Patient-Reported Outcomes and the Evolution of Adverse Event Reporting in Oncology

Andy Trotti, A. Dimitrios Colevas, Ann Setser, and Ethan Basch

ABSTRACT

Adverse event (AE) reporting in oncology has evolved from informal descriptions to a highly systematized process. The Common Terminology Criteria for Adverse Events (CTCAE) is the predominant system for describing the severity of AEs commonly encountered in oncology clinical trials. CTCAE clinical descriptors have been developed empirically during more than 30 years of use. The method of data collection is clinician based. Limitations of the CTC system include potential for incomplete reporting and limited guidance on data analysis and presentation methods. The Medical Dictionary for Regulatory Activities (MedDRA) is a comprehensive medical terminology system used for regulatory reporting and drug labeling. MedDRA does not provide for severity ranking of AEs. CTC-based data presentations are the primary method of AE data reporting used in scientific journals and oncology meetings. Patient-reported outcome instruments (PROs) cover the subjective domain of AEs. Exploratory work suggests PROs can be used with a high degree of patient engagement and compliance. Additional studies are needed to determine how PROs can be used to complement current AE reporting systems. Potential models for integrating PROs into AE reporting are described in this review. AE reporting methods will continue to evolve in response to changing therapies and growing interest in measuring the impact of cancer treatment on health status. Although integration of PROs into AE reporting may ultimately improve the comprehensiveness and quality of collected data, it may also increase the administrative burden and cost of conducting trials. Therefore, care must be used when developing health outcomes and safety data collection plans.

J Clin Oncol 25:5121-5127. © 2007 by American Society of Clinical Oncology

INTRODUCTION

Adverse event (AE) reporting is a critical component in the conduct and evaluation of clinical trials. Methods of AE reporting have evolved with the complexity of cancer treatments. Reporting the subjective manifestations of AEs has historically involved clinician elicitation, interpretation, and rating of signs and symptoms. There is growing interest in the use of patient-reported outcome tools (PROs) to collect symptom measures directly from the patient to improve the accuracy and efficiency of subjective AE data collection. In this review, we describe the strengths and limitations of the current AE reporting process, and discuss potential models for integrating PROs into AE reporting in oncology.

EVOLUTION OF AE REPORTING IN ONCOLOGY

The methods for reporting of AEs in oncology have evolved in response to new treatments and modalities. In the 1950s to 1970s, retrospective studies provided limited descriptions of adverse outcomes, and severity ranking was rare.1 The routine pursuit of clinical trials in the 1980s provided the opportunity for formal toxicity documentation and generation of toxicity profiles or safety reporting. Clinical trial AE reporting is now a multistep process that begins with a standardized terminology and grading system and ends in publication of toxicity profiles.2 Reporting has grown more structured in recent years, including regulations for serious adverse event (SAE) reporting and adoption of more uniform terminology.

The National Cancer Institute (NCI) Common Terminology Criteria (CTC) system is a longstanding empirically developed dictionary, or lexicon, designed for use in clinical trials to aid clinicians in the detection and documentation of an array of AEs commonly encountered in oncology.3,4 The CTC is maintained by the NCI Cancer Therapy Evaluation Program. It is a minimum requirement for reporting in NCI-sponsored trials, and has also been widely adopted by the pharmaceutical industry. It is the de facto standard for reporting AEs in the oncology clinical trial literature. CTC data may be
supplemented with other measures of adverse effects depending on the nature of the trial.

Development of the CTC was driven by the need to document severity and to narrow descriptions into a reasonably sized and common lexicon. Specific terms and severity descriptors were developed by expert panel consensus process. The terms have expanded and evolved during more than 30 years in response to the introduction of new agents and modalities. Limitations include lack of a formal validation process, potential for under-reporting (especially for the subjective elements), and variations in end-results data presentation. Although the CTC was designed exclusively for use in clinical trials, it is often used in routine care to guide treatment decisions including drug dosing and supportive care interventions.

