Chapter 12
Assessment and Reporting of Harm

Thomas J. Moore

Assessment of harm is more complex than the assessment of benefit of an intervention. The measures of favorable effects are or should be prespecified in the protocol and they are limited in number. In contrast, the number of adverse events is typically very large and they are rarely prespecified in the protocol. Some may not even be known at the time of trial initiation. These facts introduce analytic challenges.

Most intervention effects have three dimensions. The major objective of any trial is to measure change in the incidence or rate of a clinical event, symptom, laboratory test, or other measure. For benefit, the hoped-for difference in rate between the two study groups is prespecified and forms the basis for the sample size calculation. Such prespecifications rarely exist for the adverse events and clinical trials are often underpowered statistically for the documentation or dismissal of evidence of harm. There are two other dimensions of interest—severity of the event and recurrence or duration of its occurrence. In terms of severity, a clinical event can be uncomplicated or complicated, including being fatal, while a symptom can vary from mild to severe. There are few good objective scales for quantifying symptoms; their severity is based on the participants' perceptions. In arthritis trials, a pain scale is commonly used to determine treatment benefit although it could also be used to determine adverse effects. Their recurrence can vary substantially from occasional to being constant. As a result of these methodological limitations the reporting of harm is often limited to whether adverse events occurred or not; rarely are severity and recurrence reported.

Contributing to the complexity of assessment and reporting of harm is a common confusion about the terminology. An *adverse event* is "any untoward event that occurs during a drug or medical treatment whether or not a causal relationship with the treatment is suspected or proven" [1]. Thus, adverse events might be experienced by treated as well as untreated patients. The incidence of adverse events is assessed in and reported for both study groups. One objective of trials is to compare

T.J. Moore
Senior scientist, Institute for Safe Medication Practices

the adverse event experiences in participants receiving active intervention or control. An *adverse effect* has been described as "a noxious or unintended response to a medical product in which a causal relationship is at least a reasonable possibility" [1]. In this text we will use these definitions of adverse events and adverse effects, except that they are broadened to include not just medical treatment, but any intervention.

Harm is the sum of all adverse effects and is used to determine the benefit-harm balance of an intervention. *Risk* is the probability of developing an adverse effect. *Severe* is a measure of intensity. *Serious* is an assessment of the medical consequence (see below). *Expected* adverse events or effects are those that are anticipated based on prior knowledge. *Unexpected* are findings not previously identified in nature, severity, or degree in incidence.

Fundamental Point

Careful attention needs to be paid to the assessment, analysis, and reporting of adverse effects to permit valid assessment of harm from interventions.

Assessment of Harm

There are three categories of adverse events—serious adverse events, general adverse events and adverse events of special interest. Serious adverse events are defined by the U.S. Food and Drug Administration (FDA) as those events that are (a) life threatening (b) result in initial or prolonged hospitalization, (c) cause irreversible, persistent or significant disability/incapacity, (d) are a congenital anomaly/birth defect, (e) require intervention to prevent harm, or (f) have other medically serious consequences [2]. General adverse events are those which patients or trial participants have complained about or clinicians have observed. These may range in intensity from very mild and not of much consequence to severe. Adverse events of special interest are typically derived from studies of mechanisms of action of the intervention (for example immunosuppression), animal studies, or observations from chemically similar drugs or related interventions. Assessment of adverse events of special interest requires prospective definition, specific ascertainment, and plans for reporting. Another area of importance is the evaluation of adverse drug interactions.

Strengths

There are four distinct advantages to assessment of harm in clinical trials, as opposed to other kinds of clinical research. First, adverse events can be defined prospectively, which allows proper hypothesis testing and adds substantial

credibility. Post hoc observations, common in the area of harm, are often difficult to interpret in terms of causation and therefore often lead to controversy.

Second, randomized clinical trials by definition have a proper and balanced control group which allows for comparisons between the study groups. Randomization assures that intervention and control groups have similar characteristics—even those unknown to science at the time the trial was conceived. Other study designs have a dilemma when comparing users of a particular intervention to non-users. In observational studies, there is no guarantee that the user and non-user groups are comparable. There are clinical reasons why some people are prescribed a particular intervention while others are not. Observed group differences can be intervention-induced, due to differences in the composition and characteristics of the groups, or a combination thereof. Statistical adjustments can help but will never be able to control fully for unmeasured differences between users and non-users.

Third, clinical trials with a blinded design reduce potential biases in the collection, assessment and reporting of data on harm (Chap. 7).

Fourth, participants in clinical trials are closely and systematically assessed, including physical examinations, regular blood work, weekly or monthly clinic visits, vital signs, clinical events, and detailed assessment of concomitant medications.

Limitations

There are also four potential limitations in relying on clinical trials for evaluation of harm. First, the trial participants are a selected non-random sample of people with a given condition who volunteered for the trial. The selectivity is defined by the scope of the trial inclusion and exclusion criteria and the effects of enrolling only volunteers. In general, trial participants are healthier than non-participants with the same disease. In addition, certain population groups may be excluded, for example, women who are pregnant or breastfeeding. Trials conducted prior to regulatory agency approval of a product are typically designed to document clear findings of benefit and, therefore, often exclude from participation those who are old, have complicating medical conditions and/or are taking other medications which may affect the outcome. Trial sponsors also exclude participants at higher risk of suffering an adverse event. This reduces the incidence of such events and contributes to the likelihood of not documenting harm. The absence of serious adverse effects observed in low-risk participants in pre-approval trials is no assurance that a drug lacks harmful effects when it reaches the marketplace. Another limitation is that the ascertainment of adverse events often relies on volunteered information by the participant rather than specific, solicited information (see below). An early survey showed that most FDA-approved drugs have one serious adverse effect detected after approval when there is more exposure to higher-risk patients and longer treatment duration [3]. More recent high-profile cases of serious adverse effects not detected pre-approval are the treatments of osteoarthritis with

COX-2 inhibitors [4–7], of type II diabetes with rosiglitazone [8–10], and prevention of thromboembolic events with oral anticoagulants [11–13]. The reported high rates of new Boxed Warnings and drug withdrawals over the past two decades illustrate a limitation of FDA's current process for documenting real and potential harm pre-approval [14].

A second limitation relates to the statistical power of finding a harm, if it exists. Small sample sizes and short trial durations, as well as the focus on low-risk populations, reduce the likelihood of detecting serious adverse effects. Drug manufacturers often conduct a large number of small, short-term trials, and their large trials are often not of long duration. Due to limited statistical power, clinical trials are often unreliable for attributing causality to *rare* serious adverse events. Approximately 3,000 participants are required to detect a single case with 95% probability, if the true incidence is one in 1,000; a total of 6,500 participants are needed to detect three cases [15]. When a new drug is approved for marketing, approximately 500–2,000 participants have typically been exposed to it in both controlled and uncontrolled settings. More commonly, rare serious adverse effects are initially discovered through case reports, other observational studies or reports of adverse events filed with regulatory agencies after approval [16, 17]. However, clinical trials can detect precursors of serious adverse effects through measurements such as elevated ALT levels (acute liver failure) or prolonged QT interval on the electrocardiogram (sudden cardiac death). Vandenbroucke and Psaty [18] properly concluded that "the benefit side [of drugs] rests on data from randomized trials and the harms side on a mixture of randomized trials and observational evidence, often mainly the latter."

