

# Patient free text reporting of symptomatic adverse events in cancer clinical research using the National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE)

Arlene E. Chung,<sup>1,2,3</sup> Kimberly Shoenbill,<sup>2,4</sup> Sandra A. Mitchell,<sup>5</sup> Amylou C. Dueck,<sup>6</sup> Deborah Schrag,<sup>7</sup> Deborah W. Bruner,<sup>8</sup> Lori M. Minasian,<sup>5</sup> Diane St. Germain,<sup>5</sup> Ann M. O'Mara,<sup>5</sup> Paul Baumgartner,<sup>9</sup> Lauren J. Rogak,<sup>10</sup> Amy P. Abernethy,<sup>11,12</sup> Ashley C. Griffin,<sup>2</sup> and Ethan M. Basch<sup>1,2,3,10</sup>

<sup>1</sup>Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA, <sup>2</sup>Program on Health and Clinical Informatics, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA, <sup>3</sup>Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA, <sup>4</sup>Department of Family Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA, <sup>5</sup>National Cancer Institute, Rockville, Maryland, USA, <sup>6</sup>Alliance Statistics and Data Center, Mayo Clinic, Scottsdale, Arizona, USA, <sup>7</sup>Division of Population Sciences, Department of Medical Oncology, Dana-Farber/Harvard Cancer Center, Brookline, Massachusetts, USA, <sup>8</sup>Nell Hodgson Woodruff School of Nursing, Winship Cancer Institute, Emory University, Atlanta, Georgia, USA, <sup>9</sup>Semantic Bits, LLC, Herndon, Virginia, USA, <sup>10</sup>Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, USA, <sup>11</sup>Department of Medicine, Duke Cancer Institute, Durham, North Carolina, USA, and <sup>12</sup>Flatiron Health, New York, New York, USA

Corresponding Author: Arlene E. Chung, MD, MHA, MMCI, 5034 Old Clinic Building, CB 7110 University of North Carolina School of Medicine, Chapel Hill, NC 27599-7110, USA (arlene\_chung@med.unc.edu).

Received 3 May 2018; Revised 17 October 2018; Editorial Decision 12 November 2018; Accepted 26 November 2018

## ABSTRACT

**Objective:** The study sought to describe patient-entered supplemental information on symptomatic adverse events (AEs) in cancer clinical research reported via a National Cancer Institute software system and examine the feasibility of mapping these entries to established terminologies.

**Materials and Methods:** Patients in 3 multicenter trials electronically completed surveys during cancer treatment. Each survey included a prespecified subset of items from the National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). Upon completion of the survey items, patients could add supplemental symptomatic AE information in a free text box. As patients typed into the box, structured dropdown terms could be selected from the PRO-CTCAE item library or Medical Dictionary for Regulatory Activities (MedDRA), or patients could type unstructured free text for submission.

**Results:** Data were pooled from 1760 participants (48% women; 78% White) who completed 8892 surveys, of which 2387 (26.8%) included supplemental symptomatic AE information. Overall, 1024 (58%) patients entered supplemental information at least once, with an average of 2.3 per patient per study. This encompassed 1474 of 8892 (16.6%) dropdowns and 913 of 8892 (10.3%) unstructured free text entries. One-third of the unstructured free text entries (32%) could be mapped post hoc to a PRO-CTCAE term and 68% to a MedDRA term.

**Discussion:** Participants frequently added supplemental information beyond study-specific survey items. Almost half selected a structured dropdown term, although many opted to submit unstructured free text entries.

Most free text entries could be mapped post hoc to PRO-CTCAE or MedDRA terms, suggesting opportunities to enhance the system to perform real-time mapping for AE reporting.

**Conclusions:** Patient reporting of symptomatic AEs using a text box functionality with mapping to existing terminologies is both feasible and informative.

**Key words:** PRO-CTCAE, patient-reported outcomes, free text, symptomatic adverse events, MedDRA

## INTRODUCTION

Symptomatic adverse events (AEs) such as nausea and fatigue are common among patients enrolled in cancer clinical trials.<sup>1</sup> Historically, this information has been collected and reported into research databases by clinical staff members using a set of AE grading criteria maintained by the U.S. National Cancer Institute (NCI), called the Common Terminology Criteria for Adverse Events (CTCAE).<sup>2</sup> Recently, the NCI developed a library of items for patient self-reporting of symptomatic AEs called the Patient-Reported Outcomes version of the CTCAE (PRO-CTCAE).<sup>2</sup> The PRO-CTCAE item library includes 78 symptomatic AE terms, each of which corresponds to an existing CTCAE item and maps to a discrete term in the Medical Dictionary for Regulatory Activities (MedDRA), which is a widely used lexicon for AE reporting in pharmaceutical industry clinical trials (please see [Supplementary Table S1](#) for PRO-CTCAE terminologies).<sup>3</sup> The PRO-CTCAE library is publically available from the NCI,<sup>4</sup> and is currently in use in cancer clinical trials sponsored by the NCI and the pharmaceutical industry.<sup>5</sup>

In clinical trials using PRO-CTCAE, a subset of items is preselected from the item library by the investigative team based on the anticipated toxicities of the therapy under evaluation.<sup>6</sup> These PRO-CTCAE items are then presented to patients at regular intervals during the trials to augment CTCAE grading of toxicities by clinicians. Often, there is a desire to minimize the length of surveys to minimize patient burden. However, it is recognized that patients in a trial may experience symptomatic AEs beyond those preselected by investigators for surveillance. To fully characterize the toxicity profile and tolerability of treatment, and to avoid possible ascertainment bias introduced by PRO-CTCAE item selection, there is interest to allow patients to provide supplemental symptomatic AE information.

