlined above, the investigators can simultaneously focus on several manifestations of the disease process. Thus, the use of the combined endpoint can represent an earnest attempt by the investigators to construct a “whole” of the disease’s varied effects that may be greater than the “sum” of the combined endpoint’s components.

7.4.1 Epidemiologic Considerations

Additionally, epidemiologic assessments of singleton endpoints reveals that the isolated interpretation of a single component endpoint can be misleading. As an example, consider the correct interpretation of a clinical trial that is prospectively designed to examine the effect of an intervention on the occurrence of myocardial infarctions (MI). There is one prospectively identified primary analysis in this study, and that is the effect of therapy on the cumulative incidence rate of nonfatal MI. The experiment is concordantly executed, and, at its conclusion, the study demonstrates both a clinically significant and a statistically significant reduction in the nonfatal MI rate.

In this illustration, the randomly allocated intervention reduced the occurrence of nonfatal heart attacks. However, the intervention may not be as effective as it first appeared. By focusing solely on a nonfatal endpoint, the
investigators might miss the possibility that the intervention may have produced a harmful effect on another measure of this same disease—one that was not captured by the primary analysis of the effect of therapy on nonfatal MI. For example, it is possible that the therapy reduced the incidence of nonfatal heart attacks by increasing the incidence of fatal heart attacks (Figure 7.1). That is, even though the number of nonfatal heart attacks was reduced in the active group of the clinical trial, the total number of heart attacks was increased in the active group, and the majority of these events were fatal heart attacks.

![Figure 7.1](image.png)

**Figure 7.1.** The active group converts more nonfatal MI’s to fatal ones and therefore produces the misleading result of fewer MI’s.

Because the intervention’s influence on mortal events may be hidden if the principle analysis involves the measurement of only a morbidity endpoint, the morbidity endpoint can be combined with the mortality endpoint to provide a more complete depiction of the effect of therapy.

### 7.4.2 Sample Size Concerns

An additional motivation for the use of the composite endpoint is to insure that there is adequate power for the primary analyses of the study. Combining component endpoints permits their endpoint rates to be accumulated, and this increased event rate can be translated into a reduction in the minimum number of patients required for the clinical trial.

Recall from Appendix 6 that one of the critical factors included in the sample size formula is the control group endpoint event rate. The larger this rate is, the greater the number of endpoint events that will be accumulated in the study. Thus, if all other assumptions remain constant, we find that the greater the probability of an endpoint, the smaller the number of subjects that will be required...
to produce an adequate number of those endpoint events. It is this relationship that is taken advantage of in the construction of a composite endpoint.

Consider an investigator interested in studying the effect of a new therapy on coronary heart disease (CHD) death. His preliminary data suggest that the annual event rate in the control group of the study for the population of patients he is planning to recruit is 0.015. The investigator will be able to follow these patients for five years. Thus, the five year cumulative incidence rate of CHD death is 0.073.* The investigator believes that the randomly allocated intervention will reduce the cumulative event of CHD death by 20%. A computation† reveals that the required minimum number of patients required for this clinical trial, assuming 90% power and a two sided type I error probability of 0.05 is 12,132 patients, to be divided equally between the two groups (Table 7.1).

Table 7.1 Sample Size computation as a function of the composite endpoint

<table>
<thead>
<tr>
<th>Primary Analyses</th>
<th>Annual Event Rate</th>
<th>Cumulative Control Group Event Rate</th>
<th>Efficacy</th>
<th>Alpha (two-tailed)</th>
<th>Power</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHD Death</td>
<td>0.015</td>
<td>0.073</td>
<td>0.20</td>
<td>0.05</td>
<td>0.90</td>
<td>12132</td>
</tr>
<tr>
<td>CHD Death + Nonfatal MI</td>
<td>0.025</td>
<td>0.119</td>
<td>0.20</td>
<td>0.05</td>
<td>0.90</td>
<td>7092</td>
</tr>
<tr>
<td>CHD Death + Nonfatal MI + Unstable Angina</td>
<td>0.040</td>
<td>0.185</td>
<td>0.20</td>
<td>0.05</td>
<td>0.90</td>
<td>4260</td>
</tr>
</tbody>
</table>

Sample size decreases for combined event rates with larger annual event rates

This is a large sample size, and the investigator believes he will be unable to recruit and follow this number of patients for five years. However, if this investigator were to combine with CHD death the event of survival and nonfatal myocardial infarction, then the cumulative event rate of this combined endpoint will include not just CHD death but also the cumulative event rate for the nonfatal MI component. The cumulative annual event rate is 0.025 for this combined analysis, leading to a cumulative five year event rate of 0.119 and a sample size of 7092.