Third, inability to detect *late* serious adverse effects is another potential limitation of clinical trials. When a new compound is introduced for long-term treatment of a non-life threatening disease, the minimum regulatory standard is only several hundred participants exposed for 1 year or longer [19]. This is obviously inadequate for evaluation of drugs intended for chronic or long-term use. Moreover, a long lag time to harm must be considered for drugs that may be carcinogenic or have adverse metabolic effects. For example, the lag time for carcinogens to cause cancer may often be longer than most long-term trials. We support the view that evaluation of harm should continue the entire time a drug intended for chronic use is on the market [20].

Fourth, the investigators or sponsors may be unaware of some adverse effects because they are unexpected, or, in the case of known adverse effects, not ascertained. Potentially lethal cardiac rhythm disturbances may not be identified because electrocardiographic studies are not performed. Diabetes risk may be overlooked because laboratory testing does not include periodic assessment of Hb_{A1c}. Adverse effects related to sexual function or suicidal ideation may be underestimated because participants rarely volunteer information about sexual problems or suicidal ideation in response to general questions about changes in their health status. Ascertaining withdrawal and rebound effects require a special protocol to monitor discontinuation symptoms. Drug interactions may be overlooked because of rigid exclusion criteria in the protocol and failure to analyze

concomitant medication data in relation to adverse events. Additionally, it is very challenging to be rigorous in these analyses.

The methods for collecting information on harm should take advantage of the strengths of clinical trials and to supplement them with properly designed and conducted observational studies post-trial, especially if issues or signals of harm emerge. Establishment of such long-term safety registries as one tool for post-marketing surveillance is becoming more common [21].

Identification of Harm in Clinical Trials

As pointed out earlier in this chapter, randomized clinical trials are not optimal for the detection of rare, late and unexpected serious adverse events. Experience has shown that critical information on serious reactions comes from multiple sources.

The role of clinical trials in identifying serious adverse reactions was investigated in an early study by Venning [16], who reviewed the identification and report of 18 adverse reactions in a variety of drugs. Clinical trials played a key role in identifying only three of the 18 adverse effects discussed. Another comparison of evidence of harm of various interventions in 15 large randomized and observational studies showed that the non-randomized studies often were more likely to find adverse effects [22].

A clinical trial may, however, suggest that further research on adverse reactions would be worthwhile. As a result of implications from the Multiple Risk Factor Intervention Trial [23] that high doses of thiazide diuretics might increase the incidence of sudden cardiac death, Siscovick and colleagues conducted a population-based case-control study [24]. This study confirmed that high doses of thiazide diuretics, as opposed to low doses, were associated with a higher rate of cardiac arrest.

Drugs of the same class generally are expected to have a similar effect on the primary clinical outcome of interest. However, they may differ in degree if not in kind of adverse effects. One illustration is cerivastatin which was much more likely to cause rhabdomyolysis than the other marketed statins [25]. Longer acting preparations, or preparations that are absorbed or metabolized differently, may be administered in different doses and have greater or lesser adverse effects. It cannot be assumed in the absence of appropriate comparisons that the adverse effects from similar drugs are or are not alike. As noted, however, a clinical trial may not be the best vehicle for detecting these differences, unless it is sufficiently large and of long duration.

Genomic biomarkers have assumed an increasing and important role in identifying people at an increased risk of adverse effects from medications. A large number of FDA approved drugs have pharmacogenomics information in different sections of the labeling [26]. Thus, adverse drugs effects observed in genetically defined subgroups of people are reflected in label additions of Boxed Warnings, Contraindications, Warnings, Precautions and Drug Interactions.

Classification of Adverse Events

Since the late 1990s, adverse drug events in clinical trials and many other clinical studies around the world are classified and described with a common terminology, the Medical Dictionary for Regulatory Activities (MedDRA) [27]. It was established by the International Conference on Harmonisation, a global organization created by the pharmaceutical industry to coordinate requirements among the world's regulatory agencies.

The structure and characteristics of the MedDRA terminology have an effect on how adverse events are collected, coded, assessed, and reported in clinical trials. The most important feature is its pyramidal, hierarchical structure with highly granular terms at the bottom and 26 System Organ Classes at the top. The structure is shown in Table 12.1.

The number of Low Level Terms is very large and intended to facilitate human MedDRA coders using auto-coding software in assigning MedDRA terms to adverse event narratives by including a large number of phrases that might appear. These terms are aggregated at the Preferred Term level, the most granular level normally used in study reports. A key feature of Preferred Terms is that they do not necessarily describe an adverse event. A term could be a sign, symptom, diagnosis, surgical treatment, outcome (such as death), or person characteristic (such as bed sharing, aged parent, or surrogate mother). They are often coded based on a participant complaint noted in the medical record or data collection form.

The terminology designers sought to overcome some of the limitations of the hierarchical structure by allowing links across categories (a multi-axial structure) and the creation of Standardized MedDRA Queries (SMQs). For example an "Air embolism" has a primary link to the Vascular Disorder System Organ Class and a secondary link to Injury and Poisoning. SMQs, on the other hand, are designed to capture specifically adverse events independent of the hierarchical structure. Version 16.1 of MedDRA included 211 SMQs organized on four hierarchical levels [28].

The methodological strengths of the MedDRA terminology include the following: It is an accepted global standard with multiple language translations, which facilitates comparisons among trials. As a granular terminology it provides for detailed and accurate coding of narratives without requiring complex medical judgment in each case. The hierarchical structure and SMQs provide alternative tools for identifying adverse events.

Table 12.1 MedDRA terminology hierarchy[a]

Term	Abbrev.	Number of terms
System Organ Class	SOC	26
High Level Group Term	HLGT	334
High Level Term	HLT	1,717
Preferred Term	PT	20,307
Low Level Term	LLT	72,072

[a]MedDRA version 16.1

While the MedDRA terminology design provides for simple and accurate coding of narratives, the terms thus selected do not necessarily describe adverse events. If analysis is limited to the approximately 20,000 Preferred Terms, the result is so granular that the number of participants for each listed event often becomes too few to evaluate meaningfully. See Table 12.2 for the large number of synonyms for depression. The SMQs in particular vary widely in design, specificity, sensitivity and other features and need to be assessed specifically in each case. By design, the terminology is continuously revised, with new versions appearing twice a year. This complicates replication of previous study results and comparisons among studies, and may even require special procedures to update an ongoing clinical trial that lasts for longer than 6 months. A participant may express the same complaint differently on two clinical visits. As a result, they are likely to be recorded differently, and thus coded differently, which makes it impossible to track a particular adverse event across visits.

