



t4 Workshop Report*

From *In Vivo* to *In Vitro*: The Medical Device Testing Paradigm Shift

Dayna Kerecman Myers¹, Alan M. Goldberg¹, Albrecht Poth², Michael F. Wolf³, Joseph Carraway⁴, James McKim⁵, Kelly P. Coleman³, Richard Hutchinson⁶, Ronald Brown⁷, Harald F. Krug⁸, Anthony Bahinski⁹ and Thomas Hartung^{1,10}

¹Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; ²Dr. Knoell Consult GmbH, Mannheim, Germany; ³Medtronic plc, Minneapolis, MN, USA; ⁴NAMSA, Northwood, OH, USA; ⁵IONTX LLC, Kalamazoo, MI, USA; ⁶J&J Medical Devices, Preclinical Center of Excellence, Sommerville, NJ, USA; ⁷FDA, Center for Devices and Radiological Health, Beltsville, MD, USA; ⁸EMPA, Swiss Federal Institute for Materials Science & Technology, St. Gallen, Switzerland; ⁹Wyss Institute for Biologically Inspired Engineering at Harvard University, Boston, MA, USA; ¹⁰University of Konstanz, Konstanz, Germany

Summary

Amid growing efforts to advance the replacement, reduction, and refinement of the use of animals in research, there is a growing recognition that *in vitro* testing of medical devices can be more effective, both in terms of cost and time, and also more reliable than *in vivo* testing. Although the technological landscape has evolved rapidly in support of these concepts, regulatory acceptance of alternative testing methods has not kept pace. Despite the acceptance by regulators of some *in vitro* tests (cytotoxicity, gene toxicity, and some hemocompatibility assays), many toxicity tests still rely on animals (irritation, sensitization, acute toxicity, reproductive/developmental toxicity), even where other industrial sectors have already abandoned them. Bringing about change will require a paradigm shift in current approaches to testing – and a concerted effort to generate better data on risks to human health from exposure to leachable chemicals from medical devices, and to boost confidence in the use of alternative methods to test devices. To help advance these ideas, stir debate about best practices, and coalesce around a roadmap forward, the JHU-Center for Alternatives to Animal Testing (CAAT) hosted a symposium believed to be the first gathering dedicated to the topic of *in vitro* testing of medical devices. Industry representatives, academics, and regulators in attendance presented evidence to support the unique strengths and challenges associated with the approaches currently in use as well as new methods under development, and drew next steps to push the field forward from their presentations and discussion.

Keywords: *in vitro* testing, refinement, skin sensitization, threshold of toxicological concern (TTC), nanotoxicology

1 Introduction

Reflecting a wide spectrum of products, encompassing everything from surgical instruments, to catheters, to surgical gloves, to blood bags, to new frontiers like tissue engineering products and advanced therapy medicinal products, the medical device market is poised to grow dramatically. Worth \$390 billion in

2012, the industry is expected to expand to over \$476 billion in revenue by 2018¹. There is a high rate of research and development in this field.

As the market for medical devices grows, interest in *in vitro* testing alternatives to assess the safety of the devices is growing under the general push to advance the replacement, reduction, and refinement of the use of animals in research (the

*A report of t4 – the transatlantic think tank for toxicology, a collaboration of the toxicologically oriented chairs in Baltimore, Konstanz and Utrecht sponsored by the Doerenkamp-Zbinden Foundation; participants do not represent their institutions and do not necessarily endorse all recommendations made.

¹ Frost and Sullivan. Global Medical Devices Outlook 2013. <http://www.frost.com/sublib/display-report.do?id=NC8C-01-00-00-00>





3Rs). There is growing recognition that *in vitro* testing of medical devices can be more effective, both in terms of cost and time, than *in vivo* testing. However, acceptance of alternative testing methods has not kept pace with the technology. Regulators accept some *in vitro* tests (cytotoxicity, gene toxicity, and a few hemocompatibility assays), but many sensitization and irritation tests still rely on animals. Bringing about change will require a paradigm shift in current approaches to testing and a concerted effort to generate better data on risks to human health by exposure to medical devices, to share that data with regulatory authorities, and to boost confidence in alternative methods to test devices.

The Johns Hopkins Center for Alternatives to Animal Testing (CAAT) hosted a symposium in Baltimore, Maryland, in December 2013, which brought together a group of industry representatives, academics, and regulators ideally positioned to embrace new technologies and move the field forward. The unique strengths and challenges associated with *in vitro* approaches currently in use as well as new methods under development were considered and the next steps to advance the technology and the community's acceptance of these were determined.

2 Background

2.1 Medical devices and the ISO standards process

Medical devices are now essential to everything from diagnostics, to the prevention and treatment of diseases, to contraception. Many involve close contact with tissue or blood, or will even be implanted in the body. Therefore, to ensure that medical devices have no detrimental effects, they must undergo testing.

At this point, biological testing of medical devices relies mostly on animal models. However, as scientific knowledge advances our understanding of basic mechanisms, preference should be given to *in vitro* models – a point backed by the International Organization for Standardization (ISO): “*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 10993-1, 2009).

ISO Technical Committee 194 (TC 194) is comprised of working groups that develop and maintain standards and technical reports on biological and clinical evaluation of medical devices. These are the ISO 10993 standards on biocompatibility and the ISO 14155 (2011) standard on clinical trials.

Tab. 1: Examples of hazards modified from ISO 14971:2007

©ISO. This material is excerpted from ISO 14971:2007 with permission of the American National Standards Institute (ANSI) on behalf of the International Organization for Standardization. All rights reserved.

Examples of energy hazards	Examples of biological and chemical hazards	Examples of operational hazards	Examples of information hazards
Electromagnetic energy Line voltage – leakage current – enclosure leakage current – earth leakage current – patient leakage current Electric fields Magnetic fields Radiation energy Ionizing radiation Non-ionizing radiation Thermal energy High temperature Low temperature Mechanical energy Gravity – falling – suspended masses Vibration Stored energy Moving parts Torsion, shear and tensile force Moving and positioning of patient Acoustic energy – ultrasonic energy – infrasound energy – sound High pressure fluid injection	Biological Bacteria Viruses Other agents (e.g., prions) Re- or cross-infection Chemical Exposure of airway, tissues, environment or property, e.g., to foreign materials – acids or alkalis – residues – contaminants – additives or processing aids – cleaning, disinfecting or testing agents – degradation products – medical gasses – anesthetic products Biocompatibility Toxicity of chemical constituents, e.g.: – allergenicity/irritancy – pyrogenicity	Function Incorrect or inappropriate output or functionality Incorrect measurement Erroneous data transfer Loss or deterioration of function Use error Attentional failure Memory failure Rule-based failure Knowledge-based failure Routine violation	Labeling Incomplete instructions for use Inadequate description of performance characteristics Inadequate specification of intended use Inadequate disclosure of limitations Operating instructions Inadequate specification of accessories to be used with the medical device Inadequate specification of pre-use checks Over-complicated operating instructions Warnings Of side effects Of hazards likely with re-use of single-use medical devices Specification of service and maintenance

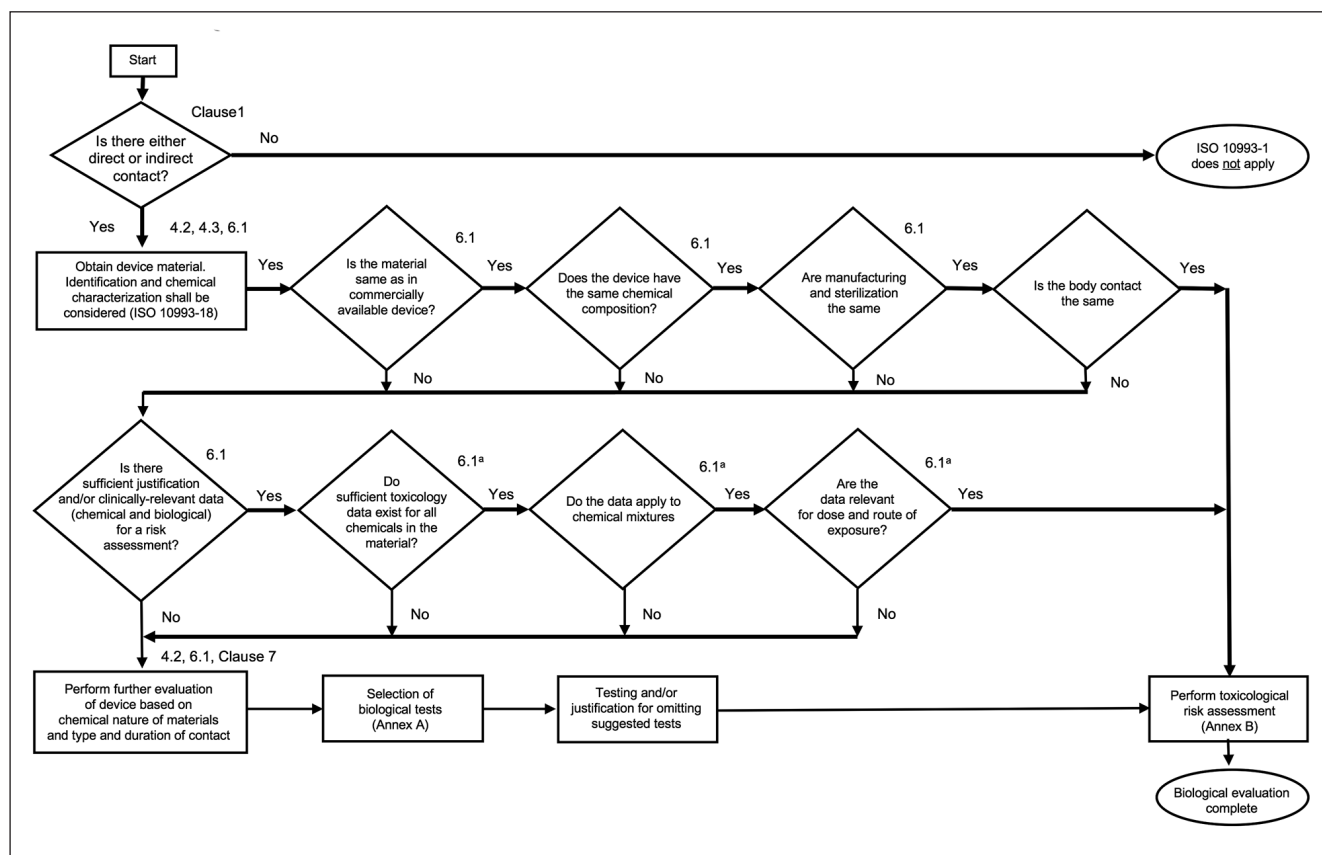


Fig. 1: Summary of the systematic approach to a biological evaluation of medical devices as part of a risk management process Modified from ISO 10993-1:2009(E) with permission of the American National Standards Institute (ANSI) on behalf of the International Organization for Standardization. All rights reserved. ©ISO 2009.

In addition, ISO TC 210 has developed standards to help medical device manufacturers implement and maintain a quality management system (ISO 13485, 2016) and manage risks (ISO 14971, 2007). The ISO 14971 standard addresses the application of risk management ideals to medical devices, to consider potential hazards (see Tab. 1) ranging from energy-related, to biological (bacterial, viral contamination) and chemical, to operational (errors in use), to information hazards (labeling, operating instructions). The risk management process has to be documented and includes risk analysis, risk evaluation, risk control, and post-production information.

Sets of tests may be necessary to determine the potential adverse effects of medical devices. To determine the best course, ISO 10993-1 (2009) offers a flow chart, reproduced in Figure 1, which guides qualified toxicologists through a series of key questions surrounding the chemical composition of the medical device, any material information, and residual chemicals from the manufacturing and cleansing process as well as from sterilization (work has started on the revision process of this part, the new version is expected by 2019).

2.2 Evaluating chemical risk

The ISO TC 194 working groups have created a number of standards to guide various aspects of the oversight of medical

devices – such as allowable limits of leachable substances, systemic toxicity, irritation/sensitization, genotoxicity, and effects on blood. Other standards cover chemical characterization and degradation product identification.

Chemical characterization involves identifying a device's materials or component chemicals and determining how much of them might leach into a patient during use. This information is provided to a toxicologist who conducts a risk assessment to determine whether any of the leachable chemicals might pose a health risk to patients. ISO 10993-17 (2002) provides guidance on establishing allowable limits for such leachable substances in a device and consideration of whether or not the material is equivalent to that already in use in another device, on quantitating composition of the medical device and on making the risk assessment. If a risk is identified, it becomes necessary to obtain (clinical) exposure and hazard data through *in vivo* tests. The systematic approach poses a series of questions to help determine whether biocompatibility testing is needed – for example, is sufficient justification or data available for a risk assessment? If yes, does sufficient data exist to determine the chemical risk? If yes, does this data apply also for mixtures? If yes to all of those questions, a final conclusion can be made without conducting any further biocompatibility testing. Otherwise, further evaluation of the device, based on



Tab. 2: Prominent points of comparison between the United States and the European Union for approval of medical devices
Modified from Kramer et al. (2012).

