Contemporary Clinical Trials 86 (2019) 105846

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Interim monitoring in a treatment strategy trial with a composite primary endpoint

Minhee Kang[a,*], Birgit Grund[b], Sally Hunsberger[c], David Glidden[d], Paul Volberding[d]

[a] Harvard T.H. Chan School of Public Health, Boston, MA, USA
[b] School of Statistics, University of Minnesota, Minneapolis, MN, USA
[c] Biostatistics Research Branch, NIAID, NIH, Bethesda, MD, USA
[d] School of Medicine, University of California, San Francisco, CA, USA

ARTICLE INFO

Keywords:
Interim monitoring
Futility
Composite endpoint
Treatment strategy trial

ABSTRACT

When a clinical trial has a composite endpoint and a comparison of treatment strategies with multiple intervention components, interim data reviews by a data safety and monitoring board (DSMB) can be challenging as the data evolve on multiple fronts. We illustrate with a study in the treatment of Kaposi sarcoma (KS), an HIV-associated cancer with a multi-faceted disease presentation. The study, ACTG-A5264/AMC-067, was a 1:1 randomized trial to compare two strategies: immediate initiation of etoposide with antiretroviral therapy (ART), or ART with delayed etoposide upon disease progression. The outcome was a composite endpoint that included the following events, ordered from worst to best in the following three categories: (1) KS progression at 48 weeks, death, initiation of alternate KS treatment, loss to study follow-up; (2) stable KS; and (3) partial or complete KS response at 48 weeks. We present the interim results on the composite endpoint and the individual components, where components favored different study arms at an interim review. To facilitate interim data monitoring for complex trials, we recommend clear communications between the study team and the DSMB prior to the initiation of the trial on the need for a composite endpoint, the intentions behind the defined strategies, and relative importance of individual components of the composite endpoint. We also recommend flexibility in the timing of data reviews by the DSMB to interpret emerging data in multiple dimensions.
Clinicaltrials.gov NCT01352117

1. Introduction

Clinical trial designs can be complex when the disease under consideration is multi-faceted. There may be multiple outcomes and competing treatment options for the disease. These complexities may lead to a *composite endpoint* and comparisons of multi-stage or dynamic *treatment strategies*. Outcome assessments can be challenging with complex designs, including interim analyses that use incomplete and evolving data while the follow-up is ongoing. With a composite endpoint, the data are evolving on each component of the composite endpoint. In a treatment strategy trial, intervention components may change over time depending on the clinical events. Taken together, interim data reviews can be especially challenging as the data evolve in multiple dimensions.

Formally established independent committees, sometimes called Data and Safety Monitoring Board (DSMB) or Data Monitoring Committee (DMC), play a key role in ensuring safety and integrity of the clinical trial and ethical conduct of human research [1–4]. They are chartered to review in confidence the interim analysis results from accumulating data on safety and efficacy, and study conduct. They make recommendations for the trial to be modified or stopped, rather than waiting until the end of the clinical trial to examine the data. Monitoring committees can make various recommendations, from continuing the trial without changes to the existing protocol, to continuation with modifications, or stopping the trial early. There can be a number of reasons to recommend stopping the trial early, such as: the study question can be answered earlier than expected with the accumulated data (in support of or against the study hypothesis); there is unexpected harm; or the trial is not likely to be able to answer the study question (futility).

In this paper, we describe the challenges of interim data reviews in a treatment strategy trial with a composite endpoint. We begin by describing composite endpoints and treatment strategy trials, and complexities they add in interim reviews. We then illustrate these points

* Corresponding author at: Harvard T.H. Chan School of Public Health, 651 Huntington Ave, Boston, MA 02115, USA.
E-mail address: mkang@hsph.harvard.edu (M. Kang).

https://doi.org/10.1016/j.cct.2019.105846
Received 3 June 2019; Received in revised form 24 August 2019; Accepted 4 September 2019
Available online 11 September 2019
1551-7144/

with a completed study on the treatment of Kaposi sarcoma (KS), conducted together by the AIDS Clinical Trials Group (ACTG) and the AIDS Malignancy Consortium (AMC). As a trial to investigate when to initiate chemotherapy for an HIV-associated cancer, the study embodied complexities of both HIV and cancer, and challenges of both the composite endpoint and the treatment strategy design. We present the interim results on the composite endpoint, its individual components where components of the composite endpoint favored different study arms at an interim review, and conditional power. We conclude with recommendations to facilitate interim data monitoring for complex trials.