Initially called Common Toxicity Criteria (1984 to 2002), in 2003 the CTC was revised substantially and relabeled as the Common Terminology Criteria for Adverse Events (CTCAE). The CTCAE includes more than 1,000 terms with improved anatomic site specificity and expanded criteria for surgical effects. The CTCAE represents the first comprehensive grading system for reporting both acute and late effects in oncology, and was the first attempt to cover AEs associated with all therapeutic interventions, including radiation and surgery.

The choice of specific terms and descriptors in the CTCAE was driven by the notion that terminology should be immediately comprehensible to those with a basic fund of medical knowledge. The severity descriptors should correspond to common-sense notions of mild, moderate, severe, or life-threatening events. Significant attention was paid to ensuring consistency in grading between similar pathophysiologic processes. For example, grading descriptors for cerebrovascular and cardiac ischemia, although not identical, were intended to capture similar indicators of severity including permanent and transient events. Similar criteria were applied to gastric versus urinary bladder perforation, for example. However, there was no attempt to ensure severity consistency across AEs as disparate as amino transferase and mood alteration, for example. The severity descriptors for AEs in the metabolic/laboratory category were chosen arbitrarily and without any intent for severity cross-comparison. For example, a bilirubin level of greater than 10 times the upper limit of normal will usually have more patient-discernable consequences than an isolated aminotransferase value exceeding 20 times the upper limit of normal, but the CTCAE assigns these abnormalities the same status on the severity scale.

As described by Bentzen (Fig 1), the universe of adverse effects of cancer treatment can be considered in four general domains that contain some overlap. Data collection using the NCI-CTCAE system involves three of these domains (excluding health-related quality of life [HRQOL]/PRO tools). Movement up the y-axis involves more specific and quantifiable end points and tests. As one moves to the right on the y-axis, end points are more directly recognized and reported by patients. Clinician-graded symptoms are reported by patients and translated by clinicians or research staff into medical terms and grades. Symptoms and subjective elements can also be reported directly by the patient using HRQOL and PRO tools. Objective elements are recognized and graded by clinicians using specific medical terms and are often noted on physical examination or part of a constellation of findings or syndromes. Analytic elements are laboratory, imaging, or other technology-based tests designed to identify and/or quantify abnormalities that may not be appreciated using other means. This model addresses traditional end points and methods used in clinical trials, and in some retrospective chart reviews. Grading descriptors for some CTCAE terms contain a mix of subjective and objective elements. Studies regulated by the US Food and Drug Administration use US Food and Drug Administration-specific guidance and may use other standardized medical terminology systems (see next section).

Traditional data collection methods for the symptom/subjective elements typically involve unstructured patient interviews. These methods may differ significantly between clinicians and specific studies. Final data capture involves clinician interpretation of patient reporting and determination of the most appropriate term and severity grade. Thus, the subjective domain may be associated with poor inter-rater reliability regarding grading consistency and completeness of capture. Symptom research has documented systematic under-reporting of symptoms (in number, severity, and time of onset/resolution) by clinicians compared with patients. Thus, an alternative approach is to collect subjective elements directly from the patient using specific HRQOL surveys or PRO tools (discussed later in this review).

The two predominant AE vocabularies are the NCI CTCAE and the Medical Dictionary for Regulatory Activities (MedDRA system). MedDRA is a medical terminology system developed by the International Conference on Harmonisation and is owned by the International Federation of Pharmaceutical Manufacturers and Associations acting as trustee for the International Conference on Harmonisation steering committee. The MedDRA system contains a large number (> 65,000) of coding terms covering a wide range of clinical information, including AEs, for all medical disciplines and covering multiple

![Fig 1. Adverse effects domains. NCI, National Cancer Institute. Adapted with permission.](image-url)
medical product areas. MedDRA does not involve severity ranking of AEs. The NCI system is a more limited dictionary of terms (~1,000) commonly encountered in oncology and emphasizes severity grading of each event, subdividing most terms into four basic grades (excluding grade zero [no event] and grade 5 [death] in some terms). CTCAE terms (but not grades) are mapped (translated) to MedDRA. Although MedDRA terminology is used for regulatory reporting purposes, CTCAE-based data presentations are the primary method of AE data reporting used in scientific medical journals and oncology medical meetings.