The NCI developed a software system to enable administration of PRO-CTCAE items to patients enrolled in NCI-sponsored clinical trials.<sup>2</sup> This software system allows investigators to build study-specific, customized surveys for patient self-reporting of preselected AE items, and allows patients to add supplemental symptomatic AE information at the end of the survey via a free text box functionality. Patients are presented with an unlimited character text box in which to type any symptomatic AE(s) they wish to self-report (see [Figure 1](#)). As patients type in text, dropdown options dynamically populate with terms derived from the PRO-CTCAE item library and MedDRA Lowest-Level Terms (LLTs) (described below).<sup>3</sup> Respondents may select from the structured dropdown terms, or they can opt to continue typing unstructured free text, which is submitted in its verbatim form (see [Supplementary Table S2](#) for definitions of the types of supplemental symptomatic AE information).

Any time a patient selects a PRO-CTCAE term from the dropdown list, they are then asked to score its frequency, severity, and/or interference with daily activities as specified for that given item in the PRO-CTCAE library.<sup>1</sup> For unstructured free text or structured MedDRA dropdown terms, no scoring is solicited, and only presence of that symptomatic AE is recorded. Patients could provide an

unlimited number of dropdown or free text entries at the end of each PRO-CTCAE survey. Then, on the subsequent completion of a PRO-CTCAE survey, their prior supplemental information entry, whether unstructured or structured (dropdown selection), is presented to them to respond if that symptomatic AE is still present or resolved.

In this study, we describe the frequency with which patients electronically enter supplemental symptomatic AE information during conduct of 3 multicenter cancer clinical studies, characterize how often unstructured free text entries could be mapped to existing PRO-CTCAE or MedDRA terms, and identify commonly reported supplemental symptomatic AE terms that may be considered for future inclusion in the PRO-CTCAE item library. To our knowledge, this study is the first to report on the integration of electronic patient-reported, supplemental symptomatic AE data in clinical trials.

## MATERIALS AND METHODS

### Data sources

PRO-CTCAE surveys were periodically administered to patients in 3 prospective, NCI-sponsored, multicenter clinical studies using the PRO-CTCAE software system. Each of these studies invited participants to complete surveys containing between 14 to 48 PRO-CTCAE questions, and to add supplemental information about additional symptomatic AEs they were experiencing after the completion of each survey. In the NRG Oncology 1012 (Manuka Honey trial) randomized controlled trial (NCT01262560),<sup>7</sup> 152 patients with lung cancer receiving chemoradiotherapy across 80 sites completed PRO-CTCAE surveys via tablet computers at 5 consecutive clinic visits. In the Alliance N1048 Preoperative Radiation Or Selective Preoperative radiation and Evaluation before Chemotherapy and TME (PROSPECT) (PROSPECT trial) randomized controlled trial (NCT01515787), 668 patients receiving multimodality therapy for locally advanced rectal cancer at 165 sites in the United States and Canada completed PRO-CTCAE surveys remotely from home via web or an automated telephone interactive voice response system at weekly intervals for up to 12 weeks during active therapy, then every 6 months post-treatment for 3 years (PROSPECT trial accrual is ongoing, but the free text analysis was approved by the Data Safety Monitoring Board as an interim analysis). In the PRO-CTCAE validation study (NCT02158637),<sup>8</sup> 940 patients that had various types of cancer and were receiving outpatient systemic treatment at 9 U.S. cancer centers and community practices completed PRO-CTCAE surveys by tablet computer at 2 consecutive clinic visits. All patients in the 3 studies were receiving systemic cancer therapy (ie, chemotherapy, immunotherapy, or targeted therapy, in some cases with concurrent radiation). The first 2 of these studies (Manuka Honey and PROSPECT trials) were prospective clinical trials that incorporated PRO-CTCAE to capture symptomatic AEs, while the third was a validation study conducted to evaluate the validity, reliability, and responsiveness of PRO-CTCAE in a diverse sample of patients undergoing cancer treatments.<sup>7,8</sup> Use of the PRO-CTCAE system in all 3 studies was



**Figure 1.** Screenshot of Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events free text functionality.

approved by Institutional Review Boards from all participating sites, and all patients provided informed consent.

In each study, participants completed study-specific, PRO-CTCAE surveys. The number and nature of the preselected terms included in each study-specific survey differed (the Manuka Honey trial included 31 PRO-CTCAE terms, the PROSPECT trial 14 terms, and the PRO-CTCAE validation study 48 terms). Thus, we anticipated that in the pooled analysis, the number of supplemental symptomatic AE entries by patients might differ by study, with more entries being offered by participants in the studies that included fewer preselected PRO-CTCAE terms in their study-specific survey. We also posited that the number of supplemental symptomatic AE entries submitted might depend on how well the investigative teams for each study anticipated the symptomatic AEs that would be experienced by participants.