2. Composite endpoints

Composite primary endpoints are increasingly used in clinical trials [5,6]. A composite endpoint is comprised of multiple single endpoints – or components. In some settings, it may be appropriate to combine key clinical outcomes that are expected to be affected by the treatment into a single variable for a number of reasons, such as: multidimensional presentation of the disease under study, rare clinically important events, no consensus for a single endpoint, and study feasibility. For example, a composite outcome of clinical events such as death, non-fatal myocardial infarction and stroke might be used in cardiovascular trials. In oncology trials, clinical disease progression and all-cause death are often combined to assess progression-free survival. Composite outcomes may also combine clinical events with laboratory markers; in HIV clinical trials, time to loss of virological response (TLOVR) is a composite outcome that incorporates virological failure that is confirmed with two consecutive viral load measurements, loss to follow-up, initiation of a new treatment and all-cause death [5]. In the last example, components have vastly different consequences for the study participant. Several papers have discussed analysis considerations for clinical trials with composite endpoints, and the FDA draft guidance to industry on multiple endpoints in clinical trials includes a discussion on composite endpoints [6–9]. They recommend analyzing individual components of the composite endpoint for a deeper understanding of the treatment effects and interpretation of the overall study results, to assess which components are driving the overall significance in the composite endpoint. The study intervention effects on some components may be of very different magnitudes or even point to different directions. It is recommended that the interim analysis reports also contain individual examination of each component so that DSMBs can examine all relevant data to obtain a clearer picture of the overall balance between benefit and harm [1].

Composite endpoints can present challenges to interim data reviews due to the multiplicity of both outcome components and interim analyses. Observed data on each component of the composite endpoint may evolve over the study period differently, as component events may occur at different rates during the follow-up. The observed treatment effect on the primary composite endpoint may be driven by one component at one interim data review, and by another component at a subsequent review. And components may not have the same impact for the participant: for instance, disease progression is less concerning than death. Decision-making process is multifactorial even for a single outcome, but there are even more dimensions to consider with composite endpoints.

3. Treatment strategy trials

In treatment strategy trials, comparisons are not simply between a drug and a placebo (or among multiple drugs), but between strategies consisting of sequences of treatments as would be applied in clinical practice. The switch or addition of drug(s) may be driven by an event, such as failure of an initial regimen. Hence, one or more components of the intervention may change over time as part of the planned strategy. In the context of an HIV trial aimed to answer which sequence of drugs was best in a regimen, treatment strategy was defined as “an approach that includes one or more combination regimens in succession”, that included selection of follow-up regimens for participants who fail the initial regimen [10].

One class of treatment strategy trials is “when to start” trials, where the study aims to determine the optimal timing of treatment initiation or addition of drug(s) to the treatment in the course of the disease. Examples include treatment guided by interim diagnostic imaging in advanced Hodgkin's lymphoma [11], earlier versus later antiretroviral treatment in patients with both tuberculosis and HIV infections [21], and more generally, trials to study early versus delayed (or deferred) treatment in various diseases areas: acute hepatitis C [12], HIV [13], prostate cancer [14], ovarian cancer [15]. While not specific to treatment strategy trials, the study endpoint is often chosen to assess the status of study participants at a fixed time after the trial initiation. In the tuberculosis example above, the primary endpoint was survival to week 48.

Interim monitoring for treatment strategy trials includes review of evolving data on intervention changes in accordance with the strategy. It is important for the DSMB to know when the intervention changes occur at a participant-level, especially relative to the study events, as well as the evolution of the treatment stages over time for the study population. This is especially important in consideration of adverse events. If the intervention change is dependent on a participant outcome – for instance, switching or adding a new drug upon a clinical event – then the treatment data are particularly informative. Treatment strategy trials are typically open-label since the strategy changes are driven by specific events, and the DSMB must be mindful of the trial conduct in an open-label setting.

