In vivo Skin Irritation Test

In the 1940s, the rabbit Draize skin irritation test was developed to predict the human dermal irritation potential of raw chemicals or formulations (i.e., finished products) (Draize, 1944). During the second half of the 20th century it became the most widely used method for assessing skin irritancy (ECB, 1967; OECD, 2002; Phillips et al., 1972). However, in the 1970s the relevance of the Draize test came into question when studies found that its predictive accuracy was poor (Gilman et al., 1978; Nixon et al., 1975). This was due to the fact that rabbit and human skin are physiologically dissimilar and thus respond differently to irritants (Marzulli, 1975; Scott et al., 1991). In addition, the subjective nature of the test’s visual scoring made inter- and intra-laboratory reproducibility difficult (Weil and Scala, 1971). During the 1990s a number of studies compared human patch test results with Draize scores and found that the rabbit test was prone to frequent over-prediction and occasional under-prediction. False positive rates as high as 44%, and false negative rates of 5% were reported (Basketter, 1977, 1999; Basketter et al., 2004; Liebsch et al., 2000; Robinson, 1999). Nevertheless, the rabbit Draize test for skin irritation has remained the accepted method for many years.

Although regulators accept the Draize skin irritation test, it has never been scientifically validated (Jirova, 2007). Concerns about the test’s predictivity and reproducibility, plus animal welfare and political pressure in Europe (e.g., the Cosmetics Directive and REACH), have prompted a search for alternative test methods. That search began in the 1990s and led to the development of a number of in vitro skin irritation models, several of which have been validated by the European Centre for Validation of Alternative Methods (ECVAM) (Hayden, 2007; Spielmann et al., 2007).

In the 1980s a modified version of the Draize test was incorporated into the Tripartite Agreement test matrix for predicting the skin irritation potential of medical devices (OECD, 1988, 2002). In the 1990s, it was included in the ISO 10093 standards for this purpose as well (ISO, 2010). Skin irritancy testing became one of three biocompatibility tests recommended for all medical devices; the others being cytotoxicity and sensitization (ISO, 2009).

Over the years the Draize test has been widely used throughout the world to screen medical device materials, components, and products for skin irritancy. The test is time-consuming (2 weeks), expensive (up to $1,700 for 2 solvents and 3-4 rabbits), and rarely produces positive findings. For example, an internal Medtronic study of Draize test results back to the 1980s found that the company’s materials passed 99% of the time (Coleman, 2012).

ECVAM Validated Methods
In 2007 and 2008, ECVAM’s Science Advisory Committee (ESAC) approved the following reconstructed human epidermis assays as stand-alone replacements for the Draize rabbit skin irritation test: EpiSkin®, SkinEthic™ RHE (SkinEthic, Nice, France), and EpiDerm™ SIT (MatTek Corp., Ashland, Massachusetts) (Alepee et al., 2010; Hartung, 2007; Kandarova et al., 2009; Kreysa, 2008). In 2009, ESAC endorsed these three methods for testing under the United Nations Globally Harmonized System. By 2010 those same methods had achieved regulatory acceptance internationally as replacements for the rabbit skin irritation test (OECD, 2010).

Applicability to Medical Devices
In 2009, ISO Technical Committee (TC) 194 published 10993-1:2009. Biological evaluation of medical devices – Part 1: Evaluation and testing within a risk management process. This updated consensus standard specifically encouraged the use of alternative tests, stating: “In vitro test methods, which are appropriately
validated, reasonably and practically available, reliable and reproducible shall be considered for use in preference to in vivo tests" (ISO, 2009).

That advice, along with the previously cited issues, prompted the medical device industry to consider the ESAC-approved in vitro methods as potential alternative tests for medical devices. However, those tests were validated using pure chemical irritants, not dilute mixtures like those extracted from medical devices.

**Key Questions**
When considering the use of in vitro skin irritation assays for medical devices, two key questions arise: (1) are in vitro skin assays capable of identifying irritants in dilute medical device extract mixtures, and (2) do in vitro assays, which use apical application of extract solutions, measure the same end point as the modified Draize test which uses intradermal injection of medical device extract solutions? The first question will be answered by the feasibility and validation studies described on the subsequent pages. The second question is answered below.