The investigator can take the additional step and add yet another singleton endpoint to this combined endpoint to gain a further reduction in the sample size. If the investigator was to consider patients who survived the study and did not experience a nonfatal MI but did experience unstable angina pectoris during the

* If the annual event rate is $r$, then the cumulative event rate over $y$ years is $1 - (1 - r)^y$. The reasoning behind this formula is that one first computes the probability of no events in five years or $(1 - r)^y$ and then computes the probability of at least one event in five years, which is $1 - (1 - r)^y$.

† The formula for this computation is provided in Appendix 6.
course of the trial as a third component endpoint, then the annual event rate for this triple composite endpoint of CHD death + nonfatal MI + unstable angina increases to 0.040. The five year cumulative incidence rate for this new composite endpoint is 0.185, and the resultant sample size is 4260 patients. By choosing a combined endpoint, the investigator was able to decrease the sample size of the trial from 12,132 to 4,260 solely by adding component endpoints to the composite endpoint. This reduction in the clinical trial’s sample size produces a clinical trial requiring fewer resources to execute.

In the assembly of the composite endpoint, each singleton endpoint contributes an event rate that is included in the event rate of the combined endpoint. Thus with each new component endpoint, the sample size will decrease because the event rate increases.

### 7.4.3 Improved resolving power

Another advantage of the well considered use of a composite endpoint derives from the increase in the number of patients with events that the trial introduces. In the previous section we saw how the increase in the control group event rate in a clinical trial decreases the sample size required for the study. One other perspective on this multifaceted sample size computation is that the larger event rate provides a more sensitive test of the effectiveness of the therapy.

As an illustration of this principle, consider an investigator interested in designing a clinical trial to detect the effect of an intervention for the treatment of heart failure. The intervention is relatively safe and is free (e.g. a change in lifestyle). The investigator is interested in demonstrating that the effect of the randomly allocated intervention in patients with heart failure will lead to a 10-12% reduction in heart failure clinical consequences. He chooses as the single primary analysis for this trial the effect of therapy on the cumulative incidence rate of CHF mortality. He anticipates being able to randomize no more than 4000-5000 patients for this study.

The investigator estimates an eighteen month CHF mortality rate will be 15% in the control group. Assuming 80% power and a two sided type I error of 0.05, the original sample size for this research effort reveals that if will take 4066 patients to demonstrate a 20% reduction in the actively treated group. However, the investigator believes that in this population, a 10-12% reduction is the minimal clinical threshold that is worth detecting. In order to be able to detect this low level of efficacy with any statistical reliability, 11,731 patients must be recruited. However, if the investigator chooses to add to the CHF mortality the clinical event of CHF hospitalization, the number of events experienced in the placebo group will increase. In fact, if the cumulative control group event rate for this new combined endpoint of CHF death/CHF hospitalization is 34%, a sample size of 4094 patients will identify a 12% reduction in events that is due to the therapy. The ability of a sample size of approximately 4100 patients to identify clinically significant but smaller levels of therapy efficacy has been improved with the use of the combined endpoint (Figure 7.2).
7.6 Component endpoint coherence

The two advantages of adding a composite endpoint are broadening the measure of the therapy’s effect on the disease, and decreasing the required sample size or increasing the trial’s ability to detect smaller efficacy levels. However, these advantages must be weighed against the problems with incorporating a combined endpoint into a clinical trial.

7.5 Properties of combined endpoints

We have just defined a combined endpoint. To appreciate the complications of embedding a composite endpoint into the structure of a clinical trial, we must develop the properties of combined endpoints. These properties can be divided into 1) coherence, 2) endpoint equivalence, and 3) therapy homogeneity.

7.6 Component endpoint coherence

There are many manifestations of chronic disease, and as we have pointed out, investigators who wish to study the effect of a therapy on the occurrence of that disease using a combined endpoint must choose the component endpoints from among these signs and symptoms. Some manifestations may measure common clinical and pathophysiologic correlates (e.g., fatal and nonfatal myocardial infarction)—others measure disparities (e.g., intraocular pressure and popliteal nerve conduction velocities in patients with diabetes mellitus). The component endpoints
which make up the combined endpoint must be coherent – they should measure the same underlying pathophysiologic process and be consistent with the best understanding of the causes of the disease the investigators wish to study. Consideration of coherence requires an examination of the degree to which different component endpoints may measure related pathology. A balance must be struck between coincidence and separation of the singleton endpoints.

7.7 Coincidence

Since the combined endpoint is “built up” from the contributions of at least two and many times several singleton endpoints, the clinical relevance of the composite endpoint derives from the clinical meaning of its component endpoints from which it was constructed. Each component endpoint must measure not just the same disease process, but the same underlying pathophysiology. When each component endpoint is measuring the progression of the same pathology, then the investigator can be assured that the component endpoint is measuring the process which is best understood to excite the production of the disease’s manifestations.

