Review: Clinical Trial Methodology

CLINICAL INVESTIGATION

Adverse event monitoring in oncology clinical trials

Clin. Invest. (2013) 3(12), 1157–1165

Jeffery S Russell & A Dimitrios Colevas*

Division of Oncology, Department of Medicine, Stanford University Medical Center, Stanford University, Stanford, CA 94305, USA

*Author for correspondence:
E-mail: colevas@stanford.edu

Treating oncology patients with potentially toxic drugs is balanced with the premise of 'do no harm.' Most drugs are approved by the US FDA based on analysis of the clinical benefit-to-risk ratio. As such, adverse event (AE) reporting is a crucial aspect of oncology clinical trials. Attribution of symptoms can be difficult because there can be a lack of clarity concerning what is disease-related, treatment-related, a co-morbid illness or a combination of all three. Additionally, as cancer progresses, both the symptoms and treatments evolve, resulting in a complex, time-dependent relationship. If toxicities that are not actually the result of an investigational agent are inappropriately attributed to that agent during Phase I development, further study of a potentially useful drug may be delayed or abandoned. In contrast, if there is an underreporting of drug-attributable AEs in Phase I development, there will be unnecessary patient risk and expense with subsequent clinical development when the relationship between the AEs and the drug emerges. Given the large amount of data generated from clinical trials, the reliability of collecting and recording these results in modern multimodality oncology trials is increasingly complicated. Accurate toxicity reporting is integral for a true assessment of drug efficacy, economic evaluation and regulatory decision making.

Keywords: adverse drug reaction reporting systems • adverse events • clinical trials • oncology

Most patients with advanced cancer have a median of 13 symptoms or complaints; this symptom burden decreases a patients' quality of life (QOL) [1,2]. Adverse events (AEs) arising in oncology clinical trials may reflect the toxicity of therapy or be a sign of the underlying disease. Documentation of symptom status is essential in understanding the benefit and safety of a new drug or treatment regimen and can have a significant long-term economic impact on the regulatory and funding environment for that agent's development. AEs in oncology clinical trials are considered as side effects, complications, toxicity and/or morbidity of a tested compound or a regimen, and can occur in the acute setting or as a late complication. AEs can be symptoms, physical exam finding, abnormal laboratory results or irregular radiology reports. In general, while many AEs may have no direct impact on a patient's QOL and may not be associated with specific symptoms, any event that may eventually lead to an undesirable experience should be considered an AE. Examples of this might be an abnormal laboratory value not associated with symptoms, such as an elevated serum alkaline phosphatase or thrombocytopenia without bruising or bleeding.

Cancer clinical trial protocols typically detail a defined list of specific AEs that investigators prospectively evaluate. Prospectively defined AEs of interest on a protocol should typically include known drug-related side effects, such as heart failure for trastuzumab or pulmonary fibrosis with bleomycin. There is also a system for reporting unexpected and emergent AEs as well, but without a prospective plan for querying these latter AEs, collection and attribution are likely to be less accurate and complete.

FUTURE SCIENCE part of fsg

AEs are represented by a medical documentation system to designate the organ system affected. Once the affected organ system is selected, an investigator must assign a grade to designate the severity of the AE. While a general paradigm of mild/moderate/severe/life threatening/fatal corresponds to grades 1/2/3/4/5 for AE grading, in reality, each AE grading scale is unique to that particular symptom or finding. For laboratory- or symptom-based AE grades, the numerical severity assignment may not be closely linked to the above paradigmatic association. Severe AEs that are usually assigned as grade 3 or higher should not be confused with the largely regulatory term. A serious AE (SAE) in this case is specifically defined as an AE that: results in death; is life threatening; requires inpatient hospitalization or prolongation of existing hospitalization; results in persistent or significant disability; or any congenital abnormality or birth defect [3].

Finally, the investigator must attribute the AE as a result of the protocol-specific intervention or another cause (that is: unrelated to drug; unlikely related to drug; possibly related to drug; probably related to drug; or definitely related to drug). Of course, this attribution can be highly subjective.

AEs are useful for monitoring the safety of drug dosing and scheduling regimens. However, one must keep in mind that the commonly used AE scales such as the Common Terminology Criteria for AEs (CTCAE) were developed for clinical trial purposes and may not actually reflect or be applicable to real-world patient experiences outside the context of therapeutic clinical trials. For example, some agents may be associated with a plethora of AEs according to laboratory assay criteria, but these so-assigned AEs based on clinical trial data may not have an actual impact on patient health or well-being. Historically, both objective data (i.e., white blood cell count or AST/ALT levels) and subjective data (i.e., muscle aches or fatigue), as observed and documented by investigators, were the basis for AE assignment, grading and attribution. More recently it has been the practice to attempt to use patient-reported outcomes (PROs) of the subjective AEs to improve the accuracy, efficiency and patient relevance of AE data collection (to be discussed below).