Data monitoring based on the MedDRA terminology has turned out to be a challenge. The small numbers of events for each term due to the granular terminology are very difficult to interpret, and the aggregation of individual granular terms into a category with more events requires judgment in order to be clinically meaningful.

Table 12.2 MedDRA preferred terms describing depression in a clinical trial

Agitated depression
Anhedonia
Childhood depression
Decreased interest
Depressed mood
Depression
Depression postoperative
Depression suicidal
Depressive symptom
Dysthymic disorder
Feeling guilty
Feeling of despair
Feelings of worthlessness
Major depression
Menopausal depression
Morose
Negative thoughts
Post stroke depression
Postictal depression
Postpartum depression
Psychomotor retardation
Tearfulness

Source: MedDRA version 16.1

The National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events v3.0 is another advanced system for reporting adverse events [29]. One strength is the 5-step severity scale for each adverse event ranging from mild to any fatal adverse event. It is available without charge.

Ascertainment

The issue often arises whether one should elicit adverse events by means of a checklist or rely on the participant to volunteer complaints. *Eliciting* adverse events has the advantage of allowing a standard way of obtaining information on a preselected list of symptoms. Thus, both within and among trials, the same series of events can be ascertained in the same way, with assurance that a "yes" or "no" answer will be present for each. This presupposes, of course, adequate training in the administration of the questions. *Volunteered* responses to a question such as "Have you had any health problems since your last visit?" have the possible advantage of tending to yield only the more serious episodes, while others are likely to be ignored or forgotten. In addition, only volunteered responses will give information on truly unexpected adverse events.

The difference in the yield between elicited and volunteered ascertainment has been investigated. In the Aspirin Myocardial Infarction Study [30] investigators first asked a general question about adverse events, followed by questions about specific complaints. The results for three adverse events are presented in Table 12.3. Two points might be noted. First, for each adverse event, eliciting gave a higher percent of participants with complaints in both intervention and placebo groups than did asking for volunteered problems. Second, similar aspirin-placebo differences were noted, regardless of the method. Thus, in this case the investigators could detect the adverse effect with both techniques. Volunteered events may be of greater severity but fewer in number, reducing the statistical power of the comparison.

Table 12.3 Percent of participants ever reporting (volunteered and solicited) selected adverse events, by study group, in the Aspirin Myocardial Infarction Study

	Hematemesis	Tarry stools	Bloody stools
Volunteered			
Aspirin	0.27	1.34[a]	1.29[b]
Placebo	0.09	0.67	0.45
Elicited			
Aspirin	0.62	2.81[a]	4.86[a]
Placebo	0.27	1.74	2.99

[a]Aspirin-placebo difference >2 S.E
[b]Aspirin-placebo difference >3 S.E
Aspirin group: $N = 2,267$
Placebo group: $N = 2,257$

Spontaneously volunteered events also may substantially undercount some types of adverse effects, notably psychiatric symptoms. For example, when specifically queried using the Arizona Sexual Experiences Scale, 46.5% reported sexual dysfunction in one study, compared to 1–2% as spontaneously ascertained in clinical trials of fluoxetine [31]. Spontaneous reports could also underestimate new onset diabetes occurring in an unrelated treatment, as well as effects that are not typically characterized medically, such as falls, anger, or tremor.

Prespecified Adverse Events

The rationale for defining adverse events in the protocol is similar to that for defining any important benefit variable; it enables investigators to record something in a consistent manner. Further, it allows someone reviewing a trial to assess it more accurately, and possibly to compare the results with those of other trials of similar interventions.

Because adverse events are typically viewed as secondary or tertiary response variables, they are not often systematically and prospectively evaluated and given the same degree of attention as the primary and secondary benefit endpoints. They usually are not defined, except by the way investigators apply them in their daily practice. A useful source is the Investigator's Brochure for the study drug. The diagnosis of acute myocardial infarction may be based on non-standardized hospital records. Depression may rely on a patient-reported symptom of non-specified severity and duration rather than a careful evaluation by a psychiatrist or responses to a standardized depression questionnaire. Thus, study protocols seldom contain written definitions of adverse events, except for those that are recognized clinical conditions. Multicenter trials open the door to even greater levels of variability in event definitions. In those cases, an adverse event may be simply what each investigator declares it to be. Thus, intrastudy consistency may be as poor as interstudy consistency.

However, given the large number of possible adverse events, it is not feasible to define all of them in advance and, in addition, many do not lend themselves to satisfactory definition. Some adverse events cannot be defined because they are not listed in advance, but are spontaneously mentioned by the participants. Though it is not always easy, important adverse events which are associated with individual signs or laboratory findings, or a constellation of related signs, symptoms, and laboratory results can and should be well-defined. These include the events known to be associated with the intervention and which are clinically important, i.e. adverse events of special interest. Other adverse events that are purely based on a participant's report of symptoms may be important, but are more difficult to define. These may include nausea, fatigue, or headache. Changes in the degree of severity of any symptom should be part of the definition of an adverse event. The methods by which adverse events were ascertained should be stated in any trial publication.

Characteristics of Adverse Events

The simplest way of recording presence of an adverse event is with a yes/no answer. This information is likely to be adequate if the adverse event is a serious clinical event such as a stroke, a hospitalization or a significant laboratory abnormality. However, symptoms have other important dimensions such as severity, duration and frequency of recurrence.

The severity of subjective symptoms is typically rated as mild, moderate or severe. However, the clinical relevance of this rating is unclear. Participants have different thresholds for perceiving and reporting their symptoms. In addition, staff's recorded rating of the reported symptom may also vary. One way of dealing with this dilemma is to consider the number of participants who were taken off the study medication due to an adverse event, the number who had their dose of the study medication reduced and those who continued treatment according to protocol in spite of a reported adverse event. This classification of severity makes clinical sense and is generally accepted. A challenge may be to decide how to classify participants who temporarily are withdrawn from study medication or have their doses temporarily reduced.

The duration or frequency with which a particular adverse event occurs in a participant can be viewed as another measure of severity. For example, episodes of nausea sustained for weeks rather than occasionally is a greater safety concern. Investigators should plan in advance how to assess and present all severity results.

Length of Follow-up

The duration of a trial has a substantial impact on adverse event assessment. The longer the trial, the more opportunity one has to discover adverse events, especially those with low frequency. Also, the cumulative number of participants in the intervention group complaining will increase, giving a better estimate of the incidence of the adverse event. Of course, eventually, most participants will report some general complaint, such as headache or fatigue. However, this will occur in the control group as well. Therefore, if a trial lasts for several years, and an adverse event is analyzed simply on the basis of cumulative number of participants suffering from it, the results may not be very informative, unless controlled for severity and recurrences. For example, the incidence could be annualized in long-term trials.