However, the component endpoints should not be so closely related that a patient is likely to experience all of them. These types of component endpoints we will term coincident endpoints. If a patient experiences one of two coincident endpoints, they are likely to experience the other. In this situation, there is no real advantage in using the combined endpoint instead of one of its component endpoints. Constructing component endpoints that are too interrelated will make the combined endpoint redundant.

For example, consider a clinical trial will study the effect of a randomly allocated medication in patients with diabetes mellitus. In this trial, patients will undergo a baseline blood sugar evaluation, receive the study medication (active or placebo), and then be followed for six months, at which time a second blood sugar evaluation will take place. The investigator could choose as the principle analysis the effect of the intervention on the combined endpoint of reduction in fasting blood glucose + reduction in glycosylated hemoglobin. There is no doubt that these two component endpoints measure the same underlying pathophysiology, i.e. the presence of glucose in the blood. A fasting blood glucose measurement reports the current level of blood sugar, a level that is transient and changes from day to day. The glycosylated hemoglobin level evaluation provides a more stable measure of blood sugar levels over approximately three months.

The randomly allocated therapy being evaluated in the trial is likely to reduce each of these measures of plasma glucose over the course of the study. At the conclusion of the study, patients who experience important reductions in their elevated blood sugar are also likely to experience reductions in their glycosylated hemoglobin. Thus, patients who experience one component endpoint are likely to experience the other. The events of reduction in blood sugar and reduction in glycosylated hemoglobin, while measuring the same underlying pathophysiology (abnormalities in carbohydrate metabolism) are measuring “too much of the same thing”.

Even though they should measure the same underlying pathophysiology, the component endpoints of a composite endpoint should be different enough that a
patient can experience either of the component endpoints, not just both. We might express this formally as follows in the case of a combined endpoint which is constructed from two component endpoints A and B each of which have incidence rates (e.g. total mortality or total hospitalizations). Recall that we defined the composite endpoint as the occurrence of either A, B, or both and denoted that event as the union of component endpoint A and component endpoint B. We wrote the combined endpoint as $A \cup B$.

Since the clinical trial will randomize a fixed number of subjects and follow them for a pre-specified period of time, measuring the new occurrence of events, the event rates the trial will measure are incidence rates. We can use the probability of the occurrence of the endpoints in the placebo group as a representation of the placebo group incidence rate. Thus, $P[A \cup B]$ is the incidence rate of the composite endpoint over the course of the trial. From elementary probability we can write

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$  \hspace{1cm} (7.1)

where $P[A]$ and $P[B]$ are the component event incidence rates for the singleton endpoints A and B. $P[A \cap B]$ is the incidence of the joint occurrence of the two of the component endpoints A and B. Recall also from elementary probability that

$$P[A | B] = \frac{P[A \cap B]}{P[B]}$$  \hspace{1cm} (7.2)

This can be written as $P[A \cap B] = P[A | B]P[B]$. Substituting this result into equation (7.1) reveals


$$= P[A] + P[B] - P[A | B]P[B]$$  \hspace{1cm} (7.3)

$$= P[A] + P[B](1 - P[A | B])$$

This last formulation directly links the probability of the occurrence of the combined endpoint to the probability of each of the component endpoints and the conditional probability of the occurrence of the component A given component endpoint B has occurred. When the $P[A | B] = 0$, the composite endpoint event rate $P[A \cup B]$ reaches its maximum value, i.e. the sum of the incidences rates for component endpoints A and B. What type of events must A and B be so that the $P[A | B] = 0$ and the cumulative incidence rate $P[A \cup B]$ reaches its maximum value?
7.8 Mutual exclusivity and disparate events

In probability, when we consider the properties of events, a useful observation to make is whether the occurrence of one event excludes the occurrence of another event. We describe such joint events which cannot occur together (or for these purposes, cannot occur in the same patient during the course of the trial) as mutually exclusive events. As an illustration of this property, consider a clinical trial in which patients who have died are categorized by their cause of death. In this example let there be only two possible causes of death; cardiovascular death, or non-cardiovascular death. If a patient is judged to have died a cardiovascular death, then that patient cannot have died a non-cardiovascular death. We say that the event of dying from a cardiovascular event and the event of dying a non-cardiovascular death are mutually exclusive. The occurrence of one of these events excludes and makes impossible the occurrence of the other event.

When component endpoints A and B are mutually exclusive, then their joint occurrence is impossible and \( P[A \cap B] = 0 \). Thus, the probability that component endpoint A has occurred, given that endpoint B has occurred is zero (i.e. \( P[A | B] = 0 \)). This implies that \( P[A \cup B] = P[A] + P[B] \). When events are mutually exclusive the cumulative event of the union of events is simply the sum of the cumulative incidence rate of the composite endpoints. Thus, when the component endpoints cannot occur in the same patient, the largest combined endpoint event rate is attained. Mutual exclusivity of component events directly translates into the smallest sample size for the trial.