System feature	United States	European Union	Potential implications
Mandate	Oversight of public health	Device safety (overseen through Competent Authorities), device approval (through Notified Bodies), and facilitation of trade	May influence dealings with industry clients, and attention paid to balance between effectiveness and risk of safety concerns
Centralization	Oversight of all device regulation by the FDA	Directives outline processes carried out by Competent Authorities and Notified Bodies	Standardization and coordination of premarketing and postmarketing evaluation are theoretically simpler and easier to enforce in the US
Data requirements	Reasonable assurance of safety and effectiveness for approval of high-risk devices, “substantial equivalence” for 510(k) clearance	Generally performance-based analysis, requiring proof that device works as intended	EU assessment made by manufacturers and Notified Bodies; provides less insight into clinical end points for high-risk devices
Transparency	Proprietary limits with public reporting of premarketing review of approved devices, recalls and adverse events	Review of Notified Bodies not made public; postmarketing data shared among Competent Authorities but not with the public	Greater public access to evidence in the US
Funding	Combination of federal appropriations (80%) and user fees (<20%)	Funding of Competent Authorities variable among countries; Notified Bodies paid directly by sponsors	Notified Bodies may be vulnerable to conflict of interest with industry client; the FDA may be influenced by changes in federal funding and political climate
Access	Clinical premarketing testing of high-risk devices delays patient access to these devices (no differences for low- and moderate-risk devices)	EU patients may have access to certain high-risk devices sooner than in the US, subject to limitations by payers	EU patients have faster access to certain devices, but these products are marketed with less rigorous proof of effectiveness and may have a greater chance of later-identified adverse events

the chemical nature of the materials and the type and duration of contact, is necessary; after selecting the toxicological endpoints, testing – or justification of the omission of certain testing – must be performed. Then the final biological evaluation can be conducted.

The ISO standards advise manufacturers to screen candidate materials early on in the development process to select materials that are sufficiently biocompatible and non-toxic, although it is still necessary to conduct biological tests on the finished device under actual conditions of use. In addition, the standards advise that any existing non-clinical and clinical data or human exposure data, as well as any historical data from similar devices or experience relevant for the medical device should be evaluated thoroughly prior to any further testing.

For a biological evaluation, many factors need to be taken into account, including effects based on the manufacturing process that can lead to changes in a device’s surface (e.g., by surface treatment or welding); additives (color pigments, lubricants, inks); contamination by cleaning, disinfection, and sterilizing agents; degradation by manufacturing, clinical use or storage; material changes – such as the abrasion and aging of materials – which can lead to the generation of particles

and degradation of materials such as prostheses; heat; and interactions with the physiological environment. For example, endoscopes can be affected by gastric fluid, dental implants can produce wear particles, and wound-dressing materials may come into contact with sweat. Many situations – such as those involving the human digestive system, or dental implants – are difficult to reproduce in a standardized way in an animal model, highlighting the need for standardized *in vitro* models.

2.3 Comparison of classification systems

The ISO 10993-1 standard incorporates language that specifically captures a tiered approach and a preference for *in vitro* over *in vivo* testing, but a gulf persists between the growing support for evolving *in vitro* methods and their actual use. Neither the EU, the US, nor Japan have reached the point of including *in vitro* testing as an acceptable replacement for *in vivo* methods for medical devices, though such tests may be used as a source of supportive data.

Classification systems for medical devices are risk-based and vary by country (Tab. 2). The centralized US approval process, under the Food and Drug Administration (FDA), requires rea-

Tab. 3: Classification of medical devices in Europe, US, and Japan

Modified from table provided by Harlan Laboratories.

Europe	US (FDA)	GHTF (Japan)	Examples of included medical devices
Class I	Class I*	Class A	Non-sterile items or sterile items with a low potential risk: surgical instruments, urine bags, stethoscopes, examination gloves
Class IIa	Class II	Class B	Sterile items, surgical gloves, urinary catheters, stomach tubes, needles, tracheal tubes, IV giving sets
Class IIb	Class II	Class C	Blood bags, condoms, non-absorbable sutures, anesthesia machines
Class III	Class III	Class D	Absorbable sutures

*with or without GMP

sonable assurance of safety and effectiveness for approval of high-risk devices². The EU's more decentralized approach, in contrast, generally requires proof through performance-based testing that the device works as intended. In terms of access, pre-marketing testing of devices in the US delays patient access, even when devices are thought to be of low or moderate risk. EU patients may have access to certain high-risk devices sooner than in the US, but such devices may be marketed with less rigorous proof of effectiveness.

In the US, medical devices are classified as Class I (mostly exempted from clearance but subject to general control requirements), II (general and specific controls), or III (pre-market approval) devices – with Class III devices representing the highest risk. In the EU, medical devices are classified as Class I (low-risk), IIa, IIb, or Class III (intervention of Notified Body is compulsory to control the design and the manufacture) devices, while active implantable medical devices are not classified, and *in vitro* diagnostic devices have their own classification system. Japan has a letter-based classification system, as illustrated in Table 3.

There are practical hurdles to broader acceptance of *in vitro* tests. For example, regarding biocompatibility testing, ISO 10993-1 recommends starting with *in vitro* tests, but given the pressure to get products to market quickly, manufacturers often feel they have no time to wait for such lengthy tests. Funding, staffing and management issues are all potential roadblocks to efforts to try to speed up the ISO process to better reflect the speed of technology. The ISO process is also slow because of its focus on reaching a common solution, which is often a lengthy and arduous process as different stakeholders are involved in the committees, complicating decision-making. The lack of a global harmonized approach hurts the ISO process; for example, the US will not accept an *in vitro* test for irritation, so developers are required to perform an *in vivo* test if they want to take a product to market. Harmonizing ISO and alternative assessment paradigms also could reduce duplication of efforts.

3 Currently used *in vivo* testing methods

Currently, ISO 10993 allows for *in vitro* methods, but negative test results – whether in a sensitization or irritation assay – must be confirmed by an *in vivo* assay. As most tests for medical devices do produce negative results, sponsors are typically reluctant to conduct two sets of tests and perform only the *in vivo* tests.

A risk-based approach to considering the use of animals in medical device testing takes into account the type and duration of contact and extraction procedures. For example, devices that only entail surface contact carry less potential risk than implantable devices, and the longer the exposure, the greater the risk and the need for more tests. Devices that only have skin contact require basic cytotoxicity, sensitization, and irritation or intracutaneous reactivity testing, whereas devices that come into contact with bone and tissue require more extensive testing.

Table 4 offers a framework to develop an assessment program. Essentially, the more invasive the device, and the longer the contact, the more toxicological endpoints need to be considered. Biological evaluation tests include cytotoxicity, sensitization, irritation, acute systemic toxicity, subacute/subchronic toxicity, genotoxicity, implantation, and hemocompatibility.

Medical devices, which typically involve an unknown mixture of materials, pose some challenges that are different from pharmaceutical and chemical testing. Drug testing involves preparing a solution of the substance where the concentration is controlled. Devices, in contrast, are extracted to perform safety testing using polar and non-polar solvents, as per guidelines described in ISO 10993-12 (2012). Standard solvents such as saline, water, culture media, and vegetable oil are used to extract chemicals or materials. Medical devices and components tend to be extracted on a surface area to solvent volume basis, however, if surface area cannot be calculated, weight to solvent volume is used.

There are some challenges to the standard methods for conducting extractions. For example, when an extract is created based on weight, the device is often cut into small pieces and

² <http://www.fda.gov/medicaldevices/productsandmedicalprocedures/deviceapprovalsandclearances/510kclearances>: “Section 510(k) of the Food, Drug and Cosmetic Act requires device manufacturers who must register, to notify FDA of their intent to market a medical device at least 90 days in advance. This is known as Premarket Notification – also called PMN or 510(k).”



Tab. 4: Devices, type of contact, and potential biological effects/endpoints

Device categories		Initial evaluation									Supplemental evaluation	
	Body contact	Contact duration	Cytotoxicity	Sensitization	Irritation or intracutaneous reactivity	Systemic toxicity (acute)/pyrogenicity	Subchronic toxicity	Genotoxicity	Implantation	Hemocompatibility	Chronic toxicity	Carcinogenicity
Surface devices	Skin	A	●	●	●							
		B	●	●	●							
		C	●	●	●							
	Mucosal Membranes	A	●	●	●							
		B	●	●	●	○	○		○			
		C	●	●	●	○	●	●	○		○	
	Breached/ compromised surface	A	●	●	●	○						
		B	●	●	●	○	○		○			
		C	●	●	●	○	●	●	○		○	
External communication devices	Blood path indirect	A	●	●	●	●				●		
		B	●	●	○	●	○			●		
		C	●	●	●	●	●	●	○	●	●	●
	Tissue/bone/dentin communicating	A	●	●	●	○						
		B	●	●	○	○	○		●			
		C	●	●	○	○	○		●		○	●
	Circulating blood	A	●	●	●	●				●		
		B	●	●	●	○	○	●	○	●		
		C	●	●	●	●	●	●	○	●	●	●
Implant devices	Bone/tissue	A	●	●	●	○		○				
		B	●	●	○	○	○	●	●			
		C	●	●	○	○	○	●	●		●	●
	Blood	A	●	●	●	●			●	●		
		B	●	●	●	●	○	●	●	●		
		C	●	●	●	●	●	●	●	●	●	●

A, limited exposure (5-24 hours); B, prolonged exposure (24 hours-30 days); C, permanent contact (> 30 days)

● FDA and ISO evaluation tests; ○ Additional tests for FDA

placed in a vial. However, a coiled catheter, for example, could have a braided wire at its core – and cutting it up might release metals that a patient would not ordinarily be exposed to during typical clinical use, adversely impacting the test results. Instead, for this type of device it would be preferable to extract the intact catheter based on surface area instead of weight.

Three common sensitization tests require the use of animals: the guinea pig maximization test (the most widely recognized, which uses 15 guinea pigs per extract), the Buehler closed patch test (also with guinea pigs), and the murine local lymph node assay (LLNA). For the guinea pig maximization test, intracutaneous injections and topical applications of polar and

non-polar medical device extracts are used for the induction phase, followed later by topical applications of extracts for the challenge phase. The appearance of the challenge skin sites is subsequently observed at 24 and 48 hours and given a score. From an animal welfare/discomfort standpoint, the LLNA is clearly a refinement method as there is less suffering than in the guinea pig maximization test, which can lead to unpleasant sores. However, the FDA currently requests work within the ASTM standard for the LLNA, which requires the inclusion of two positive controls – more than the OECD procedure. The LLNA also uses 15 animals per extract. Its variability has been assessed (Hoffmann, 2015).

For skin irritation, the Draize rabbit skin irritation test is currently the preferred method. For implanted devices, intracutaneous injections of polar and non-polar extracts of medical devices are used. At 24, 48, and 72 hours the injection sites are examined for erythema and edema and scored. This method typically uses three adult rabbits.

The acute systemic toxicity test generally requires control animals, although three more recent OECD methods (TG 420, 423, 425) do not – offering a way to reduce the number of animals. Essentially, the medical device industry trails the drug and chemical industries in *in vivo* options; medical devices use five animals versus four for chemicals, which also do not require positive controls. The lack of data for devices drives the expectation for positive controls for *in vivo* device testing. The typical test for acute systemic toxicity, developed for pharmaceutical containers, involves injecting mice with a maximum tolerated dose, typically an injection of around 50 ml/kg of an extraction or an actual liquid sample – an enormous volume, considering that blood volume in comparison for an average adult human is about 70 ml/kg, i.e., we are injecting almost an entire blood volume into our test animals. Such a high volume is injected to improve sensitivity, but such high volumes could not be imagined in drug or chemical testing. For medical devices, the doses used are not usually tiered, i.e., starting with one dose and increasing/decreasing the next dose depending on outcome as commonly done for testing chemicals. It is also rare to see an acute toxic response in the oil extract and any responses are almost always seen in the saline extract, giving opportunities for reduction. Given that the extract is given as a bolus injection in a matter of a minute, any disbalance of the electrolytes induces cardiac arrest – making this a crude screen.

There are many potential *in vitro* methods that could be used instead of the *in vivo* tests described above to assess the biological effects of implantation devices (see Section 4). It can be a little more difficult, however, to address the effects on the living tissue surrounding an implanted device with *in vitro* methods. Muscle tends to be used as a surrogate regardless of where a device is meant to be implanted, and rabbits tend to be a common implant species, although the size of the device can determine what species is used. However, tissue engineering technology, developed to replace or repair damaged tissue, is advancing, offering alternate testing systems for testing implantation devices. Tissue engineering combines cells with a porous scaffolding composed of biomaterials to support the regenera-

tion of damaged tissue – but can also facilitate the analysis of material-tissue interaction.

The five main areas of consideration for hemocompatibility tests are: thrombosis, coagulation, platelets, hematology, and complement. Most hemocompatibility tests can be done *in vitro*. 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 to be implanted in an appropriate animal model in the same fashion as described in the product information for use (IFU) instructions. This is critical because design features, supportive drugs such as anticoagulants and antiplatelet agents, and surface geometry and material characteristics can influence clinical performance and outcome. The main ISO guidance document for this area, ISO 10993 Part 4: selection of tests for interaction with blood, was published in 2017.

