Clinical Trials: – Collection of Safety Data and Establishing the Adverse Drug Reaction Profile

J. Talbot and M. D. B. Stephens

Introduction

All clinical trials should have a safety component as a primary or secondary objective. In early Phase I and Phase II trials, safety and tolerability are often a primary objective and the main reason for conducting the study. In later Phase II and Phase III trials the primary objective is usually efficacy, but safety and tolerability must also be included as a secondary objective.

Safety data from clinical trials can be broadly divided into the following three types, although these can overlap:

- adverse events (AEs)
- laboratory safety data
- vital signs and physical findings.

Safety monitoring in clinical trials can be considered as non-specific, i.e. general safety monitoring, or specific, i.e. looking for particular safety issues based on animal data, pharmacology or experience with other similar drugs or from earlier trials; see below.

Clinical trials have some general safety goals:

- To detect and characterize common adverse drug reactions (ADRs), usually type A.
- To determine tolerability in volunteers or patients, i.e. how is the ADR tolerated, does it resolve or improve on repeated dosing, is it so unpleasant that the subject has to stop treatment or can they put up with it?
- To identify any predisposing or risk factors for particular ADRs.
In studies where a range of doses is used and in studies with a pharmacokinetic component, an additional safety goal will also be to determine the relationship between ADRs and dose or plasma concentrations of the drug. The nature and frequency of the safety monitoring used in a trial will depend on experience with the drug and the type of study being conducted. In early Phase I safety and tolerability studies in volunteers, there will be extensive and frequent monitoring (electrocardiogram (ECG), blood pressure, etc.), blood sampling and questioning of the subjects. In large Phase III trials, standard questions and routine laboratory screens at appropriate time intervals are typically used. In some Phase IV studies, only certain clinical outcomes, e.g. death, hospitalization or a clinical event such as stroke, may be collected. This chapter will focus mainly on AEs. Laboratory safety data are dealt with in Chapter 5.

Adverse Events

The term AE was defined by Finney (1965) as ‘a particular untoward happening experienced by a patient, undesirable either generally or in the context of his disease’. AEs are not necessarily recognized drug reactions and a causal relationship to treatment is not implied, as in the term ADR. Thus, all ADRs are AEs but only some AEs are ADRs.

The concept of collecting AEs rather than ADRs was adopted after the failure of clinical trials to detect severe skin and eye problems with practolol. Skegg and Doll (1977) proposed that the value of clinical trials in detecting unwanted effects of new medicines would be enhanced if doctors recorded all AEs experienced by subjects, not just those regarded as adverse reactions to drugs. All events should be reported to the centre coordinating the trial and analysed in treated subjects and controls. This is the basis for how AE monitoring in clinical studies is conducted today. Events should be treatment-emergent. This can be defined as an event that was not present before the start of treatment and became apparent after treatment began, or an event that was present before the start of treatment but worsened after treatment began. In controlled studies, the profile of AEs in the different treatment groups can be compared; see Figure 4.1.

This simple example, taken from either a single parallel group, placebo controlled trial or pooled data from a number of such trials, shows only ten different AEs. In reality, there will be many more different AEs, maybe hundreds, which then need to be grouped by system organ class and preferred terms; see Chapter 12. Figure 4.1 shows that patients on both drug and placebo experience AEs, with the placebo group giving the background nature and frequency of AEs in the patient population. Some AEs are more frequent on drug; in this example, rash and diarrhoea are significantly more frequent. This serves to flag these AEs for closer scrutiny as possible ADRs, and common clinical characteristics, similar time to onset, etc. would strengthen their association. However, with such multiple comparisons, some AEs will be significantly more frequent by chance, e.g. if the frequencies for 100 different AEs were compared, five would be significantly different at $p = 0.05$ by chance; see below and also Chapter 6.

Furthermore, if an AE is not more frequent in the group receiving drug, then this does not exclude individual cases of the AE being drug induced. It could simply be an ADR of low frequency; this is often seen with serious events such as hepatitis or agranulocytosis, where the numbers in a trial are likely to be very small and not statistically significant.
Factors affecting collection of adverse events

In order to collect AEs efficiently it is necessary to know which factors might hinder their collection so that they can be circumvented. Between the occurrence of an AE and its final assessment the AE must be communicated. Factors preventing this communication include failure of the patient to recognize the AE or to communicate it to the clinical investigator and failure of the investigator to recognize or report it. The patient may fail to recognize the AE because:

- There are no symptoms (i.e. biochemical change, hypertension, etc.).
- There is a change in mood, which is only recognized by relatives or friends.
- There is a lack of intelligence or mental illness.

The patient may fail to communicate the AE to the investigator because:
• The patient does not associate the AE with the drug and, therefore, does not consider it to be relevant.

• The patient recognizes the event as a possible ADR (e.g. from the patient consent information), but presumes that one has to put up with it.

• The patient does not inform the doctor for fear of being thought neurotic or because the doctor inhibits the patient by tone of voice, interruptions or poor bedside manner.

• The patient has a poor memory or there are long intervals between patient visits.

The investigator may fail to recognize the AE because:

• The doctor does not give the patient the opportunity to communicate the AE.

• The doctor listens to the patient, but fails to consider the possibility of an AE.

• The doctor fails to take positive steps to look for AEs (i.e. does not ask questions and/or examine patient adequately).

The investigator recognizes the AE but does not report it because of:

• Complacency, thinking that it is too minor to report.

• Fear of litigation.

• Guilt at causing the patient to suffer.

• Ignorance of the mechanism of reporting (this should not happen in clinical trials).

• Lethargy – too busy.

Filtration of adverse events using protocols and case report forms

Bearing in mind the factors that can prevent the reporting of an AE to the pharmaceutical company or trial coordinator, steps can be taken to overcome them by using the study protocol and case report forms (CRFs) for collecting data in addition to the explanation by the study monitor. The wording of the protocol and the form design need to be appropriate to the indication and the stage in the drug’s development.

Collection of adverse events

A specific illness does not preclude patients from experiencing many of the same AEs as a healthy person, in addition to those due to their illness. Should details of all AEs be collected? This will depend upon several factors:

• The indication for the drug. If the drug is given for a minor illness, then even very mild AEs may be relevant. If the drug is given for a disease that leads to death, such as secondary cancer or AIDS, then minor discomforts are less relevant.
• The stage or phase of the clinical trial programme. Details of minor symptoms are relevant for Phase I studies, but once an ADR has been characterized then counting the numbers of AEs may suffice.

• The type of potential ADR. Standardized enquiry is often needed for psychiatric studies, and laboratory tests are necessary for some drug-induced diseases.

• It is important that the clinical investigator makes a diagnosis wherever possible, rather than just listing signs and symptoms. When the diagnosis is in the form of a syndrome, such as an organ failure, the cause should be recorded by the investigator whenever possible (e.g. left-sided cardiac failure due to hypertension (Nickas, 1995)).

• A signal of a possible ADR halfway through a clinical trial programme may require changes to the methods of collecting data.

• AE collection during a clinical trial programme does not need to be standardized throughout the programme, but it must be consistent. The addition of a questionnaire to one study should not interfere with the analysis as long as the other standard methods of collection are included. Different questionnaires for a single drug programme should only be used if the drug programme is for two separate indications.

**During treatment**

The basic principle is to collect AEs that have appeared whilst the patient is on treatment and in the immediate period after stopping treatment, as well as any AE that was present at baseline but has since become worse on treatment. These are sometimes called ‘treatment-emergent signs and symptoms’ (TESS). A recent history of any AE immediately before the study may influence the assessment if it (or a related event) then occurs on treatment.

It is usual practice for companies to collect AEs with specifically designed company forms, and Astra was one of the pioneers in this area (Wallander and Palmer, 1986); however, they discovered that a substantial number of AEs were recorded not on the special form, but elsewhere in the CRF (73.7 per cent were on the correct form, 26.3 per cent were found elsewhere). In older studies, the percentage found elsewhere was 33–36 per cent, whereas for studies starting in 1986 the figure was only 13 per cent. Moreover, in the older studies, those found elsewhere were often serious events: 42 per cent compared with 26 per cent on the correct form. This ratio subsequently reversed, so that by 1989 no serious AEs and only 14 per cent of non-serious AEs were not on the correct form. Drug safety staff should hence ensure that the whole of the CRF is checked for AEs and that there is a good CRF design, instructions and investigator training (Wallander *et al.*, 1992).

If a checklist or questionnaire is going to be used then it needs to cover the same time interval on each occasion, and since minor events are soon forgotten then the interval should generally not be longer than 2 weeks. Wallander *et al.* (1991) found that a questionnaire revealed events primarily from the preceding week, despite asking about the previous month. This means that there should be a baseline question, questionnaire or checklist that covers the previous 1 or 2 weeks followed by the same question, questionnaire or checklist 1 or 2 weeks later. This allows easy recognition of increased frequency of headaches for example. The subsequent intervals between visits should preferably be the same; but this is
not always practical or essential, since it is possible to compare the events occurring with the trial drug and comparator over the same period. If the study is for longer than 4 weeks then it is preferable to have at least two visits while on the drug (one in the middle of treatment and another on the last day of treatment) and a post-treatment visit. In a long study, weekly questionnaires may overburden the patient and their enthusiasm to complete them might diminish. The only way to circumvent this problem might be to use the questionnaire to highlight crucial points (e.g. the week before adding the study medication and the week after, so as to pick up the most common type A ADRs).

**The post-treatment visit**

Spilker (1984) considers that the main purpose of the post-treatment visit is to check for any withdrawal effects and to ensure the patient’s safety, but says that they may also be used to study residual effects of treatment. He lists five factors that should govern whether or not a post-treatment period is needed:

- previous experience with the drug in similar patient populations;
- whether the drug dosage is being tapered slowly or stopped abruptly;
- the clinical status of the patient;
- whether patients are in a secure and/or controlled environment;
- the pharmacokinetic characteristics of the drug.

Friedman et al. (1985) described two types of post-study follow-up. The first was a short-term follow-up, which should be considered when ‘intervention’ is stopped at the previous visit in order to find out how soon the laboratory values or symptomatology return to baseline. The second type was a long-term follow-up, monitoring possible toxicity or benefit.

The guidelines of the Fogarty Conference (Davidson et al., 1979) for the detection of hepatotoxicity recommend that, in early clinical trials of new drugs (Phase II), laboratory tests should be performed at 24 h, 5–7 days and 4–6 weeks after the last dose. In short-term studies (less than 6 months) the follow-up should be for 2 months and that following long-term studies (6 months or more) should be for 2 years, and 20 years in a small subset of patients. These recommendations are obviously intended to detect adverse reactions with a very delayed onset, such as fibrosis, cirrhosis, vascular lesions or neoplasms.

A routine post-treatment visit that includes examination and a routine laboratory screen is essential for all pre-marketing studies for the reasons outlined below:

- *To review the laboratory data from the samples taken at the last visit whilst on treatment.* The results of a laboratory test will take time to reach the investigator. If of clinical significance then it needs to be followed up until it is normal or a cause is found, since it may represent an ADR, a new disease, or a complication of an underlying disease. If it is abnormal, but of no clinical significance, then it may be an early sign of an ADR, a new disease or complication of an old disease, or a chance variation from normal. A repeated individual test value is unlikely to be abnormal by chance. Clinical enquiry and, if necessary, examination and further tests may resolve whether or not the abnormality is due to the drug or disease.
• To detect any delayed ADR. A rare type B ADR may not appear until after a drug has been stopped, e.g. jaundice or aplastic anaemia. The fialuridine disaster, in which 5 of 15 patients with chronic hepatitis B died from drug-induced hepatotoxicity, developed 9–13 weeks after treatment (McKenzie et al., 1995).

• To detect any signs or symptoms due to drug withdrawal or a rebound phenomenon. It is important to show that drugs used in psychiatric disease are not followed by drug withdrawal symptoms similar to those with benzodiazepines (Busto et al., 1986). Any rebound phenomena usually occur during the first week after stopping treatment, e.g. beta blocker rebound.

• To ascertain the response of AEs to dechallenge. The response to dechallenge is an important factor in the assessment of an AE occurring whilst on treatment.

Disadvantages of a post-treatment visit The visit is an additional inconvenience for both the patient and the investigator and an additional cost to the patient and the pharmaceutical company. These are counterbalanced by the assurance that no lasting harm has been caused to the patient. However, if a chronic disease requires replacement of the trial drug by another treatment, then this latter treatment may cause an ADR by itself and it may be difficult to distinguish between a delayed ADR due to the trial drug and that caused by the replacement treatment. This problem could be overcome by replacement with a well-established drug with known side effects.

A more frequent problem occurs when the AE itself is treated, thereby confounding the response to dechallenge; this should only occur rarely, since most ADRs reverse rapidly when treatment is stopped. Any complications or sequelae of the underlying disease may be difficult to differentiate from a delayed ADR. These can be countered, as there should be an equal incidence in the control group.

The reasons for the post-treatment visit dictate the duration of the follow up period. The results of the laboratory tests must be available at the time of the visit and sufficient time must be given for AEs to reverse and for any withdrawal symptoms and any rebound phenomena to manifest themselves. These events should have occurred within 1 week; a delayed ADR may take longer, and a 2 week interval is probably reasonable. However, cholestatic hepatitis may appear up to 5 weeks later, as in the case of co-amoxiclav. A visit at 1 month may be more reasonable.

When the study has a crossover design, the intermediary washout period should act as a post-treatment period following the first drug and be equal in duration to the post-treatment observation periods following the second drug (Stephens, 1988).

It is also reasonable to recommend a routine post-treatment visit following post-marketing clinical trials for subgroups of patients who may be prone to ADRs, such as the young, the elderly and those with organ failure. There may also be hypotheses arising from animal work or from clinical experience with similar drugs that will require long-term follow up for specific purposes. Long-term follow up to ensure that no serious ADRs, such as aplastic anaemia or a fialuridine-type problem, have occurred (i.e. at about 9 months) might be done by regular telephone interviews with the patient and/or GP.

Conclusion The increased protection for the patient, the additional information obtained for assessment of any AEs and better knowledge of the safety of a drug that should result
from routine post-treatment visits make them essential. They may need to be repeated under certain circumstances.

**Post-study serious adverse events**

AEs occurring after the final visit (usually the post-treatment visit) should not be actively sought or collected unless long-term follow up is indicated; see below. However, on occasions, investigators report deaths and other serious AEs that have occurred weeks or months after the study. It is important to establish why they have been reported, particularly serious AEs not due to underlying disease. Is there a possibility of a delayed drug effect or did the investigator misunderstand their responsibilities for reporting? When received, such post-study serious AEs will usually be entered onto the drug safety database but probably not onto the clinical study database. They are not usually included in the statistical analysis of the study but may be added as an addendum to the clinical study report (CSR) and, if appropriate, discussed in the safety section of the CSR.

**Long-term follow up**

For some drugs it may be appropriate to undertake long-term safety monitoring and follow-up of outcomes after patients have completed treatment. This has been done with some immuno-modulatory drugs such as tumour necrosis factor (TNF) antagonists, where there are concerns regarding infections, particularly tuberculosis, cancers, etc. (Klareskog *et al.*, 2001). Such a long-term follow up will either require a special protocol or an extension to the existing protocol and informed consent. Contact with the patient can be made by post, telephone or Internet and validated by the original investigator or the patient’s GP. A control group is essential, preferably patients who have been randomly assigned to a comparator or placebo in the original study.

**Separation of adverse drug reactions from placebo reactions**

Since adverse symptoms not caused by drugs are common (Reidenberg and Lowenthal, 1968) and are not easily separated from drug-induced symptoms, both must be collected for analysis if a profile of ADRs is to be made. However, this technique can only be used in controlled studies, ideally with placebo. In principle, the frequency of ADRs may be estimated by subtraction of the number of patients affected by a particular AE in the placebo group from the number affected in the active drug group as follows (Lasagna, 1984):

\[
\text{Drug group} - \text{Placebo group} = \text{Number of ADRs}
\]

However, this is an oversimplification because:

- The difference may have arisen by chance.
- Having established that there is a significant difference between the two treatment groups for the number of events and the number of patients afflicted, the severity of the ADRs in the two groups should then be compared.
- The difference may be due to bias (see below), e.g. because of inadequate blinding.
A further problem is classification. Some terms may include more than one type of abnormality. For example, the incidence of ‘blurred vision’ may be equal in both groups; but there may be several cases of tunnel vision with the trial drug, but because there is no code for tunnel vision it is coded under a more general term. Another problem is that the symptoms forming a syndrome are often reported and coded separately; individually, there may be no difference between two treatments, but when the cases are examined there may be a combination of symptoms with one drug that warrant being called a syndrome. It is essential, therefore, essential to know the individual original description of the AEs before making a judgement. This area has been explored more fully by Bernheim (1994), and he has added bias to the equation:

\[
\text{Attributable AEs} = \frac{\text{Drug group AEs}}{\text{Placebo group AEs}} \pm \text{Bias}
\]

Bias is equal to the \( B \) (baseline frequency) and severity of the AE, multiplied by pharmacological clinical activity of the drug \( A_D \), minus the pharmacological clinical activity of the placebo \( A_P \):

\[
\text{Bias} = B(A_D - A_P)
\]

ADRs that are similar to common non-drug AEs are rarely described or investigated sufficiently for a causal relationship for each individual event to be established. If they cannot be distinguished qualitatively, then the correct quantitative procedure is to compare them using non-parametric statistics, giving the confidence limits for the incidences of ADRs. Small studies (\( n < 30 \)) have little chance of separating ADRs from placebo or non-drug events unless they are very common and specific to the drug (Simpson et al., 1980). The situation can be worsened in volunteers receiving placebo who have a tendency to ‘catch’ AEs from the active drug group, therefore changing a relatively specific ADR to a non-specific event.