4. Methods: a Kaposi sarcoma study example

4.1. ACTG-A5264/AMC-067

We describe a recently completed trial in Kaposi sarcoma (KS) to illustrate some of the complexities in interim monitoring of a treatment strategy trial with a composite endpoint. KS is an HIV-associated cancer, a major complication associated with AIDS patients upon immunosuppression, with disease presentation in a continuum from mild to advanced stages. Presentation in sites such as the lung is potentially associated with mortality, whereas mortality is not commonly associated when limited to sites like the skin. As an HIV-associated cancer, KS embodies complexities of both the HIV disease and cancer [16]. In the era of potent antiretroviral therapy (ART), KS incidence has decreased in developed nations where ART is administered, but still poses a major problem in developing countries where HIV-1 incidence is high and ART is still not yet widely available. ART for the treatment of HIV is considered an essential part of HIV-associated KS treatment, but the role for chemotherapy in mild to moderate KS has not been well-defined; in some cases, limited KS may resolve with ART alone without aggressive chemotherapy. ACTG-A5264/AMC-067 was a Phase III, randomized, open-label study to compare two treatment strategies for initial treatment of mild to moderate, HIV-associated KS. The strategies included ART (combination of HIV drugs) and a low dose oral chemotherapy, etoposide (ET). Participants were randomized 1:1 to the strategy of ART with ET initiated only upon KS progression (“As-Needed”) and the strategy of immediate initiation of ET with ART (“Immediate”). The study design and the primary results of the trial have been described elsewhere [17]. Participants randomized to “As-Needed” who responded to ART remained on ART alone and did not initiate ET. Participants in “Immediate” initiated both ART and ET together. While all participants initiated ART upon randomization, the study can be thought of as a “when to start” trial with respect to ET chemotherapy.

4.2. Primary endpoint

The primary objective of the study was to compare Week 48 outcomes between As-Needed and Immediate arms. The primary endpoint was composed of the following three ordered categories from worst to best as follows.

(E1) Failure: KS progression at Week 48 compared to study entry, initiation of alternate KS treatment (treatment other than ET, triggered by worsening KS) by Week 48, all-cause death or loss to follow-up by Week 48.

(E2) Stable: Neither E1 nor E3.

(E3) Response: KS response (partial or complete) at Week 48 compared to study entry without initiation of alternate KS treatment by Week 48.

The team was interested in the participant outcome at 48 weeks after the initiation of ART, either with immediate etoposide or with etoposide added later as needed. As a "when to start" strategy trial, outcomes at a fixed time (Week 48) were of interest, since it is possible that participants in As-Needed would respond to ART alone, sparing chemotherapy and potential toxicity, or with etoposide added later as needed upon KS progression. Evaluation criteria for KS progression, partial response and complete response have been described elsewhere, which assess cancer outcomes [17]. Initiation of alternate KS treatment not provided by the study upon site clinician judgment was categorized as failure, as it was deemed to reflect failure of the study treatment. Losses to follow-up were categorized as failures, assuming the worst case scenario, similar to the TLOVR endpoint in HIV clinical trials. While KS-related deaths would certainly be considered the worst outcome from the study perspective, all deaths were considered failures due to challenges in death attributions, also similar to TLOVR. Sensitivity analyses with alternate categorizations of losses to follow-up and all-cause deaths were planned and specified in the protocol, since assuming the worst may not be correct.

4.3. Sample size and planned analyses

Unlike a single endpoint, multiple assumptions are needed for the components in a composite endpoint. Table 1 gives the rates assumed in each category for the sample size calculations. The team considered KS progression and potential initiation of non-study KS treatment together as worsening of disease. The team projected 10% reduction in progression and 15% improvement in response with Immediate.

Assuming a uniform rate of loss over time due to early study discontinuation and all-cause deaths of at most 10%, Failure (E1), Stable (E2) and Response (E3) proportions in As-Needed led to 37%, 36% and 27%, respectively, and 27%, 32% and 41% in Immediate. The sample size of 234 per arm was determined to achieve 90% power to test superiority of Immediate, in a Wilcoxon rank-based test for ordered categorical outcomes with a two-sided type I error of 5%.

Sensitivity analyses on some alternate categorizations of components were planned. Analysis on each component of the composite endpoint in the interim review reports was also planned. Haybittle-Peto stopping boundaries were specified for interim analyses on the primary endpoint to address type I error rate control for repeated analyses [18]. Conditional power analyses [19,20] were planned as tools to assess futility at interim monitoring reviews, but binding thresholds were not specified to allow flexibility in interim monitoring. As such, the sample size calculations did not take into account the possible loss of power due to the repeated interim monitoring of futility boundaries. The monitoring plan specified that the first interim efficacy review would occur when approximately 25–35% of the participants have Week 48 data available and annually thereafter.

SAS software, version 9.4 (SAS Institute Inc., Cary, North Carolina) was used for the interim data analyses, and R software, version 3.1.1 (R Foundation for Statistical Computing, Vienna, Austria) was used for conditional power calculations.