**In Vitro vs. Intradermal Irritation Testing**
For in vitro assays, test article extracts are added to the apical side of the tissues. This topical application method has been shown to produce accurate and reliable predictions of skin irritation in humans (Kandarova et al., 2009). However, since many medical devices are implants, subcutaneous irritation also needs to be considered. This is accomplished by intradermal injection of polar and non-polar extracts of the medical device (ISO, 2010).

Chemical-induced primary skin irritation, manifested by erythema and edema, is the result of a cascade of events beginning with penetration of the stratum corneum and damage to the underlying layers of keratinocytes. The dying keratinocytes release mediators that begin the inflammatory cascade, which acts primarily on fibroblasts and endothelial cells in the dermis. It is the dilation and increased permeability of the endothelial cells that produce the observed erythema and edema (Welss, Basketter, and Schroder, 2004). The in vitro assays measure the initiating events in this cascade.

The epidermis is comprised chiefly of keratinocytes while the dermis is composed of fibroblasts, endothelial cells, and adipocytes (Marks, 2006; McGrath, 2004). Both keratinocytes and fibroblasts are involved in irritation response reactions. Cell viability is a primary marker of such reactions while inflammation serves as a secondary marker (Fentem et al., 2001).

Epidermal keratinocytes act as the main switch that turns on the inflammatory response in fibroblasts. This is accomplished by their release of primary cytokines such as IL-1 alpha and tumor necrosis factor alpha (Bernhofer et al., 1999). For this reason cell viability and IL-1 alpha release have both been measured during in vitro skin irritation assays.

Lastly, the symptomology induced by apical application or intradermal injection of irritants is identical; the fact that both exposure routes produce edema, erythema, itching, and pain confirm that their irritation mechanisms are comparable (Welss, Basketter, and Schroder, 2004).

It should be clear then that the in vitro skin irritation test and intradermal injection test are measuring the same endpoint, however, the route to that endpoint is slightly different.

**Feasibility Study**
A feasibility study was conducted to resolve whether in vitro assays are capable of identifying irritants in medical device extracts (Casas et al., 2013). The primary aim of this study was to determine if the EpiDerm™ reconstructed human skin model (MatTek Corp.) could be an acceptable alternative to the ISO 10993-required rabbit skin irritation test for assessing medical device biocompatibility. Eleven medical device polymers were tested. Four extracts were prepared per polymer, two each with saline and sesame oil; half were spiked with two...
R-38 irritants, lactic acid for saline extracts and heptanoic acid for the sesame oil extracts. Tissue viability was assessed by MTT reduction and the pro-inflammatory response was assessed by IL-1α release. LOAELs of 2% for lactic acid in saline and 0.7% for heptanoic acid in sesame oil were determined. A cell viability reduction of > 50% was indicative of skin irritation. Cells exposed to saline extracts spiked with 3.25% lactic acid had significantly reduced mean cell viabilities (12.6% – 17.2%). Cells exposed to sesame oil extracts spiked with 1.25% heptanoic acid also exhibited reduced mean cell viabilities (25.5% – 41.7%). All spiked cells released substantial amounts of IL-1α (253.5 pg/ml – 387.4 pg/ml) signifying a pro-inflammatory response. These results indicate that the EpiDerm™ model may be a suitable in vitro replacement for the assessment of the irritation potential of medical device extracts.

Validation Study
To confirm the feasibility study’s findings, ISO TC 194’s Working Group 8—which is responsible for the ISO 10993-10 standard on skin irritation and sensitization—agreed to sponsor an international round robin validation study (ISO, 2010). Fifteen medical device firms, government agencies, universities, and contract research laboratories expressed interest in participating in this effort. MatTek Corporation offered to provide free EpiDerm™ assay tissue training to these labs. This validation study will include spiked medical device extracts and positive control polymers. Protocol development and tissue training was completed during the fall of 2013. The validation study will begin in early 2014.

Final Steps
The validation study will be completed by the summer of 2014. After data analysis is complete a manuscript will be prepared for publication. If the round robin is successful then TC 194 Working Group 8 plans to seek ECVAM approval and change the ISO 10993-10 standard so that in vitro human skin assays become the normative method for irritation testing in the medical device industry.

Presentation
The title of this presentation is “Success of in In Vitro Evaluation,” but in truth, it is really a work in progress.

