

Data and analysis considerations in oncology clinical trials

Abstract

Background: Oncology clinical trials are distinct from trials in other disease areas in its unique patient population, treatment and toxicity monitoring, endpoint assessment and follow-up. Promising innovative therapies to treat and prevent cancer have made oncology a fertile ground to conduct clinical research.

Purpose: To raise awareness and discuss relevant data and analysis issues that are critical to the ultimate success of oncology clinical trials.

Methods: We review data collection, cleaning, and analysis considerations in oncology clinical trials in the area of dosing, adverse events, tumor assessments, and survival follow-up. Operational issues relating to statistical analysis milestones and validation are also discussed.

Results: Upfront planning and careful considerations in data collection, monitoring, cleaning, and analyses have major impacts on the quality and conduct of the trial. Clear and appropriate data presentations not only enhance the interpretability of the results but also boost the confidence in the analyses.

Conclusion: Collaboration and coordination among multiple stakeholders and different functions of the trial sponsor are essential at all stages of a clinical trial. Special emphasis should be given to systematic approaches to collect appropriate data and to monitor data issues as early as possible to ensure quality in execution and clarity in reporting of oncology clinical trials.

Keywords: oncology, clinical trial, treatment emergent adverse event, dose intensity, RECIST criteria, data cut-off date, event-driven analysis, analysis validation

Abbreviations: AE, adverse event; CDISC, clinical data interchange standards consortium; AR, adherence ratio; CRF, case report forms; DI, dose intensity; EDC, electronic data capture; FDA, food and drug administration; ICH, international conference on harmonisation; OS, overall survival; PD, progressive disease; PD-I, prescribed dose intensity; PFS, progression free survival; P-RDI, prescribed relative dose intensity; PT, prefer term; RDI, relative dose intensity; RECIST, response evaluation criteria in solid tumors; TEAE, treatment emergent adverse event

Introduction

Clinical trials have become a more complex undertaking. Trials that enroll patients globally are commonplace. Biomarker trials involving companion diagnostics become more sophisticated and require coordination and seamless execution from multiple stakeholders. Innovative trial designs involve upfront planning and attention to details in execution. These factors along with regulatory requirements have placed a greater burden on the sponsors to ensure that the data collected from the trials have unimpeachable integrity, quality, and validity to draw conclusions with respect to the objectives of the trials.

Data are arguably the ultimate measure of clinical trial performance. Appropriate trial design, well-thought-out case report forms (CRF) and clear CRF completion instructions are among the first steps to ensure data are consistently collected in order to achieve the trial objectives. These efforts are augmented by ongoing data cleaning and monitoring to ensure data accuracy. Monitoring visits are one of the most resource-intensive activities to ensure valid data are collected against source documents. The US Food and Drug Administration (FDA) has provided guidance¹ on an integrated

approach to monitor clinical trial quality and subject safety through a risk-based centralized monitoring approach. The main idea is to implement a central, systematic, and ongoing review of data which allows the monitoring to be more focused and data-driven. Continued advances in technology, data sharing standards including Clinical Data Interchange Standards Consortium (CDISC),² and the increasing sophistication of data visualization tools promise to increase efficiency and reduce clinical trial cost.

In this article, we focus on data collection and analysis issues in oncology clinical trials. In particular, we will take a deeper dive on data related to dosing, adverse events (AEs), tumor assessments, and overall survival. Special emphases are given to considerations in collecting relevant data and in identifying data issues in oncology clinical trials. We strongly believe that the biggest bang for the buck in monitoring trial conduct and ensuring quality comes from early identification of issues and trends so preventative and remedial actions can be put in place to avoid costly damages to the trial. Analysis and reporting considerations are also discussed to highlight the need for effective and clear presentation of data.

Methods

Treatment exposure

In oncology trials, treatment exposure is not only informative in its own right but also acts as an indirect measure of tolerability and even efficacy.

Actual versus prescribed dose: Dosing information is commonly captured in a log format with key elements such as start date, stop date, actual dose, and dose change reasons. When a dose is withheld

due to AE, a record with zero actual dose and dose change reason as “withheld due to AE” is expected. When dosing records do not cover a period of time continuously, it is common to assume that there are no doses given during the gap. However, one cannot be certain if this is due to data entry error (that there is actually no dosing gap). Even though it's cumbersome, dosing records should cover the treatment period continuously with zero dose records always entered and the appropriate dose change reason provided when doses are not given (either intentionally withheld due to AE or missed by accident by patients).