5. Results: interim data reviews of the KS study

5.1. Evolving interim data on the composite endpoint

Table 2 shows the evolution of the safety data as reviewed at DSMB meetings, including the early reviews where primary endpoint data were not yet included as specified by the protocol's interim efficacy analysis plans. Since death was an important part of safety monitoring, it was presented in every report to the DSMB. Table 2 shows that there was a suggestion of an imbalance in deaths early in the study, especially among the deaths attributed to KS with more events in the As-Needed arm. Death was also a component of the primary endpoint. While the original monitoring plan did not call for primary endpoint evaluation this early in the study, the DSMB was presented with partial information on the primary endpoint that could be concerning. Therefore, a review of the primary endpoint was warranted from the safety perspective regardless of the planned schedule on the interim efficacy analyses.

We first note that when a component of the composite endpoint is related to safety, it may not be possible to separate the interim analysis timeline for the primary endpoint from the safety review that occurs routinely. Therefore, the overlap between safety and efficacy should be considered in interim monitoring plans. We make a second note that by the end of the study, the imbalance in deaths between arms resolved. This is an example that caution should be taken in interpreting early interim data. Due to small numbers, estimates based on early data are highly variable with wide confidence intervals. A third note is that primary attribution of death to KS remained higher in the As-Needed arm. Attributing the primary cause of death, however, is a challenging task and is highly subjective by nature. One might postulate that site knowledge of the treatment assignment may have affected the attribution, and if true, this exemplifies data interpretation challenges when a trial is forced to be conducted as open-label.

Table 3 shows the interim analysis results on the primary endpoint. The data were limited to the participants with Week 48 data potential at the time of the interim analysis; that is, they enrolled early enough so that they could have been followed to 48 weeks. The first efficacy review occurred earlier than initially planned, at 18% information instead of at $\geq$25%. Conditional power was assessed in two ways as planned:

Table 1
Projected event rates at study design.

	Category	As-Needed: ART alone or with delayed ET	Immediate: immediate ET with ART	Improvement in immediate
Projected disease status (excludes loss to follow-up and death)	Worsening of disease	30%	20%	10% decrease
	Stable disease	40%	35%	5% decrease
	Improvement of disease	30%	45%	15% increase
Projected composite endpoint (incorporates loss to follow-up and all-cause death)	E1 = Failure	37%	27%	10% decrease
	E2 = Stable	36%	32%	6% decrease
	E3 = Response	27%	41%	14% increase

Table 2
Interim safety data for As-Needed (ART alone or with delayed ET) and Immediate (Immediate ET with ART).

	September 2012 (4% accrual)		October 2013 (20% accrual)		September 2014[a] (28% accrual)		September 2015[a] (36% accrual)		March 2016[a] (40% accrual)	
	As-Needed (N = 10)	Immediate (N = 8)	As-Needed (N = 45)	Immediate (N = 48)	As-Needed (N = 64)	Immediate (N = 67)	As-Needed (N = 82)	Immediate (N = 87)	As-Needed (N = 93)	Immediate (N = 95)
Grade 3 Signs/ Symptoms	1 (10%)	1 (12%)	8 (18%)	5 (10%)	14 (22%)	10 (15%)	19 (23%)	14 (16%)	24 (26%)	15 (16%)
Grade 4 Signs/ Symptoms	0 (0%)	0 (0%)	1 (2%)	0 (0%)	2 (3%)	1 (2%)	3 (4%)	2 (2%)	5 (5%)	2 (2%)
Grade 3 Lab	0 (0%)	1 (12%)	11 (24%)	9 (19%)	14 (22%)	16 (24%)	17 (21%)	18 (21%)	18 (19%)	20 (21%)
Grade 4 Lab	2 (20%)	1 (12%)	5 (11%)	7 (15%)	5 (8%)	11 (16%)	7 (9%)	13 (15%)	9 (10%)	13 (14%)
Deaths	2 (20%)	1 (12%)	7 (16%)	3 (6%)	10 (16%)	8 (12%)	12 (15%)	11 (13%)	12 (13%)	12 (13%)
KS Primary	0 (0%)	0 (0%)	4 (9%)	0 (0%)	5 (8%)	0 (0%)	6 (7%)	1 (1%)	6 (6%)	1 (1%)
Other	2 (20%)	1 (12%)	3 (7%)	3 (6%)	5 (8%)	8 (12%)	6 (7%)	10 (12%)	6 (6%)	11 (12%)

[a] Interim reviews that also included efficacy analysis.