Methods for collecting symptomatic adverse events
Collection of all AEs or symptoms should only be done if:

- it is possible to compare the AEs of one group with those of another, since the ‘background noise’ of the non-drug symptoms can overwhelm the drug-induced symptoms in uncontrolled studies;
- they can be collected at the beginning and end of the study as a minimum.

Methods for collecting adverse events
- diary card,
- questionnaire,
- checklist,
- standard question.
The patient can be prompted to report all adverse symptoms if the investigator uses a diary card, a patient questionnaire or a checklist with a standard question. Since the majority of ADRs occur within the first week of drug treatment, the first visit should be within a week or so of starting treatment if all the minor events are to be collected. There is a steep fall-off in recollection of minor events, even in young intelligent volunteers, and this is likely to be greater with elderly sick patients.

**Differences in reporting rates with different instruments**

Studies with temafloxacin showed the following reporting rates (Norrby and Pernet, 1991):

- spontaneous reporting, 1.5–5.1 per cent
- standard question, 29–49 per cent
- studies using diary cards, 41.5 per cent
- studies not using diary cards, 23.5 per cent.

**Patient diary cards**

In trials where a patient diary card is used for recording patient information (e.g. daily peak flow rates in asthma trials) it can also be used for recording AEs. It is, in fact, the equivalent of answering a daily standard question or checklist. If sufficient space is allotted to the daily recording of any AEs in sufficient detail then the diary card is likely to be large. The large amount of unstructured data that is likely to be collected over any period longer than a few days would be difficult to manage. Daily recording of objective data with weekly recording of AEs makes the data easier to handle without the loss of important events.

In cancer studies, the known side effects due to chemotherapy vary from day to day and diary cards have been used very successfully. The Medical Research Council (MRC) Tuberculosis and Chest Diseases Unit, the Clinic of Oncology and the Radiotherapeutic Unit Daily Diary Card has three pages, one for each week (Fayers and Jones, 1983; Jones *et al.*, 1987); see also www.atsqol.org/ddc-mrc.asp.

The five standard questions asked each day relate to:

- sickness (vomiting)
- activity
- mood
- anxiety
- overall condition.

Each of these questions is given a range of five answers:

- Sickness – (i) none; (ii) poor appetite; (iii) felt sick but wasn’t; (iv) sick once; (v) sick more than once.
• Activity – (i) normal work/housework; (ii) normal work, but with effort; (iii) reduced activity, but not confined to home; (iv) confined to home or hospital; (v) confined to bed.

• Mood – (i) very happy; (ii) happy; (iii) average; (iv) miserable; (v) very miserable.

• Anxiety – (i) very calm; (ii) calm; (iii) average; (iv) anxious; (v) very anxious.

• Overall condition – (i) very well; (ii) well; (iii) fair; (iv) poor; (v) very ill.

Many of the problems of a patient diary have been eliminated with the electronic patient diary (Donovan et al., 1996), and it has been successfully used for rheumatology, urology, Parkinson’s, psychiatry and pulmonary disease. It works like a daily questionnaire rather than a diary, since it does not need to allow free text entries. It can have a built-in alarm to remind the patient to fill it in. It has multiple-choice questions and a visual analogue scale (VAS) can be added. Electronic patient diaries have been found to be cost effective and do not present problems for elderly patients (Lundström, 1993). Symfo S. A., from Belgium, market hand-held electronic patient diaries for clinical trials and surveys Phases I to IV; see www.symfo.com.

**Questionnaire or checklist?**

Whereas diary cards generally collect the AEs experienced by the patient in an unstructured fashion on a daily basis, the questionnaire or checklist collects the AEs in a structured fashion so that valid statistical comparisons can be made between an active drug group and a control group. When the patient diary is limited to answering set questions it acts as a daily questionnaire. In many earlier papers the terms ‘questionnaire’ and ‘checklist’ have been used loosely to mean any question, whether delivered directly to the patient (self-answering questionnaire) or via a third person (i.e. doctor, nurse, social worker). The advantages and disadvantages of these two approaches are outlined in Tables 4.1 and 4.2.

There are advantages and disadvantages to using any form of multiple-question questionnaire compared with the use of a single standard open question, such as ‘Have you had any medical problems since your last visit?’ The multiple-question questionnaire will collect symptoms present in normal healthy subjects in addition to those due to disease or drugs. A comparison between an open question and a 38-item checklist showed that 15 per cent of healthy persons had symptoms in the previous 3 days when the open question was used compared with 82 per cent when the checklist was used (Reidenberg and Lowenthal, 1968). The parallel figures for patients who had been ill or taken medication in the previous 3 days were 69 per cent for the open question and 87 per cent for the checklist. An open question tends to collect only the more severe symptoms, whereas the incidence of irrelevant complaints is higher with the multiple-question questionnaire.

The multiple-question questionnaire is likely to lead to the conclusion that the incidence of AEs with a given drug is higher than that resulting from placebo (Downing et al., 1970); this is especially true for neurotic patients (Rickels and Downing, 1970; Lapierre, 1975) and with depressed patients, where 5–10 times more side effects will be listed with the multiple questionnaire than with the open question. Relevant side effects are more likely to be detected if a checklist is not used (Huskisson and Wojtulewski, 1974; Fisher et al., 1987). Known side effects were recorded more frequently using active questioning, whereas unknown side effects were more likely to be described by spontaneous reports (Hagman et al., 1977).
These comments are not as contradictory as they may first appear if one relates the use of the open question and the multiple-question questionnaire to the phase of drug development. During Phases I and II, before the nature of the ADR profile is known, the open question is probably more appropriate. In large-scale Phase III and IV studies, however, where the relative incidence of common ADRs of the new drug can be compared with those of its main competitor, the multiple-question questionnaire is more likely to differentiate between the two.

<table>
<thead>
<tr>
<th>Table 4.1</th>
<th>Self-administered questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages</strong></td>
<td><strong>Disadvantages</strong></td>
</tr>
<tr>
<td>They can be given directly to the patient and returned to the organizer bypassing the investigator</td>
<td>Needs more organization and the costs to print, distribute, collect and analyse are greater</td>
</tr>
<tr>
<td>Questions involving sexual behaviour can be answered more frankly than by any other method</td>
<td>If the questionnaire bypasses the investigator then the patient may forget to report important AEs, forgetting that the investigator does not see the answers to the questionnaire</td>
</tr>
<tr>
<td>The involvement of the investigator can be minimal</td>
<td>Great care is needed in the wording of questions, since there is no interpretation by the investigator</td>
</tr>
<tr>
<td>Answers to very precise questions can be given and there is almost no limit to the number of questions posed</td>
<td>Tends to overestimate the real incidence of adverse events (Fisher et al., 1987)</td>
</tr>
<tr>
<td>Confidentiality can be guaranteed by the use of the patient trial number</td>
<td>May suggest symptoms to patients</td>
</tr>
<tr>
<td>It is more likely to lead to the conclusion that the incidence of side effects resulting from a given drug is higher than that resulting from placebo</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4.2</th>
<th>Checklist (questionnaire administered by a third party)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages</strong></td>
<td><strong>Disadvantages</strong></td>
</tr>
<tr>
<td>The actual words used to the patient can be tailored according to the patient’s intelligence and background</td>
<td>Unless the person administering the checklist reads out the question to the patient then there will be individual variations in terminology, etc.; therefore, the resulting answers may not be comparable</td>
</tr>
<tr>
<td></td>
<td>The number of questions is more limited, since it involves the administrator’s time</td>
</tr>
<tr>
<td></td>
<td>Questions involving sex may cause more embarrassment than with the self-administered questionnaire</td>
</tr>
</tbody>
</table>

These comments are not as contradictory as they may first appear if one relates the use of the open question and the multiple-question questionnaire to the phase of drug development. During Phases I and II, before the nature of the ADR profile is known, the open question is probably more appropriate. In large-scale Phase III and IV studies, however, where the relative incidence of common ADRs of the new drug can be compared with those of its main competitor, the multiple-question questionnaire is more likely to differentiate between the two.
Questionnaires

There are two types of questionnaire:

- A generic questionnaire that has been developed for use over a wide field, e.g. quality of life (QoL) questionnaires. Its disadvantage is that it needs to be very extensive if it is to cover the range of possible AEs. If it has a restricted number of questions then it must have an open question: ‘Were there any other AEs?’

- A questionnaire designed specifically for a trial with particular drugs. Use of such a questionnaire is only advised for early randomized clinical trials when the new drug is similar to a standard drug, since a questionnaire designed to pick up the known ADRs of the standard drug may be inadequate for identifying the as yet unknown ADRs of the new drug. This may well bias the study in favour of the new drug, since the established ADRs of the standard treatment will be well represented in the questionnaire.

Designing a questionnaire   The 10 stages in this process are (Stone, 1993):

- decide what data you need
- select items for inclusion
- design individual questions
- compose wording
- design layout
- think about coding
- prepare first draft and pre-test
- pilot and evaluate
- perform survey.

Read (Stone, 1993) and (Charlton, 2000) before starting.

Quantification of symptoms

The two main methods for quantifying a symptom are:

- descriptive scales
- Visual Analogue Scales (VAS).

Descriptive scales (fixed interval scales or Likert scales)   Denis Likert first described a five-point scale in 1932. It uses graded descriptive terms:

- Absent or present (score 0 or 1). This is the method used by most questionnaires, but it may lack sensitivity when comparing two similar drugs in a relatively small trial.
If present, is it mild, moderate or severe? (Score 2, 3, 4 respectively). Patients’ understanding of the words mild, moderate and severe is likely to differ, so they should be defined.

A five-point scale (e.g. very drowsy, slightly drowsy, normal, more alert, very alert) (Yuen et al., 1985). The WHO handbook for reporting the results of cancer treatment has a five-grade scheme (0–4) for both clinical and laboratory AEs (WHO, 1979). Available at http://whqlibdoc.who.int/publications/.

A seven-point scale (Jaeschke et al., 1990):
(i) extremely short of breath
(ii) very short of breath
(iii) quite a bit short of breath
(iv) moderately short of breath
(v) some shortness of breath
(vi) a little shortness of breath
(vii) not at all short of breath.

Another seven-point scale can be used for the frequency of the event (Guyatt et al., 1987):
(i) all of the time
(ii) most of the time
(iii) a good bit of the time
(iv) some of the time
(v) a little of the time
(vi) hardly any of the time
(vii) none of the time.

A further seven-point scale has been used with the SAFETEE general inquiry questionnaire. Patients were asked to rank each of the 76 possible symptoms on a scale of 1 to 7. The rankings were based upon the patients’ willingness to exchange their current disease state for a situation where they would now be afflicted with that symptom. A rank of 1 would indicate a complete willingness to exchange and a rank of 7 would be an absolute refusal. Using weighted values for the SAFETEE symptoms obtained from the ranking procedure, an index was created to measure the impact of non-life-threatening AEs associated with drug therapy. This adverse drug effect index has been validated (Levine et al., 1990).

A 10-point scale (Lewis et al., 1985a): the patient is asked to allocate a score between 1 and 10 for different symptoms.
Visual analogue scale A VAS is shown in Figure 4.2. It is a 10 cm scale; it may be vertical or horizontal, with some practical advantage with the horizontal scale. There should not be any intermediate points (Aitken, 1969; Scot and Huskisson, 1976) that might cause clustering. However, there is always a tendency towards clustering at the extremes of the scale, at the midpoint and at 6.18 cm (Benjafied and Adams-Webber, 1976). It has been suggested that patients should see their previous scores when making serial assessments, but others disagree (Joyce et al., 1975; Scott and Huskisson, 1979; Carlsson, 1983). Some 7 per cent of patients find the VAS difficult to understand despite instruction (Lewis, 1987; Jaeschke et al., 1990) and it has been suggested that patients who score inaccurately for all symptoms should be screened out (Lewis et al., 1985c).

![Figure 4.2 A visual analogue scale (VAS)](image)

Grant et al. (1999) compared the reproducibility and sensitivity of a VAS, Borg scale (12 points) and a Likert scale (five points) in normal subjects. The VAS performed best in terms of reproducibility for breathlessness and general fatigue and in terms of sensitivity for breathlessness.

Drug effects were significantly more marked for volunteers who were not depressed compared with those who were depressed (Peat et al., 1981). A VAS has been found to be more sensitive than a 10-point scoring system when used for testing beta adrenoceptor blockers (Lewis et al., 1985c), a five-point scale or a standard question when used to measure sedation (Yuen et al., 1985), a sleep and mood scale (Lundberg, 1980), and a four-point pain scale (Joyce et al., 1975).

Errors can occur using a VAS, and Maxwell (1978) has suggested that it should be combined with either a four-point scale or a simple global assessment. In a VAS assessment of angina using ‘No pain at all’ and ‘Pain as bad as I could ever bear’, afterwards the patients were asked which of the following did they use: number of attacks, duration of attacks, severity of an individual attack or a combination of these; it was found that the patients’ VAS scores correlated best with the severity of pain, which was considered the least clinically important variable. It is important, therefore, to take care in phrasing the question (Vandenburg, 1987). In multi-national studies it is also necessary to be aware of cultural differences, for instance in perception of pain, between countries. An alternative to using a scale is the use of a VAS meter, which gives an immediate reading without the need for subsequent measurement and produces a similar assessment to the conventional VAS (Hounslow et al., 1987).

VASs have been used for:

- Pain (Huskisson, 1974; Joyce et al., 1975; Scott and Huskisson, 1976; Carlsson, 1983; Vandenberg, 1987; McCormack et al., 1988).
• Quality of sleep (Parrott and Hindmarsh, 1978).
• Dyspnoea (Jaeschke et al., 1990).
• Subjective sensation of resistance to breathing (Aitken, 1969).
• Depression (Zealley and Aitken, 1969; McCormack et al., 1988).
• Anxiety (McCormack et al., 1988).
• Beta blocker side effects (Lewis et al., 1984; Lewis et al., 1985a,b; Lewis, 1987; Dimenas et al., 1989).
• Quality of life (Jaeschke et al., 1990).

Bulpitt and Fletcher (1990) did not use a VAS because of the difficulty in explaining the concept to many patients, the lack of data on validity and repeatability, and the difficulty in interpreting the results.

Conclusions  A VAS is best used as an efficacy assessment when the symptom is due to the underlying disease and the drug is likely to improve the symptom (i.e. a change in severity). In a large-scale placebo-controlled clinical trial using a questionnaire for AEs, there are disadvantages in having just absent or present (i.e. a tick in a box if present). It is only where the trial compares a new drug and the standard therapy under similar circumstances and where it is important to detect a difference that it is worth using a four-point scale. If a large-scale study is not possible and a limited questionnaire is used yet it is vital to discover whether the two treatments differ, then it is worth using a VAS, but probably only in a single-centre study with an enthusiastic investigator.

Questionnaire/checklist

A questionnaire can qualify a symptom by asking further questions about the type of sensation, location, duration, quality, etc., but if it aims to cover all body sensations then it will be prohibitively long. Questionnaires devised for one clinical trial may not be suitable in a different context and, therefore, the investigator needs to check whether:

• the questionnaire is acceptable to the study population;
• it is easily completed;
• it will produce responses consistent with those obtained in normal doctor–patient interviews;
• it is reproducible when administered on two separate occasions;
• it will be of value or use when completed (Lewis et al., 1984).

Questionnaires are likely to identify milder symptoms than those volunteered spontaneously or in answer to a general question and will include body sensations experienced by normal subjects. This increases the background noise and may entail the use of larger groups if ADRs are to be distinguished (Borghi et al., 1984).
Types of questionnaire Questionnaires used in clinical trials can be divided into:

- Specific questionnaires designed for a specific trial(s) or for a specific drug where possible side effects are elicited by individual specific questions.

- Generic questionnaires, which can be divided into ADR questionnaires and QoL questionnaires.

The ADR questionnaires are designed to cover all reasonable ADRs. Rare ADRs are too specific and can occur in too many areas to be covered in a questionnaire of limited size. QoL questionnaires cover the physical state, emotional wellbeing, sexual and social functioning, and cognitive acuity of patients (Croog et al., 1986). In general, an ADR questionnaire is used for identifying the AEs suffered by the patient, whereas the QoL assessment indicates how the adverse and beneficial events have affected the patient’s general wellbeing.