Some sponsors collect prescribed dose in the dosing log along with the actual dose. We think it provides useful information and suggest that care be given to the CRF design. In a hypothetical example, a patient may be prescribed with a reduced dose due to AE but accidentally doesn't take the dose. A CRF with both the prescribed and actual dose along with the respective reasons of dose change can accurately capture the scenario. Table 1 outlines a sample dosing CRF incorporating the actual dose and prescribed dose.

Pill Counts: For orally administered treatment, bottles of pills are dispensed to patients to be taken home. Patients are usually asked to keep dosing records in a diary. Upon return to the clinic, they are to return the unused pills and are dispensed with a new supply. As a part of the protocol procedure, sites are required to review and discuss the dosing diary with patients and to account for the returned pills. In addition, some sponsors design specific CRFs to capture each transaction of dispense and return of the bottles and pills. In theory, one can take the difference between each dispense and return to obtain the exact number of pills consumed and therefore the total amount of dose. In reality, such efforts do not always yield reliable results. Patients may forget to bring back a dispensed bottle at a visit; a bottle may be lost; sites may only guesstimate the number of returned pills as the actual counting is time-consuming for busy clinics. Any one of these issues poses challenges to rely on pill counts to capture actual dose administered. We view data collected from pill counts as a nice-to-have and potentially redundant to the dosing information collected in the log format described above.

Dose intensity and compliance: Dose intensity (DI) is expressed as the amount of dose per unit of time. It is calculated by dividing the actual cumulative dose by treatment duration. It is common in trials where treatment is given until disease progression. Dose intensity calculation is straightforward based on the above definition. However, in settings where treatment is given for a fixed period of time, for example in the adjuvant breast cancer setting where the standard of care is one year of trastuzumab treatment after surgery, for patients with HER2-positive disease, alternative definition of dose intensity may be relevant. If the protocol specified treatment period is for one year and a subject has stopped treatment after 6 months on drug, dose intensity may be calculated for this patient as the cumulative dose over 6 months or over one year. The former definition may be viewed as dose intensity conditional on the patient being treated. The latter definition insists on the protocol specified treatment duration regardless of the actual treatment duration (unconditional). While we are not advocating one definition of dose intensity over the other as the emphasis may be different depending on the situation, we recommend greater clarity and transparency when reporting results. In calculating treatment duration, it is common practice to add the dosing interval after the last dose. If prescribed dose is collected, prescribed dose intensity (P-DI) can also be obtained similarly.

Relative dose intensity (RDI) is calculated as the (actual) dose intensity divided by the protocol specified dose intensity. It compares the actual dose and schedule captured in dose intensity with the protocol specified intensity (protocol specified dose and schedule). With the exception of over-dose (which should be monitored and flagged), a higher RDI represents closer adherence to the protocol specified dose and schedule. Prescribed relative dose intensity (P-RDI) can also be obtained analogously reflecting prescribed rather than the actual dose and schedule relative to the protocol dosing specification. One may also be interested in calculating adherence ratio (AR), defined as the ratio between the (actual) dose intensity divided by the prescribed dose intensity. AR compares the actual dose relative to the prescribed dose with the understanding that prescribed dose has taken into account dose changes, if any, prescribed by the treating physician. It can be shown that AR may also be expressed as RDI/P-RDI.

Table 1 also includes hypothetical data of a patient to illustrate dose intensity and adherence calculations. By providing the details, we wish to highlight the nuances in summarizing dose exposure and intensity.

Dose intensity is a single summary measure incorporating both dose and treatment duration. When a treatment is given in cycles, as is common in oncology, summary measures such as dose received per cycle and percent of planned dose received per cycle may be more appealing. However, dose intensity and relative dose intensity may still be relevant as they capture delays of cycle time due to the need to manage treatment toxicities.

