



*“The less the people know about how sausages
and laws are made, the better they’ll sleep at night”*

often attributed to Otto von Bismarck

“An ounce of prevention is worth a pound of cure”

Benjamin Franklin

Food for Thought ...

Rebooting the Generally Recognized as Safe (GRAS) Approach for Food Additive Safety in the US

Thomas Hartung

Johns Hopkins Center for Alternatives to Animal Testing (CAAT), Baltimore, MD, USA and CAAT-Europe, University of Konstanz, Konstanz, Germany

Summary

The US Food and Drug Administration (FDA) has premarket review authority over food additives, but a food manufacturer may, according to the legislation, intentionally add a substance to human food or animal food without their premarket review or approval if the substance is generally recognized, among qualified experts, to be safe under the conditions of its intended use. Generally recognized as safe (GRAS) implies that the current scientific community agrees on the adequacy of how data is generated. This system has come under public pressure because of doubts as to its efficiency and the FDA’s recent GRAS rule is part of the response. The FDA guidance for testing food additives, known as the “Redbook”, is about two decades old. Work toward a new “Redbook” is on the way, but the US Grocery Manufacturer Association (GMA) also has initiated the development of an independent standard on how to perform GRAS determinations.

This review of the current guidance shows a very rigorous system for higher concern levels, but also many waiving options. Opportunities and challenges for safety evaluations of food additives are discussed. Where scientific progress has allowed improving existing and adapting new methods, these should be adopted to improve product safety and animal welfare. The continuous adaptation of such improved methods is therefore needed. Especially, there are opportunities to embrace developments within the toxicity testing for the 21st century movement and evidence-based toxicology approaches. Also, the growing understanding of the limitations of traditional tests needs to be considered.

Keywords: food additives, safety testing, US legislation, animal testing, alternative methods

1 Introduction

Food for thought... and *vice versa* this time. An earlier article in this series (Hartung and Koëter, 2008), also on food safety, was written mainly from a European perspective – unavoidably as both authors were part of the European Commission: Herman Koëter, as acting head of the European Food Safety Authority (EFSA), and I, then shortlisted to become possibly Director for Food Safety in what was the EC Directorate General Health and Consumers (DG SANCO, now DG SANTE). A decade later,

eight of these years spent in the US, this article addresses a key pillar of the US system, the generally recognized as safe (GRAS) provisions. This is how the safety of most food additives is assessed in the US and also one of the more contended aspects of US food law. Importantly, in US law, food additives include direct (intentionally added) and indirect (contaminants) additives.

I prepped this article with the famous epigram of Bismarck, a German aristocrat and statesman (the first Chancellor of Germany, 1871-1890). It makes an interesting connection between

Received December 18, 2017;
doi:10.14573/altex.1712181



This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.



law and food production. However, it should be noted that the quote is probably misattributed and John Godfrey Saxe¹ (University Chronicle, University of Michigan, 1869) or an unnamed member of the Illinois state legislature² should be credited. This brings it back into the US context and shows that such concerns engage both sides of the Atlantic and that these are not recent concerns at all.

Michael Palin (Monty Python's Flying Circus) said "*All I ask of food is that it doesn't harm me*". This is what it is all about. Nobody wants to poison customers. However, an enormous number of chemicals are added either directly or indirectly (as contaminants such as pesticides or contact materials) to US food. And many of these have no (public) data as to their safety; Neltner et al. (2013a) report that "*Less than 38% of FDA-regulated additives have a published feeding study. For chemicals directly added to food, 21.6% have feeding studies necessary to estimate a safe level of exposure and 6.7% have reproductive or developmental toxicity data in FDA's database*". So, this is once again about data-gaps and uncertainty in the existing information.

To market a new food or color additive, a manufacturer must first petition the FDA for its approval. Approximately 100 new petitions are submitted to the FDA annually, most of which are for indirect additives such as packaging materials. GRAS determination, in contrast, requires only a (voluntary) notification. Karmaus et al. (2016) found a total of 8,659 food-relevant chemicals including direct food additives, food contact substances, and pesticides, which is slightly below previous estimates of the food-use chemical universe; furthermore, only 3,888 were possible direct additives while 4,771 were food contact substances or pesticides. This shows the enormous number of substances to be handled, especially when addressing data-gaps from the past.

So, what is GRAS? As summarized by Neltner et al. (2013a): "*By law, food additives cannot be used in food without an affirmative determination that their use is safe (21 U.S.C. §321 and §348) by FDA or, in some cases, the additive manufacturer. Safety is defined as 'reasonable certainty in the minds of competent scientists that the substance is not harmful under the intended conditions of use' (21 CFR §170.30(i)). Also, 'no additive shall be deemed to be safe if it is found to induce cancer when ingested by man or animal' (21 U.S.C. §348(c)(3)(A)). Congress required affirmative determinations and set stringent standards for safety, acknowledging that the health effects of chemical additives are often hidden or take years to show up. These standards aim to protect the public as well as to encourage innovation and build public confidence in the safety of the food supply (Congress, 1958). In addition, a safety decision for a GRAS substance based on scientific procedures must be supported by published studies (whether peer-reviewed or not) (21 CFR §170.30(b)) and there can be no genuine dispute regarding the chemical's safety (Degnan, 2006). If an additive was in common use before 1958, safety may be based on that experi-*

ence rather than scientific procedures landmark 'Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food', also known as the 'Redbook' (FDA, 1982), and the development of the Priority-based Assessment of Food Additives (PAFA) system (Smith and Rulis, 1981) that included a database with toxicological information on direct additives (those added directly to food)."

Burdock and Carabin (2004) explained: "*Generally recognized as safe (GRAS) is the self-determination of safety and regulatory compliance in an otherwise stringently regulated venue. GRAS is therefore unique in technologically advanced societies and is characteristic of American reliance on self-governance*". More on this in the next section on the history of GRAS.

The GRAS process has come under recent criticism by the US Government Accountability Office (GAO)³:

"The Food and Drug Administration (FDA), which is responsible for ensuring the safety of most of the U.S. food supply, is not required to review substances, such as spices and preservatives, added to food that are generally recognized as safe (GRAS) for their intended use. Currently, companies may determine a substance is GRAS without FDA's approval or knowledge. However, a few substances previously considered GRAS have later been banned; and concerns have been raised about the safety of other GRAS substances, including those containing engineered nanomaterials, materials manufactured at a tiny scale to take advantage of novel properties.

FDA's oversight process does not help ensure the safety of all new GRAS determinations. FDA only reviews those GRAS determinations that companies submit to the agency's voluntary notification program—the agency generally does not have information about other GRAS determinations companies have made because companies are not required to inform FDA of them. Furthermore, FDA has not taken certain steps that could help ensure the safety of GRAS determinations, particularly those about which the agency has not been notified. FDA has not issued guidance to companies on how to document their GRAS determinations or monitored companies to ensure that they have conducted GRAS determinations appropriately. Lastly, FDA has yet to issue a final regulation for its 1997 proposed rule that sets forth the framework and criteria for the voluntary notification program, potentially detracting from the program's credibility. FDA is not systematically ensuring the continued safety of current GRAS substances. While, according to FDA regulations, the GRAS status of a substance must be reconsidered as new scientific information emerges, the agency has not systematically reconsidered GRAS substances since the 1980s. FDA officials said they keep up with new developments in the scientific literature and, on a case-by-case basis, information brought to the agency's

¹ https://en.wikiquote.org/wiki/Otto_von_Bismarck

² https://en.wikiquote.org/wiki/Talk:Otto_von_Bismarck

³ <https://www.gao.gov/products/GAO-10-246>



attention could prompt them to reconsider the safety of a GRAS substance. However, FDA has largely not responded to concerns about GRAS substances, such as salt and the trans fats in partially hydrogenated vegetable oils, that individuals and consumer groups have raised through 11 citizen petitions submitted to the agency between 2004 and 2008. In fact, FDA has decided on the validity of these concerns in only 1 of 11 cases. In addition, FDA does not know to what extent, or even whether, companies track evolving scientific information about their GRAS substances. FDA's approach to regulating nanotechnology allows engineered nanomaterials to enter the food supply as GRAS substances without FDA's knowledge. While some uses of engineered nanomaterials have the potential to help ensure food safety, uncertainties remain about how to determine their safety in food. After reviewing the uncertainties associated with the safety of engineered nanomaterials, FDA has decided that it does not need additional authority to regulate products containing such materials. Rather, FDA encourages, but does not require, companies considering using engineered nanomaterials in food to consult with the agency regarding whether such substances might be GRAS. Because GRAS notification is voluntary and companies are not required to identify nanomaterials in their GRAS substances, FDA has no way of knowing the full extent to which engineered nanomaterials have entered the U.S. food supply as part of GRAS substances. In contrast to FDA's approach, all food ingredients that incorporate engineered nanomaterials must be submitted to regulators in Canada and the European Union before they can be marketed."

In response, in 2016 the FDA issued a final rule to amend and clarify the criteria in their regulations for when a substance is GRAS under the conditions of its intended use in human food or animal food, and to replace the voluntary administrative procedure for petitioning FDA to affirm the GRAS status of a use of a substance in human food or animal food with a voluntary administrative procedure for notifying FDA about a conclusion that a substance is GRAS under the conditions of its intended use in human food or animal food (FDA, 2016). This article addresses the toxicology for complying with the rule Substances Generally Recognized as Safe issued by the FDA on 17 August 2016⁴ and corrected on 8 September 2016⁵. This is the legal text this article refers to if not noted specifically. In November 2017, FDA released two documents relating to guidance for conclusions for GRAS Regulatory Framework⁶ and draft guidance for conduct of GRAS Expert Panels⁷, which revise and expand the 2016 rule. For discussion see⁸.

2 Background of GRAS⁹

Under the 1958 Food Additives Amendment to the Federal Food, Drug and Cosmetic Act (FFDCA) (Congress, 1958), any substance intentionally added to food is a food additive and subject to pre-market approval by FDA unless the use of the substance is generally recognized as safe (GRAS), or otherwise exempt from the definition of food additive, e.g., color additive. In 1960, Congress passed similar legislation governing color additives. The Color Additives Amendments to the FFDCA require dyes used in foods, as well as in drugs, cosmetics and certain medical devices, to be approved by the FDA prior to their marketing. In contrast to food additives, color additives in use before the legislation were allowed to be used continuously only if they underwent further testing to confirm their safety. Of the original 200 provisionally listed color additives, 90 are listed as safe, and the remainder has either been removed from use by the FDA or withdrawn by industry.

Rulis and Levitt (2009) summarize: "The so-called 'GRAS exemption' to the statutory definition for food additives in Section 201(s) of the FD & C Act expresses Congress' recognition that many commonly used food ingredients, some with long histories of safe use in food or whose safety based on scientific procedures is widely recognized and accepted by qualified experts, need not be subjected to further government scrutiny. As we pointed out above, in the absence of such an exemption, the wide reach of the 'component part' of the food additive definition in the Act would needlessly require many common and ostensibly safe food ingredients and their uses to become the subject of new food additive petitions". A GRAS substance is thus neither safer, nor less safe than an approved food additive. Rather, the distinction between a GRAS substance and an approved food additive is that, for a GRAS substance, there is common knowledge of safety within the expert community.

By 1961, FDA had amended its regulations to include "the GRAS list" – a list of substances that are GRAS under certain conditions of use. During the 1960s, many manufacturers requested FDA's opinion on whether their conclusions of GRAS status were justified and received "opinion letters". In 1969, FDA removed cyclamate salts from its GRAS list as a result of safety questions, and President Nixon directed FDA to re-examine the safety of the GRAS substances. In the 1970s, FDA announced that it was conducting a "comprehensive review" (Select Committee on GRAS Substances (SCOGS) review) of presumed GRAS substances and established rulemaking procedures to affirm the GRAS status of substances that were either on the GRAS list or the subject of a petition ("GRAS affirmation"). In 1997, FDA proposed replacing the then-existing voluntary

⁴ <https://www.federalregister.gov/documents/2016/08/17/2016-19164/substances-generally-recognized-as-safe>

⁵ <https://www.federalregister.gov/documents/2016/09/08/C1-2016-19164/substances-generally-recognized-as-safe>

⁶ <https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/ucm583858.htm>

⁷ <https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/ucm583856.htm>

⁸ <http://burdockgroup.com/fdas-hostile-takeover-gras-ground-rules/>

⁹ see also: http://www.gmaonline.org/file-manager/GMA%20GRAS%20Backgrounder_1.pdf



but resource-intensive GRAS Affirmation Petition process the agency put in place in the 1970's, with a more streamlined voluntary GRAS notification process. The FDA made it clear that both food additives and GRAS ingredients require the same strength of evidence of safety. Rulis and Levitt (2009) point out: *"The GRAS exemption has been much misunderstood and misinterpreted over the years. ... Many people mistakenly associate GRAS with a sort of 'second' tier of safety protection, based on a less-than-rigorous standard compared to petitioned food additives. This is not true. In fact, the safety standard applicable to GRAS food ingredients is the same as for food additives, namely 'reasonable certainty of no harm' For food additives, the FDA reviews and approves the use based on evidence supplied by the petitioner in the form of a petition. For GRAS ingredient uses, the extra added element is the general recognition required that the available safety evidence is both widely known (e.g., available in the published literature or in commonly accessible knowledge sources such as text books) and that there is a consensus among qualified experts about that evidence in support of the safe use of the material in food. In fact, the GRAS criteria are in some ways more difficult to satisfy than the food additive criteria because of the additional requirement of public availability of the data and general recognition and acceptance of a safety conclusion based on those data. ... The FDA also accepts, as part of a GRAS notification, unpublished studies as providing additional support for or corroboration of the published scientific findings, but the critical data and information forming the essential basis of a GRAS determination must be in the public domain."*

Also, FDA (2016) clarifies that a *"substance cannot be classified as GRAS under the conditions of its intended use if the available data and information do not satisfy the safety standard for a food additive under the FD&C Act"*, i.e., *"General recognition of safety requires common knowledge, throughout the expert scientific community knowledgeable about the safety of substances directly or indirectly added to food, that there is a reasonable certainty that the substance is not harmful under the conditions of its intended use"*.