The US Food and Drug Administration and NCI have different charges: legal approval of drug claims versus support of clinical and basic research. The US Food and Drug Administration has largely dealt with single-agent drug approval. NCI-sponsored studies by cancer centers and cooperative groups are often concerned with evaluation of complex multimodality and multiagent treatment programs. As a result of increasing collaborations between industry and the NCI, there is growing overlap and simultaneous use of both dictionaries. This also increases data collection burdens and the cost of trial conduct. Both organizations require SAE reporting as a safety reporting process distinct from routine AE reporting. The NCI also requires expedited reporting of unexpected and selected high-grade events under the Adverse Event Expedited Reporting System.

The US Food and Drug Administration provides specific guidance regarding AE data collection, data retrieval and organization, risk assessment, data analysis, presentation for new drug applications, and labeling. The US Food and Drug Administration accepts MedDRA terminology as an industry standard, but does not require it in submissions or reporting. In practice, MedDRA is the terminology used for regulatory reporting in the United States. The CTCAE is designed as a clinical research tool. The NCI guidance focuses on terminology applications, but provides no specific guidance on data analysis or end-results reporting.

Most cancer agents used today were approved by the US Food and Drug Administration on the basis of single-agent activity. Relative to nononcology therapeutics, toxicity profiles of single agents in oncology reveal high rates of multiple AE end points. Moreover, many cancers are now treated with multiagent and multimodality therapies, generating even larger numbers of events. Traditional data collection and reporting methods are strained by the volume of AE data generated by such trials. Data regarding dozens of AE end points, each carrying four grades, quickly become difficult to summarize, interpret, and publish.

Traditional methods of AE data summary and presentation formats were borrowed from nononcology summary methods. These methods do not reflect the reality of multiple coincident and sequential events that often occur in oncology. The time dimensions of events, treatment cycles, and chronic effects are customarily collapsed into a single data point using variable forms of a worst-grade approach. Systematic under-reporting and bias using standard summary methods has been described, potentially leading to treatment regimens appearing less toxic than they actually are. There are no standard methods to summarize the overall burden generated by complex multimodality concurrent or sequential courses of treatment. Late effects influencing the HRQOL of cancer survivors are believed to be widely under-reported. The limitations of traditional methods to characterize fully the toxicity burden of cancer treatments have in part stimulated the HRQOL and PRO initiatives.

**HAVE THE CTCAE CRITERIA BEEN VALIDATED?**

Often the NCI is asked whether the CTCAE in general, or if specific AE terms/grading language, have been validated. Unlike HRQOL or PRO tools, CTCAE have not been through a formal development or validation process. However, some content validity is conferred through 30 years of widespread use and iterative improvements. In most cases, questions regarding validation concern collection, presentation, or interpretation of aggregate CTCAE data, rather than the CTCAE lexicon itself. CTCAE items have not been evaluated for inter-rater reliability or test-retest metrics.

Is there adequate consistency and accuracy of CTCAE data? Audits of clinical trials sponsored by NCI demonstrate that when events from the analytic domain are recorded and supported by primary documentation, such as a laboratory report for a metabolic/laboratory AE, or computerized tomogram for a pulmonary embolus, the AE and severity grade are reported accurately in the vast majority of cases. However, in the cases of more subjective AEs, it is impossible by audit to determine if the collection of the severity of pain, for example, was accurate.

The question of accuracy of documentation of highly subjective AEs such as pain is hampered to a certain extent by the lack of a gold standard of measurement. Before one can ask whether the CTCAE is useful as a tool for accurate capture of subjective AEs, one must ask whose truth is to be used when making that assessment. For example, is the physician’s report of a patient’s pain necessarily less accurate than a patient’s self-report? If so, under what circumstances is the patient’s report most accurate—when filling out a written questionnaire in the waiting area, when keeping a diary at home, or when speaking to the care provider? For example, Basch et al have found that for many AEs there is a high level of agreement between physician and patient assessment. However, for highly subjective AEs such as fatigue and dyspnea, the agreement is less strong and clinicians tend to assign a lower severity to these than does the patient. Conversely, an evaluation of symptoms in patients with rheumatoid arthritis treated during a randomized controlled trial suggests that patient-reported symptom improvement was superior to clinician-reported improvement. In fact, patients and clinicians may hold different concepts for subjective domains that have different or complementary applications in clinical care and research. Therefore, in such cases one can only address consistency and completeness of AE capture, rather than accuracy.