The Physiological Research Laboratories Biomaterials Dept. of Medtronic has existed for about 40 years; roughly half of Medtronic’s pre-clinical animal work done there. At the facility, located in Coon Rapids, Minnesota north of the twin cities, the only in-house ISO testing that is done there is hemolysis and cytotox screening. That testing is not done for submissions; it is called receiving and inspection, because the twin cities area has a lot of production lines, producing medical devices like catheters and such that require a lot of polymers. Before they can use them, they have to perform hemolysis and cytotox screening testing.

Testing in large animals, such as sheep, is used to test pacemakers, brain stimulators, and other things of that nature.

Medtronic’s Experience with Skin Irritation Testing
In the device industry, as Carraway did a great job explaining, primary irritation testing is similar to the old Draize test where it could be a topical application. That is not done often anymore; it is now pretty rare. Now, most of Medtronic’s skin irritation testing is intradermal injection, because most of the Medtronic devices are implants rather than surface devices.

The skin irritation test uses three animals and it does take some time. It is one of three ISO 10993 tests that are required or at least strongly encouraged. It requires the intradermal injection of three animals, for two weeks, at a cost of $1700.
Following an internal survey in 2009, Medtronic created a Biomaterials information database (BID) – a searchable relatable database cataloguing 40,000 reports going back to the 1970s of all animal tests done at Medtronic. Looking at irritation testing going back to the early 1980s, over 99% of the time they pass. On the one hand that is a really good thing; we do not want to create devices that’s causing an irritation response. Yet on the other hand, we keep doing this test, we’ve been doing it for decades, and we’re getting the same answer, spending this money and using these animals … and there must be a better way.

ECVAM Validated Methods

The birth of the initial idea to transition irritation tests on rabbits to one of the ECVAM methods came at the Society of Toxicology 48th Annual Meeting & ToxExpo in Baltimore in 2009. At the time, ECVAM had validated three RHE assays: EpiSkin®, SkinEthic™, and EpiDerm™. After discussing the possibilities with colleagues at CeeTox, who brought ideas to do feasibility and validation studies, Coleman went back to Medtronic and obtained a grant to launch a study, and began exploring some key questions.

In Vitro Model Questions

The ECVAM methods were validated for pure chemicals; medical device extracts are not. They are dilute mixtures, depending on how hard the polymers they contain are, but the concentrations are pretty low. The primary question was, are in vitro skin assays capable of identifying irritants that you would see in dilute medical device extract mixtures? Can fancy 3D tissues sniff out very low levels of irritants that you would see if they were extracted from medical devices?

The secondary questions related to primary irritation v. intradermal reactivity: Are we measuring the same endpoint? Is the in vivo endpoint that you would see with these tissues the same as you would see in vitro? (Reference: American National Standard, AAMI, Part 10.)

To answer this question, Coleman looked at the cellular response in skin. In the epidermal layer, there are keratinocytes. In the dermal layer, there are fibroblasts and endothelial cells. When you put an irritant on the surface of the skin, the stratum corneum, and it penetrate down into the keratinocytes it starts irritating them, killing some of them, and releasing a whole laundry list of some mediating factors: different cytokines, the IL-1α that are responsible for inflammation. It’s like flipping a switch, and that is the beginning of this irritation cascade. That is what happens on a surface application.

When you do an intradermal injection, sometimes intracutaneous – but you’re seeing the same endpoint, just through a different route. In irritation, the proof is in the symptoms. What are the symptoms of irritation seen in topical application v. intradermal injection? Redness, swelling, itching, pain — you’re getting the same symptoms from the two different routes, with the same chemical. Therefore, Coleman postulated that the same thing is happening inside the cells; it’s just getting there differently.

Feasibility Study (2010)

1. Assay Selection

- ECVAM-validated
  - Instead of using an open source method that had not been validated, we decided to use this.
- Ease of use
- Importation issues (we feared that if we bought tissues in France, it would take time, and also customs might x-ray them — photographic film test proved they did that sometimes)
- EpiDerm™ SIT – chosen because it was produced in the US
  - EpiDerm™ - a skin model grown from harvested human skin cells; it has been on the market for 20-25 years; the epidermal layers differentiate into layers in a way that are very similar to human
epidermis, but they are grown in a lab. Near the base, layers of cells are metabolically active; as they move up the eventually die and slough off.

