Chapter 9: Evaluation of toxicity data

Toxicology studies have been designed to determine the toxic effects associated with exposure to chemical hazards. Such studies can provide information relating to toxic effects and potential health hazards likely to arise from single or repeated exposures, in terms of predicting potentially important toxicity endpoints and identifying potential target organs or systems.

This chapter on toxicological evaluation focuses on chemical hazards assessed using traditional toxicity testing methods and, in particular, on some of the problems and pitfalls that may arise during an assessment of possible compound-related changes in the parameters measured in toxicity studies. It is intended to provide guidance on the process of hazard identification and assessment.

The basic assumption is that traditional methods of toxicity assessment will continue to be the mainstay of EHRA for some time.

It is also important to note that, over time, the scientific community is gaining a better understanding of the mechanisms of toxicity, and this is leading to changes in both methodology and interpretation of toxicity data. It is inevitable that new paradigms will be introduced as science advances (see Section 9.4).

It is therefore important to acknowledge that the analysis and evaluation of toxicity studies reflects scientific consensus at the time the data is reviewed. This means that the toxicity studies underpinning many EHRAs may contain data generated during an era when toxicity testing and the interpretation of results were less well advanced.

9.1 TOXICITY TESTING – MAJOR *IN VIVO* STUDY TYPES

Hazard identification mostly relies on the results of *in vivo* toxicity studies conducted according to standard protocols. Guidance on the conduct of toxicity tests has been promulgated by the OECD (OECD 2009). There have been 53 OECD *Test guidelines* published since they were first promulgated in 1981, and many of these have been periodically updated.

The following types of studies are defined.

Acute toxicity studies are studies that investigate the effects of single doses of a substance. The LD_{50} test, or medium lethal dose test (OECD *Test guideline* TG401), which records gross toxicity and mortality data over a 14-day post-dosing period, has been commonly employed and may still be included in many data packages. However, TG401 was formally withdrawn by the OECD in 2002 in response to animal welfare concerns. Newer tests ('limit' tests and 'up-and-down' dosing methods) are now favoured as they reduce the numbers of animals required and reduce the suffering seen in the classical LD_{50} test. OECD TG420 covers acute oral toxicity determination by the 'fixed-dose method', TG 423 by the 'acute toxic class method', and TG 425 by the 'up-and-down procedure'.

The standard acute toxicity studies include tests for: acute oral, dermal and inhalational toxicity; eye irritation; skin irritation; and skin sensitisation. Such studies may serve as the basis for classifying and labelling a particular chemical or mixture, and serve as an initial guide to possible toxic modes of action and in establishing a dosing regimen in sub-chronic toxicity studies. Substantial work has been done to develop alternative tests (mainly *in vitro*) to replace skin/eye irritancy and sensitisation tests, and some of these have now been incorporated into the OECD *Test guidelines* series (e.g. TGs 429–435, and 437–438).

Sub-chronic toxicity studies are short-term repeat-dose studies. A short-term study has been defined (WHO 1990) as 'having a duration lasting up to 10 per cent of the animal's life span, 90 days in rats and mice, or 1 year in dogs', although the US EPA considers a 1-year dog study to be a chronic study. The main purpose of sub-chronic testing is to identify any target organs and to establish dose levels for chronic exposure studies.

Chronic toxicity studies, or long-term studies, are defined as studies lasting for the greater part of the life span of the test animals, usually 18 months in mice and 2 years in rats (WHO 1987; 1990). The OECD protocols for these studies may cover the investigation of chronic toxicity (TG452) or carcinogenicity (TG451), or both (TG453). All three of the OECD *Test guidelines* were updated in 2009 to better reflect developments in animal welfare and to improve dose selection.

Reproductive toxicity studies are studies designed to provide general information about the effects of a test substance on reproductive performance in both male and female animals, such as effects on mating behaviour, gonadal function, oestrous cycling, conception, implantation, parturition, lactation, weaning and neonatal mortality. These studies may also provide some information about developmental or teratogenic effects of the test substance. The conduct of and the results from these studies are very important to assess with care, since the reproductive process is critical for perpetuation of the species and factors or agents that alter or disrupt this process can have devastating consequences, both in fact and in public perception (Korach 1998). For information on study design, refer to OECD *Test guideline* 415, *One-generation reproduction toxicity study* and *Test guideline* 416, *Two-generation reproduction toxicity study*: (OECD 2009). For guidance on evaluating reproductive toxicity studies, refer to IPCS EHC 225 *Principles for evaluating health risks to reproduction associated with exposure to chemicals* (WHO 2001).

