ORIGINAL ARTICLE

# Variation in Toxicity Reporting Methods for Early Phase Lung Cancer Treatment Trials at Oncology Conferences

Emily A. Simons, MD, MPH,[a,*] Derek E. Smith, MS,[b,c] Dexiang Gao, PhD,[b,c] D. Ross Camidge, MD, PhD[a]

[a]Division of Medical Oncology, Department of Medicine, University of Colorado School of Medicine, Aurora, Colorado
[b]Department of Pediatrics, University of Colorado School of Medicine, Aurora, Colorado
[c]Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, Colorado

Received 19 February 2020; revised 18 April 2020; accepted 20 April 2020
Available online - 27 April 2020

**ABSTRACT**

**Introduction:** Phase I and II trials provide the initial human safety and tolerability data for new drugs. Nevertheless, the methods for presenting toxicity data are not standardized. Clinicians often first encounter these data at professional conferences. We sought to characterize how the burden of adverse events (AEs) is reported at the largest professional conference in clinical oncology.

**Methods:** We collected toxicity data from all lung cancer-associated phase I and II trial presentations and posters at the American Society for Clinical Oncology annual meetings from 2017 to 2019. We captured the various AE features, including the minimum incidence used for reporting; whether AEs found were treatment emergent or treatment related, grouped by organ system or separated by individual descriptors; whether combined or separated across dose levels when a dose-escalation component was included; and whether dose-limiting toxicities, serious AE, dose-reduction rules, and denominators for laboratory tests were described.

**Results:** A total of 209 trials were analyzed. There was wide variability in toxicity reporting practices. Six different thresholds for reporting AEs of any grade were used. Treatment-related AEs were reported twice as frequently as treatment-emergent AEs. Toxicities were as likely to be reported across dose levels as by dose level. Terms such as dose-limiting toxicity and serious AE were rarely defined. Dose-reduction rules and denominators for laboratory tests were never defined.

**Conclusions:** Standardization of methods for reporting toxicities could improve the quality and ease of comparability of data on adverse effects in early phase therapeutic trials. A minimal AE data disclosure template is proposed.



*Keywords:* Lung cancer; Toxicity; Adverse effects; Conference

## Introduction

The clinical influence of early (phases I–II) trial data in oncology has dramatically increased with the advent of accelerated drug approval and the breakthrough drug classification within the Food and Drug Administration Safety and Innovation Act in 2012, which allowed drug approval based on the results of compelling, if limited, initial clinical data.[1] National guideline organizations, such as the National Comprehensive Cancer Network, have also produced recommendations on the basis of early phase trial data.[2,3] Professional conferences are a major avenue for dissemination of early phase (phases I–II) trial data and a key source of continuing medical education for oncologists who increasingly rely on early phase trial data to understand not just the efficacy but

*Corresponding author.

*Disclosure: Dr. Camidge reports honoraria from Roche/Genentech and Takeda clinical trials (institutional) sponsored by Pfizer, Roche, Genentech, and Takeda. The remaining authors declare no conflict of interest.*

Address for correspondence: Emily A. Simons, MD, MPH, Division of Medical Oncology, Department of Medicine, University of Colorado School of Medicine, MS 8117, 12801 E 17th Avenue, Aurora, CO 80045. E-mail: emily.simons@cuanschutz.edu



ISSN: 1556-0864
https://doi.org/10.1016/j.jtho.2020.04.020

also the safety and tolerability (toxicity) of novel therapies. Recently, there has been a call to improve communication of toxicities for novel therapies, particularly in oncology.[4]

Despite this, there is no systematic description of current practices for reporting toxicities at conferences, and there may be large variations in how safety and tolerability are presented in early phase trials. Listings of treatment-emergent adverse events (AEs), which do not include a filter of causality introduced by the investigators, would be expected to be associated with higher rates of AE reporting. In contrast, listings of treatment-related AEs do include such a filter and would be expected to be associated with lower rates of AE reporting from the same trial. To facilitate practical displays of AEs, frequency filters may also be used, such that only AEs occurring in a certain number or percentage of patients may be found. The use or lack of use of combined AE terms (e.g., combining impairment in memory and concentration as "cognitive effects") in the presence of frequency filters could cause toxicities to fall above or below the presentation threshold and facilitate over- or under-representation of toxicities, respectively. Finally, when multiple different doses of an agent or agents are being explored in the same trial, finding aggregated toxicity across all dose levels, rather than only at the subsequently designated recommended phase II dose and schedule (RP2D), could be considered appropriate as a greater total number of patients would have contributed to the data set. Alternatively, such an approach may be criticized because it could overestimate toxicity through contributions from cohorts dosed above the RP2D or underestimate it from the dilutional effect of data from cohorts dosed below the RP2D.

