

The FDA's Final Rule on Expedited Safety Reporting: Statistical Considerations

Janet WITTES, Brenda CROWE, Christy CHUANG-STEIN, Achim GUETTNER, David HALL, Qi JIANG, Daniel ODENHEIMER, H. Amy XIA, and Judith KRAMER, for the Biostatistics Working Group of the CTTI-IND Safety Reporting Team

In March 2011, a Final Rule for expedited reporting of serious adverse events took effect in the United States for studies conducted under an Investigational New Drug (IND) application. In December 2012, the U.S. Food and Drug Administration (FDA) promulgated a final Guidance describing the operationalization of this Final Rule. The Rule and Guidance clarified that a clinical trial sponsor should have evidence suggesting causality before defining an unexpected serious adverse event as a suspected adverse reaction that would require expedited reporting to the FDA. The Rule's emphasis on the need for evidence suggestive of a causal relation should lead to fewer events being reported but, among those reported, a higher percentage actually being caused by the product being tested. This article reviews the practices that were common before the Final Rule was issued and the approach the New Rule specifies. It then discusses methods for operationalizing the Final Rule with particular focus on relevant statistical considerations. It concludes with a set of recommendations addressed to Sponsors and to the FDA in implementing the Final Rule.

Key Words: Expedited safety reporting; Final Rule on expedited safety reporting; Serious adverse events.

1. Introduction

People not involved in clinical trials may think that identifying harms caused by a drug or a biologic prior to the product's approval is simple—companies should report bad things that occur to individuals who take the product and the United States Food and Drug Administration (FDA) will tell the public if the data raise concerns about safety. Those of us who work in clinical trials and drug development, however, know that identification of harms caused by the investigational treatment is fraught with problems. Although many participants in clinical trials experience adverse events (i.e., untoward things that happen to a patient who is taking a product), it is often challenging to distinguish those events that were adverse drug reactions (i.e., events caused by an investigational product) from those that might have happened in the absence of the product.

The FDA has long had a rule (21 CFR 312.32) calling for prompt (within 15 days) reporting of any serious unexpected (i.e., not in the investigators' brochure or labeling) adverse experience "associated with use of a drug" (i.e., if there was a reasonable possibility that the drug may have caused the event). The "associated with" language made the adverse experience a "suspected" adverse reaction. Under the previous system for reporting

adverse events prior to approval of a product, the FDA had been receiving many reports of individual serious adverse events often when there was little reason to believe in a causal relationship between the event and the product. Some of these events may have been manifestations of the underlying disease or events that occur in the study population independent of drug exposure. Moreover, the system was extremely labor intensive for sponsors, investigators, Institutional Review Boards (IRBs), and regulators. To address these problems, in September 2010, the FDA published in the Federal Register a Final Rule clarifying the approach required for expedited reporting of unexpected serious adverse reactions that occurred in studies conducted under an Investigational New Drug (IND) application, focusing particularly on what “associated with” or “suspected” meant. See the Code of Federal Regulations (CFR) for a description (U.S. Department of Health and Human Services and Food and Drug Administration 2011). Although the Rule became effective from March 28, 2011, the FDA exercised “enforcement discretion” until September 28, 2011. To help explicate the Final Rule, FDA issued the finalized Guidance in December 2012 (U.S. Department of Health and Human Services and Food and Drug Administration 2012). The Final Rule and Guidance have clarified definitions and delineated what events qualify for expedited reporting; they did not change the required timing of reporting. The Rule applies not only to trials conducted under an IND [21 CFR 312.32 and 21 CFR 312.64 (b)], but also to studies of bioequivalence and bioavailability [21 CFR 320.31 (d) (3)]. Before promulgation of the Final Rule, an event must have satisfied three criteria in order to qualify for expedited reporting. The event must have been serious, unexpected, and associated with study drug. Under the Final Rule, these criteria are still necessary but because of previous confusion about the definition of “associated,” the wording of the Final Rule now simply states that the sponsor “must report any suspected adverse reaction that is both serious and unexpected,” and that “the sponsor must report an adverse event as a suspected adverse reaction only if there is evidence to suggest a causal relationship between the drug and the adverse event.” Not only does the Rule give these definitions, but it gives examples of categories of drugs that should be reported as single occurrences and those that should be analyzed in the aggregate before deciding whether there is evidence of causality [21 CFR 312.32(a)]. The Final Rule specifies the need for the sponsor to evaluate the available evidence and to judge the likelihood that the drug actually caused the adverse event.

Prior to the Final Rule, many sponsors interpreted the criteria regarding causality conservatively, meaning that they reported events in an expedited manner if they could not rule out an association. This approach often

led to reporting essentially every serious adverse event. If either the investigator or the sponsor thought a serious unexpected event had a possible causal relationship to the study drug (i.e., a relationship could not be ruled out), the sponsor reported it expeditiously. Under the Final Rule, on the other hand, even if the investigator identifies an event as drug-related, but the sponsor finds no *evidence* of causality, the sponsor should not report the event on an expedited basis to the U.S. FDA. That is, under current U.S. regulations, the sponsor’s determination of causality defines the need to report. Sponsors do, however, routinely report these events in an expedited fashion to other regulatory agencies and often to IRBs, investigators, and Data Monitoring Committees. Conversely, sponsors should report the event expeditiously if the investigator declares an event unrelated, but the sponsor has evidence of relatedness.

FDA’s Rule and Guidance illustrate adverse events that are readily interpretable as single or small numbers of events (agranulocytosis, Stevens-Johnson Syndrome) but notes that many others would be anticipated to occur in the study population (stroke or heart attacks in patients > 65) so that a causal relationship would need a higher rate in the drug-treated group. How to deal with these “anticipated” events, and when to report them, is a critical problem to avoid routine submission of uninformative expedited reports.

Eight months after the Final Rule became effective, the Clinical Trials Transformation Initiative (CTTI) conducted a survey of pharmaceutical sponsors to understand the structures, processes, and procedures in place to monitor safety of products being studied under an IND application. In February 2012, the CTTI convened a meeting of experts to discuss the survey results, including a group of biostatisticians from academia, FDA, and industry designated as a Biostatistics Working Group. Following the meeting, the Biostatistics Working Group was asked to confer among themselves to provide statistical advice to CTTI concerning implementation of the Final Rule. Archdeacon et al. (2014) described the survey, the expert meeting, and the formation of the Biostatistics Working Group.

This article describes the Biostatistics Working Group’s understanding of the Final Rule and our recommendations. All members of the group are involved in clinical trials. Our goal is to provide a statistical framework to help guide sponsors in judging whether a serious adverse event is subject to expedited reporting to the FDA. (21 CFR 312.32 (c)(i) does not itself mention unmasking.)

The article uses many terms; we define them in a lengthy [Appendix](#). To understand our recommendations, readers need to be conversant with the many terms used in the Code of Federal Regulations (CFR). Much of the

article distinguishes three types of serious adverse events (SAE) as defined in the CFR and described in Section 2: Category A, rare events known to be strongly associated with drug exposure; Category B, events that are neither common in the exposed population nor commonly associated with drug exposure; and Category C, those events that are common in the population under study so that establishing their relationship to the investigational product requires comparing frequencies in a treated group to a control group.

We then describe how data on serious adverse events are usually collected in clinical trials (Section 3) and review the expedited reporting to the FDA typically practiced prior to the Final Rule (Section 4).

Having summarized the relevant definitions and practices, we present our collective view of specific statistical and operational aspects of data on safety: masking (Section 5), causality of adverse events (Section 6), thresholds for expedited reporting of Category B events (Section 7), identifying potential harms from combined data from completed trials (Section 8), multiplicity (Section 9), and how pharmacovigilance groups within companies and Data Monitoring Committees, respectively, deal with events of Category C (Sections 10 and 11, respectively). We end with a discussion and a set of recommendations about implementing the Final Rule (Section 12).

Throughout the article, we ground our statements either in the CFR or in the FDA Guidance that describes implementation of the Final Rule (U.S. Department of Health and Human Services and Food and Drug Administration 2012). Parenthetical citations within the body of the text refer to the CFR; other sources are listed in the References section at the end of the article (Section 13). Out of deference to our colleagues in ophthalmology, we use the term “masked” instead of the term “blinded” to refer to treatment assignment.

In this article, we are not aiming for precision nor are we looking for methods that will identify adverse events that are necessarily caused by a product under study. Instead, we aim for an approach that will identify likely harms early without flooding the IND expedited safety reporting system in the United States with noise.

Of interest to the discussion in this article are the boxes in bold in Table 1: the serious unexpected suspected adverse reactions and the serious unexpected adverse reactions.

2. Types of Serious Adverse Events

The Final Rule gives examples of three categories of adverse events that would qualify as suspected adverse

reactions. In discussing these three categories, we use the nomenclature A, B, and C described below.