4 Newer *in vitro* tests for irritation, sensitization and acute toxicity

While the medical device industry accepts some *in vitro* tests (cytotoxicity, genotoxicity, and hemocompatibility) for biocompatibility already (Coleman et al., 2012), many toxicity endpoints still require animal tests – including eye and skin irritation (most often using rabbits), skin sensitization (most often using guinea pigs), and in this industry to a very limited extent acute systemic toxicity (frequently using rodents). And yet, according to medical device manufacturers among the authors, the vast majority of *in vivo* eye irritation, dermal irritation, and sensitization tests produce negative results. Given that a large number of animals is used to demonstrate a negative result, combined with the existence of new *in vitro* technology that tends to be more accurate, there should be a way to flip the paradigm and accept negative *in vitro* results. Furthermore, the recent advance of some newer types of devices, such as the first lab-grown organ, a bladder, to be implanted into a human (Atala et al., 2006), which is not compatible with animal testing, signals a need to develop new testing methods, and new regulatory guidelines around the process.

Newer and more advanced tools and methodologies are being introduced at a rapid rate, including several novel assays to assess dermal sensitization and irritation (see Section 5), and to predict an oral LD₅₀ for measuring acute toxicity that could be applied for testing the safety of medical devices.

Several *in vitro* sensitization tests recently accepted by OECD following validation (TG 442C,D,E) show promise for their ability to predict chemical sensitizers. These include: the Direct Peptide Reactivity Assay (DPRA) – a cell-free assay that uses two tri-peptides, one with a cysteine residue and the other with a lysine residue, to measure reactivity of test compounds (Gerberick et al., 2004); the human cell line activation test (h-CLAT), which uses a human monocytic leukemia cell line (THP-1) and measures expression of CD85/54 (Sakaguchi et al., 2009); and the KeratinoSens™ assay, which uses a human keratinocyte cell line (HaCaT) and the activation of the Nrf2/Keap/ARE signaling pathway (Natsch, 2010; Natsch et al. 2011).



One method that has shown promise for predicting the acute oral LD₅₀ value is an acute toxicity assay developed in partnership by L'Oréal and CeeTox: The assay uses a rat hepatoma cell line (H4IIE), combined with concentration response measuring cellular markers of cell health and receptor binding. An important component of this assay is the incorporation of physical and chemical data that describe solubility, logP, and pKa. The assay has been evaluated in multiple blinded studies with more than 200 chemicals that covered a wide application domain. The sensitivity and specificity of the assay ranges between 84–90% depending on the type of chemicals tested. This system may be an important test for medical device extracts, though current need for acute toxicity data is rather limited; follow-up work includes improving efficiency for screening and making the assay cost-effective enough for practical use (McKim et al., 2012). “Organ-on-a-chip” systems – microchips seeded with human cells – offer one approach to replace acute systemic toxicity testing on animals (Marx et al., 2016). Using microscale technologies, organ-on-a-chip systems mimic human organ systems and replicate physiological functions. In the end, a combination of approaches may provide the most accurate and reproducible data and allow for chemicals to be placed into potency categories. The US NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) is sponsoring a multi-year program to identify and validate *in vitro* replacements for *in vivo* acute systemic toxicity tests. The US medical device industry is participating in this process (Hamm et al., 2017).

5 Example of a successful *in vitro* evaluation

Skin irritancy, along with cytotoxicity and sensitization testing, is one of three ISO 10993 biocompatibility tests recommended for all medical devices, and the Draize rabbit skin irritation test has been widely used throughout the world to screen medical device materials, components, and products.

However, prone to frequent over-prediction and occasional under-prediction, the Draize assay is not ideal. Rabbit and human skin are physiologically dissimilar and thus respond differently to irritants (Marzulli and Maibach, 1975; Scott et al., 1991). The test is also expensive and time-consuming (the two-week test costs \$1,700), and has never been scientifically validated. Despite reports of false positive rates as high as 44%, and false negative rates of 5%, it has remained the accepted method for many years. A review of irritation test results going back to the early 1980s found that over 99% of the time medical devices pass this test. After decades of conducting the same costly, time-consuming test and arriving at the same answer, Coleman and others (2012) have argued that there must be a better way. Ultimately, concerns about the test's predictivity and reproducibility, along with animal welfare concerns and political pressure in Europe, prompted a search for alternative test methods (Natsch and Emter, 2008).

That search, which began in the 1990s, led to the development of a number of *in vitro* skin irritation models – and the European Centre for Validation of Alternative Methods (ECVAM) has

validated several (Hayden, 2007; Spielmann et al., 2007). The ISO TC 194 recommendation to seek and give preference to *in vitro* test methods (ISO, 2009), along with the previously cited issues, prompted the medical device industry to consider the OECD-accepted *in vitro* methods for skin corrosion (TG 431), skin irritation (TG 439), skin sensitization (TG 442c-e), skin absorption (TG 428), phototoxicity (TG 432) and eye irritation (TG 437, 438, 460, 492) as potential alternative tests. However, those tests were validated using pure chemical irritants, not dilute mixtures like those extracted from medical devices. Additionally, many medical devices are implants, rendering the topical application methods typical for *in vitro* assays insufficient. Subcutaneous irritation also needs to be considered, and can be accomplished by intradermal injection of polar and non-polar extracts of the medical device (ISO 10993, 2010). The *in vitro* skin irritation test and intradermal injection test measure the same endpoint, however, the route to that endpoint is slightly different.

The birth of the initial idea to transition irritation tests on rabbits to one of the ECVAM-validated *in vitro* methods for medical devices came at the Society of Toxicology's 48th Annual Meeting in Baltimore in 2009. At the time, ECVAM had validated three RHE assays: EpiSkin[®], SkinEthic[™], and EpiDerm[™]. Medtronic began exploring some key questions: Are ECVAM-validated *in vitro* skin assays for pure chemicals capable of identifying irritants in dilute medical device extract mixtures? Can sophisticated 3D tissues identify very low levels of irritants that you would see if they were extracted from medical devices? And, would the *in vivo* and *in vitro* endpoints seen with these tissues match?

To answer the first question, a feasibility study explored whether or not *in vitro* assays are capable of identifying irritants in medical device extracts using the EpiDerm[™] reconstructed human skin model. Extracts were prepared for eleven medical device polymers, and half were spiked with irritants. All spiked samples caused the release of substantial amounts of IL-1 α (253.5 pg/ml - 387.4 pg/ml), which suggests a pro-inflammatory response. EpiDerm[™] consistently detected low levels of two R-38 irritants in dilute medical device extract mixtures. These results indicated that the EpiDerm[™] model could be a suitable *in vitro* replacement to judge the irritation potential of medical device extracts. Medtronic is working to expand these tests, seeking multiple laboratories to back up its results. As one example, a feasibility study using an *in vitro* skin assay to test medical device extracts conducted by Kelly Coleman provides an example of how a successful *in vitro* evaluation might offer an alternative to the Draize rabbit skin irritation test (Casas et al., 2013).

The question whether 3D tissues identify very low levels of irritants has been addressed recently. ISO TC 194's Working Group 8 – which is responsible for the ISO 10993-10 standard on skin irritation and sensitization – sponsored an international round robin validation study to confirm the feasibility of the pilot project findings. The round robin, which was completed in December 2016, involved over 20 global laboratories that tested several irritant and non-irritant containing polymers on two commercially available reconstructed human epidermis (RhE) tissue models (unpublished). The successful round robin is now

open for vote in TC 194 Working Group 8 to 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. ISO has approved a New Work Item Proposal (NWIP) for a new ISO 10993-23 standard for Determination of Skin Irritation of Medical Device Extracts using Reconstructed human Epidermis (RhE).

The last question can be answered by looking at the cellular response in skin. In the epidermal layer, there are keratinocytes. In the dermal layer, there are fibroblasts and endothelial cells. When putting an irritant on the surface of the skin, it penetrates down to the keratinocytes where it causes damage, killing some cells and releasing cytokines (e.g., IL-1 α) that are responsible for inflammation. This is like flipping a switch, which sets off an irritation cascade. With an intradermal injection, sometimes intracutaneous, the same endpoint is reached – just through a different route. In irritation, the proof is in the symptoms. In numerous studies, the symptoms of irritation – redness, swelling, itching, pain – were seen in both topical application and intradermal injection, leading to the conclusion that the same reactions were happening at cell level. Thus, there is an opportunity to introduce more humane upstream endpoints as well as for cell models to reflect the adverse outcome.

6 Exploring the use of human cells and tissues to replace animal models

The field has largely failed to take advantage of modern assays to test human proteins, human cells, and human responses to medical devices – over-relying, instead, on animal models, which in many cases are not only expensive, but are not very predictive of the kind of response that humans give.

At the highest level, the controversy over whether an *in vitro* test can predict *in vivo* performance, or an acute *in vitro* test can predict long-term performance correlation might be addressed within the standards already in place. For example, mechanisms of genotoxicity, carcinogenicity and cytotoxicity exist in the standards today (ISO 10993-3 (2014) and ISO 10993-5 (2009)), and these are mechanisms that are not any less or more complicated than thrombosis (ISO 10993-4, 2009), yet this part has been slow to incorporate contemporary *in vitro* screening methods. Thus, while there is solid support that *in vitro* tests are helpful in screening devices for safety and risk analysis (Wolf and Anderson, 2012), standards are slow to keep up with contemporary methods.

To catch up with modern tools and methods of the 21st century, members of the Working Group 9 behind the ISO 10993-4 development are preparing for a 2016 round robin study on methods that use human blood and contemporary clinical coagulation assays to evaluate blood-device/material interactions (van Oeveren, 2013). A driving force for this work is to develop an *in vitro* alternative to the non-anticoagulated venous implant (NAVI) model – a test frequently called for to assess material

thrombogenicity in medical device regulatory submissions to the FDA (Michael Wolf, Medtronic, unpublished³). The NAVI test involves catheter-shaped devices or medical materials placed into the venous vasculature of large animals for 1–4 hours, followed by basic scoring for extent of visible thrombus (Wolf and Anderson, 2012). In the *in vitro* study, six materials recognized to give low, intermediate, and high responses in the NAVI model will be tested in three different *in vitro* models (e.g., test tubes and closed loops). Here testing will involve fresh human blood taken from a representative pool of human donors. Importantly, the final evaluation method will include standard statistical methods (ANOVA and comparison of all tests) rather than simple subjective scoring. *Toxicology In Vitro* has offered to publish a special edition about the *in vitro* irritation round robin study, for which 6–7 manuscripts and an editorial are currently submitted (Kelly Coleman, personal communication).

7 Novel developments in toxicology

Validation efforts have delivered the evidence that new approaches do not necessarily lower safety standards and can be integrated into regulatory safety assessments, especially by using integrated testing strategies (ITS) rather than relying upon a single test or fixed batteries of tests (Hartung et al., 2013; Rovida et al., 2015a). In addition, alongside efforts to embrace new technologies and promote the “Tox21” push to move toxicology from the twentieth century into the twenty-first, the concept of evidence-based toxicology provides a quality assurance component (Hartung, 2009).

And yet, the field of medical device safety testing has been slow to embrace *in vitro* methodologies, in part because of some limitations inherent in *in vitro* testing and the slow pace of developing many of the alternative assays.

Other important factors to consider for medical device testing include product degradation, the inflammatory process or mechanical use of these materials – which can cause tearing – and the challenges of dealing with mixtures.

Additionally, efforts to develop tests for medical devices, such as a whole blood pyrogen assay, expose the shortcomings of standard methodologies in removing contaminants from medical devices (Hartung, 2015). Very often, it is the contaminants, rather than the materials themselves, that produce problems with medical device testing, as some examples related to pyrogen testing demonstrate (Hasiwa et al., 2007; Mazzotti et al., 2007). Pyrogens, e.g., bacterial remnants that cause fever in humans, are often absorbed onto surfaces, producing surface effects and prompting immune reactions. Complicating testing efforts even further, they are batch-dependent because typically there is a lack of consistency in the contamination of devices with bacterial materials during production.

Another issue is one of replacement, reduction, and refinement: the traditional tests for pyrogens still call for the use of about 300,000 rabbits per year worldwide. One of the most

³ <http://www.fda.gov/downloads/MedicalDevices/NewsEvents/WorkshopsConferences/UCM397145.pdf> (last accessed 7 Aug 2016)



successful alternative methods, the Limulus assay, has replaced about 90% of pyrogenicity testing. It is a very successful assay even though it has never been formally validated, and it is restricted to Gram-negative pyrogens (Hasiwa et al., 2013). Significantly for testing biofilms, for example, it misses all pyrogens from Gram-positive bacteria. It also does not cover fungi, and does not reflect the relative potency of the different pyrogens in humans (Hasiwa et al., 2013).