Examples of such questionnaires are described in Chapter 5 of the fourth edition of this book. More recent examples include:

- a new instrument to assess drug safety (Sacristan et al., 2001);
- standardizing assessment of adverse effects in rheumatology trials (Woodworth et al., 2001);
- patient-based method of assessing AEs in rheumatology clinical trials (Welch et al., 2001).

An early evaluation of side effects should take place at baseline and within 1–2 weeks of the start of a trial, since most ADRs are evident within that period. Possible long-term ADRs can be assessed using QoL 3 to 6 months later, when adaptation has occurred and placebo effects have worn off. The period covered by the questionnaire should be identical throughout the study. Where and how the questionnaire is completed is important; answering a questionnaire in the home setting may not have the same results as within the hospital environment. It is rarely possible to organize a prospective randomized study of an ADR, due to ethical problems, but it has been done with angiotensin converting enzyme (ACE) inhibitors and cough (Ramsay and Yeo, 1995; Tanser et al., 2000).

Drawbacks of adverse drug reaction questionnaires In all structured systems for the collection of AEs there is a tendency to lose information. A patient’s graphic description of an event may help to separate the drug-induced event from naturally occurring events, but information can be lost on coding. In early studies the quality of the patient’s description must be retained. Although in a structured system there should be space for additional description, there will inevitably be loss of descriptive information. There is a continuum from the individualized approach with a single event reported spontaneously to the counting of events in epidemiological surveys. There is a need to characterize the AEs in detail early in clinical development of a new drug until an accurate description is developed; thereafter they can be counted.
**Standard/open question**

This should be the standard method for all clinical trials and some of its characteristics have already been mentioned. An alternative approach is to record only ‘spontaneously volunteered side effects’. When the list of factors that can prevent reporting of AEs is studied in relationship to this method, it can be seen that it has the following disadvantages compared with a standard question:

- The doctor may not give the patient the opportunity to mention an AE.
- If the doctor assessing the spontaneously volunteered AE judges that it was not due to the drug, then it may not be recorded.

The standard question should be unambiguous and worded in such a way that it is not mistaken for a social courtesy. Examples of standard questions from clinical studies include:

- ‘Have you noticed any change in body function or had any physical complaints in the past week?’ (Avery *et al*., 1967). This was used in a study in America in hospitalized depressed patients. It very pointedly does not ask for any mental changes. The wording might not be so easily understood by other English-speaking patients.
- ‘How are you feeling?’, followed by ‘How else are you feeling?’, and finally by ‘How does the drug make you feel?’. This was used in a study of neurotic outpatients in America (Downing *et al*., 1970).
- ‘Have you noticed any new symptoms which might be related to the treatment?’ (Huskisson and Wojtulewski, 1974).
- ‘Did you experience any unpleasant effects from the medicine you took?’ (Lasagna, 1981).
- ‘Has the treatment upset you in any way?’ (Aitken, 1969).
- ‘Have you noticed any symptoms since your last examination?’ (Jackson, 1990a,b).
- ‘Have you had any health problems since we last met?’ (Wallander *et al*., 1991).

Some of these imply that patients make decisions as to causality and will, therefore, vary in their interpretation of AEs. These examples should be avoided. Two alternative standard questions are:

- ‘Have you had any medical problems since your last visit?’ (Lapierre, 1975), or ‘Have you had any problems since your last visit?’, or ‘Have you had any problems during the last week?’
- ‘Have you felt different in any way since your last visit?’
The standard question is a suitable method for all clinical trials, being usable in addition to patient questionnaires or checklists, as well as independently. If the question is worded correctly then it should collect all drug-associated events, but not stimulate the production of too many non-drug-associated events. If the standard question is worded to collect all events defined by Finney (1965), then it includes non-medical events (i.e. social). The problem of dealing with large amounts of social data in clinical studies has not yet been solved in the drug trial context. Until methodology of collecting, recording and analysing social events has advanced and the pattern established, first for the healthy population and second for disease groups, then the definition of AEs should be restricted to medical events.

**Sequence for collecting subjective adverse events**

If all four methods of collecting subjective AEs (spontaneous, standard question, checklist and questionnaire) are to be used, then instinctively they should be used in this order. After a social greeting, the patient needs to be given the opportunity to mention any medical problem bothering them. Then the standard question should be asked and then, lastly, a checklist used. A questionnaire can be handed to the patient on leaving, so that it can be filled in either in the waiting room or at home. The alternative, which is usually used with QoL questionnaires, is to have the questionnaire filled in prior to the consultation on the grounds that the doctor/nurse cannot then influence its completion. This approach has the theoretical disadvantage that having mentioned their symptoms on the questionnaire they might think it unnecessary to repeat them to the investigator and, therefore, they will not be recorded on the CRF. However, the influence of a patient-completed symptom checklist on the subsequent reporting of AEs in a clinical trial interview was examined in a study of 128 patients receiving anti-epileptic medication. The patients were randomized to receive a 16-symptom checklist either before the clinician assessed the AEs or afterwards. The difference was small and not significant; the authors suggested that giving the checklist first does not affect subsequent reporting (Wagner et al., 1994).

**Investigator assessments**

Having decided how collection of the AEs occurring in a clinical trial is to be done, a decision must be made as to whether one needs to collect the investigator’s opinion about whether the AE was due to the drug. Where a questionnaire or a checklist has been used, a statistical assessment of the numbers of each type of AE with the trial drug and the comparator drug/placebo will be made and the investigator’s opinion is not necessary. However, does one require an opinion on all spontaneous reports and those elicited by a standard question? The number of these AEs is likely to be relatively few and will probably be more severe than those elicited by questionnaire or checklist. The investigator knows more about the patients and their diseases, both past and present, and is frequently an expert in the latter. Although the investigator may be an expert in the disease under treatment, they may well not be expert in the area of the AE. Most general physicians have a good knowledge of common ADRs of drugs in general use, but physicians specialized in a branch of medicine frequently have knowledge on only a narrow range of drugs and diseases; however, their opinion may be invaluable if the AE comes within their speciality. One possibility is to collect the investigator’s opinion in all studies except where data have been collected by a questionnaire or checklist. The investigator’s opinion on all serious AEs is essential.
Choice of alternatives

Two categories include:

- Drug related or not drug related.
- P (possible or probable) or N (not assessable or unlikely) – used by the Swedish regulatory authority.
- Is there a reasonable possibility that the event may have been caused by the trial therapy? Yes or no?

Three categories include:

- Possible, probable or certain.
- Probably not, possible, probable.
- Improbable, possible, probable.

Four categories include:

- Unlikely, possible, probable, definite (Kramer and Hutchinson, 1984; Weber, 1984).
- General list – implies unlikely, possible, probable, certain (Mashford, 1984).
- Doubtful, possible, probable, definite (Naranjo et al., 1980).
- Unlikely, possible, probable, almost certain (Stephens, 1984).
- Remote, possible, probable, highly probable (Turner, 1984; Ruskin, 1985).

Five categories include:

- Unrelated, doubtful possible, probable, almost definite, definite (Emanueli, 1984).
- Unrelated, unlikely, possible, probable, definite.
- Doubtful, coincidental, possible, probable, certain.
- Appears to be excluded, doubtful, possible, probable, very probable (French regulatory authority).
- Unrelated, conditional, possible, probable, almost definite, definite (Karch and Lasagna, 1977).
- Negative, coincidental, possible, probable, causative.
- Not related, remote, possible, probable, definite.
How many alternatives and which terms?

There are certain principles, as follows:

- The terms themselves should not require explanation or definition. Their lack of definition is an advantage.
- The investigator should not be forced into an either/or situation because of a lack of alternatives. In the clinical world there are all shades of opinion and the choices should cover the whole range.
- The more alternatives there are the narrower the use of each term becomes.
- Absolute terms, such as ‘unrelated’ and ‘definitely not’, can only be used in exceptional circumstances because they are almost impossible to prove and, therefore, should be avoided.
- No term has an absolute meaning and terms mean different things to different people. The very indistinct limits are suitable for an area where differences of opinion are extremely common and the data are very rarely reliable.

Decisions regarding collection of adverse events

Consider:

- Animal toxicology and pharmacology and potential type A effects or class effects.
- ADRs of other treatments for the same indication and possible control groups.
- The drug development plan – Phases I–IV. Consider which methods of collection would be suitable, whether translation into different languages will be necessary, and whether the results can be pooled if more than one method is being used.
- All studies should include the opportunity for spontaneous reporting and a standard question. If a diary card, checklist or questionnaire is added to these two standard methods, then have the implications been considered?
- Check with data management personnel for design of forms, coding, etc. Will the data be collected by more than one company, e.g. a clinical research organization (CRO)? If so, then problems of standardization and coordination increase.
- Does the method of collecting AEs suit the aim of the study?
- Can the method be simplified?
- Is the method consistent with methods used in the rest of the trial programme?
- Have all the essential staff involved in the study been approached?
Pre-marketing studies

The pre-marketing programme for the detection of new ADRs can be divided into:

- The general non-specific search for ADRs. This aims to detect all ADRs, not just those previously foreseen.
- The specific search for ADRs that may be foreseen for historical, toxicological, pharmacological or clinical reasons.

Non-specific monitoring

This is the search for ADRs that is undertaken for all drugs and excludes the specific search for particular ADRs that might be foreseen from previous information. An ADR may manifest itself either by subjective symptoms or objective findings, or a combination of both.

Subjective symptoms  The Phase I and early Phase II studies should aim to identify the minor AEs, which may be fairly common. If a particular AE is shown to be more common in the drug group than in the placebo group, then later studies can be planned with this in mind.

Most minor AEs are described inadequately by clinicians (e.g. ‘headache’). However, if they are recognized early in the clinical trial programme, then a specific questionnaire or form can be designed to obtain a full description. If the minor AE has some special characteristics, then its relationship to the drug may be recognized and possibly the drug will not need to be stopped unnecessarily. The questionnaire or form must, therefore, try to identify the particular clinical characteristics of the event.

The background characteristics of the patients suffering the AE must be identified to see whether a susceptible subgroup can be identified as being more at risk (e.g. the elderly, those with renal failure). Further investigation of these patients may also indicate the mechanism of action.

Treating the AE often confounds the effect of stopping the drug, so that stopping should precede treatment whenever possible. Minor events are usually completely reversible on stopping the drug, but it is helpful if the speed of reversibility can be ascertained.

Objective findings  These are usually covered by standard laboratory investigations and the standard clinical investigations, but then are of special importance when considering the effect of the drug on the course of any chronic diseases that may be present in addition to the primary disease. In these circumstances, it is important to measure the effect of the drug on chronic disease and the parameters usually used for its diagnosis and prognosis.

Specific monitoring

Some ADRs may be predicted because of the known pharmacology of the drug, experience with drugs of the same class, animal toxicology, or previous use in humans. Their discovery will be a function of the number of patients studied and the investigations undertaken.
Phase I studies

The purpose of Phase I studies is to obtain information on:

- initial safety
- tolerability
- bioavailability
- pharmacokinetics
- drug metabolism
- drug interactions
- proof of principle.

Certain physiological parameters that measure the function of an organ with limited power of repair and regeneration are ideally evaluated in Phase I. These include the special senses and the central and peripheral nervous system (i.e. ophthalmological screen, audiometry, etc. (Goldberg et al., 1975).

Volunteer studies uncover the following types of ADR:

- ADRs masked or modified in patients by disease or concomitant medication.
- ADRs likely to be induced by higher doses of study drug (i.e. in dose escalating studies).
- ADRs that are extensions of the pharmacological action of the drug.
- ADRs related to interaction between drug and common events in normal life, e.g. alcohol (Idänpää-Heikkilä, 1983).

Rarity of serious adverse drug reactions in volunteers

Serious ADRs are extremely rare in volunteers. In a survey in the USA there was only one drug-related sequel in one of 29 162 volunteers used over 12 years and only one clinically significant medical event occurring in every 26.3 years of an individual volunteer’s participation (Zarafonetis et al., 1978). However, in 1985 there were two deaths. The first occurred in Dublin, when, unbeknown to those in charge of the study, a volunteer had received a depot injection of flupenthixol the day before receiving an injection of trial drug. The volunteer had not revealed that he was under a psychiatric clinic (Darragh et al., 1985). The second death occurred in Cardiff and was due to aplastic anaemia 9 months after taking part in the study of a new benzodiazepine. It was not possible to say whether the disease was due to the investigational drug (Anon., 1985). These two deaths resulted in a re-examination of the problems of research in healthy human volunteers.

More recently, a 24-year-old female research volunteer at Johns Hopkins University died of adult respiratory distress syndrome in 2001, following inhalation of hexamethonium used to block bronchial nerve ganglia in an asthma study. The Food and Drug Administration
(FDA) said that Johns Hopkins had not sought FDA approval to use an unlicenced drug and that they had failed to report a previous unanticipated ADR (persistent cough) in an earlier volunteer (Altman, 2001). The report of the internal investigation into this death has been published: www.hopkinsmedicine.org (report of internal investigation).

The Association of the British Pharmaceutical Industry (ABPI) published the results of a 1984 survey of its member companies on their experience in this area, including both in-house and external studies (Royle and Snell, 1986). The number of serious suspected reactions in in-house studies was five (0.27/1000 subjects exposed) and in external studies it was eight (0.91/1000). A reappraisal in 1986 from the USA cited a further death in a volunteer who had anorexia nervosa unbeknown to the investigator. The author of that paper, therefore, suggests that investigators ask themselves a series of questions prior to the study:

- Do some studies pay subjects so much that they are willing to give inaccurate histories?
- Are you so busy that you do not actually participate in screening subjects or in conducting the study?
- Is the study conducted in an environment and with appropriate medical supervision to respond to a medical emergency?
- Is the research question asked either trivial or predictable in outcome?
- Should the research question be asked in the population for which the drug is intended?
- Have you made adequate provision to cover medical expenses of the subject and liability expenses for yourself? (Powell, 1986).

A survey among members of the British Pharmacological Society in 1987 showed that 69 per cent of volunteers (n = 8163) had AEs. They were moderately severe in 0.55 per cent and these, in order of frequency, were postural hypotension, abdominal pain, nausea and vomiting, palpitations, bronchoconstriction, drowsiness and headache. There were three severe life-threatening effects; viz. anaphylaxis, perforated duodenal ulcer and a skin reaction, but all made a complete recovery. Orme et al. (1989) suggested that GPs should be given exclusions for studies and thereby prevent those who match an exclusion criterion from entering the study.

Another study from France in 430 volunteers had an overall incidence rate of 13.5 per cent, covering 69 different types of AE. Severe AEs accounted for only 0.36 per cent of AEs. A total of nine deaths and life-threatening AEs had been reported in clinical research up to 1992 (Sibille et al., 1992). The same group have updated this study with 1015 volunteers. The incidence rate for AEs with active drugs was 13.7 per cent and for placebo it was 7.9 per cent (Sibille et al., 1998).

The ABPI, in 1998, issued a booklet that summarizes the situation, Medical Experiments in Non-patient Human Volunteers. The Royal College of Physicians’ report on research in healthy volunteers was published in 1986 (RCP, 1986).
Problems with including women in Phase I studies

It is not generally expected that reproductive toxicology studies will be completed before Phase II studies. Therefore, it is essential that adequate reproductive precautions are taken after counselling. Current reproductive toxicology information must be given to the participant. This should cover the risk of study participation, the importance of adequate contraception, lists of contraceptive choices with risk–benefit drug interactions, information on teratology and the need to avoid pregnancy during drug exposure (Johnson-Pratt and Bush, 1996).

The new drug must be shown not to be an enzyme-inducing agent, lest the volunteer/patient on oral contraceptives should become pregnant whilst on the new drug. If this is not possible, then the contraceptive pill cannot be relied on. Normal contraceptive failure rates are high; therefore, females of childbearing potential should have a pregnancy test and a test for recent occurrence of ovulation before and at appropriate intervals during and at the end of the study (Larson et al., 1982).

A survey of 28 clinical pharmacology units showed that whereas 79 per cent conducted clinical trials in women of childbearing potential only 86 per cent did pregnancy tests routinely at screening and immediately before the first dose and only 18 per cent did so at the post-study visit. Although 91 per cent specified that volunteers should be using a reliable form of contraceptive, only 9 per cent on the oral contraceptive were asked if they had missed any pills, and of those using other methods only 36 per cent were asked if they had used adequate contraception while in the study. The following recommendations were made:

- All female volunteers should have their menstrual history taken at screening.
- All post-menopausal females should have their hormone levels checked at screening.
- All women of childbearing potential should be tested for pregnancy at screening and this should be repeated at the beginning of each study session and again at the post-study visit (Higginbotham and Rolan, 1997).

Women in Phase I studies have 2.3 times the frequency of AEs as men, with a higher percentage due to laboratory abnormalities (males:females, 15 per cent: 26 per cent) (Vomvouras and Piergies, 1995). In addition, the presence of women suffering from premenstrual tension could increase the background noise in the control and trial drug groups, and if distributed unevenly between the two groups in controlled studies this could produce misleading ADR data. This can be dealt with by stratifying the randomization by sex.