However, even in the case of high-fidelity collection of objective AEs such as laboratory values, final reporting of these values can be incomplete and inconsistent. For example, a study comparing data submitted to a regulatory sponsor versus data submitted for publication demonstrated inconsistencies in approximately 25% of all high-grade AEs. We know of no similar comparison using verbatim versus MedDRA-coded AE data or PRO-type data.
Use of PROs As Study-Specific End Point Tools

In the early development of HRQOL tools, data were collected predominantly via staff-administered surveys. After a period of questionnaire development and validation, patient self-reporting became the preferred method. PROs evolved from HRQOL tools to become more disease-specific and symptom oriented. PROs recently have become the favored method for collecting symptom end point data in trials to assess specific treatment benefits. These patient-reported end points have now been used as the basis for drug labeling claims for a limited number of oncology drugs including mitoxantrone (pain) and more recently eculizumab (HRQOL).

The use of PRO end points in clinical trials is an area of increasing interest to regulatory agencies. To clarify the role of patient-reported data in the drug approval process and to refine standards for PRO instrument development, in February 2006 the US Food and Drug Administration issued a draft guidance on PRO Measures. The NCI responded to the guidance by sponsoring the NCI Patient Reported Outcomes Assessment in Cancer Trials conference in September 2006 (described in this issue of the Journal of Clinical Oncology), at which the position was articulated that for subjective domains the patient’s own account should be considered the gold standard.

Use of PROs in Routine AE Monitoring in Clinical Trials

Unlike HRQOL and formal symptom end point assessment for which PROs are considered standard, AE data collection is performed exclusively by clinical staff during cancer treatment trials. Staff members are required not only to aggregate analytic AE data (ie, laboratory values) and objective AEs (ie, physical examination findings), but also subjective patient experiential information (ie, symptoms). In the latter case, the process of data collection is complex, with multiple steps of information transfer and vulnerability to errors of misinterpretation or omission (Fig 2). Alternatively, shifting to a model in which patient-reporting is the primary mechanism for monitoring subjective AEs may improve the quality and completeness of collected data, provide a more comprehensive picture of the patient experience, and improve the efficiency of clinical operations.

Initial research suggests that the majority of cancer patients are willing and able to self-report symptom AEs at clinic visits for up to a year (including those with heavy symptom burdens), and that clinicians will accept this information as the basis for management decisions and regulatory documentation. Comparisons of patient versus clinician CTCAE symptom reporting have found high levels of agreement for the majority of items, with most cases of disagreement by one point, and only rare cases in which a discrepancy would alter clinical management or regulatory reporting. An ongoing study in the NCI cooperative group, Cancer and Leukemia Group B (protocol 70501), is evaluating the feasibility and potential benefits of this approach in the setting of a multicenter treatment trial, with results expected in late 2008.

Use of PROs in Routine Cancer Care

As with clinical trials, patient self-reporting of symptoms has been suggested to enhance patient-clinician communication and assist in the early detection of toxicities in the routine care setting. Initial research has demonstrated the feasibility of electronically collecting symptoms from patients during chemotherapy and delivering printouts of this information to clinicians at appointments. However, at present, patient self-reporting is not commonly used to monitor patient status during routine cancer care.

PRO-Related CTCAE Items

The current version of the CTCAE consists of slightly more than 1,000 individual terms. For purposes of PRO tool application, these terms can be divided into four general categories: analytic technology-based measurements (such as laboratory tests or imaging results); objective items that rely on clinical expertise and judgment (such as physical examination findings); subjective items that reflect patient experience and can therefore only truly be known by the patient (symptoms); and mixed subjective/objective items that combine patient-reporting and clinician interpretation (such as drug reactions experienced by patients but requiring clinician confirmation). The latter two categories are potentially amenable to a patient-reporting model, including the approximately 127 (12%) subjective and 128 (12%) mixed subjective/objective items in the current version of the CTCAE (version 3.0). Thus, PROs potentially could play a role in 24% of all AE terms. However, these terms may represent a larger portion of all events collected, depending on the nature of the trial.