The American Society for Clinical Oncology (ASCO) annual meeting is the largest conference in oncology, with approximately 5500 US-based oncologists attending the 2019 ASCO annual meeting, which is equivalent to 40% of the 13,216 medical hematologists and oncologists registered with Medicare, together with a large number of international oncologists.[5,6] We characterized parameters used to display toxicity in lung cancer-associated trials at the ASCO annual meetings from 2017 to 2019.

## Materials and Methods

We reviewed all abstracts of phase I and II medical trials for the years 2017 to 2019 available at the ASCO virtual meeting website under the sections "Lung cancer" and "Developmental therapeutics." Abstracts that had posters or slides available to assess detailed toxicity information were included. We downloaded all materials for presentations on phase I to II trials that presented toxicity data and included at least one patient with lung cancer. We excluded presentations that did not provide toxicity data or investigated radiation alone or in combination owing to the different toxicity profile of radiation compared with systemic agents (Fig. 1).

On the basis of our experience in implementing early phase clinical trials, we developed a list of parameters that could potentially affect the interpretation of toxicity results. This list was iteratively revised and expanded after an exploratory analysis of 1 year of ASCO presentations, resulting in the reporting elements listed in Table 1. We categorized trials labeled as phase I/II as phase I if the trial included a dose-escalation component and as phase II if the trial did not include dose escalation. Owing to the increased sensitivity of a small, absolute number of patients compared with an incidence threshold for reporting toxicities and the qualitative difference in the number of toxicities reported by trials that used a threshold of one to two patients compared with incidence, we categorized reporting thresholds as one to two patients and then the remainder of the thresholds for reporting as 5% to 10%, 15% to 20%, and greater than or equal to 25%. We considered a term to represent consolidated conditions if an organ system term (e.g., gastrointestinal disorders) was used, if the author specifically annotated that a term combined different AEs, or if the author referenced use of standard terminology (see subsequent text) but reported more generalized terms (e.g., elevated liver function in place of elevated aspartate transaminase or alanine transaminase). We considered a terminology as standardized if the terms were consistent with the Common Terminology Criteria for Adverse Events (CTCAE) or the preferred terms of the Medical Dictionary for Regulatory Activities (MedDRA).

We used unordered multinomial logistic regressions to evaluate whether the minimum incidence for reporting toxicities varied by phase (dichotomized as phase I and phase II) or trial size (dichotomized as $<40$ patients and $\geq 40$ patients), using phase I and $<40$ patients as the referent group. Phase and trial sizes were evaluated separately. For all other elements listed in Table 1, we used logistic regression to evaluate whether these characteristics varied by phase (using phase I as the referent group) or trial size (using $<40$ patients as referent group) and evaluated the magnitude of the effect. The $p$ values reported are not corrected for multiple hypotheses testing. The Bonferroni $p$ value, corrected for multiple hypotheses testing of the 16 elements in Table 1, is 0.003; thus, only regression results with $p$ less than 0.003 would have met rigorous standards for detecting differences between groups, which was not considered practical given the data set available for analysis.

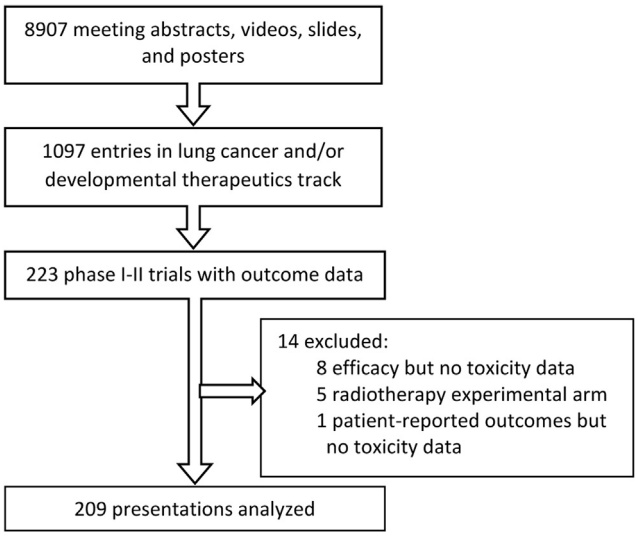


**Figure 1.** CONSORT diagram.

## Results

The ASCO 2017 to 2019 Annual Meetings provided 8907 meeting slides and posters. Of these, 1370 presentations were on lung cancer or developmental therapeutics. Phase I to II trials with outcome data comprised 223 entries, of which 209 disclosed toxicity data and were included in our analysis (Fig. 1). Of these 209 presentations, 120 (57%) were phase I and 89 (43%) were phase II trials (Table 2). Trials that broadly enrolled patients on the basis of having a solid tumor, which included at least one patient with lung cancer, were the subject of 68 of the 209 presentations (33%) whereas the remaining 141 of the 209 presentations (67%) represented trials that exclusively enrolled patients with lung cancer. The median trial size was 38 patients; phase I trials enrolled a median of 34 patients whereas phase II trials enrolled a median of 49 patients (the mean number of patients was 53 and 86 patients, respectively). Although trial phase and size were associated, they were not colinear and occasionally had unique associations with other trial characteristics as evaluated in the subsequent text.