Several questions arise when considering the collection and reporting of AEs: first, who are AEs for? Initially, AEs, as specified in formal AE grading tools, were used by clinical researchers and drug developers to determine the toxicity burden of a new treatment. Second, are these AE capture tools validated? Third, is the collection of data best performed by the investigator or do patient-reported outcomes enhance the quality of data? In addition, AE reporting will be different based on the focus of the trial (therapeutic vs. supportive interventions) as well as the size of the trial (i.e., a large Phase III vs a small pilot Phase I). This article aims is to provide an overview of the development of the current AE reporting criteria in use with oncologic therapeutic trials and outline a general discussion of limitations and future directions of AE capture, reporting and analysis.

Defining patient symptoms & AEs

AEs range from minor subjective complaints and asymptomatic clinical changes to life-threatening injuries or death. In 1979, the WHO came out with a handbook and guidelines in order to standardize the data reported for investigational agents as well as to determine the benefits of a particular therapy versus the cost in toxicity [4]. This was a first attempt to develop a common language to describe the outcome and response to cancer treatments and was developed as the result of several international conferences and research organizations.

In the USA, the NCI established the Common Toxicity Criteria system (CTC v1.0) in 1983 in order to evaluate toxicities of chemotherapy (Table 1) [5,6]. Its goal was to provide standards for the description and exchange of safety data reported in oncologic clinical trials. Grading criteria for the CTC v1.0, while often subjective, were designed to separate what is tolerable injury from life-threatening events (Table 2). According to the CTC scale, grade 1 events are minor, asymptomatic

Table 1. Development of the common terminology criteria for adverse events.

Year	Name	Adverse events terms (n)	Notes
1982	CTC v1.0	>40	Acute adverse events of chemotherapy
1998	CTC v2.0	>300	Expansion of terms
2003	CTCAE v3.0	>1000	Inclusion of adverse events relevant to all modalities (medical, surgical and radiation) as well as special populations (pediatrics)
2009	CTCAE v4.0	>3000	Updated mapping to MedDRA†; minor edits

†MedDRA is a proprietary coding scheme only available by paid subscription.

CTC: Common terminology criteria; CTCAE: Common Terminology Criteria for Adverse Events; MedDRA: Medical Dictionary for Regulatory Activities.

Table 2. Common terminology criteria for adverse events v4.0 grading scale examples.

Grade	Generic adverse events	Nonanalytical adverse events (anorexia)	Analytical adverse events (hypokalemia)
0	No adverse events or within normal limits	–	–
1	Mild	Loss of appetite without alteration of eating habits	<LLN–3.0 mmol/l
2	Moderate	Oral intake altered without significant weight loss or malnutrition; oral nutritional supplements indicated	<LLN–3.0 mmol/l; symptomatic; intervention indicated
3	Severe, not life threatening	Associated with significant weight loss or malnutrition; tube feedings or TPN indicated	<3.0–2.5 mmol/l; hospitalization indicated
4	Life threatening, urgent intervention required	Life-threatening consequences; urgent intervention indicated	<2.5 mmol/l; life-threatening consequences
5	Death	Death	Death

LLN: Lower limit of normal; TPN: Total parental nutrition.

occurrences and do not impair functional end points. Interventions or medications are generally not required. Grade 2 events are moderate in nature, usually symptomatic and local treatment or outpatient medications may be used. These events may limit some patient function but should not impair activities of daily living. Grade 3 events are severe and consist of multiple distressing symptoms. Often hospitalization, intravenous medications or even surgery may be necessary. Grade 4 events are life threatening, may be disabling and can include the loss of an organ or organ function. Grade 5 events result in death. One major exception to this grading system is laboratory-based grade assignment. For example, a patient may have an absolute neutrophil count of <500 cells/mm^3, and this would be assigned a grade 4 severity, even though a particular patient may experience no actual undesirable experience from this level of neutropenia. Additionally, the CTC grading system does not account for possible worsening of the severity grade, but merely captures the initial severity experienced by the patient.

Additional guidelines, such as the Late Effects of Normal Tissue Scale, were developed in 1995 in order to try to capture late toxicity effects focused on radiation therapy trials [7,8]. Similarly, the Radiation Therapy Oncology Group (RTOG) and the European Organization for Research and Treatment of Cancer developed AE guidelines to monitor the acute and late side effects associated with radiotherapy [9]. While these two radiation therapy-specific scales enhanced investigators' understanding of AEs within individual trials, comparing results between trials was not possible because there was no straightforward 'translation' or mapping of AEs across these multiple grading severity scales. Thus, the need for a comprehensive AE monitoring system was clearly identified.

In 1998, an updated CTC v2.0 contained a more comprehensive dictionary of AEs including those associated with radiotherapy treatments; however, this version's radiation relevant AEs focused only on acute radiotherapy injuries [5]. In 2003, CTCAE v3.0 was further revised to evaluate pediatric, surgical and late AEs and to improve AE-reporting mechanisms [10]. The name of the NCI AE tool was changed from CTC to CTCAE to emphasize that AEs were not necessarily toxicities. It was believed at the time that 'toxicity' was a biased expression in terms of AE attribution that might have led to skewed AE reporting and assignment. Additional v3.0 guideline recommendations included acute and late effects criteria merged into a single AE without a time-based reference and instead investigators were urged to report all events as they occurred during treatment. AEs should be applied globally and not have a specific modality identification (except for certain intra-operative injury AEs). Duration of AEs would not be captured within the CTCAE but instead by iterative evaluations over time and the CTCAE should not be used to rank AEs (i.e., AEs should only describe the specific AE; grade 2 diarrhea is not necessarily better or worse than grade 2 nausea or grade 2 thrombocytopenia).