A newer assay, based on human blood, (Hartung and Wendel, 1995, 1996; Fennrich et al., 1999; Schindler et al., 2009; Hartung, 2015) involves simply bringing the material – whether it is a drug or a solid material – in contact with blood, and measuring cytokine release. Proving that alternative methods can outperform animal tests, the validation study showed clearly that the assay is more sensitive than the rabbit assay (Hoffmann et al., 2005). It is also a quantitative assay; numerous studies showed that it is a less variable assay, and less prone to mistakes than rabbit assays. Unlike current tests, which do not adequately control pyrogen contamination of medical devices because they do not measure on the surface, the whole blood assay can be performed in direct contact with the medical device instead of using extracts. As human blood is used, the test reflects the fever response of humans – which is not the case for the Limulus assay. The pyrogenicity assay seems to be especially useful in testing nanomaterials (Hartung and Sabbioni, 2011), as the whole blood assay can handle nanoparticles, which cannot be tested in the Limulus assay.

After the validation study, the assay advanced into FDA acceptance as a Monocyte Activation Test (MAT) – though with some limitations. The FDA accepts it only as a replacement for the Limulus assay, not for the rabbit assay, and does not believe that the validation study proved coverage for other pyrogens. This is one of the big hindrances for the assay (Hasiwa et al., 2013; Hartung, 2015). It is commercially available also using cryo-preserved human blood (Schindler et al., 2004, 2006).

An important lesson learned from the MAT validation study is that commercial concerns cannot be ignored (Hartung, 2015). An assay that is not commercially available has a low level of standardization, and a number of challenges face the researchers and the labs that wish to carry it out. Without kits being available, it is very difficult for most laboratories to implement this kind of assay.

Its limited acceptance in the US – and the hesitancy of other countries to adopt it – leads to very limited use of the assay. Testing in the EU still uses 170,000 rabbits per year, which could all be replaced. There is not a single example so far for any product where pyrogen testing in animals could not have been replaced, or where the material was not compatible with the non-animal tests.

As for the limitations for *in vitro* medical device testing in general, the length of time it takes to develop many of the alternative assays is an impediment to progress. Fifteen years from contriving and publishing the assay to its regulatory acceptance is too long and costly for fast-moving industries and their testing needs. Validation has served us for traditional tests, but no one wants to wait that long for the next assay to be approved. Now is the time to start to support some of these alternative method-

ologies, and to use them to support regulatory decision-making. Getting the pyrogen test accepted by ISO is a matter of defining the results the test needs to show, as well as the availability and suitability of test materials.

A pragmatic step might be to look at the toxicological endpoints mentioned in ISO 10993-1 (2009), and determine how to include *in vitro* tests within these toxicological endpoints that must be considered for medical devices. For certain toxicological endpoints, for example for irritation or sensitization, the likely conclusion would be that the necessary systems are already available, which may not be the case for other toxicological endpoints.

Industry cooperation – in the form of providing more devices for testing – could also be key. Regulators, including the FDA, are interested in seeing the *in vitro* data. Now, the industry needs to come together and begin to share data with regulators, and that includes international harmonization. One way forward is to develop a mechanism to provide *in vitro* data to the FDA, perhaps in the context of ISO Working Group 15 (the group tasked with discussing emerging concepts).

8 Threshold of toxicological concern initiatives for medical devices

Another path that exists to advance alternatives to animal testing is through the use of analytical chemistry to demonstrate that the extract solutions typically tested in animals contain no chemicals above the threshold of sensitivity for the animal model. The threshold of toxicological concern (TTC) (Hartung, 2017) is one concept that can be used to set such an analytical threshold below which animal testing is futile, as no meaningful amounts of leachable chemicals are present. TTC is a statistical method used to estimate safe exposure levels for chemicals lacking toxicological data. It is based on chemical structure and known safety data for structurally related chemicals (Kroes et al., 2004; Kroes, 2006).

Understanding the potential toxicity of leachable substances is often required in biocompatibility risk assessments. When the structure of a leachable chemical is identified, toxicological conclusions may be drawn if there is sufficient data available for that chemical. TTC is used when there is limited or no data on a given compound but the human exposure is so low that undertaking toxicity studies is not warranted (Renwick, 2004). The concept is already in use to evaluate impurities in pharmaceuticals, and to assess contaminants in consumer products and environmental contaminants (Rovida et al., 2015b). The application of this concept to leachables from medical devices presents several challenges common to many risk-assessment procedures that may be addressed by the inclusion of appropriate uncertainty factors as described in ISO 10993-17 (see Fig. 2).

Another interesting option is the use of read-across, i.e., the use of toxicity information on similar substances to conclude on properties of untested ones (Patlewicz et al., 2014). Current developments toward Good Read-Across Practice guidance (Ball et al., 2016) and automated read-across (Hartung, 2016) are expected to be very helpful in promoting the acceptance of

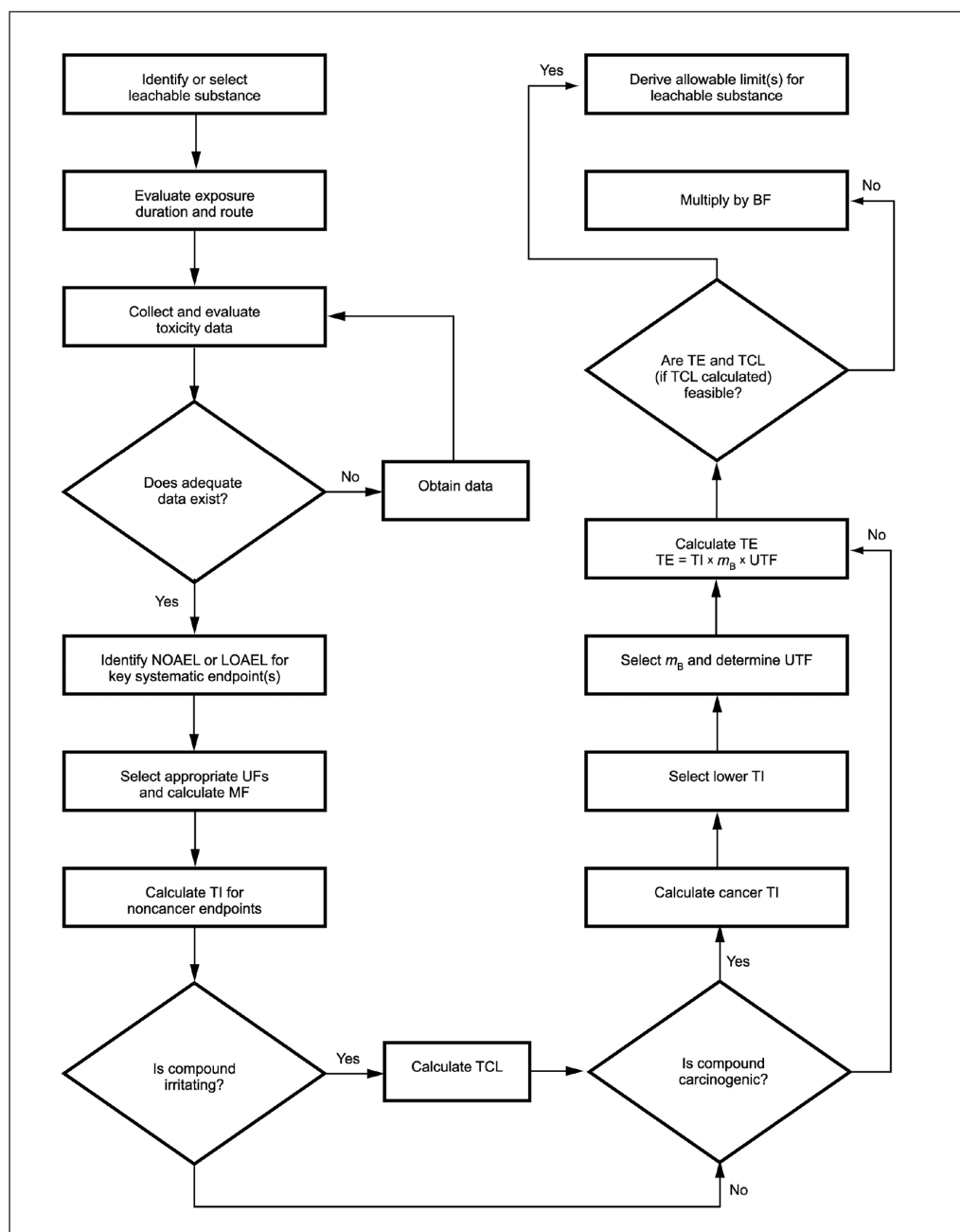


Fig. 2: Flowchart from ISO 10993-17 Biological Evaluation of Medical Devices – Part 17: Methods for the establishment of allowable limits for leachable substances

©ISO. This material is excerpted from ISO 10993-17:2002 with permission of the American National Standards Institute (ANSI) on behalf of the International Organization for Standardization. All rights reserved.

this concept by regulators.

It is worth noting that TTC values have been derived for food and cosmetic products that have no therapeutic benefit to consumers; that should be taken into consideration in determining acceptable risk for medical devices. Second, most TTC values are based on oral toxicity studies and adjustments for other exposure routes are needed. Because many medical devices involve parenteral exposure, inter-route extrapolation must be accounted for (Kroes et al., 2005). For example, oral TTC values have been shown to be appropriate for inhalation and dermal routes (Ball et al., 2007; Blackburn et al., 2005). Notably, the experts responsible for ICH M7 discounted the administration

route consideration and applied the same TTC thresholds to all drugs regardless of their route of administration, arguing that the values are already so conservative that additional safety factors are not appropriate. Third, most TTC values are based on continuous lifetime exposure to chemicals. Because a conservative approach is used when assessing chronic chemical exposure, the relevant TTC value should also be adjusted for exposure duration. When a limited contact medical device or an absorbable biomaterial is being evaluated, the exposure time is finite, so it is necessary to adjust the total dose to a daily dose.

The process for demonstrating the biocompatibility of medical devices often involves a staged approach, starting with a



thorough understanding of the chemical composition of the device, then progressing through *in vitro* and *in vivo* bioassays and ultimately continuing through the lifecycle of the product.

ISO 10993-1 (2009): Evaluation and testing within a risk management process aims to “...serve as a framework in which to plan a biological evaluation which, as scientific knowledge advances our understanding of the basic mechanisms of tissue responses, minimizes the number and exposure of test animals by giving preference to chemical constituent testing and *in vitro* models, in situations where these methods yield equally relevant information to that obtained from *in vivo* models.”

A flow chart entitled “Use of International Standards ISO-10993” (FDA, 2013)⁴ (Fig. 3) presents the same principles a little differently. It asks whether the device contacts the patient directly or indirectly, whether or not the materials are the same as those in marketed devices, and whether or not the manufacturing processes are the same. The key is to demonstrate technical rigor to prove that the chemical compositions are the same.

This approach facilitates the development of sound scientific evidence for the safety of the product and recognizes chemical characterization as perhaps the most important stage of the evaluation. The stepwise approach also contributes to reduction, replacement, and refinement of animal use by building in a decision point before *in vivo* bioassays take place. The TTC approach facilitates making decisions early on in this process. By identifying all the chemical constituents present above a certain level, and then performing a risk assessment on each of those constituents, TTC offers a way to calculate acceptable levels in just about any endpoint – answering a recurring question over how to calibrate the dose-response relationship surrounding efforts to shift medical device testing to *in vitro* methods.

Approaches to demonstrate the safety of a medical device using risk assessment principles and relying heavily on analytical chemistry could help the medical device industry make more use of TTC. The ISO TC 194 draft standard addresses some of these questions. In the hope that more real-world examples could encourage regulatory acceptance of the idea, Richard Hutchinson conducted a hypothetical situation for the symposium to demonstrate how TTC can be applied to real-life situations and to consider practical applications of the TTC approach for compounds released from devices. The scenario placed all 656 leachables relevant to medical plastics (Jenke, 2009) into a spreadsheet. A random generator selected 15, delivering by chance a near-ideal positive and two negative controls – i.e., a known carcinogen and two chemicals from the FDA GRAS (generally recognized as safe) list (Tab. 5). Next, known chemical information about each of these chemicals was selected with assistance from structure-activity relationship (SAR) software packages including DEREK, SARAH, ToxTree, and QSAR Toolbox (shown in Tab. 6).

For performances of the SAR packages in the hypothetical experiment, only DEREK and the ToxTree Cramer model flagged the known human carcinogen. For the one GRAS compound,

the QSAR toolbox predicted it was mutagenic, over-predicting the accepted negative controls in the example. For the second GRAS leachable, all of the SAR packages predicted it was carcinogenic. This showed that all of the packages are tuned to be very sensitive, and to pick out as many risks as possible.

The guidance document states that this methodology clearly needs to be applied by qualified, skilled toxicologists who will carefully determine whether this methodology is appropriate, consider all the relevant endpoints and routes of exposure, and prepare a risk assessment that goes through all data logically and argues why the decision is appropriate. Therefore, the use of Computer SAR packages should only serve as confirmation after evaluation by a competent toxicologist.

