Chapter 2
Analyzing Safety Data

Abstract The safety of a drug is assessed at all stages of its life cycle, from drug discovery through nonclinical development, preapproval clinical development, and all the time it is subsequently on the market. The focus of this chapter is safety assessment during later-stage clinical trials. When a drug is granted marketing approval by a regulatory agency and therefore becomes available to prescribing physicians and their patients, the best available information upon which the doctor and patient can form answers to questions concerning the drug’s safety profile is the safety information gathered during clinical trials. This information is provided to the clinician and all patients receiving the drug in the drug’s package insert, or label. The content of the label will have been agreed upon by the Sponsor and the regulatory agency granting marketing approval, and it summarizes the best available data about the drug’s safety at that point in time.

Keywords Drug safety • Adverse events • Adverse drug reaction • Exclusion of unacceptable risk • Antidiabetic drugs • Risk ratio

2.1 Introduction

The safety of a drug is addressed at all stages in its life history. This starts long before the drug is administered to humans during clinical trials. By the time Phase I trials commence the Sponsor will have collected considerable amounts of safety data during the nonclinical development program. Safety will be assessed throughout the clinical development program, and, if the drug is approved for marketing, it will be assessed throughout the time the drug is on the market. As in other chapters, the focus here is on assessments during Phase III trials.
2.2 Providing Safety Data to Prescribing Physicians and Patients

Jumping ahead to the point where a new drug has been approved, consider a scenario in which a prescribing physician is discussing with a patient whether the new drug is a good treatment option (therapeutic decisions are ideally made by a ‘health team’ comprising the physician and the patient). What kinds of information might the doctor and the patient find useful? Examples include [1]:

- How likely is the patient to experience an adverse drug reaction? (The term adverse drug reaction is employed once the drug is on the market. Before that, as we will see shortly, the term adverse event is used in clinical trials).
- Are the typical adverse drug reactions temporary or permanent in nature?
- How likely is the patient to suffer an adverse drug reaction that is extremely serious, or even life-threatening?
- How might the risk of an adverse drug reaction vary with different doses of the drug?
- How might the risk of an adverse drug reaction change with increasing length of treatment?
- Are there any specific clinical parameters that should be monitored more closely than usual in patients receiving this drug?

When the drug is marketed and therefore becomes available to prescribing clinicians, the best available information upon which a clinician and patient can form answers to such questions is the safety information gathered during clinical trials. This information is provided to the clinician, and all patients receiving the drug, in the drug’s package insert or label. The content of the label will have been discussed by the Sponsor and the regulatory agency granting marketing approval, and it summarizes the best available data about the drug’s safety at that point in time.

2.3 The Clinical Study Report

Clinical trial data, both safety and efficacy, are typically presented in a clinical study report (CSR) that is suitable for submission to a regulatory agency. Describing, or summarizing, the tremendous amount of data that are collected in a clinical trial is typically a useful first step in reporting the results of the trial. Simple descriptors such as the total number of participants in the trial, the numbers that received the drug treatment and the placebo treatment, respectively, information concerning sex and ethnicity, and the average age of the participants in each treatment group help to set the scene for more detailed reporting. This information can be usefully summarized in in-text tables that are placed in the body of the text in CSRs. In each case, the source of the information presented will need to be cited. The source is typically one of many listings that are appended to
2.3 The Clinical Study Report

Table 2.1 Subject accountability (Clinical Trial ABC789)

<table>
<thead>
<tr>
<th>Completion status</th>
<th>Number (%) of subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Drug (N = 400)</strong></td>
<td><strong>Placebo (N = 400)</strong></td>
</tr>
<tr>
<td>Completed study</td>
<td>320 (80)</td>
</tr>
<tr>
<td>Withdrew prematurely</td>
<td>80 (20)</td>
</tr>
<tr>
<td><strong>Premature withdrawals</strong></td>
<td></td>
</tr>
<tr>
<td>Adverse event</td>
<td>40 (10)</td>
</tr>
<tr>
<td>Withdrew consent</td>
<td>20 (5)</td>
</tr>
<tr>
<td>Protocol violation</td>
<td>20 (5)</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
</tr>
</tbody>
</table>

Other: 1. [Description]. 2. [Description] Source Table: XYZ

the report. Listings are comprehensive lists that provide all information collected during the trial.

2.4 Subject Demographics and Accountability

Each CSR has various sections, including subject demographics and accountability, safety data, and efficacy sections. An in-text table may be presented for demographic characteristics. Specific characteristics that are important can vary from study to study, but typical ones include age, sex, ethnicity, and baseline data of relevance, e.g., weight, blood pressure, and heart rate. Information concerning the use of concomitant or concurrent medications and evaluations of subject adherence or compliance with the trial’s treatment schedule is also typically presented.