In 2009, CTCAE v4.0 was redesigned for the adoption of the Medical Dictionary for Regulatory Activities (MedDRA), a set of medical terminology agreed upon by the NCI, industry and regulatory agencies including the US FDA and European Medicines Agency (formerly the European Agency for the Evaluation of Medicinal Products) [11]. This system, developed by the International Conference on Harmonization, defines AEs as any unfavorable and unintended sign, symptom or disease temporally associated with the use of a medical treatment or procedure that may or may not be relevant to

the medical treatment or procedure. MedDRA allows for more descriptive mapping of medical terms, but lacks the severity scales previously associated with the CTCAE. As a result, AE terms in MedDRA are organized into System Organ Class (SOC) groupings (Box 1). SOC groupings are based on anatomical, physiological, etiology or purpose (i.e., laboratory values). The CTCAE v4.0 has been reorganized along SOC lines so that CTCAE terms are easily mappable to MedDRA terms. CTCAE subsections are then used to designate a severity grading scale, which are not in the MedDRA definitions. Furthermore, the SOC naming system is dynamic and can be expanded or modified as the need arises.

AE categories cover a diverse spectrum of data including laboratory values, radiologic findings and subjective patient symptoms. Since there is no duration parameter in either the CTCAE or MedDRA, there must be another layer of data capture, requiring multiple AE assessments over time, in order to best capture the time-dependent nature and severity of an AE. Until recently, the practice was to collect and report only the 'worst grade/severity' in a particular 'treatment cycle' or span of time. This methodology led to under representation of the AE burden especially for ongoing AEs or AEs that recur in the specified time interval. For example, the clinical relevance of a grade 4 oral mucositis during radiation for 7 days duration is obviously less severe than the same grade AE lasting 4 weeks, but until recently these events would be captured and reported as identical, unless there were prespecified plans to capture the AE burden over time. One such AE evaluation scheme developed by Trotti *et al.* called 'TAME' demonstrated the increasing toxicity burden to altered fractionation and chemoradiation treatments for head and neck cancer patients had gone undetected with the traditional AE capture methods [12]. In addition, this slant toward 'worst grade/severity' case reporting may result in 'AE migration' and an under representation of low-grade, but early and persistent AEs.

Box 1. Common terminology criteria for adverse events V4.0: system organ classes.

- Blood and lymphatic system disorders
- Cardiac disorders
- Congenital, familial and genetic disorders
- Ear and labyrinth disorders
- Endocrine disorders
- Eye disorders
- Gastrointestinal disorders
- General disorders and administration site conditions
- Hepatobiliary disorders
- Immune system disorders
- Infections and infestations
- Injury, poisoning and procedural complications
- Investigations (laboratory tests)
- Metabolisms and nutritional disorders
- Musculoskeletal and connective tissue disorders
- Neoplasms benign, malignant, and unspecified
- Nervous system disorders
- Pregnancy, puerperium and perinatal complications
- Psychiatric disorders
- Renal and urinary disorders
- Reproductive system and breast disorders
- Respiratory, thoracic and mediastinal disorders
- Skin and subcutaneous tissue disorders
- Social circumstances
- Surgical and medical procedures
- Vascular disorders

Grading the severity of AEs follows the general principles, as mentioned above, of grades 0–5, with grades 1–4 representing the vast majority of events. However, certain toxicity findings require special analysis and often these evaluations have not been completely validated. Cut-offs between severity grades for many AEs are arbitrary and are not necessarily clinically relevant. As an example, what makes diarrhea of seven small volume stools per day (a grade 3 AE) worse than diarrhea of six large volume stools per day (a grade 2 AE)? Additionally, what is the impact of revisions to the CTCAE lexicon when evaluating AEs? Liu *et al.* evaluated oral mucositis by CTCAE v3.0 versus CTCAE v4.0 using validated head and neck QOL surveys with respect to oral mucositis in a population of nasopharyngeal patients treated with induction chemotherapy followed by chemoradiation [13]. They found that CTCAE v4.0 had higher correlation coefficients to the QOL surveys than CTCAE v3.0, but one must also keep in mind that no formal gold standard for objectively measuring oral mucositis is presently available.

Another example, is hearing impairment (ototoxicity) which has often been difficult to quantify clinically as well as objectively. Two recent published standards (The Chang Scale and The Brock Scale) for evaluating ototoxicity in the pediatric population have been published, but there is no universal gold standard that is known to be validated with respect to clinically relevant outcomes [14,15]. CTCAE v3.0 determines hearing impairment based on the numeric loss of decibels of hearing at any frequency; however, the revised CTCAE v4.0 now includes decibel loss at specific adjacent frequencies, as the loss of lower frequency hearing ranges has a larger impact on patient symptoms than the loss of higher frequencies [16]. It is also important to note that the lack of gold standards when determining grade severity has resulted in many AE definitions being based on consensus recommendations from committee review rather than from results of evidence-based medicine trials.