Duration of follow-up is also important in that exposure time may be critical. Some drugs may not cause certain adverse effects until a person has been taking them for a minimum period. An example is the lupus syndrome with procainamide [32]. Given enough time, a large proportion of participants will develop this syndrome, but very few will do so if treated for only several weeks. Other sorts of time patterns may be important as well. Many adverse effects even occur soon after initiation of treatment. In such circumstance, it is useful, and indeed prudent, to monitor carefully

participants for the first few hours or days. If no effects occur, the participant may be presumed to be at a low risk of developing these effects subsequently.

In the Diabetes Control and Complications Trial (DCCT) [33], cotton exudates were noted in the eyes early after onset of the intervention of the participants receiving tight control of the glucose level. Subsequently, the progression of retinopathy in the regular control group surpassed that in the tight control group, and tight control was shown to reduce this retinal complication in insulin-dependent diabetes. Focus on only this short-term adverse effect might have led to early trial termination. Fortunately, DCCT continued and reported a favorable long-term benefit-harm balance.

Figure 12.1 illustrates the first occurrence of ulcer symptoms and complaints of stomach pain, over time, in the Aspirin Myocardial Infarction Study [30]. Ulcer symptoms rose fairly steadily in both the aspirin and placebo groups, peaking at 36 months. In contrast, complaints of stomach pain were maximal early in the aspirin group, then decreased. Participants on placebo had a constant, low level of stomach pain complaints. If a researcher tried to compare adverse effects in two studies of aspirin, one lasting weeks and the other several months, the findings would be different. To add to the complexity, the aspirin data in a study of longer duration may be confounded by changes in aspirin dosage and concomitant therapy.
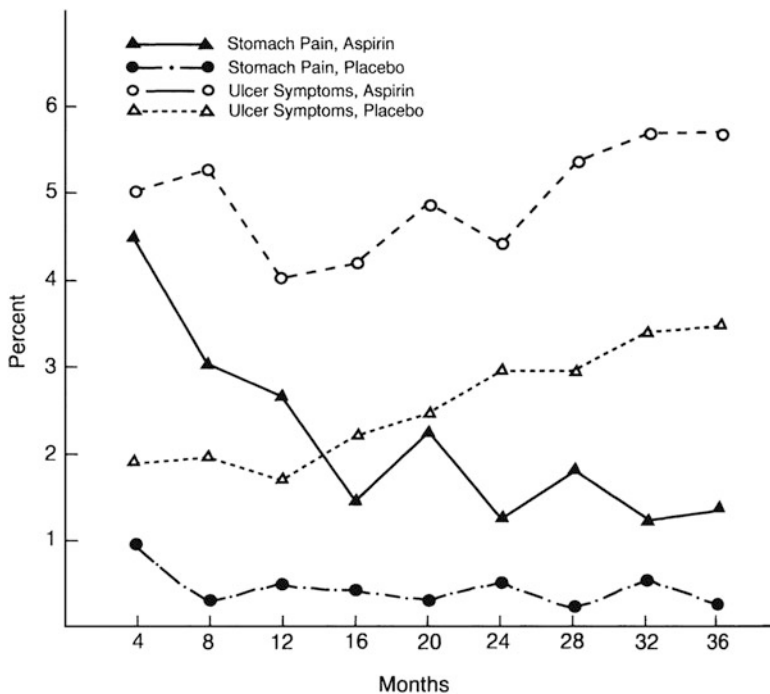
Fig. 12.1 Percent of participants reporting selected adverse events, over time, by study group, in the Aspirin Myocardial Infarction Study

An intervention may cause continued discomfort throughout a trial, and its persistence may be an important feature. Yet, unless the discomfort is considerable, such that the intervention is stopped, the participant may eventually stop complaining about it. Unless the investigator is alert to this possibility, the proportion of participants with symptoms at the final assessment in a long-term trial may be misleadingly low.

Analyzing Adverse Events

Analysis of adverse events in clinical trial results depends in part on the intended use of the analysis. On one hand, drug regulators may provide detailed specifications for both required format and content of information of harm. On the other, peer reviewed journals typically provide space limited to a single table and a paragraph or two in the Results section (although electronic publication can allow considerably more space). Analysis will also depend on specifics of the participant population and intervention under study. Collection, analysis and reporting for prevention in a largely healthy population may differ substantially from an intervention in hospitalized patients with pre-existing heart failure. Nevertheless many trials provide important opportunities unavailable outside a clinical study setting to evaluate potential harm of interventions and public health is served by thorough analysis, even if results are reported in appendixes or on-line supplements.

This section will review four basic types of analysis: standard reporting of adverse events occurring in the trial, prespecified analysis of adverse events of interest, post hoc data-mining, including other exploratory analysis and meta-analysis.

Standard Reporting

The most basic form of assessment of harm is a complete accounting for all participants including those who did not complete the trial. Overall dropout rates are a useful measure of the tolerability of the drug or other interventions, and can be compared across many interventions. Dropout reporting is typically divided into at least three subcategories: dropout due to adverse events, dropouts for lack of efficacy and dropouts for administrative reasons. Assignment of a case to these subcategories may be more subjective than it appears. Lack of efficacy dropouts may rise because symptomatic adverse events might persuade some participants that they are not getting enough benefit to continue. Withdrawals of consent or other administrative departures may conceal problems with the drug, or the conduct of the trial. The overall dropout rate across all categories should be presented. If the dropouts have characteristics over time (such as dropouts related to short-term, early onset adverse events), some form of survival analysis of dropout rate over time may provide useful insights for managing treatment or suggest a need for dose titration.

Another standard analysis consists of a table of reported adverse events at the MedDRA level of Preferred Terms, with control and each intervention arm forming a column for easy comparison across groups. To make the list manageable in length, investigators typically set a threshold value for a subset of adverse events that total more than 1%, 5%, or 10% of patients. This has the major drawback of excluding less common adverse events which may be the more serious ones. Tests of statistical significance may be presented, but must be interpreted cautiously. Longer tables are usually organized by body system using the MedDRA System Organ Class. These standard event tables do not distinguish the severity and frequency of adverse events and are typically dominated by frequently occurring symptoms such as headache, nausea or dizziness.

Standard safety analysis may also include a listing of deaths, serious adverse events, clinically significant laboratory abnormalities, and changes in vital signs.

Prespecified Analysis

Possible adverse effects that could be reasonably expected from the known mechanism of action of the evaluated intervention, prior studies, or underlying participant conditions could be defined and analyzed from the perspectives of ascertainment, classification, and in particular statistical power, but these are rarely done. An investigator needs to consider prospectively and in the analysis the possibility of Type I or Type II error in the context of all three.