It is now easier to think through the problem with coincident component endpoints. If the component endpoints are very coincident, then the conditional probability \( P[A | B] \) is close to one, and the cumulative event rate for the combined endpoint is only marginally larger than the incidence rate for the component endpoint A. The larger the conditional probability \( P[A | B] \), the more coincident the component endpoints become. This in turn leads to a lower frequency of occurrence of the combined endpoint (i.e. smaller \( P[A \cup B] \)) and a larger sample size for the clinical trial.

7.9 The problem with mutual exclusivity

The prior discussion suggests that the selection of mutually exclusive component endpoints would be advantageous in constructing a combined endpoint. However, the difficulty with the use of mutually exclusive singleton endpoints is that they can measure different characteristics of the same disease that physicians are unaccustomed to linking together. This can produce serious problems in the interpretation of the results. Even though the choice of mutually exclusive component endpoints minimizes the required sample size for an evaluation of the effect of the intervention on the combined endpoint, care should be taken to assure that the component endpoints are not too disparate.

As an example of the problems that disparate component endpoints can produce, consider a clinical trial that will measure the effect of a randomly allocated intervention on the signs and symptoms of diabetes mellitus. The
combined endpoint for this trial is blood sugar levels > 300 mg/dl + the occurrence of peripheral neuropathy. This is a difficult composite endpoint to defend. There is no doubt that diabetes mellitus produces both elevations in blood glucose and peripheral neuropathy. However, the two events are not closely linked. Changes in blood sugar can be acute, while the development of peripheral neuropathy is chronic, appearing after many years of exposure to the complex metabolic derangements produced by the disease. It is difficult to make clinical sense out of a combined measure of two manifestations of a disease that themselves are not very clearly linked together pathophysiologically. If the component endpoints become too disparate, it can become very difficult to describe exactly what the combined endpoint is measuring that is of direct clinical relevance. Choosing component endpoints, several of which are likely to occur in the same patient, may not produce the combined endpoint that leads to the smallest sample size but it can make the trial’s results much easier to interpret.

The occurrence of multiple component endpoints in the same patient during the course of the trial, however, admits a possible problem with the use of combined endpoints that the trial planners must overcome. Since it is possible that a patient can experience each of the components of a combined endpoint, care must be taken to insure that the patient is not considered to have reached the combined endpoint multiple times. Consider a clinical trial in which the prospectively defined primary analysis is the effect of therapy on the combined endpoint of total mortality + hospitalization due to congestive heart failure. It is possible, (perhaps even likely) that a patient who meets the inclusion/exclusion criteria of the study could experience a hospitalization for CHF (perhaps experience multiple distinct hospitalizations for CHF) and then subsequently die. In the analysis of this endpoint, even though the patient has satisfied the criteria for the combined endpoint more than once, that patient can only be counted as having reached this endpoint once.* Commonly, the prospective determination is made that the first time the patient reaches the primary endpoint of the study, the patient is considered to have reached the endpoint. In our example, the patient who suffered multiple hospitalizations for CHF and subsequently died during the hospitalization is considered to have reached the primary endpoint upon the first hospitalization.

7.10 Balancing the separation

Component endpoints in clinical trial are commonly not mutually exclusive – patients can experience combinations of the singleton endpoints which make up the component endpoint. However, the component endpoints of a composite endpoint

* This does not mean that in a randomized clinical trial in which patients are expected to be followed for five years, a patient who reaches one nonmortal component of the prospectively defined combined endpoint early in their follow-up should not be followed for the duration of the study. Post event measurements, which include the occurrence of adverse events and the possible occurrence of secondary endpoints subsequent to the occurrence of the primary endpoint are two of many reasons why patients should continue to be followed until the end of the study.
should be contributory and coherent — they must make sense. Each of the components endpoints should measure the same underlying pathophysiology, but be different enough that they add a dimension to the measurement of the disease process that has not been contributed by any other component endpoint.

As an example of a combined endpoint whose component endpoints reflect a balance of distinct component endpoints, consider the design and results of the Cholesterol and Recurrent Event (CARE) trial[1]. CARE evaluated the effect of the HMG-CoA reductase inhibitor pravastatin on the reduction in morbidity and mortality in patients at risk of developing atherosclerotic disease. The CARE trial recruited 4159 patients with a history of recent myocardial infarction and with low density lipoprotein (LDL) cholesterol levels between 115 mg/dl and 174 mg/dl. These patients were randomized to either standard care or standard care plus pravastatin 40 mg once a day. The prospectively chosen primary endpoint that was the only primary analysis of this study was fatal coronary heart disease (CHD) + nonfatal myocardial infarction (MI). Each of the two component endpoints of this composite endpoint is an important manifestation of the same atherosclerotic cardiovascular disease process. Each singleton endpoint is in its own right an important clinical manifestation of ischemic heart disease.