Another underdeveloped area of the CTCAE is patient-based input on subjective symptom reporting. During the medical interview, a physician will frequently categorize

patient symptomatic complaints. Data demonstrate that this method of recording subjective data results in an under-reporting of the severity of patient symptoms [17,18]. A symptom can be defined as a patient-observed subjective evaluation of disease or a physical alteration. The severity of those symptoms on the patient and the impact on normal function are considered as the 'symptom burden' [19], but an investigator functioning as interpreter and collector often applies a filter, consciously or not, to the subjective data from the patient.

PROs have been demonstrated to improve the accuracy of AE reporting. Several studies had found that physicians and nurses underestimate symptom onset, frequency and severity in comparison with patient reports. Fromme *et al.* found that physician evaluation of chemotherapy-related symptoms in prostate cancer patients was neither specific nor sensitive in detecting AEs [20]. Many AEs are quite subjective and only the patient can appropriately quantify these symptoms. As such, the NCI has developed a web-based platform to collect patient reports of symptoms during treatment to enhance AE reporting [101]. Currently, 81 symptoms from CTCAE v4.0 have been evaluated as appropriate for patient reporting. The goal of the PRO-CTCAE system is to generate a patient-reported AE system within cancer trials that is widely accepted and used by clinicians, investigators and regulatory bodies [21,22].

In 2006, Kirkova *et al.* performed a meta-analysis of assessment tools to collect oncology patient symptoms and found 21 instruments, ranging from the collection of two symptoms to 75 symptoms, with approximately 15 studies having undergone some type of validation testing [23]. Items such as pain, fatigue and anorexia were the most common symptoms included and instruments were a combination of numerical, visual or verbal scale assessments. One of the limitations of the majority of these tools was establishing appropriate timelines of symptoms with many focused on 'at the present time' or 'within the last few weeks' and included poor documentation of symptom duration. No single instrument was determined as 'ideal' for assessing oncologic symptoms; however, several of the tools were able to demonstrate benefit in specific symptom-focused assessments. Recently, an independent working group (Assessing the Symptoms of Cancer Using Patient-Reported Outcomes) has also established guidelines and recommendations for the measurement of cancer-related fatigue in clinical trials as well as to include patient symptoms as an end point measurement in oncology clinical trials [19,24,25].

QOL assessments have been used previously to monitor patient-perceived outcomes in a variety of trials. However, two limitations exist. First, not all QOL tools fully capture AEs [26]. In addition, some AEs such as anemia or thrombocytopenia are not patient reportable and may be under-reported. In contrast, the CTCAE approach to AEs does not directly capture a true QOL measure. Spitzer *et al.* developed a visual scale to represent patient-perceived QOL within the last week, which has been validated; but this tool does not provide any details of symptoms or patient experience [27,28].

One must ask, what is relationship between QOL and AEs? Huschka *et al.* found a broad range of agreement between QOL assessment and AE evaluation in a pooled analysis of lung cancer clinical trials. Using multi-item assessments the range was 44–74%. Interestingly, the AE with the least agreement to QOL rating was anorexia, while the AE with the highest agreement to QOL was constipation [26]. AE evaluations are not designed to report patient-perceived problems, but to record the incidence and severity of AEs observed by clinicians that would require medical intervention or lead to a change in the design of clinical trials. QOL studies often also evaluate the dimensions of emotional and spiritual events that cannot be reflected as AEs. Clearly, QOL assessment and AE reporting have areas of overlap as well as distinct areas of separation; however, a further discussion of these topics would require a separate review.

Due to the large amount of data that could be collected from patient-reported symptoms, some clinicians have identified a subset of symptoms that are similar across many cancer types and are associated with significant patient distress [19]. Symptoms such as fatigue, pain, poor appetite, nausea, distress, disturbed sleep and depression are often considered as sentinel symptoms [29–34]. A balance between limited patient reporting of symptoms as well as some general measure of QOL should be included in the full evaluation of AEs encountered during a clinical trial. It is likely that for each trial, if prospectively determined PROs, clinical and laboratory AEs are selected, the choice will be trial specific, as a generic set of symptoms is not likely to be optimal for most clinical trials.

Data collection instruments for AEs

One significant question around use of the CTCAE is whether the system has been formally validated. While CTCAE had not undergone a formal validation analysis, its widespread use for over 30 years supports its relevance and usefulness as a tool in oncology clinical trials. Accuracy of CTCAE data is highest with objective data, such as laboratory values and lowest with subjective events. Data collection typically occurs during unstructured patient interviews. Often patient symptoms are recorded by physicians, extracted from paper or electronic medical records (EMRs), transcribed into trial databases and frequently reported to trial sponsors for placement into another database. Each step uses time and resources and could possibly introduce several sites for errors. The physician collection methods can be thought of as a passive collection tool. In contrast, a formal structured survey

provided to the patient to collect a specific list of AE items would be considered as an active collection tool.