Difference between adverse events in volunteers and patients

AEs occurring in healthy volunteers need to be considered separately from those occurring in patients for the following reasons:

- Due to absence of disease. The volunteer does not have the disease that the drug can correct. Therefore, there may be an exaggerated pharmacological reaction in volunteers that would not occur in patients or would occur to a lesser extent, thus producing either objective or subjective effects, e.g. hypertension.
• Some of the AEs in volunteers are due to the environment in which the studies are conducted, e.g. no caffeine, confinement.

• Due to incorrect dosage. The dosage used may be outside the subsequent recommended therapeutic range.

• Due to different age, intelligence and psychological make-up. Excluding the absence of disease, volunteers are likely to differ from patients who subsequently take the drug.

• Due to a different relationship with the investigator. The relationship between the volunteer and the person responsible may be that of a junior (the volunteer) to a senior investigator, and this could affect the reporting of AEs.

Because of these differences between patients and volunteers, it has been advocated that volunteer patients should be used more often (Oates, 1972; Weissman, 1981).

**Screening**

The screening of new volunteers before they participate in clinical studies is essential and will cover medical history of the volunteer, history of allergy, family medical history, smoking and drinking. Inclusion criteria will cover age, weight, fluency of literacy in relevant language, absence of significant physical abnormality, laboratory examination for hepatitis and for drugs of abuse plus the usual pre-study laboratory screen (clinical chemistry, haematology and urinalysis) (Jackson, 1990a,b), ECG, ophthalmological examination and chest X-ray. These are followed by exclusion criteria relating to abnormalities in the medical history, alcohol and current treatment, if any.

**Drug abuse**

Screening for drug abuse is essential. In the USA 7.7 per cent of volunteers used illicit drugs, 5.8 per cent cannabinoids, 3.6 per cent amphetamines, 1.2 per cent barbiturates, 1.1 per cent cocaine, 1.0 per cent opiates and 0.3 per cent benzodiazepines (De Vries et al., 1991). Screening for drugs of abuse would normally be carried out at least at the pre-study medical and randomly thereafter. They should cover morphine, amphetamines, cocaine, cannabis, benzodiazepines and barbiturates.

**Electrocardiography**

Normal values for ECGs in healthy volunteers have been published (Adamson et al., 1998). The value of 24 hour continuous ECG monitoring was shown in a paper from Simbec Research Ltd, where the use of a 24 hour Holter monitoring in 57 healthy male volunteers aged 18–46 revealed:

• sinus bradycardia in 53
• sinus tachycardia in 54
• sinus arrhythmia in 30
• ventricular ectopics in 25
• atrial premature beats in 43
• first-degree block in seven
• pause of more than 2 seconds in three
• second block in five
• atrial bigeminy in one
• two consecutive unifocal ventricular ectopics in two
• three consecutive unifocal ventricular ectopics in one

The authors recommend 24 hour ambulatory monitoring before drug treatment with new chemical entities (Barrington et al., 1990).

Screening of 156 volunteers found that only 20 (13 per cent) had normal sinus rhythm throughout, 83 per cent had supraventricular ectopics, 11 per cent had ventricular ectopics, 2 per cent had unsustained ventricular tachycardia and 6.5 per cent had sinus pauses. One volunteer was in atrial fibrillation throughout. The authors also gave some guidelines for the management of ambulatory cardiac monitoring in volunteers (Stinson et al., 1995).

**QTc interval prolongation**

Extension of the QTc interval on the ECG has been an issue of much debate in recent years following withdrawal of several non-cardiac drugs, e.g. terfenadine, astemizole, droperidol and cisapride, which caused QTc prolongation to such a degree that potentially life-threatening or fatal ventricular arrhythmias, e.g. torsades de pointes, may occur. These were either with the drugs alone or as a result of drug–drug interactions (De Ponti et al., 2000). See also Chapter 2.

Phase I studies should be directed at the early identification of a change in repolarization, usually QT or QTc (corrected) prolongation. Important questions include (Haverkamp et al., 2000):

• What electrocardiographic repolarization signal should be measured?

• What threshold QTc signal is of concern and what are the clinical implications of a small, yet statistically significant, QTc prolongation?

• Since QTc quantifies a complex relationship between the duration of ventricular repolarization and heart rate, what are the heart rate correction issues for drugs that slow or accelerate heart rate?

• Which heart rate correction is preferred, e.g. Bazett, Fredericia?

The European Committee for Proprietary Medicinal Products (CPMP) have published guidelines on monitoring of the QTc interval during early clinical studies (CPMP, 1997) as have the International Society for Holter and Non-invasive Electrocardiology (ISHNE),
October 2001 (http://www.ishne.org) and more recently the FDA (2002) have published a preliminary concept paper for discussion. Malik and Camm (2001) have written an extensive review on this subject and its implications for drug approval and labelling.

**Screening for viral infectivity**

Taking a good medical and social history and clinical chemistry will detect potential volunteers in high risk groups, and many centres would routinely exclude such subjects from participating in clinical studies (Harry *et al.*, 1995).

Serological tests for hepatitis B surface antigen (HBsAg) are required to exclude those with a positive test who may risk infecting others. Three volunteers had acute hepatitis B and another had positive serology following a trial where this screening was not done. One of the volunteers was a carrier who infected the others, probably via contamination of the gloves worn by the staff (Mehlman *et al.*, 1994).

Screening for previous hepatitis C infection is also required prior to participation in Phase I clinical studies.

Although screening for hepatitis B and C is usual, there has been much more debate about screening routinely for HIV because of the perceived consequences for potential volunteers of a positive test (Thompson *et al.*, 1993). It has been recommended that all volunteers should be tested (Sanchez *et al.*, 1994; Vickers *et al.*, 1994), but this policy has not been widely adopted. An academic unit reported that all of its 500 subjects tested negative for HIV1 and HIV2 antibodies, despite drawing 41.8 per cent of volunteers from South Africa (Jagathesan *et al.*, 1995).

Because of the low rate of genuine positive results in the ‘healthy volunteer’ population the problem of false positive results is relatively greater than for patients.

**Monitoring during the study**

Phase I studies can be separated into two different categories:

- Studies that occur between first dose in man and the first patient studies.
- Studies that are carried out later in the programme in parallel with Phases II and III.

Studies that occur in parallel with Phases II and III contribute to the general AE profile but are usually carried out to address specific questions, such as the effect of renal impairment on pharmacokinetic profile. These studies are not considered further.

Phase I studies that occur prior to initiation of patient studies (hereafter described as ‘early Phase I studies’) have as their primary aim the establishment of the maximum tolerated single and multiple doses, and hence the likely therapeutic margin. The therapeutic margin is generated based on the maximum tolerated dose compared with the efficacious plasma concentrations or pharmacodynamic end-points. This will be vital information in determining whether a drug progresses beyond initial patient studies.
Safety end-points in early Phase I studies

Adverse event recording In many ways, AE recording during early Phase I studies is the same as recording AEs at other stages of drug development. The same issue of the weakness of data in open studies applies, and early studies should be carried out as double-blind placebo-controlled studies. During dose escalation the AE profile and other safety parameters should be reviewed between each dose level and provision made for unblinding of the data if safety concerns arise.

A specific feature of the early Phase I studies is the lack of prior human experience with the drug. In contrast to the later studies, the only data on which to predict the AE profile are data from animals, which has limited predictivity; see Chapter 3. The main burden of data comes from toxicology studies, but an understanding of the pharmacology is also important. Furthermore, there needs to be an appreciation that certain AEs, e.g. nausea or dizziness, may be hard to detect in animals and may occur at lower doses than associated findings in animals, e.g. where only vomiting and gross unsteadiness are observable.

Vital signs and electrocardiogram recordings These are a standard component of early Phase I studies. It is usual to make frequent recordings of vital signs, i.e. pulse and blood pressure, during early Phase I studies; the timing of such recordings is determined from an understanding of the predicted pharmacokinetic profile and a knowledge of the safety pharmacology studies, which will have documented the occurrence of changes in vital signs in animals at low multiples of the therapeutic dose. Often, the standing, as well as lying, blood pressure is recorded to look for postural changes. Vital signs may be expanded to include respiratory rate and/or temperature, depending on the observation in safety pharmacology studies and the known pharmacology of the compound being tested.

Likewise, ECG recording will often be frequent, depending on preclinical findings, and follow regulatory guidance (CPMP, 1997). Holter monitoring for 24 hours is recommended.

Safety bloods It is usual for safety bloods to be taken during the screening process and to include full blood count, clinical chemistry and urinalysis before dosing and then at 24 and 48 or 72 hours after dosing.

Pharmacodynamic end-points It may be appropriate to include safety assessments that evaluate predictable pharmacodynamic end-points, such as cognitive function testing after dosing a central nervous system (CNS) active compound. Increasingly, genomics, proteomics and metabonomics are being used to assess pharmacodynamic effects in early studies and these may, through advanced technology, be able to evaluate patterns of change across a range of assessed parameters.

Possible improvements in adverse event collection from volunteers It is important that maximum use is made of the opportunity to collect AEs efficiently. This may be improved by:

- A list of specific symptoms, depending upon the pharmacology and toxicology – a VAS could be used (Jackson, 1990a,b). The use of a questionnaire so early in the clinical trial programme means that only predictable ADRs can be covered, unless a very exhaustive questionnaire is used. There is always the concern that new minor AEs may not be
mentioned when a questionnaire is used. The relationship in a Phase I unit between staff and volunteers is usually such that minor AEs are mentioned with the use of a single standard question (Jackson, 1990a,b).

- Use of an anonymous questionnaire in controlled studies on possible sexual effects. The completed questionnaire, identified only by a number, can be placed in an envelope to be given to a person outside the unit, who also has the randomization code.

- The use of a follow-up question 24 hours after a single-dose study.

- The use of a standard question in all studies, i.e. ‘Have you noticed any physical or mental changes during the study?’

- Rechallenge – the consideration of rechallenge using placebo and active drug in volunteers with non-serious type A symptoms only.

- An initial single-blind placebo period, which might reduce the number of placebo reactions whilst in the controlled part of the study. In a small study using an ‘All Body Organs and Functions’ (ABOF) questionnaire there were 13 AEs in the initial placebo period and nine during the controlled part of the study (Nony et al., 1994)

The Institute of Medicine in the USA has published a comprehensive report recommending better protection for participants in clinical research. The report, Responsible research: a systems approach to protecting research participants, is available at http://national-academies.org.

**Phase II studies**

The purpose of Phase II studies is to test whether the new drug is effective for one or more clinical indications and to determine doses for further study. These studies should only include patients with the target disease, and preferably those with no other concomitant disease. The number of patients involved is relatively small. These studies should detect the most frequent ADRs and may predict target organ systems for other ADRs to be found later in Phase III studies. They are seldom able to define any precise or comparative incidence of ADRs or discover ADRs that typically appear in a subgroup of patients (Idänpää-Heikkilä, 1983).

The monitoring of patients in Phase II studies is similar, whenever possible, to that used in Phase I studies. Where possible, there should be two laboratory examinations before starting treatment. The first should be 1–2 weeks before the study, so that the results are available before treatment is started and patients can be excluded if necessary. The second is immediately before treatment is taken on day one. Problems met in early clinical trials have been described in the CIOMS (1983) review of safety issues in early clinical trials of drugs:

- Collaboration between sponsor and investigator may be less interactive than is desirable.
• Potentially serious ADRs may be poorly documented and the relevant forms not completed properly.

• Patients who do not fulfil the selection criteria may be admitted to trials, making interpretation of potential AEs difficult.

• Pretreatment assessment is inadequate and incomplete and, therefore, it is difficult to interpret AEs occurring during treatment.

• Appropriate actions that are needed to evaluate a potential ADR properly are often not instituted.

• Treatment may be instituted before appropriate safety assessments have been taken or before the results have been received.

• Rechallenge after an AE may be undertaken without appropriate monitoring.

• Parochial attitudes may be adopted and the sponsoring organization and experts in the evaluation of ADRs may not be involved until very late.

• Occasionally, ADRs are published without adequate investigation and may lead to problems in determining the true situation.

• Often, there seems to be a lack of appreciation of the regulatory and legal implications of handling ADRs sensibly.

These can be overcome by the protocol and CRF clearly describing the exact procedures to be undertaken if there is an AE including a rigorous follow up procedure for withdrawals. This emphasizes the need for good and thorough protocol and CRF design and development.

A single serious AE in an early Phase II study may abort an entire new drug programme if sufficient information is not collected and a proper assessment is not made. Spilker (1984) covers the various safety parameters that can be measured in clinical studies covering laboratory tests, ophthalmological testing, psychological and performance tests in his guide to clinical studies and developing protocols.

The death of a clinical trial patient during the pre-marketing phase may be due to many reasons, but if due to the drug then it may be disproportionately catastrophic for the future of that compound. The steps to take in these circumstances are dealt with by Cato and Cached (1988).

Some ECG monitoring will continue in Phase II studies and maybe into Phase III to confirm safety in the target population with co-morbidities and concomitant medications. The nature and extent of this will be based on the type of drug and the ECG findings, including QTc prolongation, from Phase I; see above. ECGs should be statistically analysed and reported according to predefined standards; Steare and Morganroth (2002) advocate centralized ECG collection and analysis.

For chronic treatments there is a requirement to have treated at least 100 patients for 1 year before a marketing application. A survey of 27 drugs showed that only 4 per cent (25
serious AEs) of first occurrences appeared in the second 6 month period. Of these 25 ADRs there were 13 type A reactions; the ADRs in this context were AEs possibly, probably or definitely due to the drug (Brown et al., 1996). It is sensible, therefore, to try to recruit patients from Phase II studies into long-term extension studies.

**Phase III studies**

The FDA comments concerning the pre-marketing programme and the resultant New Drug Application (NDA) (Temple, 1991a) were:

- Sponsors exclude patient populations deemed ‘too sick’ even when the drug is clearly going to be used in such populations.

- Some safety databases are not adequate enough to support the planned dose or the planned duration of use.

- Existing databases are not examined because doing so would require more time.

- Sometimes there is complete indifference to finding the right dose of the drug.

- There is a tendency to focus on the good effect of the drug and to forget that it may also have other less desirable effects that need study.

A subsequent review in 1995 of the 1993 cohort of new molecular entities approved by the FDA said that the FDA did not receive data analysed in the way it would like to have it. With ‘nuisance’ AEs there were seldom any subgroup summaries, by age, gender, race, or ethnicity, and analyses seldom accounted for the length of exposure or dose. If the AE limits the dose or causes discontinuation, then, although reported, it was seldom analysed statistically. There was also little assessment of the range of discomfort or the timing relative to the start of treatment or relative to the disease state.

With serious AEs there was usually no attempt to analyse them statistically and they were not discussed in relation to the patient’s condition, demographics or pharmacokinetics. The time course of therapy and the AE were usually not examined systematically. As far as laboratory data were concerned, movements within the normal range were generally ignored (Fairweather, 1996).

In March 2003, the FDA published for comment a Concept Paper on pre-marketing risk assessment that focused on clinical development, particularly Phase III studies. The design of the trials programme is critical to ensure that sufficient safety data are generated to allow for product approval, proper risk management and to inform post-marketing safety assessment. The size of the safety database supporting a new product depends on its novelty, the intended patient population, proposed indication and intended duration of use.

Ideally, programmes should include long-term controlled safety studies to allow for comparisons of event rates and for accurate attribution of AEs, particularly for detecting changes in rates of frequent events in the population and especially when the AE could be part of the disease being treated. The safety database in Phase III should include a diverse population, with only patients with obvious contraindications excluded. Broadening inclusion criteria could enable the findings to be more generalized to the population likely to
receive the product post-marketing. Using a range of doses in Phase III trials would
characterize better the relationship between exposure and the resulting clinical benefit and
risk.

Potential interactions (drug–drug, drug–demographic, drug–disease, drug–food and
drug–dietary supplement) should be addressed during controlled trials and specific safety
studies. It is recommended that the potential for the following serious safety effects be
assessed as part of all drug development programmes:

- QTc prolongation
- Liver toxicity
- Drug–drug interactions
- Polymorphic metabolism

Temporal associations between drug exposure and AEs are seen as critical to risk
assessment, as they provide clues for determining whether the AE was related. Time-to-
event analyses are appropriate for clinically important events that occur on a delayed basis
and AEs that occur at initiation of treatment but diminish in frequency over time.

**Patient subgroups**

The summary of product characteristics (SPC) or product labelling will be primarily based
on the results of the Phase III studies. The inclusion and exclusion criteria in Phase III
protocols should, therefore, reflect the intended labelling for the new product. If all patients
over 65 have been excluded from the pre-marketing trials then they should be excluded in
the SPC. At least two specific subsets of patients will require specific monitoring:

- Elderly patients.
- Patients in whom disease will modify absorption, distribution, metabolism or excretion
  of the drug, though these tend to be the very patients who are excluded by the protocol
  in randomized clinical trials (Riegelman, 1984).