Adverse events

The collection of adverse events include description of the event, start date, stop date, seriousness, severity, relationship to study drug, actions taken and outcome. Typically, AEs are collected during the screening phase after informed consent has been signed and continues until the end of the protocol specified safety follow-up time. Ensuring the quality of safety data capture is the foundation of all pharmacovigilance activities. Common checks of the AE data include review of overlapping dates when more than one episodes of the same event of different severity are reported. Actions taken due to an AE should also be cross-checked against dosing data if it leads to a dose reduction or against the end of treatment reason if it leads to the discontinuation of treatment. A more complete suite of cross checks with other data in the database such as labs, concomitant medication, death, or hospitalization should be a part of the standard built-in data cleaning process.

Treatment-emergent adverse events: Because AEs are collected after the date the informed consent is signed which may be prior to the initiation of treatment, treatment emergent adverse events (TEAEs) are often the primary focus of the safety analyses. The idea of TEAE is to include any AE that occurs or worsens after the initiation of treatment and before a pre-defined period of time (e.g. 30 days) after the last dose of treatment. ICH E9 [3] defines TEAE as an event that emerges during treatment but having been absent during pre-treatment, or worsens relative to the pre-treatment state. There are some controversies when implementing this. The following (Figure 1 & Table 2) present four cases that illustrate the key issues. It seems that cases 1 and 3 may not be considered as TEAEs since case 1 may be considered as a continuation of the same event and case 3 may be a continuation of the same event but with a lower grade. However, for cases 2 and 4, the AEs are recorded by investigators as separate episodes starting on Day 5. Because the severity grades are no worse

than the pre-existing episode, one may argue the events are pre-existing and should not be counted as TEAEs. Following this logic, what about events of the same severity which start on Day 15 rather than on Day 5? One may be hard pressed to not include separate events long after a subject has received the first dose and has continued to receive doses. We advocate a simple and inclusive approach in that any AEs with a start date on or after the first dose date (and within the pre-defined period after the last dose) should be considered as TEAE regardless of pre-treatment conditions. The TEAE summary should be complemented by a separate summary of any AEs not considered

as treatment emergent (AEs occurred pre-treatment or post the pre-defined period after the last dose). When applicable, and depending on the disease area, more in-depth summaries of TEAE taking into account of pre-treatment conditions may be explored. Finally, we want to emphasize the importance of providing clear and specific CRF completion instructions to the sites on adverse events that should be recorded as separate events and events that should be considered as continuing without an end date. In our experiences, this is an area that sponsors should proactively engage the sites during the study setup phase.

Table I A sample dosing CRF and dose intensity calculations

Start date	Stop date	Prescribed daily dose	Actual daily dose (mg)	Reason for prescribed dose change	Reason for administered Dose change (if different from prescribed dose)
1/1/2000	1/2/2000	200	200		
1/3/2000	1/4/2000	200	0		01
1/5/2000	1/5/2000	200	200		
1/6/2000	1/10/2000	100	100	01	
Prescribed dose change codes:					
01 Reduced due to adverse events					
02 Withhold due to adverse events					
03 Resumed to previous dose level					
04 Other (specify)					
Administered dose change codes;					
01: adverse event					
02: Other					
In the above example,					
$DI = (200\text{mg/day} \times 2 + 200\text{mg/day} \times 1 + 100\text{mg/day} \times 5)/10 = 110 \text{ mg/day}$					
$P-DI = (200\text{mg/day} \times 5 + 100\text{mg/day} \times 5)/10 = 150 \text{ mg/day}$					
$RDI = DI / (200\text{mg/day}) = 55\%$					
$P-RDI = P-DI / (200\text{mg/day}) = 75\%$					
$AR = DI/P-DI = 73.3\%$					
If the protocol specified treatment duration is 20 days, then the patient stopped after only 10 days of treatment. An alternative (unconditional) dose intensity and relative dose intensity may be derived as the following:					
$DI = (200\text{mg/day} \times 2 + 200\text{mg/day} \times 1 + 100\text{mg/day} \times 5)/20 = 55 \text{ mg/day}$					
$RDI = DI / (200\text{mg/day}) = 27.5\%$					