The symptom tracking and reporting (STAR) system has been used by patients to directly report disease- and treatment-related symptoms [35]. PRO measures will often require upfront training sessions as well as monitoring of compliance. Through the STAR system, physicians have been able to review patient-reported data and modify as indicated to complete CTCAE grading. Basch and colleagues have reported that clinicians agreed with patient assessment of AEs approximately 92% of the time, lowering severities 5% of the time and raising them 3% of the time [36]. A recent report using the STAR system in an active clinical trial demonstrated that patients were able to complete on-site web-based questionnaires approximately 99% of the time and clinicians were able to review data and complete CTCAE assessments over 98% of the time. Additionally, they felt the STAR system was easy to use and spent an average of 3 min reviewing and assigning CTCAE grades [37].

Dorr *et al.* performed a review of several clinical trials of imatinib to compare the quality of clinically documented SAE outcomes to those reported from institutional review board AE documentation [38]. SAE descriptions were more complete (95 vs 40.3%) and were more able to assign causality (93 vs 26%) in the primary clinical data than reports from the institutional review board AE descriptions. These quality differences were primarily due to unstructured AE reporting forms. Earley *et al.* demonstrated that in a random sampling of clinical trial records, multiple instances of missing data as well as discrepancies in deaths reported in the clinical trials records versus the final publications were noted [39,102]. Scharf and Colevas reviewed 22 clinical trials and found that 27% of the published AEs could not be matched to agent-attributable AEs in the NCI clinical data update system [40]. Furthermore, in 14 out of the 22 articles, the number of high-grade AEs in clinical data update system differed by 20% or more when compared with the published trial data. Another concern is the reporting of AEs after a clinical trial has been completed and FDA has approved the drug. As an example of under-reporting of AE events in a postclinical trial, nononcologic setting, Moore *et al.* determined that only approximately 1% of >33,000 AEs with warfarin were formally reported to the FDA monitoring system [41].

As the authors are aware, technology in medicine has made significant advances with the EMR as well as a variety of web-based recording tools and databases. Surprisingly, many centers still use a large volume of paper forms when collecting data for clinical trials. Many different groups including local institutions, governmental agencies, as well as commercial entities, have developed a wide variety of electronic record-tracking systems for clinical trials. London *et al.* reported that an automated, computer-based AE tracking system resulted in faster confirmation by the principle investigator, improved workflow and resulted in more comprehensive AE reporting [42]. Within the NIH and NCI, an automated clinical trials suite was developed (Cancer Biomedical Informatics Grid), and a separate module, the cancer AE reporting system [103], was implemented to track AEs based on the CTCAE and MedDRA lexicons. It is likely that in the future there will be more and more direct extraction of AEs from the patient via PROs collected electronically, as well as more sophisticated extraction of clinical and laboratory AEs from EMRs.

Analytic tools for AEs & metrics

Presenting AE data was traditionally through the use of summary tables of incidence. Data across multiple time points are frequently compressed into a 'worst grade/severity' approach. This method, while simplifying the data, does not reflect multiple or sequential events and results in bias and a systematic under-reporting of toxicity reflecting on treatment regimens that may appear less toxic than they actually are. A complete analysis of toxicity may report the general incidence of an AE event, group AEs into general subsections (i.e., hematologic vs nonhematologic), and/or present AEs grouped by reporting by low grade versus high grade events.

Limitations of current AE recording systems are that they are unable to fully capture AEs generated at different time points (duration) and poorly capture intermittent, dynamic events (number of episodes) during a course of treatment. For example, Machtay *et al.* found that over 40% of head and neck cancer patients treated on RTOG protocols experienced late toxicity events that were not captured in AE assessments during the original trial [43]. Trotti and Bentzen found that using three different grading systems of late AEs in head and neck patients undergoing chemoradiation found the reporting of grade 3 or 4 toxicity to be between 14 and 82% [44]. In addition, there is no current method of summarizing AEs into specific patient risk definitions. TAME analysis was developed to focus on short-term acute toxicity, adverse long-term events, treatment-related mortality and to present results in an end-result summary format [11]. Analysis of five RTOG trials in patients with head and neck cancer found an increase of approximately 500% of acute toxicity burden compared with prior analysis methods. Wang *et al.* demonstrated the development of four unique temporal clustering of symptoms during the course of treatment in a population of non-small-cell lung cancer patients undergoing chemoradiation [45]. Thus, the use of a longitudinal assessment of symptoms during multimodality treatment is useful to clearly define patient-reported experiences.

Statistical comparisons of incidence of AEs either intra- or inter-trial have not been formally assessed. If reporting of AEs was standardized, then inter-trials comparisons could have more validity when, for example, comparing data from two separate Phase II trials in order to compare toxicities of new agents or comparing toxicities of particular sequences of therapy. Obviously, additional rigorous research is required to confirm these concepts.