Category A. An event that is uncommon and known to be strongly associated with drug exposure (e.g., angioedema, hepatic injury, Stevens-Johnson Syndrome, and agranulocytosis). Such events are informative as single cases when fully investigated and characterized. The occurrence of even one of these events in an experimental arm of a trial would meet the definition of a *suspected adverse reaction* unless there is incontrovertible evidence of noncausation. Sponsors should unmask each of these events and, if it occurs in the investigational drug arm, report it. We note, however, that even these events can occur spontaneously, albeit rarely. For example, not all hepatic injury is drug-induced, particularly in patients with certain underlying diseases such as heart failure. A separate guidance, which describes drug-induced liver injury, calls for active monitoring of suspect Hy’s Law cases with sufficient follow-up to assess causation (U.S. Department of Health and Human Services and Food and Drug Administration 2009b).

Category B. An event that is not commonly associated with drug exposure and is uncommon in the population exposed to the drug (e.g., tendon rupture in any recipient of a product or a myocardial infarction in a study of young women). In such cases, a few occurrences are sufficient to attribute causality. If the event occurs in association with other factors strongly suggesting causation (e.g., clear temporal association or reoccurrence on rechallenge), a single case may be sufficiently persuasive to meet the definition of a *suspected adverse reaction*. Often, more than one occurrence of a specific event is necessary before the sponsor can judge that there is a reasonable possibility that the drug caused the event.

Category C. Events that an aggregate analysis of data from a clinical development program indicates occur more frequently in the drug treatment group than in a concurrent or historical control. Candidate events for the aggregate analysis may be known; they may come from masked or unmasked data. These events may be known consequences of the underlying disease or condition under investigation or other events that commonly occur in the study population independent of drug therapy, but in both cases their occurrence is anticipated, so that interpretation of individual events will rarely be possible. Events in this category are the most problematic for determining whether they need expedited reporting because of the difficulty of identifying whether an individual event is causally related to the drug. Most serious adverse events that occur during drug development are known to occur in some frequency in the population under study regardless of drug exposure; consequently, a single case of a Category C event is not sufficient to infer a reasonable possibility that the drug caused the event.

Table 1. Classification of adverse events occurring in clinical trials (FDA definitions)

		Serious			Not serious		
		Is there evidence that the investigational product caused the event?					
		No	Yes, but not certain	Yes, certain	No	Yes, but not certain	Yes, certain
Is the event expected?	No	Serious unexpected adverse event	Serious, unexpected suspected adverse reaction	Serious, unexpected adverse reaction	Unexpected adverse event	Unexpected suspected adverse reaction	Unexpected adverse reaction
	Yes	Not applicable	Serious, expected suspected adverse reaction	Serious, expected adverse reaction	Expected adverse event	Expected suspected adverse reaction	Expected adverse reaction

NOTE: The two cells in bold describe the events that are to be reported as expedited events. Note that some other events should be reported in an expedited fashion (e.g., findings from epidemiological studies, animal studies, in vitro testing); see 21 CFR 312.32 (c)(1)(ii–iv).

Under the Guidance describing the Final Rule (U.S. Department of Health and Human Services and Food and Drug Administration 2012, p. 11–12), instead of reporting each Category C event individually (because it will not be possible to conclude that the event is a suspected adverse reaction), the sponsor should compare “at appropriate intervals” (p. 11) the number of such events in each arm of a trial and report the events to the FDA expeditiously as IND safety reports if an observed excess rate in the treatment arm relative to control suggests a reasonable possibility that the drug caused the adverse event. If the trial is part of a larger development program, the sponsor should evaluate the entire clinical trial database (both masked and unmasked studies) to determine whether to submit the events as expedited reports. Periodic reviews that include study endpoints (e.g., mortality or major morbidity) may require careful handling of Type I error rate.

The boundary between Categories B and C is sometimes murky. For example, a heart attack in a 25-year-old woman would likely fall into Category B while one in a 75-year-old man would fall into Category C. Some people would assign a heart attack in a 35-year-old man to Category B and others would see it as falling into Category C. For events of Category C, even experienced reviewers may find it difficult to identify an excess of a particular event if they look only at individual reports of unexpected, but anticipated, events. In general, when the boundary between B and C is unclear, erring on the side of B is probably the preferred action.

3. Collection and Classification of Data on Serious Adverse Events

In contrast to the active, systematic approach to collection of data standard for efficacy, collection of data

on adverse events typically relies on undirected, passive query of the participants in clinical trials. When an event is identified as “serious,” however, the investigator generally collects supportive documentation, for example, hospital records and, if the participant died, death certificates to help determine the actual type of event and its likely relationship to study drug. The actual occurrence of serious adverse events in clinical trials, however, may still be subject to a lack of timely collection. Further, if the protocol-defined study visits are infrequent, some serious adverse events may be missed because the study participant fails to report them.

Typically, to collect data on safety, the investigator or delegate asks participants at each clinic visit or telephone call about any medical event that occurred since the previous contact. A member of the investigator’s team records as an adverse event any event this question elicits. With the exception of ongoing events, the investigator usually does not prompt participants about events they had reported previously. Some studies prompt for adverse events of special interest or for symptoms commonly associated with such events of special interest.

When an investigator reports an adverse event to a sponsor, a human or machine coder classifies the event according to a dictionary; currently the one most frequently used in the Medical Dictionary for Regulatory Activities (MedDRA). The reported event receives a “lowest level” term that is then mapped to a “preferred term,” which in turn is mapped to a “high level term” and a “high level group term.” Finally, it is classified into a “System Organ Class” (SOC). The current version of MedDRA (v. 16.1) has over 72,000 lowest level terms, over 20,000 preferred terms, 1717 high level terms, 334 high level group terms, and 26 system organ classes. Standardized MedDRA Queries (SMQs) are available for grouping events that may be related to each other and intended to reflect relevant medical concepts (some individual events

identified by the terms within an SMQ will not actually be cases of the medical concept of interest). See www.meddra.org for a description of MedDRA. Note that a single event may be reported as a diagnosed single event or as individual symptoms and signs, further complicating coherent review.

Reviewing tables of MedDRA coding, even of serious adverse events, can be confusing. Reports that present MedDRA “preferred” terms have many rows with counts that are so small that a signal may be missed (e.g., many lines have fewer than a handful of events). For SAEs, sponsors typically review not only the MedDRA codes but each SAE’s case description. This review is valuable as many events are similar; they often represent different terms describing the same medical phenomenon but the structure of the tables make discerning their relationship to each extremely difficult. For example, “pulmonary edema” will be classified in the Respiratory SOC and “heart failure” in the Cardiovascular SOC even though they may represent the same medical condition. This type of splitting, while allowing precise description of a given event, often makes it difficult to identify all of the cases that reflect a medical concept of interest. Often it is more useful to combine events or groups of events reflecting identical or very similar underlying pathophysiology in spite of potentially different identifying MedDRA terms.” The SMQs address this type of problem. For example, if cases of heart failure are labeled “heart failure,” “congestive heart failure,” “right heart failure,” “left heart failure,” “dyspnea,” and “pulmonary edema,” each event may occur with a low frequency, but the totality of the events may occur with a worrisome frequency (Littlejohn et al. 1991). The difficulty is further exacerbated in multinational trials where nomenclature of symptoms and events, and even cultural propensity to report, may vary considerably.

To simplify the tables, some sponsors in reporting to Data Monitoring Committees, publications, and FDA advisory panels list only those events that occur with a frequency more than 1% (or 5% or 10%). Often the FDA for drug labels and journals require such simplification. Specifically for Data Monitoring Committees, however, we strongly recommend that sponsors report all SAEs but that they devise systems to lump similar events. Below we suggest some approaches to such lumping.

Looking only at higher order term classifications (i.e., lumping) produces tables with higher counts for many listed events and thus more stability in the estimated event rates; however, overaggressive lumping may commingle adverse events of real clinical concern with many others that are not or events with no common pathophysiology. Often, the least clinically relevant items can be the most frequent in the higher order classification. Thus, noise can overwhelm important signals of harm and so reviewers of these tables may miss the important signals. The

choice of when, and how, to lump will depend on clinical judgment related to the particular disease, the particular intervention, the likely mechanism of action, the size of the study, and, in multinational studies, even the way in which physicians from different countries describe events. As mentioned earlier and further discussed later, sometimes a Standardized MedDRA Query (SMQ) provides a useful starting basis for lumping events that can facilitate cross-study and cross-company standardization.

Another source of confusion is that what gets reported and coded may change with each patient visit, depending on how the patient describes the experience, how the physician or nurse reports it, and how the coder codes it. A serious adverse event that is identified at hospitalization may be initially reported in terms of symptoms or by a presumptive diagnosis. When a diagnosis is confirmed, the serious adverse event form and adverse events case report form may both be corrected; the amount and type of data collected may diverge; or changes may show up only in the comments or text part of the report.