Under ISO 10993-17 (2002), the appropriate next steps would include individual risk assessments for each of these chemicals. For example, benzene has a known cancer slope factor, so the next steps would be to characterize the dose the patient is exposed to, calculate the patient risk, and then write up a document considering whether or not the medical benefit of the device outweighs the potential risk (i.e., whether it is an acceptable risk). The final decision could depend on the purpose of the device, e.g., if it is a lifesaving device that would influence whether or not the risk is acceptable.

From a regulatory perspective, and building upon Hutchinson's example, some key points need to be considered in the TTC's practical application to devices: The FDA has used the TTC approach in relation to food additives for at least 20 years. The International Conference on Harmonization (ICH) has published guidance (see Tab. 7) on how to use this approach to qualify genotoxic impurities in drug substances and combinations of drugs with devices (e.g., auto injector pens) (ICH, 2014), and there is truly a paradigm shift underway over the evolving applications of TTC. TTC is heading in the direction of not merely being used to determine whether more testing is needed, but to provide default tolerable intake (TI) values, according to Ronald Brown, whose description of the landscape can be summarized as follows: Rather than automatically testing the biocompatibility of medical devices in animal models, there are some alternatives – including the push to evaluate extracts chemically instead of biologically. There is compelling, albeit limited, data suggesting that for non-cancer endpoints, adverse effects are unlikely to occur when the individual components in the mixture are present at levels well below their respective threshold (Seed et al., 1995). Furthermore, “...exposure to a combination of chemicals compared with exposure to the single chemicals does not constitute an evidently increased hazard provided each individual chemical is administered at a level similar or slightly lower than its own NOAEL” (Feron et al., 1995).

Beyond just using TTC to decide when compounds need additional testing, the medical device community is generally comfortable with the idea of using the TTC as a default exposure limit. It may also be used as a cutoff for the limit of detection (i.e., an analytical evaluation threshold). Given com-

⁴ <http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm348890.pdf>

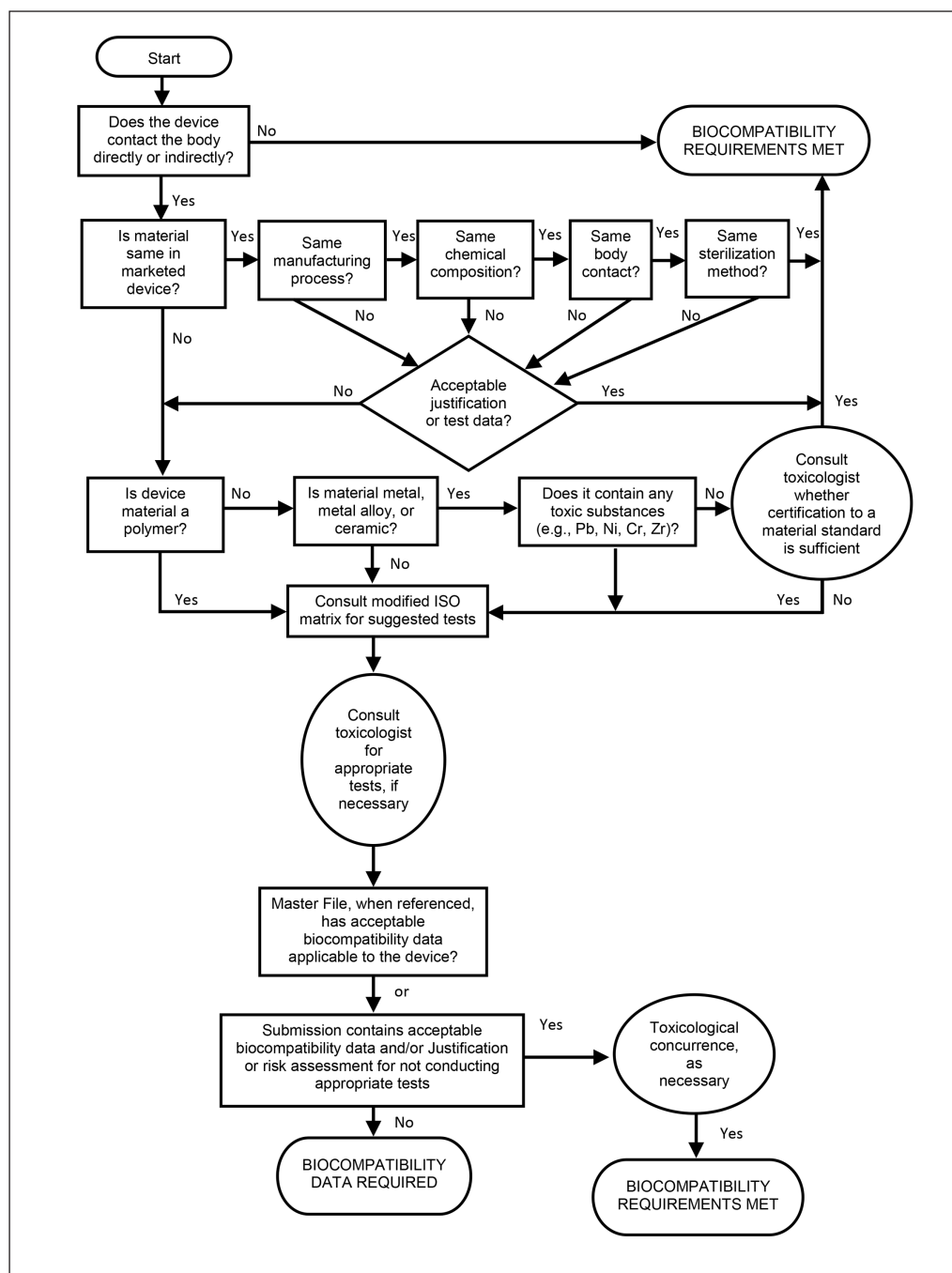


Fig. 3: Flow chart from FDA-CDRH: Use of International Standard ISO-10993, “Biological Evaluation of Medical Devices Part 1: Evaluation and Testing.” Draft Guidance for Industry and FDA Staff

pounds released below the TTC value, there is little point in identifying them any further if they are not expected to be of toxicological concern.

Still, using this approach for compounds that are released from medical devices poses a number of challenges. One challenge relates to the duration of exposure: how to adjust the TTC approach for less than lifetime exposure to compounds, because for many medical devices patients are exposed for a relatively short time. The ICH document only provides short-term exposure limits for carcinogenic compounds; for non-can-

cer effects, compounds can be stratified in terms of potency and sorted into different categories based on their structure. One possible solution is to use the tiered approach for short-term exposure used by the pharmaceutical industry.

The TTC is also a possible screening-level risk assessment approach. To date, default assumptions of dose or response additivity have been used to characterize the toxicity of chemical mixtures. Before a screening-level approach can be used, it is essential to know whether synergistic interactions can occur at low, environmentally relevant exposure levels. After evaluation

**Tab. 5: Fifteen leachables selected by a random generator for a hypothetical situation**

Modified from table originally created by Richard Hutchinson to test the TTC approach for medical devices.

CAS#	Common name	IARC class	GRAS
71-43-2	Benzene	Group 1: Carcinogenic to humans	no
123-95-5	N-butyl stearate	n/a	no
112-80-1	9-Octadecanoic acid	n/a	yes
111-46-6	Diethylene glycol	n/a	no
628-97-7	Ethyl palmitate	n/a	no
106-68-3	3-Octanone	n/a	no
3892-00-0	2,6,10-Trimethylpentadecane	n/a	no
141-79-7	3-Methyl-4-methylene-hexane-2-one	n/a	no
79-09-4	Propionic acid	n/a	yes
117-81-7	Di-(2-ethylhexyl) phthalate	Group 2B: Possibly carcinogenic to humans	no
589-38-8	3-Hexanone	n/a	no
638-67-5	Tricosane	n/a	no
1565-94-2	Bisphenol A glycerolate dimethacrylate	n/a	no
5581-32-8	Bisphenol A bis(2,3-dihydroxypropyl)ether	n/a	no
638-36-8	2,6,10,14-Tetramethylhexadecane	n/a	no

Tab. 6: Fifteen chemicals selected by the random generator and the SAR packages tested in Richard Hutchinson's hypothetical test of the TTC approach for medical devices

The alerts hit with the different packages are highlighted. Modified from table originally created by Richard Hutchinson.

CAS	Tox Tree (Cramer Class)	Tox Tree (skin irritation)	Tox Tree (skin sensitization)	QSAR Toolbox (eye irritation)	QSAR Toolbox (sensitization)	QSAR Toolbox (skin irritation)
71-43-2	III	unknown	no	no	no	no
123-95-5	I	irritating	no	no	no	yes
112-80-1	I	irritating	no	no	no	no
111-46-6	I	irritating	no	no	no	no
628-97-7	I	irritating	no	no	no	yes
106-68-3	II	irritating	no	no	no	yes
3892-00-0	I	unknown	no	no	no	no
141-79-7	I	irritating	yes	no	no	yes
79-09-4	I	irritating	no	no	no	yes
117-81-7	I	no	no	no	yes	no
589-38-8	I	irritating	no	no	no	yes
638-67-5	I	unknown	no	no	no	no
1565-94-2	III	no	yes	no	yes	no
5581-32-8	III	no	no	no	no	no
638-36-8	I	unknown	no	no	no	no

Tab. 7. Acceptable individual and total impurities

Modified from the International Council on Harmonization M7 Draft Consensus Guidance (2013), finalized 2014.

Acceptable intakes for an individual impurity

Duration of treatment	≤ 1 month	> 1-12 months	> 1-10 years	> 10 years to lifetime
Daily intake (µg/day)	120	20	10	1.5

Acceptable intakes for total impurities

Duration of treatment	≤ 1 month	> 1-12 months	> 1-10 years	> 10 years to lifetime
Daily intake (µg/day)	120	60	10 (30*)	5

of the six studies that provided useful quantitative estimates of synergy, the magnitude of synergy at low doses did not exceed the levels predicted by additive models by more than a factor of 4 (Boobis et al., 2011).

Another challenge surrounds the route of exposure, and more research is needed to confirm whether oral values are adequately protective for intravenous or other parenteral routes that might be relevant for medical devices. Conservatism was built into deriving the oral TTC values, leading those on the pharmaceutical side to believe they would be protective for all exposure routes. However, some compounds could be far more potent upon IV exposure than by the oral route.

It is necessary to look at inhalation values separately; one study, by Fraunhofer Institute investigators (Escher et al., 2010) looked at the RepDose database to determine the distribution of inhalation potency values and compare three approaches. There is a need for more examples that study compounds known to emerge from medical devices, as often studies are done with very potent compounds (pesticides, for example) not likely to leach from devices. A strategy needs to be developed on how standardized inhalation tests can be performed with 3D *in vitro* test systems (Alépée et al., 2014; Gordon et al., 2015; Marx et al., 2016) to demonstrate the feasibility of combining TTC with upcoming technologies.

There are some unique considerations to keep in mind when applying the TTC approach to compounds released from device materials: the duration of exposure, the route of exposure, mixtures, metal containing compounds, and applicable endpoints. A lot of progress is being made, and real-world practical examples of how to apply the TTC approach, such as Hutchinson's hypothetical study, will be very helpful in validating the approach, and in addressing concerns such as questions surrounding inhalation values.

This is especially true for local effects, which – depending on the device – might be the most clinically relevant effects. TTC values are not intended to be protective for local effects. The TTC approach cannot serve as a fix-all alternative for all testing. It is a powerful tool, and the numbers are science-based, so

there is a considerable comfort level associated with them, but it cannot cover testing of all different endpoints. It cannot be an alternative to cytotoxicity testing, which is pretty sensitive, for example, and might produce effects at doses lower than the TTC values.

There is a need for a pragmatic approach that suggests the purposes for which the TTC approach is applicable and practical – e.g., subacute, sub-chronic toxicity, and chronic toxicity, and possibly including genotoxic endpoints. This approach has been discussed at various TC 194 meetings, with international buy-in.

Also, in terms of applicability to certain endpoints, trained toxicologists are necessary to perform such assessments. The development of a guidance document would help to define the use and limitations of the approach and further expert discussion as well as to protect against inappropriate use of the approach.

9 Novel developments in toxicology: nanomaterials

It is clear that the landscape of medical device testing is changing. There are also radical changes underway in the composition of medical devices, with important implications for *in vitro* testing. The explosion of products containing nanoparticles in the past decade is one example – nanomaterials promise to play an important role in everything from drug delivery, to implants and artificial tendons, to dental ceramics, to cancer treatments, to tissue regeneration. Their potential is great, but nanomaterials have different physico-chemical properties because of their size – and that is probably true in biological systems as well. Although human skin is a very effective barrier for nanomaterials, they can be taken up through inhalation, ingestion and medical applications such as intravenous or intraperitoneal injection, and they may penetrate into the bloodstream and be distributed throughout the body, finding target organs where they could trigger side effects – including the induction of oxidative stress and inflammatory responses.