After randomization, patients were followed for a median duration of time of five years. During that time, the investigators worked to ensure that the investigators assigned to determine whether any of the singleton endpoints had been reached were blinded from knowing whether the patient had been assigned active therapy or placebo therapy. In CARE, this also meant that investigators were not to receive information about the patient’s plasma lipid levels. Treatment guidelines based on lipid levels were provided by the trial’s coordinating center to a matched patient. This matched patient received study medication in addition to mask any additional therapy required by a patient at the same clinical center in the opposite randomized group.* From the results of CARE [2] we can assess the degree to which the component endpoints measure the same event(Table 7.2)

<table>
<thead>
<tr>
<th>Component Endpoint</th>
<th>Number of Events</th>
<th>Incidence Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatal CHD</td>
<td>215</td>
<td>0.052</td>
</tr>
<tr>
<td>NonFatal MI</td>
<td>308</td>
<td>0.074</td>
</tr>
<tr>
<td>Joint Occurrence</td>
<td>37</td>
<td>0.009</td>
</tr>
</tbody>
</table>

In CARE, there were 215 patients who experienced the fatal CHD component of the primary endpoint. The nonfatal MI component endpoint was

* This required tremendous effort in the trial, but insulated the trial from the criticism that it was unblinded de facto.
observed in 308 patients, and 37 patients had both a nonfatal MI and a CHD death. We can utilize formula (7.2) to compute the probability a patient has a CHD death given they experienced a nonfatal myocardial infarction as $0.009/0.074 = 0.12$. If these two component endpoints were coincident, then there would have been many more patients who experienced both a nonfatal myocardial infarction and a fatal coronary heart disease death. In CARE, even though the component endpoints measure the same underlying pathophysiology, these singleton endpoints are not coincident.

Perhaps a useful rule of thumb in determining whether there is sufficient difference in what the component endpoints are measuring is whether the component endpoints require different documentation. Within the clinical trial mechanism, the occurrence of multiple events in the same trial participant often translates into the requirement of different and distinct documentation to confirm the occurrence of each of the singleton endpoints. In the example of a clinical trial for which the composite endpoint is total mortality + hospitalization for CHF, the documentation for the occurrence of total mortality is insufficient to document the occurrence of a hospitalization for CHF. Each of these two component endpoints requires its own standard of documentation. The total mortality singleton endpoint will require information such as a death certificate and perhaps eye witness accounts that describe the circumstances surrounding the patient’s death. Documentation of the occurrence of a hospitalization for congestive heart failure will at least require a hospital discharge summary. The difference in the type of documentation required by these two endpoints is a reflection of the distinctions between these morbid and mortal events. Thus, although the component endpoints of a combined endpoint should be coherent, i.e. they should be reflections of the same disease process, require different documentation as a demonstration of their distinctive features.

7.11 Component endpoint equivalence

There is no doubt that component endpoints have been very useful in the design and analysis of clinical trials. However the techniques and tools of analysis of these endpoints can be complex. In some circumstances, the complications induced by the analyses of these complicated endpoints can undermine and even negate any advantage the combined endpoint itself offered.

The component endpoints that make up the composite endpoint are typically of two types. One type is the endpoint measurement that is itself a number e.g. change in blood pressure. Since a patient’s blood pressure change over time can assume any value (within a reasonable range), including fractions, this type of endpoints is defined as a continuous endpoint. Other examples of continuous endpoints are changes in left ventricular ejection fraction and reductions in glycosylated hemoglobin levels. Other endpoint measures are not continuous. They either occur or they do not occur. The simplest and best example is death—a patient either dies or survives. In this case there is no intermediate value. An endpoint such as death is described as a dichotomous(or 0-1) endpoint. While death is the clearest
example of a dichotomous endpoint, other examples are hospitalization (a patient is either hospitalized or not), or the requirement of chemotherapy*

Analysis tools for component endpoints which are either continuous or dichotomous are well described [3,4]. However analysis tools for combinations of these endpoints can be complex, and sometimes may not be generally accepted. Even in the simplest cases, the analysis of the composite endpoint may make some questionable assumptions. As an illustration, consider the circumstance of a clinical trial whose prospectively defined combined endpoint is assembled from two dichotomous component endpoints. The first component endpoint is death and the second component of this composite endpoint is hospitalization. The patient is considered to have met the criteria for the combined endpoint (said to have “reached” the combined endpoint) if they have either died during the course of the trial, or they survived the trial but were hospitalized during the study. In the case of a patient who is hospitalized and then dies during the clinical trial, only the first endpoint is counted. As described earlier, this analytic tool avoids the problem of counting a patient more than once if they have experienced multiple hospitalizations.

