

Measurement Uncertainty in Warm Blooded Multiend point studies.

Patricia D. Royal

Introduction:

Measurement uncertainty is a standardized calculation to determine the accuracy of a measurement with predetermined contributors over time. When dealing with multi endpoint studies conducted in warm blooded animal this requirement is challenging for a variety of reasons. It is vital that all A2LA assessors that are assigned to these laboratories understand the unique aspects of these studies, and that our expectations are consistent. With the goal of gaining consensus and an active dialogue, I would like to discuss the unique challenges that warm blooded animal studies encounter when determining measurement uncertainty.

We currently have several labs that are affected. These are mainly toxicology or medical device laboratories, along with laboratories that test creams or vehicles to assess bioavailability of drugs and growth of organisms regulated under Good Laboratory Practice (GLP) and to a certain extent Good Manufacturing Practices (GMP). These are regulated studies under FDA requirements. Equally affected are the clinical veterinary laboratories and animal drug testing laboratories where reference materials or contributors to the test are often derived as a metabolite from a treated animal.

This discussion is limited to toxicology studies and medical device studies. This discussion does not include mutagenic and cytotoxicity tests conducted in-vitro at this time, because these do not include live animals. It is the live animal studies where there are some problems that need to be evaluated.

Background:

Toxicology laboratories conduct a variety of tests on a single analyte or test article for multiple endpoints. Medical device studies also evaluate the toxic effects of a device, in addition to dissolution studies where the elution of a drug from the device is studied in an animal and/or analytically. Some of the challenges in estimating uncertainty of measurement stem from understanding these types of studies.

These types of studies are conducted to assess the safety of drugs and medical devices in a host of animals; and ultimately as a drug or medical device in humans or in animals as a veterinary product. Studies are conducted on a tier effect, starting with simple one end point studies and moving to more complex studies testing. Each test is conducted by a unique Protocol signed by a Study Director. Protocols to a certain extent have been standardized by regulating agencies. Testing is generally conducted on unregistered products thus there is no certified reference material. In animal drug studies conducted to assess the toxic response, the reference material usually comes from the client or FDA submitter. The pure reference material is used to determine the concentration,

homogeneity and stability of the drug/ drug substance mixed with the carrier or vehicle; however, it can also be used as a high dose of the test article. The reference material is usually characterized analytically by the manufacturer or submitter to FDA. The laboratory then conducts a verification or method modification to assure that it can measure the test article under the conditions of testing in the matrix to be used. This is approved by the submitter before testing begins.

Tests conducted to evaluate the toxic response of a drug or medical device may include mutigenicity tests, acute tests, subchronic, 2-year chronic, reproductive tests and metabolism tests. These tests are generally conducted only once, toxicity is evaluated and compared with preceding tests. Decisions are then made to proceed based on the results of one test as part of the series of tests conducted to support registration of the product.

The use of live warm blooded animals is restricted by the federal Animal Welfare Act. So using excess animals to conduct repeated tests to determine measurement uncertainty is limited. Rodent studies usually include 4-5 groups of animals of 20 animal each (10 males/10 females) which include a control, sometimes a matrix control, low, mid, and high dose group. The testing unit is analyzed by sex and the group of animals rather than each animal. As tests move up the ladder to use rabbits, dogs, sheep, pigs, goats, and monkeys the number of animals may be reduced, but the testing unit is still used to determine the toxic response as measured against the control group(s). They also become more expensive to run.

Multi-end point studies include an initial randomization based on source of animals, age, and weigh, so at the starting point, the groups are uniform. Additional measurements then are made through out the study on body-weight, food consumption, clinical chemistry on blood and urine, hematology, necropsy (including organ weight and clinical signs), behavior and animal observations, and histopathology. The end point is an evaluation or a comparison of toxicity compared to the control group(s). The individual measurements are not necessarily weighted, but used to develop a diagnosis. The end result, or diagnosis is based on a comparison of toxic responses with those of the controls, resulting in a No Observable Effect Level or NOEL. For the most part end points are qualitative scores used to evaluate or measure the toxic response against the negative control group. Pathology evaluations even though reported as “mild, moderate, and severe” are based on a scoring process as compared to the control groups.