### *Frequency of AEs*

There was wide variability in the thresholds used for reporting AEs, with 12 different thresholds used to report AEs of any grade; thresholds ranged from one patient to greater than or equal to 40% incidence. These were simplified to the five categories in Table 3. Utilizing unordered multinomial logistic regression and the five categories of reporting thresholds in Table 3, our study revealed that a larger trial size ($\geq$40 patients) was associated with a 1.6-fold to twofold increase in the odds of using a reporting threshold of 5% to 10% or 15% to 20% incidence compared with one to two patients for AE of any grade (OR 1.6 and 2.0 respectively, overall $p < 0.001$). Phase II trials tended not to use higher incidence (i.e., less sensitive) thresholds for reporting as compared with phase I trials (OR $-0.3$ and 0.6, respectively, for odds of using a reporting threshold of 5%–10% or 15%–20% incidence compared with 1–2 patients, overall $p = 0.03$ using unordered multinomial regression).

**Table 1.** Elements of Reporting Toxicities and Outcome Assignments

| Reporting Element | Outcome Assignments |
|---|---|
| What minimum incidence was used for reporting toxicities of any grade? | 1 patient, 2-3 patients, 5% incidence, 10% incidence, 15%-20% incidence, $\geq$25% incidence, not reported |
| What minimum incidence was used for reporting toxicities grade $\geq$3? | 1 patient, 2-3 patients, 5% incidence, 10% incidence, 15%-20% incidence, $\geq$25% incidence, not reported |
| Were treatment-related toxicities reported? | Y/N/Not reported |
| Were treatment-emergent toxicities reported? | Y/N/Not reported |
| Was a dose-escalation component included? | Y/N |
| Were AEs combined across dose levels if dose escalation was included? | Y/N |
| Were DLTs reported if the trial included a dose-escalation component? | Y/N |
| Were key DLT criteria described? | Y/N |
| Were serious AEs reported? | Y/N |
| Were key serious AE criteria described? | Y/N |
| Were AEs leading to treatment change reported? | Y/N |
| Was standardized terminology (CTCAE or MedDRA) used? | Y/N |
| Were grouped AE terms used? | Y/N |
| Was overall incidence of AEs at organ system level reported? | Y/N |
| Was the total number of patients with any AE any grade reported? | Y/N |
| Was the total number of patients with any AE grade $\geq$3 reported? | Y/N |

AE, adverse event; CTCAE, Common Terminology Criteria for Adverse Events; DLTs, dose-limiting toxicities; MedDRA, Medical Dictionary for Regulatory Activities; N, no; Y, yes.

**Table 2.** Trial Characteristics (N = 209)

| Characteristics | Number of Trials (%) | Median Number of Patients (Mean) |
|---|---|---|
| All trials | 209 (100) | 38 (range 5-550); mean 67 |
| <40 patients | 109 (52) | 24 (25) |
| ≥40 patients | 100 (48) | 80 (114) |
| Cancer Type | | |
| Lung | 141 (67) | 39 (74) |
| Multiple cancer types | 68 (33) | 37 (53) |
| Phase | | |
| I | 120 (57) | 34 (53) |
| II | 89 (43) | 49 (86) |

In the 141 of the 209 trials that provided the thresholds for reporting both AEs of any grade and grade greater than or equal to 3, the following two different methods were used for displaying toxicity counts of different grades: (1) 100 of the 141 trials reported AE incidence grades greater than or equal to 3 only where the overall incidence across grades reached a given threshold (see column 2 in Table 4), and (2) 41 of the 141 trials used a lower incidence threshold for reporting AE grades greater than or equal to 3 than AEs of any grade (see columns 3 and 4 in Table 4). As an example of the former method, a trial reported AEs of any grade that reached an incidence of 10% and reported how many of those events were grade greater than or equal to 3. Consequently, this excluded AEs with grade greater than or equal to 3 that did not meet a 10% incidence overall. In the latter method, a trial reported AEs of any grade that reached 10% incidence but also reported all AEs with grade greater than or equal to 3 (i.e., a threshold of one patient).