Identifying potential Category C events that should be reported is complicated by the changing patterns of hospitalization where events that in the past would have been categorized as serious by virtue of causing hospitalization are now treated on an outpatient basis and are therefore not “serious.”

Pharmaceutical sponsors typically record serious adverse events in both a clinical trial database and a global safety database for the purpose of regulatory reporting (Archdeacon et al. 2014). They add these events to the global safety database as soon as the investigator reports them to the sponsor. This may be as soon as the event occurs if the investigator is the patient’s primary care physician, but it may not be recorded until the next scheduled visit otherwise. Because the event may not appear in the clinical database for weeks or even months, a global safety database is very important for implementing IND safety reporting.

Accurate estimation of the frequency of specific events requires careful data collection. To support the FDA’s Final Rule, many investigators will need more consistent education so they know what to report and how to do so. If investigators are inadequately trained about the expected level of detail in reporting symptoms or diagnoses, identical events will continue to be scattered across different preferred terms and SOC. Identical events will be classified both by diagnosis and symptom complexes. These difficulties are most problematic for nonserious adverse events (see Crowe et al. 2013). One approach to aid investigators is to remind them to look at the way in which they previously reported an event for a specific person.

An additional source of noise in the system is the frequent inaccurate reporting of onset dates. This problem

is especially important for events like cancer or bacterial pneumonia where the date reported as the time of onset is often the date of confirmation of the diagnosis by biopsy, X-ray, or culture. This divergence is particularly problematic when comparing the date of onset with the date of start of trial medication. When onset preceded randomization or the start of treatment, the question becomes whether treatment exacerbated the event rather than whether treatment caused it.

4. How the System Previously Worked

Prior to the Final Rule, the system for reporting expedited serious adverse events worked in practice as follows. Investigators in a clinical trial received a protocol, an Investigator's Brochure describing the product including the preclinical and prior clinical experience with it and the adverse events associated with its use, and paper or electronic case report forms with instructions on how to fill them out. During the trial, the investigator was to record on the case report form each adverse event the participant experienced. The data collected were to include the investigator's assessment of the relationship of the event to the product, its severity (mild, moderate, or severe), and whether it was "serious" or not. The investigator was required to report each fatal or life-threatening serious adverse event to the sponsor "immediately" [21 CFR 312.64(b)] and each serious adverse event "promptly." The sponsor then made its own assessment of causality and whether the event was unexpected.

Sponsors were to report to the FDA each unexpected fatal or life-threatening suspected event associated with the study drug within seven days of learning about it and each unexpected, serious, and suspected event within 15 calendar days. Sponsors typically used the more conservative of either the sponsor's or investigator's assessment of causality to make this determination (i.e., an event would be considered "suspected" if either the sponsor or the investigator indicated that it was likely related to drug). By regulation, these reports were to go not only to the FDA but also to participating investigators who, in turn, should decide whether to report these events to their IRB or Ethics Committee. Many sponsors told the investigators whether an event should be reported to the IRBs and ECs. Although not required by regulation, if a trial had a Data Monitoring Committee (DMC) many sponsors sent these expedited reports to the Chairperson, and sometimes to all members, of the DMC.

Importantly, under the former system, both the investigator and the sponsor contributed to assessing the final determination of causality (i.e., "relatedness"). The sponsor did not generally downgrade an investigator's assessment for the purpose of IND safety reporting, but

might judge an event "related" even if the investigator did not.

If the data came from an ongoing masked trial, the pharmacovigilance group in some companies unmasked the treatment assignment before sending the report to the FDA; however, in many cases the report from the investigator to the sponsor and the report from the sponsor to the FDA was masked, that is, no one—neither the investigator, the sponsor, nor the FDA—knew whether the reported event occurred to a participant on the investigational or the control arm. This process led to a huge burden on the sponsor and the FDA, as well as on many investigators, Institutional Review Boards, and Data Monitoring Committees. A large proportion of the many reports they received described events that had not even occurred on the investigational arm; and, among those on the investigational arm, many (and perhaps most) were probably unrelated to treatment; for anticipated events, there is no way to assess the likelihood that the event was drug related. That assessment requires comparing rates in drug-treated and control patients.

For development programs studying drugs in reasonably healthy persons, or programs studying orphan diseases where the total development program includes at most several hundred participants, only a few events required a 15-day report; however, for large studies in patients with serious underlying disease—in particular trials of cancer patients or long-term Phase 3 trials in cardiovascular disease with thousands or even tens of thousands of participants, hundreds or even thousands of such events may have occurred.

Many of the changes and clarifications in the Final Rule will require sponsors to modify their previous approach to reviewing, analyzing, and reporting safety data. In particular, event rates in treatment and control groups will need to be compared. For that reason, sponsors, investigators, IRB members, and DMCs will need to be fully educated about the Final Rule and its subtleties.

5. Masking

In the past, various sponsors have had different systems for unmasking serious adverse events. Some sponsors have routinely unmasked all serious adverse events and sent to the regulatory bodies information only on those events that occurred in the experimental group. Other sponsors have kept all staff masked except on a need-to-know basis, where need-to-know was defined as necessary for the safety of a participant in the trial. As a consequence of the Final Rule's emphasis on the need for evidence of causality, sponsors must know whether the event in question occurred to a participant who received the investigational product, and for most adverse

events, must compare rates in treated and untreated patients. To that end, the Guidance recommends unmasking the treatment code for events potentially subject to expedited reporting. The FDA should not receive expedited reports for serious adverse events occurring to a placebo participant or to a participant in the active arm who did not actually receive study drug. CIOMS VI (Council for International Organizations of Medical Sciences (CIOMS) Working Group VI 2005) and SPERT (Crowe et al. 2009) provide advice for planning these processes. CIOMS VI discusses safety management teams and aggregate safety reviews. SPERT describes Program Safety Analysis Plans (PSAPs). The program-level reviews can help to identify reportable Category C events.

The Final Rule explicitly directs sponsors not to submit reports of study endpoints as expedited reports even when the data are unmasked; thus, a myocardial infarction occurring in a trial with MACE (heart attack, stroke, and cardiovascular death) as a protocol-designated endpoint should not be considered a serious adverse event subject to expedited reporting.

Many sponsors are uneasy about unmasking, probably because of concerns about overall study integrity. Consider first events of Categories A and B. To maintain the integrity of the trial, only personnel who are not on the study team should unmask adverse events. In many companies, the individuals performing these tasks are members of an internal safety organization that is administratively separate from the clinical part of the organization involved in the operations of the trial. In this article, we call the group responsible for unmasking safety reports within a company the “Safety Review Committee.” A sponsor that desires to keep investigators and study team masked could send expedited cases to the FDA unmasked but send them masked to the investigators and IRBs. Another approach, which is consistent with the FDA’s Guidance, would send the expedited cases unmasked to the investigators; however, some other regulatory bodies are currently opposed to sending unmasked reports to investigators. Individual sponsors will have to evaluate their own processes and make decisions regarding unmasked reporting in a way that is consistent with the often conflicting advice from different regulatory agencies.

Events of Category C present a more complicated problem than events of Categories A and B. Sponsors do not usually unmask individual events of Category C. While unmasking unexpected serious suspected adverse events of Category C is clearly critical to determining whether a serious unexpected suspected adverse reaction occurred, a relevant question is whether unmasking of investigators for Category C events, or unmasking even the Safety Review Committee, could introduce bias into the assessment of efficacy and of subjective safety outcomes in confirmatory (Phase 3) trials. (The potential for this

type of bias is much less relevant for Category A and B events.) Site investigators who are unmasked to treatment assignment for Category C events would be able to compare study groups with respect to outcome assessments for a number, not necessarily a small number, of patients. We do not favor routine unmasking of Category C events at the investigator level. Instead, we favor assigning a fire-walled committee the responsibility for unmasking these events. For large companies, a group internal to the company (which in many companies is the Safety Review Committee itself) can serve this role; small companies may need an external committee to keep a firewall intact. For trials or drug development programs with a DMC, that committee can play an important role in unmasking (see Section 11). We summarize our recommendations for unmasking Category C events in Section 12.

6. General Considerations on Assessing Causality

Sponsors need processes in place to review evolving safety data periodically and they must expeditiously report events with medical evidence of causation. The process of review has several levels. First, sponsor staff reviews individual cases soon after the report of the event enters the company’s safety system. Next, staff reviews rates of serious adverse events of interest in individual trials, masked while a masked trial is ongoing but fully unmasked when the trial is complete. For trials with a DMC, the DMC is generally unmasked to the ongoing data (and we urge that sponsors and DMC members who object to unmasked DMCs reconsider their stance). Finally, for important serious adverse events, sponsors periodically review safety data for the entire drug development program. For many open-label trials, although the investigator and the participant know the treatment given, only a limited number of team members, specifically those who visit the clinical sites, are aware of the treatment assignment. Many sponsors go to great lengths to keep most of the remaining clinical and statistical team members masked to treatment assignment. Of course, for one-arm trials, everyone involved is aware of treatment assignment.