Conversion of CTCAE Items for Patient Reporting

The CTCAE was originally designed for clinician reporting, and the language used in question items therefore includes medical jargon and technical terminology. To provide versions of CTCAE items amenable to patient self-reporting, a layperson-friendly language adaptation of selected CTCAE subjective items (from CTCAE version 3.0) was created using methods described in the US Food and Drug Administration guidance for ensuring content validity of PRO instruments, including patient focus groups and cognitive debriefing sessions. A goal of this patient language adaptation was to preserve the concepts represented in each response option compared with those of the original CTCAE (Table 1). Although a high level of patient agreement for the majority of items, with most cases of disagreement by one point, and only rare cases in which a discrepancy would alter clinical management or regulatory reporting.

INTEGRATING PROs INTO THE CTCAE

![Diagram of symptom information flow in cancer treatment trials. CRA, clinical research assistant. Reprinted with permission.](image-url)
acceptance and comprehension of content was observed, it was found that reading and responding to items can be time consuming for patients, taking an average of 4.5 minutes for 15 items, leading to respondent questionnaire fatigue. CTCAE items include long descriptions for each grade criterion in the original design to ensure precision, which increases the risk of incomplete questionnaires and missing data. Therefore, briefer symptom items (for example, including only simple ordinal or visual analog scales) ultimately may be preferable for optimizing patient response rates (e.g., self-rating of worst pain during the last 24 hours, using a 10-point scale with zero representing “no pain” and 10 representing the “worst pain imaginable”). Development and evaluations of such briefer CTCAE PRO symptom items (and studies mapping responses to these back to the CTCAE) are needed.

Clinician acceptance of patient toxicity symptom self-reporting has been high, with most surveyed physicians and nurses believing that patient self-reports are an accurate reflection of clinical status; expressing willingness to base clinical decisions on these data; and agreeing that PROs should be collected routinely to monitor patients receiving chemotherapy.32,36 In an ongoing study at Memorial Sloan-Kettering Cancer Center (New York, NY), it is being assessed whether clinicians will alter their own symptom toxicity reporting when exposed to patient self-reports.

Potential Models

In the clinical trial setting, a potential model to integrate PROs could include a symptom AE checklist provided to enrollees at clinic visits (paper or electronic) with selected items drawn from the subjective and mixed subjective/objective CTCAE categories. Specific items could be chosen to include a core set of symptoms relevant to cancer patients receiving toxic therapies, as well as additional disease-specific and treatment-specific items. A free-text or drop-down option could be included to allow for patients to add other symptoms relevant to them individually. In this paradigm, subjective items would be directly reported by patients via toxicity reports unfiltered by clinicians, whereas items in the mixed category would undergo clinician interpretation before reporting. Any free text items added by patients would likely require clinician interpretation and severity grading, especially if used for regulatory reporting.

This model conceivably could be extended to the routine care setting as well, in which real-time review of this information could be required by clinicians to prompt discussion with patients or inform management decisions. Furthermore, electronic home patient-reporting could be integrated to monitor patient status between visits, not only at the time of office appointments.

### Table 1. Example Adaptation of CTCAE Item Into Patient Language

<table>
<thead>
<tr>
<th>Grade</th>
<th>NCI Description</th>
<th>Original Language</th>
<th>Patient Adaptation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mild</td>
<td>Mild pain not interfering with function</td>
<td>I have had mild pain, but it does not interfere with my normal functioning</td>
</tr>
<tr>
<td>2</td>
<td>Moderate</td>
<td>Moderate pain; pain or analgesics interfering with function, but not interfering with ADL</td>
<td>I have had moderate pain, and my pain or my use of pain medications interferes with my normal functioning. But I am still able to carry out my daily activities.</td>
</tr>
<tr>
<td>3</td>
<td>Severe</td>
<td>Severe pain; pain or analgesics severely interfering with ADL</td>
<td>I have had severe pain, and my pain or my use of pain medications severely interferes with my normal daily activities</td>
</tr>
<tr>
<td>4</td>
<td>Disabling</td>
<td>Disabling</td>
<td>My pain has been disabling</td>
</tr>
</tbody>
</table>