So, in theory, there should be public information equivalent to an FDA petition in the public domain backing any GRAS determination. But as nicely stated (and misattributed to Yogi Berra¹⁰), *"In theory, there is no difference between practice and theory. In practice, there is"*. I already referred to Neltner et al. (2013a) above.

In the scientific literature, however, there are many positive voices about the GRAS system (admitting that advocacy and watchdog groups rarely write scientific articles):

Burdock and Carabin (2004): *"The most resounding testimony to the viability of GRAS is the fact that few GRAS determinations by experts have been overturned and the number of GRAS substances found unsafe is vanishingly small when*

seen in the perspective of the large numbers of substances approved."

Rulis and Levitt (2009): *"Today, the agency's procedures for premarket evaluation of food additive safety have evolved into a scientifically rigorous, sound and dependable system whose objective and independent evaluations by FDA scientists assure that new food additives are safe for their intended uses before they arrive on the consumer's plate."*

Roberts and Haighton (2016): *"In spite of criticisms, the FDA GRAS notice process is clearly defined, efficient, and cost-effective, and there have been no known public health issues following its implementation."*

This would suggest that a key problem is perception. The lack of transparency of the process is certainly critical here. But there is also a lack of clear guidance on how to do a GRAS assessment. Rulis and Levitt (2009): *"What the statute does not provide, however, is an easy-to-follow roadmap describing precisely how one satisfies this standard for GRAS ingredients, other than to say that they must either have a long history of safe use in food by a sufficiently large population of eaters, or that, based on scientific procedures there is a consensus among experts qualified by scientific training and experience to judge such matters that the use of the ingredient is safe"*. This underlines the importance of either clear guidance from the FDA, as currently emerging, or appropriate standards by the industry, which are actually enforced.

3 Rebooting GRAS – FDA update of guidance and GMA standard development

I have chosen the term "reboot" in the title of this article as it describes the process of restarting (a computer) often after some glitches or when installing an update but not as starting some completely new (an overhaul). This is what I see happening at the moment. Two processes currently aim at providing guidance on how to perform GRAS assessments: The FDA has started to revise their Redbook of 2000 (updated 2007) and issued rules and guidance over the last two years and the GMA in August 2014 launched a five-part initiative to help modernize the process of making GRAS assessments of food ingredients. The FDA held a public meeting in December 2014 and collected public comments with a deadline in 2015, but there is no public information on further progress. The GMA launched a Code of Practice in October 2014¹¹, corresponding to the modernizing initiative, which includes five points¹²:

1. GMA will take the lead in defining a standard that will provide clear guidance on how to conduct transparent state of the art of ingredient assessments. These advanced procedures will be documented in a Publicly Available Standard (PAS) for GRAS determinations. The PAS will be a science-based

¹⁰ <https://www.snopes.com/quotes/berra/practicetheory.asp>

¹¹ http://www.gmaonline.org/file-manager/GRAS%20Code%20of%20Practice_1_091015.pdf

¹² http://www.gmaonline.org/file-manager/GMA%20GRAS%20Modernization%20Initiative_1.pdf

framework that specifies a rigorous and transparent ingredient safety assessment process. The procedures included in the PAS will also ensure GRAS assessments meet the regulatory requirements of the Food, Drug and Cosmetic Act. The PAS will be developed by an independent body of technical experts in an open public process that included interested stakeholders. The PAS will be suitable for accreditation using an independent official accreditation body.

2. *GMA is establishing a program to ensure the FDA has increased visibility to the ingredients that are assessed as GRAS by members of the food industry. The increased visibility will be made possible through the establishment of a GMA-sponsored database that will list information on all GRAS assessments conducted by the food industry.*
3. *GMA will expand its curriculum of GRAS education and training programs in order to further increase the capability of scientists who assess the GRAS status of ingredients used by the consumer packaged goods industry.*
 - a. *GMA's broad-based educational programs provide GMA members and other interested stakeholders a clear understanding of the scientific procedures that need to be followed in order to complete a robust and transparent safety regulations so new ingredients are fully compliant with U.S. food additive law and regulations.*
 - b. *GMA has taken the lead in establishing the Center for Research and Ingredient Safety (CRIS) at Michigan State University launched in the spring of 2014. CRIS will serve as an independent academic center of expertise on the safety of ingredients used in foods and consumer products. Their expertise on ingredient safety and independent analysis will be available to all interested stakeholders.*
4. *GMA members have committed to drive improvement in the GRAS assessment process by adopting a Code of Practice at the GMA Board of Directors Meeting held on August 22, 2014. The Code outlines the commitments GMA members have made to conduct assessments according to the procedures defined in the PAS, to maintain the database with up to date information and to ensure that their employees are fully trained on GRAS procedures.*
5. *GMA will execute a communications outreach program to inform stakeholders and consumers of the steps being taken by industry to increase the integrity of procedures used to assess ingredient safety.*

NSF International, an American product testing, inspection and certification organization based in Ann Arbor, Michigan, was charged with developing this standard. Notably, the author is part of this group, led the subgroup for evidence integration and has drafted a toxicological part. This already indicates that I am very supportive of this initiative. Independent of progress on the regulatory side, which is not clear under the current administration, this gives this important industry an opportunity to raise safety standards to the level of the 21st century by creating a new transparency to earn consumer trust.

The standard under development cannot substitute for or modify the FDA rule, but will complement it by defining how to interpret the rule in order to claim compliance with

the standard. In the author's personal view, it is imperative that the standard must be committed to both sound and the most humane science. It must embrace the developments of evidence-based toxicology and the 3Rs, i.e., the principles of reducing, refining and replacing animal use where reasonably possible. The foundations of evidence-based methodologies were laid in medicine over the last few decades. The primary tool of evidence-based medicine (EBM) is the systematic review, which includes a variety of steps: framing the question to be addressed and deciding on how relevant studies will be identified and retrieved, which studies will be excluded from the analysis, how the included studies will be appraised for quality/potential for bias, and how the data will be synthesized across studies (e.g., meta-analysis). Such reviews are also characterized by EBM's hallmark tenets of transparency, objectivity, and consistency.

The 3Rs in relation to science are guiding principles for more ethical use of animals in testing (Russell and Burch, 1959). Their implementation in any test guidance should be a matter of course but, noteworthy, they are not mentioned in the 2000 FDA Redbook. As a guiding principle, they also provide leverage to adapt testing requirements and have a broader scope than simply encouraging alternatives to animal testing, but aim to improve animal welfare and scientific quality where the use of animals cannot be avoided. In the US, the 3Rs are embraced legislatively in the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and National Toxicology Program's Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) (Schechtman, 2002; Casey et al., 2015) as well as by the Toxic Substances Control Act (TSCA) reauthorization of 2016, the Frank R. Lautenberg Chemical Safety for the 21st Century Act. Notably, FDA is a crucial part of the ICCVAM process.

The standard under development should commit to the application of the 3Rs principles, i.e., to use alternative approaches where possible, going beyond the requirements of the FDA GRAS ruling, but staying in line with general FDA policy (Hamburg, 2011). However, for the standard test requirements for food additives, very few such alternatives are already available. An inventory produced by Adler et al. (2011) – though for the purpose of cosmetic regulation in Europe – and confirmed in an independent assessment (Hartung et al., 2011) shows very few sufficiently developed alternative methods applicable for the systemic toxicity testing required under GRAS.

4 Safety evaluations under GRAS compliant with the standard to be developed

No standard is needed simply to comply with existing legislation – it has to go beyond this for the safety and trust of consumers. Specifically, the standard should detail:

- How to compile, generate and evaluate the toxicological information needed
- How to adapt this process to scientific progress
- Where it is recommended to go beyond the standard set by the FDA ruling



- How to document, share and communicate such results
- How to apply quality assurances such as Good Laboratory Practice and evidence-based principles
- How and when to review an evaluation in the future
- How to make older GRAS assessments and the revised GRAS ruling compliant with the standard
- How to bridge this to international requirements, for example the process of the European Food Safety Authority (EFSA)

Systematic evaluation of safety is key to the GRAS process. Burdock and Carabin (2004) remark: “*Importantly, a substance must be found safe; it is insufficient to show only that there is no conclusive evidence that the use of a substance is unsafe*” citing (Hutt and Merrill, 1991). This means that the substance of interest must be experimentally challenged to elicit possible adverse effects using scientific procedures known to do so.

Rulis and Levitt (2009) summarized: “*Unlike the approval of new drugs, the law for food additives does not permit FDA to consider ‘benefits’ from the use of the additive in its decision – rather; it is a safety per se standard. Furthermore, unlike industrial chemicals and pesticides, food additives are generally members of a class of chemicals of relatively low toxic potential, i.e., they have relatively little acute toxicity. Rather, the adverse effects of interest to FDA safety reviewers are usually more subtle and likely to be observed most clearly in animal feeding studies only after a lifetime of exposure (24 months typically). Ultimately, once approved, food additives must be safe for everyone – children and the very young; teenagers and adults; the elderly; pregnant and lactating women. It is assumed that every population subgroup may potentially be exposed to the additive in their diet, and possibly for their entire lifetime. If the use of an additive that is safe for most consumers could present special risks for certain subpopulations, such as those who might be allergic to a particular ingredient or who may have an inborn metabolic deficiency such as phenylketonuria, for example, then FDA can require special labeling so those consumers are properly informed. In addition, unlike the case of drugs, FDA promulgates ‘generic’ regulations for food additives, not a ‘product-specific’ approach as with drugs. Except for the case where a use of an additive is protected by a patent, any company that is in compliance with the conditions of use of the additive specified in the permitting regulation in the Code of Federal Regulations (CFR) may use the additive in food in the way prescribed and within the purview of that regulation.*”

“*Common knowledge*” can be based on either “*scientific procedures*” or on experience based on common use of a substance in food prior to January 1, 1958, which does not fall under the standard to be developed. FDA (2016) requests: “*General recognition of safety through scientific procedures must be based upon the application of generally available and accepted scientific data, information, or methods, which ordinarily are published, as well as the application of scientific principles, and may be corroborated by the application of unpublished scientific data, information, or methods.*”

In general, any GRAS toxicity evaluation has thus to be on par with the FDA recommendations for the evaluation of food additives as published as Guidance for Industry¹³. The guidelines developed for the toxicity tests are valid both for direct and indirect additives. These requirements change with concern level, which is based on chemical structure and exposure (notably, the tiered system applied for direct additives is separate from the tiered system used for indirect additives):

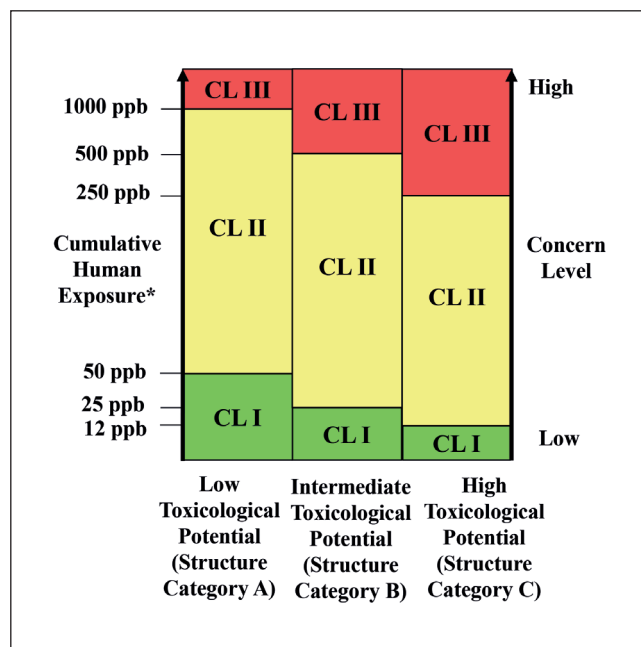


Fig. 1 Concern levels (CL) as Related to Human Exposure and Chemical Structure

Modified from the FDA website¹³.

*Cumulative human exposure is expressed as parts per billion (ppb, equivalent to microgram per kg diet) of daily dietary consumption. Conversion of ppb to microgram per kg-body weight per day, divide by 20, assuming 3 kg daily diet.

For any GRAS consideration, establishing the concern level is thus fundamental. Given the fact that no risk-benefit argument can be applied, but a no-risk approach is required, the studies cited serve the exclusion of a risk. Most substances will therefore fall *a priori* into concern level I, where testing costs amount to about \$80,000. Concern level II will easily surpass \$1 million and concern level III can reach multi-million \$, especially if human studies are required. Notably, the addition to most testing needs “*if indicated by available data or information*” leaves room for a lot of waiving of testing. The concern level is, furthermore, driven by the expected exposure level (Barraj et al., 2016). It is advisable, where there are concerns identified either from structural read-across or biological profiling to directly address this concern and not plod on through the “laundry list” of tests. In such a tiered approach, evident

¹³ <https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/ucm054658.htm>

unsuitability of a substance in a given test should lead to a stop of further testing.