Conclusion

Capturing all grade and durations of AEs in multimodality cancer treatment is not necessary or even possible. The aim of AE reporting is to provide estimates of risk to help with both investigational and clinical decision making. Trials are likely to become more complex as combinations of standard chemotoxic drugs, molecularly targeted agents and biological modulators become the norm. In addition, the implementation of rapid 'real-time' toxicity monitoring programs will be necessary for earlier recognition of AE patterns to allow for rapid change in trial designs concerning patient interventions [46]. It is likely that as tools for PROs and automated prospective collection and analysis of AE data directly from EMRs becomes widespread, investigators and clinicians will face both the benefit of a more accurate and complete AE dataset and the challenge of interpreting and digesting much larger datasets.

The authors would like to provide the following general recommendations for the clinical investigator:

- Collect all severe AEs without regard to causality;
- Collect only intervention-associated low-grade AEs;
- Specify in advance the subset of trial-specific AEs of high priority;
- Develop a systematic tool to investigate high priority AEs;
- High-priority AE tools should use patient-reported outcomes for nonanalytical AEs;
- Collect AEs at baseline, with every treatment cycle and after treatment completion;
- Ensure there is a plan for determining recurrent versus persistent high-grade AEs;
- Ensure all AE data collected are reported consistently in the literature and to regulatory authorities;
- AEs should be presented consistently regardless of trial outcome.

Of course, some of these recommendations may need to be adapted based on the intent of the trial (supportive vs therapeutic) as well as size of the trial or purpose of the trial (Phase I vs III). Being aware of AEs or toxicities from preclinical or Phase 0/I trials may help focus intervention-associated AEs in later phase trials. Thus, many questions still remain to be answered about the best methods to evaluate, capture and report AEs in oncology clinical trials. Research is ongoing to determine if patient-reported outcomes are better for certain subjective AEs; two recent editorials by Basch *et al.* stress the importance of patient-reported outcomes and recommendation for PRO instruments to be developed for trials and to additionally have PRO performance measures to evaluate outcomes for accountability and quality improvement [47,48].

While AE-capturing systems have entered the digital age, no uniform or standard approach has been identified. Furthermore, the analysis of AE reporting techniques as well as the use of patient symptoms or AEs as clinical trial end points is a current area of research in need of additional development. It is important to remember that the CTCAE is only a lexicon used to define AEs and assign severity. Additional guidelines addressing the collection, presentation and analytical methods of AE evaluation in oncology clinical trials are needed to capture comprehensive AE data in order to provide useful results for the research community.

Future perspective

Therapeutic clinical trials are necessary to confirm the benefit of experimental drugs with acceptable levels of toxicity. Large trials are becoming more and more expensive to implement, and thus, trials need to be optimally designed, which includes capturing AEs from both patient reports and investigator evaluations. The authors predict that over the next 5–10 years the vast majority of trials will be captured electronically with a focus on patient reported outcomes. Electronic collection will increase the accuracy of documenting AEs and through appropriately designed interfaces, both the patient, investigator and regulatory oversight should be able to be performed with high levels of efficiency. The authors feel that new technology such as touch-screen tablets and secure cloud computing will become standard tools for assessing and monitoring AEs in oncology clinical trials.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Defining patient symptoms & adverse events

- The NCI established the common terminology criteria for adverse events (AEs) to capture AEs in therapeutic clinical trials as a result of chemotherapy, surgical or radiation treatments.
- AEs cover a spectrum of patient symptoms, laboratory values, clinical findings and radiologic examinations.
- Particular AEs are classified by grade or severity, which is specific to that individual AE.

Data collection instruments for AEs

- The growth of the electronic medical record has allowed for easier collection of AEs.
- However, currently there are no standard guidelines for reporting AEs in clinical trial publications.

Analytic tools for AEs & metrics

- Reporting of 'worst' grade or severity of AEs results in bias and frequently under-reports the true toxicity of an intervention.
- The use of the common terminology criteria for adverse events has limitations on effectively evaluating AEs over time or from multimodality treatments.
- Statistical comparisons of AEs between published trials has not been addressed.

References

1 Homsi J, Walsh D, Nelson KA *et al.* The impact of a palliative medicine consultation service in medical oncology. *Support. Care Cancer* 10(4), 337–342 (2002).

2 Portenoy RK, Thaler HT, Kornblith AB *et al.* Symptom prevalence, characteristics and distress in a cancer population. *Qual. Life Res.* 3(3), 183–189 (1994).

3 NCI. *NCI guidelines for investigators: adverse event reporting requirements for DCTD (CTEP and CIP) and DCP INDs and IDEs.* NCI, MD, USA (2012).

4 WHO. *WHO handbook for reporting results of cancer treatment.* WHO, Geneva, Switzerland (1979).

5 Trotti A, Byhardt R, Stetz J *et al.* Common toxicity criteria: version 2.0. An improved reference for grading the acute effects of cancer treatment: impact on radiotherapy. *Intl J. Radiat. Oncol. Biol. Phys.* 47(1), 13–47 (2000).