In the absence of convincing medical evidence relating drug use to an individual event, one must rely on statistical evidence of a between group difference for most serious adverse events. Both medical and statistical considerations influence how large an observed increase in event rate in the treatment arm relative to control should lead to the judgment to report the event on an expedited basis, and such matters as consistency of effects in different studies will be germane. Some events may be reportable even if the nominal p -value comparing the

treated group to control is high. For example, three occurrences of an event in a treated group and none in its equally sized control group gives a one-sided p -value of 0.125, but if the event in the population at large is very rare, the event is serious, and the causality even remotely associated (i.e., the event is of Category B), the sponsor may opt to report it (and the FDA may expect a report of such an event).

On the other hand, the sponsor may decide not to report some events with a low p -value because the relationship to study drug is deemed remote and the probability of a false positive is high (see Section 9).

Various strategies can help evaluate the likelihood that an observed increase in event rate in the experimental group represents likely causality. Lumping closely related events may provide insight into whether the event likely reflects causality. Information regarding the mechanism of action, the pharmacology of the drug, and nonclinical toxicology may provide evidence that the increase represents evidence or merely reflects the play of chance. The degree of observed difference in the event rates as measured by the p -value, the confidence interval, and the relative risk represent classical statistical approaches to assessing likely relationship.

Chance baseline imbalances in other relevant biological measures may contribute to an observed imbalance in the frequency of a particular serious adverse event. In that case biological factors may be contributing to the event more in one group than the other. Such an insight is not possible if a sponsor is viewing masked data or is only unmasking incident cases and has not adopted a mechanism for a more thorough review of unmasked data in an ongoing clinical study. In some cases, it may be feasible to unmask only relevant data. For example, if one is considering some serious adverse events related to liver injury, it would seem to be easy to create an unmasked dataset that contains only liver laboratory values. What may appear easy in principle, however, may be difficult in practice—the feasibility of such separation will depend crucially on the structure of the database and the experience of the data managers.

Data from completed studies or findings from other sources (e.g., literature on drugs in the same or related class) can sometimes help define whether an event should be reported. Setting a low threshold for sending in 15-day reports will mean that noise may swamp true signals, but setting the bar too high will delay detection of important events. The challenge is to find a “sweet spot.” Some sponsors routinely produce a so-called rolling Integrated Summary of Safety that includes data only from completed studies, but we authors believe that delayed exploration of large ongoing studies is undesirable. Another possible approach is a “program level safety review” that includes a review of all safety data for the compound from both

completed and ongoing studies. In some companies, this review is unmasked for completed studies but masked for ongoing studies. The pharmacovigilance groups in other companies unmask some serious unexpected adverse events if the clinical reviewers deem unmasking important.

In addition to the reporting of events observed in clinical trials, sponsors should report a potential harm that emerges from other types of studies (e.g., epidemiological studies) in cases where the harm, if caused by the product, would influence the investigator’s decision to enroll a patient in a trial or the investigator’s approach to managing the patients in the trial. This determination may depend on the outcome of patients experiencing the event in question. For example, the threshold for when an investigator needs to be informed about a possible harm will depend on whether the event may be fatal or permanently disabling rather than an event that, although serious by regulatory definition, resolves in nearly all patients. The sponsor should err on the side of reporting if knowledge of the suspected, unexpected serious adverse reaction would help the physician prevent, identify, or manage the event.

For example, a 3/100 incidence on active versus 0/100 on control may represent a reportable event if other data (e.g., mechanism, related outcomes) suggest that the product likely caused the event but the same frequencies do not constitute much evidence when they come from an analysis of 100 types of SAEs from a program-wide analysis or when the event is one that commonly occurs in the population under study. Even a single Category B event might be reportable if the event and its case description show sufficiently compelling evidence of causation.

The trigger for reporting events should be well articulated. Some organizations assign a safety lead to a product. This safety lead, a member of the clinical team associated with a product, has access to all SAEs reported on the product. Working with the clinical team, the safety lead may assess data from individual and multiple trials and decide whether the Safety Review Committee should unmask the data. The safety lead may, instead, ask the DMC, if there is one, to provide an unmasked assessment of the likely causality of the event in question. Other escalation procedures are also possible, depending on the organization.

Internal committees that review unmasked data, especially when some of the data come from ongoing masked trials, should operate under strict procedures about how to handle the unmasked safety data keeping in mind that these are not formal “interim results” (i.e., no effectiveness data are being evaluated). Procedures should also describe how to communicate between the committee and any other existing review committee (such as an internal or external DMC). All committee members must

understand the importance of guarding the confidentiality of any unmasked data and maintaining the integrity of the trial while protecting patient safety. They must understand that they must not share or discuss the unmasked data with clinical teams.

Small companies that cannot maintain a properly fire-walled committee should consider convening an external advisory board whose members have detailed knowledge of the compound and are willing to devote the time and effort necessary to assess emerging issues with short notice. External advisory boards need guidelines on how to communicate interim results and decisions with company personnel. Ideally, such guidelines should be articulated in a charter.

7. Guidelines for Expedited Reporting Category A and Category B Events

By their very nature, Category A events are uncommon. The sponsor should not report a Category A event as a 15-day report if it occurred on the control arm, or if there is strong evidence of noncausation. The sponsor should report, unmasked, all other Category A events.

In programs with very few serious adverse events, the same criteria used for Category A may apply also to Category B. For programs with a sizeable number of events, one approach to determining when to report Category B events expeditiously is to define guidelines for expedited reporting as part of the protocol or as part of a program-wide safety analysis plan. For Category B events, the guidelines are most likely based on a medical review of individual cases as well as comparisons to the rate in the population under study where that rate is calculated from the rate in the control group (if the study has a control group) or from epidemiological or natural history data. Because these events are unanticipated, the calculated rates cannot be prepared a priori; rather, they need to be calculated ad hoc (see, e.g., Lachenbruch and Wittes 2007).

The following table provides suggested guidelines for those drug development programs where the study team decides not to report all Category B events. The criteria for a given investigational product, disease, and event will depend on the judgment of the product team. We offer the sample guidelines as a structure for thinking about Category B events. Assuming sponsors are judicious in what they send to the FDA, and assuming that single reports are interpretable, in accordance with the Final Rule, sponsors should expect that investigators would forward to the appropriate IRB most of those that are interpretable as single reports.

The cumulative incidence of suspected unexpected serious adverse reactions will depend on the patient pop-

ulation (healthy vs. diseased, patients with stable vs. life-threatening conditions, young vs. elderly, immunocompromised vs. immunointact, etc.), length of follow-up, number of patients in the database, the ease of differential diagnosis, and so on.

For uncommon serious adverse events, the statistical power for signal detection will be very small even without any multiplicity adjustment to p -values. Therefore, for rare events in an ongoing development program, the interpretation will usually be based more on clinical judgment than on statistical significance of p -values. Bayesian approaches (Berry and Berry 2004; Xia et al. 2011b) could be useful in this context, especially when a reasonable amount of prior information is available (e.g., for the event rate in the control arm).

8. Detecting Potential Evidence of Harms in Data From a Set of Clinical Trials

The Final Rule has stimulated interest in appropriate ways to aggregate data from a set of clinical trials to identify serious adverse events subject to 15-day reports. The adverse events of interest here are the anticipated events, where the issue is whether the rate of these events is increased in drug-treated patients. We use the term “aggregate” to refer to analyses that summarize events of interest at the study level as well as analyses intended to combine comparisons of events across studies. Aggregate analyses are likely to be exploratory in nature in the early development period, but may become more targeted as events of potential interest are identified. They become the focus of statistical analyses for which meta-analytic type methods might be useful.

The CIOMS VI working group (Council for International Organizations of Medical Sciences (CIOMS) Working Group VI 2005), Crowe et al. (2009), and Xia et al. (2011a) have recommended periodic aggregate analysis of clinical trial safety data. A Program Safety Analysis Plan (PSAP) can be helpful for prospective planning. Such a plan sets the framework for more extensive safety analyses performed later in the drug development process as well as in other safety analyses for regulatory submissions. Importantly, it encourages standardization of terminology for adverse events of special interest. For description of PSAPs and how to implement them, see Xia and Jiang (2014), Crowe et al. (2015), and Chuang-Stein and Xia (2013). The frequency of these periodic analyses will depend on many factors, most prominently, the population being studied, the diseases under investigation, the specific drug and drug class, and the nature of the adverse reactions of concern.

Analysis of data that come from randomized trials must preserve the structure of the design of those trials.