FDA advises to estimate “*potential toxicity based on structural similarity to known toxicants in the absence of toxicological information of an additive*”. This process using structural similarity to known toxicants is often referred to as read-across. Lately, initiatives to establish a Good Read-Across Practice (GRAP) have resulted in initial guidance (Ball et al., 2016); GRAP should be used in order to comply with the standard to be developed and any deviation should be justified.

Increasingly, within the context of the Tox21 alliance between EPA, FDA and NIH, also biological characterization of substances is used for priority setting. This might in the future help determine concern levels and could also be combined with read-across considerations (Zhu et al., 2016). However, up to now, neither a publicly available process for subjecting a substance of interest to such standardized testing nor a process of deriving concern levels has been established.

The guidance to industry includes a *Summary Table of Recommended Toxicological Testing for Additives Used in Food*. It is included in Table 1. FDA refers to current guidelines on these toxicological studies in the relevant final chapters in Redbook

2000, or chapters published in the draft 1993 Redbook II. With respect to this article, these thus have to be considered as point of reference. However, as stated by Kruger (2016): “*it is critical to remember that Redbook does not provide this guidance as a checklist. The reason for the flexibility in approach is that science evolves. Thinking evolves. Risk assessments should incorporate and embrace these new advances*”. Progress over the last two decades to improve these scientific procedures should be used to attain the same or higher level of safety, but will require case-by-case justification and, where available, validation. As international acceptance including the US, as the consensus process of OECD requires, can be seen as the strongest evidence that a scientific procedure and its results are “generally recognized as safe”, preference is given by the author to OECD test guidelines and their continuous updates, which safeguards also that scientific progress is implemented. Future revisions of these guidelines should be considered for incorporation under this standard, if no obvious reasons hinder this. This includes the continuous development of novel test guidelines for new and alternative approaches. These novel methods, sometimes now referred to as new approach methods (NAM), will be discussed in more detail below.

Tab. 1: Information requirements according to the FDA website triggered by concern levels, currently available options for adaptation, and some pertinent opportunities for future improvements

Endpoint	Required from Concern Level	Current Requirement Redbook	Adaptation to Technical Progress	Future Opportunities
Genetic toxicity tests	Low (I)	2-3 tests ^a if daily intake in the diet >50 ppb (150 µg per person per day): Gene mutations in bacteria and <i>in vitro</i> test for chromosomal damage in mammalian cells or <i>in vitro</i> mouse lymphoma thymidine kinase ^{+/-} and an <i>in vivo</i> test for chromosomal damage using hematopoietic cells	Micronucleus test Comet assay Reduction of false-positives (maximum concentrations, p53 competent cells, etc.) Transgenic rodents TG488 with fewer false positives and possibly combined with repeated-dose testing	Refinement and reduction of animal numbers for <i>in vivo</i> tests <i>In silico</i> approaches Mutagenicity in skin models
Short-term toxicity tests with rodents	Low (I) if indicated by available data or information	14 or 28 days (one month) ^b including neurotoxicity and immunotoxicity screens	Updated TG 407 Combination with <i>in vivo</i> mutagenicity Combination with developmental toxicity screening assays (TG 422)	Human-on-chip approaches Integrated testing strategies Systems toxicology
Subchronic toxicity studies with rodents	Intermediate (II) if indicated by available data or information	90 days (3 months) up to 12 months including neurotoxicity and immunotoxicity screens	Updated TG 408	Possible prediction from 28-day study Assessment factors

^a <https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/IngredientsAdditivesGRASPackaging/ucm078321.htm>

^b <https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/IngredientsAdditivesGRASPackaging/ucm078339.htm>



Endpoint	Required from Concern Level	Current Requirement Redbook	Adaptation to Technical Progress	Future Opportunities
Subchronic toxicity studies with non-rodents	Intermediate (II) if indicated by available data or information	Usually dogs, generally conducted for 90 days (3 months) including neurotoxicity and immunotoxicity screens	Updated TG 409	
One-year toxicity studies with non-rodents	High (III) if indicated by available data or information	Usually dogs for a minimum of 12 months (one year) including neurotoxicity and immunotoxicity screens	Possibly abandon based on pesticide experience	
Chronic toxicity or combined chronic toxicity/carcinogenicity studies with rodents	High (III) if indicated by available data or information	Rodents, conducted for a minimum of 12 months (one year) including neurotoxicity and immunotoxicity screens	Updated TG 452 Combined with carcinogenicity TG 453	
Carcinogenicity studies with rodents	High (III)	Carcinogenicity bioassays in two rodent species (usually rats and mice)	Updated TG 451 TG 453 combined with chronic toxicity Consider cell transformation assays (OECD guidance available) Consider abandoning completely based on poor performance Abandon mouse bioassay	Integrated testing strategy
Reproduction studies	Intermediate (II) if indicated by available data or information	Studies of substances given orally to rodents including neurotoxicity and immunotoxicity screens	Define suitable OECD TG TG 443 (Extended One-Generation Reproductive Toxicity Study) instead of TG 416 (Two-Generation Reproduction Toxicity Study) Consider abandoning completely based on poor performance	Zebrafish reproductive toxicity assay Integrated testing strategy
Developmental toxicity studies	Intermediate (II) if indicated by available data or information	Stand-alone study, or may be part of a multigeneration reproduction study, orally given to rat, mouse, hamster, or rabbit including neurotoxicity and immunotoxicity screens	TG 421 or TG 422	<i>In vitro</i> embryotoxicity studies especially where second species is considered
Metabolism and pharmacokinetic studies	Intermediate (II) if indicated by available data or information	Determination of metabolic pathways and the rates of metabolism in different test species – see draft Redbook 1993		Integrated testing strategy of diverse assays available Gut-on-chip
Human studies	High (III) if indicated by available data or information	Not defined – see draft Redbook 1993		

The FDA website further states “*Minimum toxicological testing recommended to support the safety of a novel additive might include studies generally recommended for a Concern Level III additive, irrespective of its chemical structure and exposure. Additionally, toxicological testing may be needed for metabolites or degradation products, as well as for possible impurities, of an additive to establish safety of these components*”. This means that for any GRAS evaluation, metabolites, degradation products and impurities need to be considered. In many cases, *in vivo* testing of such compounds is not practically feasible with respect to their number, the quantity available (extended animal testing requires several kg of the substance) and associated costs.

For compliance with a standard to be developed, the concern level for metabolites, degradation products and impurities needs to be established, favorably by validated (Q)SAR and read-across (Luechtefeld and Hartung, 2017). The use of a combination of chemical analytical and *in vitro* bioassays especially for contaminants has been proposed (Severin et al., 2017). Additional testing and risk assessment needs have to be established case-by-case considering expected doses (also considering the evaluation of metabolism and pharmacokinetic studies of the mother compound), nature and extent of the hazard of concern, and whether the effect is more severe than that of the mother compound.

4.1 Limiting testing needs

All available information on a given substance needs to be considered before and when carrying out a GRAS evaluation. A systematic review of the literature (Rooney et al., 2014; Stephens et al., 2016; Hoffmann et al., 2017) represents the most appropriate tool here.

Any substance that has already received an accepted GRAS evaluation must not be registered again. However, no complete list is available, only extensive examples from FDA¹⁴. With the new mandatory registration of GRAS evaluations, and the database promised by GMA, a comprehensive list should be provided. In compliance with the standard to come, a substance that has undergone an FDA food additive or GRAS evaluation should not be retested. Similarly, studies and risk assessments available from other uses or internationally should be used to the extent possible. Kruger (2016) remarks “*In the EU and Australia, regulatory guidance provides a framework for risk assessment applied to novel foods, nutritive substances and food additives is similar to that used in the US for food additives and GRAS ingredients*”. A pertinent example is the assessment framework developed by the EFSA Panel on Food Additives and Nutrient Sources Added to Food (ANS) (2014), which is largely in line with GRAS assessments and because of its scientific rigor could be considered as generally recognized as safe. The legitimate access to such data, however, needs to be clarified. The use of such preexisting studies and assessments comes, however, with the problem of varying study quality. A formal evaluation of study quality is required (for an overview

see Samuel et al., 2016) to allow proper use. For example, the Tox-R-Tool (Schneider et al., 2009) allows systematically assigning Klimisch scores of study quality for *in vivo* and *in vitro* studies.

The US FDA introduced a threshold of regulation (TOR) approach for indirect food additives (FDA, 1995). The TOR represents a pragmatic way to address the safety of food packaging materials that have the potential to migrate into food at a level that is considered to be sufficiently low to be considered toxicologically insignificant, even in the absence of chemical-specific toxicity data. This was based on a statistical analysis of the Carcinogenic Potency Database (CPDB) of Gold et al. (1984, 1989) and the TOR of 0.5 ppb in the diet, corresponding to 1.5 µg/person/day (FDA, 1995, 2001) or 0.025 µg/kg body weight/day, was set (Rulis, 1986, 1989). This formed the starting point for thresholds of toxicological concern (TTC) applied to impurities in drugs and food, but increasingly also other sectors. Since development of the TTC approach in 1995, the GRAS Expert Panel of Flavor and Extract Manufacturers Association of the United States (FEMA) has adopted the TTC principle according to the chemical classes and thresholds (Hallagan and Hall, 2009). The GMP regulation limits the amount of food and color additives used in foods: manufacturers may use only the amount of an additive necessary to achieve the desired effect. Noteworthy, EFSA with its Scientific Panel published a *Scientific opinion on exploring options for providing advice about possible human health risks based on the concept of Threshold of Toxicological Concern (TTC)* (EFSA, 2012).

With endpoint-specific TTC and raised TTC values for certain chemical classes, they become increasingly suitable also to limit testing for low level substance uses (Hartung, 2017), such as direct food additives. The first to use the TTC for flavoring agents was the Joint FAO/WHO Expert Committee on Food Additives (JECFA, 1993, 1995, 1999). The TOR approach was later expanded into a tiered TTC decision tree (Kroes et al., 2005; Munro et al., 2008). Both JECFA and EFSA have been using TTC for evaluating risk assessment of flavoring agents with very low levels of exposure through the diet (JECFA, 1997; EFSA, 2012; Renwick, 2004). The TTC is based on the concept that an exposure threshold value can be established below which a very low probability of an appreciable risk to human health (or the ecosystem) exists (Hennes, 2012). TTC represent rather simple limits, which could help prioritize resources for other cases. They definitively lend themselves to metabolite, degradation product and impurity evaluations.

Earlier in this series (Hartung, 2017), it was suggested that read-across / (Q)SAR and biological characterization help set concern levels for a substance and a given hazard. The combination of exposure information formalized as TTC *versus* intended use and comparing with no-effect levels of other (similar) chemicals can then help to decide whether and how extensive the testing and risk assessment need to be.

¹⁴ <https://www.ecfr.gov/cgi-bin/text-idx?SID=e956d645a8b4e6b3e34e4e5d1b690209&mc=true&node=pt21.3.182&rgn=div5>



4.2 Quality assurance

The FDA requests that toxicological information is generated under Good Laboratory Practice (GLP)¹⁵. Currently, complementary guidance for *in vitro* tests is being developed: Good In Vitro Method Practices (GIVIMP) for the Development and Implementation of In Vitro Methods for Regulatory Use in Human Safety Assessment¹⁶. This corresponds closely with attempts to develop a revised Good Cell Culture Practice (Pamies and Hartung, 2017; Pamies et al., 2017; Eskes et al., 2017; Pamies et al., 2018). Please consult these references for more details. However, it needs to be stressed that rigorous quality control of *in vitro* methods represents the door-opener to a world still largely relying on animal tests.

4.3 Endpoint-specific considerations

In the following, the current requirements and perspectives to modernize GRAS assessments are discussed. Table 1 shows the information requirements according to the FDA website triggered by concern levels (Fig. 1) and defined by the Redbook of 2000 and, in part, its draft update of 1993. Currently available options for adaptation are given next (explained in the text) and some pertinent opportunities for future improvements are mentioned last. All together, these testing requirements are very light for low concern levels and intense for high concern levels. Notably, many information needs include the notion “*if indicated by available data or information*”, which opens up opportunities for waiving.

4.3.1 Genetic toxicity tests

The FDA-recommended test battery for food ingredients whose cumulative estimated daily intake exceeds 50 ppb in the diet (150 µg per person per day) generally includes genetic toxicity testing. The Agency prefers the mouse lymphoma tk[±] assay because this assay measures heritable genetic damage arising by several mechanisms in living cells and is capable of detecting chemicals that induce either gene mutations or heritable chromosomal events, including genetic events associated with carcinogenesis. In performing the mouse lymphoma tk[±] assay, either the soft agar or the microwell method is acceptable.

The current FDA guidance for food additives reflects the state of the art of the year 2000. Most regulatory agencies and international authorities today recommend a test scheme consisting of *in vitro* and *in vivo* methods to identify genotoxic/mutagenic substances. A tiered test scheme would likely start with computer-based prediction using (quantitative) structure-activity relationships [(Q)SAR] or read-across and *in vitro* testing. The International Conference on Harmonization (ICH) for drug safety testing recommends a standard core battery for pharmaceuticals, which was adopted by FDA (2012). This guidance, however, is more demanding (either bacteria test + cytogenetic test + *in vivo* one tissue or bacteria test + *in vivo* in two tissues), and considers

already measures to reduce false positives. The latter is of critical importance for possible reassessments of substances already in use, as false-positives could trigger enormous concerns and testing needs. The ICH approach could be considered a prudent approach for testing of food additives, but notably has not been endorsed by FDA for this purpose.

The current OECD test portfolio and the respective US validation statuses are given in Table 2. The table illustrates the enormous variety of genotoxicity tests in use.