Issues:

How do we determine measurement uncertainty? What are we going to ask these laboratories to do to meet these requirements? Various sections of the ISO 17025 must be evaluated to assure consistent application for these types of studies.

Section 5.4.6: Estimation of uncertainty of measurement

States that a “laboratory shall have and shall apply a procedure to estimate measurement of uncertainty”.

5.4.6.2 States that “in certain cases the nature of the test method may preclude rigorous, metrologically and statically valid calculations of uncertainty of measurement. In these cases the laboratory shall at least attempt to identify all components of uncertainty and make a reasonable estimation, and shall ensure that the form of reporting of the results does not give a wrong impression of the uncertainty. Reasonable estimation shall be based on knowledge of the performance of the method and on the measurement scope and shall make use of, for example previous experience and validation of data.”

“Previous experience” can include the number of times a laboratory has run a particular type of test, such as a subchronic rat study or skin irritation test, as a technology, rather than as an individual test on a particular medical device or drug, in addition to the training and qualifications of its staff.

5.4.6.3 States that “When estimating uncertainty of measurement all uncertainty components which are important in the given situation shall be taken into account using appropriate methods of analysis. “

When evaluating uncertainty of measurement an assessor can assure that the laboratory has written a procedure that addresses measurement uncertainty, and then assesses procedures for animal observations, control charts for clinical chemistry and hematology, along with balance calibration, temperature and humidity controls in animal rooms, and with randomization of animals by weight and age, along with education and training of the staff. The assigned Study Director and pathologist should be identified in the study specific Protocol.

Beyond that, we run into qualitative evaluation, including toxic response based on educated judgment to determine the toxic response. We must remember that the biggest contributor to uncertainty of measurement in these types of studies is the animal. That is why the testing unit of measurement is based on a uniformed randomization of 10 animals per sex per dose group.

P103b: Life Science Policy on Uncertainty of Measurement:

The current Policy was our initial attempt to qualify life science testing into three compartments; those being purely qualitative, semi-quantitative, and quantitative. While a good initial attempt for most tests, for studies conducted in warm blooded animal studies of the type described, these requirements and notes may need some revision. While all endpoints are contributors, the ultimate endpoint is a qualitative evaluation based on the educated evaluation of a pathologist, the Study Director and educated technician, and comparing toxic responses with the control groups.

In the Life Science Policy, after the description of Category 1, qualitative tests, there is a note or explanation in the current Policy that goes beyond the ISO 17025

Standard that may need revision. The note states that for Category 1 tests that “report +, ++, or +++ indicating relative intensity of a response... These examples are all subject to effects of measurement uncertainty”. It appears that this note may need revision to provide a realistic means for these types of studies to meet these requirement. Multi-end point studies are “well recognized tests” and could also be classified as Category II. Therefore, the following proposal is given:

Proposal for dealing with Uncertainty of Measurement in Multi-endpoints Studies

- 1) There must be a procedure that describes the process of determining uncertainty of measurement
- 2) Contributors must be reasonably identified.
- 3) Category 1 Studies: Multi-endpoint studies where the ultimate end point is visual should be qualified as qualitative. These tests include those with pathology evaluation.
- 4) Category II: Visual observation of irritation and sensitization where visual observations and scoring are well defined.
- 5) Category III: Studies with defined analytical endpoints, such as radiometric uptake.

In Conclusion:

- Requirements for calculating measurement uncertainty in tests using warm blooded animals have certain challenges with meeting the requirements of ISO 17025, Sections 5.4.6, and the current Life Sciences Policy P103b.
- Types of tests affected include toxicology studies, medical device studies, veterinary programs, and animal drug testing programs.
- The use of excessive animals is restricted by the Animal Welfare Act.
- Certified reference materials are limited or non-existent in many cases.
- Accuracy of the end result is based to a large extent on the education and training of the staff along with the facility and equipment.
- The current policy should be revised to accommodate these unique scientific disciplines and to assure the ongoing accessibility of accreditation to these types of testing programs.
- We must agree on a consistent interpretation and application by assessors to meet these goals.