Pre-clinical In Vivo Testing Methods for Medical Devices
Joseph W. Carraway, DVM, MS
Director, Toxicology, NAMSA

Dr. Carraway, who holds a doctorate in veterinary medicine, shared his strong understanding of how we use animals currently in testing, and what is the state-of-the-art in terms of animal test methods for medical devices. He reviewed testing methods based on the type of contact and contact duration, extraction procedures, and current in vivo test methods.

ISO 10993-1: Device Categories

Testing categories are based on the type of body contact, and the contact duration.

Type of body contact:

- Surface contacting device
- Externally communicating devices (part of the device is inside, part is outside – catheters, fixation pin in a bone)
- Implant devices

Contact Duration:

- Limited: ≤ 24 hours
- Prolonged: > 24 hour, up to 30 days
- Permanent: > longer than 30 days of exposure

Biological Evaluation Process

Thinking about it from a risk-based approach, devices that only entail surface contact carry less potential risk than implantable devices, and the longer you have exposure, the greater the risk. Thus, you tend to see that devices that are limited to surface contact have fewer tests than implantable devices, and as you go out in duration the number of tests increases as well.

Table of devices, type of contact, and potential biological effects/endpoints
Biological Evaluation Tests

- Consider tests described in the table for the duration and nature of contact
- Consider any additional specific areas of concern based on the product
- Important considerations for the biological tests:
  - Testing shall be done on the finished device: We can look at raw materials, but we always want to run tests on the finished device, because they have been through the manufacturing process; things could happen along the way that could impact results
  - Extracts of the device shall be prepared based on ISO 10993 Part 12
  - Specific chemicals of concern and type of patient exposure to be considered

Biological Evaluation Process

Biological Evaluation Tests

- Cytotoxicity – this is a purely in vitro assay; all the others are either primarily in vivo assays, or in vitro assays are very limited, with the possible exceptions of genotoxicity and hemocompatibility
- Sensitization
- Irritation
- Acute Systemic Toxicity
- Subacute/Subchronic Toxicity
- Genotoxicity
- Implantation
Hemocompatibility

Additional effects to consider include chronic toxicity, carcinogenicity, and reproductive toxicity, biodegradation, immunotoxicity, toxicokinetics, etc.

How a device can be tested
Testing medical devices poses some different challenges in contrast to pharmaceutical and chemical testing, which involves preparing a solution where you can control the concentration. Typically, device testing involves an unknown mixture of materials. Extracting the device is typically the way to get it into a test system. Standard vehicles such as saline, water, culture media, and vegetable oil are used to pull out non-polar or lipid-soluble chemicals, and other solvents. Extracts tend to be tested on a surface-area basis if surface area can be calculated, if not, the default is to test by weight.

- Standard vehicles used:
  - Polar – saline, water, culture media (w/o serum)
  - Non-polar – vegetable oil, alcohol
  - Additional – alcohol/saline, DMSO, culture media (w serum)

- Standard ratio test/control article to vehicle
  - <0.5 mm thick = 120 cm²/20 mL
  - >0.5 mm thick = 60 cm²/20 mL
  - Irregular= 4 grams/20 mL
  - Elastomeric = 25 cm²/20 mL

- Time and temperature – use standardized times & temps for extractions
  - 37°C for 72 hours, 37°C for 24 hours (cytotoxicity only)
  - 50°C for 72 hours
  - 70°C for 24 hours
  - 121°C for 1 hours

How a device extract is prepped
In general, packaging material is not included for testing — just the device itself. It is important to include only the portion with possible patient contact, base the extraction ratio on surface area if possible, ensure that the entire device is included (or, if that is not possible, representative portions), and that the device is subdivided into small pieces.

Image slide (Slide 9): A coiled catheter will be prepped on a surface area basis, upon deciding whether or not to test the entire device or just representative pieces. When an extract is created and put it into a vial, it is typically chopped into small pieces (as a method indicated by the standard). That raises an important consideration to keep in mind, because sometimes that could have an impact. For example, imagine that this catheter has a wire braid in the core, and cutting it up might expose a component that would not be exposed ordinarily in its typical clinical use, and that could adversely impact the test outcome.

Most of the in vivo test methods are adapted from typical chemical and pharmaceutical test methods that have been around for a long time. As such, sometimes the adaptations work well, and sometimes there are challenges.

