Developmental and Reproductive Toxicology

A Practical Approach

Third Edition

Edited by
Ronald D. Hood

informa
healthcare

21 Understanding statistics in developmental and reproductive toxicology

Tom Vidmar, Les Freshwater, and Roger Collins

INTRODUCTION

From a statistical perspective, developmental and reproductive toxicology (DART) studies offer unique issues not found in other biological research. The objective of DART studies is to evaluate the effects of test article administration not only on a first generation of animals (F_0), but also on a second (F_1), and sometimes a third (F_2). It is the information collected on the F_1 (and later generation) animals that present unique statistical issues. In particular, two aspects of the study designs and objectives of such studies set them apart from other non-clinical, and even clinical, studies from a statistical perspective: F_1 animals (*i*) are not all independent of one another, and (*ii*) in most cases, are not directly exposed to the test article.

The objective of this chapter is to develop concepts to aid in the understanding of statistics as it applies to DART. First, the DART study design will be reviewed from a statistical perspective, followed by some basic statistical concepts presented in a toxicological context. Next, statistical methodologies will be presented for data collected on the F_0 generation. These same methodologies will then be adapted to the unique issues presented by the F_1 generation. Finally, related issues are considered, such as specific details of the statistical aspects.

This chapter is not intended to provide cookbook recipes for applying statistical methods to data collected in DART studies. As such, formulas and calculations are kept to a minimum unless used to illustrate a concept. Rather, this chapter is intended to provide a basic foundation of statistical principles that will enable the reader to critically appraise the design, conduct, statistical methodology, and interpretation of DART studies in the literature and in a regulatory environment.

The example data used to illustrate concepts have been constructed to reflect real data in terms of magnitude and variation.

STUDY DESIGN

The goal of statistical experimental design is to get as close as possible to the ideal experiment, given the constraints of feasibility, finite resources, inherent variability in animals, and the ability to extend the results beyond the current experiment. The ideal experiment will generate data to answer the question it was proposed to answer.

Requirements for a good experiment are listed below (see Cox for further details): (1)

Absence of systematic error. This is a requirement to provide an unbiased estimate of the treatment effect through the comparison of nearly equivalent groups, each receiving differing treatments. Most often, equivalent groups are accomplished through randomization. Randomization is one of the most important concepts when discussing experimental design. Without proper randomization, conclusive evidence can't be derived from an experiment.

Precision. If the systematic error is eliminated, then the results obtained in the experiment should differ from the underlying true values by random variation only. To add to precision, the amount of random error should be made as small as possible.

Range of validity. As mentioned earlier, can the conclusions from an experiment be applied in a larger context?

Simplicity. If the experiment is overly complex, it will be prone to errors and large variability.

The calculation of uncertainty. Do the results of the experiment allow the assessment of the uncertainty or error associated with the estimated effects of the treatments applied?

The experimental unit is the smallest division of the experimental material that can receive different treatments. Using this definition, the experimental unit for data collected from the F_0 generation is the animal receiving the treatment. The experimental unit for the F_1 generation is also the animal receiving the treatment, and not the pups from which data is collected. This makes these experiments unique, and statistical accommodations for this phenomenon will be discussed in the subsection on Litter Effects. Although the number of experimental units to be used in these experiments is recommended by regulatory agencies, it is instructive to discuss how that number is usually determined. The proper number of experimental units impacts the statistical design principle of precision and avoids wasting precious resources.

The size of the sample must be large enough so that an effect big enough to be of scientific relevance will also be statistically significant. It is just as important that the number of animals not be too large, where an effect of little scientific importance becomes statistically detectable. Sample size is also important for economic reasons. An experiment with too few animals can be a waste of resources, as it does not have the capability to produce useful results, while an oversized experiment consumes more resources than is necessary. Since the experiments in this chapter involve animals, sample size is a pivotal issue for ethical reasons. An undersized experiment exposes the subjects to potentially harmful treatments without advancing knowledge, while an oversized experiment exposes an unnecessary number of animals.