## Attribution of AEs

A total of 182 of the 209 presentations (87%) indicated whether AEs were treatment emergent or treatment related (Table 5). Treatment-related AEs were reported in 139 of the 209 (67%) presentations. We found no evidence that the odds of reporting treatment-related AEs were affected by phase or trial size ($p = 0.08$ and 0.49, respectively). Overall treatment-emergent AEs were included in 70 of the 209 presentations (33%). Phase II trials were associated with lower odds of reporting treatment-emergent AEs than phase I trials (OR 0.3, 95% confidence interval 0.2–0.6, $p < 0.01$.) A total of 39 of the 209 (19%) reported both treatment-emergent and treatment-related AEs; 27 of the 209 (13%) did not specify whether AEs were treatment related or treatment emergent.

## Grouping of AEs

The overall incidence of any AE of any grade was reported in 92 of the 209 studies (44%), and the overall incidence of AEs with grade greater than or equal to 3 was reported in 92 of the 209 studies (44%) (Table 5). A total of 48 of the 209 (23%) included a term that represented a consolidated group of disease entities, such as infection, instead of component terms of urinary tract infection, lung infection, etc. Of these 48 studies, seven presented AEs only by organ system (e.g., endocrine, skin, or gastrointestinal) without identifying the component terms (Table 5).

## Dose Level Grouping, Definitions of Actionable AEs, and Relevant AE Denominators

Standardized terminology, consistent with the Common Terminology Criteria for Adverse Event or the preferred terms of the Medical Dictionary for Regulatory

**Table 3.** Threshold for Reporting AE of Any Grade (N = 209)

| Number of Patients With AE | Number of Trials With AE of Any Grade by Threshold (%) | Association With Trial Size and Phase (OR, 95% CI) | |
|---|---|---|---|
| | | Phase II vs. I | Trial Size ≥40 vs. <40 |
| 1-2 patients | 46 (22) | – | – |
| 5%-10% incidence | 66 (32) | −0.3 (−1.0 to 0.5) | 1.6 (0.7-2.4) |
| 15%-20% incidence | 30 (14) | 0.6 (−0.3 to 1.6) | 2.0 (0.9-3.0) |
| ≥25% incidence | 5 (2) | −15 (−1713 to 1684) | 0.9 (−1.0 to 2.8) |
| Not reported | 62 (30) | 0.3 (−0.5 to 1.0) | 1.2 (0.4-2.1) |
| Total | 209 | | |

AE, adverse event; CI, confidence interval; OR, odds ratio.

**Table 4.** Concordance of Reporting Thresholds Between AEs of Any Grade and Grade Greater Than or Equal to 3 (N = 141[a])

| Number of Patients With AE | Same Threshold AE of Any Grade and Grade ≥3 (N = 100), N (%) | Discordant Thresholds Between AEs of Any Grade and Grade ≥ 3 (N = 41)[b] | |
|---|---|---|---|
| | | Any Grade, N (%) | Grade ≥3, N (%) |
| 1-2 patients | 39 (39) | 5 (12)[c] | 32 (78) |
| 5%-10% incidence | 44 (44) | 19 (46) | 9 (21) |
| 15%-20% incidence | 15 (15) | 14 (34) | – |
| ≥25% incidence | 2 (2) | 3 (7) | – |
| Total | 100 | 41 | 41 |

[a]A total of 141 of 209 (67%) displayed reporting thresholds for both AEs of any grade and grade greater than or equal to 3. A total of 68 trials did not disclose the threshold used for reporting either or both AEs of any grade and grade greater than or equal to 3.
[b]A total of 41 studies had different thresholds for AEs of any grade versus grade greater than or equal to 3 which are represented in both columns.
[c]These five studies used a threshold of two patients to report AEs of any grade and one patient to report AEs of grade greater than or equal to 3.
AE, adverse event.

Activities, was used in 180 of the 209 presentations (86%) (Table 5). All but four studies reported toxicity grade. Of 99 trials that included a dose-escalation component, 46 presented AEs only as a combined data set across all dose levels, rather than separated out by individual dose levels. Dose-limiting toxicities (DLTs) were reported in 67 of the 99 trials (68%) with a dose-escalation component, but only 21 of these 67 defined the criteria for DLT. Serious AEs were reported in 68 of the 209 trials (33%), none of which defined the trial-specific criteria for serious AE. AEs that resulted in dose decrease, delay, or discontinuation were reported in 89 of the 209 trials (43%) and were 2.5 times more likely to be reported in trials with greater than or equal to 40 patients (95% confidence interval 1.4–4.4, $p < 0.01$). Although the individual toxicities leading to decisions on dose interruptions were often noted, the trial-specific criteria for such decisions (e.g., grades ≥3 and at 1 or >1 occurrence in a patient) were never provided in the available presentations. Although not part of our