- Rabbit skin tends to over-predict for irritation, because it is more porous than human skin, with maybe 1,500 hairs per square inch.
- These 3-D tissues also over-predict, for a different reason — they have no hair, they’re avascular — but because of a term that David Baskettter mentioned in one of his papers, called “lipid cement.” These 3-D assays do not have “lipid cement,” made of cholesterol, triglycerides, and more, which can be likened to mortar between the bricks. Human skin that has been growing for a long time builds up these substances between the cells that make the skin more impermeable to liquids that could come through it. The 3-D models lack this, and thus they are more permeable than regular human skin. Therefore, they tend to over-predict — not as dramatically as you see with rabbit skin, which might reach a 40% false positive rate; but still these models are more porous than regular human skin.

2. Materials Selection

- Extractions on polymers: It is difficult to extract anything from very hard polymers such as polycarbonates; more can be extracted out of soft polymers (silicons, rubbers, etc.). To ensure fairness in the irritation study, Medtronic took care to select samples that would be similar to what would be seen with extractions on medical devices.

**Gravimetric Extraction Results for Polymers**

![Diagram showing gravimetric extraction results for polymers]

Medtronic came up with a list of materials; opting not to use any metals or ceramics; you never see those things failing the tests. Instead, knowing that extractions are likely to be taken out of polymers or adhesives, about a dozen materials were selected: adhesives, soft polymers, silicon rubbers, and a couple of hard ones that we would extract (polycarbonate).

3. Irritant Chemicals

Next, in picking the irritant chemicals, Coleman’s team was mindful of polar and non-polar extractions, and also sought to ensure that they had a connection of some sort to medical devices or drugs.

- Lactic acid – used for drug delivery
- Heptanoic acid (used in steroid drugs; also an R34, corrosive)
- Both are R38s; these two strong irritants were selected, with the plan to dilute them down to the 1-4 % range

4. Spiking & Extraction
Biology v. Chemistry – When you do an extraction, you’re trying to simulate what might happen if you implant something. Where will it be implanted? In muscle or tissue? Will it be exposed to blood? Therefore you use saline, or vegetable oil. We’re trying to simulate with the chemical extraction processes what happens under biological conditions.

Spiking v. Extraction
The debate was, do we want to try to cook up some polymers that have irritant chemicals in them, or do we want to just extract some known materials that have passed the rabbit test, and do a titration curve – figure out how many irritant chemicals it would take to actually fail this epiderm assay, and take that amount and spike it into those extracts. The extracts would have dilute solutions that you come up with when you extract polymers and you get monomers, colorants, antioxidants, etc.

So in the interest of time, money, real quantifiable and reproducible concentration of these irritant chemicals, we decided to go with the spiking procedure. As far as extraction goes, performed the standard three-day, 37°C extraction process; the one unique element introduced was that at the end, we had two sets of extracts (two sets of polar, two sets of non-polar); one was kept as a control, the others were spiked with irritants.

The first step was a titration curve (really a range-finding study) to try to figure out the NOAELs and LOAELs for these two chemicals by using a range of concentrations, dosing them on these tissues, and looking at cell viability. In terms of results, Heptanoic acid came out at a lower number than lactic acid, not surprisingly, because it is a slightly nastier compound.

5. Range-finding Study

- Goal: Determine EC50s and NOAELs, using a range of concentrations

- Lactic acid
  - NOAEL=1%
  - EC50=2.87%

- Heptanoic acid
  - NOAEL=0.5%
  - EC50=0.85%

- MTT Range-Finding (4%)* Much higher concentration than what was needed for lactic and heptanoic acid to fail.
  - The graph below shows results at 24 hours and 48 hours. As hoped, the negative control (phosphate-buffered saline) is not killing the cells, and our positive control (1% Triton x-100) really wiped out these cells. There was no problem with the vehicles, saline or sesame oil. The two results that are most interesting are the lactic acid in saline, which it is highly soluble in, and heptanoic acid in sesame oil, which it is highly soluble in. At those levels, which are lethal levels, they are clearly killing the cells. For this test, a 50% threshold is the cut-off for irritation. If you kill more than 50% of the cells, that is clearly an irritant.
  - The other two results that you see are worth showing because of the solubility issue. Lactic acid is not very soluble in oil; thus, you see higher cell viability because the irritant is not really dissolved in the solvent; it’s more like salad dressing. Then, you can also see heptanoic acid, which is not very soluble in saline in an emulsion situation. Therefore, you see higher cell viability, as expected.
IL-1α Range-Finding (4%)

- This is like the mirror image of the MTT-Range finding; looking at IL-1α release instead of cell viability. In the positive control, you can see that the cells are kicking out a lot of IL-1α as a sign of stress. After that, heptanoic acid in sesame oil kicks out the most cells; not surprising as heptanoic is known to be a very nasty irritant.