### Organization of the MedDRA dictionary

MedDRA is a clinically validated international medical terminology dictionary and thesaurus, which is widely used by regulatory authorities and the pharmaceutical industry for AE classification in clinical trials.<sup>3</sup> Typically, AE terms recorded by clinicians in medical record documentation are extracted verbatim by data managers, and are mapped to terms in the MedDRA dictionary. To our knowledge, MedDRA has not been used previously for structuring patient-reported free text AEs, particularly in premarket surveillance, although many of its terms are colloquial and therefore amenable to this process. The MedDRA dictionary is organized by a 5-level hierarchy. "System Organ Class" (SOC), which is the highest level of the terminology, represents an anatomical or physiological system, etiology, or purpose.<sup>3</sup> Subordinate to the SOC is the High-Level Group Term, then the High-Level Term, and then the Preferred Term (PT), which represents a single medical concept for a symptom, sign, or disease diagnosis. LLTs are the lowest level of the terminology and are linked to a single PT. Each PT has at least 1 LLT and potentially synonyms,

lexical variants, or quasi-synonyms.<sup>3</sup> In clinical trials, any given verbatim AE from the medical record is coded for data entry at the level of the most specific LLT, and then is mapped up to a PT code for reporting. All terms in the NCI's CTCAE and PRO-CTCAE have been previously mapped to MedDRA LLTs and PTs by the NCI.

### Data abstraction and mapping

All structured dropdown symptomatic AE terms and unstructured free text submissions from web-based surveys were extracted from the PRO-CTCAE software system for each of the 3 studies. Structured dropdown symptomatic AE terms were categorized by MedDRA LLT and mapped to their corresponding MedDRA PT, and when applicable, to a NCI PRO-CTCAE term. Spanish free text submissions were translated into English for analysis by a native Spanish speaker who was credentialed in medical Spanish translation.

Standardized methods endorsed by MedDRA were used for mapping terms within the unstructured free text submissions,<sup>3</sup> which included selection of the "LLT that most accurately reflects the reported information." In general, every term within the free text entry is coded as a separate symptom. In this study, each unstructured free text entry was first independently analyzed for the presence of 1 or more symptomatic AE terms by 2 physician researchers (A.E.C. and K.S.) with adjudication by a third physician researcher (E.M.B.) for any disagreements. For example, a single unstructured free text entry might include a sentence with multiple potential terms (eg, "I am having chest pain and nausea"). Once symptomatic AE term(s) had been extracted from the free text entry (eg, chest pain and nausea from the prior example), these terms were then coded to the closest MedDRA LLT that matched the description of the symptom noted in the free text by searching for the term using the MedDRA web-based browser. The MedDRA web-based browser allows coders to search for symptom terms and their associated PT by entering specific words or combinations of words

**Table 1.** Sample characteristics by study

	Manuka study	PROSPECT study	Validation study	Total across studies
Participants	152	668	940	1760
Age, y				
Median	66	56	59	60
Range	37–85	19–91	19–91	19–91
Sex				
Female	71 (47)	230 (34)	539 (57)	840 (48)
Male	81 (53)	438 (66)	401 (43)	920 (52)
Race				
White	127 (84)	579 (87)	675 (72)	1381 (78)
Black or African American	19 (12)	33 (5)	203 (22)	255 (14)
American Indian or Alaskan Native	2 (<1)	3 (<1)	0 (0)	5 (<1)
Asian	4 (3)	17 (2)	42 (4)	63 (4)
Other or multiple races reported	0 (0)	0 (0)	8 (1)	8 (<1)
Missing	0 (0)	33 (5)	12 (1)	45 (3)
Cancer type				
Lung, head or neck	152 (100)	0 (0)	329 (35)	481 (27)
Breast	0 (0)	0 (0)	260 (28)	260 (15)
Genitourinary or gynecologic	0 (0)	0 (0)	172 (18)	172 (10)
Gastrointestinal <sup>a</sup>	0 (0)	668 (100)	95 (10)	763 (43)
Hematologic	0 (0)	0 (0)	47 (5)	47 (3)
Other or unknown	0 (0)	0 (0)	37 (4)	37 (2)

Note: Values are n (%), unless otherwise reported.

<sup>a</sup>Includes gastroesophageal, hepatic, pancreatic, small bowel, colon, and rectal cancers. Notably, all rectal cancer cases came from the PROSPECT trial, which included 668 patients, constituting 43% of total accruals across trials.

(<https://tools.meddra.org/wbb>). The browser presents possible terms within a hierarchical structure along with synonyms, which allows for mapping to both a LLT and the associated PT. A benefit of using the MedDRA web-based browser is that it allows a coder to enter in verbatim terms that patients used to describe their symptoms for coding.

After identifying the LLT and associated PT for each symptom term, we assessed it for potential mapping to a PRO-CTCAE term (corresponding PT and SOC). If the MedDRA LLT did not map to the MedDRA PT associated with the PRO-CTCAE term, then the MedDRA PT was recorded along with its corresponding SOC. For any terms that were not classified as symptomatic AEs, these were coded as nonsymptomatic AE items (eg, sentence or word fragments that were not interpretable or “laptop freezing”).

### Data analysis

Descriptive statistics were used to characterize the frequency of structured dropdown symptomatic AE terms and unstructured free text entries per survey, per patient, and overall in each of the 3 clinical studies. The proportions of unstructured free text terms mapping to PRO-CTCAE and MedDRA PTs, respectively, were tabulated. The overall number of MedDRA and PRO-CTCAE unstructured symptomatic AE terms was also tabulated. Additionally, the symptomatic AE terms reflected in the supplemental information that were not currently included in the PRO-CTCAE item library or in current study-specific surveys was assessed. The number of unstructured free text entries that were not considered to be symptomatic AEs (eg, sentence fragments) was also tabulated. Also, the number of disagreements between the 2 primary coders and the number of terms that required adjudication by a third coder were recorded.

## RESULTS

Table 1 shows the characteristics of patients enrolled in the 3 included clinical studies. All patients had been diagnosed with a malignancy and were receiving systemic cancer treatment. There were 1760 patients enrolled in these studies with an overall median age of 60 years (range, 19–91 years), 48% were women, and 78% were White. Multiple cancer types were represented. Lung and rectal cancers were overrepresented in the pooled sample, as 2 of the included trials focused on a single disease site, specifically lung cancer (Manuka Honey trial) and rectal cancer (PROSPECT trial).