(1) assuming that the future data would follow the observed trend, and (2) assuming that the future data would follow the design assumptions. While the observed estimates were not as assumed in the study design, a decrease of nearly 16% in Failure (E1) and nearly 10% increase in Response (E3) in the Immediate arm led to a high conditional power. The second efficacy review occurred when there was 28% information one year later. The conditional power decreased sharply from the first interim efficacy review due to the emerging data on the initiation of alternate KS treatments, since such initiations counted towards Failure in the composite endpoint. Unexpectedly, there were more alternate KS treatment initiations in the Immediate arm, which led to more similar Failure proportions in the two study arms. It serves as another reminder that interim data are evolving data. And with a composite endpoint, the data are evolving on multiple fronts.

5.2. Intervention effects on the components of the composite endpoint

Fig. 1 shows components of the composite endpoint in the interim efficacy analysis that occurred in March 2016. At this time, 188 participants were enrolled in the study (40% accrual), and 152 participants had sufficient follow-up to be included in the analysis of the primary endpoint (32% information). The study had very few losses to follow-up and the proportions across arms were similar, so there were no concerns regarding these components. KS outcomes, deaths, and initiation of alternate KS treatment were examined closely. The component analyses of the primary endpoint suggested decreased KS response and increased KS progression in the As-Needed arm among the participants with Week 48 KS evaluations and without initiation of alternate KS treatment. These components suggested that the Immediate strategy might be favorable. However, there was more initiation of alternate KS treatment by Week 48 in Immediate compared to As-Needed, which was categorized as Failure (E1). The latter component dampened the E1 difference between the two study arms to the point of reversing the direction. This is an example where the treatment effects on the components of the composite endpoint are not all aligned in the same direction. Next section illustrates how sensitive the primary endpoint analysis was to the component on alternate KS treatment initiation.

5.3. Sensitivity analysis

We present a sensitivity analysis from the March 2016 interim report where alternate KS treatment is not counted as a failure. This analysis was meant to give insight into the role of the alternate KS treatment initiation in the composite endpoint. The following categories defined a new, ad-hoc composite endpoint.

Failure: KS progression at Week 48 compared to study entry, all-cause death or loss to follow-up by Week 48.

Stable: Neither Failure nor Response.

Response: KS partial or complete response at Week 48 compared to study entry.

Table 4 shows the observed estimates for the protocol-defined and the ad-hoc endpoints. Leaving out the initiation of alternate KS treatment in the composite endpoint changed the endpoint category estimates substantially in Immediate arm, where the Failure estimate changed from 53.9% in the protocol-defined endpoint to 35.5% in the ad-hoc endpoint. The Response category estimate changed from 34.2% to 47.4% in the same arm. These led to a substantial difference between the two arms, and the p-value in the comparison test changed from 0.965 to 0.085 in favor of Immediate in the ad-hoc analysis.

This sensitivity analysis highlighted the effect of the initiation of alternate KS treatment component in the composite endpoint. It should be noted that the analysis using the ad-hoc composite endpoint does not answer the original study question on the benefit of the immediate versus delayed administration of etoposide, but perhaps a different

Table 3
Observed estimates for As-Needed (ART alone or with delayed ET) and Immediate (Immediate ET with ART) arms and conditional power for the primary composite endpoint (N = 468) at each interim efficacy look. E1 = Failure (includes all-cause death, alternate KS treatment, loss to follow-up), E2 = Stable and E3 = Response.

Interim efficacy look	Information[a]	Observed estimates in as-needed	Observed estimates in immediate	Conditional power (Observed Effect)	Conditional power (Design Assumption)
1 September 2014	18% (N1 = 85)	E1 = 65.9% (N = 27) E2 = 7.3% (N = 3) E3 = 26.8% (N = 11)	E1 = 50.0% (N = 22) E2 = 13.6% (N = 6) E3 = 36.4% (N = 16)	91%	90%
2 September 2015	28% (N2 = 130)	E1 = 54.0% (N = 34) E2 = 19.0% (N = 12) E3 = 27.0% (N = 17)	E1 = 53.7% (N = 36) E2 = 9.0% (N = 6) E3 = 37.3% (N = 25)	14%	75%
3 March 2016	32% (N3 = 152)	E1 = 52.6% (N = 40) E2 = 15.8% (N = 12) E3 = 31.6% (N = 24)	E1 = 53.9% (N = 41) E2 = 11.8% (N = 9) E3 = 34.2% (N = 26)	2%	57%

[a] Limited to participants with potential Week 48 data. Maximum information is 100% at N = 468.