Irritation/sensitization covered in ISO 10993-10:

Sensitization and irritation testing is required for all patient-contacting devices. Sensitization refers to an allergenic response, an immune-mediated response; irritation refers to a localized, non-specific inflammatory response. In the case of irritation testing, test methods are generally selected based on the intended use of the device, but many irritation tests are conducted in rabbits.
Currently, ISO allows for in vitro methods, but as they are written right now, if you have a negative result in those in vitro assays — whether it’s a sensitization or irritation assay — the standards indicate that needs to be confirmed in an in vivo assay. Most tests for medical devices do produce negative results. Sponsors are typically reluctant to do two sets of tests, when they assume the outcome is going to be a negative test result.

**ISO 10993-10: Tests for irritation & sensitization**

The three most common sensitization test methods are the guinea pig maximization test, the Buehler closed patch (in guinea pigs), and the murine local lymph node assay (LLNA).

By far most common method that NAMSA uses, the guinea pig maximization test, is also the most widely recognized. Typically done with 15 guinea pigs per extract (10 with the test extract, five with the control). These tests involve several phases, and are usually done with both polar & non-polar extracts – typically two extracts per device.

Induction I Phase of the maximization test method, the intradermal injection phase, involves a series of intradermal injections to the guinea pigs; some with the extract with the test article and those that have been mixed with Freund’s adjuvant as an immunopotentiator. In the Induction II Phase – also known as the topical induction phase – the skin is pre-treated with sodium laurel sulfate to increase erythema and improve absorption. Then, a topical application of the extract is applied and occluded for 48 hours, and hopefully some of the chemicals in that extract are absorbed through the skin. In the final phase (Challenge Phase), another topical exposure is applied after a 14-day rest, occluded for 24 hours, and then evaluated for two days for erythema & edema.

The endpoint to evaluate is a subjective assessment. We’re looking for erythema & edema, and there’s a grading scale — but still, its subjective. If the results are somewhat inconclusive, there is also an option for a re-challenge to see if you can get confirmatory results. Early, at NAMSA, if we had 10 test guinea pigs we might see one with a reaction score of 1, and maybe nothing in the control. With experience, if NAMSA sees a 1, it almost automatically triggers a re-challenge, and surprisingly often that leads to a repeat or even increased response. Some fairly weak responses that are reproducible are picked up through this testing.

The second major sensitization test method, the Buehler or Closed Patch method, also uses guinea pigs. This method is only used for devices that have contact with intact skin, maybe with mucous membranes (although the FDA prefers to limit this to intact skin).

Essentially, this involves a topical exposure three days in a week, versus the chemical method that would be just one day a week for three weeks. Since this method is typically looking at fairly weak, or low concentrations of an unknown, the device modification is done three times a week. After three weeks of exposure, or nine topical inductions, we kick over to essentially the same as the maximization test — a rest period, then come back with a
challenge round, then score for erythema and edema. This method is less common than it was 10 years ago; in fact now it is fairly rare.

The Murine local lymph node assay (LLNA) measures the proliferation of lymphocytes in the lymph nodes after they are exposed to a sensitizing test substance. The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) has approved the LLNA. It is listed as an option in ISO Part 10. NAMSA did this test method for a number of years, then concerns surfaced about whether this test method picks up various sensitizers, and it was not widely accepted by FDA. Currently, the FDA accepts it on a case-by-case basis. Although it uses mice rather than guinea pigs, it calls for a similar number of animals: five per test extract plus five negative controls and five positive controls. There are variations where you can just do periodic positive controls, but the guidance standards generally advise the inclusion of a positive control with every assay.

The method involves applying extracts topically to the skin on the back of a mouse’s ear for three consecutive days, followed by a two-day rest period. Then, the mice are injected with tritiated thymidine (3H), for radioactive labeling of the cells. If there are rapidly dividing cells in the lymph nodes, then that thymidine gets incorporated into those cells. Radioactivity can then be measured in those cells, making this a quantitative assay (the cutoff point to look for would be three times the negative controls). It is a more definitive assay, because it is quantitative, compared to the qualitative guinea pig assays. Despite some of its advantages in terms of animal welfare, it is not widely being used.

Tests for Irritation

Intracutaneous Irritation Test
For implantable devices, there is not a good way to expose the animals to a device. Instead, a device is extracted and injected. This method really is used primarily for implantable or breached, compromised skin surfaces. Commonly, intradermal injections of an extract is administered over a three-day period, then the animal is evaluated for a response and assigned a score for erythema and edema over a three-day period, and compared it to the negative control. This is a common method, used primarily for implants but also externally communicating devices as well.

Skin Irritation Test
The topical skin irritation assay, used for devices that leave skin intact or only minimally compromise tissue, is also done in rabbits. Albino rabbits are commonly used for the skin assays. They lack pigmentation, so it is very easy to see erythema and edema in these animals.