We list three cases of aggregate analysis in order of their analytic complexity. First are data from ongoing drug development programs in which the total exposure to the investigational product is less than generally needed to assess safety. Second are data from completed drug development programs where the information come from analyses that will form part of, or that constitute, an integrated summary of safety (ISS). The third, and most difficult, type of analysis comes from programs with a mixture of completed and ongoing studies, especially where at least one of the ongoing studies is a large Phase 3 trial. The first and third types are relevant to identifying serious adverse events subject to expedited reporting. When all trials for a particular submission are complete, aggregate analyses are useful to identify events of Category C; however, our interest is identifying Type C events while studies are ongoing.

Analyses of two or more studies should use valid statistical methods that stratify by study so as to preserve the randomization scheme. When events are so rare that they do not occur in all studies, the statisticians must decide how to include those studies, how to calculate incidence, and how to make comparisons. We recommend using metrics that assess relative differences (e.g., odds ratios, relative risks, or hazard ratios) because absolute rates can mask signals and because they are more vulnerable to violations of the assumption of homogeneity of effect. For example, if a set of trials is comprised of many short-term small trials and a few long-term trials, looking at the absolute rate from the trials with events may identify a signal, but absolute incidence from all trials including short-term trials without any events may attenuate the effect and result in missing the signal. After creating a meta-analysis using relative measures, one can convert the combined result to risk differences to help with clinical interpretability (Berlin et al. 2013).

When clinical trials expose subjects for different durations of treatment, either by design or because subjects withdraw from study before the planned study completion, the denominator to calculate and estimate an adverse event rate is not constant and should be evaluated using appropriate statistical methods for estimating event rates. Possible approaches are exposure-time-adjusted incidence when the event rates are assumed randomly distributed over time and Kaplan–Meier analysis or life tables when they are not. These types of analyses might have to be performed either at the end of the study when the data are unmasked or during the study by an independent Data Monitoring Committee. In assessing whether an observed increase constitutes a trigger for further analysis, a variety of methods may be used to assess how sensitive the conclusions are to different assumptions. Age, gender, dose, baseline disease severity, and concomitant

medications or diseases may be relevant risk factors to assess. Despite the International Conference on Harmonization's recommendations for how to present data in an Integrated Summary of Safety (International Conference on Harmonisation Expert Working Group 2002), we discourage crude pooling methods that simply present numerators, denominators, and percentages in tables as their analysis of data from multiple studies, without showing the data by each study so that appropriate adjustments and weightings might be used to combine rates from different studies (Lievre et al. 2002; Chuang-Stein and Beltangady 2011). Tables with simple aggregate data from multiple clinical trials can lead to incorrect interpretations (Lievre et al. 2002; Chuang-Stein, et al. 2011). Simple unstratified lumping can lead to over- or under-estimates of the effect of treatment, even to complete reversal in the direction of the effect. The problem is a reflection of Simpson's paradox (see Armitage and Colton 1998 for a description of the paradox). We recognize that in spite of its bias, many regulators request this type of analysis.

The most difficult situations are drug-development programs when some of the trials are completed but the ongoing trials are large and long-term. If a potential harm is identified during clinical development, additional actions may be needed to confirm or refute the apparent finding. Actions could include, but are not limited to, using a Standardized MedDRA Query (SMQ) for analysis (or developing a customized MedDRA query), developing a new Case Report Form for current and/or future studies or implementing cross-compound adjudication for the event.

In general, completion of an aggregate safety analysis of one or more randomized clinical trials leads to an estimate of the excess rate at which serious adverse events of interest occur in the drug group. This information can be helpful in interpreting results of ongoing and still masked clinical trials where the events of interest may also be observed but for which the study group assignment has not been determined. Trials with a DMC that understands the Final Rule can report an excess rate of an unexpected serious adverse event to the sponsor's Safety Review Committee, thus triggering an aggregate analysis as described above. When the data reach a clinically and statistically important threshold, the event needs to be reported expeditiously. As described above, the sponsor should report a summary of all the events of the type identified in both the treated and control groups even if the decision to report was based on a subset of the events. The summary report, like the meta-analyses described above, should separate those events that occurred as part of controlled trials from those that occurred in uncontrolled trials or in open-label treatment extensions.

9. The Final Rule and Multiplicity With Particular Reference to Category C Events

In trials of diseases that are not life-threatening in a healthy young population, few serious adverse events of Category C unrelated to the product under study are likely to occur; however, many will occur in trials of life-threatening diseases or in trials of populations with serious or life-threatening comorbidities. In these cases, most serious adverse events are likely to reflect the disease process itself or the general risk in the population under study; therefore, they are expected to occur at equal rates in the study arms. The challenge is to distinguish small imbalances that are due to chance from those due to the investigational product under study. While we focus here on the statistical aspects of this process of noting imbalances, we stress the importance of medical judgment in helping to address causality.

Frequentist-based approaches of flagging adverse events from unadjusted p -values or confidence intervals (CIs) can lead to a high rate of false positives. On the other hand, traditional multiplicity adjustments may lead to a high rate of false negatives, which are of particular concern when evaluating safety. Therefore, it is critical to strike the right balance between no adjustment and over-adjustment.

Statisticians have proposed both frequentist and Bayesian approaches for evaluating accumulating safety data. Bayesian approaches are particularly well-suited to continuous monitoring of the data because Bayesian decisions are made not on the basis of p -values, but rather by summarizing the posterior or predictive distribution (Berry et al. 2004; Xia et al. 2011b), which requires some knowledge of the prior distribution of events. Moreover, Bayesian hierarchical methodology allows modeling the biologic relationship among adverse events, thus obviating the need for considerations of multiplicity.

Emphasis on controlling the Type 1 error rate is often less important for safety than for efficacy analyses. Instead, the Safety Review Committee or DMC might interpret a p -value below 0.1 or 0.2 as sufficient evidence for concern.

If presumptive evidence of a harm emerges (e.g., an increase in heart failure in the treated group), we recommend that the physicians on the Safety Review Committee review the list of serious adverse events that have occurred and create a combined list of similar events as suggested in Section 3.

Our sample decision tree for expedited reporting of Type C events (Table 3) is similar to that of the tree of Type B (Table 2) events except that in the sample tree the threshold for reporting events is a relative risk of 2

compared to control while our sample tree for Type B used a relative risk of 3. The difference is due to the likelihood that Type B events are very rare. If the threshold were 2, a trial with 3 events of a specific type, 2 in one arm and 1, would satisfy the relative risk of 2 that could lead to unnecessary expedited reporting. For Type B events, the clinical picture is typically more relevant than the relative risk. For events of Category C, a doubling of event rate for common events is potentially a large increase in events.

Importantly, these numbers are guidelines for thinking about what to report; decisions must be based on judgment.

10. Identifying Events for Expedited Reporting of Category C Events

Without unmasking the treatment arm and in the absence of an aid of a threshold for reporting, it is difficult to decide whether the frequency of a serious adverse event is unequal in the two (or more) study groups. Some sponsors and investigators have in the past relied on the total number of reports (combined over study groups). If that rate exceeds what would have been anticipated in the population, the data on the specific event may be unmasked. The projection of the expected number of reports could be based on the labels of other products for the same disorder. For malignancies, one could obtain background cancer rates from SEER (Surveillance Epidemiology and End Results). In the latter case, one could use the standardized incidence ratio to define thresholds. Unfortunately, often obtaining a reliable estimate of the expected number of events in the population, especially for novel indications or rare diseases, is difficult. Moreover, the populations in clinical trials are different from the general population because the inclusion and exclusion criteria create specific study populations.

More important, these approaches suffer a fundamental problem: the combined rate might be hiding an important difference in the treated and control group.

Rather than trying to rely on the masked overall rates to identify events of Category C for expedited reporting, many sponsors use either an external independent Data Monitoring Committee (DMC) or an internal Safety Review Committee to review by-group safety data. Most of these committees are convened for individual studies. For a large development program that includes several indications, asking a single committee to review all ongoing studies regularly would be challenging. In some committees, the Sponsor transmits to the DMC Chairperson, or all members of the DMC, a copy of each expedited report sent to the FDA. The Chair then apprises the rest of the committee if a concern arises.