Pooled across the 3 studies, patients completed a total of 8892 PRO-CTCAE surveys (Table 2). Supplemental entries were provided by 1024 of 1760 (58.2%) patients, with an average of 2.3 unique entries per patient, which averages to 0.45 unique entries per survey completed. This included 1474 of 8892 (16.6%) dropdown structured terms and 913 of 8892 (10.3%) unstructured free text entries (notably, some free text entries included more than 1 individual term, as shown in Table 2).

Mapping of the unstructured free text terms is shown in Figure 2. Among the 1357 free text terms, 87.5% were categorized as symptomatic AEs, of which 384 (32.4%) mapped to an existing PRO-CTCAE term, and all remaining 803 (67.6%) could be mapped to MedDRA LLTs and PTs. The remaining 170 (12.5%) terms that could not be mapped consisted of nonclinical information (eg, “I did not have chemotherapy today”) or sentence fragments that could not be interpreted (eg, “right kidney”). The frequencies of reporting for each individual symptomatic AE term from unstructured free text entries across the 3 studies, and its mapping to PRO-CTCAE or MedDRA (if the term did not map to PRO-CTCAE) are provided in Supplementary Tables S3 and S4. The frequencies of each PRO-CTCAE or MedDRA structured dropdown symptomatic AE terms are shown in Supplementary Table S5.

**Table 2.** Supplemental symptomatic AE information entries by study (unstructured free text entries and structured dropdown terms)

	Manuka study	PROSPECT study	Validation study	Total across studies
Participants	152	666	940	1760
PRO-CTCAE surveys completed	843	5794	2255	8892
Average PRO-CTCAE surveys completed per participant	5.5	8.7	2.4	5.1
Supplemental symptomatic AE information entries: unstructured free text entries and structured dropdown symptomatic AE terms				
Supplemental information entries (unstructured free text + structured dropdown)	61	1407	919	2387
Participants who submitted either unstructured free text or structured dropdown entries	72 (47)	557 (84)	395 (42)	1024 (58)
Average supplemental symptomatic AE information entries per participant	0.8	2.5	2.4	2.3
Supplemental symptomatic AE information entries: unstructured free text entries				
Unstructured free text entries <sup>a</sup>	46 (6)	720 (12)	147 (7)	913 (10)
Participants who entered unstructured free text entries	32 (21)	247 (37)	111 (12)	390 (22)
Average unstructured free text entries submitted per participant	1.4	2.9	1.3	0.5
Total items within unstructured free text entries <sup>a</sup>	69	1134	154	1357
Nonsymptomatic AE items	11	154	5	170
Symptomatic AE items	58	980	149	1187
Mapped to PRO-CTCAE term	20	330	34	384
Unmapped to PRO-CTCAE term (mapped to MedDRA)	38	650	115	803
Supplemental symptomatic AE information entries: structured dropdown symptomatic AE terms				
Total structured dropdown symptomatic AE terms	15	687	772	1474
PRO-CTCAE	12	395	729	1136
MedDRA	3	292	43	338
Participants who selected structured dropdown symptomatic AE terms	13 (9)	310 (47)	284 (30)	634 (36)
Average structured dropdown symptomatic AE terms per patient	1.2	2.2	2.7	2.3

Note: Values are n (%), unless otherwise indicated.

AE: adverse event; MedDRA: Medical Dictionary for Regulatory Activities; PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events.

<sup>a</sup>A single free text entry might include a sentence with multiple symptom terms, and therefore would be mapped individually.

Tables 3 and 4 show the most frequently provided supplemental symptomatic AE entries by patients, including both structured dropdown term selections using MedDRA PTs that are or are not associated with PRO-CTCAE terms (Table 3) and unstructured free text terms that were mapped post hoc to existing MedDRA PTs that are or are not associated with PRO-CTCAE terms (Table 4). Notably, there are MedDRA PTs not associated with PRO-CTCAE terms that appear similar to PRO-CTCAE terms, such as the MedDRA PTs for musculoskeletal pain (vs myalgia), gastroesophageal reflux (vs heartburn), or dysuria (vs painful urination). However, MedDRA separates these concepts into distinct PT codes, and as a PRO-CTCAE term can only map to a single MedDRA PT, they must be considered mutually exclusive.

For example, the most common symptomatic AE term from unstructured free text submissions was “muscle spasms” (reported 40 times), which was also selected multiple times as a structured dropdown term selection (7 times). Muscle spasms could be conceptually related to the PRO-CTCAE term “aching in my muscles (myalgia),” but is likely considered differently by patients. Notably, MedDRA includes distinct PTs for ache or pain and for cramp/spasm.