Some key developments since the Redbook of 2000 include the successful validation of the *in vitro* Micronucleus Test (MNT) as an alternative to the standard *in vitro* Chromosome Aberration Test (CAT) (Corvi et al., 2008). It is less expensive and time-consuming, requires less investment in training, allows a greater statistical power, and has the potential to enhance the basic package of *in vitro* tests to detect aneugens. The validation study has supported the finalization of OECD TG 487¹⁷ and its regulatory acceptance.

Ongoing work by ECVAM and Cosmetics Europe aims to establish and validate new methods for genotoxicity testing in reconstructed human 3D skin models (micronucleus test and comet assay) (Maurici et al., 2005; Aardema et al., 2010; Reus et al., 2013), again with the key goal of reducing false positives. The Comet assay is now being validated in a joint effort between Cosmetics Europe and BfR, Germany. The Japanese Centre for the Validation of Alternative Methods (JaCVAM) validated the Comet assay *in vivo* and the respective OECD TG 489¹⁸ has been approved.

The high false positive rate of the established *in vitro* genotoxicity tests leads to an increased number of follow up *in vivo* tests needed for the confirmation or more often falsification of these results. ECVAM organized a workshop in April 2006 aimed at discussing how to improve current *in vitro* tests and at reviewing the development of new tests. The recommendations of this workshop (Kirkland et al., 2007) paved the way for several international initiatives (e.g., Cosmetics Europe, ILSI/HESI, ECVAM, NC3Rs, JaCVAM). Testing at high concentrations was identified as one possible source of false positives. Similarly, the right choice of cell line is critical (Fowler et al., 2012). To address this issue, an analysis of published data for top concentration considerations in mammalian cell genotoxicity testing was carried out, which suggested that the top concentration could be reduced without any loss of sensitivity in detecting rodent carcinogens (Parry et al., 2010). Some OECD TGs for genotoxicity have thus recently been revised, taking also into account the knowledge gained during the last decades of testing and the recent activities related to false positives. The recommendations made in the revised TGs will most probably enhance the quality of the data that will be produced. ICH now recommends a maximum top concentration of 1 mM or 0.5 mg/ml,

¹⁵ <http://www.oecd.org/chemicalsafety/testing/oecdseriesonprinciplesofgoodlaboratorypracticeglpandcomplianceandmonitoring.htm>

¹⁶ http://www.oecd.org/env/ehs/testing/OECD%20Draft%20GIVIMP_%2020161028.pdf

¹⁷ http://www.oecd-ilibrary.org/environment/test-no-487-in-vitro-mammalian-cell-micronucleus-test_9789264091016-en

¹⁸ http://www.oecd-ilibrary.org/environment/test-no-489-in-vivo-mammalian-alkaline-comet-assay_9789264264885-en

Tab. 2: Current OECD test portfolio and the respective US validation statuses

OECD Test Guideline	Test	ICCVAM validation status ^a
TG 471	Bacterial Reverse Mutation Test (Ames Test)	
TG 472	Genetic Toxicology: <i>Escherichia coli</i> , reverse assay	
TG 473	<i>In Vitro</i> Mammalian Chromosome Aberration Test	Accepted
TG 474	Mammalian Erythrocyte Micronucleus Test	
TG 475	Mammalian Bone Marrow Chromosome Aberration Test	
TG 476	<i>In Vitro</i> Mammalian Cell Gene Mutation Test	
TG 477	Genetic Toxicology: Sex-linked Recessive Lethal Test in <i>Drosophila melanogaster</i>	
TG 478	Genetic Toxicology: Rodent Dominant Lethal Test	
TG 479	Genetic Toxicology: <i>In Vitro</i> Sister Chromatid Exchange Assay in Mammalian Cells	
TG 480	Genetic Toxicology: <i>Saccharomyces cerevisiae</i> , Gene Mutation Assay	
TG 481	Genetic Toxicology: <i>Saccharomyces cerevisiae</i> , Mitotic Recombination Assay	
TG 482	Genetic Toxicology: DNA Damage and Repair, Unscheduled DNA Synthesis in Mammalian Cells <i>In Vitro</i>	
TG 483	Mammalian Spermatogonial Chromosome Aberration Test	
TG 484	Genetic Toxicology: Mouse Spot Test	
TG 485	Genetic Toxicology: Mouse Heritable Translocation Assay	
TG 486	Unscheduled DNA Synthesis (UDS) Test with Mouse Liver Cells <i>In Vitro</i>	
TG 487	<i>In Vitro</i> Mammalian Cell Micronucleus Test	Accepted (replacement of animal use)
TG 488	Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays	
TG 489	<i>In Vivo</i> Mammalian Alkaline Comet Assay	Accepted (reduction of animal use)
TG 490	<i>In Vitro</i> Mammalian Cell Gene Mutation Tests Using the Thymidine Kinase Assay	Accepted

^a <https://ntp.niehs.nih.gov/pubhealth/evalatm/iccvam/acceptance-of-alternative-methods/index.html>

whichever is lower, when not limited by solubility in solvent or culture medium or by cytotoxicity. With respect to cytotoxicity, ICH now recommends for *in vitro* cytogenetic assays for metaphase chromosome aberrations or for micronuclei that cytotoxicity should not exceed a reduction of about 50% in cell growth, and for the mouse lymphoma assay there should be 80-90% cytotoxicity at the top dose as measured by a relative total growth between 10 to 20%.

The Ames test (OECD TG 471¹⁹), conducted in bacteria, is the most commonly used genotoxicity test within the *in vitro* battery as it is considered able to reveal DNA reactivity and DNA reactive compounds. Despite the many activities investigating false positive results in *in vitro* mammalian cell tests, positive results in the Ames test have not been analyzed in the same way. In this context, EURL ECVAM held a workshop in January 2013 and initiated a project with the aim of evaluating

the predictivity of the Ames test for *in vivo* genotoxicity and carcinogenicity when considered alone or in association with a mammalian cell assay for the detection of chromosome damage and/or gene mutations, and better characterizing the cases where the Ames test seems to lead to irrelevant (false positive) results (e.g., chemical classes, type of bacterial strains, magnitude of effects). A possible impact on the recommendations of a standard has to be awaited.

Since substantial *in vivo* testing is still required by authorities for confirmation of the genotoxic prediction *in vitro*, it became clear that it is crucial to address issues related to the reduction and refinement of *in vivo* genotoxicity tests. The collection of relevant data might be considered as a basis for possible amendments of guidelines to reduce animal consumption. Recommendations on opportunities to reduce the number of animals in genotoxicity tests are published by Pfuhler et al. (2009).

¹⁹ http://www.oecd-ilibrary.org/environment/test-no-471-bacterial-reverse-mutation-test_9789264071247-en



The use of transgenic rodents (TGR) might bear advantages for *in vivo* mutagenicity testing and OECD has adopted TG 488²⁰: Transgenic Rodent Somatic and Germ Cell Gene Mutation Assays in 2013. TGR gene mutation assays measure mutations induced in genetically neutral genes recovered from virtually any tissue of the rodent. These assays, therefore, circumvent many of the existing limitations associated with the study of *in vivo* gene mutation in endogenous genes (e.g., limited tissues suitable for analysis, negative/positive selection against mutations). The TG anticipated that in the future it may be possible to combine a TGR gene mutation assay with a repeat dose toxicity study (TG 407), see below.

Approaches to improve the GRAS process by addition or replacement of standard procedures used for genotoxicity and carcinogenicity testing with alternative assays have been made (Williams et al., 2016), but not yet evaluated by FDA.

4.3.2 Short-term toxicity tests with rodents

For food additives, FDA (2012) expects short-term toxicity studies with rodents, which are generally conducted over 14 or 18 days. Results of these studies (1) can help predict appropriate doses of the test substance for future subchronic or chronic toxicity studies, (2) can be used to determine NOELs for some toxicology endpoints, and (3) allow future studies in rodents to be designed with special emphasis on identified target organs.

An updated OECD TG 407²¹ is available since 2008, which should represent the default requirement under a standard to come. A combination with the *in vivo* mutagenicity TGR tests should be considered (Akagi et al., 2015). The more recent micronucleus and Comet assays also might be combined as they can be carried out *ex vivo* on blood lymphocytes, but no such validations or guideline inclusions are available.

A combination with developmental toxicity screening assays, i.e., TG 422²²: Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test, is also possible.

No alternative methods for this test have been validated by ICCVAM/NICEATM and/or accepted by US agencies.

Human-on-chip approaches (Andersen et al., 2014; Marx et al., 2016) and integrated testing strategies / integrated approaches to testing and assessment (ITS/IATA) (Hartung et al., 2013; Rovida et al., 2015), possibly designed led by adverse outcome pathways (AOP) (Tollefsen et al., 2014), represent the most promising approaches but are in their infancy. Long-term visions include systems toxicology approaches (Hartung et al., 2017). See also the roadmap for this endpoint developed by Basketter et al. (2012).

4.3.3 Subchronic toxicity studies with rodents

Subchronic toxicity studies with rodents are generally conducted for 90 days (3 months), but they may be conducted for up to 12 months. Results of these studies (1) can help predict appropriate doses of the test substance for future chronic toxicity studies, (2) can be used to determine NOELs for some toxicology endpoints, and (3) allow future long-term toxicity studies in rodents and non-rodents to be designed with special emphasis on identified target organs. Subchronic toxicity studies usually cannot determine the carcinogenic potential of a test substance.

Taylor et al.'s (2014) research suggests a redundancy between 28-day and 90-day repeated dose toxicity tests, proposing that the latter can be predicted given the former. They propose that non-toxicity (NOAEL \geq 1000 mg/kg b.w.) in the 28-day test is a strong predictor for non-toxicity in the 90-day study. Luechtefeld et al. (2016) confirmed in a much larger dataset that the 28-day NOAEL is predictive (albeit imperfectly) of 90-day NOAELs, however, the constraints suggested by Taylor et al. did not affect predictivity. The interrelationship of 28-day and 90-day NOAEL is also used very pragmatically within the European REACH legislation in line with ECHA's test guidance to estimate from a 28-day study a 90-day derived no-effect level (DNEL) using an assessment factor of 3 (ECETOC, 2010). Luechtefeld et al. (2016) used the dataset also to test this assumption. 133 substances had both 90-day and 28-day key studies. Only 11 chemicals had NOAEL below one third of the 28-day NOAEL, while 122 (91.7%) were within this limit. Given the reproducibility issues of repeated dose studies (Wang and Gray, 2015), this is strong confirmation of this pragmatic approach. The approach has not been evaluated for food additives though.

There has been no update of OECD TG 408²³: Repeated Dose 90-Day Oral Toxicity Study in Rodents, which therefore should be the default.

No alternative methods for this test have been validated by ICCVAM/NICEATM and/or accepted by US agencies.

4.3.4 Subchronic toxicity studies with non-rodents

Subchronic toxicity studies with non-rodents (usually dogs) are generally conducted for 90 days (3 months), but they may be conducted for up to 12 months. Results of these studies (1) can help predict appropriate doses of the test substance for future chronic toxicity studies, (2) can be used to determine NOELs for some toxicology endpoints, and (3) allow future long-term toxicity studies in rodents and non-rodents to be designed with special emphasis on identified target organs. Subchronic toxicity studies usually cannot determine the carcinogenic potential of a test substance.

There has been no update of OECD TG 409²⁴ from 1998: Repeated Dose 90-Day Oral Toxicity Study in Non-Rodents, which therefore should be the default.

²⁰ http://www.oecd-ilibrary.org/environment/test-no-488-transgenic-rodent-somatic-and-germ-cell-gene-mutation-assays_9789264122819-en

²¹ http://www.oecd-ilibrary.org/environment/test-no-407-repeated-dose-28-day-oral-toxicity-study-in-rodents_9789264070684-en

²² http://www.oecd-ilibrary.org/environment/test-no-422-combined-repeated-dose-toxicity-study-with-the-reproduction-developmental-toxicity-screening-test_9789264070981-en

²³ http://www.oecd-ilibrary.org/environment/test-no-408-repeated-dose-90-day-oral-toxicity-study-in-rodents_9789264070707-en

²⁴ http://www.oecd-ilibrary.org/environment/test-no-409-repeated-dose-90-day-oral-toxicity-study-in-non-rodents_9789264070721-en

No alternative methods for this test have been validated by ICCVAM/NICEATM and/or accepted by US agencies.

4.3.5 One-year toxicity studies with non-rodents

Long-term, one-year toxicity tests with non-rodents (usually dogs) should be conducted for a minimum of 12 months (one year). Results of these tests can be used to (1) characterize the toxicity of the test substance in non-rodents and (2) determine the dose of the test substance that produces no observed adverse effects (NOEL or NOAEL) for some toxicological endpoints. One-year toxicity tests are not conducted for the purpose of assessing carcinogenicity, although data from these tests may reveal information about the carcinogenicity of the test substance.

Extensive experience with these assays exists only for pesticides. A number of analyses showed very convincingly that the test does not add to the 90-day-study in dogs (Appleman and Feron, 1986; Contrera et al., 1993; Gerbracht and Spielmann, 1998; Spielmann and Gerbracht, 2001; Box and Spielmann, 2005; Doe et al., 2006; Dellarco et al., 2010; Kobel et al., 2010, 2014; Linke et al., 2017). While this challenges the value of this animal test also for food additives, there is no similar analysis. Given this extensive analysis, however, it appears to be prudent to carry out this study only in case of specific concerns.

There is no OECD test guideline for one-year toxicity tests with non-rodents (usually dogs); thus, the default, if the study is warranted, is the original Redbook guidance²⁵.

No alternative methods for this test have been validated by ICCVAM/NICEATM and/or accepted by US agencies.