Adjudication is a tool frequently used when adverse events are of particular importance or are difficult to define. Adjudicated events are typically assessed by expert panels blinded to study group and following written protocols. While adjudicated results are typically seen as increasing the credibility and objectivity of the findings they may also reduce already limited statistical power by discarding cases with incomplete information. Adjudication can also be abused to suppress adverse event counts though unreasonably specific and restrictive case definitions. In addition, bias may be introduced if the adjudicators are not fully blinded. In the Randomized Evaluation of Long-Term Anticoagulation Therapy (RE-LY) trial, a team of adjudicators reviewed the outcome documents after reported to have been blinded [11]. Subsequently, the FDA took a closer look at the documents and concluded that information on intervention group assignment was available in 17% of the cases [34]. The credibility of adjudication results can be enhanced by accounting for possible but excluded cases.

Post Hoc Analysis

All post hoc analyses of adverse events may be subject to the criticism that it introduces bias because the analyses were not prospectively defined. Bias may also be introduced by problems of ascertainment and classification. These concerns are

valid, but must be considered in light of two factors. First, analyses of prespecified events may themselves have biases and additional, even post hoc, analyses may provide further insight. Second, good clinical research is expensive, difficult to conduct, and seldom repeated without addressing new scientific issues. Therefore, post hoc analysis may yield important information and clues not otherwise obtainable.

One straightforward post hoc analysis addresses limitations of adverse event classification that occur due to the underlying MedDRA terminology. With approximately 20,000 Preferred Terms to describe an adverse event, this terminology permits substantial precision at the cost of disaggregating adverse events and is raising issues about accuracy. For example, at the Preferred Term level, a case of depression could be coded into any of 22 different terms (Table 12.2). Problems of gastrointestinal tolerability might be divided into nausea, vomiting, dyspepsia, and various forms of abdominal pains. Adverse event tables can be examined at all three key levels of the MedDRA hierarchy (Preferred, High Level and High Level Group Terms) as well through other categories created or Standardized MedDRA Queries. Additional understanding of adverse events could be expanded through examining time to reaction, effect duration, or severity. While these post hoc analyses may provide valuable insights into the harm of drugs and medical interventions, they should be specifically identified as separate from prospectively defined analyses.

Statistical techniques for data mining may provide additional opportunities to detect new signals of harm overlooked by clinical investigators in blinded trials. These techniques can be applied initially to the analysis of spontaneous adverse event reports but can be used both for signal detection in individual clinical trials and pooled data sets. With large populations, repeated visits, multiple outcome measures, many concomitant medications, and measures of underlying disease severity, the accumulated data are often too massive to exploit effectively with a prospective data analysis plan. However, the results of data mining analysis should be regarded as hypothesis generating that, after evaluation, would require additional investigation. Such signals may provide a useful basis for additional post hoc studies of existing data or enable prespecified analysis in future clinical trials. Data mining results may also provide context and focus to interpret particular results that were prespecified. Statistical tools such as the false discovery rate estimation [35] can help identify reliable associations in larger spontaneous reporting databases; other analysis might point to the need to explore associations that appeared borderline initially.

Meta-analysis

When individual trials are inconclusive, one approach is the combination of data on harm from multiple trials in a meta-analysis or systematic review (see Chap. 18).

Meta-analyses or pooled analyses conducted by manufacturers are commonly included in New Drug Applications submitted to regulatory agencies. Meta-analyses of treatment harm are now being published in leading medical journals.

Singh and colleagues published three meta-analyses showing that rosiglitazone and pioglitazone double the risk of heart failure and fractures (in women) in type 2 diabetes [36, 37] and that rosiglitazone, in contrast to pioglitazone, also increases the risk of heart attacks [38]. None of these adverse effects was recognized at the time of regulatory approval of these drugs. Singh and colleagues concluded that cumulative clinical trial data revealed increased cardiovascular harm associated with rofecoxib a couple of years before the drug was withdrawn from the U.S. market. It has been recommended that cumulative meta-analysis be conducted to explore whether and when pooled adverse effect data reveal increased harm [39].

It is important to keep in mind that meta-analyses of harm have many limitations. Adverse event data in published studies are usually limited and event ascertainment seldom disclosed. Individual trials revealing unfavorable results may never be reported or published leading to publication bias and underestimation of the true rate of adverse effects. Experience has shown that conclusions from meta-analyses of a large number of small trials are not always confirmed in subsequent large trials.

Even though the clinical trials were prospectively designed, meta-analysis for harm is vulnerable to all the biases of a post hoc study design about a controversial safety issue when both the relevant trials and the number of events in each trial are already known by the study investigators. Small differences in inclusion or exclusion criteria can have large effects on the relative risk calculation, but are not evident in published results.

A substantial problem arises when investigators report that a meta-analysis of numerous trials detected no evidence of an adverse drug event reported using other methods. The failure to disprove the null hypothesis (no difference observed) is then claimed to be an assurance of safety. In this setting, additional evidence is required to rule out a simple Type II statistical error—that a difference existed but could not be detected in this study. In comparative clinical trials with an active drug control this problem is managed with relatively rigorous statistical standards for demonstrating non-inferiority. No such standards exist for meta-analysis of drug adverse events. Finally, when the magnitude of reported harm is small (for example a relative risk <2) all these imperfections in this technique mandate caution in interpreting the results.

Reporting of Harm

Selecting the appropriate and relevant data about harm from the large amount of data collected is a substantial challenge and may vary by the type and duration of the clinical study.

The usual measures of harm include:

(a) Participants taken off study medication or device removed;
(b) Participants on reduced dosage of study medication or on lower intensity of intervention;
(c) Type, severity and recurrence of participant symptoms or complaints;
(d) Abnormal laboratory measurements, including X-rays and imaging;

(e)   Clinical complications
 (f)   In long-term studies, possible intervention-related reasons participants are hospitalized;
(g)   Combinations or variations of any of the above.

   All of these measures can be reported as the number of participants with the occurrence at any point during the trial. Presenting data about how frequently these occurred in the same participant requires more detailed data and may consume considerable space in tables (again, electronic publication may allow considerably more space). Another method is to select a frequency threshold and assume that adverse events which recur less often in a given time period are less important. As an example, of ten participants having nausea, three might have it at least twice a week, three at least once a week, but less than twice, and four less than once a week. Only those six having nausea at least once a week might be included in a table, with the criteria fully disclosed.
   Severity indices may be used. It can be assumed that a participant who was taken off study drug because of an adverse event had a more serious episode than one who merely had his dosage reduced. Someone who required dose reduction probably had a more serious event than one who complained, but continued to take the dose required by the study protocol. Data from the Aspirin Myocardial Infarction Study [30], using the same adverse events as in the previous example, are shown in Table 12.4. In the aspirin and placebo groups, the percent of participants complaining about hematemesis, tarry stools, and bloody stools are compared with the percent having their medication dosage reduced for those adverse events. As expected, numbers of participants complaining were many times greater than those prescribed reduced dosages. Thus, the implication is that most of the complaints were for relatively minor occurrences or were transient in nature.
   As mentioned above, another way of reporting severity is to establish a hierarchy of consequences of adverse events, such as permanently off study drug, which is more severe than permanently on reduced dosage, which is more severe than ever on reduced dosage, which is more severe than ever complaining about the effect. Unfortunately, few published clinical trial reports present such severity data.