6 Cox JD, Stetz J, Pajak TF. Toxicity criteria of the Radiation Therapy Oncology Group (RTOG) and the European Organization for Research and Treatment of Cancer (EORTC). *Intl J. Radiat. Oncol. Biol. Phys.* 31(5), 1341–1346 (1995).

7 Rubin P, Constine LS 3rd, Fajardo LF, Phillips TL, Wasserman TH. EORTC Late Effects Working Group. Overview of late effects normal tissues (LENT) scoring system. *Radiother. Oncol.* 35(1), 9–10 (1995).

8 Denis F, Garaud P, Bardet E *et al.* Late toxicity results of the GORTEC 94–01 randomized trial comparing radiotherapy with concomitant radiochemotherapy for advanced-stage oropharynx carcinoma: comparison of LENT/SOMA, RTOG/EORTC, and NCI-CTC scoring systems. *Intl J. Radiat. Oncol. Biol. Phys.* 55(1), 93–98 (2003).

9 No authors listed. Late effects consensus conference: RTOG/EORTC. *Radiother Oncol.* 35(1), 5–7 (1993).

10 Trotti A, Colevas AD, Setser A *et al.* CTCAE v3.0: development of a comprehensive grading system for the adverse effects of cancer treatment. *Semin. Radiat. Oncol.* 13(3), 176–181 (2003).

11 NCI. *Common Terminology Criteria for Adverse Events v.4.0.* US Department of Health and Human Services, NIH, 28 May 2009.

12 Trotti A, Pajak TF, Gwede CK *et al.* TAME: development of a new method for summarising adverse events of cancer treatment by the Radiation Therapy Oncology Group. *Lancet Oncol.* 8(7), 613–624 (2007).

13 Liu YJ, Zhu GP, Guan XY. Comparison of the NCI-CTCAE version 4.0 and version 3.0 in assessing chemoradiation-induced oral mucositis for locally advanced nasopharyngeal carcinoma. *Oral Oncol.* 48(6), 554–559 (2012).

14 Chang KW, Chinosornvatana N. Practical grading system for evaluating cisplatin ototoxicity in children. *J. Clin. Oncol.* 28(10), 1788–1795 (2010).

15 Brock PR, Bellman SC, Yeomans EC, Pinkerton CR, Pritchard J. Cisplatin ototoxicity in children: a practical grading system. *Med. Pediatr. Oncol.* 19(4), 295–300 (1991).

16 Chang KW. Clinically accurate assessment and grading of ototoxicity. *Laryngoscope* 121(12), 2649–2657 (2011).

17 Bruner LH, Carr GJ, Curren RD, Chamberlain M. Validation of alternative methods for toxicity testing. *Environ. Health Perspect.* 106(Suppl. 2), 477–484 (1998).

18 Varricchio CG, Sloan JA. The need for and characteristics of randomized, Phase III trials to evaluate symptom management in patients with cancer. *J. Natl Cancer Inst.* 94(16), 1184–1185 (2002).

19 Cleeland CS, Sloan JA, Cella D *et al.* Recommendations for including multiple symptoms as endpoints in cancer clinical trials: a report from the ASCPRO (Assessing the Symptoms of Cancer Using Patient-Reported Outcomes) Multisymptom Task Force. *Cancer* 119(2), 411–420 (2013).

20 Fromme EK, Eilers KM, Mori M, Hsieh YC, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. *J. Clin. Oncol.* 22(17), 3485–3490 (2004).

21 Basch E, Iasonos A, Barz A *et al.* Long-term toxicity monitoring via electronic patient-reported outcomes in patients receiving chemotherapy. *J. Clin. Oncol.* 25(34), 5374–5380 (2007).

22 Basch E, Jia X, Heller G *et al.* Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. *J. Natl Cancer Inst.* 101(23), 1624–1632 (2009).

23 Kirkova J, Davis MP, Walsh D *et al.* Cancer symptom assessment instruments: a systematic review. *J. Clin. Oncol.* 24(9), 1459–1473 (2006).

24 Cleeland CS, Sloan JA. Assessing the Symptoms of Cancer Using Patient-Reported Outcomes (ASCPRO): searching for standards. *J. Pain Symptom Manage.* 39(6), 1077–1085 (2010).

25 Barsevick AM, Cleeland CS, Manning DC *et al.* ASCPRO recommendations for the assessment of fatigue as an outcome in clinical trials. *J. Pain Symptom Manage.* 39(6), 1086–1099 (2010).

26 Huschka MM, Mandrekar SJ, Schaefer PL, Jett JR, Sloan JA. A pooled analysis of quality of life measures and adverse events data in north central cancer treatment group lung cancer clinical trials. *Cancer* 109(4), 787–795 (2007).

27 Spitzer WO, Dobson AJ, Hall J *et al.* Measuring the quality of life of cancer patients: a concise QL-index for use by physicians. *J. Chronic Dis.* 34(12), 585–597 (1981).

28 De Boer AG, Van Lanschot JJ, Stalmeier PF *et al.* Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? *Qual. Life Res.* 13(2), 311–320 (2004).