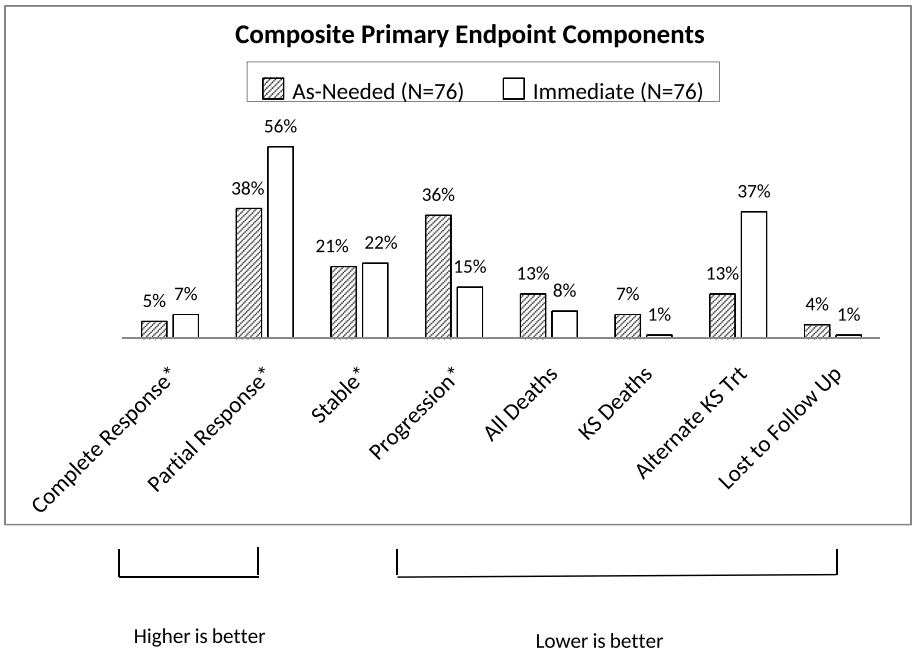

Fig. 1. Individual Components of the Primary Endpoint at the Final Interim Look.

question on the benefit of initiating chemotherapy early (which can allow earlier introduction of other, more potent chemotherapy drugs). Importantly, if a new endpoint, such as the ad-hoc composite endpoint above, is defined after the study is initiated upon review of the emerging data and presented to the DSMB, the study team should not be informed of such analysis while the study is ongoing. Since the study team is blinded to all interim data, the team would not understand why such analysis is needed and could inappropriately draw conclusions that could affect the conduct of the study. This emphasizes the need to keep the interim data reviews confidential.

Other sensitivity analyses specified in the protocol using alternative categorizations for losses to follow-up and deaths were also conducted. Because of few losses to follow-up and similar rates between the two arms, those results were not substantially different from the primary endpoint analysis results.

5.4. Conditional power

Table 5 shows the conditional power estimates for the composite endpoint analyses presented at the March 2016 DSMB meeting, using the protocol-defined primary endpoint and the ad-hoc sensitivity analysis endpoint. In addition, two key clinical endpoints that were part of the composite endpoint were included: KS progression and KS response (partial and complete combined). The conditional power for the protocol-defined composite endpoint under the assumption that the future data would follow the observed trends was low at 2%. In contrast, the conditional power for the ad-hoc endpoint was high at 89%, further highlighting the treatment strategy effect on the initiation of alternate KS treatment component. Considering KS progression as an endpoint on its own, the conditional power was also high at 85%. Similarly, considering KS response as a study endpoint also gave a high conditional power.

Based on the interim data and power discussions, the DSMB recommended closing the study, making it the final interim data review. The conditional power for the primary endpoint was so low at only 2% that the study was deemed futile to continue. At that time, the study

Table 5
Conditional power for various endpoints at the final interim review (N = 152, 32% information).

Endpoint	Observed estimates in As-needed	Observed estimates in Immediate	Conditional power (Observed Effect)
Primary endpoint	E1 = 52.6% E2 = 15.8% E3 = 31.6%	E1 = 53.9% E2 = 11.8% E3 = 34.2%	2%[a]
Ad-Hoc endpoint for additional sensitivity analysis	E1 = 48.7% E2 = 15.8% E3 = 35.5%	E1 = 35.5% E2 = 17.1% E3 = 47.4%	89%
KS progression[b]	48.7%	35.5%	85%
KS response[b]	35.5%	47.4%	75%

[a] Conditional power under the design assumptions was 57%.
[b] Using a key component of the composite endpoint as the endpoint, rather than a composite endpoint.