If the drug is intended for chronic usage, then a minimum of 100 patients should be
followed up for 1 year. However, if a subgroup of 100 patients shows no serious AE, then
how certain can one be that no serious ADR will occur in that subgroup in subsequent use?
The following simple rules calculate the risk for different size subgroups:

- Rule 1 – if none of $n$ patients shows the AE then we can be 95 per cent confident that
  the chance of the event is at most $3/n$ (i.e. $3/n$). The two corollaries for this rule are:
- Rule 2 – for the 99 per cent confidence interval the figure is 4.6 (i.e. $4.6/n$).
- Rule 3 – for the 99.5 per cent confidence interval the figure is 6.9.

The aptly entitled paper ‘If nothing goes wrong is everything all right’, which contains these
rules, is worth reading (Hanley and Lippman-Hand, 1983). For any material change in risk
discovery the increase in size of the denominator should be in the order of magnitudes (i.e. 1000, 10 000, 100 000, etc.).

Idänpää-Heikkilä (1983) said that Phase III studies lasting 2 weeks or less uncover the most frequent and the acute ADRs, but that for ADRs with longer latent intervals these trials need to be extended to 3 months. There will still be some important ADRs that will only be detected in studies lasting 6 months. All long-term studies are dogged with increasing numbers of withdrawals, and the protocol must make sufficient provision for a determined effort to establish their fate. The diminishing numbers can give a false impression of the incidence of ADRs with a long latent period, and life table analysis should be used to cope with this situation (O’Neill, 1988). The advantages put forward for the method are:

- It permits estimation of the cumulative ADR rate over a specific time interval.
- It handles losses from observation adequately.
- It allows determination of whether a time-specific ADR is occurring at a constant, decreasing or increasing rate.
- It may allow combining safety data from more than one clinical trial. Severe and easily recognized ADRs will be as easily detected in uncontrolled as in controlled studies (Idänpää-Heikkilä, 1983).

**Controlled trials**

**Protocol**

The protocol needs to consider all aspects of the clinical trial and it is useful to use a checklist to make certain that all the essentials have been considered. The following are the questions concerning ADRs that should be considered when writing the protocol. A safety monitoring evaluation plan should be in place before the start of any clinical trial.

**Aim**

- Do the aims of the study accurately describe the purpose of the trial as far as ADRs are concerned?
- Can estimates of the expected incidence of death and serious AEs in the study population, based on the disease and/or concomitant medications used to treat the disease, be provided?
- Any death or serious AEs that exceed these estimates would presume a drug relatedness and require notification of the regulatory authorities; remember fialuridine (Nickas, 1997).
**Patient selection**

Does the selection of patients bias the study as far as ADRs are concerned?

- Have patients who have previously had one of the trial drugs been excluded? If not, how will they be dealt with? There is more likely to be a problem in comparative controlled trials in chronic diseases where the alternatives are the new drug and the standard therapy. A patient who has already had the standard therapy without any problem is very unlikely to have an ADR in the study, whereas the patient who had developed an ADR while previously on the standard therapy will be excluded from the study. The statistical analysis of the trial results can weight the effect of previously having the standard against not having had it. Exclusion of these patients from the study could have an adverse effect on trial recruitment. The CRF must, therefore, have a space with the questions ‘Has the patient ever had any of the trial medications? If so, with what result?’ In hospitals there has been inadequate history of ADRs taken (Cook and Ferner, 1993; Shenfield et al., 2001).

- Has the background noise been reduced to a reasonable level?

- Does the choice of investigator or hospital bias the selection of patients as far as type, severity or resistance to treatment of the target illness?

- If the target illness is chronic, then will the clinical features and relevant laboratory investigations be shown to be stable before the study?

- Are the inclusion and exclusion criteria at the right level so that the trial results can be extrapolated to a reasonable population of patients on the drug (e.g. women of childbearing age, etc.).

**Trial design**

Features to consider include the following:

- If the underlying disease is likely to produce AEs that might be confused with drug toxicity, then consider the use of a formal control group.

- If the study is not double-blind, will the lack of blindness bias the occurrence/collection of AEs in favour of one of the trial drugs?

- Will the known ADR of the trial drugs unblind the study? If so, how will this problem be overcome? Any controlled clinical trial where the type A ADRs of the drugs might allow the patients or investigator to identify which drug they are taking should be assessed for maintenance of blindness (see 4th edition).

- The protocol should state under what circumstances the treatment code can be broken by the investigator for medical emergencies.
Does the consent form adequately represent the risks of the study?

Does the protocol detail notification of serious AEs to the trial Ethics Committee/Institutional Review Board (IRB)?

Should a Data Monitoring Committee be used and how should it operate (Ellenberg et al., 2002)? See also Chapter 11.

**Concurrent therapy**

Factors to consider include:

- What concurrent therapy should be permitted or forbidden?
- Is sufficient provision made for recording of concurrent therapy on the CRF?
- With the increasing number of drugs changing from prescription-only medicines (POM) to over-the-counter (OTC) drugs there should be special provision made to record all OTC drugs.
- Herbal remedies.

**Baseline characteristics**

It is important to document any characteristic that might be important in the analysis or assessment of any associated factors of a TESS. These may involve areas that are very sensitive for the patient, who may be reluctant to mention them spontaneously or even if prompted. They include:

- OTC products.
- Herbal medicines.
- Other alternative or complementary treatments, e.g. reflexology, aromatherapy, vitamins, etc. In 1997 42 per cent of the American population used alternative medicines (Eisenberg et al., 1998).
- Alcohol consumption. Abnormal liver function tests (LFTs) during a study may be related to alcohol excess or abuse. The use of the Alcohol Use Disorders Identification Test (AUDIT) is probably sufficiently sensitive (89 per cent) and specific (91 per cent) if a cut-off score of >8 is used (Hearne et al., 2002). Moderate alcohol consumption in elderly patients is associated with an increased risk of ADRs (Onder et al., 2002).
- Mental problems.
- Sexual orientation. This is sensitive and its documentation will only be necessary in a very relevant indication.

The patient should be asked to give consent that their GP be contacted to enquire whether there is any reason known to him/her why the patient should not be entered in the study.
Adverse events (symptomatic)

Questions to ask include:

- Have the AEs to be collected been defined?

- How are the AEs to be collected:
  (i) Diary card.
  (ii) Questionnaire with or without analogue scale. Has the questionnaire been validated? (Offerhaus, 1979).
  (iii) Checklist.
  (iv) Standard question. Is the wording in the protocol and CRF?
  (v) Quality of life?
  (vi) Other?

- Does the protocol require the investigator to investigate fully all AEs, including seeking the aid of specialists where necessary? There is sometimes failure of companies to follow up serious AEs sufficiently. It is vital to obtain all the data that must be collected and would help decide on the cause of the AE.

- Does the protocol request the physician to make any interim diagnosis or drug attribution before breaking the drug code?

- Does the protocol allow for a sample of blood to be taken for drug levels in the case of serious AEs?

- Does the protocol inform the investigator that all serious AEs must be notified immediately to the company?

- Does the CRF allow sufficient space and require sufficient details for assessment of all types of AE?

- Does the protocol request full follow up on patients who have had a serious AE, stopped a trial drug due to an AE, or who have a laboratory abnormality or an AE at the last visit while on the drug?

- Does the protocol require full details of treatment of any AE to be recorded?

- Is the frequency of trial visits adequate to pick up the AEs and is the timing of the post-study visit suitable considering the disease and the drug?

- Who is to assess the causality of the AE?

- How are the AEs to be analysed and compared? Clinically only? Statistically? If the latter, then how?
Adverse events (asymptomatic)

Questions to ask are:

- Do the laboratory and other objective investigations cover potential ADRs?
- Is the frequency of sampling adequate and has a post-study sampling been agreed?
- How is the handling of patients with asymptomatic abnormal laboratory investigations to be dealt with (Sackett and Gent, 1979; DeMets et al., 1980):
  (i) By repeating tests?
  (ii) By further confirmatory tests?
  (iii) By clinical examination?
  (iv) By dechallenge?
  (v) By rechallenge?
- How are the laboratory examination results to be assessed:
  (i) According to the normal laboratory range as normal or abnormal?
  (ii) By the clinician and/or company physician?
  (iii) Clinically as well as statistically?
- Will the samples be analysed centrally in a multi-centre trial? If analysed in the hospital where the trial takes place, has the laboratory been approached? Have the details of storage and transport been decided? See Chapter 5.

Withdrawals

How are withdrawals to be investigated/followed up to make certain that the cause was not drug intolerance? Other questions to consider are:

- Is there a financial disincentive for the investigator to follow up withdrawals (i.e. withdrawals not paid for)?
- Is there provision for following up patients who change GP or move, so that they can be contacted for long-term follow up?
- If the study is in the UK and is long term, have arrangements been made with the NHS Central Register in Southport for patients to be flagged so that deaths will be identified (Cancer Research Campaign Working Party, 1980)?
- Is there a procedure for follow up of withdrawals?
**Third-party ‘interference’**

Have sufficient arrangements (e.g. provision of a letter to be held by the patient) been made to ensure cooperation with the management of the patient during the trial, but outside the confines of the trial? Relevant factors to consider are:

- Provision for notification of trial with request to GP for cooperation.
- Provision for possible emergency admission to another hospital.
- Recording the use of OTC products by the patient.
- Provision of a card for the patient giving details of the trial and contact telephone numbers/addresses, requesting information from any doctor consulted.

**General**

What is the possibility that the trial will fulfil its aim regarding ADRs (i.e. what is the size of the type 2 error)? This involves considering:

- What incidence of adverse reactions (95 per cent confidence limits) appearing with only one of the comparative drugs could be detected?
- Can measurable safety outcomes be defined (e.g. a laboratory value above a defined point)?
- What difference in incidence of AEs in the investigational drug group compared with the control group, or the expected incidence of AEs from previous epidemiological studies, would the trial detect?
- Has it been agreed that adequate space will be given to the reporting of AEs in any subsequent papers (Ioannidis and Lau, 2001), including statements about the power of the study?
- Will all patients who have been randomized and taken even a single dose be analysed and accounted for as far as AEs are concerned (May et al., 1981)?
- Since all pre-marketing studies have some risk to the patient, all studies need to be monitored for any undue risk and in studies of any length the possibility of interim analysis must be considered. The protocol should clearly specify safety variable outcomes that could necessitate patient discontinuation from the study or complete study termination (Enas et al., 1989).
- There has been increasing use of CROs for carrying out part of the development of a drug and also an increase in the licensing out of products to other manufacturers. Inadequate preparation for these multiple sources of AE data can result in chaos when finally it is all put together for a licence application. Some form of template is necessary to develop a detailed agreement covering AE collection, processing, distribution and
reporting (Fieldstad et al., 1996; Society of Pharmaceutical Medicine, Pharmacovigilance Group Working Party, 1998).

Uncontrolled trials

These may vary from the use of the drug by a physician in a single patient resistant to other therapy to relatively large-scale dose–titration studies. All these studies should be governed by a protocol. All patients must be accounted for and detailed records kept (as for controlled studies).

From the point of view of ADRs, uncontrolled trials pose several problems:

- Without a control group the AEs that are symptoms alone and can occur in normal persons without drugs often cannot be attributed to the drug.

- Since the type of patient admitted to the study may not be tightly controlled, patients are entered with disease or complications other than the target disease, giving rise to the difficulty in deciding whether an AE is due to the natural course of the concurrent disease (or a complication) or to the drug.

- Concurrent therapy is often permitted in these uncontrolled studies and therefore attribution of AEs to the investigational drug may be difficult.

- The consent form, if it lists potential ADRs, may well bias reporting of AEs (Levine, 1987; Myers et al., 1987).

An unaccountable fatal outcome occurring in an early phase uncontrolled study may, quite unjustly, be attributed to the drug and all further studies stopped or delayed. The practice of allowing investigators involved in controlled trials to use the drug in an uncontrolled study must be strictly limited and the investigator must be prepared to monitor the patients as strictly as if in a formal controlled trial. All AEs, no matter how trivial, must be documented and the pharmaceutical company notified of any serious AE without delay. This can be done using either a specifically designed record card, which is returned at regular intervals, or by the use of a standard AE form, which is returnable immediately after the event has occurred. The latter has the advantage that the event report is not delayed until the main CRF is returned.

Recording adverse events in clinical trials

From the point of view of recording AEs the essential division is into serious or non-serious. It is unusual that non-serious AEs or symptoms can be causally related to the drug on an individual basis; however, they can be grouped together and their incidence compared statistically in treatment groups. Clinicians are only likely to supply very limited information concerning non-serious events, and an extensive AE form may be inappropriate. A delay until the CRF is returned to the company may be important for non-serious AEs if they have caused the drug to be stopped and there is a possibility that they are causally related. They may need to be assessed individually for causality and more detailed information will be required. There are unlikely to be many similar serious AEs in the two
treatment groups and they will, therefore, not be suitable for statistical comparison. The withdrawals possibly due to the drug should probably be notified to the company when they occur and, therefore, before the patient finishes the study. They will require subsequent follow up.

**Non-serious adverse events**

These are AEs that do not satisfy the definition of serious. These can be recorded on a special form, and the date of onset, duration, frequency, severity (mild, moderate or severe) and outcome should be noted. These will be assessed when the CRF is returned. The investigator must be encouraged to give full descriptions of these events, as a single word is not usually sufficient. Where the event comprises one or more symptoms with little objective data to back it up, it becomes very important for the investigator to record the patient’s own description on the CRF. The diagnosis that the investigator makes should be recorded in their own words.

**Serious adverse events**

Serious AEs should be recorded on a serious AE form, and a copy despatched immediately to the company. The investigator must record all serious AEs, regardless of causality, but should be asked to make a causality judgement while blind to treatment. For serious AEs the seriousness and courses of action open may demand that the code be broken and the initial causality judgement may need to be changed once the results of all tests are known. These cases must be followed up by the company. The latter must be emphasized, because otherwise the investigator may delay sending the form until he has more information, and there is always a difficulty in deciding whether to notify a serious AE as soon as the possibility arises or to wait until all the data are available and thus establish a causal relationship (Dangoumau *et al.*., 1978; Boisseau *et al.*, 1980).

**Follow up**

Follow up is vital in order to obtain:

- Causality assessment.
- Response to dechallenge and possibly to rechallenge.
- Further investigations that would help in causality assessment, but may not be required for the patient’s clinical management.
- Final outcome.

Follow up may also allow the trial monitor to help the investigator by giving previous experience with the event or suggesting reduction of dosage rather than stoppage for type A reactions.

The response to follow up requests from companies is often very poor, and this is true for Europe, USA and Japan. Some companies state that if there is no response to a follow up then it should be repeated either once or twice, but it is important to explain what
information is needed and why it is needed. There is no reason to accept an inadequate reply, since investigators are under contract to the company and it should have been specifically referred to in the protocol. There are several points that should be considered:

- A local company physician could indicate what examinations are likely to have been done and what details should be available for each AE.

- A specific questionnaire may be advisable for certain areas, e.g. liver, haemolysis, kidney, thrombocytopenia and dermatology. See Bénichou’s *Adverse Drug Reactions* (see Bibliography).

- If allergy is likely to have been involved, consider using the ENDA Drug hypersensitivity questionnaire (Demoly *et al.*, 1999).

- Consider extra payment for time spent investigating really important cases.

- Request photocopies of hospital records where applicable.

- If an AE falls outside the expertise of the investigator request him to ask the appropriate specialist to give an opinion at the company’s expense.

- A site visit by a company physician can be very effective.

- The communications line may be long: HQ drug safety → HQ clinical research → local company clinical research → local company drug safety → ? CRO → local clinical research associate (CRA).

- The follow up may need to continue if the results of dechallenge and rechallenge are to be collected, or if the event is prolonged and it is necessary to continue until complete resolution.

**Should withdrawals due to adverse events be notified before the end of a trial?**

One can question the advantages of using solely the serious criteria as a reason for immediate notification. Drug withdrawals due to AEs may indicate too high a dosage regimen and need further investigation before further patients are recruited.

Withdrawals due to AEs may be for several reasons:

- The patient or physician may have thought the AE to be due to the drug.

- The patient or physician may have thought the AE was not due to the drug, but that the AE made continuing with the drug undesirable (e.g. if the AE was renal impairment from natural causes, treatment with a renally excreted drug may be inappropriate).

- The physician may not know the cause of the event and, therefore, has stopped all treatments.
• The patient might have had several different signs and symptoms, but only one of these might be the actual reason for stopping the drug.

• The drug may have been ineffective.

• The dosage of the drug may be incorrect. This may be due to poor prescribing or the Investigator’s Brochure (IB) gives a dosage regimen that is too high.

If type A ADRs due to an incorrect dose for an individual patient cause a patient to withdraw from a Phase II study, their notification at the end of the study may be too late for a change in dose for Phase III studies. This may be particularly relevant to indications requiring long-term therapy. Similarly, notification during Phase III studies would allow studies at the correct dose to run in parallel so that early in the post-marketing phase a lower dose would be available. It would also be sensible to take blood for drug levels in any patients complaining of type A events during Phase I and Phase II studies.

Incorrect dosage at marketing requiring change after marketing

Patients can experience ADRs that may have been preventable if a better estimation of the dose at marketing had been possible.