Incidence by severity grade: A popular way to present AE incidence table is to layer in the severity grade. For illustration purpose, we provide a simple example to highlight the issues when incorporating worst grade in the AE summary. In this example, ten subjects are given treatment A and have experienced AEs with system organ class of nervous system disorders (with preferred terms of headache and dizziness) and eye disorders (with preferred terms of dry eye and eye pain). Table 3 appears to yield some surprising results where, in the column of worst grade 1, 5 patients have reported grade 1 headache which is larger than the 2 patients reporting any grade 1 AE and the 4 patients reporting grade 1 nervous system disorders as the system

organ class. This may be counter intuitive but is not an error. A patient experiencing a grade 1 headache is counted as grade 1 under headache (preferred term), but the same patient also has a grade 3 event (e.g., vision blurred) and is counted in the “Any AE” row under worst grade 3. The same applies when a patient has different severity of two preferred terms associated with the same system organ class. This can be, admittedly, quite confusing. We recommend that AE tables by severity grade using the layout as in Table 3 be shown with preferred terms (PTs) only, skipping the customary first row of “Any AEs” and the system organ class rows. A separate summary table on system organ class may be provided. An alternative layout (Table 4)

avoids the issue by displaying the by-grade summary for PT only in rows rather than in columns. This is the recommended layout if one

wants to include both the system organ class and preferred terms in the table.

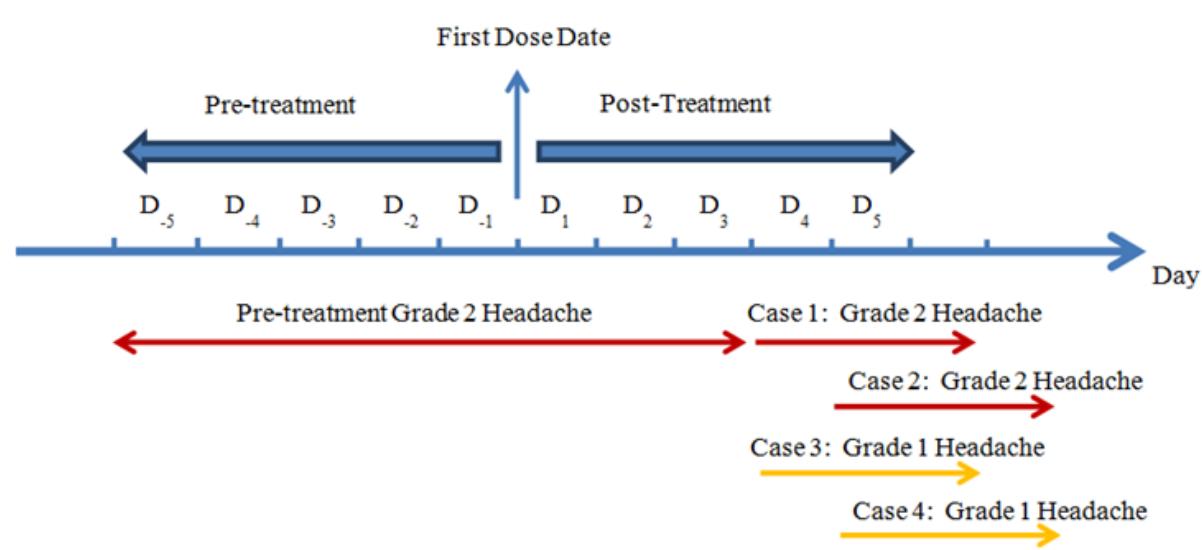


Figure 1 Determining treatment emergent adverse events (TEAEs).

Table 2 Determining treatment emergent adverse events (TEAEs)

Pre-treatment AE records	Post treatment AE (recorded as a separate entry from the pre-treatment record)	Comments
Grade 2 headache started 5 days before the first dose, ended on Day 3 (Day 1 is the first dose)	Case 1 Grade 2 headache started on Day 4. Case 2 Grade 2 headache started on Day 5 Case 3 Grade 1 headache started on Day 4. Case 4 Grade 1 headache started on Day 5.	Data query, potentially the same event as the pre-treatment. A new event of the same grade. A continuing event of a lower grade. A new event of a lower grade.

Tumor assessments

Tumor may be assessed over time by multiple modalities in oncology trials. These assessments contribute to the definition of key endpoints such as response rate and progression free survival (PFS). Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 [4] is a set of standardized and widely accepted criteria used to assess tumor. Any modifications by the sponsor of the criteria should be documented in the protocol and in a radiology review charter.

