

What Types of Efficacy Outputs Does the DMC Receive?



David Kerr and Nand Kishore Rawat

Abstract This chapter goes into detail about the outputs that are both comprehensive and comprehensible for the DMC – this chapter focuses on efficacy. We advocate for efficacy data to be provided – even if non-inferential and just a proxy for the primary endpoint. This aids the DMC in a more comprehensive assessment of risk-benefit to provide more suitable DMC recommendations. The hazards of repeated assessment of efficacy data are noted, as are the hazards of overinterpreting early trends or lack of early trends. The principles of interpreting p-values are provided.

Keywords Inferential efficacy · Non-inferential efficacy · Risk-benefit · Endpoints · p-values · Kaplan-Meier figure · Repeated assessments · Alpha hit · Alpha spending

Non-inferential efficacy is a critical component to many Data Monitoring Committee (DMC) outputs even for meetings that some might ascribe as ‘safety reviews’. Without access to some form of efficacy, the DMC cannot fully assess risk-benefit. The result would be the DMC recommending many studies stop for safety concerns when there is actually a good chance of efficacy eventually being demonstrated that more than offsets the safety concern. For example, many chemotherapy treatments are ‘unsafe’ given the severe side effects that are seen but are valued by the clinical community because of the efficacy provided.

This philosophy regarding the inclusion of efficacy data (even if non-inferential) is most compelling if the endpoint itself is clinically compelling – one that measures

Figure by Bill Coar.

D. Kerr
Seattle, WA, USA
e-mail: david.kerr@cytel.com

N. K. Rawat (✉)
Lantheus Medical Imaging (United States), King of Prussia, PA, USA

© The Author(s), under exclusive license to Springer Nature
Switzerland AG 2023

N. K. Rawat, D. Kerr (eds.), *Data Monitoring Committees (DMCs)*,
https://doi.org/10.1007/978-3-031-28760-2_12

how the subject 'feels, functions, and survives'. An endpoint that is a biomarker that has not yet been confirmed as a correlate for a clinically compelling endpoint is not as necessary to include. In these cases, the DMC might exclusively use the safety data to guide their recommendations.

The non-inferential efficacy can be a proxy to the primary endpoint. It could use data that is not fully cleaned or complete. There could be many approximations or other issues. But the purpose is simply to assuage DMC concerns that, in the presence of safety concerns, there is still some offsetting benefit or hope for benefit. As an example, in a study where the primary endpoint is disease progression confirmed by adjudication committee (e.g., BICR), it might be sufficient to have the DMC outputs include the site-proposed disease progression, given the expected high correlation between site-proposed disease progression and the disease progression confirmed by the adjudication committee. Or the primary endpoint might be a composite, but the DMC is presented with just the individual components of the composite. Or the endpoint is a time-to-event analysis, but the DMC is provided with the simple counts of the events, rather than being provided the Kaplan–Meier curve.

Including information on death can be viewed as both a safety output, as well as an efficacy output. It is typical to include an ad hoc Kaplan–Meier figure of OS. Note that the safety output will include a table using treated population, whereas the efficacy section may also include table of deaths using randomized population which could additionally include deaths from subjects that were never treated. It is relatively unlikely, but there could be concerns of 'reverse efficacy' where the trends in the endpoint variables are in the unexpectedly opposite direction and be indicative of safety concerns. That is another reason to include some measure of efficacy to the DMC.

There have been statistical concerns stated about including efficacy data as a routine matter. That is true for both non-inferential efficacy data but especially if inferential statistics are included (hazard ratio and confidence interval and p-value, for example) due to a perceived 'alpha hit'. There is a valid statistical concern that repeated assessments of data with the potential of early stopping do need to be accounted for in assessing p-values at interim and final analysis. But these efficacy outputs are not for potential stopping and therefore do not impact the interpretation of the final p-value at the final analysis. However, if objections persist, a simple approach is to build in that the alpha-spent at each review is some extremely small value, say, 0.0001. The end result on the final p-value is negligible. This does account for the unlikely but theoretical possibility that amazingly good results are seen during the study and the DMC may feel ethically compelled to recommend informing the sponsor and allowing subjects in the study and beyond to get access to this efficacious treatment earlier. Of course, as with all DMC recommendations, there can and should be discussion with the Sponsor Liaison and others (such as regulatory agencies) before any final decisions are made.