6. Main Study

- Controls & Spiked Samples
- 3.25% Lactic Acid
- 1.25% Heptanoic Acid
- 100µL of Extract (standard dosing level) → Tissues
- 24-Hour Incubation @ 37 °C
- 25X Rinse with PBS
Starred bars signify significant IL-1α release. The results are similar for both substances: strong irritants are killing cells, the vehicles are not killing the cells, a lot of IL-1α is being kicked off when the cells are really stressed.

**Summary**

- EpiDerm™ SIT consistently detected low levels of two R-38 irritants in dilute medical device extract mixtures
- MTT was the best endpoint predictor
- IL-1α provides additional insights
- However, questions remain

*Results published, 2013 (Toxicology in Vitro): “In vitro human skin irritation test for evaluation of medical device extracts. JW Casas, GM Lewerenz, EA Rankin, JA Willoughby, Sr. LC Blakeman, Jim McKim, Jr., and KP Coleman.*
Validation Study

- ISO TC 194 Working Group 8 is sponsoring a validation study of this work
- 15 labs are interested; 8 have signed up so far
- Free EpiDerm™ training provided by MatTek
- Positive control polymers & spiked extracts
- Working with ECVAM contacts

Next Steps (2014)

It is hoped that the validation study will be completed and published by mid-2014; if successful, it will seek ECVAM approval and change ISO 10993-10 standard to make it an option.

Discussion

Is testing necessary if materials don’t contain severe irritants?
While some wondered whether or not irritancy testing would be needed if a material does not contain an R34 or higher severe irritant, the FDA likes for you to test the final devices. A survey of past Medtronic reports showed there were only a couple of oddball hydrogels that the material itself actually failed. Almost all the time, it’s some kind of processing aid, something in the way they were manufactured, or a contaminant rather than the base material itself. This in vitro round robin won’t be able to test pacemakers and brain stimulators; instead, the focus is on using materials that go into them, to try to figure out whether the assay will work.

Taking a look at the NOAEL numbers for the two irritants we used, heptanoic acid was one of the nastiest things we could find; it has an R34 and an R38 which means it causes temporary and permanent damage … and we were getting a .5% NOAEL, so I think you’d be challenged to find other stronger irritants. You could use sodium hydroxide, or other than acids for instance that are probably going to be nastier than that … there are going to be corrosives that are even worse than heptanoic acid, but it’s unlikely they’d be used in a medical device. Furthermore, you could have a polymer that has a strong irritant in it, but if it’s a hard polymer, it’s not going to come out.

Next steps in the validation process
Human skin equivalent in vitro irritancy tests are already included in Part 10, in an annex, with the promising results from this study, moving the skin results to the main part seems desirable. While it is uncertain whether or not the round robin study will be sufficient to provide evidence that this method can move in the main part, or that regulators can be convinced, as much information as possible has been collected to make that case.

The existing animal results already, the in vitro data, and the chemical extractions on these will combine to form a very strong suite of data on all these chemicals to write up in a report that tells a more complete story. That is something that has been missing; previous reports have had a lot of data gaps. So we’re trying to cover all the bases in a timely, financially responsible fashion.

More data is always better and there has to be a finite limit, since this method has been validated twice with a large number of chemicals in a very long-running assay (at least 20, 25 years), so it’s a pretty well known method. We’re trying to repurpose this method for a slightly different use. Our goal with this feasibility study was just to see if it could sniff out strong irritants at really low levels down in this LOAEL range concentration, which it did. So now we just want to expand this and use multiple laboratories to prove what we did was correct.

References


Hartung, T., 2007. Statement on the Validity of In-Vitro Tests for Skin Irritation Europoean Centre for the Validation of Alternative Methods, Ispra, Italy.