Table 2. Sample decision tree for reporting category B events

If the event is definitely not caused by the investigational product (e.g., it occurred in the control arm or, for example, it was a tendon rupture caused by trauma), do not report it on a 15-day report. Otherwise, continue below.		
Is increase in rate clear?		
<ul style="list-style-type: none"> Does a one-sided 80% confidence interval of the difference between the observed and anticipated include 0? (or, does the one-sided confidence interval of a relative measure include 1?) <ul style="list-style-type: none"> The anticipated rate may come from the rate in the control group. The anticipated rate may come from epidemiological or natural history data. The observed rate may come from a single trial or from a combined analysis of all trials of the product. Is the relative risk less than 3? (For studies without controls, the risk is relative to expectation in a relevant historical population; for studies with controls, the risk is relative to the controls or to the historical population, or both.) Does lumping similar events make the signal disappear? 		
If the answers to all three are “yes,” the data do not show sufficient evidence of imbalance to report. If the answers to all three are “no,” the evidence of causality is clear enough to report the event. Otherwise, the imbalance is unclear.		
No increase in rate DO NOT SEND 15-DAY REPORT	Increase unclear CONSIDER SENDING A 15-DAY REPORT IF AT LEAST ONE OF THESE CONDITIONS HOLD Other safety outcomes (e.g., adverse events and laboratory data) support causality. The mechanism of action supports causality. Investigators need information about this event to manage patients. Information could influence: <ul style="list-style-type: none"> A patient’s willingness to participate in the study. Behavior of patients outside the trial. Pooling similar endpoints strengthens signal. Event is potentially fatal or disabling. Otherwise, do not report but reanalyze as new events occur.	Clear increase SEND A 15-DAY REPORT

NOTE: The choice of 3 for relative risk is arbitrary. In some situations the study team may opt for a different value. Similarly, use of a one-sided 80% confidence interval is a suggestion; some teams would opt not to use confidence intervals at all and some would use 90% or even 95% confidence intervals. The level used for the confidence interval may also depend on the clinical relevance of the event itself.

Table 3. Sample decision tree for expedited reporting of type C events

Analysis of Category C events Based on observing increased frequency relative to control		
Is imbalance clear?		
<ul style="list-style-type: none"> Does a one-sided 80% confidence interval of the difference between observed and control (perhaps using meta-analysis of all related completed and ongoing studies) include 0? (or does the 80% confidence interval for a relative metric include 1?) Is the relative risk compared to control less than 2? (For studies without controls, the risk is relative to expectation in a relevant historical population; for studies with controls, the risk is relative to the controls or to the historical population, or both.) Does lumping similar events make the signal disappear? 		
If the answers to all three are “yes,” the data do not show sufficient evidence of imbalance to file a 15-day report. If the answers to all three are “no,” the evidence of causality is clear enough to file a 15-day report. Otherwise, the increase is unclear.		
No increase DO NOT SEND 15-DAY REPORT	Unclear increase CONSIDER SENDING 15-DAY REPORT IF Other safety outcomes (e.g., adverse events and laboratory data) support causality. OR The mechanism of action supports causality. OR Information could influence: <ul style="list-style-type: none"> The way investigators manage patients A patient’s willingness to participate in the study. Patients outside the trial OR Pooling similar endpoints strengthens signal. OR Event is potentially fatal or disabling. Otherwise, do not report but reanalyze as new events occur.	Clear increase SEND 15-DAY REPORT

NOTE: This is a sample table. For a given drug, disease, or event, the sponsor might select a different confidence interval (i.e., not 80%) or a different relative risk (i.e., not 2).

For multiple ongoing studies or multiple ongoing indications, a standing committee (the Safety Review Committee) with deep knowledge about the compound can serve as a referral body when an important safety question surfaces. This Safety Review Committee can examine all available data (e.g., preclinical studies, completed and ongoing clinical trials, literature, class labels, and post-marketing reports if applicable), consult with available advisory councils, and bring in experts to assess the safety concern, and decide if follow-up actions are needed. Other variations on this theme are possible. For example, some sponsors request the clinical team to examine all the available data and only refer cases to the Safety Review Committee when the team has determined that an imbalance between groups would constitute a reportable event. In this case the Safety Review Committee assesses the imbalance between groups. CIOMS VI (Council for International Organizations of Medical Sciences (CIOMS) Working Group VI 2005) and SPERT (Crowe et al. 2009) provide advice for planning these processes. CIOMS VI discusses safety management teams and aggregate safety reviews. SPERT describes Program Safety Analysis Plans (PSAPs). The program-level reviews can help to identify reportable Category C events.

11. Identifying Events for Expedited Reporting: The Role of the DMC

Data from clinical trials form the basis for aggregate analysis to look for an excess of events that can provide evidence of causation of serious adverse events. This aggregate analysis cannot be informative without consistently collected and coded data.

This problem has challenged data monitoring committees (DMCs) whose mandate is to review accumulating safety data periodically to assess the balance of risk and benefit to recommend whether the trial should continue or be modified. MedDRA-type coding systems are often the main, and sometimes the only, source of safety data presented to the DMC. Unfortunately, DMCs often receive undigested reports consisting of reams of pages. Such reports are not conducive to identifying harms (or anything else!).

The fact that data accumulate over time can make interpretation of data very difficult for the DMC. If DMCs are to contribute meaningfully to implementation of the Final Rule they need data presented in a better way than they usually see and they need to understand the implications of the Final Rule. In large Phase 3 trials, the DMC is often the only group that has routine access to unmasked safety data and they are therefore the only group likely to be able to detect an increase in rate of event that is suggestive of a causal relation between the product under

study and the event in question. Typically, DMCs have not reported observed increases in event rates if the overall balance of benefit to risk is likely to be favorable by the end of the trial unless the committee is nearly certain that the product under study caused the adverse event in question. To be helpful to sponsors in implementing the Final Rule, a DMC needs to report an observed excess of events where the evidence of causality is based on an increase in rates relative to the control but with the recognition that some of these events may turn out not to have been caused by the investigational drug. One approach that a DMC can use is to adopt a guideline similar to that in Table 3 as a trigger for reporting the event to the sponsor's internal Safety Review Committee. That group, having access to all data from the program, can use the DMC's report as a suggestion to review the Category C event in the program at large. If the Safety Review Committee considers the totality of the evidence from all the data of the program to constitute evidence of causality, the event should be reported as 15-day reports. "Day 0" would then be the day that the group judges the event as having been probably caused by the product. Because for Type C events it will be difficult, usually impossible, to attribute a specific event to the product, the 15-day report should be comprised of a summary of the data leading to the judgment of causality.

12. Summary of Recommendations

The Working Group has the following set of recommendations. The recommendations here are slight modifications of the ones listed in Archdeacon et al. (2014). Some of these are included, at least in part, in the December 2012 Guidance (U.S. Department of Health and Human Services and Food and Drug Administration 2012).

- In masked randomized clinical trials, prompt detection of evidence of a causal association between an investigational product and serious adverse events is important. Prompt detection can, however, be difficult if the type of event in question is one that occurs commonly in the study population. Moreover, it may be hard to distinguish between these events and recognized manifestations of the disease under study. Sponsors therefore need a capability for conducting periodic, unmasked, aggregate analyses to detect differential frequency of such events by treatment group. When proper firewalls can be established internally, sponsors should consider convening an internal Safety Review Committee with standard operating procedures to ensure proper implementation. For small companies where a properly firewalled committee is not

possible, the companies should consider convening an external Safety Review Committee comprised of members with skilled knowledge about the compound. Committee members must be willing to devote the time and effort necessary to assess emerging issues with short notices. Companies need to have guidelines or Standard Operating Procedures specifying how the internal or external board communicates interim results and decisions. Whether the committee is internal or external, it should have at most minimal involvement with the clinical development team or those interacting with investigative sites.

- The sponsor may ask this firewalled committee to evaluate unexpected serious adverse events to determine if there is evidence of a causal association with the investigational product. Unmasking will likely be necessary to determine whether these events are subject to an expedited report.
- Sponsors, investigators, IRBs, DMCs, safety assessment committees, and FDA reviewers must be thoroughly educated regarding the Final Rule. In investigator meetings for clinical trials, the sponsor should offer training about collecting safety data as carefully as training about collecting efficacy data.
- The sponsor should develop a plan (e.g., a Program Safety Analysis Plan) that allows incorporation into aggregate analyses the totality of data on the investigational product across its development program or programs, including not only serious adverse events, but also vital signs, laboratory results and other relevant measures.
- When appropriate, meta-analysis of safety data from all completed studies should be performed. Results from such analysis could be of interest in their own right in assessing whether there is evidence that the study drug caused the events or could help interpret observations in ongoing studies. In some cases, the meta-analysis might include unmasked data from ongoing studies. To the extent feasible, these analyses should preserve the randomization of the individual studies and should account for differences in the study designs, the nature of the control groups, and the duration of exposure time. When comparing rates in a drug treatment group to a control group for some serious unexpected adverse events, lumping or combining similar events may help in the evaluation of the strength of a safety signal. In addition, the individual case information must be reviewed as well. These analyses, which are intended to identify reportable serious adverse events, should not

formally correct for multiplicity, nor should a specific p -value be the sole criterion for expedited reporting.

The sponsor, investigators, IRBs, and the FDA should recognize that the approach described in this article is likely to identify some events that turn out not to have been caused by the product in question. Similarly, development programs may not identify some rare events actually caused by the product or some common events with only a small increase in rates among individuals receiving the product. Nonetheless, reporting in an unmasked fashion only those events with evidence of association to the investigational product should increase the specificity of the process of identifying serious unexpected adverse reactions. Moreover, limiting the number of expedited events as described above should focus attention on events that are likely to have been caused by the investigational product, thereby potentially increasing positive predictive value of the reports.