Abbreviations: CTCAE, Common Terminology Criteria for Adverse Events; NCI, National Cancer Institute; ADL, activities of daily living. Adapted with permission.17

The current model of AE reporting in oncology clinical trials allows for clinician discretion in choosing which items from the CTCAE merit reporting at any given patient visit (although some items may be required in protocols based on expected drug-related AEs). In the absence of specific preplanned safety questions, toxicity reporting is largely broad and passive. Large numbers of events of all grades are collected. Using standard summary methods, most low-grade events are disregarded in the reporting of final results, suggesting some inefficiency in data collection efforts. Safety reporting has many stakeholders, including patients, institutional review boards, sponsors, regulators, administrators, investigators, and industry. Multiple reporting requirements have increased data collection and reporting burdens. The introduction of HRQOL and PRO tools may add to the burden of data collection and reporting. Considering finite resources, determining the optimal mix of data collection and reporting requirements is challenging. One cooperative group has initiated policies proactively to actively manage the deluge of AE data in order to limit data collection and reporting burdens.42

### BALANCING DATA COLLECTION AND REPORTING BURDENS

The inherent complexity of AE outcomes in oncology makes the development of comprehensive reporting standards extremely challenging. Currently, NCI does not have specific guidance on the methods of AE data collection, data organization, analysis, or end-results reporting. This approach has advantages and disadvantages. Wide variations in reporting methods have arisen, making it difficult to compare safety outcomes among trials or cooperative groups.43 Conversely, overly specific guidance might constrain the flexibility needed to adapt reporting to study and agent-specific issues as they arise. Nonetheless, some additional reporting guidance would likely enhance the interpretability and comparability of AE end results.

The development and enforcement of reporting standards could also consume considerable time and resources. Whether such guidance should come from regulators, study sponsors, administrative...
centers, or special interest groups is an open question. One recent report suggests this should be done at the level of journal editors, using the CONSORT model.6,44

The use of more uniform approaches focusing on the most important safety issues in the context of a given disease and given protocol may generate more reliable and useful safety data. This need is brought into focus by the development of guidance for PROs, where the use of prespecified data collection tools, preplanned analyses, and formal analytic methods is expected. This model may also be useful in the development of more specific AE data analysis and reporting plans.

In an effort to improve the consistency and completeness of AE reporting in head and neck cancer, the Radiation Therapy Oncology Group has recently begun piloting new data collection methods using prespecified lists of AEs to guide patient screening (Radiation Therapy Oncology Group trials 0522 and 0615). Core acute and late neck AE end points are captured and graded by clinical staff (eg, dysphagia, mucositis, skin; chronic dysphagia, fibrosis, xerostomia, bone necrosis) in addition to drug specific issues (epidermal growth factor receptor–induced rash from cetuximab; bleeding from bevacizumab). Data collection of prespecified safety end points are required for each report period to enhance the completeness of critical AE data capture. Whether this approach will improve the completeness or interpretability of AE data will be evaluated and reported in the future.

**REFERENCES**

12. MSSO M: Mapping of CTC to MedDRA

**AUTHORS’ DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST**

The author(s) indicated no potential conflicts of interest.

**AUTHOR CONTRIBUTIONS**

Conception and design: Andy Trotti, Ethan Basch
Provision of study materials or patients: Ann Setser, Ethan Basch
Collection and assembly of data: Ann Setser, Ethan Basch
Data analysis and interpretation: Ann Setser, Ethan Basch
Manuscript writing: Andy Trotti, A. Dimitrios Colevas, Ann Setser, Ethan Basch
Final approval of manuscript: Andy Trotti, A. Dimitrios Colevas
PROs and the CTC