Table 2.1 provides an example of an in-text table from a hypothetical clinical trial that summarizes subject accountability.

Several comments about these hypothetical data are appropriate. First, it is possible but unlikely that the number of subjects for the two treatment groups would be identical in a real study. Presenting percentages as well as absolute numbers is therefore useful, since the percentages allow for differing totals of subjects in each group. Second, the number of subjects in the individual categories must add up to the respective group totals. Third, explanation of (any) other reasons for premature withdrawal should be presented, either in text form above or below the table or in footnote form immediately underneath the table.

Documentation of premature withdrawals from a study is important for various reasons. The implications of premature withdrawals are different in the analysis of safety and in the analysis of efficacy. From a safety perspective, these data relate to tolerability of the drug. From an efficacy perspective, dropouts lead to missing data, and the way(s) that missing data are addressed is important from the point of view of full interpretation of the analysis presented.
2.5 Safety Parameters Measured

Safety data sets comprise enormous numbers of parameters. These include laboratory tests (e.g., liver function tests such as albumin, bilirubin, and alanine aminotransferase), electrocardiogram (ECG) waveforms, vital signs (e.g., blood pressure, heart rate, and weight), and possibly imaging data. These data are typically presented descriptively. The mean, minimum, and maximum values may be presented in many cases, which would capture measures of both central tendency (the mean) and dispersion (the range in this case). There may also be a count of ‘values higher than the normal range’ and ‘values lower than the normal range’ for a parameter.

2.6 Adverse Events

General safety assessments are wide-ranging, and are typically also presented descriptively. Safety-related data can be considered at four levels: extent of exposure to the drug; adverse events; laboratory tests; and vital signs. We will focus on adverse events.

During the course of a clinical trial it is likely that many subjects will have some form of adverse events (AEs). The longer the study and the sicker the subjects are, the more AEs there will be. When reporting the results from the clinical trial, therefore, it is of interest to know about the frequency of all adverse events. Several different kinds of adverse events can be distinguished:

- Pretreatment adverse events;
- Treatment adverse events;
- Drug-related adverse events;
- Serious adverse events (SAEs);
- Adverse events of special interest;
- Adverse events leading to withdrawal from the study.

There are various ways to provide this information, including listings and in-text summary tables. Descriptive statistics for AEs typically include rates of occurrence of the events in exposed groups overall and among subgroups of subjects (e.g., according to age and sex) to look for any potential patterns of differential rates of adverse events.

Often, a sponsor will report adverse events that occurred in at least a given percentage of the subjects in either group (information for both treatment groups is presented in each case). The meaning of the term most common must be defined every time it is used. In this example it is defined by the statement “Greater or equal to 10%.” It is likely that the incidence of AEs will not be identical in the two treatment groups. Some AEs that occur in 10% or more of the subjects in the drug treatment group may occur in less than 10% of subjects in the placebo treatment group, and vice versa. Data for both treatment groups employed will be provided.
for any AE listed for either group. Data are often presented in descending order of occurrence. Similar presentations of these data are included in package inserts for marketed products.

2.7 From Descriptive Statistics to Inferential Statistics

The analysis of general safety data collected in Phase III trials is not particularly rigorously defined, and as a result the presentation of safety data, as noted, is largely descriptive. This situation is now changing, certainly for more serious events. Examples are therefore provided from the field of cardiovascular safety in which inferential statistics plays a central role.

As noted in the book’s Abstract, the ultimate purpose of the results from a clinical trial is not to tell us precisely what happened in that trial, but to gain insight into likely drug responses in patients who would be prescribed the drug should it be approved for marketing. One branch of the discipline of Statistics, inferential statistics, enables us to do this. The precise data collected from the specific group of subjects who participated in a given trial are used to infer what the likely responses would be in the general population of patients with the disease or clinical condition of concern.

2.8 Prospective Exclusion of Unacceptable Risk

The prospective exclusion of unacceptable cardiac and cardiovascular risk has become a critical component of drug safety evaluation in new drug development. A useful approach to prospectively excluding such risk can be conceptualized in a three-component model comprising clinical, regulatory, and statistical science [2]:

- **Clinical science:** Clinical judgments concerning absolute and relative risks are needed.
- **Regulatory science:** Based on clinical evidence, regulatory agencies have to choose the thresholds of regulatory interest, which are, at least from an absolutist point of view, the criteria of demarcation between acceptable risk and unacceptable risk. (In practice, regulators are given more latitude based on other aspects of the drug and the disease it is intended to treat).
- **Statistical science:** Once a threshold of regulatory concern has been established by a regulatory agency (or one proposed by ICH has been adopted), accepted statistical methodologies for determining whether or not these thresholds have been breached is required. Confidence intervals play a central role in this methodology.