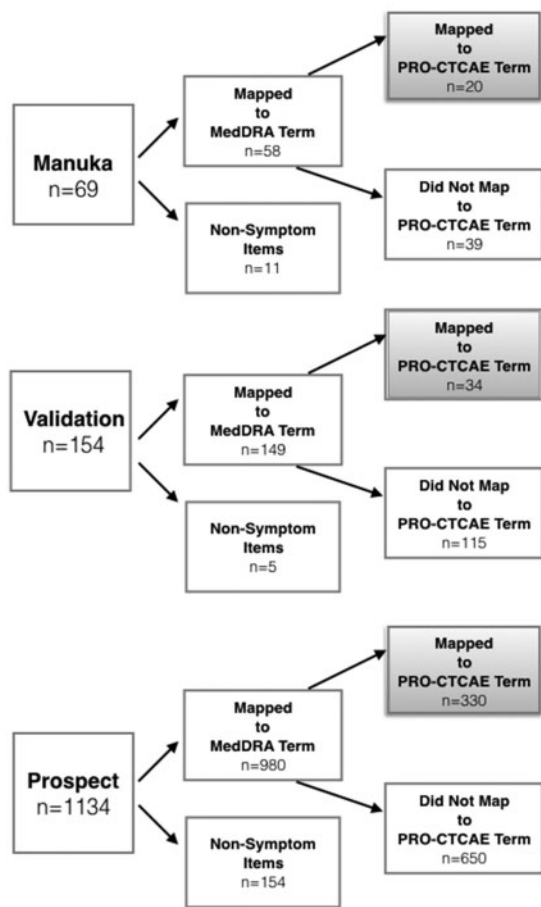
Pain-related terms were also common, and accounted for 27% of symptomatic AEs from unstructured free text entries and 19% of structured dropdown selections. In many cases, these focused on

specific locations of the body (eg, chest or jaw pain), suggesting an opportunity to make the PRO-CTCAE pain assessment more granular, for example with subitems or a graphic depiction of the body.

The MedDRA PTs that did not map to PRO-CTCAE terms could help identify new candidate symptomatic AE terms that may need to be added to the PRO-CTCAE item library to better capture the full range of symptomatic toxicities experienced by patients receiving a diverse range of cancer treatments. For example, rectal bleeding, rectal pain, and painful bowel movement terms were common submissions, which largely were reported in the rectal cancer treatment trial (PROSPECT trial). Of note, the PROSPECT trial did not elicit these symptomatic AEs in the preselected study-specific survey. Frequent bowel movements were also reported by patients separately and would seem distinct from the existing PRO-CTCAE diarrhea item, which asks about “loose or watery stools” and might not be considered by patients to encompass frequent, nonwatery stools. Other MedDRA PTs identified in this study that could be considered for addition to the PRO-CTCAE library in the future (converted when necessary into lay terms comprehensible for patients) include temperature intolerance or feeling cold, upper airway secretions or rhinorrhea (mucus), dry eyes, diplopia, gait disturbance or balance disorder, tremor, and decreased oral intake.

For the analysis of the 3 included studies, a total of ~62 hours were required for coding spread between the 2 coders and





**Figure 2.** Mapping of unstructured free text entries. MedDRA: Medical Dictionary for Regulatory Activities; PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events.

1 adjudicator for 8892 surveys that included a total of 913 free text entries, which is an average of 4.1 minutes per free text entry or ~2 minutes per entry for a single coder. Adjudication was rarely necessary, only for 36 of 913 (<4%) of terms, as there was agreement between the coders for 96% of the terms.

## DISCUSSION

This study describes supplemental symptomatic AE information provided by cancer clinical research participants using a free text functionality offered at the end of electronic PRO-CTCAE surveys and the feasibility for structuring this supplemental AE information. This free text functionality enabled patients either to self-structure responses to dropdown options from MedDRA and PRO-CTCAE terminologies, or to provide an unstructured free text entry. Use of the free text functionality was common—despite preselected static PRO-CTCAE items being provided to patients before offering patients the free text option. Among the 8892 PRO-CTCAE surveys completed by patients, supplemental entries were provided in 2387 (26.8%). More than half of (58.2%) patients provided supplemental symptomatic AE entries during trial participation. These findings suggest that for patients with advanced cancers undergoing treatment in clinical trials, symptomatic AEs are common and diverse,

necessitating more than a static survey if the intention is to capture the full spectrum of the patient experience for AE monitoring.

Notably, rates of providing supplemental entries differed between the 3 included clinical studies. In the Manuka Honey and PRO-CTCAE validation studies, which included more preselected PRO-CTCAE terms in their surveys (31 and 48, respectively), relatively fewer patients added supplemental entries (47% and 42%, respectively) compared with the PROSPECT study. The PROSPECT study included only 14 preselected terms in the static study-specific survey, and 84% of participants provided supplemental entries. This finding highlights the tradeoff between the number of preselected terms in a trial-specific PRO-CTCAE survey, and the number of supplemental entries that can be expected. In a given trial, there are other considerations for the number of selected terms and in general, briefer surveys are considered desirable to minimize patient burden—but this analysis demonstrates that such an approach will miss symptomatic AEs in the absence of a free text functionality in which patients can provide supplemental information.

Supplemental symptomatic AEs provided additional information about the patient experience, and highlighted AEs that could be further explored or considered when selecting PRO-CTCAE items for future therapeutic trials in these clinical subgroups. For example, in the Manuka Honey trial, which included patients with lung cancer receiving chemoradiotherapy, common free text supplemental entries were chest discomfort, throat tightness, and oropharyngeal pain—all of which could be anticipated to be associated with radiation treatment to the chest, although none were included in the study-specific PRO-CTCAE survey. Muscle spasm was also commonly reported. Although muscle spasms have not been previously described in association with chest irradiation to our knowledge and do not have a readily identifiable mechanistic link, this observation generates hypotheses that could be explored prospectively. In the PROSPECT trial, which included patients with rectal cancer receiving chemotherapy, common symptomatic AEs noted were dysphagia, hematochezia, and anorectal discomfort, most which have a clear mechanistic association with the disease or treatment but were not included the study-specific PRO-CTCAE survey. Musculoskeletal issues were also commonly reported (eg, pain in the jaw, muscle spasm, musculoskeletal pain, joint stiffness). These symptomatic AEs were not anticipated based on prior experience with this therapeutic regimen.