4.3.6 Chronic toxicity or combined chronic toxicity/carcinogenicity studies with rodents

Chronic toxicity studies with rodents should be conducted for a minimum of 12 months (one year). Results of these tests can be used, 1) to characterize the toxicity of a food ingredient following prolonged and repeated exposure, and 2) to determine toxicological dose-response relationships needed to establish the maximum dose that produces no adverse effects (i.e., NOEL or NOAEL).

An updated OECD TG 452²⁶ Chronic Toxicity Studies was published in 2009; this should represent the default requirement. Also, there is a revised TG 453²⁷ Combined Chronic Toxicity/Carcinogenicity Studies from 2009, which is preferred with respect to the number of animals used.

No alternative methods for this test have been validated by ICCVAM/NICEATM and/or accepted by US agencies.

4.3.7 Carcinogenicity studies with rodents

Carcinogenicity studies (bioassays) in two rodent species (usually rats and mice) are recommended for food ingredients with the highest levels of concern (e.g., Concern Level III direct food additives, food contact substances with cumulative exposure at

or greater than 1 ppm). The carcinogenicity studies (preferably in rats) may be combined with chronic toxicity studies. The FDA recommends that *in utero* exposure be included in carcinogenicity studies due to the fact that exposure to food ingredients occurs during all stages of life. These studies are designed to determine whether a food ingredient possesses carcinogenic activity when administered to rodents in regularly repeated oral doses for the “lifetime” of the test animal.

Both the Food Additives and Color Additives Amendments include a provision that prohibits the approval of an additive if it is found to cause cancer in humans or animals (the Delaney Clause).

Noteworthy, the cancer bioassay is seen increasingly critically as to its performance characteristics, reproducibility and relevance to humans (Alden et al., 2011; Basketter et al., 2012; Paparella et al., 2017). The use of mice as second species has been questioned for pesticides (van Ravenzwaay, 2010), but no similar analysis is available for direct food additives. As there is no process of counterbalancing especially false-positive results and due to their immediate impact on the usability as food additive, carcinogenicity studies are highly problematic. Recommendations to abandon the test have been made (Basketter et al., 2012) and, noteworthy, there are first activities to replace the test by an ITS (Luijten et al., 2016).

The *in vitro* cell transformation assays (CTAs) have been shown to closely model some key stages of the *in vivo* carcinogenesis process and have been in use for more than four decades to screen for potential carcinogenicity as well as investigate mechanisms of carcinogenicity. Moreover, they are faster and more cost-efficient than the *in vivo* rodent carcinogenicity assay, providing a useful approach for screening of chemicals with respect to their carcinogenic potential.

CTAs are considered to provide additional useful information to more routinely employed tests for assessing carcinogenic potential and are therefore listed in various recent guidelines and testing strategies for such purposes (SCCP, 2010²⁸; Jacobson-Kram and Jacobs, 2005; EC, 2008; Pfuhrer et al., 2010). Since regulatory agencies may receive and review CTA data and these assays are used for internal risk assessment of various chemicals, there was a need within the scientific community for standardization of these test methods and technical guidance on their conduct and use. This need was already addressed in 1998 by a workshop organized by ECVAM on CTAs (Combes et al., 1999) and by the OECD that produced a detailed review paper (DRP) on the CTAs for the detection of chemical carcinogens (OECD, 2007). As with some other assays with a long history of use, CTAs had not undergone formal validation in accordance with current standards (OECD GD 34, 2005). Therefore, ECVAM coordinated an international study that was designed to address issues of CTA protocol standardization, transfer-

²⁵ <https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/IngredientsAdditivesGRASPackaging/ucm078348.htm>

²⁶ http://www.oecd-ilibrary.org/environment/test-no-452-chronic-toxicity-studies_9789264071209-en

²⁷ http://www.oecd-ilibrary.org/environment/test-no-453-combined-chronic-toxicity-carcinogenicity-studies_9789264071223-en

²⁸ https://ec.europa.eu/health/scientific_committees/consumer_safety/docs/scs_s_004.pdf



ability and reproducibility (Vanparys et al., 2010). The study assessed two protocol variants for the SHE CTA (at pH 6.7 and pH 7.0) and the BALB/c 3T3 assay. This study was peer-reviewed by the EURL ECVAM Scientific Advisory Committee (ESAC) that issued an ESAC opinion, leading to the publication of an EURL ECVAM Recommendation (ECVAM, 2012) on three CTAs for assessment of the carcinogenic potential of chemical substances. An OECD Test Guidance document on the CTA in SHE cells has been published (OECD, 2015). Two Guidance Documents on cell transformation assays have been drafted at the OECD to allow the scientific and regulatory communities to use the described method as part of a weight-of-evidence approach in the testing of substances for carcinogenic potential. These are the “In vitro Syrian hamster embryo cell transformation assay”, which has recently been adopted (OECD, 2015) and the “In vitro Bhas 42 cell transformation assay” (the Bhas 42 cell line was established by the transfection of the v-Ha-ras oncogene into the BALB/c 3T3 A31- 1-1 cell line). The carcinogenic potential of a substance cannot be derived from a stand-alone CTA. The complete study results as well as the recommended CTA protocols and photo catalogues developed during the ECVAM study are published in a special issue of Mutation Research on CTA (Corvi and Vanparys, 2012). The Japanese Centre for the Validation of Alternative Methods coordinated the validation of the Bhas 42 CTA, a system derived from the BALB/c 3T3 CTA. The study addressed two protocols: a 6-well method and the 96-well method and has been peer reviewed by ESAC. Based on the validation report and the ESAC Opinion, EURL ECVAM issued a Recommendation on the Bhas 42 CTA.

Approaches to improve the GRAS process by addition or replacement of standard procedures used for genotoxicity and carcinogenicity testing with alternative assays have been made (Williams et al., 2016) but not yet evaluated by FDA.

An updated OECD TG 451²⁹: Carcinogenicity Studies is available since 2009, which should represent the default requirement. Also dating from 2009, there is a revised TG 453³⁰ Combined Chronic Toxicity/Carcinogenicity Studies.

No alternative methods for this test have been validated by ICCVAM/NICEATM and/or accepted by US agencies.

Future opportunities, as laid out in the roadmap for replacement (Basketter et al., 2012; Leist et al., 2014), include abandoning the test based on poor performance, and including ITS around cell transformation assays.

4.3.8 Reproduction studies

The guideline for reproduction studies pertains to substances given orally to rodents. It is designed to evaluate the effects of a test substance on the reproductive systems of both males and females, the postnatal maturation and reproductive capacity of offspring, and possible cumulative effects through several generations. A study can provide information concerning the effects of a substance on gonadal function, estrous cycles, mating behavior, conception, parturition, neonatal morbidity, mortality, lactation, weaning, growth and development of the offspring, and target organs in the offspring. The study may also serve as a guide for subsequent tests. The endpoints evaluated and the indices calculated must provide sufficient information and statistical power to permit the FDA to determine whether the chemical is associated with changes in reproduction and fertility.

Current OECD guidelines addressing reproductive toxicity effects in mammals include Prenatal Developmental Toxicity Study (TG 414³¹ from 2001), Two-Generation Reproduction Toxicity Study (TG 416³² from 2001), One-Generation Reproductive Toxicity Study (TG 415³³ from 1983), Reproduction/Developmental Toxicity Screening Test (TG 421³⁴ from 2016), Combined Repeat Dose Toxicity Study with the Reproduction/Developmental Screening Test (TG 422³⁵ from 2016), Extended One-Generation Reproductive Toxicity Study (TG 443³⁶ from 2012) and Developmental Neurotoxicity (TG 426³⁷ from 2007).

OECD TG 443 is based on the International Life Science Institute (ILSI)-Health and Environmental Sciences Institute (HESI), Agricultural Chemical Safety Assessment (ACSA) Technical Committee proposal for a life stage F1 extended one-generation reproductive study (Cooper et al., 2006). Because of its animal saving potential, its partial validation and the inclusion of immunotoxicity as well as neurodevelopmental endpoints (for which only guidance from the 1993 draft Redbook is otherwise available) as triggered options, TG 443 is the preferred reproductive toxicity test, which makes also additional developmental toxicity studies obsolete.

No animal-free alternative methods for this test have been validated by ICCVAM/NICEATM and/or accepted by US agencies.

Future opportunities as laid out in the roadmap for replacement (Basketter et al., 2012; Leist et al., 2014) include abandoning the test based on poor performance, and including ITS including zebrafish egg and (human) stem cell embryotoxicity tests. See also Bremer et al. (2007).

²⁹ http://www.oecd-ilibrary.org/environment/test-no-451-carcinogenicity-studies_9789264071186-en

³⁰ http://www.oecd-ilibrary.org/environment/test-no-453-combined-chronic-toxicity-carcinogenicity-studies_9789264071223-en

³¹ http://www.oecd-ilibrary.org/environment/test-no-414-prenatal-development-toxicity-study_9789264070820-en

³² http://www.oecd-ilibrary.org/environment/test-no-416-two-generation-reproduction-toxicity_9789264070868-en

³³ http://www.oecd-ilibrary.org/environment/test-no-415-one-generation-reproduction-reproduction-toxicity-study_9789264070844-en

³⁴ http://www.oecd-ilibrary.org/environment/test-no-421-reproduction-developmental-toxicity-screening-test_9789264264380-en

³⁵ http://www.oecd-ilibrary.org/environment/test-no-422-combined-repeated-dose-toxicity-study-with-the-reproduction-developmental-toxicity-screening-test_9789264264403-en

³⁶ http://www.oecd-ilibrary.org/environment/test-no-443-extended-one-generation-reproductive-toxicity-study_9789264185371-en

³⁷ http://www.oecd-ilibrary.org/environment/test-no-426-developmental-neurotoxicity-study_9789264067394-en

4.3.9 Developmental toxicity studies

The developmental toxicity test may be done as a stand-alone study, or may be part of a multigeneration reproduction study. If it is combined with a reproduction study, assessment of teratological effects may be performed on either the first or second generation, but it is usually performed on the last litter of the generation to maximize exposure to the test agent. As part of a multigeneration study, the fetuses may be exposed to the test substance from conception. In a stand-alone study, treatment must begin early enough to include organogenesis for the species used and should continue to the day prior to the expected day of parturition. This guideline may be used with substances given orally to the rat, mouse, hamster, or rabbit. If the test substance is believed to have the capacity to alter the rate of its own metabolism through induction of metabolizing enzymes or as a result of damage incurred by the liver, then consideration should be given to evaluating the teratogenic potential of the compound by using a separate study.

There is ongoing discussion on the shortcomings of developmental and reproductive toxicity testing (Makris et al., 2011; Basketter et al., 2012). The second species could possibly be replaced by zebrafish tests (ongoing validation via systematic review by the Evidence-Based Toxicology Collaboration³⁸) or validated alternative methods such as the embryonic stem cell test (EST), the limb bud assay or whole rat embryo culture (Genschow et al., 2004; Piersma et al., 2004; Spielmann et al., 2004, 2006), but none of these methods has yet been accepted for this purpose. Notably, a variant of the EST using either human embryonic stem cells or human induced pluripotent stem cells and metabolite measurements, which were identified by metabolomics, was introduced by Stemina (Madison, WI).

The FDA refers to the 1993 draft Redbook as guidance³⁹. The different OECD guidelines under Reproduction Toxicity all include developmental effects. As a standalone test, TG 422 is preferred, or TG 421 can be considered if repeat dose toxicity is addressed separately. Possibilities to substitute the second species with new approach methods should be considered in a weight-of-evidence evaluation. However, no such alternative methods for this test or the substitution of a second species have been validated by ICCVAM/NICEATM and/or accepted by US agencies.

4.3.10 Metabolism and pharmacokinetic studies

Results from animal toxicity studies are used by FDA to determine dose-response characteristics for any effects observed in the evaluation of the safety of food and color additives. Since the delivered dose of a substance to any affected tissue or organ is determined by the pharmacokinetics and metabolism of the substance in the test animal, toxicity studies are more easily interpreted, likely to achieve target doses, and avoid excessive toxicity if data from metabolic and pharmacokinetic studies are available during the planning of short-term, subchronic and/or chronic toxicity studies. Early determination of meta-

bolic pathways and the rates of metabolism in different test species may provide explanations for species differences in any effects that are observed, and suggest biochemical or pharmacologic experiments which might be used to test explanations of such phenomena. The FDA recommends that petitioners submit data that will enable their scientists to evaluate: 1) the extent of absorption, 2) tissue distribution, 3) pathways and rates of metabolism, and 4) rate(s) of elimination of the parent substance and any metabolites formed for all Concern Level II and III substances. The FDA may recommend submission of additional metabolic and pharmacokinetic data based on the extent to which a chemical is metabolized, the potential toxicity of the metabolites, and the extent to which observed toxic effects seem to correspond to the presence of the parent substance or its metabolites.