**Table 5.** Reporting Element Results and Association With Trial Phase and Size

| Reporting Element | Yes/Total Trial Number (%) | OR (95% CI)[a] | |
|---|---|---|---|
| | | Phase II vs. I | Trial Size ≥40 vs. <40 |
| Were treatment-related toxicities reported?[b] | 139/209 (67) | 0.5 (0.3-1.1); $p = 0.08$ | 0.8 (0.4-1.6); $p = 0.49$ |
| Were treatment-emergent toxicities reported?[b] | 70/209 (33) | **0.3 (0.2-0.6); $p < 0.01$** | 1.0 (0.5-1.8); $p = 0.94$ |
| Was a dose-escalation component included? | 99/209 (47) | NA[c] | **0.4 (0.2-0.7); $p < 0.01$** |
| Were AEs combined across dose levels if dose escalation was included? (n = 99) | 46/99 (46) | NA[c] | 1.9 (0.8-4.5); $p = 0.12$ |
| Were dose-limiting toxicities reported if the trial included a dose-escalation component? (n = 99) | 67/99 (68) | NA[c] | 0.7 (0.3-1.7); $p = 0.45$ |
| Were key dose-limiting toxicity criteria described? | 21/67 (31) | NA[c] | 1 (0.3-3.1); $p = 0.95$ |
| Were serious AEs reported? | 68/209 (33) | **0.4 (0.2-0.7); $p < 0.01$** | 0.9 (0.5-1.6); $p = 0.65$ |
| Were key serious AE criteria described? | 0/68 (0) | – | – |
| Were AEs leading to treatment change reported? | 89/209 (43) | 0.9 (0.5-1.5); $p = 0.59$ | **2.5 (1.4-4.4); $p < 0.01$** |
| Was standardized terminology (CTCAE or MedDRA) used? | 180/209 (86) | 1.1 (0.5-2.3); $p = 0.89$ | 1 (0.4-2.1); $p = 0.96$ |
| Were grouped AE terms used? | 48/209 (23) | 1.6 (0.9-3.2); p = 0.13 | 1.9 (1.0-3.7); $p = 0.05$ |
| Was overall incidence of AEs at organ system level reported? | 7/209 (3) | 1.8 (0.4-8.4); $p = 0.43$ | 2.8 (0.5-14.9); $p = 0.20$ |
| Was the total number of patients with any AE any grade reported? | 92/209 (44) | 0.8 (0.5-1.5); $p = 0.54$ | 1.7 (1-3); $p = 0.05$ |
| Was the total number of patients with any AE grade ≥3 reported? | 92/209 (44) | 1 (0.6-1.7); $p = 0.96$ | **2.6 (1.5-4.5); $p < 0.01$** |

[a]Using unadjusted logistic regression. Bold indicates $p$ values less than 0.05.
[b]A total of 27 studies did not specify if AE were treatment related or treatment emergent; these 27 studies were excluded from logistic regression.
[c]Dose escalation by definition includes only phase I trials.
AE, adverse event; CI, confidence interval; CTCAE, Common Terminology Criteria for Adverse Events; MedDRA, Medical Dictionary for Regulatory Activities; NA, not applicable; OR, odds ratio.

initial data element capture (Table 1), we noted that the denominator was not included for laboratory abnormalities and we could not distinguish whether all patients were tested for these laboratory parameters or only a subset based on other criteria, such as clinical suspicion.

## Discussion

Early phase trial data at clinical conferences may represent the first time the efficacy and toxicity of a new drug or drug combination are revealed to healthcare providers, and the same data set may well be the dominant one available for education when highly active drugs are rapidly licensed. In addition, a substantial amount of time may elapse from abstract presentation to manuscript publication delaying access to more complete data sets.[7] Beginning with CONSORT recommendations in 1996, consistent efficacy reporting standards to permit cross-trial and cross-drug comparisons have been developed and periodically refined, such as the Response Evaluation Criteria in Solid Tumors guidelines and the Response Assessment in Neuro-Oncology for central nervous system (CNS) efficacy end points.[8–10] Nevertheless, standardization of toxicity reporting has received less attention. Guidelines for reporting harms were published 8 years after the first CONSORT recommendations, but these have been poorly adhered to in oncology trial publications.[11,12] No guidelines exist for presentation of toxicity data at conferences. ASCO provides recommendations for abstract submission, but of the 26 recommendations for reporting data from phase I to III trials, the only recommendation related to safety and tolerability is that phase I trial abstracts should report DLTs, a criterion that was not met in nearly a third of ASCO presentations in our data set.[13] Variation in toxicity reporting practices could dramatically alter the initial perceptions of the safety and tolerability of new drugs and drug combinations.