While this analysis is useful, it makes the explicit assumption that each of the two components of this combined endpoint is analytically equivalent to the other. Whether a patient meets the hospitalization part of the endpoint or the mortality part of the endpoint doesn’t matter as far as the analysis is concerned. But is a hospitalization the same as a death? Is this assumption of equivalence a true reflection of clinical reality? While one might possibly make the argument that a patient who is admitted to a hospital in stage four heart failure is close to death, an investigator would not need to look very far to find someone who disagrees with the assumption that this complicated hospitalization is equivalent to death. Obviously, less sick patients can be hospitalized but survive to lead productive lives, a circumstance that is clearly not the clinical equivalent of death.

A similar debate might be sparked in the consideration of the equivalence assumption for patients who reach the prospectively defined composite endpoint of fatal or nonfatal myocardial infarction. Since patients who suffer and survive heart attacks can live for years, be involved in gainful employment, participate in community activities, enjoy their families, and even be enrolled in subsequent clinical trials whose entry criteria require a prior myocardial infarction, is it reasonable to assume that MI and subsequent survival is equivalent to MI with immediate death?

This equivalence can be a troubling assumption and can complicate acceptability of the combined endpoint. Of course, there are alternative algorithms that are available that would provide different “weights” for the occurrence of the various component endpoints of a combined endpoint. For example, one might assume that for the combined endpoint of death + hospitalizations a death is three times as influential as a hospitalization. However, it is very difficult for

* Sometimes a continuous endpoint is converted to a dichotomous endpoint prospectively. An example of this type of conversion would be reduction in blood sugar, in which case the endpoint is not the magnitude of the reduction in blood sugar, but whether the blood sugar has been reduced by a prespecified amount, e.g. 25 mg/dl.
investigators to reach a consensus on the correct weighting scheme to use, and any selection of weights that the investigators choose that is different from equal weighting for each of the components of the combined endpoint is difficult to defend. Unfortunately, at this point there is no commonly accepted way out of this analytic enigma in clinical trials.

The situation only worsens when continuous and dichotomous component endpoints are combined in the same composite endpoint. How would one construct an analysis tool for the combined endpoint that has two component endpoints 1) death or 2) reduction by at least ten units in left ventricular ejection fraction? Not only is there the equivalence issue, but there is also the fact that, while the exact date of the patient’s death is known, the date when the patient first experienced a ten unit reduction in their ejection fraction after they were randomized is not known.* Complicated analysis procedures that address this issue have been developed [5]. However, as revealed at conversations held by the CardioRenal Advisory Panel of the F.D.A, these endpoints can be difficult to understand, and their acceptance by the medical community is guarded at best[6].

### 7.12 Therapy homogeneity

As pointed out in the previous section, an important trait of a combined endpoint is that each of its component endpoints should reflect an important clinical manifestation of the disease. However, the purpose of the selection of a combined endpoint is that it not only provide a persuasive depiction of disease morbidity and mortality, but that it also be a useful metric against which the effect of the clinical trial’s intervention will be tested. Therefore, from the investigators’ point of view, it would be most useful if the composite endpoint is sensitive to the therapy that will be assessed in the clinical study. This situation is most likely to occur if each of the component endpoints that make up the combined endpoint is itself responsive to the therapy to be tested in the clinical trial. The homogeneity of therapy effect for each of the singleton endpoints permits a fairly straightforward assessment of the prediction of the effect of therapy on the combined endpoint; this is a necessary feature in the traditional sample size computation of the trial†. In addition, therapy homogeneity helps to avoid interpretative difficulties when the medical community considers the intervention’s effect at the conclusion of the study. In addition, as recalled from chapters five and six, the presence of therapy homogeneity can be exploited in allocating type I error among dependent statistical hypothesis tests.

As an example of a combined endpoint whose component endpoints reflect consistent therapy homogeneity, we can return to the example of the CARE trial. Recall that CARE evaluated the effect of the HMG-CoA reductase inhibitor pravastatin on the reduction of morbidity and mortality in patients at risk of developing atherosclerotic disease. The primary analysis in CARE was the

* This could only be known if the patient had an ejection fraction measured each day of the trial.

†A brief primer on sample size computations is provided in Appendix 6, and an exploration of sample size computations in the presence of heterogeneity of singleton endpoints is provided in Appendix 8.
evaluation of the effect of pravastatin on the cumulative incidence rate of fatal coronary heart disease (CHD) + nonfatal myocardial infarction (nonfatal MI). As pointed out earlier in this chapter, each of the two component endpoints of this composite endpoint measures the same pathophysiology; however, the endpoints are distinct enough from each other to capture different manifestations of the same pathology.