Table 12.4 Percent of participants with drug dosage reduced or complaining of selected adverse events, by study group, in the Aspirin Myocardial Infarction Study

	Aspirin ($N = 2,267$)	Placebo ($N = 2,257$)
Hematemesis		
Dose reduced	0.00	0.00
Complaints	0.27	0.09
Tarry stools		
Dose reduced	0.09	0.04
Complaints	1.34	0.67
Bloody stools		
Dose reduced	0.22	0.04
Complaints	1.29	0.45

Scientific Journal Publication

Published reports of clinical trials typically emphasize the favorable results; the harmful effects attributed to a new intervention are often incompletely reported. This discordance undermines an assessment of the benefit-harm balance. A review of randomized clinical trials published in 1997 and 1998 showed that reporting of harm varied widely and, in general, was inadequate [40]. Adverse effect reporting was considered adequate in only 39% of 192 clinical trial articles from seven therapeutic areas. The 2001 CONSORT statement included a checklist of 22 items that investigators ought to address in the reporting of randomized clinical trials. However, it only included one item related to adverse events which recommended that every report presents "All important adverse events or side effects in each intervention group" [41].

In 2004 [42], the checklist was extended to include ten new recommendations related to the reporting of harm-related issues and accompanying explanations (Table 12.5). The authors encouraged the investigators to use the term "harm" instead of "safety", which is a reassuring term. In the first two years after the publication of the 2004 CONSORT guidelines the impact was negligible. Pitrou et al. [43] analyzed 133 reports of randomized clinical trials published in six general medical journals in 2006. No adverse events were reported in 11% of the reports.

Table 12.5 Endorsed recommendations regarding better reporting of harms in randomized trials [42]

Recommendation	Description
1	If the study collected data on harms and benefits, the title or abstract should so state
2	If the trial addresses both harms and benefits, the introduction should so state
3	List adverse events with definitions for each (with attention, when relevant, to grading, expected vs. unexpected reactions, reference to standardized and validated definitions, and description of new definitions)
4	Clarify how harms-related information was collected (mode of data collection, timing, attribution methods, intensity of ascertainment, and harms-related monitoring and stopping rules, if pertinent)
5	Describe plans for presenting and analyzing information on harms (including coding, handling of recurrent reactions, specification of timing issues, handling of continuous measures, and any statistical analyses)
6	Describe for each arm the participant withdrawals that are due to harms and the experience with the allocated treatment
7	Provide the denominators for analyses on harms
8	Present the absolute risk of each adverse event (specifying type, grade, and seriousness per arm), and present appropriate metrics for recurrent reactions, continuous variables and scale variables, whenever pertinent
9	Describe any subgroup analyses and exploratory analyses for harms
10	Provide a balanced discussion of benefits and harms with emphasis on study limitations, generalizability, and other sources of information on harms

Eighteen percent did not provide numerical data by treatment group and 32% restricted the reporting to the most common events. The data on severity of adverse events were missing in 27% of the publications and almost half failed to report the proportion of participants withdrawn from study medication due to adverse events.

Ioannidis [44] proposed six explanations for inadequate reporting of adverse events that reflects diverse motives. (1) the study design ignored or undervalued adverse events, (2) collection of adverse events during the trial was neglected, (3) reporting of adverse events was lacking, (4) reporting of adverse events was restricted, (5) reporting of adverse events was distorted, and (6) the evidence of harm was silenced. The same recommendations are included in the 2010 CON-SORT statement [45].

This is clearly an area in reporting of trial results that is not handled well. It is imperative that investigators devote more attention to reporting the key data on harm from their clinical trials. If not in the main results article, additional data on harm could be included in appendices to this paper or, if possible, covered in separate articles.

Regulatory Considerations

The regulatory issues related to the reporting of harm and efficacy in clinical trials are discussed in more detail in Chap. 22 (Regulatory Issues). Guidance for safety evaluation can be found in documents issued by the US Department of Health and Human Services [46–51].

The purpose of premarketing assessment of harm is to identify adverse effects prior to regulatory approval for marketing. This assessment is typically incomplete for several reasons. Very few early phase studies are designed to test specified hypotheses about harm. They are often too small to detect less common serious adverse events or adverse events of special interest. Additionally, the assessment of multiple adverse events raises analytic questions regarding multiplicity and thus proper significance levels. Moreover, the premarketing trials tend to focus on low-risk participants by excluding elderly persons, those with other medical conditions, and those on concomitant medications, which also reduces the statistical power.

The major drug regulatory agencies in the world have requirements for expedited reporting of adverse events in clinical trials. These requirements apply to serious, unexpected, and drug-related events. As described earlier, a serious adverse event is defined as death, life-threatening event, hospitalization initial or prolonged, persistent or significant disability, congenital anomaly/birth defect, or required intervention to prevent harm or other medically serious event. Unexpected means an effect is not listed in the Investigator's Brochure or product label at the severity observed. The unexpected events in trials registered with the FDA must be reported by the trial sponsor in writing within 15 calendar days of being informed. For an unexpected death or life-threatening reaction, the report should be made within 7 days of notification. The regulations do not specify deadlines for sites to report

these reactions to the study sponsor, although sponsors typically establish their own deadlines.

To deal with often limited information on harm, special regulatory attention is given to adverse trends in the data. The regulatory term *safety signal* [49] is defined as "a concern about an excess of adverse events compared to what would be expected to be associated with a product's use." These signals generally indicate a need for further investigation in order to determine whether they are drug-induced or chance findings. As part of the approval decision, the sponsor may be required to conduct post-approval phase IV studies.

Rules for reporting adverse events to the local ethics review committees vary. Many require that investigators report all events meeting regulatory agency definitions. These committees have, based on the safety report, several options. These include making no change, requiring changes to the informed consent and the trial protocol, placing the trial on hold, or terminating approval of the trial. However, the committees seldom have the adequate expertise or infrastructure to deal with serious adverse event reports from multicenter trials, or even local trials. When the trial is multicenter, different rules and possible actions from different ethics committees can cause considerable complications. These complications can be reduced when the ethics review committees agree to rely on safety review by a study-wide data monitoring committee.