If the device is something that can be applied directly, such as a patch (the article must be thin and pliable) – or for extracts, applied with gauze, it is applied, then typically scored for erythema and edema through 72 hours. This assay is calculated based on a Primary Irritation Index; the device is assigned an irritancy rating based on a scale.

Mucosal Irritation
These methods are probably a little bit less mainstream, but for devices with contact with urogenital mucosal, extracts can be exposed to the rabbit model penis or vagina – looking for macroscopic observations. This type of irritation assay involves euthanizing the animal at the end, and collecting tissue samples, and then conducting final microscopic evaluations on H&E stained sections of tissue. The tests are compared to control animals with a scoring scheme for the evaluations, and an irritancy ranking based on the score.

Ocular Irritation Tests
An ocular irritation test might entail an ocular solution applied directly to the eye, or an extract. They typically use a standard Draize scoring scheme. At least for devices, while it is not extremely unusual to see some scores
just on the intracutaneous irritation test, it tends to be less common to see scores for the skin irritation assay and also for the ocular irritation assays. The things that tend to be seen, perhaps a new contact lens solution or an extract of a contact lens, are all ocular products that are well known, and well-characterized. The sponsors would be surprised to see a response to already accepted ocular products. An exception might be if a client is tinkering with the raw materials going into a contact lens; some of those materials might trigger new responses.

ISO 10993-11: Tests for systemic toxicity

Acute systemic toxicity testing

The test that is typically done for medical devices is an adaptation of a USP test for systemic toxicity. Mice are injected with a single dose – you might say a maximum tolerated dose, with a typically injection of around 50 mL/kg of an extraction or an actual liquid sample. That is an enormous volume, considering that blood volume in an average adult human is about 70 mL/kg. This method was developed for pharmaceutical containers. In this situation, it might be difficult to draw out enough of a potentially adverse substance, such as toxic chemicals from a drug vial. Therefore, to improve sensitivity, a high amount must be injected; that is where the 50 mL/kg dose was derived.

That works well for most devices, although it can be a problem for devices that are soluble, given that kind of volume. We’ll typically inject saline extracts intravenously, and sesame oil extracts are administered by the intraperitoneal route. After the injection, the animals are evaluated for evidence of toxicity (vitality, clinical signs, weight loss) over a 72-hour period.

The doses are not usually tiered. It is rare to see a response in the oil extract; if a response is seen, it’s almost always in the saline extract. It might not be obvious that something is dissolving; the solution could be clear, but if you have excess calcium or some electrolytes. Given that it is a bolus injection given in a matter of a minute, if the electrolytes are off the heart will stop — it is a crude screen.

Subacute, subchronic, and chronic toxicity

Subacute, subchronic, and chronic toxicity studies are most often done as implants. A portion of the device is taken and implanted subcutaneously (typically in a rat model) with the goal of exaggerating the clinical dose. After calculating the human dose and converting to mg/kg and plugging in a safety factor, ideally somewhere between 10-100x, then the device is implanted and the animal is observed for endpoints.

Implantation

There are a lot of potential in vitro methods to look at the biological effects of implantation. But implantation, the device-tissue-interface, is a little more difficult to address with in vitro assays in implants, when trying to see what happens to the living tissue surrounding the device. Muscle tends to be used as a common surrogate for a variety of tissues regardless of where it is implanted, but for some devices we will try to implant in the site of use, but there are multiple tissues that a device traverse through, and hence multiple tissues represented. Rabbits tend to be a common implant species, but the size of the device can determine what species is used.

While the primary focus of this part of the standard, ISO 10993-6, is on local tissue biocompatibility, typically these studies can provide insight into functional and performance aspects as well. Implantation/local effect assays typically run 2-12 weeks, with short-term or long-term options. For degradable materials, the last interval occurs when the material is fully absorbed by the body. The evaluation includes macroscopic and microscopic scoring of implant sites, looking at what happens around the implant material and comparing that to the control article.

ISO 10993-4: Selection of tests for interaction with blood

Most of the tests for hemocompatibility can be done in vitro. There are five main areas that may be considered: thrombosis, coagulation, platelets, hematology, and complement. Even where testing in this category may not
appear to be required, any device with bone contact will also have blood contact. Hemolysis is the only well defined assay, and the only test likely to be needed for an orthopedic device.

Thrombosis, however, is an exception; addressing thrombosis requires the actual device, or something that represents it well, because surface geometry and materials characteristics can influence the outcome. Factors such as the wide variation in size, location of use, exposure time make it difficult to say one test method ideal for all. There is work underway within ISO to come up with some in vitro assays for thrombosis, but for the time being, in vivo assays are commonly used. The challenge is that these assays require large animals — canine, swine, sheep — are typically needed, because with thrombosis, you want to try to implant in the vasculature that mimic clinical use as much as possible (variety of species, method, test challenges). At the end, animal typically euthanized and scoring for thrombosis formation that may be present on the device.