Appendix: Definitions

“Potato, patahto, tomato, tomahto. Let’s call the whole thing off.”—Ira Gershwin

A plethora of words are associated with adverse events. Regulatory language has a few nouns or noun phrases: event, experience, harm, reaction, reasonable possibility; a host of adjectives: adverse, serious, severe, anticipated, unanticipated, expected, unexpected, suspected, fatal, life-threatening, related, not related; and a few adverbs or adverbial phrases: immediately, promptly, within 7 days, and within 15 days. This section attempts to define these words as the FDA uses them; the remainder of this article adopts the definitions in this section. Especially confusing to many investigators is the distinction between “severe” (a measure of intensity) and “serious” (a regulatory assessment of medical consequence) and the distinction between “unanticipated” and “unexpected”. We urge you, the reader, to relinquish your ordinary understanding of these words and use the definitions below when reading this article.

Adverse event (AE): “any untoward medical occurrence associated with the use of a drug in humans, whether or not considered drug related” [21 CFR 312.32 (a)]. Older versions of the CFR and some FDA guidances (e.g., U.S. Department of Health and Human Services and Food and Drug Administration 2009a) use the terms *adverse event*, *adverse effect*, and *adverse experience* as synonyms. The Guidance describing the Final Rule does note that an adverse event is “also referred to as an adverse experience” (U.S. Department of Health and Human Services and Food and Drug Administration 2005, p. 3).

Adverse reaction: Any adverse event caused by a drug [21 CFR 312.32(a)].

Suspected adverse reaction: “[A]ny adverse event for which there is a reasonable possibility that a drug caused the adverse event.” For the purposes of IND safety reporting, “reasonable possibility” means there is evidence to suggest a causal relationship between the drug and the adverse event. *Suspected* adverse reaction implies a lesser degree of certainty about causality than an *adverse reaction*” [21 CFR 312.32(a)]. (Lest readers think they clearly understand the above definitions, note that European Union states that *adverse reaction* and *suspected adverse reaction* are synonymous (European Commission and European Medicines Agency 2008). This article uses the FDA definitions above, which do not consider the terms synonymous.)

Reasonable possibility: There is evidence to suggest a causal relationship between the drug and the adverse event [21 CFR 312.32(a)]. The difference between an adverse event and a suspected adverse reaction depends on this somewhat subjective concept of *reasonable possibility*. The protocols for many trials define a taxonomy of relatedness that aims to capture both biology and timing. Investigators are typically asked to describe the relationship of a particular event to the product under study with terms like definite, probable, possible, unlikely, or not related; the definitions of those adjectives depend on the study.

Prior to the Final Rule, sponsors would report in an expedited fashion serious unexpected events that the investigator classified as “definitely,” “probably,” and in many cases “possibly” related to any study medication. Under the Final Rule, the sponsor should consider the investigator’s assessment of causality, but use its own judgment about causality to decide whether to submit an expedited report of a serious unexpected event. For example, in a masked trial, suppose an investigator classifies an event as “definitely related.” The sponsor should not report this event in an expedited fashion if it learned, after unmasking the treatment allocation, that the participant had not received the test drug (e.g., the participant had been on the placebo arm or, although randomized to tested drug, had never received it). The sponsor would also not report the event as a suspected adverse reaction if, for example, the event was common in the population and there was no particular reason to consider it causally related to the use of the test drug.

Severity: Severity is a measure of intensity. Different protocols may define severity differently but a typical approach to categorization is use of a graded set of terms such as mild, moderate, and severe. Many, but not all, serious adverse events are “severe”; many severe adverse events (e.g., severe headache) do not satisfy the regulatory definition of “serious.”

Serious adverse event or serious adverse reaction: “Seriousness” has a formal regulatory definition. An adverse event or a suspected adverse reaction is considered *serious* if, in the view of either the investigator or sponsor, it results in any of the following outcomes: death, a life-threatening adverse experience, an inpatient hospitalization or a prolongation of an existing hospitalization, a persistent or significant incapacity or substantial disruption of the ability to conduct normal life functions, or a congenital anomaly or birth defect in an offspring of the person taking the product. The definition further states, “Important medical events that may not result in death, be life-threatening, or require hospitalization may be considered serious when, based on appropriate medical judgment, they may jeopardize the patient or subject and may require medical or surgical intervention to prevent one of the outcomes listed in this definition” [21 CFR 312.32(a)]. Examples of such events include allergic bronchospasm requiring intensive treatment in an emergency room or at home; blood dyscrasias or convulsions that do not result in inpatient hospitalization; the development of drug dependency or drug abuse; or cancers, but not basal cell carcinoma.

Unexpected adverse event or unexpected suspected adverse reaction: An adverse event or suspected adverse reaction is considered “unexpected” if it is not listed in the investigator brochure or, if a similar event is in the brochure, it is not listed at the observed specificity or severity. The FDA Guidance states (U.S. Department of Health and Human Services and Food and Drug Administration 2012, p. 15), “The investigator brochure should not list adverse events that are unlikely to have been caused by the drug because such lists could dilute the importance of clinically meaningful risk information.”

The Regulations say, “For example, under this definition, hepatic necrosis would be unexpected, by virtue of its greater severity, if the investigator brochure referred only to ‘elevated hepatic enzymes’ or ‘hepatitis.’ Similarly, cerebral thromboembolism and cerebral vasculitis would be unexpected, by virtue of greater specificity, if the investigator brochure listed only ‘cerebral vascular accidents’” [21 CFR 312.32(a)]. *Unexpected*, as used in this definition, also refers to adverse events or suspected adverse reactions that the investigator brochure mentions as occurring with a class of drugs or is consistent with the pharmacological properties of the drug but is not specifically mentioned as occurring with the particular drug under investigation [21 CFR 312.32(a)].

For a product that is not yet approved, if an investigator brochure is not available, *unexpected* means the event or reaction is not consistent with the information concerning risk described in the general investigational plan or elsewhere in the current documents describing the studies investigating the product (U.S. Department of Health

and Human Services and Food and Drug Administration 2005, p. 5).

Two other adjectives used to describe adverse events are “anticipated” and “unanticipated.” In the context of the Final Rule, FDA describes anticipated events to include known consequences of the underlying disease or condition under investigation (e.g., symptoms, disease progression, comorbidities) as well as events unlikely to be related to the underlying disease or condition under investigation but common in the study population independent of drug therapy (e.g., cardiovascular events in an elderly population). Still other anticipated events are those that are consistent with the pharmacological properties of the drug, even though they have not been observed with the drug under investigation (U.S. Department of Health and Human Services and Food and Drug Administration 2005, p. 6). The Final Rule requires monitoring such anticipated serious adverse events at appropriate intervals, generally by comparing the rate of events in the drug-treated and control arms in a trial.

The sponsor must report an unexpected serious adverse event to the FDA expeditiously as an IND safety report if the individual event is clearly caused by the product under study. As the rule makes clear, that is possible when the serious event rarely occurs spontaneously and is a known consequence of drug therapy. Where the event is anticipated (i.e., known to occur in the population), its occurrence would be plausibly considered an unexpected adverse event, and thus reportable as a suspected adverse reaction, if it occurs at a higher rate in the treated than in the control group. How much higher leads to the difficult question of defining reporting thresholds. Judging an anticipated serious event to be unexpected and reportable would require the sponsor to review the completed and ongoing studies in an IND to allow an assessment of whether an aggregate analysis of these studies shows that the likelihood is reasonably high that the drug caused the adverse event [21 CFR 312.32(c)(1) (i)(C)]. We, the authors of this article, recommend that making such a judgment on the basis of an observed imbalance between arms that is large enough to suggest a reasonable possibility that the drug caused the adverse event.

In contrast to FDA, which assigns the final reporting decision to the sponsor alone, the European Medicines Agency (EMA) requires expedited reporting on the basis of the investigator's or sponsor's judgment of relatedness (i.e., if either the sponsor or the investigator thinks the event is related, reporting is required). The FDA does not use the term “adverse effect” but the EMA uses “adverse effect” synonymously with “adverse reaction.”

Another set of definitions deals with timing of reports.

Seven-day reports. The sponsor must report any unexpected fatal or life-threatening suspected adverse reaction to the FDA within 7 calendar days after originally receiving the information [21 CFR 312.32(c)(2)]. (Note that

a previous version of the safety reporting rule specified time in weeks or business days.)

Fifteen-day reports. The sponsor must send a safety report for a serious, unexpected, suspected adverse reaction to the FDA within 15 calendar days of determining “that the suspected adverse reaction... qualifies for reporting” [21 CFR 312.32(c)(1)].

This article deals with the 7- and 15-day reports, both of which are called “expedited.”