Recommendations for Assessing and Reporting Harm

Substantial improvements are needed in the ascertainment, analysis, and reporting of harm in clinical trials. One advance would be to match better sample size, patient population, and trial duration to clinical use, especially when long-term treatment is intended.

Second, to meet higher standards in the evaluation of harm, efforts should be made in pre-approval trials to prespecify and collect data on known or potential intervention-induced adverse effects. The data ought to be solicited with special questions asked of the participants rather than left completely open-ended and be based on a volunteered response. Asking participants whether they had any general problem since the last contact will underestimate the true rate of reported adverse events, especially those that are sensitive. Collection of known adverse effects is also important in trials of new populations or when new indications are investigated in order to permit determination of the benefit-harm balance. If groups of participants are believed to be susceptible to adverse events or effects, prespecified subgroup analyses ought to be identified in the protocol. As stated above, subgrouping based on genetic variations has been very informative.

Third, limiting the assessment of harm to the simple frequency of adverse events is a crude approach. As stated above, many adverse events have additional dimensions—severity, time of onset, and duration. By ignoring these, one episode of a mild adverse symptom is given equal weight to a severe, constant symptom leading

to discontinuation of the intervention. As a minimum, the number of participants taken off the study intervention due to an adverse event, the number who had their dose reduced and those who continued treatment according to protocol in spite of an adverse event, ought to be assessed and reported in publications.

Fourth, all serious events should be fully disclosed, by study group. There is no reason to omit, restrict or suppress these events especially if they are of a serious nature. Even non-significant imbalances are important. In the disclosure, it is also essential to account for all randomized participants.

Fifth, we endorse the ten CONSORT recommendations regarding better reporting in the literature of harms in randomized trials (Table 12.5). There should be a full and open accounting of all important adverse effects in the main trial publications.

Sixth, we support cooperation with investigators who are pooling and analyzing adverse effect data from multiple clinical trials. This type of data sharing has strong support in the academic community [52–58]. Support for data sharing has also been given by industry [59–61], funders of research [62], major organizations [63] and medical journals [64]. A 2015 report from the Institute of Medicine recommends responsible data sharing for completed trials, with focus on data used in trial publications as well as data used in the complete study report submitted for regulatory review [65]. More details of this report are presented in Chap. 20.

Seventh, we have limited sympathy for investigators who question the existence of adverse effects unless clearly documented in randomized clinical trials. Other types of studies, systematically analyzed case reports, and use of registries have a role in the identification of serious adverse effects. A detailed discussion of these falls outside the scope of this book. Very large observational studies have been successfully used in the past [22]. Spontaneous adverse event reporting continues to be a critical and primary source for identifying new serious adverse drug reactions that were not fully evident in clinical trials. One study of all new major safety warnings from the FDA in 2009 showed that 76% of new Boxed Warnings in the drug label were based on spontaneous reports [17]. A subsequent paper from the FDA confirmed that spontaneous reports accounted for over half of all safety-related drug label changes [66]. Thus, these data can establish associations, but the incidence of such adverse effects needs to be determined through additional studies.

References

 1. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Clinical Safety Data Management: Definition and Standards for Expedited Reporting E2A, October 27, 1994.
 2. United States Code, Code of Federal Regulations 21 CFR 314.80(a) Postmarketing reporting of adverse drug experiences. Definitions.
 3. US General Accounting Office. FDA Drug Review: Postapproval Risks, 1976-85. Washington, DC: US General Accounting Office; April 26, 1990. GAO/PEMD-90-15.

4. Bombardier C, Laine L, Reicin A, et al. for the VIGOR Study Group. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *N Engl J Med* 2000;343:1520-1528.

5. Bresalier RS, Sandler RS, Quan H, et al. for the Adenomatous Polyp Prevention on Vioxx (APPROVe) Trial Investigators. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. *N Engl J Med* 2005;352:1092-1102.

6. Solomon SD, McMurray JJV, Pfeffer MA, et al. for the Adenoma Prevention with Celecoxib (APC) Study Investigators. Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. *N Engl J Med* 2005;352:1071-1080.

7. Psaty BM, Furberg CD. COX-2 inhibitors – Lessons in drug safety. *N Engl J Med* 2005;352:1133-1135.

8. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 2007;356:2457-2471.

9. Food and Drug Administration. FDA briefing document: advisory committee meeting for NDA 21071 Avandia (rosiglitazone maleate) tablet July 13 and 14, 2010. www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/Endocrinologicand MetEndocrinologicandMetabolicDr/UCM218493.pdf.

10. Nissen SE. Rosiglitazone: a case of regulatory hubris. The FDA's defensiveness over its decisions means further drug safety disasters may occur. *BMJ* 2013;347:f7428.

11. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2009;361:1139-1151.

12. Moore TJ, Cohen MR, Furberg CD. Quarterwatch 2012 Quarter 2. www.ismp.org/quarterwatch/pdfs/2012Q2.pdf

13. Eikelboom JW, Connolly SJ, Brueckmann M, et al. Dabigatran versus warfarin in patients with mechanical heart valves. *N Engl J Med* 2013;369:1206-1214.

14. Frank C, Himmelstein DU, Woolhandler S, et al. Era of faster FDA drug approval has also been increased black-box warnings and market withdrawal. *Health Affair* 2014;33:1453-1459.

15. Furberg BD, Furberg CD. *Evaluating Clinical Research. All that Glitters is Not Gold* (2nd edition). New York, NY: Springer, 2007, pp. 17-18.

16. Venning GR. Identification of adverse reactions to new drugs. II: How were 18 important adverse reactions discovered and with what delays? *Br Med J* 1983;286:289-292 and 365-368.

17. Moore TJ, Singh S, Furberg CD. The FDA and new safety warnings. *Arch Intern Med* 2012;172:78-80.

18. Vandenbroucke JP, Psaty BP. Benefits and risks of drug treatments. How to combine the best evidence on benefits with the best data about adverse effects. *JAMA* 2008;300:2417-2419.

19. Guideline for Industry: The extent of population exposure to assess clinical safety for drugs intended for long-term treatment of non-life-threatening conditions. International Conference on Harmonization, Geneva, March 1995.

20. Committee on the Assessment of the US Drug Safety System. Baciu A, Stratton K, Burke SP (eds.). *The Future of Drug Safety: Promoting and Protecting the Health of the Public*. Washington, DC: The National Academies Press, 2006.

21. Furberg CD, Levin AA, Gross PA, et al. The FDA and drug safety. A proposal for sweeping changes. *Arch Intern Med* 2006;166:1938-1942.

22. Papanikolaou PN, Christidi GD, Ioannidis JPA. Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. *CMAJ* 2006;174:635-641.

23. Multiple Risk Factor Intervention Trial Research Group. Baseline rest electrocardiographic abnormalities, antihypertensive treatment, and mortality in the Multiple Risk Factor Intervention Trial. *Am J Cardiol* 1985;55:1-15.