Advances in the field have been recently summarized (Tsaïoun et al., 2016). ADME (absorption, distribution, metabolism, elimination) has rapidly evolved over the past two decades, creating a unique interdisciplinary interface between medicinal chemists, biologists, formulators, toxicologists, clinicians, and regulators across industries, but has advanced most rapidly in the pharmaceutical industry. The implementation of ADME profiling of drug candidates, in conjunction with biological efficacy and safety optimization, has dramatically reduced pharmacokinetic drug failures in clinical trials and has become a *lingua franca* between disciplines that are involved in drug development. Recently, a roadmap to integrate *in silico* and *in vitro* approaches in the area of safety assessment of food and food ingredients was published (Blaauboer et al., 2016), in which the emphasis was also on the kinetic behavior of the compounds under study. Progress in developing the most relevant *in vitro* models for intestinal uptake has been recently summarized (Gordon et al., 2015): “*Despite their varying complexity, non-animal models of the intestine described so far still remain overly simplified compared to the in vivo situation, as they do not include feedback mechanisms or aspects of the gut microbiota, immune system, innervation or specific hormonal controls. Further efforts and technological innovations are therefore needed to further improve and exploit intestinal models in order to meet the growing requirements of industry and researchers.*”

It is recommended to generate ADME information in order to help the interpretation of findings also at low concern levels. The FDA refers to the 1993 draft Redbook for guidance³⁹, which is somewhat outdated. However, as no OECD test guidelines are available, pertinent guidance from the drug and pesticide sector might be consulted or, especially for the non-mandatory generation of such information for low concern level substances, new approach methods should be considered.

For the future, the reader is referred to the roadmap developed for ADME replacement methods (Basketter et al., 2012; Leist et al., 2014).

³⁸ <http://www.ebtox.org/work-groups/zebrafish-group/index.html>

³⁹ <https://www.fda.gov/downloads/Food/GuidanceRegulation/UCM078741.pdf>



4.3.11 Human studies

The FDA does not require petitioners to conduct human clinical studies to support the safety of direct food additives or color additives used in food. However, petitioners may elect to perform such studies in certain circumstances, such as when the proposed additive will be consumed by humans at relatively high levels.

Guidelines from the 1993 draft Redbook are available⁴⁰. This evaluation cannot advise on such human studies, which will have to be designed case-by-case.

5 Future perspectives

5.1 Integrated testing strategies

Despite the fact that toxicology uses many stand-alone tests, a systematic combination of several information sources very often is required. Examples include: when not all possible outcomes of interest (e.g., modes of action), classes of test substances (applicability domains), or severity classes of effects are covered in a single test; when the positive test result is rare (low prevalence leading to excessive false positive results); when the gold standard test is too costly or uses too many animals, creating a need for prioritization by screening. Similarly, tests are combined when the human predictivity of a single test is not satisfactory or when existing data and evidence from various tests will be integrated. Increasingly, kinetic information also will be integrated to make an *in vivo* extrapolation from *in vitro* data.

ITS, also called IATA in the OECD context, offer the solution to these problems. ITS have been discussed for more than a decade (for summary see Hartung et al., 2013; Rovida et al., 2015), and some attempts have been made to develop test guidance for regulations. Despite their obvious potential for revamping regulatory toxicology, however, we still have little guidance on the composition, validation, and adaptation of ITS for different purposes. Similarly, weight of evidence and evidence-based toxicology approaches require different pieces of evidence and test data to be weighed and combined. Currently, the field of skin sensitization is pioneering the regulatory use of ITS/IATA, which is however not an information requirement under FDA GRAS.

While no ITS/IATA relevant to food additives have yet been accepted, it is advisable to consider them when developing a test strategy for a given information need.

5.2 Evidence integration

Evidence integration for a given information need (as well as the overall risk assessment) should be carried out by independent experts; notably, the lack of independence of the evalua-

tion was a major criticism to the reality of the GRAS process (Neltner, 2013b; Maffini et al., 2011, 2013). Evidence-based toxicology (EBT) (Hoffmann and Hartung, 2006; Hartung, 2009) and the EBT Collaboration⁴¹ are spearheading these efforts. Notably, both EFSA and FDA CFSAN representatives are members of the board of trustees. A joint workshop of EBTC and EFSA on evidence integration was held in October 2017 (report in preparation).

The use of all available data and the encouragement to use alternative methods to minimize animal use and suffering prompts the need for integrating various forms of evidence. Evidence integration is understood to be the process of combining different kinds of evidence relevant to hazard identification. This process is often referred to as weight of evidence (WoE) evaluations. Linkov et al. (2015) summarize “*Weighing available evidence in the process of decision-making is unavoidable, yet it is one step that routinely raises suspicions: what evidence should be used, how much does it weigh, and whose thumb may be tipping the scales? ... The inexact science of converting existing environmental health and toxicology knowledge into risk management decisions and policy relies on a growing volume of increasingly diverse scientific data. Past work in this area was guided by a small number of experimental techniques and models. Today, however, toxicity data is much more diverse. It can be collected by different modes, compounded with experiments or models, and often point in different directions regarding the same assessment endpoint. Individual lines of evidence constitute an information base from which conclusions must be drawn regarding public health and economic development. Weight of evidence (WoE) is an approach that, by means of qualitative or quantitative methods, integrates individual lines of evidence to form a conclusion (Linkov et al., 2009) and has been widely used in both ecological and human health risk assessments to collate heterogeneous information and justify selection of regulatory benchmarks.*” Notably, the National Research Council (NRC) review of the EPA’s Integrated Risk Information System (IRIS) process⁴² concluded that WoE “*has become too vague and is of little scientific use*” (NRC, 2014). In its place, the NRC proposes alternative methods of varying quantitative nature such as read-across and systematic review. Linkov et al. (2015) concur, stating: “*The present state of WoE, and evidence integration in general, is such that approaches to hazard assessments in toxicology differ greatly between applications, among both qualitative and quantitative assessments. ... WoE in the traditional Bayesian or MCDA [multi criteria decision analysis] style offers an opportunity to redirect evidence integration in toxicological assessments onto a more quantitative, transparent and objective path.*” Making the WoE evaluation as quantitative as possible is recommended.

⁴⁰ <https://www.fda.gov/downloads/Food/GuidanceRegulation/UCM078753.pdf>

⁴¹ <http://www.ebttox.org>

⁴² <https://www.ncbi.nlm.nih.gov/books/NBK230072/>

The National Toxicology Program Office of Health Assessment and Translation (OHAT) carries out literature-based evaluations about potential human health hazards to examine the state of the science (Rooney et al., 2014). In 2012, OHAT began exploring and developing an approach for implementation of systematic review methodology to define the OHAT Evaluation Process⁴³ with opportunities for external scientific, public, and interagency input. This includes the until now only formal approach for evidence integration from different data streams; though not developed or evaluated for GRAS evaluation, it can serve as guidance for such tasks.

5.3 Revising earlier GRAS evaluations

The ruling by FDA (2016) represents a clarification, not a change, in the safety standard of a food additive, be it evaluated by GRAS or petitioning the FDA. The critical aspect is the public availability of data. To quote the FDA guidance for industry⁴⁴: “*The difference between the criteria for eligibility for classification as GRAS through scientific procedures (21 CFR 170.30(b) and 21 CFR 570.30(b)) and FDA’s approval of a food additive (21 CFR 171.1 and 21 CFR 571.1) relates to who has access to the data and information and who has reviewed those data and information. For a substance to be GRAS under the conditions of its intended use, the data and information relied on to establish the safety of the use of the substance must be generally available (e.g., through publication in the scientific literature) and there must be a basis for a person to conclude that the substance is generally recognized, among qualified experts, to be safe under the conditions of its intended use. In contrast, for FDA’s approval of a food additive privately held data and information about the substance under the conditions of its intended use are sent by the sponsor to FDA and FDA reviews those data and information to determine whether they demonstrate that the substance is safe under the conditions of its intended use.*”

Earlier studies have identified deficits in this public availability (Neltner et al., 2013a), as discussed above. However, as most findings will be negative, such publication in the scientific literature is often difficult. The creation of a database to make the summary results publicly available is therefore recommended.

In adapting the ruling, GMA code of conduct and the standard to be developed, companies are urged to go through a process of sharing the respective documentation on earlier GRAS evaluations, which should include a review of:

- the completeness of the evaluation;
- possible changes in use scenarios (exposure) and consequently concern levels – note that classification as GRAS is

- for the use of a substance, rather than the substance itself;
- new information about the substance of interest impacting on the assessment;
- whether metabolites, degradation products and impurities have been adequately addressed.

Where the information is found to be incomplete or requiring expansion, the guidance for safety assessments should be applied. Given, however, that additional use experience for the substance is now available, this would allow a weight-of-evidence evaluation of all available information first and, in many instances, these use experiences combined with *in silico* and *in vitro* screening tools can suffice.

The dynamic of substance uses and toxicological information becoming available is strongly accelerating. Kruger (2016) remarks: “*There are important examples of things used in our food supply like processing aids or flavoring ingredients, where original approvals may have been very narrow and innovation has expanded uses. Re-evaluation and documentation of the safety for the expansion of these uses is critical. Reassessment of safety is very important and amendments incorporating reevaluations of GRAS ingredients are a key step in assuring the safety of our food supply. There should have [sic] more discussion and additional guidance in the framework of GRAS evaluations on when and where and how to go back through what we have in the food supply to make sure that current uses are covered by the risk assessments that were done*”. It is therefore recommended to review any GRAS determination every 10 years in order to confirm exposure estimates, concern levels, and safety of the substance.

6 Suggested process of carrying out a GRAS evaluation

Step 1: Determine the GRAS eligibility of the substance:

- Is there safe use information before January 1, 1958?
- Is the substance a color additive (which does not fall under FDA’s GRAS definition)?
- Has the substance already undergone adequate GRAS evaluations⁴⁵ or other risk assessments for food additives?
- What is the use scenario and resulting exposure⁴⁶?
- Determine the concern level (and thereby the information needs), favorably by (Q)SAR or automated read-across.

Step 2: Collect all available information on the substance separately for every information need, including:

- Is the substance listed on any positive or negative list for the use as food additive?
- Is there available toxicological data and exposure informa-

⁴³ <https://ntp.niehs.nih.gov/pubhealth/hat/review/index-2.html>

⁴⁴ <https://www.fda.gov/downloads/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/UCM525233.pdf>

⁴⁵ CFSAN’s Web site entitled “Generally Recognized as Safe (GRAS),” and CVM’s Web site entitled “Generally Recognized as Safe (GRAS) Notification Program,” each contain a list of substances that have been the subject of a GRAS notice to FDA, whether to CFSAN (for intended use in human food) or to CVM (for intended use in animal food).

⁴⁶ FDA, “Guidance for Industry: Estimating Dietary Intake of Substances in Food,” 2006; <http://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/ucm074725.htm>



tion (including other industries and other geographical regions) favorably by systematic review?

- Assess the quality of studies.
- Review the exposure and toxicological information to possibly refine the concern level
- Carry out a weight of evidence assessment to evaluate the completeness of available information; if the information is complete and unambiguous to allow satisfying the information need, consider formal evidence integration.

Step 3: Consider a TTC approach:

- Based on the exposure scenario and endpoint-specific TTC values for the class of chemicals, assess whether thresholds of regulatory concern are reached.

Step 4: Develop a test strategy, including:

- Can the available information be complemented with new approach methods (*in silico* and *in vitro*) to possibly conclude on the safety of the substance for the given information need?
- Identify the adequate study/studies in compliance with FDA guidance and the standard to be developed and justify deviations.
- Identify the sequence of information needs to be assessed – as food additive regulation follows a no-risk paradigm, any finding of toxicity relevant to the human use scenario will likely end the GRAS evaluation, thus it is advisable to address areas of concern early on.

Step 5: Carry out the test strategy and the respective risk assessment:

- Acquire the information and evaluate the risk by information need.
- If no individual findings prohibit use, carry out the overall risk assessment.
- The assessment should be done by an independent panel, i.e., a group of knowledgeable individuals with no conflict of interest for the given substance's use (note the 2017 FDA guidance on this).

Step 6: Consider mixture effects, sensitive subpopulations and extreme use scenarios:

- Address possible interactions with the food matrix as well as direct and indirect food additives, possibly even other foods.
- Consider possible sensitive subpopulations (infants, elderly, diseased populations, e.g., suffering from liver or kidney disease).
- Consider extreme uses.

Step 7: Evaluate metabolites, degradation products and impurities:

- Making use to the extent possible of (Q)SAR, automated read-across and TTC evaluate metabolites, degradation products and impurities addressing all information needs for the food additive itself; please note that the concern level for metabolites, degradation products and impurities can be higher.

Step 8: Document the process and share the results with FDA and the general public:

- The GRAS notification procedure is a voluntary procedure under which any person may notify FDA of a conclusion that a substance is GRAS under the conditions of its intended use in human food (21 CFR part 170, subpart E) or animal food (21 CFR part 570, subpart E). Although the GRAS notification procedure was voluntary, FDA strongly encouraged any person who intends to market a food substance on the basis of the GRAS provision to submit a GRAS notice to FDA⁴⁷. With the 2017 guidance, it is required to notify FDA.
- As GRAS requires public availability of the underlying data, safeguard the publication of all studies at least as robust summaries.
- The FDA operates an Adverse Reaction Monitoring System (ARMS) to help serve as an ongoing safety check of all additives, monitoring and investigating all complaints by individuals or their physicians that are believed to be related to specific foods, food and color additives, or vitamin and mineral supplements. The ARMS computerized database helps officials decide whether reported adverse reactions represent a real public health hazard associated with food, so that appropriate action can be taken. Companies should report such complaints to FDA ARMS.

7 Where should a standard go beyond current FDA guidance?

The value of a standard to be developed lies in where it goes beyond the law and FDA ruling. By doing so, it aims at providing the consistency, scientific rigor and transparency to earn consumer trust.