The FDA Guidance on implementing the Final Rule states,

“The sponsor must report in an IND safety report any suspected adverse reaction to study treatment (i.e., including active comparators) that is both serious and unexpected [21 CFR 312.32(c)(1)(i)]. Before submitting an IND safety report, the sponsor needs to ensure that the event meets all three of the definitions:

- Suspected adverse reaction
- Serious
- Unexpected

If the adverse event does not meet all three of the definitions, it should not be submitted as an IND safety report” (U.S. Department of Health and Human Services and Food and Drug Administration 2012, p. 8–9).

Acknowledgments

The authors are grateful to Ms. Jennifer Fletcher and Ms. Sunjana Supekar for their tireless efforts in helping us interpret the CFR and the relevant FDA guidances as well as for their editorial support. The authors are grateful also to Cheryl Grandinetti and Patrick Archdeacon, both of the U.S. FDA, who helped us navigate through the regulations and guidances. The authors thank Aloka Chakravarty, David DeMets, Kerry Lee, Devan Mehrotra, Robert O'Neill, and Steven Snapinn, members of the Workgroup, for their contributions to the discussions leading to early versions of the article as well as Cheri Janning for her encouragement and support throughout the process of writing this article. In-kind contributions of effort by authors and other member of the Workgroup supported this article. In addition, the Clinical Trials Transformation Initiative (www.ctti-clinicaltrials.org) provided some staff effort and meeting support derived both from pooled fees from member organizations and from a cooperative agreement, U19FD003800, awarded to Duke University by the U.S. Food and Drug Administration. The views expressed herein represent those of the authors and do not necessarily represent the views or practices of the authors' employers or of any other party.

[Received August 2014. Revised February 2015]

References

- Archdeacon, P., Grandinetti, C., Vega, J. M., Balderson, D., and Kramer, J. M. (2014), “Optimizing Expedited Safety Reporting for Drugs and Biologics Subject to an Investigational New Drug Application,” *Therapeutic Innovation and Regulatory Science*, 48, 200–207. [175,178,186]

- Armitage, P., and Colton, T. (eds.) (1998), *Encyclopedia of Biostatistics*, West Sussex, UK: Wiley. [183]
- Berlin, J. A., Crowe, B. J., Whalen, E., Xia, H. A., Koro, C. E., Kuebler, J. (2013), "Meta-analysis of Clinical Trial Safety Data in a Drug Development Program: Answers to Frequently Asked Questions," *Clinical Trials*, 10, 20–31. [183]
- Berry, S. M., and Berry, D. A. (2004), "Accounting for Multiplicities in Assessing Drug Safety: A Three-Level Hierarchical Mixture Model," *Biometrics*, 60, 418–426. [182,184]
- Chuang-Stein, C., and Beltangady, M. (2011), "Reporting Cumulative Proportion of Subjects With an Adverse Event Based on Data From Multiple Studies," *Pharmaceutical Statistics*, 10, 3–7. [183]
- Chuang-Stein, C., and Xia, H. A. (2013), "The Practice of Pre-Marketing Safety Assessment in Drug Development," *Journal of Biopharmaceutical Statistics*, 23, 3–25. [182]
- Council for International Organizations of Medical Sciences (CIOMS) Working Group VI (2005), *Management of Safety Information From Clinical Trials*, Geneva: Council for International Organizations of Medical Sciences (CIOMS) Working Group VI. [180,182,186]
- Crowe, B., Brueckner, A., Beasley, C., and Kulkarni, P. (2013), "Current Practices, Challenges, and Statistical Issues With Product Safety Labeling," *Statistics in Biopharmaceutical Research*, 5, 180–193. [178]
- Crowe, B., Xia, A., Nilsson, M., Shahin, S., Wang, W., and Jiang, Q. (2015), "The Program Safety Analysis Plan: An Implementation Guide," in *Quantitative Evaluation of Safety in Drug Development: Design, Analysis, and Reporting*, eds. Q. Jiang and A. Xia, London: Chapman and Hall, pp. 55–68. [182]
- Crowe, B. J., Xia, H. A., Berlin, J. A., Watson, D. J., Shi, H., Lin, S. L., Kuebler, J., Schriver, R. C., Santanello, N. C., Rochester, G., Porter, J. B., Oster, M., Mehrotra, D. V., Li, Z., King, E. C., Harpur, E. S., and Hall, D. B. (2009), "Recommendations for Safety Planning, Data Collection, Evaluation and Reporting During Drug, Biologic and Vaccine Development: A Report of the Safety Planning, Evaluation, and Reporting Team," *Clinical Trials*, 6, 430–440. [180,182,186]
- European Commission and European Medicines Agency (2008), *Volume 9a of the Rules Governing Medicinal Products in the European Union – Guidelines on Pharmacovigilance for Medicinal Products for Human Use*. [188]
- International Conference on Harmonisation Expert Working Group (2002), *The Common Technical Document for the Registration of Pharmaceuticals for Human Use: Efficacy M4e(R1)*. [183]
- Lachenbruch, P. A., and Wittes, J. (2007), "Sentinel Event Methods for Monitoring Unanticipated Adverse Events," in *Advances in Statistical Methods for the Health Sciences*, eds. J. Auget, N. Balakrishnan, M. Mesbah, and G. Molenberghs, Boston, MA: Birkhauser, pp. 61–74. [182]
- Lievre, M., Cucherat, M., and Leizorovicz, A. (2002), "Pooling, Meta-Analysis, and the Evaluation of Drug Safety," *Current Controlled Trials in Cardiovascular Medicine*, 3, 6–10. [183]
- Littlejohn, J. K., Lucas, D. O., Batson-Fowler, G., and Edwards, S. (1991), "Adverse Experience Collection: Perspective from a Biological Development Program," *Drug Information Journal*, 25, 175–180. [178]
- U.S. Department of Health and Human Services and Food and Drug Administration. (2005), *Guidance for Industry: Good Pharmacovigilance Practices and Pharmacovigilance Assessment*. [187,189]
- (2009a), *Guidance for Clinical Investigators, Sponsors, and IRBs: Adverse Event Reporting to IRBs — Improving Human Subject Protection*. [187]
- (2009b), *Guidance for Industry Drug-Induced Liver Injury: Premarketing Clinical Evaluation*. [176]
- (2011), *Code of Federal Regulations Title 21 Food and Drugs Chapter I Food and Drug Administration Department of Health and Human Services Subchapter D Drugs for Human Use Part 312 Investigational New Drug Application*. [174,175]
- (2012), *Guidance for Industry and Investigators: Safety Reporting Requirements for INDs (Investigational New Drug Applications) and BA/BE (Bioavailability/Bioequivalence) Studies*, Rockville, MD: Food and Drug Administration. [175,176,186,188,189]
- Xia, H. A., Crowe, B. J., Schriver, R. C., Oster, M., and Hall, D. B. (2011a), "Planning and Core Analyses for Periodic Aggregate Safety Data Reviews," *Clinical Trials*, 8, 175–182. [182]
- Xia, H. A., and Jiang, Q. (2014), "Statistical Evaluation of Drug Safety Data," *Therapeutic Innovation and Regulatory Science*, 48, 109–120. [182]
- Xia, H. A., Ma, H., and Carlin, B. P. (2011b), "Bayesian Hierarchical Modeling for Detecting Safety Signals in Clinical Trials," *Journal of Biopharmaceutical Statistics*, 21, 1006–1029. [182,184]

© J. Wittes, B. Crowe, C. Chuang-Stein, A. Guettner, D. Hall, Q. Jiang, D. Odenheimer, H. Amy Xia, J. Kramer

This is an Open Access article. Noncommercial re-use, distribution, and reproduction in any medium, provided the original work is properly attributed, cited, and is not altered, transformed, or built upon in any way, is permitted. The moral rights of the named author(s) have been asserted.

About the Authors

Janet Wittes, Ph.D., Statistics Collaborative, Inc., 1625 Massachusetts Ave., NW, Suite 600, Washington DC 20036 (E-mail: janet@statcollab.com). Brenda Crowe, Eli Lilly and Company, Indianapolis, IN (E-mail: crowe_brenda.j@lilly.com). Christy Chuang-Stein, Pfizer Inc., Kalamazoo, MI (E-mail: christy.j.chuang-stein@pfizer.com). Achim Guettner, Novartis Pharma AG, Basel, Switzerland (E-mail: achim.guettner@novartis.com). David Hall, Boehringer-Ingelheim, Danbury, CT (E-mail: david.hall@boehringer-ingelheim.com). Qi Jiang (E-mail: qjiang@amgen.com) and H. Amy Xia (E-mail: hxia@amgen.com), Amgen, Thousand Oaks, CA. Daniel Odenheimer, Drug Development Consultant, Potomac, MD (E-mail: dan.odenheimer@gmail.com). Judith Kramer, Duke University, Durham, NC (E-mail: judith.kramer@duke.edu).