Overall, we found highly variable practices for reporting toxicities in this analysis of presentations at the largest annual conference in clinical oncology. Trials reported treatment-related AEs twice as frequently as treatment-emergent AEs (67% versus 33%, respectively). Conceivably, use of treatment-related approaches might be a sensible way of improving the signal-to-noise ratio in early phase trials, particularly when an experimental approach is being given as part of a combination with a standard therapy with known associated toxicities of its own. Alternatively, novel interactions with the standard therapy, which increase the standard therapy's toxicity, or AEs that were not expected to be related to a novel agent by its known mechanism of action could be underreported as a result of this approach. Arguably, treatment-emergent AEs are less subject to bias, whereas treatment-related AEs may be particularly valuable in differentiating the effects of the novel agent over and above those associated with other agents with known toxicities in a combination. Nevertheless, in most cases, we were unable to determine exactly how "treatment-related AEs" were defined within the trials of combination therapies studied among the ASCO presentations.

The reporting of toxicity combined across dose levels, rather than broken down by dose level, occurred in nearly half of trials with a dose-escalation component (46%). Although in theory this could either overestimate or underestimate toxicities at the recommended dose because of contamination by patients treated above or below the recommended dose, in general, it is probably more likely to underrepresent AEs, as more patients are likely to have been treated at dose levels below the nontolerated dose than at the nontolerated dose.

The utilization of frequency filters to limit the presentation of toxicity data may be a rational way of presenting only the most relevant data. However, these thresholds were highly variable ranging from one patient to greater than or equal to 40% of patients for all grade toxicities and from one patient to greater than or equal to 10% of patients for toxicities grade 3 or higher (Table 3). Excluding 62 trials that did not specify what frequency filter was used, most trials (112 of 147, 76%) used a frequency filter of one patient to 10% of patients for AE of any grade.

Data presented at professional conferences are not intended to supply the compendium of information required to fully inform treatment decisions. The material provided by posters and slides are incomplete snapshots of trial data in which authors may not have the space to present the amount of data needed to fully represent drug toxicity. These materials are also accompanied by oral presentations in which the authors may anticipate supplying the additional context needed to interpret their slides and posters. Nonetheless, conferences represent the first exposure to novel agents for many providers and researchers and have the potential to influence therapeutic and research choices. Furthermore, oral disclosures of information are unlikely to be as widely promulgated as the "hard" data accessed from the poster or slide presentation itself. We suspect that the content of early phase clinical presentations at the ASCO annual meetings in relation to lung cancer is likely to reflect information provided across multiple tumor types at many, if not most, other general oncology conferences, such as the annual meetings of the European Society of Medical Oncology, the American Association of Cancer Research, and disease-specific conferences, such as the World Conference on

Lung Cancer. Consequently, we believe that our findings will be applicable to all kinds of communication in clinical trial research.

In general, a lack of standardization in toxicity data presentation leaves the field open to intentional or unintentional abuse. If any grade, or even high grade, adverse events are not reported because they are variably defined as below the limit required for presentation, the opportunity to find key clinical data worthy of communication to future prescribers, decision-makers, or other clinical influencers may be missed. Relevant toxicity may be pushed below such presentation thresholds simply through the use of treatment-related versus treatment-emergent attribution filters; by combining data across multiple dose levels, potentially "diluting" the appearance of the toxicity signal of the recommended dose level; or by not utilizing appropriately grouped terminologies such that highly-related toxic effects are fractionated, and thus, each then appears below a predefined presentation threshold. As a recent example of the latter phenomenon, the ASCO 2018 phase II presentation of the toxicities associated with lorlatinib, a third-generation ALK and ROS1 tyrosine kinase inhibitor, only presented treatment-related AEs occurring in greater than or equal to 10% of patients. Lorlatinib is now known to have adverse effects on numerous higher CNS functions, including memory, speech patterns, sleep patterns, seizures, mood, visual and auditory perceptions, and cognition. In the phase I and II ASCO presentations on lorlatinib in 2017 and 2018, only some grouped terms were used, including cognitive effects (23.3% in 2018) and mood effects (16% in 2018).[14] The final prescribing information states that overall CNS effects, including mood, cognition, speech, vision, and sleep-related AEs, occurred in 54% of patients receiving lorlatinib.[15] Vision disorders, sleep effects, and speech effects were not included in the initial ASCO presentation possibly because these fell below the 10% frequency threshold for the ASCO data set or may not have been considered treatment-related at the time.