Efficacy data will, of course, be presented if the purpose of the DMC meeting is a formal assessment of specified criteria. That criteria could be for recommendation of stopping early for benefit, recommendation for stopping early for futility, or for other reasons. In many cases, just one or two outputs are included (one table, and

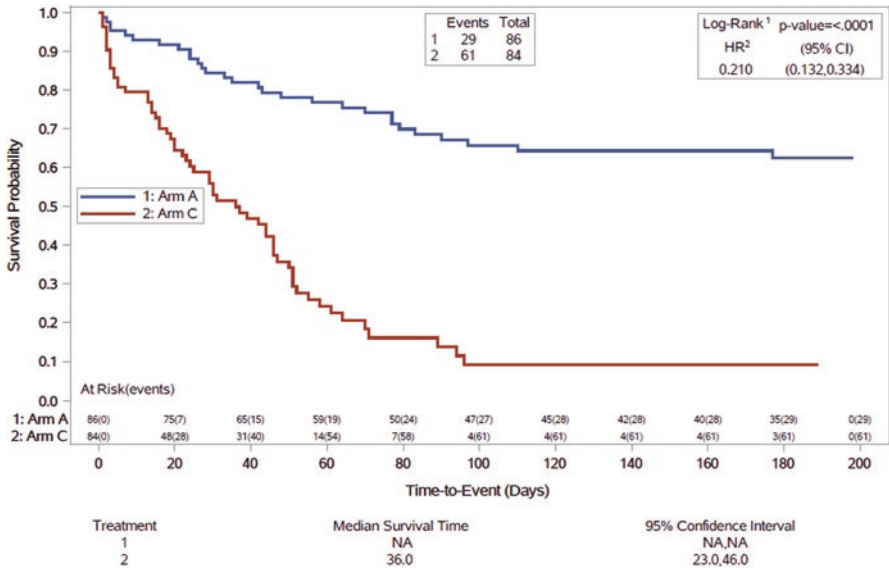


Fig. 1 Time to progression-free survival

one Kaplan–Meier figure, for example). Or it could be even more reduced – just to essentially one conditional power metric (likelihood of eventual success). The DMC will assess if the specified criteria were met and act accordingly, influenced (as agreed upon in the charter and previous discussions with the sponsor) by the totality of the data including the outputs focused on safety and study integrity. Fig. 1 shows what information might be contained in the Kaplan–Meier figure of time to progression-free survival (PFS) at a formal assessment of efficacy. Information in the top-right corner and the bottom might be removed if this PFS was being presented simply as a way to present efficacy as a possible counter-balance for safety concerns.

In other cases, additional efficacy data is provided to help the DMC fully assess the situation. This could include secondary endpoints and sensitivity analyses. The DMC, for example, would want information on the components of a composite endpoint to understand which of the components is driving a treatment effect (is it the most clinically compelling, or one that is less so). Sometimes, subgroups are presented in table or graphically as a forest plot so that the DMC can assess consistency of treatment effect across different subgroups. Some sponsors are dogmatic in how the DMC should operate – enforcing a binding recommendation on the data. Others are more flexible and provide the DMC the philosophy of the decision-making but encourage the DMC to use their full judgment. In either case, hopefully there is a Sponsor Liaison who can, if needed, have a confidential but frank discussion with the DMC if the recommendation to be made is not clearly obvious.

The DMC might see, for example, Kaplan–Meier curves where the curves cross and a trend for efficacy emerges later in the study follow-up. Based on that, the DMC might decide not to recommend stopping for futility even if numerically the

results have met the futility criteria. Or, at minimum, bring this up for discussion with the Sponsor Liaison. The DMC might use information such as events that have not yet been through adjudication to help with decision making if the results using only the adjudicated data are close to the specified threshold. Another example is where the study demonstrated impressive early results that met the criterion for success, but the DMC proposed (and Sponsor Liaison agreed) that the safety database was not quite mature enough and that there was value continuing the study another short period of time to gain enough safety data that would be convincing to regulators and the clinical community.

If p-values are included in any way, be it for safety, informal assessment of efficacy, formal assessment of efficacy, or otherwise, it is important that the DMC interpret those p-values properly. There is a long history of p-values being misinterpreted. Here are the six principles from the American Statistical Association regarding p-values (Wasserstein and Lazar, 2016) [5]:

- (i) P-values can indicate how incompatible the data are with a specified statistical model.
- (ii) P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
- (iii) Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
- (iv) Proper inference requires full reporting and transparency.
- (v) A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
- (vi) By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.