24. Siscovick DS, Raghunathan TE, Psaty BM, et al. Diuretic therapy for hypertension and the risk of primary cardiac arrest. *N Engl J Med* 1994;330:1852-1857.

25. Psaty BM, Furberg CD, Ray WA, Weiss NS. Potential for conflict of interest in the evaluation of suspected adverse drug reactions: use of cerivastatin and risk of rhabdomyolysis. *JAMA* 2004;292:2622-2631.

26. www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm
27. Introductory Guide MedDRA Version 16.1. Maintenance and Support Services Organization, Chantilly (MSSO), VA, 2013.
28. Introductory Guide for Standardised MedDRA Queries (SMQs) Version 16.1. MedDRA Maintenance and Support Services Organization (MMSSO), Chantilly, VA, 2013.
29. NCI Guidelines for Investigators. http://ctep.cancer.gov.
30. Aspirin Myocardial Infarction Study Research Group. A randomized, controlled trial of aspirin in persons recovered from myocardial infarction. *JAMA* 1980;243:661-669.
31. Lee K, Lee Y, Nam J, et al. Antidepressant-induced sexual dysfunction among newer antidepressants in a naturalistic setting. *Psychiatry Investig* 2010;7:55-59.
32. Dalle Vedove C, Simon JC, Girolomoni G. Drug-induced lupus erythematosus with emphasis on skin manifestations and the role of anti-TNFα agents. *J Dtsch Dermatol Ges* 2012;10:889-897.
33. The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 1993;329:977-986.
34. Beasley N, Thompson, A. Clinical Review of NDA 022-512 Dabigatran (Pradaxa), August 24, 2010. (Amended October 17, 2010). US Food and Drug Administration, Center for Drug Evaluation and Research, Silver Spring, MD, page 42.
35. Ahmed I, Dalmasso C, Haramburu F, et al. False discovery rate estimation for frequentist pharmacovigilance signal detection methods. *Biometrics* 2010;66:301-309.
36. Singh S, Loke YK, Furberg C. Thiazolidinediones and heart failure: A teleo-analysis. *Diabetes Care* 2007;30:2148-2153.
37. Loke YK, Singh S, Furberg CD. Long-term use of thiazolidinediones and fractures in type 2 diabetes: a systematic review and meta-analysis. *CMAJ* 2009;180:32-29
38. Singh S, Loke YK, Furberg CD. Long-term risk of cardiovascular events with rosiglitazone. *JAMA* 2007;298:1189-1195.
39. Ross JS, Madigan D, Hill KP, et al. Pooled analysis of rofecoxib placebo-controlled clinical trial data. Lessons for postmarket pharmaceutical safety surveillance. *Arch Intern Med* 2009;169:1976-1984.
40. Ioannidis JPA, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001;285:437-443.
41. Moher D, Schulz KF, Altman DG, for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-1194.
42. Ioannidis JPA, Evans SJ, Gøtzsche PC, et al. for the CONSORT Group. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141:781-788.
43. Pitrou I, Boutron I, Ahmad N, Ravaud P. Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med* 2009;169:1756-1761.
44. Ioannidis JPA. Adverse events in randomized trials. Neglected, restricted, distorted, and silenced. *Arch Intern Med* 2009;169:1737-1739.
45. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.
46. Department of Health and Human Services, Food and Drug Administration: International Conference on Harmonisation; Guideline on clinical safety data management: Definitions and standards for expedited reporting, Notice. *Federal Register* 60 (1 March 1995): 11284-11287.
47. Department of Health and Human Services, Food and Drug Administration. International Conference on Harmonisation; Draft guidance on E2D postapproval safety data management: Definitions and standards for expedited reporting, Notice. *Federal Register* 68 (15 September 2003): 53983-53984.
48. U.S. Department of Health and Human Services. Food and Drug Administration. Guidance for Industry. Premarketing risk assessment. March 2005. www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm126958.pdf.

49. U.S. Department of Health and Human Services. Food and Drug Administration. Guidance for Industry. Good pharmacovigilance practices and pharmacoepidemiologic assessment. March 2005. http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126834.pdf.

50. U.S. Department of Health and Human Services. Food and Drug Administration. Reviewer Guidance. Conducting a clinical safety review of a new product application and preparing a report on the review. March 2005. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM072974.pdf.

51. European Medicines Agency, ICH Topic E 2 A: Clinical Safety Data Management: Definitions and Standards for Expedited Reporting. European Medicines Agency, London, UK. June 1995.

52. Gøtzsche PC. Why we need easy access to all data from all clinical trials and how to accomplish it. *Trials* 2011;12:249.

53. Boulton G, Rawlins M, Vallance P, Walport M. Science as a public enterprise: the case for open data. *Lancet* 2011;377:1633-1635.

54. Loder E. Sharing data from clinical trials. Where we are and what lies ahead. *BMJ* 2013;347: f4794.

55. Mello MM, Francer JK, Wilenzick M, et al. Preparing for responsible sharing of clinical trial data. *N Engl J Med* 2013;369:1651-1658.

56. Zarin DA. Participant-level data and the new frontier in trial transparency. *N Engl J Med* 2013;369:468-469.

57. Eichler H-G, Pétavy F, Pignatti F, Rasi G. Access to patient-level data—a boon to drug developers. *N Engl J Med* 2013;369:1577-1579

58. Krumholz HM, Peterson ED. Open access to clinical trials data. *JAMA* 2014;312:1002-1003.

59. Wellcome Trust. Sharing research data to improve public health: full joint statement by funders of health research. 2011. http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm.

60. PhRMA (Pharmaceutical Research and Manufacturers of America) and EFPIA (European Federation of Pharmaceutical Industries and Associations). Principles for responsible clinical trial data sharing: Our commitment to patients and researchers. 2013. http://phrma.org/sites/default/files/pdf/PhRMA PrinciplesForResponsibleClinicalTrialDataSharing.pdf.

61. Nisen P, Rockhold F. Access to patient-level data from GlaxoSmithKline clinical trials. *N Engl J Med* 2013;369:475-478.

62. NIH (National Institutes of Health). Final NIH statement on sharing research data. 2003. http://grants.nih.gov/grants/policy/data_sharing.

63. IOM. Sharing clinical research data: Workshop summary. 2013. Washington, DC: The National Academies Press.

64. Godlee, F, Groves T. The new BMJ policy on sharing data from drug and device trials. *BMJ* 2012;345:1-3.

65. Institute of Medicine Committee on Strategies for Responsible Sharing of Clinical Trial Data. *Sharing clinical trial data: maximizing benefits, minimizing risk*. Washington, D.C.: The National Academies Press, 2015.

66. Lester J, Neyarapally GA, Lipowski E, et al. Evaluation of FDA safety-related drug label changes in 2010. *Pharmacoepidemiol Drug Saf* 2013;22:302-305.