Data on dose delays, dose modifications, and dose discontinuations are potentially highly useful as they could normalize the differences between reporting toxicities into pragmatic, clinically impactful rates of toxicity and were reported in 43% of trials. Nevertheless, without knowing the specific dose modification rules within a given trial and the events that lead to these dose reductions, even these data can be misleading. In our analysis, the trial-specific criteria for dose modifications were never provided in the available presentations. The same challenges in toxicity presentation also extend to later-phase trials. For example, the dose-reduction rate for alectinib versus crizotinib in the first-line ALEX trial was 16% but was 29% for brigatinib versus crizotinib in the comparable ALTA-1L trial.[16,17] Pragmatically, this suggests that alectinib is better tolerated than brigatinib. Nevertheless, 63% of the dose reductions for brigatinib in ALTA-1L were only for laboratory abnormalities defined as actionable within the protocol by specific blood levels rather than by symptoms.[17] Without specification of whether, for example, a laboratory abnormality was symptomatic, whether all or just a subset of patients were tested for a given laboratory abnormality, and what the similarities or differences in protocol-mandated requirements for dose reduction are between studies, concluding one agent is more "tolerable" than another based on trial dose-reduction rates, in fact, may or may not be valid. Dose-reduction rates in the real world, postlicensing, may offer additional insights into tolerability if dose decision-making in this setting were more free than the kind of constraints used in a trial protocol. Alternatively, other measures of "overall tolerability" may be important, such as health-related quality of life scores. Intriguingly, in this regard, despite the higher dose-reduction rate in ALTA-1L of brigatinib as compared with alectinib in the ALEX trial, only brigatinib and not alectinib achieved superior benefits over crizotinib in terms of health-related quality of life.[18,19]

Importantly, during our analyses, we noted that when laboratory abnormalities were presented, whether all patients were tested for these laboratory parameters or only a subset based on other criteria, such as clinical suspicion, was not described. To illustrate the potential need to define a relevant denominator (of all patients or only those tested) when quoting the percentage of patients with a laboratory abnormality, in the reports of reduced testosterone found in patients treated with crizotinib, nonindustry studies in both single- and multicenter settings reported a drug-associated drop in 100% and 84%, respectively, of men in whom these levels were checked.[20,21] In contrast, in the industry-provided medical information available from Pfizer, the manufacturer of crizotinib, this rate is reported as occurring in less than 1% of patients (across both sexes combined) in both the PROFILE 1014 and 1007 studies.[22] Beyond the obvious issue of including both sexes in the denominator, neither of these industry trial sources included routine assessments of testosterone in enrolled patients, which suggests that the denominator used would be more informative if modified from all patients treated with the drug to only those assessed for the laboratory value, and, in this example, also separated out by the patient's sex.[23,24]

**Table 6.** Suggested Minimum Disclosure Elements for AE Reporting in Oncology Clinical Trials

| | |
|---|---|
| 1. | Report both treatment-emergent and treatment-related AEs; specify criteria for defining "treatment-related." |
| 2. | Report AEs of any grade that occur in ≥10% of patients. |
| 3. | Report all AEs grade ≥3. |
| 4. | For closely related toxicities, assess both individual and overall incidence of AEs by CTCAE or MedDRA organ system class. |
| 5. | If using grouped terms other than organ system class, define components in each grouped term. |
| 6. | Indicate if a dose-escalation component was included. If so, describe AEs separately by dose level and at RP2D. |
| 7. | If dose escalation was included, report DLT criteria. |
| 8. | Report number and percentage of patients whose dose was reduced, dose was held, or discontinued treatment owing to AEs, by dose level where appropriate. |
| 9. | If reporting serious AEs, AEs of special interest and/or AEs leading to dose reduction, dose delay, or treatment discontinuation, provide criteria for these determinations. |
| 10. | Report denominators of those tested for laboratory abnormalities or other investigations, qualified by relevant clinical features where appropriate. |

AE, adverse event; CTCAE, Common Terminology Criteria for Adverse Event; DLT, dose-limiting toxicity; MedDRA, Medical Dictionary for Regulatory Activities; RP2D, recommended phase II dose and schedule.

We propose the introduction of minimum elements for adequate reporting of toxicities for phase I and II clinical trials, listed in Table 6. Although we did not include phase III trials in our analyses to focus on early data set presentations and enrich for trials with dose-escalation components, these minimum disclosure elements could also be extended and optimized across later-phase trial data sets and in any approved drugs prescribing information. By requiring authors to standardize or, at the very least, mandatorily disclose their AE reporting methodology, the quality of trial data shared at conferences and the resulting dissemination of relevant, comparable data could be significantly improved.

## Acknowledgments

The work of Dr. Camidge was partially supported by the University of Colorado Lung Cancer SPORE (P50CA058187). Drs. Gao and Smith were partially supported by the University of Colorado Cancer Center Support Grant (P30CA046934).