Automated response assessment: There are three components in the overall tumor response assessment when RECIST criteria are used: target lesion response, non-target lesion response, and the appearance of a new lesion. Each individual lesion identified at the baseline (target or non-target lesion) are tracked and entered into the CRF. Investigator assessments of response of the target, non-target, and new lesions along with an overall response assessment are provided at each protocol specified tumor assessment time. Typically, investigators are expected to do their own derivations following the

RECIST criteria and then provide the response assessment. Invariably, some errors occur along the way as some tumor burden calculations are in reference to baseline and some are in reference to the nadir (e.g., progressive disease). Sponsors may pick up these errors during the data cleaning process and issue queries. Since these queries may arrive long after the assessments have been done, investigators may or may not recall all the details why a particular response assessment is given. The time lag can be overcome by implementing a real time automated response derivation within the EDC so the investigators will get immediate feedback on any derivation errors or inconsistencies with the RECIST. Investigators should be allowed to overwrite the automated response assessment and ideally provide reasons for the discrepancy. We believe such an automated response system implemented in real time will help improve the efficiency of data cleaning and more importantly enhance the quality of the data. In addition, the investigators are given the flexibility to differ from the automated derivation to reflect the autonomy of investigator assessments.

Table 3 AE by worst grade with system organ class and preferred term

	Arm A (N=10)					
	All grades	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Any AEs	10	2	3	3	1	1
Nervous system disorder	7	4	2	1	0	0
Headache	6	5	1	0	0	0
Dizziness	6	3	2	1	0	0
Eye disorder	8	1	3	2	1	1
Vision blurred	6	1	3	1	1	0
Eye Pain	5	1	2	1	0	1

Table 4 AE by worst grade with system organ class and preferred term

	Arm A (N=10)
Any AEs	10
Nervous System Disorder	7
Headache	6
Grade 1	5
Grade 2	1
Grade 3	0
Grade 4	0
Grade 5	0
Dizziness	6
Grade 1	3
Grade 2	2
Grade 3	1
Grade 4	0
Grade 5	0
Eye Disorder	8
Vision Blurred	6
Grade 1	1
Grade 2	3
Grade 3	1
Grade 4	1
Grade 5	0
Eye Pain	5
Grade 1	1
Grade 2	2
Grade 3	1
Grade 4	0
Grade 5	1

Date of response and date of progression: In addition to the overall response assessment at each time point during the study, assessment date is also critical when it comes to defining a time-to-event endpoint such as PFS. Some sponsors require the date associated with the overall response assessment be provided in the CRF. Other sponsors require dates in the individual CRFs of target or non-target lesions but not dates of the overall response assessment. Regardless of the approach, clear instructions will need to be provided to the sites on how to fill in the dates when they are expected to provide an overall response assessment date based on potentially different dates from each component of the assessment. The same instructions should be provided to data management and statistical programming staff tasked to derive the dates for analysis purposes. As a general principle, consistent with the FDA guidance,⁵ for progression based on unequivocal progression of non-target lesion or a new lesion, the disease progression (PD) date is the earlier date when PD is detected. If progression is based on target lesions where scans may be obtained at different times, the target lesion PD date is the date of the later scan when complete information is available to establish PD. For the overall response assessment when PD may be due to target lesions, non-target-lesion, or new lesions, the earliest date when PD is identified from any source is the date of PD. However, for time-to-response endpoint (partial response or complete response), the date of response is the last date when complete information is available to establish response.

Central radiology review: While RECIST is the standard for solid tumor assessments, its implementation in practice and consistency across global sites in a clinical trial setting can be challenging. As a result, independent radiology review is set up to assess response in a centralized and blinded fashion (although the trial may be open label).^{6,7} For the central review to be successful, a great deal of coordination and collaboration is essential among the study sites, sponsors, vendors, and central review providers. A charter is required to outline all the requisite steps and the handoffs. Data quality checks are important to be imbedded in the central review process.

Recognizing the independent nature of the central review, we recommend minimum sponsor involvement in influencing the data review. However, we encourage a thorough testing of the central review system at the beginning of the trial using a suite of pre-determined test cases to ensure that the data checks in place are compliant with the RECIST so the quality of the data can be relied upon by the sponsor at the time of the analysis.

Overall survival: Overall Survival (OS) is an objective endpoint and it's the gold standard to demonstrate clinical benefit in oncology trials.