Therefore, analyzing free text entries early in a drug development program might provide a signal about symptomatic AEs of interest that could be incorporated in subsequent trials via a study-specific PRO-CTCAE survey. Elicitation of supplemental symptomatic AE information from patients may be particularly valuable in early phase trials where less may be known about the expected AEs associated with a particular regimen. Nevertheless, as observed in this analysis, even in phase III trials in which there is prior knowledge of the profile of AEs associated with a given regimen, capture of the patient perspective through the use of a free text functionality provided important additional symptomatic AE information. In postmarketing surveillance, using a free text functionality with dropdown selections and text box could also enhance spontaneous AE reporting.

Our findings suggest potential improvements toward the future. Specifically, we demonstrate the ability to structure patient-entered unstructured free text entries containing symptomatic AEs through mapping by clinicians to MedDRA and PRO-CTCAE terms. We found that 62% of patients' supplemental items were successfully structured via the software into dropdown menu entries. As the

**Table 3.** Most frequently occurring structured dropdown symptomatic AE terms from PRO-CTCAE or MedDRA (reported  $\geq 5$  times)

MedDRA Preferred Terms associated with PRO-CTCAE TERMS (PRO-CTCAE term)	n	MedDRA Preferred Terms without any associated PRO-CTCAE term	n
Dysgeusia (problems with tasting food or drink)	92	Rectal bleeding	13
Fatigue (fatigue, tiredness, or lack of energy)	74	Fever	10
Headache (headache)	72	Frequent bowel movements	8
Insomnia (insomnia including difficulty falling asleep, staying asleep, or waking up early)	67	Scalp pain	8
Concentration impairment (Problems with concentration)	58	Bowel cramps	7
Generalized pain (pain)	48	Dehydration	7
Peripheral sensory neuropathy (Numbness or tingling in your hands or feet)	47	Muscle spasms	7
Dry mouth (dry mouth)	41	Jaw cramp	5
Depression (feelings that nothing could cheer you up; sad or unhappy feelings)	39	Leg cramps	5
Constipation (constipation)	38	Nasal congestion	5
Anorexia (decreased appetite)	36		
Dyspepsia (heartburn)	35		
Anxiety (anxiety)	33		
Epistaxis (nosebleeds)	31		
Nausea (nausea)	29		
Rash maculo-papular (rash)	28		
Dyspnea (shortness of breath)	25		
Edema limbs (arm or leg swelling)	24		
Dizziness (dizziness)	24		
Alopecia (hair loss)	21		
Flatulence (increased passing of gas [flatulence])	17		
Abdominal pain (pain in abdomen [belly area])	17		
Diarrhea (loose or watery stools [diarrhea])	15		
Mucositis oral (mouth or throat sores)	15		
Myalgia (aching muscles)	13		
Arthralgia (aching joints [elbows, knees, shoulders])	12		
Bloating (bloating of the abdomen [belly])	12		
Tinnitus (ringing in your ears)	12		
Blurred vision (blurry vision)	11		
Hiccups (hiccups)	11		
Cough (cough)	9		
Dry skin (dry skin)	9		
Urinary frequency (frequent urination)	8		
Hoarseness (hoarse voice)	8		
Pruritus (itchy skin)	7		
Memory impairment (problems with memory)	7		
Chills (shivering or shaking chills)	7		
Urinary tract pain (pain or burning with urination)	6		
Palpitations (pounding or racing heartbeat [palpitations])	6		
Dermatitis radiation (skin burns from radiation)	6		
Libido decreased (decreased sexual interest)	5		
Urticaria (hives [itchy red bumps on the skin])	5		

AE: adverse event; MedDRA: Medical Dictionary for Regulatory Activities; PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events.

majority of the free text entries could be manually coded to MedDRA or PRO-CTCAE terms (as demonstrated in this study), future PRO software systems would also benefit from functionalities that better assist patients in finding a structured term that corresponds to the symptomatic AE they wish to report.

A challenge with the current free text functionality is that the unstructured free text entries require manual review and mapping to PRO-CTCAE or MedDRA. This limits the opportunity for these data to be submitted in a timely manner to AE reporting systems. Some degree of automation might both accelerate the availability of this information to near real-time, and reduce the personnel effort for coding terms. Natural language processing tools could allow for mapping of unstructured free text into structured MedDRA PTs.

Prior research has applied various natural language processing and text mining tools to extract AE data from clinical narratives from electronic health records, social media, web search logs, and spontaneous AE reports.<sup>9–15</sup> Of note, the majority of the free text submissions examined in our study (Manuka Honey and PROSPECT trials) were collected during the premarket phase of drug development in which elicitation of AEs directly from patients is not standard practice. It is plausible that the colloquial nature of how patients describe their adverse symptoms in our study would be similar to text submitted in spontaneous AE reports submitted by patients during postmarketing surveillance. Therefore, natural language processing and text mining methods in the context of spontaneous AE reports by patients could be potentially applied to the free

**Table 4.** Most frequently occurring terms from unstructured free text entries mapped post hoc to PRO-CTCAE or MedDRA (reported  $\geq 5$  times)