The statistical science component is straightforward and a precise answer to the research question is provided every time, although interpretation of the results can
be less black-and-white. However, there is less certainty involved in the choice of a regulatory threshold. Certainly, clinical science and clinical judgment can be taken into account, but it is still neither a straightforward nor an easy decision for a regulatory agency to make. Nonetheless, for this risk exclusion model to be implemented, a decision must be made.

2.8.1 Assessment of Unacceptable Cardiac Risk

Assessment of the proarrhythmic liability of noncardiac drugs, i.e., drugs not intended for cardiac indications, has assumed a central importance in the development of noncardiac drugs. While not perfect, QT interval prolongation as seen on the ECG is the central parameter of interest in cardiac safety assessment. Figure 2.1 provides a schematic representation of the normal QT interval and a prolonged interval.

The arrhythmia Torsades de Pointes, which is very rare, often self-correcting, but potentially fatal, is associated with inherited QT prolongation, and also with drug-induced QT prolongation.

The ICH E14 Guideline addressing this issue was released in 2005 [3]. It describes the Thorough QT/QTc (TQT) study, a clinical trial devoted to the rigorous assessment of an investigational drug’s QT/QTc liability, i.e., its propensity to prolong the QT interval. (Since QT varies with heart rate independently of any drug-induced effect, the measured QT interval is “corrected” for heart rate by one or more of several correction formulas, leading to the term QTc.)

TQT studies are typically conducted using healthy subjects. An indication of typical QT intervals is given by ranges of 350–460 ms for men and 360–470 ms for women [4]. The TQT study is designed to look for drug-induced increases of “around 5 ms” [3]. Operationally, this degree of prolongation is defined by placing confidence intervals (CIs) around the observed mean prolongation, in this context called the treatment effect point estimate. In this case, two-sided 90% CIs
are placed around the treatment effect point estimate, and attention falls on the upper limit of the confidence interval, which is deemed the “threshold of regulatory concern” [3]. A threshold of regulatory concern is the product of the interaction between clinical and regulatory science. In this case, it is operationalized as an upper limit of a two-sided 90% CI placed around the treatment effect point estimate (point estimate of cardiac risk) obtained from a single study of 10 ms or greater.

A drug’s treatment effect can be defined as the mean response due to the drug minus the mean response due to a placebo: In many experimental circumstances, although the placebo is not pharmacologically active, the mean response to its administration will still be a small change in the biological parameter of interest, here the QT interval. The treatment effect point estimate provides precise information about the degree of QT interval prolongation seen in the single study of interest. However, we wish to use all of the data collected in the study to infer what might be the responses of patients prescribed the drug should it be approved for marketing. Placement of confidence intervals around the treatment effect point estimate facilitates this inference.

Consider a scenario in which the treatment effect point estimate for QT interval prolongation in a TQT trial was 8.00 ms, and the lower and upper limits of the two-sided 90% CI placed around this point estimate are 6.50 ms and 9.50 ms, respectively. This result can be written as 8.00 (6.50, 9.50). We can now make a statement concerning the true but unknown treatment effect in the general population from which the sample that participated in the trial was drawn:

- The data from this single TQT study are compatible with a treatment effect (prolongation of the QT interval) as small as 6.50 ms and as large as 9.50 ms in the general population, and our best estimate is a treatment effect of 8.00 ms.

Recalling that the regulatory threshold of concern for QT interval prolongation is an upper limit of 10 ms, it can be seen that the value of 9.50 ms falls below this threshold. Therefore, this drug would not be deemed to be associated with unacceptable cardiac risk.

### 2.8.2 Assessment of Unacceptable Cardiovascular Risk

In December 2008 the FDA issued a guidance entitled *Diabetes Mellitus—Evaluating Cardiovascular Risk in New Antidiabetic Drugs to Treat Type 2 Diabetes* [5]. This guidance detailed additional research related to cardiovascular safety that sponsors must complete before submitting an NDA for a new antidiabetic drug for type 2 diabetes mellitus (T2DM). A draft Guidance containing similar cardiovascular safety requirements was released by the European Medicines Agency in January 2010 [6].