1. Post-marketing dosage changes over the period 1982–2000 showed 115 changes – 39 per cent increases (predominantly in the 1980s) and 61 per cent decreases (later on, mostly post-1993) (Heerdink et al., 2002).

2. Of 499 new medical entities (NMEs) approved by the FDA between 1980 and 2000, 354 were evaluable; of dose changes with 73 NMEs, 58 were safety motivated (79 per cent). ‘Post-marketing changes to labelled dosage regimens may reflect suboptimal drug development’. ‘The rate of these changes is greater for newer drugs than older drugs’ (Cross et al., 2002).

3. Of drugs approved in the USA since 1979, 20 per cent required dose reduction after approval (Bashaw, 1992).

The reasons put forward were:

• The dose is commonly fixed at the level that has been shown to be effective in 90 per cent of the population provided that the unwanted effects at this dosage are considered acceptable. In 25 per cent of patients a smaller dose will be effective.

• Digit preference (e.g. a correct dose of over 50 mg may be rounded up to 100 mg).

• To avoid dose titration (Herxheimer, 1991; Venning, 1991).

• Once Phase III studies have started at the wrong dose, repeating the studies at the correct dose would produce an unacceptable delay in the Marketing Authorization Application (MAA)/NDA.
• ‘The common practice of selecting the highest possible dose for use in large trials may result in an unacceptable incidence of unwanted and potentially serious adverse outcome’, e.g. Hirudin (Conrad, 1995).

• Effective low doses determined in pre-marketing studies or in post-marketing studies are often omitted from the Physicians’ Desk Reference (Cohen, 2001).

**Effects of drugs on skilled performance**

All new drugs should be tested for effects on the CNS at some stage (Jackson, 1990). The danger of drugs affecting car driving ability has attracted more attention in recent years. The use of medicinal drugs such as benzodiazepines and tricyclic antidepressants has been shown to more than double the risk of involvement in traffic accidents resulting in injuries (Ray *et al.*, 1992; Neutel 1995; Barbone *et al.*, 1998). In the EU, traffic accidents result in 50 000 fatalities and 1.5 million injuries each year, with a total cost of more than €70 billion (Cornelisissen, 1997). The contribution of medicines to these figures is substantial, since an average of 10 per cent of the adult population is frequently using impairing medicinal drugs. At a very conservative estimate, if 10 per cent of the adult population is driving under the influence of impairing medication at twice the risk of being involved in traffic accidents, then those drugs are causing 4500 deaths, 135 000 injuries and €6.3 billion damage to the European society each year (De Gier, 1995).

Determination of a drug’s effect on psychomotor performance can be derived from experimental psychopharmacology (Ramaekers, 2003). This discipline has made tremendous contributions to the development of relatively safer drugs for drivers. The methods used ranged from psychometric test batteries (e.g. the Digit Symbol Substitution Test) to psychomotor tests (e.g. reaction time, tracking, and Critical Flicker Fusion) and cognitive tests for measuring mnemonic functions. But more importantly, the simulation of real-life performance in driving simulators, closed-course driving and actual driving tests in real traffic conditions revealed the impairing properties of medicinal drugs, especially when compared with the impairing effects of alcohol in various blood concentrations. Although many studies are conducted with healthy volunteers not suffering from illnesses that might impair performance, the comparative data from studies with patients revealed that both volunteers and patients experience similar side effects of psychoactive drugs (Van Laar *et al.*, 1992; O’Hanlon, 1995; Ramaekers *et al.*, 1997).

Information concerning the increased potential for crash risk as a consequence of using potentially hazardous medicines must be meaningfully communicated to patients. The simplest way to achieve this would be by means of clear warning labels on the package. Most EU member states, however, do not require exterior warnings on packaging, and patients are informed about impairing effects only by the patient information leaflet. Since 1992, European legislation has required warnings regarding the ability to drive or use machines, written in lay language, to be part of the content of the patient drug information leaflet (Council Directive 92/27/EEC).

**European Union ‘Note for Guidance’**

A warning system based on consensus among scientists was introduced in 1991 (Wolschrijn *et al.*, 1991); the major improvement of the system was its scheme for categorizing drugs...
according to their potential for impairing driving skills. The European authorities adopted a Note for Guidance for the, SPC(III/9163/90-EN, final approval by the CPMP on 16 October 1991).

In 1998, a survey was conducted to determine how responsible regulatory authorities in the different European countries have reacted to this Note for Guidance (De Gier, 1998a,b). The report of survey findings highlights that the European Ministers of Transport resolve that EU member states should encourage the systematic printing of a warning symbol on the packaging of medicines likely to impair driving (ECMT/CM93/5/Final). The Note for Guidance provides the framework needed to categorize drugs in order to provide three different warning symbols reflecting the following categories of Article 4.7 of the SPC ‘Effects on ability to drive and use machines’, on the basis of:

- the pharmacodynamic profile, reported ADR and/or
- impairment of driving performance or performance related to driving, the medicine is:
  1. presumed to be safe or unlikely to produce an effect;
  2. likely to produce minor or moderate adverse effects;
  3. likely to produce severe effects or presumed to be potentially dangerous.

For situations 2 and 3, special precautions for use/warnings relevant to the categorization should be mentioned.

Procedures for assessing warnings and guidelines for allowing categorization

Generally speaking, the drug regulatory authorities review the data provided by the drug manufacturers and assess whether these data support statements in the SPC. There are no criteria stipulating the number of reports or the kind of tests that are needed as a basis for their assessment; it is a case-by-case assessment. Furthermore, the methodology of experimental research on drug effects is still poorly described, although adequate descriptive information exists based on the consensus of scientific opinion.

Some harmonization has been achieved, e.g. in the provision of a specific Note for Guidance on the SPC of benzodiazepines as anxiolytics; the recommendation for information to be contained in the warning about the effects of these drugs on the ability to drive and use machines has been standardized. It offers no opportunity, however, for distinguishing between various benzodiazepines when data from experimental and/or pharmaco-epidemiological research demonstrate different behavioural toxicities. Unfortunately, this situation has not been recognized by the drug regulatory authorities as being an obstacle to accurate categorization. Hence, it is recommended that better guidelines be established to assist drug manufacturers to select appropriate drug testing methodologies and to reconsider the use of standard information for the warning section in the SPC.

A major problem in the categorization of drugs may be the lack of support from pharmaceutical companies that have to submit the relevant data. Even if a standardized methodology is applied to test a drug’s impairing properties, there is still likely to be debate about the meaning of results. If a drug has been found to be impairing, then the issue will be whether it will be assigned to a different category than non-impairing drugs of the same therapeutic class.
Some drug regulatory authorities have indicated that experimental research alone is not sufficiently convincing evidence to support the formulation of different warnings. They have suggested that revision of the warning system should be based on results obtained from studies of large populations who have used the drugs. They propose that the study investigators should assess the risk potential of accident involvement for each individual drug. It is unclear whether there will be a need for the European Medicines Evaluation Agency (EMEA) to provide specific expertise in this area. Although regulatory authorities may feel much more comfortable selecting their own experts, some would welcome the specific expertise provided by EMEA. It is recommended that EMEA initiate an investigation to decide whether or not it should coordinate large-scale, case controlled pharmaco-epidemiological surveys. These would use existing databases in different EU member states to determine the relative risks of traffic accidents for users of all drugs identified as potentially dangerous.

Opportunities to improve warnings

A categorization system could improve the effectiveness of warnings, compared with the long lists of side effects that are currently provided as warnings. The new European Guideline on this subject (III/5218/97, final approval September 1998) provides an opportunity to improve the readability of the label and patient information leaflets with specific guidelines for certain categories of medicinal drugs. Clear statements are prescribed, and pictograms may be used as an additional measure if they make the message clearer to the patient.

There are current movements towards categorization systems for the purpose of improving warnings in at least five EU member states: Belgium, Germany, the Netherlands, France and Spain. Belgium was the first country that officially introduced the categorization system in April 1999, at the time that the Traffic Law was changed into a ‘zero tolerance’ law for illicit drugs (Charlier et al., 1999). Medicinal drugs were not included in this new law, but the Belgian Minister of Transport considered these to be dealt with by preventive measures, such as prescribing and dispensing guidelines and a clear patient information leaflet. Furthermore, professional organizations in these countries have applied the same system in their efforts to support physicians and pharmacists in selecting relatively safer drugs for patients who drive.

The categorization system was originally proposed in 1991 by a group of international experts who wanted a system allowing healthcare providers and patients to understand more easily the severity of impairment by medicinal psychotropic medicines (Wolschrijn et al., 1991).

In 2001, Spain became the second country in Europe to introduce an official categorization system for drugs having a potentially dangerous effect on driving (Del Rio Garcia, 2001).

In order to make the users of the categorization system aware of the meaning of each category, a comparison with the effects of alcohol, which are well known, was suggested by researchers in experimental psychopharmacology in the Netherlands, based on the views on test validation expressed several years ago (O’Hanlon et al., 1986). Data collected in experimental research, in which over-the-road driving tests have been applied with most frequently used medicinal drugs and alcohol (as ‘calibration’), have allowed researchers to
interpret weaving effects by any drug as equivalent to that produced by a particular blood alcohol concentration (see Table 4.3).

### Table 4.3  Categorization of warnings

<table>
<thead>
<tr>
<th>Category</th>
<th>Impairment description for medicinal drugs</th>
<th>Comparison with blood alcohol concentration (BAC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Presumed to be safe or unlikely to produce an effect</td>
<td>Equivalent to BAC &lt; 0.5 g/l (&lt;0.05%)</td>
</tr>
<tr>
<td>II</td>
<td>Likely to produce minor or moderate adverse effects</td>
<td>Equivalent to BAC 0.5–0.8 g/l (0.05–0.08%)</td>
</tr>
<tr>
<td>III</td>
<td>Likely to produce severe or presumed to be potentially dangerous</td>
<td>Equivalent to BAC &gt; 0.8 g/l (&gt;0.08%)</td>
</tr>
</tbody>
</table>

The most important advantage of the three-tier system over ‘older’ dichotomous drug class-based systems or systems based on quotations of long lists of side effects is the focus on the least impairing medications in each therapeutic class. Since these initiatives will have an impact on the views of patients, physicians and pharmacists, it would be advisable for them to be in accordance with present European Directives and Guidelines aimed at improving the readability of labels and patient information leaflets. Categorization and warning symbols, based on scientific consensus, have been shown to be feasible. By investigating the acceptance of a new warning symbol among patients, healthcare providers and drug manufacturers, the drug regulatory authorities could become more proactive in response to the actual needs of those who use the information presented.

**Special subgroups**

**Children**

Most drugs prescribed for children have not undergone extensive clinical trials in children, particularly in children under 2 years of age. Only between 20 and 30 per cent of approved drugs are labelled for paediatric use. Consequently, much drug prescribing in children is so called ‘off label’ use. The EMEA in October 2002 published a concept paper on the conduct of pharmacovigilance for medicines used by children (CPMP/PhVWP/4838/02), this identified the key issues as:

- Childhood diseases may be qualitatively and quantitatively different from adult diseases.
- Efficacy in children cannot always be assumed from adult efficacy data.
- Children may have different pharmacokinetics and dynamics to adults and, therefore, have particular vulnerability to ADRs.
- Children may have different drug metabolism, and consequently a different drug interaction profile compared with adults. Owing to specific ethical considerations, drug metabolism data in children may be very sparse at the time of registration.
• Children are growing and may, therefore, be susceptible to developmental disorders, as well as delayed ADRs not seen in adults.

• Certain ADRs may only be seen in children.

• Lack of clinical trials in children limits the safety data available.

• Lack of kinetics data may lead to under- or over-dosing in some age groups.

• Under-dosing may result in lack of benefit or development of resistance.

• Over-dosing may result in an increase of type A reactions.

• Lack of appropriate formulations may lead to incorrect dosing and use of products of less controlled quality.

• Children may be more susceptible to ADRs from specific excipients.

• Medicines used off-label may have inadequate product information to support safe use in children.

The EU guidance for clinical investigation of medicinal products in the paediatric population is ICH 11 (CPMP/ICH/2711/99); this has been followed in 2002 by a consultation document, Better medicines for children, proposed regulatory actions on paediatric medicinal products. In the USA, guidance for industry is also based on ICH E11, Investigation of medicinal products in the paediatric population, which was published in December 2000. There is the paediatric exclusivity provision of the FDA Modernization Act of 1997, which encourages paediatric studies by extending patent protection. This was followed by the FDA's Paediatric Rule, requiring paediatric studies starting in December 2000, but this was challenged in the courts and the FDA has issued a ‘Call for Comments’ on whether the rule needs to be changed.

The five age groups in ICH E11 are:

• pre-term newborn infants (born at <36 weeks gestation);

• term newborn infants (age 0 to 27 days);

• infants and toddlers (age 28 days to 23 months);

• children (age 2 to 11 years);

• adolescents (age 12 to 16–18 years, dependent on region).

The four groups of medicinal products are:

• To treat diseases in adults and children for which treatment exists.

• To treat diseases in adults and children that have no current treatment.

• To treat diseases that mainly affect children or are of particular gravity in children or have a different natural history in children.

• For diseases only affecting children.
On the whole, the adult experience of dose–effect relationships will provide a framework for dose titration and ADR monitoring. In other circumstances the medicinal product should be contraindicated for children until further data are available after initial marketing authorization (Jeffreys, 1995).

A survey of new molecular entities approved by the FDA from 1984 to 1989 showed that 80 per cent had no information regarding paediatric use (Roberts, 1996), and this continued in 1992 (Kearns, 1996). Drugs for use in children may be accompanied by problems not seen in adults or cause ADRs that are more frequent than in adults, e.g. antibiotic toxicity in neonates (sulphonamides, chloramphenicol), hepatotoxicity with sodium valproate and Reye’s syndrome with aspirin use for viral infections.

The metabolism of drugs differs in young children. The activity of many P450 enzymes is reduced in the neonatal period and the variation in maturation of different enzymes makes it difficult to predict dosage requirements accurately at different ages. Glucuronidation is a major process of drug elimination in adults, but this is significantly reduced in neonates. Renal excretion is impaired in the first few weeks of life. After the neonatal period, renal function is normal (Choonara et al., 1996). See also Chapter 2.


The elderly

Patients over 65 years of age comprise about 14 per cent of the population in most industrialized countries but they consume nearly 35 per cent of the drugs (Avorn, 1997). The EU/ICH E7 guidelines are provided in Note for Guidance on Studies in Support of Special Populations: Geriatrics (CPMP/ICH/379/95). The usual definition of elderly is over 65 years of age, but the FDA definition is over 60 years. There is a threefold increase in the incidence of ADRs in patients over 60 years compared with patients under 30, and 1 in 10 hospital admissions of older patients are for ADRs (Swafford, 1997).

The healthy old person does not seem to be more susceptible to ADRs compared with the young, but as a group the elderly have many factors predisposing them to ADRs.

Concomitant disease  Increased morbidity means that the elderly are often taking several drugs. Drug expenditure for the elderly accounts for 40 per cent of UK drug expenditure (O’Brien, 1995). There is also a parallel increase in the use of OTC drugs, resulting in an increased chance of drug interactions. The elderly are also targeted by manufacturers of alternative medicines.

Altered pharmacokinetics  Although drug absorption does not change, the drug distribution depends on the lean/adipose body mass ratio, and this declines with age. Protein-bound drugs are distributed differently due to reduced serum albumin concentration. Drugs excreted via the kidneys tend to have a lower clearance rate due to a decline in renal function. The hepatic blood flow and liver mass decline with age and oxidative metabolism may be impaired, although CYP4502D56 does not appear to change with age alone. See Chapter 2.

Altered pharmacodynamics  Beta blockers have less clinical effect for a given concentration in the elderly, but the elderly are more susceptible to the sedative effects of benzodiazepines. Anticoagulants are more likely to cause bleeding. In general, the elderly
have less effective homeostatic mechanisms (e.g. temperature and blood pressure control) (Beard, 1991; Pollock, 1996).

**Poor compliance**  This may be partly due to polypharmacy, poor sight or failing memory. If a product is likely to be used by the elderly then studies need to be performed, especially if the following factors are present:

- a low therapeutic index
- the drug is excreted renally
- there is a possibility of interactions
- there are problems with drugs of the same class
- deterioration in organ function, which may affect pharmacokinetics or pharmacodynamics.

Interaction studies are usually recommended with the following groups of drugs:

- digoxin and oral anticoagulants
- hepatic enzyme inducers
- drugs metabolised by cytochrome P450 enzymes
- other drugs likely to be used with the investigational drug.


**Inclusion of elderly patients in clinical trials**

- Any drug likely to be used by elderly people should be included in pre-marketing studies with an age distribution comparable to that anticipated in routine use.
- Pre-marketing evaluation should include assessing whether important age-related differences exist in efficacy and toxicity.
- Since unexpected differences may emerge in effectiveness or ADRs when a drug is used by large numbers of elderly patients, especially those too frail to be included in trials, plans for post-marketing surveillance (PMS) should be required at the time a drug is approved (Avorn, 1997).

Suggested further reading is Kitler (1989).