MedDRA Preferred Terms associated with PRO-CTCAE term (PRO-CTCAE term)	n	MedDRA Preferred Terms without any associated PRO-CTCAE term	n
Peripheral sensory neuropathy (numbness or tingling in your hands or feet)	34	Muscle spasms	40
Dysgeusia (problems with tasting food or drink)	20	Temperature intolerance	33
Epistaxis (nosebleeds)	20	Hematochezia	23
Diarrhea (loose or watery stools [diarrhea])	18	Pain in jaw	22
Fatigue (fatigue, tiredness, or lack of energy)	15	Pain in extremity	18
Insomnia (insomnia including difficulty falling asleep, staying asleep, or waking up early)	14	Hemorrhoids	15
Palmar-plantar erythrodysesthesia syndrome (hand-foot syndrome [rash of the hands or feet that can cause cracking, peeling, redness, or pain])	14	Muscular weakness	13
Generalized pain (pain)	13	Proctalgia	13
Arthralgia (aching joints [eg, elbows, knees, shoulders])	13	Hypoesthesia	12
Constipation (constipation)	12	Frequent bowel movements	11
Nausea (nausea)	12	Painful defecation	11
Dyspepsia (heartburn)	11	Increased upper airway secretion	11
Headache (headache)	10	Joint stiffness	10
Abdominal pain (pain in abdomen [belly area])	9	Chest discomfort	9
Flatulence (increased passing of gas [flatulence])	9	Nasopharyngitis	9
Alopecia (hair loss)	9	Weight decreased	9
Edema limbs (arm or leg swelling)	8	Musculoskeletal pain	9
Dermatitis radiation (skin burns from radiation)	8	Balance disorder	9
Urinary tract pain (pain or burning with urination)	8	Paresthesia	9
Hiccups (hiccups)	8	Rhinorrhea	9
Watering eyes (watery eyes [tearing])	7	Abdominal pain upper	8
Dysphagia (difficulty swallowing)	6	Gastroesophageal reflux disease	8
Fecal incontinence (loss of control of bowel movements)	6	Feeling cold	8
Vomiting (vomiting)	6	Rectal hemorrhage	8
Anorexia (loss of appetite)	6	Back pain	8
Urinary frequency (frequent urination)	6	Rash	8
Dry mouth (dry mouth)	5	Paresthesia oral	7
Chills (shivering or shaking chills)	5	Tremor	7
Myalgia (aching muscles)	5	Dysuria	7
Pruritus (itchy skin)	5	Nasal dryness	7
Skin hyperpigmentation (unusual darkening of the skin)	5	Erythema	7
		Blepharospasm	6
		Dry eye	6
		Asthenia	6
		Pyrexia	6
		Gait disturbance	6
		Neuralgia	6
		Oropharyngeal pain	6
		Skin erythema	6
		Chest pain	5
		Diplopia	5
		Feces discolored	5
		Impaired healing	5
		Parosmia	5

*Note:* Symptomatic AE terms within each unstructured free text entry were first mapped to the closest MedDRA Lower-Level Term and then up to its associated Preferred Term. If that Preferred Term corresponded to a PRO-CTCAE term, that was preferentially selected for the mapping to PRO-CTCAE (column 1). If no PRO-CTCAE term was found, then the MedDRA Preferred Term was selected (column 2).

AE: adverse event; MedDRA: Medical Dictionary for Regulatory Activities; PRO-CTCAE: Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events.

text examined in our study. Additionally, our study findings provide further support for the application of existing standardized terminologies such as MedDRA and PRO-CTCAE, which have not been previously applied to unstructured free text elicited directly from patients during surveillance with trial-specific symptomatic AE surveys.

Study findings also identified new candidate symptomatic AE terms that should be considered for future inclusion in the

PRO-CTCAE item library. When the PRO-CTCAE library was developed, terms were derived from an existing library of clinician-reported AE terms in cancer, the CTCAE, with a focus on chemotherapy clinical trials.<sup>2</sup> Moving forward, a process for identifying new candidate terms, including evaluation of supplemental symptomatic AE data in clinical trials, may improve the representativeness of the PRO-CTCAE library for application to broader populations. Moreover, supplemental information from trials in



specific populations could assist in developing standardized symptomatic AE sets that are context specific. Future evaluations can also explore whether severity grading could be added to the free text functionality for MedDRA dropdown entries to move beyond simply identifying the presence of a particular symptomatic AE.

Within the unstructured free text entries, we found that patients sometimes chose to report a symptomatic AE about which they had already been asked in the study-specific PRO-CTCAE survey that they completed (70 terms in trial Manuka Honey trial, 113 in the PROSPECT trial, and 164 in the PRO-CTCAE validation trial). This suggests that patients may sometimes forget that they were asked a particular question in a survey, or may wish to report the same symptomatic AE more than once or in a different way, such as in their own terms by entering an unstructured free text entry.

We found that 12.5% of free text items were nonsymptomatic AEs, including logistical information about treatment or technical difficulties and sentence fragments. Future iterations of the software could better direct patients to other means of communicating non-symptom information such as patient portals. This finding emphasizes the value patients place on communicating the information that they regard as relevant, and underlines the importance of providing patients with opportunities to communicate such information in addition to symptomatic AEs in clinical trials.

There are several limitations of this study. 2 of the 3 analyzed clinical trials were disease-specific (lung and rectal cancer), which might have skewed reporting to disease symptoms and symptomatic AEs common in these populations. Nonetheless, diverse symptomatic AEs were reported across each study. Moreover, one of the included trials was a large study encompassing participants with diverse cancer types (PRO-CTCAE validation trial). Additionally, there could be possible bias in the coding processes utilized to extract symptom terms or to map the terms to MedDRA LLTs. To reduce the potential for coding biases and to gauge the extent of agreement between coders, 2 physicians independently coded each symptomatic AE term within unstructured free text entries with adjudication by a third physician coder. Disagreements between the 2 coders were uncommon and generally minor, and reconciliation by a third rater was rarely necessary. This observation suggests that a single clinician coder may be adequate to reproducibly map unstructured free text entries to existing terminologies. In network or multi-site trials, this function could be centralized to achieve comparable reliability and ensure that coders have adequate familiarity with MedDRA and PRO-CTCAE terminologies. Finally, there is a potential risk with structured dropdown suggestions of “leading” patients to select a particular AE term, and this risk must be weighed vs the convenience, efficiency, and opportunity for standardization using the dropdown functionality.