Just to touch on genotoxicity as a category, most of the available tests are in vitro. There is a variety of non-mammalian Ames assay or reverse bacterial mutation assay or mouse lymphoma, chromosomal aberration for mammalian cell in vitro assays. In the US, there is a request for an in vivo assay, which is typically the mouse, micronucleus assay.

**Discussion**

**Practical questions on preparing extracts of medical devices**

In a device like a catheter, some pieces might only have indirect patient contact, such as external tubing and connectors. There are cases where only the fluid path sections would undergo testing; a simple example would be a hemodialyzer. But it’s not always black and white; there are case-by-case common sense decisions made according to what is practical. A wrinkle is that ISO standards encourage surface area tests whenever possible. Recent comments from some Asian regulatory agencies, however, indicate that surface area may provide less of a challenge than weight; in some cases they’ve asked for the sponsor to go back and do testing on a weight basis, finding that weight basis actually provided a greater challenge in some cases.

**How well do in vitro results translate to human clinical applications?**

Data is not readily available to show the percentage of tests that would test positive in animals/negatives in humans and vice versa. When we see positives in certain tests, they can be positive due to the test method. Let’s say for acute systemic toxicity, we might see a positive response with a bone graft pillar made of tricalcium phosphate. But clinically, when you put that in a bone it and it’s at a reasonable dose that you lose over time, nothing happens. In contrast, if there is a response in a sensitization assay, we can be reasonably confident that if we see some animals that are sensitized, a certain percentage of people are sensitive to a substance. We see different things that will give us a positive response, and those probably translate into a percentage of the population. But with some of the other assays, it’s important to look at the results and saying how well does that test method translate to the clinical use, and what’s the risk? Anytime you have positive responses, you have to ask, is it clinically relevant?

**Challenges and limitations of LLNA for medical devices**

The LLNA was originally designed to pick up positives, and not to have negatives, and so it leads to a lot of false positives … in fact, so many that the FDA says, if you’re using LLNA you almost always need to go to guinea pig test … so let’s shift to in vitro. This is now well accepted in cosmetic industry, throughout the FDA. Additionally, some of the original disapproval of the LLNA test stems from the fact that it was originally validated for single chemicals, and the vehicles were oil-based or non-polar vehicles. But there has since been additional work that has led a relaxing of that acceptance; for example, the pesticide industry has validated it for mixtures as well as for aqueous vehicles. Still, the FDA’s consensus page signals a case-by-case basis.

Coleman: Local lymph node assay; you mentioned FDA acceptance case by case. Could you explain what that means?
Hartung: What Alan said was mainly for human drugs, that’s where there is some hesitation.

From an animal welfare/discomfort standpoint, the LLNA is clearly a refinement method; there is less suffering than in the guinea pig maximization test. In that test, the guinea pigs can develop unpleasant sores sometimes. In fact, the ISO -10993 10, published in 2010, actually increased the number of rabbits from two-three for the intracutaneous reactivity test, which is not in line with animal welfare requirements. Concerning the LLNA, the FDA currently requests that we work within the ASTM standard, which requires the inclusion of two positive controls — a number of animals more than the OECD procedure. Concerning the acute systemic toxicity test, as it is performed according to the USP, control animals are needed. Now, however, there are three OECD methods available that might be included in the acute systemic toxicity test without including control animals. One participant suggested this could be a real way forward to reduce the number of animals. Some participants noted that the current in vivo options for medical devices are not up to date compared to testing done by the drug and chemical industries; it was noted that for medical devices, five animals are used, but for chemicals, only four, and also no mandatory positive controls were required for test for chemicals. The request, or expectation, for positive controls for in vivo device testing stemmed from the lack of data for devices. The conservative approach was to always have a positive, plus to spike with a positive control to see if there is any inhibition or enhancement. There was uncertainty surrounding how medical devices, which are essentially mixtures, might perform in the assay.

There is a disconnect between the language in the ISO Part 1 standard that recognizes a potential tiered approach to consider, giving more weight to the in vitro data – and the reality, which is that it is not being translated into regulatory decision-making. One participant suggested taking the core group that sees some merit in the in vitro methods, and discussing how to start using in vitro methods in a tiered fashion. This is a big task when you start to consider all the endpoints, because there are some special challenges with genotoxicity and for hemocompatibility, but one possible angle is to limit the discussion ways to use in vitro cytotoxicity to help make some decisions about whether to do acute systemic toxicity testing, or irritation testing or implantation testing.