**Pregnancy**

Approximately 35 per cent of pregnant women in the UK take some drug during pregnancy, and this includes about 9 per cent who take OTC products (Rubin, 1995). The figure for the USA in 1993 was 68 per cent (Rubin *et al.*, 1993), for Italy 80 per cent in 1995 (Maggini
et al., 1997), and in 1996 for Brazil 94.6 per cent (Fonseca et al., 1997). The animal toxicology will probably be the only evidence for or against teratogenicity until the drug is on the market, and from then on the evidence will be anecdotal unless large-scale epidemiological studies are undertaken. In the UK, the background incidence of major malformations is about 2.3 per cent at birth, rising to 4.5 per cent by 5 years of age. The incidence of spontaneous abortions in clinically recognized pregnancies is 10–20 per cent (Lack et al., 1968). Drugs and chemicals together are thought to account for only about 4–6 per cent of malformations with the cause of 65–70 per cent unknown (Wilson, 1977). All drugs given to the mother, except high molecular weight compounds such as insulin and heparin, can cross the placenta.

Many physiological changes that may affect drug levels occur during pregnancy, including:

- Increased plasma, extracellular fluid, and fat stores
- Increased hydroxylation capacity by steroids, which may alter metabolism
- Albumin binding and binding to specific receptors may be altered
- The 40–50 per cent increase in glomerular filtration rate (GFR), which may affect drug excretion
- Gastrointestinal absorption of oral preparations may be impaired (Redmond, 1985)

The risks of drug use in pregnancy have been classified. The European SPC Guideline states that the following should be mentioned:

- Facts on human experience and conclusions from preclinical toxicity studies that are of relevance for the assessment of risks associated with exposure during pregnancy.
- Recommendations on the use of the medicinal product at different times during pregnancy in respect of gestation.
- Recommendations on the management of the situation of an inadvertent exposure, where relevant.

Examples of wording are provided in an appendix of the European SPC Guideline. The US pregnancy categories are given in 21 CFR 201.57, Specific Requirements on Content and Format of Labelling for Human Prescription Drugs, and are summarized as follows:

A – Studies in pregnant women have not shown that (drug) increases the risk of foetal abnormalities if administered during the first (second, third, or all) trimesters of pregnancy. If this drug is used during pregnancy, the possibility of foetal harm appears remote.

B – Reproduction studies have been performed in animals at doses up to \( x \) times the human dose and have revealed no evidence of impaired fertility or harm to the foetus. There are, however, no adequate and well-controlled studies in pregnant women.
C – (Drug) has been shown to be teratogenic in species when given in doses $x$ times the human dose. There are no adequate and well-controlled studies in pregnant women. (Drug) should be used during pregnancy only if the potential benefit justifies the potential risk to the foetus.

D – (Drug) can cause foetal harm when administered to pregnant women.

X – See Contraindications. (Drug) may cause harm when administered to a pregnant woman.

In Australia the guidance is:

A – Drug taken by many pregnant women with no harmful effects.

B1 – Limited human exposure shows no increase in harmful effects. Animal studies show no harmful effects.

B2 – Limited human exposure shows no increase in harmful effects. Animal studies inadequate or lacking.

B3 – Limited human exposure shows no increase in harmful effects. Animal studies show evidence of foetal damage.

C – Human studies may have shown harmful effects, but no malformations. Effects may be reversible.

D – May cause human foetal damage or malformations.

X – Should not be used during pregnancy, due to harmful effects.

The proposal to increase the number of women in clinical trials may result in more pregnancies in the early stages of a drug’s development. Since there is a 3 per cent chance of major abnormalities in live births (DeLap et al., 1996) there will be the possibility that mothers with babies with abnormalities will sue the company.

In June 2001 the CPMP published two concept papers on ‘The development of notes for guidance on the use of medicinal products during pregnancy: need for post-marketing data’ (CPMP/EWP/PhVWP/1417/01) and on ‘Risk assessment of medicinal products on human reproductive and development toxicities: from data to labelling’ (CPMP/SWP/373/01).

**Pregnancy registries**

The goal of pregnancy registries is to provide clinically relevant human data that can be used in a product’s labelling to provide useful information for treating and counselling patients who are pregnant or planning to become pregnant.
The criteria for selecting drugs for a pregnancy registry are:

- Issues arising from conduct of animal studies.
- Expectation of AEs during pregnancy based on structure–activity relationships.
- Findings of concern from case reports in the literature or identified from PMS.
- Expectation of high use pattern in women of childbearing age.
- Treatment needed for conditions associated with high morbidity or mortality.
- Inability to discontinue treatment ethically during pregnancy (e.g. new drugs in epilepsy where there is a two to three times greater incidence of malformations, mostly due to older drugs; Craig and Morrow, 1997).

A pregnancy registry is desirable for a drug when:

- Inadvertent exposures in pregnancy are likely to be common, for instance when a drug has a high likelihood of use by women of childbearing potential.

- It presents special circumstances, such as the potential for infection of mother and foetus by administration of live, attenuated vaccines. See FDA Guidance for Industry, Establishing Pregnancy Exposure Registries, August 2002.

**Examples of pregnancy registries** These include:

- National birth defects registries (Schardein, 1993).

- UK National Teratology Information Service, which is part of the European network of teratology information services (Bateman and McElhatton, 1997).

- International Clearing House for Birth Defects Monitoring Systems, which covers Australia, France, Israel, Italy, Japan and South America, is referred to as the MADRE project (malformation drug exposure surveillance), and has been collecting cases since 1990.

- The Pegasus project – all pregnancies in Munich – 14 000 births per year and 85 per cent of women used at least one drug during pregnancy (Hasford, 1996; Cornelia et al., 1997).

Suggested further reading is Mitchell (2000).

**The Investigator’s Brochure**

Under the Code of Federal Regulations on Food and Drugs, Title 21, part 312.55 it is a requirement that all investigators are given an Investigator’s Brochure (IB). The International Conference on Harmonization (ICH) *Guideline for Good Clinical Practice* (GCP) E6, 1996 defines the IB as ‘A compilation of the clinical and non-clinical data on the investigational product(s) which is relevant to the study of the investigational product(s)
in human subjects’. In the safety section of the IB, ‘Tabular summaries of ADRs for all the clinical trials would be useful. Important differences in ADR patterns/incidences across indications or sub-groups should be discussed’. There is no mention of AEs. ‘Guidance should be provided on the recognition and treatment of possible overdose and ADR that is based on previous human experience and on the pharmacology of the investigational product’; see also Chapter 9, ICH E6 – Good clinical practice.

The IB should provide a description of the possible risks and ADRs to be anticipated on the basis of previous experiences with the product under investigation and with related products. Klincewicz et al. (2001) have reviewed and discussed the regulations and guidelines concerning the safety sections of the IB.

**Updating the Investigator’s Brochure**

The IB should be updated at least annually, but the frequency will depend upon whether there is any relevant new information (ICH E6). There have been some instances where the IB has not been updated during clinical studies (Mikhail, 1993). The question on how often the IB should be updated is difficult. It is easy to say that this should be whenever there is any relevant new data. In between regular updates a letter can be sent to all investigators and the Ethics Committees announcing any serious AEs, etc. (Mikhail, 1993). Since the term ‘expected’ refers to its mention in the IB, some companies take the view that the more AEs that are put in the IB the less they will need to report. The ICH recommend that the IB is kept current (e.g. through amendments/attachments), particularly for medically important safety data (ICH E6). The aim should be to have a single IB for use in all countries with regular updates, including anything relevant from new animal studies.

**Unblinding**

The importance of maintaining blindness as to the study drug is that it maintains the integrity of the study, and this approach is often emphasized by statisticians. However, ICH E2A advises that the blind should be broken for serious, unexpected ADRs; see Chapter 9. This is encouraged by authorities such as the UK MHRA and is usually done by Drug Safety staff, but without unblinding others involved in the study. The problem with this approach is that the frequency of the ADR in a large study might reach proportions that require further rapid action. The advantages of unblinding to facilitate expedited reporting are that it:

- allows ongoing safety evaluation of the product in development;
- facilitates updating of the IB;
- avoids need to update safety database after the study is complete;
- avoids expedited reporting of placebo/comparator cases;
- meets regulatory requirements.

Gait and Goldsmith (2000) argue that the ICH recommendation is not appropriate and that routine alerting should be based on submission of blinded, serious, unexpected related cases. It is recommended that the ICH E2A guideline should be followed unless the regulatory
Clinical study reports

These are governed by the *Structure and Content of Clinical Study Reports* (ICH E3, 1995) and CPMP guidelines *Note for Guidance on Structure and Content of Clinical Study Reports* (CPMP/ICH/137/95).

The safety evaluation is considered on three levels:

- the extent of exposure (dose, duration and number of patients);
- the more common AEs;
- serious AEs and other significant AEs.

Three kinds of analysis and display are called for:

- Summarized data using tables and graphical presentations.
- Listings of individual subject data; for AEs these should include the subject identifier, age, sex, weight, severity, dose, seriousness, time to onset, action taken, outcome, causality assessment, concomitant treatment, etc.
- Narrative descriptions of events of particular interest.

Large studies will require detailed analysis of subgroups, whereas small studies may only require minimal analysis. The study report is best written by the person who has been dealing with the AEs during the clinical trial and who is familiar with the protocol and the data.

For further reading on this subject, see Gait *et al.* (2000).

Pooling of safety data

Many clinical trials are too small to reveal differences in AE rates between the investigational drug and controls, and even the larger studies are unlikely to be large enough to show differences in subgroups of patients. Pooling of safety data is required to maximize the usefulness of the total exposure to a new drug. Where possible, this should be done as the development programme progresses, not just at the end. Anyone dealing with pooling of data should be familiar with the *Guidelines for the Format and Content of the Clinical and Statistical Sections of New Drug Applications* from the Centre for Drug Evaluation and Research, FDA, Department of Health Sciences, July 1988. The relevant section is ‘Integrated Summary of Safety Information’.


More recently, the ICH have developed Multidisciplinary Topic M4 on the format and preparation of the Common Technical Document (CTD) to harmonize applications that will be submitted to regulatory authorities (ICH, 2002). This includes the Clinical Overview
(module 2.5) and Clinical Summary (module 2.7), within which are the Overview of Safety (2.5.5) and Summary of Clinical Safety (2.7.4). Sub-headings are:

2.7.4.1 Exposure to the drug
2.7.4.1.1 Overall safety evaluation plan and narratives of safety studies
2.7.4.1.2 Overall extent of exposure
2.7.4.1.3 Demographic and other characteristics of study population
2.7.4.2 Adverse events
2.7.4.2.1 Analysis of adverse events
2.7.4.2.1.1 Common adverse events
2.7.4.2.1.2 Deaths
2.7.4.2.1.3 Other serious adverse events
2.7.4.2.1.4 Other significant adverse events
2.7.4.2.1.5 Analysis of adverse events by organ system or syndrome
2.7.4.2.2 Narratives
2.7.4.3 Clinical laboratory evaluations
2.7.4.4 Vital signs, physical findings and other observations related to safety
2.7.4.5 Safety in special groups and situations
2.7.4.5.1 Intrinsic factors
2.7.4.5.2 Extrinsic factors
2.7.4.5.3 Drug interactions
2.7.4.5.4 Use in pregnancy and lactation
2.7.4.5.5 Overdose
2.7.4.5.6 Drug abuse
2.7.4.5.7 Withdrawal and rebound
2.7.4.5.8 Effects on ability to drive or operate machinery or impairment of mental ability
2.7.4.6 Post-marketing data

The aim of pooling data is:

- to evaluate serious AEs too rare to be seen in each individual study;
- to discover whether any particular subgroup is more susceptible to an AE.

Analysis of subgroups that have not been defined prior to the study (so called ‘data dredging’) should not be performed as a part of efficacy analysis but is essential for safety
analyses. However, when ‘subgroups defined by the data are generated by the study results: often an effect is so suggested and then confirmed with statistical significance on the same data set’ (Scott and Campbell, 1998). Ideally, any ‘subgroup ADRs’ found by this method should be confirmed by a subsequent study, as well as by group causation methods.

The first aim is subject to the total number of patients receiving the drug while the second depends upon looking at those subgroups of patients by variables. These may be patient variables:

- age
- sex
- weight
- race, country or centre
- concomitant disease (e.g. renal failure)
- indication and severity
- alcohol intake and smoking

drug variables:

- dose
- formulation
- frequency of administration
- route
- duration
- comparative therapy
- concomitant therapy
- blood/plasma level

or trial variables:

- type of study (controlled, uncontrolled or named patients)
- method of collection of AEs (diary card, checklist, questionnaire, general question or spontaneous reporting).

There are two approaches to the examination of the pool by variables: pool all studies and a hierarchical approach.
Pool all studies

This is only suitable if the variables that one wishes to examine are similarly represented in all studies (e.g. age, sex) and the studies are similar in design. With this approach the effect of small subgroups may be swamped by the majority of the patients (Figure 4.3).

Hierarchical approach

If some studies have a common characteristic then they form a subgroup, which may have a different ADR pattern from the remainder of the studies. The first step is to look at the drug and indication concerned and then to list the variables, starting with those that are likely to have the greatest effect down to those not expected to have an effect. The former should be at the top of the hierarchical tree and the latter at the bottom. For example, one might expect more AEs in trials against an active drug control rather than one against placebo, since only mild disease cases are likely to be in the latter, and uncontrolled studies may contain patients not acceptable in controlled studies (which have stricter inclusion and exclusion criteria). Groups with a concentration of one variable that may show differences could be, for example, as shown in Figure 4.4. Groups of variables that need to be considered are:

- Patient – Race or nationality (Joelson et al., 1997); concomitant disease (e.g. renal failure with a drug excreted renally); indication; drug; dose/blood level; duration; route.
- Trial – collection of AEs by questionnaire/checklist (these should always be analysed separately).
- Country – despite the ICH it is not possible to eliminate national medical characteristics from influencing clinical trials. A survey in Japan found that there was a different attitude to the USA and Europe in regard to: entry criteria, prohibited concomitant

Figure 4.3  Schematic representation of ‘pool all studies’

Figure 4.4  Schematic representation of hierarchical approach
drugs, informed consent and completion of CRFs (Ono et al., 2002). Investigators in Japan have little incentive to perform clinical trials as they are not paid for this work (Miyazaki and Saito, 2002). The frequency of ADRs in Japan was significantly greater than in the USA and the EU (Homma et al., 1994). In France and Germany the sales of drugs are twice as high as in the UK or USA, and so there is likely to be under-reporting of concomitant medication (see Chapter 1). Patients in eastern Europe tend to be more compliant in taking medication because of the culture in those countries, whereby the ‘doctor knows best’ (Neal, 2001). In any country where this paternalistic culture is present the patients may be reluctant to complain of AEs.

Using the above example as an illustration, the process of combining the studies starts at the roots of the hierarchical tree. If there is more than one placebo-controlled study then the incidence rates and the types of AEs are compared and if they are similar they can be combined. If they differ, then the trials are examined to see if there is another variable that might be responsible for the difference (e.g. one study, unknown to the company, had used a questionnaire and had thereby increased the incidence rate of all ADRs). If no explanation is forthcoming then the data are combined and a note made of the discrepancy. It may be that a similar discrepancy will be noted in other studies and point to a variable not previously considered. This process of comparing and then combining is continued to the top of the tree. When a definite difference is identified and the responsible variable found, then this group may be left out of the final pooling and treated separately. Those variables not examined on the way up the tree are examined in the complete pool of the data for any common AEs. The way in which this will be done will depend upon the variable and size of the pooled data (e.g. age may be looked at by decades or divided into below 65 years and above 65 years).

The analysis of healthy volunteer studies can be carried out in the same way and may give an early indication about which variable will be important for the analysis of patient data. However, volunteer data should not be pooled with patient data.

The advantages of the hierarchical approach are:

- Analysis can frequently be started earlier (i.e. when a subgroup is complete, rather than waiting for the whole database to be available).
- The effect of variables not apparent from analysing the whole database may be discovered.
- The incidence rates are likely to be more accurate, since variation due to a known subgroup will be excluded.

**Serious adverse events**

Owing to the small size of the pool of patients treated before Marketing Authorization, serious AEs are likely to be rare; and even with a database of 4000 patients, in some indications there are only likely to be a few serious AEs. Therefore, a meaningful comparison with a control group will probably not be possible. This generalization does not apply to potentially toxic drugs, such as cancer chemotherapy, and will of course depend upon the background noise from the disease being treated. If the AEs are type B events then the
denominator must only include those patients who have taken the drug for a sufficient length of time, which is usually more than 5 days.

**Withdrawals**

All withdrawals should be listed with their cause. Those due to adverse effects of treatment should be treated similarly to serious AEs; this includes those withdrawn due to abnormal laboratory tests.

**Deaths**

All deaths should be listed and assessed as for serious AEs, and compared with the number of deaths expected in the study population. They should be examined for any common variables.