Thus, there are unique challenges in working with nanomaterials in *in vitro* tests (Hartung, 2010; Hartung and Sabbioni, 2011) – and the use and testing of medical devices containing nanomaterials must be approached with caution (Silbergeld et al., 2011). A specific example is PUR, a wound-healing, polyurethane-based surface treated foam (Rottmar et al., 2015): Once introduced into a chronic wound, neighboring cells move in to the scaffolding and close the wound (see Fig. 4). Degradation experiments have been carried out according to ISO 10993-13, but still the safety of the foam, which degrades, must be considered as the optimum surface coating is still under investigation.

Instead of developing entirely new assay systems, existing assays can be modified for assessment of nanomaterials, but it is important to recognize and respect the specific properties of these materials. As an example, it was shown that initial assessments of the cytotoxicity of carbon nanotubes were false positive because the nanotubes bind the dye used in the MTT assay, thus confounding results (Wörle-Knirsch et al., 2006). These findings strongly suggest the need to verify cytotoxicity

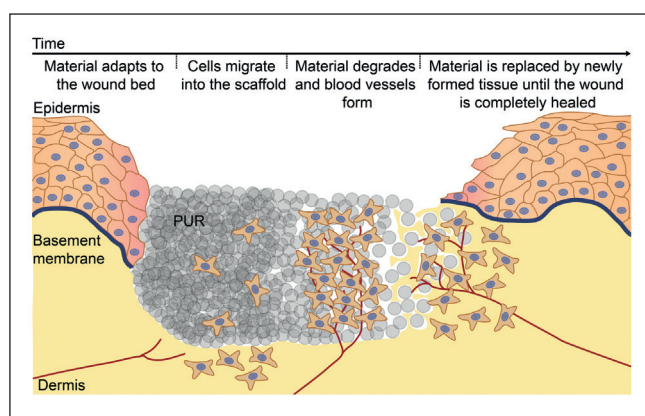


Fig. 4: The *in vivo* tissue engineering based wound treatment concept

The series of events envisioned during wound repair supported by the biodegradable PUR scaffold. Upon setting, the PUR scaffold adapts to the wound bed, allowing cells to easily migrate into the scaffold. The scaffold is degraded and blood vessels form until the material is completely replaced by newly formed tissue, eventually leading to a healed wound. This figure is derived from Rottmar et al. (2015).

1. Seed cells (RAW 264.7) in plate, adhere for 24 h
2. Replace media with media dosed with NP or chemicals
3. Incubate for 24 h
4. Remove media with NP, chemicals and dead cells
5. Add MTS reagent, wait 60 min, read absorption on UV plate reader

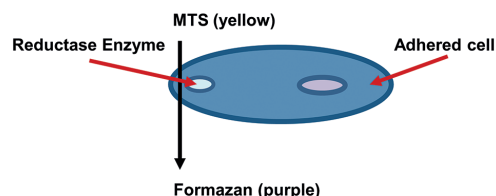


Fig. 5: 5-Step assay for viability testing of cells in culture – so-called MTS-assay

MTS: 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H tetrazolium. Illustration created by Matthias Rösslein (EMPA).

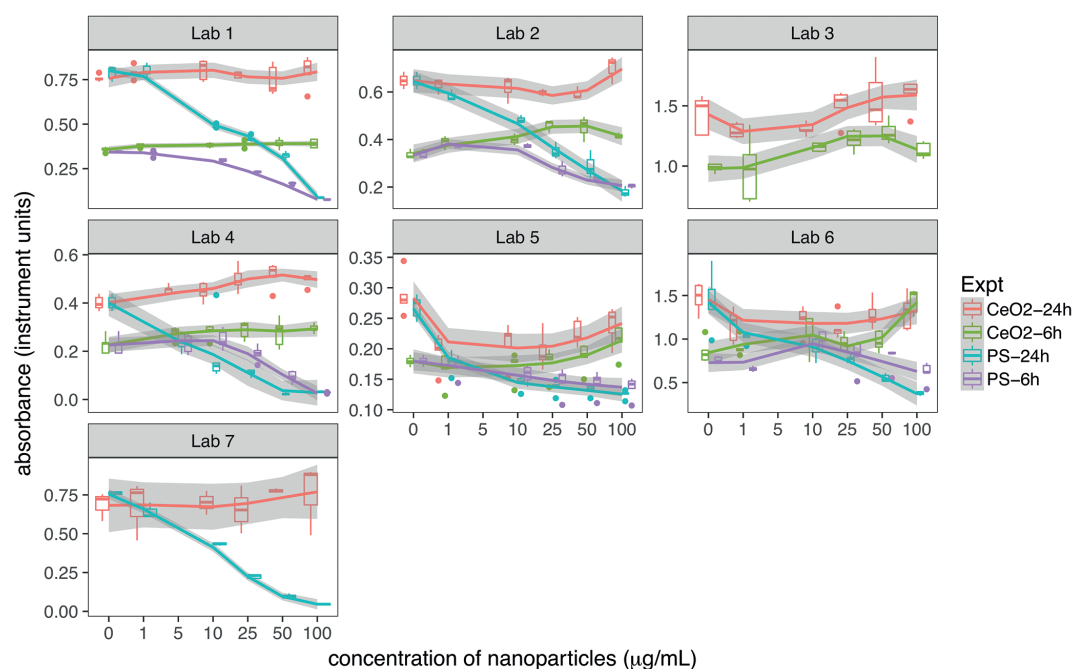


Fig. 6: MTS assay – preliminary stage 3 results from seven labs in a round robin

Focus only on the red curves (non-toxic CeO_2 -nanoparticles, 24 h treatment) and the blue curves (slightly toxic polystyrene nanoparticles with amino-functionalization, 24 h treatment; missing for Lab 03). A 549 cells have been tested by all labs under the same conditions. Illustration created by Matthias Rösslein (EMPA).

data in devices containing nanomaterials with at least a second method – as also suggested in the ISO standards. Furthermore, reference materials and standardized protocols should be used to ensure comparability of results.

Despite a dramatic increase in papers related to nanotoxicology, many of the published studies are not well designed and their results cannot support human toxicity risk assessments (Hirsch et al., 2011; Hristozov et al., 2012). A paper by Hristozov et al. illustrates the problem, showing that of nearly 300 papers delivering toxicological information for titanium dioxide and zinc oxide (the two most prominent nano-sized materials for use in cosmetics and other applications), only around 40 papers provided valid data on the physical and chemical properties of these materials (Hristozov et al., 2012). Comparable results have been found in another comprehensive literature study, which describes many of the pitfalls and flaws of nanotoxicological studies (Krug, 2014), including also sample purification (most nanomaterials are sterile but are contaminated with endotoxins, which cause inflammatory effects) and disregard of interferences of nanomaterials in toxicity tests (owing to absorbance, color, etc.), which must be determined upfront with appropriate positive and negative controls. Depending on the desired endpoint, a wide variety of toxicology tests, each with their own set of endpoints and controls, are available (Locascio et al., 2011).

In response to these challenges, EMPA, the Swiss Federal Institute for Material Sciences and Technologies, set out to establish a platform to assess nanoparticle toxicity *in vitro*. Focused especially on four endpoints (viability, inflammation, genotoxicity, and oxidative stress), this effort aimed to integrate the pathways of toxicology and determine which activities related to nanoparticles merit closer attention. The initiative “VIGO”, funded by industry and the Swiss government, has provided a set of standard operating procedures (SOPs) to harmonize the protocols for the testing of nanomaterials for these four biological endpoints, and it is hoped that this will encourage those who publish in this area to use them⁵.

EMPA and other groups launched a voluntary alliance, the International Alliance for NanoEHS (IANH) for the harmonization of protocols, including 11 laboratories in the US, Europe, and Asia. The first data published focused on characterization methods of nanoparticles.

However, the first effort to develop a protocol for a toxic effect encountered problems. Starting with the very simple, five-step MTS assay (Fig. 5), IANH delivered materials to the participating laboratories. The study used RAW 264.7 cells, murine transformed lymphoma cells. Seven labs joined the first round robin experiment, using the same protocols and the same materials, however, there were severe discrepancies between the results (Fig. 6). In the end, it was discovered that the experiments had been performed by grad students and inexperienced PhDs. Therefore, 5 international institutes (NIST, KRISS, JRC, EMPA, NANOTEC) decided to perform another round, establishing an in-depth cause-and-effect analysis (Rösslein et al.,

2015). The results of the second round were far better, with the EU institutes producing very similar values. The Asian and US participants deviated slightly, probably because they were not permitted to import the same cell culture serum used by the EU institutes (Toman et al., 2016). The assay is suitable for different cell types (Elliott et al., 2016).

9.1 New models for nanoparticles

Besides the carefully established testing protocols, a solid understanding of tissue barrier systems (Gordon et al., 2015), such as the blood-air and blood-brain barrier, is very important for judging the effects of nanomaterials. Models to investigate such tissue barriers are very limited and only the human placenta can be used as a primary tissue model. This system reflects the human tissue barrier that is very important for the development of the fetus. To determine if the placenta is an efficient barrier for nanoparticles, EMPA studied human placentas obtained from clinics and provided the first evidence that nanoparticles can pass through this specific human tissue barrier (Wick et al., 2010) up to a size of 250 nm (Grafmüller et al., 2013, 2015).

The above mentioned studies and experimental designs for the investigation of biological effects of nanoparticles reveal a number of challenges inherent in nanotoxicology, including a need for correct methodologies and study designs, a need to use SOPs or adapted guidelines and harmonize protocols, the use of dose metrics for *in vitro* as well as for *in vivo* studies, the need to measure nanoparticles *in situ*, the need to select suitable biological models, and the need to use suitable controls and reference materials.

9.2 Toxicology in the 21st century for nanomaterials

A number of investigators are publishing work on common pitfalls in nanobiotechnology: Scott McNeil and team focus on specific problems like endotoxin, characterization of manufacturing residue, sterility, batch-to-batch consistency, etc. (Crist et al., 2013); Peter Hoet has concentrated on the biological systems or cell density, assay methods, serum and solvents that have been used (Geys et al., 2010); EMPA has published ideas on how to proceed to establish some reliable test systems, calibrating reference materials, comparison of results achieved with different methods, inter-laboratory comparisons (Hirsch et al., 2011); and Klaus Wittmaack drew attention to overdosing of nanomaterials in *in vitro* testing (Wittmaack, 2011). These efforts all contribute to the need to build a stable body of reliable and comparable results, built on quality-management systems, validation, measurement uncertainty, and traceability.

A binational project called DaNa⁶ funded by the governments of Germany and Switzerland includes information on current projects in Europe and a knowledge base, including ideas about materials, applications, and methodology. The information for this knowledge base was taken from the literature, but the publications were first evaluated by international experts using

⁵ free download at <http://www.nanoobjects.info/en/nanoinfo/methods/992-operating-instructions>

⁶ <http://www.nanoobjects.info>



a literature criteria catalogue to guarantee high quality of the information provided on this website. The criteria catalogues can be downloaded from the website, which also includes SOPs for harmonized protocols.

The testing protocols established by the international consortium described above will be delivered to the OECD and ISO system, and hopefully will eventually lead to harmonized international protocols in collaboration with US and Asian institutes.

In terms of next steps, the ISO TC 194 working group “Biological and clinical evaluation of medical devices”⁷ – has been established to address nanomaterials in medical devices. From a safety point of view, analytics – to demonstrate that nanoparticles, if they are integrated into nano-structured surfaces, are to some extent stable in the biological system – is perhaps the most important area of focus for this working group to consider. Abrasion or other physical/mechanical treatments of these devices during their operation lifetime should not deliver more particles than traditional materials do, but this must be judged on a case-by-case basis, depending on what task nanoparticles fulfill in a medical device. The analytics and characterization are of greater importance than the exposure route or dose.

In the context of evidence-based toxicology, the ToxRTool, a systematic way of assigning Klimisch scores to both *in vivo* and *in vitro* toxicology studies, has been developed (Schneider et al., 2009). It is not a perfect tool, but it is now used by the National Toxicology Program for judging the quality of existing studies as a first step. The FDA is evaluating the ToxRTool, in addition to other evaluation criteria for studies published on device materials, and has so far expressed some confidence in it.

10 Organ-on-a-chip models

Human organ-on-a-chip models offer great promise to change the world of *in vitro* testing. The Wyss Institute for Biologically Inspired Engineering at Harvard University, for example, is developing multiple organ chips, including a breathing lung-on-a-chip model, as well as liver, skin, gut, and kidney models and a beating heart-on-a-chip.

The technology was inspired by the poor predictive power of existing preclinical animal models, which often leads to the failure of drugs late in their development – after they have already entered the human clinical trial phase. Given the tremendous costs of drug development and the long timelines involved, major pharmaceutical companies and government agencies are expressing particular interest in the organ-on-a-chip models (Esch et al., 2015).