Table 4
Primary endpoint and ad-hoc sensitivity analysis endpoint at the final interim review. For the primary endpoint, Failure includes all-cause death, alternate KS treatment, loss to follow-up. For the additional ad-hoc sensitivity analysis endpoint, Failure does not include alternate KS treatment.

	Primary endpoint			Ad-Hoc sensitivity analysis endpoint		
	As-Needed (N = 76)	Immediate (N = 76)	p-Value	As-Needed (N = 76)	Immediate (N = 76)	p-Value
Failure	40 (52.6%)	41 (53.9%)	0.965	37 (48.7%)	27 (35.5%)	0.085
Stable	12 (15.8%)	9 (11.8%)		12 (15.8%)	13 (17.1%)	
Response	24 (31.6%)	26 (34.2%)		27 (35.5%)	36 (47.4%)	

had accrued 40% of the target sample size. The board concluded that the study was unlikely to answer the research question as hypothesized in the protocol, even if the study were to continue to complete the planned enrollment and follow-up. That is, there was sufficient evidence from the interim data that the study, if continued, would be unlikely to conclude that immediate etoposide strategy is superior to As-Needed in the primary endpoint analysis. An additional, important consideration for the DSMB was the proportion of participants who initiated non-study chemotherapy, and the imbalance between the two study arms.

6. Discussion

Clinical trial monitoring is a complex process which aims to address multiple concepts on ethics, participant safety, intervention efficacy and study integrity. There is no one simple algorithm for the review process that incorporates all the concepts, and it is not possible to foresee all potential scenarios that may arise during monitoring. Complex diseases can lead to complex interventions and endpoints – and monitoring. When the trial design is complicated, the emerging data are also likely to be complicated, which can make it more difficult for the DSMB to interpret the data to make decisions. We have described how composite endpoints and treatment strategy trials add complexity, in particular, at the intersection of HIV and cancer. We have illustrated some of the complexities by sharing the interim looks on the evolving data from a recently completed trial to compare treatment strategies for HIV-associated Kaposi sarcoma.

In the KS study example, the primary endpoint embodied aspects of both HIV and cancer. While the main component was KS evaluation, the strategy nature of the treatment led to a composite endpoint that incorporated participant statuses at a fixed time, and the failure definition had similarities to TLOVR in HIV studies. As a treatment strategy trial, the trial was conducted as open-label and was subject to provider's discretion on when to initiate alternate KS treatment outside of the study. When the study was designed, providing KS treatment outside of the study was anticipated to be rare in the setting where the trial was conducted and in the limited-disease study population.

We have a few recommendations to facilitate the monitoring process when composite endpoints are necessary, especially in the context of treatment strategy trials. Most importantly, the process needs to start with clear communication by the study team to the DSMB prior to the initiation of the study about the following.

(1) The need and intentions behind the composite endpoint and the strategy trial design.
(2) The relative importance of individual components of the composite endpoint and how each component of the composite endpoint and the treatment strategy should be assessed at interim reviews.

Ideally, the DSMB should be given guidance on the relative importance of individual components prior to the initiation of study monitoring. Then, if the observed treatment effect is unexpectedly driven by a component perceived to be of lesser clinical significance, the DSMB has sufficient information to exercise judgment on how to weigh that in making decisions for the study. Close communication between the study team and the DSMB must occur prior to the study. Once the study is open, the DSMB must keep the discussions arising from the emerging data confidential. It may be useful to examine various potential scenarios with different assumptions on the components of the composite endpoint; it would be a useful exercise not just with the DSMB, but also within the study team prior to finalizing the protocol. The number of potential scenarios is multiplied when the disease is complex and a composite endpoint is needed. When the study is introduced to the DSMB, the study investigators should be ready to communicate clear guidelines for modifying or stopping the trial under various scenarios, with the understanding that not all scenarios can be anticipated. In the KS study example, the frequency in the initiation of alternate KS treatment was unexpected, but it was made clear at the DSMB meeting when the team introduced the study that the team intended this to be considered failure of the study drug etoposide.

Since flexibility is especially important in the monitoring of complex studies, we also recommend the following.

(3) Planning the interim data monitoring using methods that allow for increasing the number of endpoint reviews as needed.

In our example of the KS study, deaths were a component of the primary endpoint, and it was difficult to fully monitor the study without interim looks on both deaths and the composite endpoint when deaths were presented as part of routine safety monitoring. This resulted in primary endpoint analysis at every subsequent review. A flexible timeline for data reviews is likely to be necessary for the DSMB to interpret emerging data from complex trials.