A key consideration in trial design and analysis when OS endpoint is involved is to minimize loss to follow-up. The trial protocol must draw clear distinctions between withdrawal from treatment and withdrawal from follow-up. Patients may withdraw treatment for any reason; however, every effort should be made to ensure that patients are followed up for OS and long term safety after terminating treatment. It's common to focus on deaths when the number of deaths may be used to trigger the analysis. Equal emphasis should be given to the adequacy of follow-up to patients who are censored. When there are substantial gaps in follow-up, censoring patients using the snapshot date is not advised. In addressing the loss to follow-up, some sponsors undertake public record search to identify death information. Such an approach is not without controversies as it may introduce bias because of different local regulatory and legal requirements. If data from public records are used in the analyses, we recommend the information obtained from this source be clearly flagged in the data to allow for additional sensitivity analyses.

Statistical analysis

Design and analysis methods: For pivotal oncology trials, group sequential designs⁸⁻¹⁰ allow for flexibility of one or more interim analyses and are the most commonly used designs. A complexity is introduced when more than one primary endpoint is involved in a group sequential design. The group sequential Holm procedure¹¹ has been proposed to handle both multiple primary endpoints (e.g. PFS and OS) and multiple analyses while controlling for the overall type I error rate.

Study adaptation¹²⁻¹⁴ has been an area of fertile research although it has not gained wide-spread adoption in practice. Adaptations such as sample size re-estimation, switch of hypotheses (e.g. between non-inferiority and superiority), switch of analysis populations defined by a biomarker, dropping of an arm in a seamless phase 2/3 design are among the most researched. The growing interests in adaptive designs have prompted the U.S. FDA to issue a guidance document.¹⁵

Bayesian designs and analyses have been adopted in phase 1 or phase 2 oncology trials with aims including dose finding or identification of winners among multiple treatments. For example, the I-SPY 2¹⁶ trial is a phase 2 randomized trial which used an adaptive design based on biomarker subtypes to evaluate novel agents in the neoadjuvant breast cancer setting. It employed response adaptive randomization based on Bayesian modelling of accumulating data. I-SPY 2 has the potential to speed the process of screening drugs with the goal of bringing safe and effective new drugs to market more efficiently.

Event trigger and data cut-off: In oncology trials, analyses are commonly event driven, i.e. the analyses will only occur when a pre-specified number of events (e.g. OS or PFS) are reached. Event goals are specified in the protocol and statistical analysis plan to ensure appropriate study power. In tracking the events, sponsors typically employ a statistical model¹⁷⁻¹⁹ to project the time the event goal may be reached based on accumulating data. When the event goal is nearly reached, sponsors may project a data cut-off date in the future upon which all events up to the date will be included in the analysis whether the final number of events is more or less than the event goal. Alternatively, a sponsor may decide a data cut-off date upon actually observing the event goal. The data cut-off date needs to be communicated to all sites to ensure not only the event goals are reached as close as possible but also a clear understanding from sites that all data on or prior to the cut-off date must be entered and

verified in a timely fashion. For interim analysis, if a different number of events from the protocol is obtained, some adjustment is necessary in the analysis, for example, the interim boundaries.

Data snapshot and filtering: For a live clinical database with EDC, data entry across global sites may occur at any moment in time either before or after the data cut-off date. Since the data cleaning and query resolution are unlikely to finish on the same day of data cut-off, a period of several weeks are commonly required to accomplish these activities. By the time a snapshot of the database is taken, a mixture of cleaned data (data prior to the cut-off date) and uncleaned data (data post the cut-off date and up to the snapshot date) is unavoidable.

This begs the question whether all data contained in the snapshot should be the basis for the analysis. Strictly speaking, the intended analysis is to include data up to the data cut-off date both from the perspective of event triggering and data cleaning. The fact that one ends up with additional data post the cut-off date has more to do with practicality than the intent. Sponsors opting to emphasize the original intent will necessarily want to filter out data contained in the snapshot that are beyond the data cut-off date. This is not a simple undertaking and requires careful planning. For example, an adverse event may have a start date before the data cut-off date but with a stop date post the data cut-off date. The stop date should not be used in calculating duration of the AE if one wants to strictly enforce the data cut-off in the analysis. It's also possible that an event (e.g. death) occurred post data cut-off. Assuming it's agreed that the death should not be counted as event in the main analysis, it's not always clear whether knowing the subject's death post data cut-off can be used to infer that the subject is alive on the data cut-off date (therefore the subject is censored on the date). We recommend using only data within the declared data cut-off date for all analysis purposes since there's no assurance that all sites enter data in a timely manner post the data cut-off date.