Clinical development programs for new drugs for T2DM have traditionally included small and relatively short (12-week) therapeutic exploratory studies and a
12–24-month therapeutic confirmatory study. Subjects recruited into the trials were typically relatively healthy, certainly healthier than the eventual target patient population for the approved drug. The results of all trials conducted were then presented to a regulatory agency. However, the guidances have considerably changed the nature of the required trials. A typical clinical development program will now include:

- Small, short, dose-finding early trials;
- Larger and longer late therapeutic exploratory trials;
- Larger and longer therapeutic confirmatory trials that include subjects at high risk for cardiovascular events.

The FDA guidance recommends collecting clinical trial data that are more directly applicable to the patients likely to be prescribed the drug if marketing approval is given. One consideration for achieving this goal is to make the trial protocol’s inclusion criteria less restrictive, thereby opening enrolment to all patient populations that would be reflected in the approved drug’s label. A second consideration is the inclusion of high-risk subjects. The guidance states that therapeutic exploratory and confirmatory trials should include subjects “at higher risk of cardiovascular events, such as patients with relatively advanced disease, elderly patients, and patients with some degree of renal impairment” [5].

In contrast to the employment of QT interval prolongation as a cardiac safety biomarker, clinical endpoints are of interest when addressing cardiovascular safety for T2DM drugs. The Major Adverse Cardiovascular Events (MACE) composite endpoint includes non-fatal myocardial infarction, non-fatal stroke, and cardiovascular death. An expanded MACE endpoint might include hospitalization for unstable angina and urgent percutaneous coronary intervention/ coronary artery bypass graft surgery. Other possible secondary endpoints include carotid revascularization and lower extremity amputations/revascularization. In this context, the traditional MACE endpoint is acceptable to the FDA.

Cardiovascular events reported during the trial must be centrally adjudicated by independent experts [7]. Having all cardiovascular events for all subjects in the trials at all investigational sites evaluated against precisely the same criteria by the same group of independent experts considerably enhances the validity of the data. The guidance notes that sponsors should establish “an independent cardiovascular endpoints committee to prospectively adjudicate, in a blinded fashion, cardiovascular events” during all Phase II and Phase III trials [5]. These events should include cardiovascular mortality, myocardial infarction, and stroke, and can include hospitalization for acute coronary syndrome, urgent revascularization procedures, and possibly other endpoints.

Like the ICH Guideline E14 discussed in the previous section, the guidance takes the approach of prospectively excluding unacceptable risk by the employment of a threshold of regulatory concern and associated statistical methodology employing confidence intervals. However, the analysis to be conducted does not use data from a single trial. Rather, data from essentially all Phase II and Phase III trials are used via employment of a meta-analysis (this topic is covered in more
detail in Chap. 5). It is of regulatory interest to determine to what extent (if any) the drug may increase the occurrence of the cardiovascular safety endpoints in the patient population who would be prescribed the drug if approved. In this scenario, the clinical endpoints of interest are categorical, and typically dichotomous: for example, either a myocardial infarction occurred, or it did not. In this setting, relative risk is of interest, which can be captured by a risk ratio.

Consider a hypothetical single trial in which the drug treatment group receives a new drug for T2DM, and another group of subjects receives a control treatment. The question of interest here is: Does the drug lead to an unacceptable increased risk of cardiovascular events? To answer this question, a (safety) treatment effect will be calculated in the form of a risk ratio (the term treatment effect can refer to efficacy treatment effects and safety treatment effects). A relative risk ratio point estimate is calculated by considering the number of cardiovascular events in the drug group as the numerator and the number of cardiovascular events in the control group as the denominator. If the number of events in each treatment group happened to be precisely the same (the likelihood of this is vanishingly small) the value of the point estimate would be unity, represented here as 1.00. If the number of events in the drug treatment group is greater than the number for the control treatment group, the value will be greater than 1.00.

Going one step further, since multiple Phase II and Phase III trials will have been completed, a meta-analysis can be conducted, producing a relative risk point estimate that is based on data from all of the studies rather than from just a single study. Having done this, a confidence interval can be placed around this relative risk point estimate. In this scenario, a two-sided 95% CI is employed. This information will then be used to infer what the true but unknown population relative risk is. The logic for conducting a meta-analysis first is that the inference based on a larger database (the data from all Phase II and Phase III trials combined) will be more informative than an inference based on any single trial.

Imagine that the value of the relative risk ratio point estimate computed from the meta-analysis is 1.20, and that the lower and upper limits of the confidence interval are 1.15 and 1.26. This can be written as $1.20 (1.15, 1.26)$. These confidence intervals allow us to estimate the true but unknown relative risk for the general patient population. We can now make the following statement:

- The data obtained from this meta-analysis are compatible with as little as a 15% increase and as much as a 26% increase in the risk of cardiovascular events in the general population, and our best estimate is an increase of 20%.