29 Cella D, Paul D, Yount S *et al.* What are the most important symptom targets when treating advanced cancer? A survey of providers in the National Comprehensive Cancer Network (NCCN). *Cancer Invest.* 21(4), 526–535 (2003).

30 Lorenz KA, Dy SM, Naeim A *et al.* Quality measures for supportive cancer care: the Cancer Quality-ASSIST Project. *J. Pain Symptom Manage.* 37(6), 943–964 (2009).

31 Patrick DL, Ferketich SL, Frame PS *et al.;* National Institutes of Health State-of-the-Science Panel. National Institutes of Health State-of-the-Science Conference Statement: Symptom management in cancer: pain, depression, and fatigue, July 15–17, 2002. *J. Natl Cancer Inst. Monogr.* (32), 9–16 (2004).

32 Bennett MI, Rayment C, Hjermstad M, Aass N, Caraceni A, Kaasa S. Prevalence and aetiology of neuropathic pain in cancer patients: a systematic review. *Pain* 153(2), 359–365 (2012).

33 Fiorentino L, Rissling M, Liu L, Ancoli-Israel S. The symptom cluster of sleep, fatigue and depressive symptoms in breast cancer patients: severity of the problem and treatment options. *Drug Discov. Today. Dis. Models* 8(4), 167–173 (2011).

34 Martinelli F, Quinten C, Maringwa JT *et al.* Examining the relationships among health-related quality-of-life indicators in cancer patients participating in clinical trials: a pooled study of baseline EORTC QLQ-C30 data. *Expert Rev. Pharmacoecon. Outcomes Res.* 11(5), 587–599 (2011).

35 Basch E, Artz D, Dulko D *et al.* Patient online self-reporting of toxicity symptoms during chemotherapy. *J. Clin. Oncol.* 23(15), 3552–3561 (2005).

36 Basch E, Bennett A, Pietanza MC. Use of patient-reported outcomes to improve the predictive accuracy of clinician-reported adverse events. *J. Natl Cancer Inst.* 103(24), 1808–1810 (2011).

37 Pietanza MC, Basch EM, Lash A *et al.* Harnessing technology to improve clinical trials: study of real-time informatics to collect data, toxicities, image response assessments, and patient-reported outcomes in a Phase II clinical trial. *J. Clin. Oncol.* 31(16), 2004–2009 (2013).

38 Dorr DA, Burdon R, West DP *et al.* Quality of reporting of serious adverse drug events to an institutional review board: a case study with the novel cancer agent, imatinib mesylate. *Clin. Cancer Res.* 15(11), 3850–3855 (2009).

39 Earley A, Lau J, Uhlig K. Haphazard reporting of deaths in clinical trials: a review of cases of ClinicalTrials.gov records and matched publications-a cross-sectional study. *BMJ open* 3(1), pii: e001963 (2013).

40 Scharf O, Colevas AD. Adverse event reporting in publications compared with sponsor database for cancer clinical trials. *J. Clin. Oncol.* 24(24), 3933–3938 (2006).

41 Moore TJ, Bennett CL. Under-reporting of hemorrhagic and thrombotic complications of pharmaceuticals to the US Food and Drug Administration: empirical findings for warfarin, clopidogrel, ticlopidine, and thalidomide from the Southern Network on Adverse Reactions (SONAR). *Semin. Thromb. Hemost.* 38(8), 905–907 (2012).

42 London JW, Smalley KJ, Conner K, Smith JB. The automation of clinical trial serious adverse event reporting workflow. *Clin. Trials* 6(5), 446–454 (2009).

43 Machtay M, Moughan J, Trotti A *et al.* Factors associated with severe late toxicity after concurrent chemoradiation for locally advanced head and neck cancer: an RTOG analysis. *J. Clin. Oncol.* 26(21), 3582–3589 (2008).

44 Trotti A, Bentzen SM. The need for adverse effects reporting standards in oncology clinical trials. *J. Clin. Oncol.* 22(1), 19–22 (2004).

45 Wang XS, Fairclough DL, Liao Z *et al.* Longitudinal study of the relationship between chemoradiation therapy for non-small-cell lung cancer and patient symptoms. *J. Clin. Oncol.* 24(27), 4485–4491 (2006).

46 Goldberg RM, Sargent DJ, Morton RF, Mahoney MR, Krook JE, O'Connell MJ. Early detection of toxicity and adjustment of ongoing clinical trials: the history and performance of the North Central Cancer Treatment Group's real-time toxicity monitoring program. *J. Clin. Oncol.* 20(23), 4591–4596 (2002).

47 Basch E, Torda P, Adams K. Standards for patient-reported outcome-based performance measures. *JAMA* 310(2), 139–140 (2013).

48 Basch E. Toward patient-centered drug development in oncology. *N. Engl. J. Med.* 369(5), 397–400 (2013).

Websites

101 National Cancer Institute. Applied research: cancer control and population sciences. http://outcomes.cancer.gov/tools/pro-ctcae.html

102 Clinical trials. www.clinicaltrials.gov

103 caBig® adverse event reporting system. http://cabig.nci.nih.gov/tools/caAERS