For the prospectively defined analysis in CARE, patients were considered to have met the composite endpoint if they either 1) died from coronary artery disease or 2) survived the trial, but during the course of the trial, experienced a nonfatal myocardial infarction*. The effect of pravastatin was assessed at the conclusion of the trial (Table 7.3).

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>Placebo (n=2078)</th>
<th>Active (n=2081)</th>
<th>Relative Risk</th>
<th>95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined Endpoint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatal CHD* + Nonfatal MI**</td>
<td>274</td>
<td>212</td>
<td>0.76</td>
<td>[0.64 - 0.91]</td>
<td>0.003</td>
</tr>
<tr>
<td>Component Endpoints</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatal CHD</td>
<td>119</td>
<td>96</td>
<td>0.80</td>
<td>[0.61-1.05]</td>
<td>0.1</td>
</tr>
<tr>
<td>Nonfatal MI</td>
<td>173</td>
<td>135</td>
<td>0.77</td>
<td>[0.61-0.96]</td>
<td>0.02</td>
</tr>
</tbody>
</table>

* coronary heart disease
** myocardial infarction

Pravastatin therapy lowered the mean low density lipoprotein (LDL) cholesterol level by 32% and maintained average levels of LDL cholesterol of 97 to 98 mg/dl throughout the five years of follow-up. During follow-up, the LDL cholesterol level was 28% lower in the pravastatin group than in the placebo group, the total cholesterol level was 20% lower and the high density lipoprotein (HDL) cholesterol level was 5 percent higher.

CARE was a positive study. The frequency of the composite primary endpoint was 10.2% in the pravastatin group and 13.2 percent in the placebo group, reflecting a 24% reduction in risk. The use of pravastatin produced a relative risk of 0.76, representing a 24% reduction in the incidence of the combined endpoint of fatal CHD + Nonfatal MI. Note, however, the degree to which pravastatin affected each of the two components of the composite endpoint. Pravastatin produced a relative risk of 0.80 for the fatal CHD component of the primary endpoint, and a 0.77 relative risk for the nonfatal myocardial infarction component of the endpoint.

* A patient could of course suffer both a nonfatal myocardial infarction and subsequently die from coronary heart disease. In the analysis of a composite endpoint which is not continuous (e.g. blood pressure change, which can have a large number of values) but death, (which either occurs or does not, described as dichotomous), the time to the event is taken into account in the analysis. In this circumstance, when a patient experiences each of two dichotomous component endpoints, the patient is not counted twice. Instead, the earliest occurring of the two endpoints is the endpoint is counted.
Not only was the combined endpoint coherent, but it also demonstrated therapy homogeneity.

7.13 Composite endpoint measurement rules
The previous sections of this chapter discussed the considerations that the investigators must give to the details of the composite endpoint’s construction. However there are additional requirements that must be satisfied for the successful incorporation of a composite endpoint into a clinical trial. These additional requisites will now be reviewed.

7.14 Prospective identification
As pointed out in the earlier chapters of this book, the incorporation of an endpoint into the primary analysis of a clinical trial must follow certain principles. These principles require the prospective identification of the endpoint and the plan for its analysis. The motivations for this rule have been discussed in detail in chapters two through four of this book. Although that discussion focused on a single endpoint (e.g. total mortality), the guiding concept also applies to the evaluation of the effect of a randomly allocated intervention in a clinical trial on a composite endpoint.

As was the case for the single endpoint, the composite endpoint must be specified in great detail during the design phase of the trial. This description must include how each of the composite endpoint’s components will be ascertained. In addition, a committee of investigators is commonly chosen to determine whether a component endpoint has occurred. The procedures put in place to blind or mask these investigators from the identity of the randomly allocated therapy to which the patient was assigned should be elucidated. In addition, the analysis plan for the combined endpoint must also be detailed. Any weighting scheme that will be used in assessing the contribution each component endpoint makes to the combined endpoint must be determined \textit{a priori}, and should be acceptable to the medical and regulatory community. If there are plans to submit the results of the clinical trial to a regulatory agency, then that agency should be completely and fully informed about the details of both the construction of the combined endpoints and its analysis before the experiment begins.

The requirement of concordant trial execution is critical to the successful implementation of the composite endpoint in a study. Just as it is unacceptable to change the definition of the endpoints used in a study’s principle analyses, it is equally crucial to keep the constitution of the clinical trial’s combined endpoint fixed. Specifically, the component endpoints of a composite endpoint should be prospectively chosen and locked in. New component endpoints should not be added nor should established components be removed. The same chaotic effects* that can weaken and destroy the interpretation of a clinical trial whose principle analyses involve a single endpoint can also wreak havoc on the evaluation of a composite endpoint primary analysis.