Human organ-on-a-chip models utilize microscale engineering technologies combined with 3D cultured living human cells (Alépée et al., 2014; Andersen et al., 2014; Marx et al., 2016) to create microfluidic devices that simulate the physiological and mechanical microenvironment of whole living organs. For example, the breathing lung-on-a-chip recreates the mechanical

aspect that is often missing from *in vitro* assays. Dongeun Huh and Donald Ingber largely developed the initial proof-of-concept of this organ-on-chip technology (Huh et al., 2010). Despite the name, it really is an alveolus-on-a-chip, representing the key functional unit of a lung – where gas exchanges occur, where drug delivery occurs, and where metastatic cancers can occur. The microsystem reproduces breathing movements and the associated cyclic strain and flow experienced by cells at the alveolar-capillary interface. It is also possible to introduce particulate matter into the lung-on-a-chip microdevice to evaluate (nano-) particle transport, induction of an inflammatory response, acute toxic response, and then pharmacologically modulate that response. The human inflammatory response can be recreated and visualized in this small biomimetic microdevice, as well as disease responses such as IL-2-induced pulmonary edema.

Another advantage is that this microengineering approach allows for the development of the simplest models possible that retain physiological relevance (Rossini and Hartung, 2012), with the flexibility to add complexity to the system as necessary – a feature not possible with animal models. These models can enhance our fundamental understanding of complex biological processes and bring about more rapid, accurate, cost-effective, and clinically relevant testing of drugs, as well as cosmetics, chemicals, environmental toxins – and medical devices.

11 Conclusions and recommendations

Amid the opportunities and goals identified by speakers at the symposium were some areas where there is consensus to move ahead. Some action-oriented items can fit into the current ISO structure, while in some cases, new working groups might be able to help build momentum.

11.1 Refinement opportunities

In the course of sharing details about *in vitro* tests, the symposium participants identified many opportunities to refine *in vivo* tests where the *in vitro* alternatives are not yet sufficient to replace testing on animals. They noted that the current *in vivo* options for medical devices are not always up-to-date compared to drug and chemical industry testing. For example, there might be an opportunity to reduce the number of animals used in the LLNA for medical devices from five to the four required in tests for chemicals.

Rabbits are still used in the vast majority of hemolysis testing that takes place today. If it can be proven that human blood can generate the same results, this could offer another opportunity to reduce the use of animals in future testing. A round robin study exploring this question was taken up by members of WG 9 of the ISO 10993-4 (2009). This study has been concluded, and publication of the results is expected soon.

There is a need to collaboratively agree upon and organize refinement techniques – perhaps by placing them on the agenda

⁷ <https://www.iso.org/committee/54508.html>

for the next ISO meeting. There also might be some merit in creating a refinement working group for medical devices, building on the success of recent JHU-CAAT sponsored workshops (Zurlo and Hutchinson, 2013). For example, in Europe, Novartis and Roche hosted a workshop on refinement opportunities in the pharmaceutical industry (report in preparation).

There is a disconnect between the language in the ISO 10993-1 standard, which recognizes a potential tiered approach, giving more weight to the *in vitro* data, and the reality that this is not being translated into regulatory decision-making. For those who see merit in the *in vitro* methods, discussing a tiered approach could be a starting point. However, this is a big task for all endpoints, especially considering the special challenges of genotoxicity and hemocompatibility. Limiting the discussion to ways of using *in vitro* cytotoxicity assays to help decide whether to do *in vivo* acute systemic toxicity testing, irritation testing, or implantation testing is one possible angle.

11.2 Sharing data

One area where industry, academia, and regulators can cooperate is in filling in holes in the data. As most of the data is from animal studies, there is a need to generate, and share, more data on humans. Answers can be found by using existing *in vivo* assays to build algorithms that can help translate that information to humans. What is missing is the quantitative information – the NOAELs and LOAELs – that are needed for risk assessment, which will likely be developed in the next few years.

In addition to gathering data, curating data according to well-accepted benchmarks (see Section 9.2) before introducing it into a database is important as well.

Industry representatives need to find ways to share data with regulators – not in a product-specific way, but in terms of interpreting what the standard tells us, and how it might be used. Regulators, including the FDA, are interested in seeing the *in vitro* data. For example, an FDA group looking at systemic toxicity has requested more data. A mechanism to share data with the regulators could take advantage of information-sharing structures already developed by legislations like REACH, the European chemicals legislation – efforts that protect intellectual property, but still encourage the sharing of anonymized data.

11.3 Harmonizing standards and approaches

The need to harmonize protocols and efforts – between various branches of industry, between branches of science, between industry and the regulators, and internationally – emerged as a clear theme from the symposium. A new steering group could help set the agenda and push for results. This steering group could also help facilitate sharing of assessment data in the pre-competitive phase, and create an opportunity for regulators to weigh in on whether or not new methods are likely to be accepted or not.

The effort should be in concert with ISO efforts, to ensure consensus is international and not just a Euro-American approach. However, there is also a need for an independent broker who can identify and bring together industry and regulatory representatives. The participants agreed that the ISO process

remains crucial to document the developments underway. But hearing the various perspectives shared at this symposium, it is also clear that there is a pressing need to accelerate the approval process to better reflect the speed of technology.

References

- Alépée, N., Bahinski, A., Daneshian, M. et al. (2014). t⁴ workshop report: State-of-the-art of 3D cultures (organs-on-a-chip) in safety testing and pathophysiology. *ALTEX* 31, 441-477. doi:10.14573/altex1406111
- Andersen, M., Betts, K., Dragan, Y. et al. (2014). Developing microphysiological systems for use as regulatory tools – challenges and opportunities. *ALTEX* 31, 364-367. doi:10.14573/altex.1405151
- Atala, A., Bauer, S. B., Soker, S. et al. (2006). Tissue-engineered autologous bladders for patients needing cystoplasty. *Lancet* 367, 1241-1246. doi:10.1016/S0140-6736(06)68438-9
- Ball, D., Blanchard, J., Jacobson-Kram, D. et al. (2007). Development of safety qualification thresholds and their use in orally inhaled and nasal drug product evaluation. *Toxicol Sci* 97, 226-236. doi:10.1093/toxsci/kfm058
- Ball, N., Cronin, M. T. D., Shen, J. et al. (2016). Toward good read-across practice (GRAP) guidance. *ALTEX* 33, 149-166. doi:10.14573/altex.1601251
- Blackburn, K., Stickney, J. A., Carlson-Lynch, H. L. et al. (2005). Application of the threshold of toxicological concern approach to ingredients in personal and household care products. *Regul Toxicol Pharmacol* 43, 249-259. doi:10.1016/j.yrtph.2005.08.007
- Boobis, A., Budinsky, R., Collie, S. et al. (2011). Critical analysis of literature on low-dose synergy for use in screening chemical mixtures for risk assessment *Crit Rev Toxicol* 41, 369-383. doi:10.3109/10408444.2010.543655
- Casas, J. W., Lewerenz, G. M., Rankin, E. A. et al. (2013). In vitro human skin irritation test for evaluation of medical device extracts. *Toxicol In Vitro* 27, 2175-2183. doi:10.1016/j.tiv.2013.08.006
- Coleman, K., Dai, X., Deng, X. et al. (2012). Medical device biocompatibility evaluation: An industry perspective. In J. Boutrand (ed.), *Biocompatibility and Performance of Medical Devices*. Cambridge, UK: Woodhead Publishing Ltd. doi:10.1533/9780857096456.2.201
- Crist, R. M., Grossman, J. H., Patri, A. K. et al. (2013). Common pitfalls in nanotechnology: Lessons learned from NCI's Nanotechnology Characterization Laboratory. *Integrat Biol* 5, 66-73. doi:10.1039/C2IB20117H
- Elliott, J. T., Rösslein, M., Song, N. W. et al. (2016). Toward achieving harmonization in a nano-cytotoxicity assay measurement through an interlaboratory comparison study. *ALTEX* 34, 201-218. doi:10.14573/altex.1605021
- Esch, E. W., Bahinski, A. and Huh, D. (2015). Organs-on-chips at the frontiers of drug discovery. *Nat Rev Drug Discov* 14, 248-260. doi:10.1038/nrd4539
- Escher, S. E., Tluczkiewicz, I., Batke, M. et al. (2010). Evaluation of inhalation TTC values with the database Rep-



- Dose. *Regul Toxicol Pharmacol* 58, 259-274. doi:10.1016/j.yrtph.2010.06.009
- FDA (2013). Use of International Standards ISO-10993: Biological evaluation of 3 medical devices Part 1: Evaluation and testing. Center for Devices and Radiological Health (CDRH). <http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm348890.pdf>
- Fennrich, S., Wendel, A. and Hartung, T. (1999). New applications of the human whole blood pyrogen assay (PyroCheck). *ALTEX* 16, 146-149. <http://www.altex.ch/All-issues/Issue.50.html?iid=53&aid=8>
- Feron, V. J., Groten, J. P., Jonker, D. et al. (1995). Toxicology of chemical mixtures: Challenges for today and the future. *Toxicology* 105, 415-427. doi:10.1016/j.yrtph.2010.06.009
- Gerberick, G. F., Vassallo, J. D., Bailey, R. E. et al. (2004). Development of a peptide reactivity assay for screening contact allergens. *Toxicol Sci* 81, 332-343. doi:10.1093/toxsci/kfh213
- Geys, J., Nemery, B. and Hoet, P. (2010). Assay conditions can influence the outcome of cytotoxicity tests of nanomaterials: Better assay characterization is needed to compare studies. *Toxicol In Vitro* 24, 620-629. doi:10.1016/j.tiv.2009.10.007
- Gordon, S., Daneshian, M., Bouwstra, J. et al. (2015). Non-animal models of epithelial barriers (skin, intestine and lung) in research, industrial applications and regulatory toxicology. *ALTEX* 32, 327-378. doi:10.14573/altex.1510051
- Grafmüller, S., Manser, P., Krug, H. F. et al. (2013). Determination of the transport rate of xenobiotics and nanomaterials across the placenta using the ex vivo human placental perfusion model. *J Vis Exp* 76, e50401. doi:10.3791/50401
- Grafmüller, S., Manser, P., Diener, L. et al. (2015). Transfer studies of polystyrene nanoparticles in the ex vivo human placenta perfusion model: Key sources of artifacts. *Sci Technol Adv Mater* 16, 044602. doi:10.1088/1468-6996/16/4/044602
- Hamm, J., Sullivan, K., Clippinger, A. J. et al. (2017). Alternative approaches for identifying acute systemic toxicity: Moving from research to regulatory testing. *Toxicol In Vitro* 41, 245-259. doi:10.1016/j.tiv.2017.01.004
- Hartung, T. and Wendel, A. (1995). Die Erfassung von Pyrogenen in einem humanen Vollblutmodell (Detection of Pyrogens using human whole blood). *ALTEX* 12, 70-75. <http://www.altex.ch/All-issues/Issue.50.html?iid=34&aid=4>
- Hartung, T. and Wendel, A. (1996). Detection of pyrogens using human whole blood. *In Vitro Toxicol* 9, 353-359.
- Hartung, T. (2009). Toxicology for the twenty-first century. *Nature* 460, 208-212. doi:10.1038/460208a
- Hartung, T. (2010). Food for thought ... on alternative methods for nanoparticle safety testing. *ALTEX* 27, 87-95. doi:10.14573/altex.2010.2.87
- Hartung, T. and Sabbioni, E. (2011). Alternative in vitro assays in nanomaterial toxicology. *WIREs Nanomed Nanobiotechnol* 3, 545-573. doi:10.1002/wnan.153
- Hartung, T., Luechtefeld, T., Maertens, A. and Kleensang, A. (2013). Integrated testing strategies for safety assessments. *ALTEX* 30, 3-18. doi:10.14573/altex.2013.1.003
- Hartung, T. (2015). The human whole blood pyrogen test – Lessons learned in twenty years. *ALTEX* 32, 79-100. doi:10.14573/altex.1503241
- Hartung, T. (2016). Making big sense from big data in toxicology by read-across. *ALTEX* 33, 83-93. doi:10.14573/altex.1603091
- Hartung, T. (2017). Thresholds of toxicological concern – Setting a threshold for testing where there is little concern. *ALTEX* 34, 331-351. doi:10.14573/altex.1707011
- Hasiwa, M., Kullmann, K., von Aulock, S. et al. (2007). An in vitro pyrogen safety test for immune-stimulating components on surfaces. *Biomaterials* 28, 1367-1375. doi:10.1016/j.biomaterials.2006.11.016
- Hasiwa, N., Daneshian, M., Bruegger, P. et al. (2013). Evidence for the detection of non-endotoxin pyrogens by the whole blood monocyte activation test. *ALTEX* 30, 169-208. doi:10.14573/altex.2013.2.169
- Hayden, P. (2007). The way forward for in vitro skin irritation testing. <http://alttox.org/the-way-forward-for-in-vitro-skin-irritation-testing/> (accessed 07.08.2016).
- Hirsch, C., Roesslein, M., Krug, H. F. et al. (2011). Nanomaterial cell interactions: Are current in vitro tests reliable? *Nano-medicine (Lond)* 6, 837-847. doi:10.2217/nmm.11.88
- Hoffmann, S., Peterbauer, A., Schindler, S. et al. (2005). International validation of novel pyrogen tests based on the human fever reaction. *J Immunol Meth* 298, 161-173. doi:10.1016/j.jim.2005.01.010
- Hoffmann, S. (2015). LLNA variability: An essential ingredient for a comprehensive assessment of non-animal skin sensitization test methods and strategies. *ALTEX* 32, 379-383. doi:10.14573/altex.1505051
- Hristozov, D. R., Gottardo, S., Critto, A. et al. (2012). Risk assessment of engineered nanomaterials: A review of available data and approaches from a regulatory perspective. *Nanotoxicology* 6, 880-898. doi:10.3109/17435390.2011.626534
- Huh, D., Matthews, B. D., Mammoto, A. et al. (2010). Reconstituting organ-level lung functions on a chip. *Science* 25, 1662-1668. doi:10.1126/science.1188302
- ICH – International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (2014). M7 Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk, Geneva, Switzerland. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multi-disciplinary/M7/M7_Step_4.pdf
- ISO 10993-17 (2002). Biological evaluation of medical devices – Part 17: Establishment of allowable limits for leachable substances. International Organization for Standardization, Geneva, Switzerland. <https://www.iso.org/obp/ui/#iso:std:iso:10993:-17:ed-1:v1:en>
- ISO 10993-1 (2009). Biological evaluation of medical devices – Part 1: Evaluation and testing within a risk management process. International Organization for Standardization, Geneva, Switzerland. <https://www.iso.org/obp/ui/#iso:std:iso:10993:-1:ed-4:v1:en>