- The standard should be committed to both sound and the most humane science. It should embrace the developments of evidence-based toxicology and the 3Rs.
- It should recommend the use of all available data, favorably by systematic review, quality scoring and evidence integration within a (quantitative) weight-of-evidence evaluation.
- It should recommend the use of Thresholds of Toxicological Concern (TTC) for low concentration additives, metabolites, degradation products and impurities.
- It should recommend the use of (automated) read-across and Q(SAR) to objectively determine concern levels.
- It should recommend developing ITS for the execution of testing.
- It should recommend OECD test guidelines and their updates, going beyond the Redbook 2000 guidance still in use by FDA. This includes the use of test guidelines for combined chronic/carcinogenicity, subchronic/mutagenicity and chronic/developmental screening tests, as well as the use of the extended-one-generation reproductive toxicity study.
- It should recommend generating toxicokinetic information (including *in vitro* and *in silico* approaches) also for low concern level additives.

⁴⁷ <https://www.fda.gov/ForIndustry/FDAeSubmitter/default.htm>

- It should discourage the use of the one-year non-rodent study, the use of mouse as a second species in cancer bioassays, and the cancer bioassay as a whole based on increasing concern about their regulatory value and lacking human relevance.
- It should encourage the substitution of the second species (where required) for developmental toxicity with embryonic stem cell, zebrafish developmental toxicity or whole rat embryo culture tests.

The following requirements are already part of the FDA 2017 guidance:

- The independence of the risk assessors (GRAS panels) from the manufacturer/user of the substance of interest.
- The urge to review earlier GRAS determinations and requirement; going beyond this, the standard should include provisions to regularly (e.g., every 10 years) revise the determination.
- Making GRAS notifications to FDA mandatory.

8 Conclusions

The goal of all these efforts is to make the safest food available to customers. The lay public has a love/hate relationship to manufactured food. This is exemplified by two of my favorite authors: George Orwell, who stated “*We may find in the long run that tinned food is a deadlier weapon than the machine-gun*”, and George Bernard Shaw, who wrote “*There is no sincerer love than the love of food*”. The distinction made and implicit in the quotes is often between the “bad” manufactured versus the “good” home-made food. This is not always justified, especially when thinking of microbiological risks, which are typically well controlled in manufacturing but challenged at home when storing, processing and consuming food. Similarly, large manufacturers can control the toxicity of ingredients and contaminants (i.e., direct and indirect food additives) to an extent not possible in an individual household. However, the lack of control over what is added instills fear when it is outside of consumer control. The fact that we appear to lack data about the safety of these compounds adds to these fears. This strongly contrasts with the relatively few scandals and identified risks. An investment into the safety assessment of chemicals and its communication thus is an important means to regain consumer trust. People want to know – or as Dan Glickman, former US Secretary of Agriculture, said “*One of the trends we’re seeing in food and agriculture is more and more consumers wanting to know things about their food and where and how it’s grown and what’s in it*”.

US food has an especially bad reputation, to which the perceived easy-go of GRAS has added. The American author Fran Lebowitz (1950-) cynically said: “*If you’re going to America, bring your own food*”. Even a former Commissioner of the FDA (1990 to 1997), David Aaron Kessler, is quoted for “*The history of government regulation of food safety is one of government watchdogs chasing the horse after it’s out of the barn*”. With GRAS as a key mechanism to pragmatically

regulate food additives, the current reboot represents an opportunity to create a paradigm, in which the entire food chain collaborates in order to produce food of a safety standard consumers can trust.

The efforts detailed here of FDA and manufacturers show how a system that is quite rigorous in principle shall be amended to serve these functions, while not giving up on the elements of self-regulation of the industry as the agency could hardly handle the enormous number of otherwise necessary petitions and risk assessments. It will require from the industry a more stringent and transparent application of the GRAS process. These changes represent at the same time an opportunity to bring testing requirements up to date and thereby consider new approaches. Such discussion is complex, but promises, in the end, advantages for all parties involved including, ultimately, the consumer.

References

- Aardema, M. J., Barnett, B. C., Khambatta, Z. et al. (2010). International prevalidation studies of the EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay: Transferability and reproducibility. *Mutat Res* 701, 123-131. doi:10.1016/j.mrgentox.2010.05.017
- Adler, S., Basketter, D., Creton, S. et al. (2011). Alternative (non-animal) methods for cosmetics testing: Current status and future prospects. *Arch Toxicol* 85, 367-485. doi:10.1007/s00204-011-0693-2
- Akagi, J.-I., Toyoda, T., Cho, Y.-M. et al. (2015). Validation study of the combined repeated-dose toxicity and genotoxicity assay using gpt delta rats. *Cancer Sci* 106, 529-541. doi:10.1111/cas.12634
- Alden, C. L., Lynn, A., Bourdeau, A. et al. (2011). A critical review of the effectiveness of rodent pharmaceutical carcinogenesis testing in predicting for human risk. *Vet Pathol* 48, 772-784. doi:10.1177/0300985811400445
- 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
- Appleman, L. M. and Feron, V. J. (1986). Significance of the dog as “second animal species” in toxicity testing for establishing the lowest “no-toxic-effect-level”. *J Appl Toxicol* 6, 271-279. doi:10.1002/jat.2550060408
- 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
- Barraj, L., Murphy, M., Tran, N. and Petersen, B. (2016). Chemistry, manufacturing and exposure assessments to support generally recognized as safe (GRAS) determinations. *Regul Toxicol Pharmacol* 79, S99-S104. doi:10.1016/j.yrtph.2016.07.003
- Basketter, D. A., Clewell, H., Kimber, I. et al. (2012). A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing. *ALTEX* 29, 3-89. doi:10.14573/altex.2012.1.003



- Blaauboer, B. J., Boobis, A. R., Bradford, B. et al. (2016). Considering new methodologies in strategies for safety assessment of foods and food ingredients. *Food Chem Toxicol* 91, 19-35. doi:10.1016/j.fct.2016.02.019
- Box, R. J. and Spielmann, H. (2005). Use of the dog as non-rodent test species in the safety testing schedule associated with the registration of crop and plant protection products (pesticides): Present status. *Arch Toxicol* 79, 615-626. doi:10.1007/s00204-005-0678-0
- Bremer, S., Pellizzer, C., Hoffmann, S. et al. (2007). The development of new concepts for assessing reproductive toxicity applicable to large scale toxicological programs. *Curr Pharm Des* 13, 3047-3058. doi:10.2174/138161207782110462
- Burdock, G. A. and Carabin, I. G. (2004). Generally recognized as safe (GRAS): History and description. *Toxicol Lett* 150, 3-18. doi:10.1016/j.toxlet.2003.07.004
- Casey, W., Jacobs, A., Maull, E. et al. (2015). A new path forward: The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and National Toxicology Program's Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). *J Am Assoc Lab Anim Sci* 54, 170-173.
- Combes, R., Balls, M., Curren, R. et al. (1999). Cell transformation assay as predictors of human carcinogenicity. *Altern Lab Anim* 27, 745-767.
- Congress (1958). 85th U.S. Congress second session. H.R. report no. 2284. Congressional record Washington, DC.
- Contrera, J. F., Aub, D., Barbehenn, E. et al. (1993). A retrospective comparison of the results of 6 and 12 month non-rodent toxicity studies. *Adverse Drug React Toxicol Rev* 12, 63-76.
- Cooper, R. L., Lamb, J. C., Barlow, S. M. et al. (2006). A tiered approach to life stages testing for agricultural chemical safety assessment. *Crit Rev Toxicol* 36, 69-98. doi:10.1080/10408440500541367
- Corvi, R., Albertini, S., Hartung, T. et al. (2008). ECVAM retrospective validation of in vitro micronucleus test (MNT). *Mutagenesis* 23, 271-283. doi:10.1093/mutage/gen010
- Corvi, R. and Vanparys, P., guest editors (2012). Special issue: International prevalidation study on cell transformation assays. *Mutat Res* 744, 1-116. doi:10.1016/j.mrgentox.2012.02.004
- Degnan, F. H. (2006). *FDA's Creative Application of the Law: Not Merely a Collection of Words*. 2nd edition. Washington, DC, USA: FDLI.
- Dellarco, V. L., Rowland, J. and May, B. (2010). A retrospective analysis of toxicity studies in dogs and impact on the chronic reference dose for conventional pesticide chemicals. *Crit Rev Toxicol* 40, 16-23. doi:10.3109/10408440903401529
- Doe, J. E., Boobis, A. R., Blacker, A. et al. (2006). A tiered approach to systemic toxicity testing for agricultural chemical safety assessment. *Crit Rev Toxicol* 36, 37-68. doi:10.1080/10408440500534370
- EC (2008). Commission Regulation (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). *OJL* 142, 1-1900. <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02008R0440-20160304>
- ECVAM (2012). Recommendation concerning the cell transformation assays using Syrian hamster embryo cells (SHE) and the BALB/c 3T3 mouse fibroblast cell line for in vitro carcinogenicity testing. Annex I: ESAC opinion on the ESAC peer review of an ECVAM-coordinated prevalidation study concerning three protocols of the cell transformation assay (CTA) for in vitro carcinogenicity testing. *Carcinogenicity. Altern Lab Anim* 33, Suppl 1, 177-182.
- EFSA Scientific Committee (2012). Scientific opinion on exploring options for providing advice about possible human health risks based on the concept of Threshold of Toxicological Concern (TTC). *EFSA J* 10, 2750. doi:10.2903/j.efsa.2012.2750
- EFSA ANS Panel – EFSA Panel on Food Additives and Nutrient Sources Added to Food (2014). Statement on a conceptual framework for the risk assessment of certain food additives re-evaluated under Commission Regulation (EU) No 257/2010. *EFSA J* 12, 3697. doi:10.2903/j.efsa.2014.3697
- Eskes, C., Boström, A.-C., Bowe, G. et al. (2017). Good cell culture practices & in vitro toxicology. *Toxicol In Vitro* 45, 272-277. doi:10.1016/j.tiv.2017.04.022
- FDA (1982). Toxicological principles for the safety assessment of direct food additives and color additives used in food. U.S. Food and Drug Administration. U.S. Department of Commerce. National Technical Information Service. Springfield, VA.
- FDA (1995). Federal Register Food Additives; Threshold of Regulation for Substances Used in Food Contact Articles. Department of Health and Human Services, Food and Drug Administration, 21 CFR Parts 5, 25, 170,171, and 174. Docket Nos 77P-0122, 92N-0181.
- FDA (2001). 21 CFR, Ch.I, Subpart B. Food Additive Safety, 170.20, 170.22, 170.30.
- FDA (2012). International Conference on Harmonisation; guidance on S2(R1) Genotoxicity Testing and Data Interpretation for Pharmaceuticals intended for Human Use; availability Notice. Fed Regist (United States), Jun 07 2012, 77(110) p33748-9.
- FDA (2016). Substances Generally Recognized as Safe – A Rule by the Food and Drug Administration on 08/17/2016. 81 Fed Regist (United States) 54959. <https://www.gpo.gov/fdsys/pkg/FR-2016-08-17/pdf/2016-19164.pdf>
- Fowler, P. P., Smith, K. K., Young, J. J. et al. (2012). Reduction of misleading (“false”) positive results in mammalian cell genotoxicity assays. I. Choice of cell type. *Mutat Res* 742, 11-25. doi:10.1016/j.mrgentox.2011.10.014
- Genschow, E., Spielmann, H., Scholz, G. et al. (2004). Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Altern Lab Anim* 32, 209-244.
- Gerbracht, U. and Spielmann, H. (1998). The use of dogs as second species in regulatory testing of pesticides. I. Inter-