* These effects are described in chapter two.
7.15 Combined endpoint ascertainment

The accurate assessment of the component endpoint’s interpretation in a clinical trial is both critical and complicated. To understand this new complexity introduced by the use of a composite endpoint, first consider a clinical trial that has the effect of the intervention on the cumulative incidence of coronary heart disease (CHD) death as its sole primary analysis. At the conclusion of the research, the study’s investigators must classify the experience of each of the randomized patient’s as one of 1) survival, 2) death due to a non-CHD cause, or 3) death due to CHD.

In well conducted clinical trials, specific documentation is collected to confirm that a patient reported by an investigator to have died is actually dead, and, if they are dead, the cause of that death. These confirmatory steps are taken in order to insure that living patients are not mistakenly assumed to have died. However, the investigators must also collect data confirming that a patient believed to be alive is in fact alive, a check that avoids the opposite mistake of assuming that a dead patient is actually living. While this last step is a straightforward matter for patients who have attended each visit, there is commonly a subset of patients who have missed several of the most recent visits and from whom no information has been collected. It is on these patients that intense activity is exerted to determine if they are either alive (as suspected) or have died.

The situation if much more complicated when a composite endpoint is to be part of the primary analysis of a clinical trial. If, in the above illustration, the investigators chose as a primary endpoint not just CHD death, but CHD death + nonfatal myocardial infarction (MI), the investigators have an additional inspection to complete. Not only must they assure themselves of the vital status of each patient; they must also determine whether a myocardial infarction has occurred in all patients. Of course, specific documentation will be collected from patients who volunteer the information that they have suffered an MI. However not every patient who experiences a heart attack reports the event to the investigators.* Occasionally, in some patients, the myocardial infarction may have produced no symptoms at all (silent MI’s). Mistakenly assuming that these patients who have experienced an MI were infarct free would lead to an inaccurate count of the number of patients who had this event.

The provision of assurance that patients who did not report a myocardial infarction in fact did not experience an MI can be an expensive task. Investigators, after determining that a patient has survived the trial, must also ask that surviving patient if they suffered a heart attack during the course of the study that the patient did not previously report. † The determination of the occurrence of silent myocardial infarctions can be especially problematic. Although many of these silent

---

* For example, the myocardial infarction and associated hospitalization might have occurred while the patient was on vacation and the patient was out of contact with the clinical trial’s investigator.
† As we stated earlier in this chapter, the time to first component event experienced is the critical measure used in the analysis of the effect of the intervention on the composite endpoint, thus infarct ascertainment is also important for patients who died during the course of the study.
events can be identified by requiring every patient to undergo annual electrocardiograms, obtaining and interpreting these evaluations is expensive. Also, if a silent MI is found to have occurred, its exact date can be impossible to determine.

As a final complication, consider the task awaiting investigators who have prospectively chosen the combined endpoint of CHD death/nonfatal MI/unstable angina pectoris. The evaluation of the unstable angina component, whose occurrence is commonly unrecognized and unreported, can add an overwhelming logistical burden onto the clinical trial apparatus. However, the study’s investigators must complete this onerous task. Recall that, in the analysis of the composite endpoint, the occurrence of unstable angina is just as critical as the occurrence of the other two components of the combined endpoint. If each of the component endpoints is important, then each must be measured with the same high standard of accuracy and precision. Clearly the greater the number of component endpoints in the study, the more work the investigators must complete in order to assure both themselves, the medical and the regulatory communities that they have an accurate count of the number of endpoints which have taken place during the course of the study. This is one additional problem of embedding a complicated composite endpoint into a clinical trial that has limited financial and logistical resources.

7.16 Conclusions

The implementation of composite endpoints in clinical trials holds both great promise and great danger. A carefully constructed combined endpoint can helpfully broaden the definition of a clinical endpoint when the disease being studied has different clinical consequences. This expansion commonly increases the incidence of the endpoint and this property can be used to either reduce the sample size of the trial, or if a larger sample size is maintained, increase the sensitivity of the experiment to detect moderate levels of therapy effectiveness. However, if the combined endpoint is too broad it can become uninterpretable and ultimately meaningless to the medical and regulatory communities. Thus the combined endpoint should be both broad and simultaneously retain its interpretability. We have termed this property coherence. Additionally, there should be some experiential evidence or at least theoretical motivation justifying the expectation that the therapy to be studied in the trial will have the same effect on each of the component endpoints of the combined endpoint. This we have termed homogeneity of therapy effect.

In the next chapter we will review some examples of the use of composite endpoints in clinical trials and describe how the number of confirmatory evaluations that derive both from an evaluation of the effect of therapy for the composite endpoint and its components may be expanded.
References


6. Transcript of the CardioRenal Advisory Committee to the F.D.A. Captopril. February 16, 1993