**Incidence rate**

The crude incidence rate is the number of subjects with the AE divided by the number exposed to the drug. In pre-marketing clinical trials the numerator is assessed by the collection of AEs and the number of each AE on the investigational drug is compared with the number on placebo to get the relative risk compared with placebo. Note, however, that it is the relative risk of that AE, not ADR. This will be dealt with further later, but first it is necessary to look at the problems collecting the numerator, the AEs.

**The numerator**

The incidence rate of AEs will depend upon the method used for their collection. Presuming that an AE occurs in a patient whilst on the investigational drug or placebo, there are numerous factors that will influence whether or not it is analysed at the end of the clinical trial.

- The patient experiencing an AE may not remember it, depending upon its severity, or may presume that it is nothing to do with the drug and, therefore, not report it, or may be too embarrassed to report it if it concerns a ‘taboo’ area. If it is severe and the patient thinks that it is due to the drug, then the patient may stop the drug and not report back to the doctor, thereby becoming a ‘withdrawal’.

- The doctor may not allow a patient to report the AE if he is hurried or has a brusque manner. The doctor may not record the AE if he or she does not think that it is due to the drug, despite the protocol’s exhortations.

Other methods, such as diary cards, questionnaires and checklists, collect a different set of the AEs that actually occurred to the patient; therefore, it is therefore not possible to mix them together to calculate an incidence rate.

Let us presume that all of the AEs are collected by spontaneous reporting and a standard question and that the AE we are concerned with is ‘headache’. A decision must be made about whether it is better to calculate the numerator as the number of patients complaining
of headache during the trial or whether the number of headaches per patient or both is needed (Cato et al., 1983). Looking first at the ‘number of patients complaining of headache’. Unfortunately, patients on placebo may ‘catch’ the same type of AE as the patients on active drug, and this will distort the analysis. Presuming that the headache has occurred as part of flu and has been reported, then the doctor may record it as: headache; headaches, fever and general aches; headaches, fever, general aches: diagnosis flu; or just flu or ‘viral infection’. So the same patient may have different words recorded, depending upon the investigator, as follows:

- headache
- headache, fever and general muscular aches
- flu or viral infection.

When it comes to analysis of the AEs the method of counting is important. Grouping of symptoms when they have different mechanisms or aetiologies is wrong in the first instance. A drug-related effect may be drowned by unrelated events with different mechanisms. The other end of the scale, where the AEs are split into very small groups using the patient’s actual words, is too far in the other direction, for instance counting separately right- and left-sided headaches. These processes are sometimes referred to as ‘lumping and splitting’. The degree of grouping must make clinical sense, having the same mechanism or pathology (i.e. events that probably represent the same phenomenon). This is best done by scanning all the AEs to see whether there is a natural clinical grouping. The same symptom complex in many cases may represent a new ADR and should be counted separately from those cases with just a single symptom. So there may be headaches occurring in three categories:

- headaches alone
- headaches with other symptoms
- headaches as part of a diagnosis.

This makes it important that the investigator is instructed to group concurrent signs and symptoms together and give a diagnosis where possible. There is nothing to prevent grouping the first two categories together if, separately, they are equally represented in both the investigational drug group and the control group. The next hurdle is the classification of the AE using a dictionary. If different coders classify the same event under different terms then the incidence rates will of course be incorrect. The bigger the study, the bigger the problem. See also Chapter 12.

In large-scale placebo-controlled studies there is a problem with the number of different types of AE; there can easily be over 100 different types of event. When these are compared statistically between the investigational drug and the control group using $p < 0.05$ there are likely to be several statistically different results purely by chance. One of the solutions to this problem of multiple comparisons has been to adjust the $p$ value required for each comparison so that the overall value remains at $p < 0.05$ (e.g. Bonferroni inequality) (Dunnett and Goldsmith, 1981). This sacrifices some sensitivity for specificity. Another approach is to use $p < 0.05$ for each comparison, realizing that there will be 1 in 20 false positives, but using it to generate hypotheses rather than test them (Enas, 1991). Hence, if
formal testing is performed, then the nominal significance level should be used to flag the AE without adjustment in order to maximize power (Phillips et al., 2000); see also Chapter 6. Further examination of the clinical details may then provide data to prove or rebut the hypothesis, e.g. common clinical characteristics, time to onset, etc.

So far there is the presumption that the unit of time for the event to occur has been short and constant (i.e. the same number finished the study as started). This is very rarely true. In this circumstance the crude incidence rate is satisfactory, but in longer studies there are likely to be withdrawals; this will alter the denominator and perhaps the numerator. If it is presumed that satisfactory statistical significance has been shown between the incidence rate in the investigational drug group compared with the control group, then the attributable incidence rate for the drug-related event should not be arrived at solely by subtracting the placebo rate from that of the active group, but should be given with confidence intervals. If specific ADRs have a negligible background incidence with objective data then it is possible to calculate an ADR incidence rate, but these are extremely rare.

The denominator

The total denominator in UK licence applications (1987–1989) varied from 43 to 15,962; median 1528 (Rawlins and Jefferys, 1991). The numbers of patients included in studies before MAA/NDA appear to be increasing substantially. A study of 15 products in the USA from 1994 to 2000 gave the figures shown in Table 4.4.

Table 4.4 Mean and median total number of subjects per application for biopharmaceuticals, new medical entities and new active substances (Reichert, 2001). Reproduced by permission of Drug Information Association

<table>
<thead>
<tr>
<th>Products</th>
<th>Years</th>
<th>Number</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biopharmaceuticals excluding recombinant protein products</td>
<td>1994–2000</td>
<td>12</td>
<td>1014</td>
<td>960</td>
</tr>
<tr>
<td>Biopharmaceuticals including recombinant protein products</td>
<td>1994–2000</td>
<td>15</td>
<td>5160</td>
<td>1007</td>
</tr>
<tr>
<td>New medical entities</td>
<td>1998</td>
<td>17</td>
<td>5697</td>
<td>4325</td>
</tr>
<tr>
<td>New medical entities</td>
<td>1999</td>
<td>19</td>
<td>4980</td>
<td>5435</td>
</tr>
<tr>
<td>New active substances</td>
<td>1995–1999</td>
<td>23</td>
<td>4478</td>
<td>NA</td>
</tr>
</tbody>
</table>

Incidence rates with confidence intervals only apply to the group of patients tested, and it may not be possible to extrapolate this to the general population or to a subsection of that population. Pre-marketing clinical trials have rigorous exclusion criteria, often excluding the very old, the young, those who are pregnant, those with concomitant disease and those with abnormal laboratory tests.

Having established that the investigational drug has a certain incidence rate of AEs (with confidence intervals), within the trial groups there may be subgroups with very large differences in incidence rates varying from almost 0 per cent to 100 per cent. These must, therefore, be looked at as well as the factors that govern them. Many factors or variables need consideration, not all being relevant for all studies. These include:
• Patient factors – age, sex, weight, race, country or centre, concomitant disease, indication, severity, alcohol and/or tobacco usage.

• Drug factors – dose or blood level, formulation, route of administration, duration, comparator therapy and concomitant therapy.

The number of patients in each subgroup may be so small that the power of the study is insufficient to show any differences and, therefore, it will be necessary to pool experience from as many of the pre-marketing trials as possible. However, this may still leave many subgroups inadequately represented.

During the pre-marketing studies, serious AEs will have been notified rapidly whilst the minor AEs will accumulate slowly as the studies progress and the CRFs are returned to the sponsor. At any one time the numerator for the serious AEs will be known accurately, but there will be only an estimate of the total denominator until the end of the study.

As already mentioned, many subgroups will not have been exposed to the drug by the time of marketing, and life table analysis will be required for exposures of different durations. The time to onset of a type A ADR depends upon the pharmacokinetics of the drug. If the various factors in the pharmacokinetic equation remain equal, then all the patients who are going to have a particular reaction should have had it by the time that the steady state has been reached for a particular tissue, the incidence rate after that time becoming low. However, for type B ADRs, depending upon hypersensitivity, the chance of hypersensitivity occurring before 5 days of treatment at the first exposure to a drug is low and then increases after the first week. These ADRs with a varying hazard rate should use the number of patients exposed for a sufficient length of time as a denominator. On the other hand, there are other ADRs with an almost constant hazard rate (e.g. thromboembolism with oral contraceptives), and under this circumstance the denominator is patient weeks/months.

Although an incidence rate for common type A AEs can be given for the investigational drug and possibly for placebo or standard drug from the clinical trial data, rare type A or B ADRs may be absent or only present in ones or twos in the whole trial programme. In this case, the small denominators for placebo and/or comparator drug will not allow a valid comparison, and a cohort study of historical controls will often be necessary to establish the background incidence in the population (Guess, 1991, 1994).

Final analysis of data

The global index of safety

This method uses a group of specialists who assign scores, 1–5, rating the ‘severity’ of each type of event. It was used for an antipsychotic drug for schizophrenia (Sacristan et al., 2001). The acknowledged limitations of this method mean that it cannot be recommended.

Suggested method

The first essential is to have the resolve to find any ADRs present. Although the main aim will be to establish the presence of type A reactions, the data may contain individual type B ADRs. The grouping of trials in preparation for the final analysis, as described in the 4th edition of this book, may suggest areas for special attention. When all the data from pre-
marketing studies are available they must be analysed by a blend of statistical and group causality techniques.

Statistical (descriptive): the incidence of each AE with the investigational drug is compared with the control. Since there may be over a 100 different types of AE in the database, every 100 comparisons should produce five significant differences at the $p = 0.05$ level purely by chance, i.e. very sensitive. Therefore, a further statistical comparison should be done using a technique to account for these multiple comparisons, e.g. Bonferroni, i.e. more specific. This will result in three groups.

Group 1 contains those AEs where there is no significant difference between the patient incidence in the investigational drug group and any control group. It is unlikely that any of these will prove to be a true ADR.

Group 2 contains those AEs where there is a statistical difference, which disappears when account has been taken of the multiple comparisons. These may include ADRs, but it is not very likely.

Group 3 contains those AEs where there is a statistical difference after allowance has been made for multiple comparisons. These are likely to be ADRs.

The next step is to look at Group 3 AEs for evidence of other clinical, demographic and laboratory differences between the events occurring in the investigational drug group and the control drug group; this may help to confirm or refute a causal relationship:

- Are there a priori hypotheses from drugs of the same class, animal or human pharmacology, animal toxicology, concurrent disease or disease characteristics, analysis of individual studies or individual cases during clinical development?

- Are there demographic differences (age, sex, race, nationality, etc.)?

- Is there a dose relationship?

- Is there temporal clustering related to the expected time to onset of the particular event, i.e. five times the half-life? Consider a statistical comparison for that time period (hazard function).

- Are there differences in the AE: severity, concurrent medication, associated symptoms, associated laboratory abnormalities, course of AE including withdrawals?

- Is there investigator causality?

- Is there biological plausibility?

This should then be repeated with Group 2 events. Further statistical analysis may be appropriate, e.g. event incidence for AEs that recur during treatment.

In Group 1 there may be a rare ADR. There may be a few extra events in the investigational drug group but not reaching the level of significance. These may include type B reactions or interactions with other drugs or diseases.

Finally, a decision should be reached as to whether a certain event is an ADR or not, and for ADRs whether any particular group is susceptible.
Inadequate reporting of safety data from clinical trials

It is essential that whoever is responsible for the safety aspects of a trial ensures that all the appropriate measures are included in the protocol and that the findings are included in the clinical study report. A survey of 192 published randomised drug trials in seven different indications (total 130 074 patients) showed that the quality and quantity of safety reporting to be variable and largely inadequate (Ioannidis and Lau, 2001). Severity of AEs and laboratory-determined toxicity was adequately defined in only 39 per cent and 29 per cent of reports respectively. Only 46 per cent of trials stated the frequency of specific reactions for stopping treatment due to toxicity, and the median space allocated to safety results was only 0.3 of a page.

Another survey of 185 clinical trials showed that 14 per cent made no mention of ADRs; 32 per cent could not be fully evaluated, either because numbers were not given for each treatment arm (52 per cent) or because a generic statement was made without full details (48 per cent). Details as to how clinical events had been recorded was given in only 15 per cent of cases, and similar details on patient symptoms in 17 per cent of trials. Only 49 per cent stated how severity had been defined. The median amount of space used for safety data in the results and discussion sections was 5.8 per cent (Loke and Derry, 2001).

Seventeen licensing applications for drugs used for post-menopausal hormone therapy were obtained from the Finnish Drug Agency, despite opposition from some pharmaceutical companies. They were examined for cardiovascular and thromboembolic events or superficial phlebitis. The trials and their reporting of unanticipated AEs were mostly inadequate. Many trials were of very short duration, the methods by which AEs were monitored and how reasons for withdrawals were assessed were unclear, and sometimes reporting was superficial. Adverse effects resulting in withdrawals, or the group in which withdrawals occurred, were not always reported. Examples of superficial reporting included ‘one death, unrelated to treatment’ (the reason was not specified) and ‘one cancer’ (type unspecified); a common statement was that ‘there were no (serious) adverse effects’. How this conclusion was obtained was usually unclear, as was whether the fate of all patients allocated was known and whether only events thought to be drug related were included. In one application, a trial report stated that ‘during the study, four serious adverse experiences, all in the placebo group, were reported’. Only when looking through the detailed tables did the reviewers find three cases of cardiovascular events in the hormone group. Either the authors regarded cancer but not angina or palpitations requiring hospital treatment as serious, or the summary text was purposefully misleading (Hemminki and McPherson, 2000).

Common errors made in reporting safety information (Ioannidis and Lau, 2002) are:

- Not reporting any safety data at all.
- Making only vague statements, such as ‘the medication was well tolerated’.
- Not specifying a breakdown of events per study arm.
- Lumping different kinds of adverse effects under broad categories.
- Not providing severity, or lumping together numbers for different severity levels, and failing to define the scales used for categorizing severity.
- Giving p-values for comparison of events without numbers per severity level.
• Reporting only the most common events.
• Not providing information on AEs that lead to discontinuation of treatment.
• Providing data on subgroups without that for the total population.
• Over-interpreting the absence of adverse effects for small sample sizes.
• Reporting events without proper data on the observational unit.

Possible solutions for the improvement of the study of AEs associated with highly active antiretroviral therapy in AIDS have been proposed by Carr (2002). These include developing regulatory guidance and consensus methods, active reporting of AEs and a greater number of patients.

Conclusions

As far as ADRs are concerned, the clinical trial programme for a new drug needs to be based on the known problems found with similar drugs and the results of animal studies. The results of the healthy volunteer studies need to be assessed carefully for potential ADRs and then interpreted with previous findings from the animal studies. Each clinical trial will contribute information concerning the ADR profile of the drug, and each study protocol should be considered carefully in order to make the most of the drug exposure. The overall development programme plan must be balanced, such that questionnaires and checklists are used at the right stage with sufficient numbers of patients in order to have a reasonable chance to distinguish between the investigational drug and its controls.

The early clinical trials should be designed to distinguish the ADRs of the new drug from those of placebo-treated patients where this is possible. Later clinical trials will concentrate on comparisons with its potential competitors. Certain clinical trials may be allotted the task of monitoring for specific ADRs.

All patients in clinical trials should be asked a standard open question to elicit AEs at each study visit and given the opportunity to reply. Adverse symptoms should be followed up by clinical examination and/or investigations to search for objective confirmation. In those cases where the drug is withdrawn, the results of dechallenge should be observed, preferably without addition of other treatment either in substitution for the original drug or for treatment of the AE. Any investigations found to be abnormal at the time of the AE should be repeated. Uncontrolled studies before marketing should be restricted and should be documented and monitored as closely as the normal double-blind randomized clinical trials.

Whoever is responsible for the analysis of safety data should be able to review the protocols and have an input into their design with future safety analyses in mind. There will be many reasons why the clinical trials cannot be completely standardized, but there should be a ‘by default’ standard trial protocol that can be altered when necessary. The aim should not be to have completely standardized protocols, but to have them consistent: i.e. using standard safety modules where possible, and where this is not possible making certain that the analyses will not be made more difficult. This is especially true of uncontrolled trials, where a standard protocol can often be used. Similarly, named patients can still follow a
standard protocol designed solely for that purpose. Uncontrolled studies should not be an opportunity for investigators to follow their own whims.

**Future aspirations**

These include:

- More attention to the lowest effective dose.
- Removal of legal restriction on access to company pre-marketing data to enable further research similar to that of Hemminki and McPherson (2000).
- Reappraisal of the pros and cons of measuring various laboratory parameters.
- Improvement in individual case follow-up.
- More information on the preferences of regulatory authorities as to the quantity, quality and presentation of safety data for the MAA/NDA.
- Better publication of information on adverse effects in clinical trial publications.

**Acknowledgements**

We are grateful to Dr Andrew Lockton for his contribution to the section on Phase I studies and to Dr Han de Gier for contributing the entire section on the ‘Effects of drugs on skilled performance’ and finally to Dr Marianne Keisu for her helpful comments.

**References**


REFERENCES


REFERENCES


REFERENCES


O’Hanlon JF (1995). Zopiclone’s residual effects on psychomotor and information processing skills involved in complex tasks such as car driving: a critical review. Eur Psychiatry 10: 137S–144S.


REFERENCES