Of course, any increase in risk is not ideal. However, the question here is: Is there compelling evidence of an unacceptable increase in risk, given the serious nature of the disease? (The more serious the disease being treated, the more risk one may be prepared to take in order to obtain the desired therapeutic benefit). To permit this question to be answered requires us to compare the result with the thresholds of regulatory concern presented in the FDA guidance (such thresholds are not explicitly stated in the EMA draft guidance). Three scenarios are described:
• If the upper limit of the two-sided 95% CI is equal to or greater than 1.80, the drug would be deemed to have an unacceptable risk. In this case, the Sponsor would be expected to conduct more research and collect more data whose analysis would yield an upper limit of less than 1.80 before requesting marketing approval.

• If the upper bound is equal to or greater than 1.30 and also less than 1.80, and the overall benefit-risk analysis presented at submission supports marketing approval (benefit-risk is considered in more detail in Chap. 4), “a postmarketing trial generally will be necessary to definitively show that the upper bound of the two-sided 95 percent confidence interval for the estimated risk ratio is less than 1.3.”

• If the upper limit is less than 1.30 and the overall benefit-risk analysis presented at submission supports marketing approval, “a postmarketing cardiovascular trial generally may not be necessary.”

Returning to our hypothetical result from the meta-analysis conducted for the drug’s Phase II and Phase III trials, the results were a relative risk of 1.20 (1.15, 1.26). As in the case of the TQT study and its investigation of QT interval prolongation, interest focuses on the upper limit of this two-sided confidence limit. This value, i.e., 1.26, falls below the regulatory threshold of concern of 1.80, and, moreover, also falls below the more stringent threshold of concern of 1.30. This result therefore falls in the territory covered by the third bullet point in the previous list. It is likely that, given a favorable overall benefit-risk profile, the drug will be approved and there will not be a requirement to conduct a postmarketing trial to further evaluate the relative risk.

2.8.3 A More Realistic Scenario and its Unintended Consequences

The relative risk point estimate and its associated two-sided 95% CI in the example in the previous section were deliberately chosen to illustrate the scenario where the cardiovascular risk associated with the drug was not deemed to be unacceptable, and was not therefore prospectively excluded by denying the drug marketing approval. However, this example represents the ‘ideal case.’ A more realistic example would be a case where the upper limit falls below the 1.80 threshold of regulatory concern, but above the 1.30 threshold. It is therefore likely that the Sponsor would be asked to conduct a postmarketing trial to show definitively that, based on a much larger sample of data, the upper limit actually falls below 1.30.

The intent of the guidance is an extremely laudable one: It is to ensure that new antidiabetic drugs do not unacceptably increase the risk for cardiovascular events in patients would receive a new drug should it be approved for marketing. The importance of this intent is underscored by observations that diabetes increases the risk of heart disease and stroke and that the majority of patients with T2DM die from cardiovascular disease and not from their hyperglycemia per se. However,
there may be unintended consequences. A postmarketing cardiovascular trial of the kind discussed in the guidance requires a large number of subjects (perhaps in the order of five times as many as participated in an earlier Phase III trial), and such trials cost a considerable amount of money (tens to hundreds of millions of dollars). In addition to the costs of the ‘regular’ Phase II and Phase III trials, this represents a financial challenge to many sponsors. It is true that the FDA should not be concerned with such costs: their job is to protect and promote public health. However, an unfortunate and unintended consequence may be that fewer drugs for this serious disease will be developed by Sponsors who are concerned that trying to develop a drug for this indication may be beyond their financial resources.

Caveney and Turner [8] noted that an inspection of relevant data on www.clinicaltrials.gov reveals a general increase in the number of diabetes trials between 2005 and 2008, followed by a leveling off in 2009. Additionally, late 2008 and 2009 saw several smaller biotechnology companies abandon their diabetes programs because of the cost increase. It is therefore possible to speculate that the new regulatory mandates (or burdens) are leading all biopharmaceutical companies, regardless of size, to re-examine their diabetes pipelines and re-forecast their predicted return on the investment necessary to bring a drug to market. As these authors commented, “With so many patients suffering with diabetes, the unclear pathogenesis of the disease, and patients not meeting professional goals for optimal care, the field is ripe for more research discoveries and the market is open for further drug developments. Regulators, policy-makers, and industry leaders will need to be vigilant and work together to ensure that the new regulatory guidance does not stifle the development of antidiabetic agents” [8].

References