- ISO 10993-4 (2009). Biological evaluation of medical devices – Part 4: Selection of tests for interaction with blood. International Organization for Standardization, Geneva, Switzerland. <https://www.iso.org/obp/ui/#iso:std:iso:10993:-4:dis:ed-3:v1:en>
- ISO 10993-5 (2009). Biological evaluation of medical devices – Part 5: Tests for in vitro cytotoxicity. International Organization for Standardization, Geneva, Switzerland. <https://www.iso.org/obp/ui/#iso:std:iso:10993:-5:ed-3:v1:en>
- ISO 10993-10 (2010). Biological evaluation of medical devices – Part 10: Tests for irritation and skin sensitization. International Organization for Standardization, Geneva, Switzerland. <https://www.iso.org/obp/ui/#iso:std:iso:10993:-10:ed-3:v1:en>
- ISO 10993-12 (2012). Biological evaluation of medical devices – Part 12: Sample preparation and reference materials. International Organization for Standardization, Geneva, Switzerland. <https://www.iso.org/standard/53468.html>
- ISO 10993-3 (2014). Biological evaluation of medical devices – Part 3: Tests for genotoxicity, carcinogenicity and reproductive toxicity. International Organization for Standardization, Geneva, Switzerland. <https://www.iso.org/obp/ui/#iso:std:iso:10993:-3:ed-3:v1:en>
- ISO 13485 (2016). Quality management for medical devices. International Organization for Standardization, Geneva, Switzerland. <https://www.iso.org/obp/ui/#iso:pub:PUB100377>
- ISO 14155 (2011). Clinical investigation of medical devices for human subjects – Good clinical practice. International Organization for Standardization, Geneva, Switzerland. <https://www.iso.org/obp/ui/#iso:std:iso:14155:ed-2:v1:en>
- ISO 14971 (2007). Application of risk management to medical devices. International Organization for Standardization, Geneva, Switzerland. <https://www.iso.org/obp/ui/#iso:std:iso:14971:ed-2:v2:en>
- Jenke, D. R. (2009). *Compatibility of Pharmaceutical Solutions and Contact Materials: Safety Assessments of Extractables and Leachables for Pharmaceutical Products*. New Jersey, USA: John Wiley & Sons.
- Kramer, D., Xu, S. and Kessleheim, A. (2012). Regulation of medical devices in the United States and European Union. *New Engl J Med* 366, 848-855. doi:10.1056/NEJMhle1113918
- Kroes, R., Renwick, A. G., Cheeseman, M. et al. (2004). Structure-based thresholds of toxicological concern (TTC): Guidance for application to substances present at low levels in the diet. *Food Chem Toxicol* 42, 65-83. doi:10.1016/j.fct.2003.08.006
- Kroes, R., Kleiner, J. and Renwick, A. (2005). The threshold of toxicological concern concept in risk assessment. *Toxicol Sci* 86, 226-230. doi:10.1093/toxsci/kfi169
- Kroes, R. (2006). The threshold of toxicological concern concept in risk assessment. *Toxicol Lett* 164, S48. doi:10.1016/j.toxlet.2006.06.102
- Krug, H. F. (2014). Nanosafety research – are we on the right track? *Angew Chem Int Ed* 53, 12304-12319. doi:10.1002/anie.201403367
- Locascio, L. E., Reipa, V., Zook, J. M. et al. (2011). Nanomaterial toxicity: Emerging standards and efforts to support standards development. In V. Murashov and J. Howard (eds.), *Nanotechnology Standards: Nanostructure Science and Technology* (179-208). New York, NY, USA: Springer-Verlag. doi:10.1007/978-1-4419-7853-0_8
- Marx, U., Andersson, T. B., Bahinski, A. et al. (2016). Biology-inspired microphysiological system approaches to solve the prediction dilemma of substance testing using animals. *ALTEX* 33, 272-321. doi:10.14573/altex.1603161
- Marzulli, F. N. and Maibach, H. I. (1975). The rabbit as a model for evaluating skin irritants: A comparison of results obtained on animals and man using repeat skin exposures. *Food Cosmet Toxicol* 13, 533-540. doi:10.1016/S0015-6264(78)80051-0
- Mazzotti, F., Beuttler, J., Zeller, R. et al. (2007). In vitro pyrogen test – a new test method for solid medical devices. *J Biomed Mat Res* 80, 276-282. doi:10.1002/jbm.a.30922
- McKim, J. M. Jr., Keller, D. J. III and Gorski, J. R. (2012). An in vitro method for detecting chemical sensitization using human reconstructed skin models and its applicability to cosmetic, pharmaceutical, and medical device safety testing. *Cutan Ocul Toxicol* 31, 292-305. doi:10.3109/15569527.2012.667031
- Natsch, A. and Emter, R. (2008). Skin sensitizers induce antioxidant response element dependent genes: Application to the in vitro testing of the sensitization potential of chemicals. *Toxicol Sci* 102, 110-119. doi:10.1093/toxsci/kfm259
- Natsch, A. (2010). The Nrf2-Keap1-ARE toxicity pathway as a cellular sensor for skin sensitizers – functional relevance and a hypothesis on innate reactions to skin sensitizers. *Toxicol Sci* 113, 284-292. doi:10.1093/toxsci/kfp228
- Natsch, A., Bauch, C., Foertsch, L. et al. (2011). The intra- and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict skin sensitizers in vitro: Results of a ring-study in five laboratories. *Toxicol In Vitro* 25, 733-744. doi:10.1016/j.tiv.2010.12.014
- Patlewicz, G., Ball, N., Becker, R. A. et al. (2014). Read-across approaches – misconceptions, promises and challenges ahead. *ALTEX* 31, 387-396. doi:10.14573/altex.1410071
- Renwick, A. G. (2004). Toxicology databases and the concept of thresholds of toxicological concern as used by the JECFA for the safety evaluation of flavouring agents. *Toxicol Lett* 149, 224. doi:10.1016/j.toxlet.2003.12.034
- Rösslein, M., Elliott, J. T., Salit, M. et al. (2015). Use of cause-and-effect analysis to design a high-quality nanocytotoxicology assay. *Chem Res Toxicol* 28, 21-30. <http://doi.org/10.1021/tx500327y>
- Rossini, G. P. and Hartung, T. (2012). Towards tailored assays for cell-based approaches to toxicity testing. *ALTEX* 29, 359-372. doi:10.14573/altex.2012.4.359
- Rottmar, M., Richter, M., Mader, X. et al. (2015). In vitro investigations of a novel wound dressing concept based on biodegradable polyurethane. *Sci Technol Adv Mater* 16, 034606. doi:10.1088/1468-6996/16/3/034606
- Rovida, C., Alépée, N., Api, A. M. et al. (2015a). Integrated testing strategies (ITS) for safety assessment. *ALTEX* 32, 25-40. doi:10.14573/altex.1411011



- Rovida, C., Asakura, S., Daneshian, M. et al. (2015b). Toxicity testing in the 21st century beyond environmental chemicals. *ALTEX* 32, 171-81. doi:10.14573/altex.1506201
- Sakaguchi, H., Ashikaga, T., Miyazawa, M. et al. (2009). The relationship between CD86/CD54 expression and THP-1 cell viability in an in vitro skin sensitization test – human cell line activation test (h-CLAT). *Cell Biol Toxicol* 25, 109-126. doi:10.1007/s10565-008-9059-9
- Schindler, S., Asmus, S., von Aulock, S. et al. (2004). Cryopreservation of human whole blood for pyrogenicity testing. *J Immunol Meth* 294, 89-100. doi:10.1016/j.jim.2004.08.019
- Schindler, S., Spreitzer, I., Loschner, B. et al. (2006). International validation of pyrogen tests based on cryopreserved human primary blood cells. *J Immunol Meth* 316, 42-51. doi:10.1016/j.jim.2006.07.023
- Schindler, S., von Aulock, S., Daneshian, M. et al. (2009). Development, validation and applications of the monocyte activation test for pyrogens based on human whole blood. *ALTEX* 26, 293-305. doi:10.14573/altex.2009.4.265
- Schneider, K., Schwarz, M., Burkholder, I. et al. (2009). “ToxR-Tool,” a new tool to assess the reliability of toxicological data. *Toxicol Lett* 189, 138-144. doi:10.1016/j.toxlet.2009.05.013
- Scott, R. C., Corrigan, M. A., Smith, F. and Mason, H. (1991). The influence of skin structure on permeability: An intersite and interspecies comparison with hydrophilic penetrants. *J Invest Dermatol* 96, 921-925. doi:10.1111/1523-1747
- Seed, J., Brown, R. P., Olin, S. S. et al. (1995). Chemical mixtures: Current risk assessment methodologies and future directions. *Regul Toxicol Pharmacol* 22, 76-94. doi:10.1006/rtp.1995.1071
- Silbergeld, E. K., Contreras, E. Q., Hartung, T. et al. (2011). Nanotoxicology: “The end of the beginning” – signs on the roadmap to a strategy for assuring the safe application and use of nanomaterials. *ALTEX* 28, 236-241. doi:10.14573/altex.2011.3.236
- Spielmann, H., Hoffmann, S., Liebsch, M. et al. (2007). The ECVAM international validation study on in vitro tests for acute skin irritation: Report on the validity of the EPISKIN and EpiDerm assays and on the Skin Integrity Function Test. *Altern Lab Anim* 35, 559-601.
- Toman, B., Rösslein, M., Elliott, J. T. et al. (2016). Estimation and uncertainty analysis of dose response in an inter-laboratory experiment. *Metrologia* 53, 40-45. doi:10.1088/0026-1394/53/1/S40
- van Oeveren, W. (2013). Obstacles in haemocompatibility testing. *Scientifica* 2, 1-14. doi:10.1155/2013/392584
- Wick, P., Malek, A., Manser, P. et al. (2010). Barrier capacity of human placenta for nanosized materials. *Environ Health Perspect* 118, 432-436. doi:10.1289/ehp.0901200
- Wittmaack, K. (2011). Excessive delivery of nanostructured matter to submersed cells caused by rapid gravitational settling. *ACS Nano* 5, 3766-3778. doi:10.1021/nn200112u
- Wolf, M. F. and Anderson, J. M. (2012). Practical approach to blood compatibility assessments: General considerations and standards. In J. Boutrand (ed.), *Biocompatibility and Performance of Medical Devices*. Cambridge, UK: Woodhead Publishing Ltd. doi:10.1533/9780857096456.2.159
- Wolf, M. F., Coleman, K. and Lewerenz, G. (2013). In vitro assessment of cell and tissue compatibility. In B. D. Ratner, A. S. Hoffman, F. J. Schoen and J. E. Lemons (eds.), *Biomaterials Science: An Introduction to Materials in Medicine*. 3rd edition. New York, USA: Elsevier Inc. doi:10.1016/b978-0-08-087780-8.00053-x
- Wörle-Knirsch, J. M., Pulschke, K. and Krug, H. F. (2006). Oops they did it again! Carbon nanotubes hoax scientists in viability assays. *Nano Lett* 6, 1261-1268. doi:10.1021/nl060177c
- Zurlo, J. and Hutchinson, E. (2013). Refinement. *ALTEX* 31, 4-10. doi:10.14573/altex.1312191

Conflict of interest

Several authors work for companies producing medical devices. Thomas Hartung holds patents as inventor of the whole blood pyrogen test and the use of cryopreserved blood, which are licensed to Merck-Millipore; he receives royalties from Merck-Millipore from sales of the kit version; he also consults Underwriters Laboratories (UL) on computational toxicology and receives shares of their revenues.

Correspondence to

Thomas Hartung, MD PhD
Center for Alternatives to Animal Testing
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe Str.
Baltimore, MD, 21205, USA
e-mail: thartun1@jhu.edu