Toward the future, a challenge for clinical trials that include this type of free text functionality will be to determine how and when collected supplemental symptomatic AE information should be interpreted, who should see it, and within what time frame. In the usual AE reporting workflows, clinicians generally are grading the symptoms and severity, so reporting of these AEs back to the clinicians is not necessary. However, for patient-reported symptomatic AEs, there may be clinical value for this information to be evaluated in near real-time and be conveyed to clinicians along with the solicited PRO-CTCAE items.<sup>16</sup> In international trials being conducted at numerous locations across languages, this may be particularly complex and challenging to implement. This question remains to be determined and therefore should still be considered optional, but evidence from this study can help inform this discussion.

## CONCLUSION

In conclusion, this study demonstrates the feasibility and value of patient-reported supplemental information about symptomatic AEs using a free text box functionality that includes mapping to existing AE terminologies. Our findings also elucidate reproducible methods for coding of unstructured free text entries for AE reporting, and suggest directions for future research to strengthen this approach to patient self-reporting in clinical trials.

## FUNDING

Funding for this project was provided through contracts HHSN261201000043C and HHSN261201000063C, which were awarded by the National Cancer Institute. A.E.C. also received support from the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant 1KL2TR001109 during part of this research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## CONTRIBUTIONS

A.E.C. and E.M.B. conceived of the study concept for free text analysis. A.E.C. prepared the initial draft of the manuscript, and developed the study design and methods for free text analysis. Overall study design contributions were from D.S., D.W.B., A.P.A., A.C.D., L.M.M., A.M.O., and P.B. A.E.C., K.S., E.M.B., A.C.G., A.C.D., and S.A.M. contributed to data analysis. Extraction and/or collection of data was contributed to by A.C.D., D.S., D.W.B., L.J.R., P.B., and E.M.B. All authors provided input into the final manuscript, and critically revised and approved the final version.

## SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

*Conflict of interest.* None declared.

## REFERENCES

1. Basch EM, Reeve BB, Mitchell SA, *et al.* Electronic toxicity monitoring and patient-reported outcomes. *Cancer J* 2011; 17 (4): 231–4.
2. Basch EM, Reeve BB, Mitchell SA, *et al.* Development of the National Cancer Institute’s patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *J Natl Cancer Inst* 2014; 106 (9): 244–231.
3. Understanding MedDRA. *The Medical Dictionary of Regulatory Activities*. Available at: <https://www.meddra.org/sites/default/files/page/documents/meddra2013.pdf>. Accessed January 4, 2018.
4. Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE™). Available at: <https://healthcaredelivery.cancer.gov/pro-ctcae/>. Accessed January 4, 2018.
5. Kluetz PG, Papadopoulos EJ, Johnson LL, *et al.* Focusing on core patient-reported outcomes in cancer clinical trials-response. *Clin Cancer Res* 2016; 22 (22): 5618.
6. Basch E, Dueck AC. Patient-reported outcome measurement in drug discovery: a tool to improve accuracy and completeness of efficacy and safety data. *Exp Opin Drug Discov* 2016; 11 (8): 753–8.
7. Basch EM, Pugh SL, Dueck AC, *et al.* Feasibility of patient reporting of symptomatic adverse events via the patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE) in a chemoradiotherapy cooperative group multicenter clinical trial. *Int J Radiat Oncol Biol Phys* 2017; 98 (2): 409–18.

8. Dueck AC, Mendoza TR, Mitchell SA, *et al.* Validity and reliability of the US national cancer institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *JAMA Oncol* 2015; 1 (8): 1051–9.
9. Harpaz R, Callahan A, Tamang S, *et al.* Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug Saf* 2014; 37 (10): 777–90.
10. Combi C, Zorzi M, Pozzani G, Pozzani G, Arzenton E, Moretti U. Normalizing spontaneous reports into MedDRA: some experiments with MagiCoder. *IEEE J Biomed Health Inform* 2018 [E-pub ahead of print Jul 30]. doi: 10.1109/JBHI.2018.2861213.
11. Ventola CL. Big data and pharmacovigilance: data mining for adverse drug events and interactions. *PT* 2018; 43 (6): 340–51.
12. Wang X, Hripcsak G, Markatou M, Friedman C. Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study. *J Am Med Inform Assoc* 2009; 16 (3): 328–37.
13. Harpaz R, Vilar S, DuMouchel W, *et al.* Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *J Am Med Inform Assoc* 2013; 20 (3): 413–9.
14. Cocos A, Fiks AG, Masino AJ. Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. *J Am Med Inform Assoc* 2017; 24 (4): 813–21.
15. Rolfes L, van Hunsel F, van der Linden L, Taxis K, van Puijenbroek E. The quality of clinical information in adverse drug reaction reports by patients and healthcare professionals: a retrospective comparative analysis. *Drug Saf* 2017; 40 (7): 607–14.
16. Kim J, Singh H, Ayalew K, *et al.* Use of PRO measures to inform tolerability in oncology trials: Implications for clinical review, IND safety reporting and clinical site inspections. *Clin Cancer Res* 2017; 24 (8): 1–5.