Developmental toxicity studies are studies that examine the spectrum of possible *in utero* outcomes for the conceptus, including death, malformations, functional deficits and developmental delays (Tyl & Marr 1997). Exposure during sensitive periods may alter normal development resulting in immediate effects, or may subsequently compromise normal physiological or behavioural functioning later in life. Since some developmental processes can occur perinatally or postnatally, protocols for developmental studies are being modified and extended to address developmental toxicity during the period covering major organogenesis as well as covering the perinatal and early postnatal period. This could include delayed toxicity associated with epigenetic effects during sensitive phases of foetal development. Such attention to the critical timing of exposure also accords with a growing emphasis on understanding early-life susceptibility to carcinogenesis (see Section 5.5.2).

Genotoxicity studies are designed to determine whether test chemicals can perturb genetic material to cause gene or chromosomal mutations. A large number of assay systems, especially *in vitro* systems, have been devised to detect the genotoxic or mutagenic potential of agents (IARC 1999). Most authorities consider that a minimum set of data is required to define a mutagen/non-mutagen. This data usually consists of gene mutations in bacteria and mammalian cells, and *in vitro* and *in vivo* cytogenetics. Newer assays that could provide additional information include the comet assay, mutations in transgenic animals, fluorescent *in situ* hybridisation and cell transformation. Guidance on the conduct and interpretation of *in vivo* and *in vitro* genotoxicity assays, and integration of their results, is also available in a UK Department of Health document (see COM 2000). Interpretation of the results of *in vitro* genotoxicity tests for the purposes of identifying potential human genotoxins and, by inference, potential human carcinogens, needs to be done within a well-defined science policy context (Thybaud et al. 2007).

Other tests: The OECD *Test guideline* series now includes special tests for such endpoints as neurotoxicity (TG424) and developmental neurotoxicity (TG426). It has also addressed animal welfare issues through the development of a range of validated short-term *in vivo* tests and *in vitro* tests, which may complement, or possibly substitute for, the conventional animal tests that have been used for many years. These include tests for skin absorption (TG428) and tests for endocrine-related endpoints (*in vivo* tests TG440, 441 and *in vitro* test TG455).

9.2 GUIDANCE ON EVALUATING AND INTERPRETING TOXICITY TESTS

Supplementary guidance on the evaluation and interpretation is provided in more detail in Appendix 1. This guidance is aimed primarily at experienced toxicologists who may be asked to provide a weight-of-evidence (WoE) analysis of the extent to which toxicity tests are able to define the hazard identification component of an EHRA, and to provide useful information on dose–response relationships. It may also be of value to less experienced readers seeking further detail on using conventional toxicity tests.

9.3 EVALUATING THE WEIGHT OF EVIDENCE AND CONSIDERING THE TOXICOLOGY DATABASE IN TOTO

The essential purpose of toxicity studies is detecting valid biological evidence of the hazard potential of the substance being investigated. Evaluation of the weight of evidence (WoE)[3] produced by toxicity studies is the process that considers the cumulative data pertinent to arriving at a level of concern about the potential adverse effects of a substance. It is composed of a series of judgements concerning the adequacy, validity, and appropriateness of the methods used to produce the database, and those judgements that bring into causal, complementary, parallel or reciprocal relationships, all the data considered. Because our knowledge about mechanisms of toxicity is still developing, because good epidemiological evidence is seldom available, and because animal studies are not always conclusive, the information available at a given time may provide only 'persuasive' rather than 'hard' evidence of a defensible presumption (one way or the other) about the potential health effects of a substance under given conditions of exposure. Therefore, it is necessary to succinctly discuss the rationale for judgements and conclusions contained in risk assessments together with any associated uncertainties. This becomes important when new data or new scientific knowledge requires re-evaluation of the database or a change in a previous risk assessment or regulatory action.

3 Strength of evidence' is commonly taken to mean the degree of conviction regarding the outcome of an experiment such as NTP's 'clear evidence', 'some evidence', 'equivocal evidence' and 'no evidence' of carcinogenicity. 'Weight of evidence' involves integration of all available data, not just one study.