While "data filtering" described above is tempting with sound rationales, the benefits may not outweigh the time and efforts involved. This is especially true for interim analyses when data cleanliness is not of paramount concern compared to data timeliness. For final analyses, these extra efforts may be worthwhile.

Validation: Statistical analysis of the final or interim data is the climax of a clinical trial. The final analysis represents the fruit of labor from the patients, sites, sponsor and numerous collaborators around the world. The stakes are even higher when a phase 3 trial is reported. Since the results may be considered material for many publicly traded pharmaceutical and biotech companies, time is a luxury that the sponsors do not have. To ensure accuracy of the results released to the sponsor management and then to the public, meticulous planning and seamless execution are essential. In our experiences, analysis outputs are typically generated by statistical programmers working in collaboration and under the guidance of statisticians. Programmers are expected to follow sponsor standard procedures in performing validation which forms the first line of defense. However, this alone may not be adequate as programmers tend to focus on data processing from raw data to analysis data (e.g. data conforming to CDISC standards) and from analysis data to analysis displays. We recommend a parallel, independent validation path from raw data directly to analysis results performed by statisticians or programmers (more than one validators if necessary) for key endpoints so analyses are connected to the raw data. It reduces the risk of relying on the intermediate datasets and boosts the confidence on the final results.

Conclusion

Oncology clinical trials are among the most exciting and challenging. Enormous amount of interest, resources, time, and money are invested in oncology drug development. In the end, it all boils down to the data. Data quality and integrity are the cornerstones of all clinical trials. We have highlighted specific issues to consider in collecting, monitoring, and analyzing a wide range of data collected in oncology trials. These are by no means exhaustive nor should they be limiting. We have deliberately emphasized on issues that may be taken for granted in practice. We hope by sharing our experiences and recommendations, more awareness is raised toward designing and executing high quality oncology trials.

Acknowledgement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest statement

The Authors declare that there is no conflict of interest. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. <http://www.fda.gov/RegulatoryInformation/Guidances/ucm127004.htm> 2006.
2. <http://www.cdisc.org/standards-and-implementations>.
3. <http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/statistical-principles-for-clinical-trials.html>.
4. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45(2):228–247.
5. <http://www.fda.gov/downloads/Drugs/Guidances/ucm071590.pdf>.
6. <http://www.dovepress.com/subjective-endpoints-in-clinical-trials-the-case-for-blinded-independent-peer-reviewed-article-OAJCT#.pdf>.
7. <http://www.fda.gov/RegulatoryInformation/Guidances/UCM268555.pdf>.
8. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika*. 1977;64(2):191–199.
9. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979;35(3):549–556.
10. Lan KKG, Demets DL. Discrete sequential boundaries for clinical trials. *Biometrika*. 1983;70(3):655–660.
11. Ye Y, Li A, Liu L, Yao B. A group sequential Holm procedure with multiple primary endpoints. *Stat Medicine*. 2013;32(7):1112–1124.
12. Bauer P, Köhne K. Evaluation of experiments with adaptive interim analyses. *Biometrics*. 1994;50(4):1029–1041.
13. Cui L, Hung HMJ, Wang SJ. Modification of sample size in group sequential clinical trials. *Biometrics*. 1999;55(3):853–857.
14. Müller HH, Schäfer H. Adaptive group sequential designs for clinical trials: Combining the advantages of adaptive and of classical group sequential approaches. *Biometrics*. 2001;57(3):886–891.
15. <http://www.fda.gov/RegulatoryInformation/Guidances/UCM201790.pdf>.
16. Park JW, Liu MC, Yee D, et al. Adaptive Randomization of Neratinib in Early Breast Cancer. *Engl J Med*. 2016;375(1):11–22.
17. Ying GS, Heitjan DF. Weibull prediction of event times in clinical trials. *Pharm Stat*. 2008;7(2):107–120.
18. Ying GS, Heitjan DF, Chen TT. Nonparametric prediction of event times in randomized clinical trials. *Clinical Trials*. 2004;1(4):352–361.
19. Bagiella E, Heitjan DF. Predicting analysis times in randomized clinical trials. *Statistics in Medicine*. 2001;20(14):2055–2063.