Acknowledgements

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Numbers UM1 AI068634, UM1 AI068636, UM1 AI106701, and P30 AI027763. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We appreciate the support of ACTG-A5264/AMC-067 study chairs and co-chairs in presenting the study as an example, and we are grateful to the study team, site investigators and the trial participants.

References

[1] D.L. DeMets, C. Furberg, L.M. Friedman, Data Monitoring in Clinical Trials: A Case Studies Approach, Springer. xxvi, New York, NY, 2006, p. 374.
[2] S.S. Ellenberg, T.R. Fleming, D.L. DeMets, Data Monitoring Committees in Clinical Trials: A Practical Perspective, J. Wiley & Sons, Chichester, West Sussex, England; Hoboken, NJ, 2002 (1 online resource (xiii, 191 pages).
[3] R. Holubkov, et al., The role of the data and safety monitoring board in a clinical trial: the CRISIS study, Pediatr. Crit. Care Med. 14 (4) (2013) 374–383.
[4] J. Zuckerman, B. van der Schalie, K. Cahill, Developing training for data safety monitoring board members: a National Institute of Allergy and Infectious Diseases case study, Clin. Trials 12 (6) (2015) 688–691.
[5] L. Wittkop, et al., Methodological issues in the use of composite endpoints in clinical trials: examples from the HIV field, Clin. Trials 7 (1) (2010) 19–35.
[6] A.J. Sankoh, H. Li, R.B. D'Agostino Sr., Use of composite endpoints in clinical trials, Stat. Med. 33 (27) (2014) 4709–4714.
[7] FDA, Multiple Endpoints in Clinical Trials: Guidance for Industry, https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegrulatoryInformation/Guidances/UCM536750.pdf, (2017).
[8] G. Gomez, S.W. Lagakos, Statistical considerations when using a composite endpoint for comparing treatment groups, Stat. Med. 32 (5) (2013) 719–738.
[9] A.J. Sankoh, H. Li, R.B. D'Agostino Sr., Composite and multicomponent end points in clinical trials, Stat. Med. 36 (28) (2017) 4437–4440.
[10] L.M. Smeaton, et al., ACTG (AIDS Clinical Trials Group) 384: a strategy trial comparing consecutive treatments for HIV-1, Control. Clin. Trials 22 (2) (2001) 142–159.
[11] P. Johnson, et al., Adapted treatment guided by interim PET-CT scan in advanced Hodgkin's lymphoma, N. Engl. J. Med. 374 (25) (2016) 2419–2429.
[12] K. Deterding, et al., Delayed versus immediate treatment for patients with acute hepatitis C: a randomised controlled non-inferiority trial, Lancet Infect. Dis. 13 (6) (2013) 497–506.
[13] J.D. Lundgren, et al., Initiation of antiretroviral therapy in early asymptomatic HIV infection, N. Engl. J. Med. 373 (9) (2015) 795–807.
[14] G.M. Duchesne, et al., Timing of androgen-deprivation therapy in patients with prostate cancer with a rising PSA (TROG 03.06 and VCOG PR 01-03 [TOAD]): a randomised, multicentre, non-blinded, phase 3 trial, Lancet Oncol. 17 (6) (2016) 727–737.
[15] G.J. Rustin, et al., Early versus delayed treatment of relapsed ovarian cancer (MRC OV05/EORTC 55955): a randomised trial, Lancet 376 (9747) (2010) 1155–1163.
[16] R. Yarchoan, T.S. Uldrick, HIV-associated cancers and related diseases, N. Engl. J. Med. 378 (11) (2018) 1029–1041.

[17] M.C. Hosseinipour, et al., As-needed Vs immediate etoposide chemotherapy in combination with antiretroviral therapy for mild-to-moderate AIDS-associated Kaposi sarcoma in resource-limited settings: A5264/AMC-067 randomized clinical trial, Clin. Infect. Dis. 67 (2) (2018) 251–260.
[18] J.L. Haybittle, Repeated assessment of results in clinical trials of cancer treatment, Br. J. Radiol. 44 (526) (1971) 793–797.
[19] C. Jennison, B.W. Turnbull, Group Sequential Methods with Applications to Clinical Trials, Chapman & Hall/CRC, Boca Raton, 2000 (xviii, 390 p).
[20] J.M. Lachin, A review of methods for futility stopping based on conditional power, Stat. Med. 24 (18) (2005) 2747–2764.
[21] D.V. Havlir, M.A. Kendall, P. Ive, et al., Timing of antiretroviral therapy for HIV-1 infection and tuberculosis, N Engl J Med 365 (16) (2011) 1482–1491.