## References

1. Sherman RE, Li J, Shapley S, Robb M, Woodcock J. Expediting drug development-the FDA's new "breakthrough therapy" designation. *N Engl J Med*. 2013;369:1877–1880.
2. U.S. Food and Drug Administration. 2019 FDA Approves Third Oncology Drug That Targets a Key Genetic Driver of Cancer, Rather Than a Specific Type of Tumor. https://www.fda.gov/news-events/press-announcements/fda-approves-third-oncology-drug-targets-key-genetic-driver-cancer-rather-specific-type-tumor. Accessed November 13, 2019.
3. Ettinger DS, Wood DE, Aggarwal C, et al. NCCN guidelines version 7.2019 non-small cell lung cancer. https://www.nccn.org/patients/guidelines/cancers.aspx#nsclc. Accessed November 13, 2019.
4. Sacks CA, Miller PW, Longo DL. Talking about toxicity – "what we've got here is a failure to communicate" *N Engl J Med*. 2019;381:1406–1408.
5. American Society of Clinical Oncology (ASCO). 2019 Previous meeting demographics. https://meetings.asco.org/am/previous-meeting-demographics. Accessed October 17, 2019.
6. Data. Medicare.Gov. 2019 Physician compare datasets. https://data.medicare.gov/data/physician-compare. Accessed October 17, 2019.
7. Qunaj L, Jain RH, Atoria CL, Gennarelli RL, Miller JE, Bach PB. Delays in the publication of important clinical trial findings in oncology. *JAMA Oncol*. 2018;4:e180264.
8. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45:228–247.
9. Wen PY, Chang SM, Van den Bent MJ, Vogelbaum MA, Macdonald DR, Lee EQ. Response assessment in neuro-oncology clinical trials. *J Clin Oncol*. 2017;35:2439–2449.
10. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996;276:637–639.
11. Sivendran S, Latif A, McBride RB, et al. Adverse event reporting in cancer clinical trial publications. *J Clin Oncol*. 2014;32:83–89.
12. Ioannidis JPA, Evans SJW, Gøtzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med*. 2004;141:781–788.
13. American Society for Clinical Oncology. Abstract biostatistical guidelines. https://meetings.asco.org/am/abstract-biostatistical-guidelines. Accessed November 11, 2019.
14. Besse B, SB, Felip E, et al.2. Lorlatinib in patients with previously treated ALK+ advanced non-small cell lung cancer (NSCLC): updated efficacy and safety. Paper presented at: ASCO 2018 Annual Meeting; June 1–5, 2018; Chicago, IL.

15. U.S. Food and Drug Administration. LORBRENA® (lorlatinib) tablets, for oral use. Initial U.S. approval: 2018. https://www.accessdata.fda.gov/drugsatfda_docs/label/2018/210868s000lbl.pdf. Accessed November 13, 2019.
16. Huber RM, KD, Ahn MJ, et al. Brigatinib in crizotinib-refractory ALK+ non-small cell lung cancer: efficacy updates in the exploratory analysis of CNS ORR and overall ORR by baseline brain lesion status. Paper presented at: ASCO 2018 Annual Meeting. June 1–5, 2018; Chicago IL.
17. Camidge DR, Kim HR, Ahn MJ, et al. Brigatinib versus crizotinib in ALK-positive non-small-cell lung cancer. *N Engl J Med*. 2018;379:2027–2039.
18. Pérol M, Pavlakis N, Levchenko E, et al. Patient-reported outcomes from the randomized phase III ALEX study of alectinib versus crizotinib in patients with ALK-positive non-small-cell lung cancer. *Lung Cancer*. 2019;138:79–87.
19. Campelo RC, Lin HM, Perol M, et al. Health-related quality of life (HRQoL) results from ALTA-1L: phase 3 study of brigatinib vs crizotinib as first-line (1L) ALK therapy in advanced ALK+ non-small cell lung cancer (NSCLC). *J Clin Oncol*. 2019;37(suppl 15):9084.
20. Weickhardt AJ, Rothman MS, Salian-Mehta S, et al. Rapid-onset hypogonadism secondary to crizotinib use in men with metastatic nonsmall cell lung cancer. *Cancer*. 2012;118:5302–5309.
21. Weickhardt AJ, Doebele RC, Purcell WT, et al. Symptomatic reduction in free testosterone levels secondary to crizotinib use in male cancer patients. *Cancer*. 2013;119:2383–2390.
22. Pfizer. 2019 Product Monograph: Xalkori Crizotinib Capsules. Pfizer. https://www.pfizer.ca/sites/default/files/201904/XALKORI_PM_215123_27March2019_E.pdf. Accessed January 13, 2020.
23. Shaw AT, Kim DW, Nakagawa K, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med*. 2013;368:2385–2394.
24. Solomon BJ, Mok T, Kim DW, et al. First-line crizotinib versus chemotherapy in ALK-positive lung cancer. *N Engl J Med*. 2014;371:2167–2177.