- species comparison. *Arch Toxicol* 72, 319-329. doi:10.1007/s002040050509
- Gold, L. S., Sawyer, C. B., Magaw, R. et al. (1984). A carcinogenesis potency database of the standardized results of animal bioassays. *Environ Health Perspect* 58, 9-319. doi:10.1289/ehp.84589
- Gold, L. S., Slone, T. H. and Bernstein, L. (1989). Summary of carcinogenic potency and positivity for 492 rodent carcinogens in the carcinogenic potency database. *Environ Health Perspect* 79, 259-272. doi:10.1289/ehp.8979259
- 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
- Hallagan, J. B. and Hall, R. L. (2009). Under the conditions of intended use – New developments in the FEMA GRAS program and the safety assessment of flavor ingredients. *Food Chem Toxicol* 47, 267-278. doi:10.1016/j.fct.2008.11.011
- Hamburg, M. A. (2011). Advancing regulatory science. *Science* 331, 987-987. doi:10.1126/science.1204432
- Hartung, T. (2009). Food for thought... on evidence-based toxicology. *ALTEX* 26, 75-82. doi:10.14573/altex.2009.2.75
- Hartung, T. and Koëter, H. (2010). Food for thought... on alternative methods for food safety testing. *ALTEX* 25, 259-264. doi:10.14573/altex.2010.2.87
- Hartung, T., Blaauboer, G. J., Bosgra, S. et al. (2011). An expert consortium review of the EC-commissioned report “alternative (non-animal) methods for cosmetics testing: Current status and future prospects – 2010”. *ALTEX* 28, 183-209. doi:10.14573/altex.2011.3.183
- 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. (2017). Opinion versus evidence for the need to move away from animal testing. *ALTEX* 34, 193-200. doi:10.14573/altex.1703291
- 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
- Hartung, T., FitzGerald, R., Jennings, P. et al. (2017). Systems toxicology – Real world applications and opportunities. *Chem Res Toxicol* 30, 870-882. doi:10.1021/acs.chemrestox.7b00003
- Hennes, E. C. (2012). An overview of values for the threshold of toxicological concern. *Toxicol Lett* 211, 296-303. doi:10.1016/j.toxlet.2012.03.795
- Hoffmann, S. and Hartung, T. (2006). Toward an evidence-based toxicology. *Human Exp Toxicol* 25, 497-513. doi:10.1191/0960327106het6480a
- Hoffmann, S., de Vries, R. B. M., Stephens, M. L. et al. (2017). A primer on systematic reviews in toxicology. *Arch Toxicol* 91, 2551-2575. doi:10.1007/s00204-017-1980-3
- Hutt, P. B., Merrill, R. A. and Grossman, L. A. (1991). *Food and Drug Law, Cases and Materials* (1354 pp). 2nd edition. Westbury, NY, USA: University Casebook Series, Foundation Press.
- Jacobson-Kram, D. and Jacobs, A. (2005). Use of genotoxicity data to support clinical trials or positive genotox findings on a candidate pharmaceutical or impurity ... now what? *Int J Toxicol* 24, 129-134. doi:10.1080/10915810590952933
- JECFA (1993). Evaluation of certain food additives and contaminants. Safety evaluation of flavouring agents. Forty-first report of the Joint FAO/WHO Expert Committee on Food Additives. WHO Technical Report Series 837. World Health Organisation, Geneva, Switzerland.
- JECFA (1995). Evaluation of certain food additives and contaminants. Safety evaluation of flavouring agents. Forty-fourth report of the Joint FAO/WHO Expert Committee on Food Additives. WHO Technical Report Series 859. World Health Organisation, Geneva, Switzerland.
- JECFA (1997). Evaluation of certain food additives and contaminants (Forty-sixth report of the Joint FAO/WHO Expert Committee on Food Additives). WHO Technical Report Series 868, 1997. World Health Organisation, Geneva, Switzerland.
- JECFA (1999). Evaluation of certain food additives and contaminants. Procedure for the safety evaluation of flavouring agents. Forty-ninth report of the Joint FAO/WHO Expert Committee on Food Additives. WHO Technical Report Series 884. World Health Organisation, Geneva, Switzerland.
- Karmaus, A. L., Filer, D. L., Martin, M. T. and Houck, K. A. (2016). Evaluation of food-relevant chemicals in the ToxCast high-throughput screening program. *Food Chem Toxicol* 92, 188-196. doi:10.1016/j.fct.2016.04.012
- Kirkland, D., Pfuhler, S., Tweats, D. et al. (2007). How to reduce false positive results when undertaking in vitro genotoxicity testing and thus avoid unnecessary follow-up animal tests: Report of an ECVAM Workshop. *Mutat Res* 628, 31-55. doi:10.1016/j.mrgentox.2006.11.008
- Kobel, W., Fegert, I., Billington, R. et al. (2010). A 1-year toxicity study in dogs is no longer a scientifically justifiable core data requirement for the safety assessment of pesticides. *Crit Rev Toxicol* 40, 1-15. doi:10.3109/10408440903300098
- Kobel, W., Fegert, I., Billington, R. et al. (2014). Relevance of the 1-year dog study in assessing human health risks for registration of pesticides. An update to include pesticides registered in Japan. *Crit Rev Toxicol* 44, 842-848. doi:10.3109/1040844.2014.936550
- 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
- Kruger, C. (2016). The relevance of international assessments to GRAS determinations. *Regul Toxicol Pharmacol* 79, Suppl 2, S119-S123. doi:10.1016/j.yrtph.2016.06.010
- Leist, M., Hasiwa, N., Rovida, C. et al. (2014). Consensus report on the future of animal-free systemic toxicity testing. *ALTEX* 31, 341-356. doi:10.14573/altex.1406091
- Linke, B., Mohr, S., Ramsingh, D. and Bhuller, Y. (2017). A retrospective analysis of the added value of 1-year dog studies in pesticide human health risk assessments. *Crit Rev Toxicol* 47, 581-591. doi:10.1080/10408444.2017.1290044
- Linkov, I., Loney, D., Cormier, S. et al. (2009). Weight-of-ev-



- idence evaluation in environmental assessment: Review of qualitative and quantitative approaches. *Sci Total Environ* 407, 5199-5205. doi:10.1016/j.scitotenv.2009.05.004
- Linkov, I., Massey, O., Keisler, J. et al. (2015). From “weight of evidence” to quantitative data integration using multicriteria decision analysis and Bayesian methods. *ALTEX* 32, 3-8. doi:10.14573/altex.1412231
- Luechtefeld, T., Maertens, A., Russo, D. P. et al. (2016). Analysis of public oral toxicity data from REACH registrations 2008-2014. *ALTEX* 33, 111-122. doi:10.14573/altex.1510054
- Luechtefeld, T. and Hartung, T. (2017). Computational approaches to chemical hazard assessment. *ALTEX* 34, 459-478. doi:10.14573/altex.1710141
- Luijten, M., Olthof, E. D., Hakkert, B. C. et al. (2016). An integrative test strategy for cancer hazard identification. *Crit Rev Toxicol* 46, 615-639. doi:10.3109/10408444.2016.1171294
- Maffini, M. V., Alger, H. M., Bongard, E. D. and Neltner, T. G. (2011). Enhancing FDA’s evaluation of science to ensure chemicals added to human food are safe: Workshop proceedings. *Compr Rev Food Sci Food Saf* 10, 321-341. doi:10.1111/j.1541-4337.2011.00165.x
- Maffini, M. V., Alger, H. M., Olson, E. D. and Neltner, T. G. (2013). Looking back to look forward: A review of FDA’s food additives safety assessment and recommendations for modernizing its program. *Compr Rev Food Sci Food Saf* 12, 439-453. doi:10.1111/1541-4337.12020
- Makris, S. L., Kim, J. H., Ellis, A. et al. (2011). Current and future needs for developmental toxicity testing. *Birth Defects Res B Dev Reprod Toxicol* 92, 384-394. doi:10.1002/bdrb.20335
- 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
- Maurici, D., Aardema, M., Corvi, R. et al. (2005). Genotoxicity and mutagenicity. *Altern Lab Anim* 33, Suppl 1, 117-130.
- Munro, I. C., Renwick, A. G. and Danielewska-Nikiel, B. (2008). The threshold of toxicological concern (TTC) in risk assessment. *Toxicol Lett* 180, 151-156. doi:10.1016/j.toxlet.2008.05.006
- Neltner, T. G., Alger, H. M., Leonard, J. E. and Maffini, M. V. (2013a). Data gaps in toxicity testing of chemicals allowed in food in the United States. *Reprod Toxicol* 42, 85-94. doi:10.1016/j.reprotox.2013.07.023
- Neltner, T. G., Alger, H. M., O’Reilly, J. T. et al. (2013b). Conflicts of interest in approvals of additives to food determined to be generally recognized as safe: Out of balance. *JAMA Intern Med* 173, 2032-2036. doi:10.1001/jamainternmed.2013.10559
- NRC – National Research Council (2014). Review of EPA’s Integrated Risk Information System (IRIS) Process. Washington, DC, USA: National Academies Press.
- OECD (2005). Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. *OECD Series on Testing and Assessment* 34. [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono\(2005\)14](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2005)14)
- OECD (2007). Detailed review paper on cell transformation assays for detection of chemical carcinogens. *Series on Testing and Assessment* 31. [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono\(2007\)18](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)18)
- OECD (2015). Guidance document on the in vitro Syrian hamster embryo (SHE) cell transformation assay. *Series on Testing & Assessment* 214. <http://www.oecd.org/env/ehs/testing/Guidance-Document-on-the-in-vitro-Syrian-Hamster-Embryo-Cell-Transformation-Assay.pdf>
- Pamies, D. and Hartung, T. (2017). 21st century cell culture for 21st century toxicology. *Chem Res Toxicol* 30, 43-52. doi:10.1021/acs.chemrestox.6b00269
- Pamies, D., Bal-Price, A., Simeonov, A. et al. (2017). Good cell culture practice for stem cells and stem-cell-derived models. *ALTEX* 34, 95-132. doi:10.14573/altex.1607121
- Pamies, D., Ball-Price, A., Chesne, C. et al. (submitted). Good cell culture practice for human primary and stem cell-derived models and organoids.
- Paparella, M., Colacci, A. and Jacobs, M. N. (2017). Uncertainties of testing methods: What do we (want to) know about carcinogenicity? *ALTEX* 34, 235-252. doi:10.14573/altex.1608281
- Parry, J. M., Parry, E., Phrakonkham, P. and Corvi, R. (2010). Analysis of published data for top concentration considerations in mammalian cell genotoxicity testing. *Mutagenesis* 25, 531-538. doi:10.1093/mutage/geq046
- Pfuhler, S., Kirkland, D., Kasper, P. et al. (2009). Reduction of use of animals in regulatory genotoxicity testing: Identification and implementation opportunities – Report from an ECVAM workshop. *Mutat Res* 680, 31-42. doi:10.1016/j.mrgentox.2009.09.002
- Pfuhler, S., Kirst, A., Aardema, M. et al. (2010). A tiered approach to the use of alternatives to animal testing for the safety assessment of cosmetics: Genotoxicity. A COLIPA analysis. *Regul Toxicol Pharmacol* 57, 315-324. doi:10.1016/j.yrtph.2010.03.012
- Piersma, A. H., Genschow, E., Verhoef, A. et al. (2004). Validation of the postimplantation rat whole-embryo culture test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Altern Lab Anim* 32, 275-307.
- 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, 223-234. doi:10.1016/j.toxlet.2003.12.034
- Reus, A. A., Reisinger, K., Downs, T. R. et al. (2013). Comet assay in reconstructed 3D human epidermal skin models – Investigation of intra- and inter-laboratory reproducibility with coded chemicals. *Mutagenesis* 28, 709-720. doi:10.1093/mutage/get051
- Roberts, A. and Houghton, L. A. (2016). A hard look at FDA’s review of GRAS notices. *Regul Toxicol Pharmacol* 79, Suppl 2, S124-S128. doi:10.1016/j.yrtph.2016.06.011
- Rooney, A. A., Boyles, A. L., Wolfe, M. S. et al. (2014). Systematic review and evidence integration for literature based environmental health science assessments. *Environ Health*

- Perspect* 122, 711-718. doi:10.1289/ehp.1307972
- Rovida, C., Alépée, N., Api, A. M. et al. (2015). Integrated testing strategies (ITS) for safety assessment. *ALTEX* 32, 171-181. doi:10.14573/altex.1506201
- Rulis, A. M. (1986). De minimis and the threshold of regulation. In C. W. Felix (ed.), *Food Protection Technology* (29-37). Lewis Publishers.
- Rulis, A. M. (1989). Establishing a threshold of regulation. In J. J. Bonin and D. E. Stevenson (eds.), *Risk Assessment in Setting National Priorities* (271-278). Plenum Publishing Corporation. doi:10.1007/978-1-4684-5682-0_28
- Rulis, A. M. and Levitt, J. A. (2009). FDA's food ingredient approval process. *Regul Toxicol Pharmacol* 53, 20-31. doi:10.1016/j.yrtph.2008.10.003
- Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. London, UK: Methuen. http://altweb.jhsph.edu/pubs/books/humane_exp/het-toc
- Samuel, G. O., Hoffmann, S., Wright, R. et al. (2016). Guidance on assessing the methodological and reporting quality of toxicologically relevant studies: A scoping review. *Environ Int* 92-93, 630-646. doi:10.1016/j.envint.2016.03.010
- Schechtman, L. M. (2002). Implementation of the 3Rs (refinement, reduction, and replacement): Validation and regulatory acceptance considerations for alternative toxicological test methods. *ILAR J* 43, Suppl, S85-S94. doi:10.1093/ilar.43.Suppl_1.S85
- 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
- Severin, I., Souton, E., Dahbi, L. and Chagnon, M.-C. (2017). Use of bioassays to assess hazard of food contact material extracts: State of the art. *Food Chem Toxicol* 105, 429-447. doi:10.1016/j.fct.2017.04.046
- Smith, M. and Rulis, A. (1981). FDA's GRAS review and priority-based assessment of food additives. *Food Technol* 1981, 71-74.
- Spielmann, H. and Gerbracht, U. (2001). The use of dogs as second species in regulatory testing of pesticides. Part II: Subacute, subchronic and chronic studies in the dog. *Arch Toxicol* 75, 1-21. doi:10.1007/s002040000195
- Spielmann, H., Genschow, E., Brown, N. A. et al. (2004). Validation of the rat limb bud micromass test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Altern Lab Anim* 32, 245-274.
- Spielmann, H., Seiler, A., Bremer, S. et al. (2006). The practical application of three validated in vitro embryotoxicity tests. The report and recommendations of an ECVAM/ZEBET workshop (ECVAM workshop 57). *Altern Lab Anim* 34, 527-538.
- Stephens, M. L., Betts, K., Beck, N. B. et al. (2016). The emergence of systematic review in toxicology. *Toxicol Sci* 152, 10-16. doi:10.1093/toxsci/kfw059
- Taylor, K., Andrew, D. J. and Rego, L. (2014). The added value of the 90-day repeated dose oral toxicity test for industrial substances with a low (sub)acute toxicity profile in a high-quality dataset. *Regul Toxicol Pharmacol* 69, 320-332. doi:10.1016/j.yrtph.2014.04.008
- Tollefsen, K. E., Scholz, S., Cronin, M. T. et al. (2014). Applying adverse outcome pathways (AOPs) to support integrated approaches to testing and assessment (IATA). *Regul Toxicol Pharmacol* 70, 629-640. doi:10.1016/j.yrtph.2014.09.009
- Tsaioun, K., Blaauuboer, B. J. and Hartung, T. (2016). Evidence-based absorption, distribution, metabolism, excretion and toxicity (ADMET) and the role of alternative methods. *ALTEX* 33, 343-358. doi:10.14573/altex.1610101
- Vanparrys, P., Corvi, R., Aardema, M. et al. (2010). ECVAM prevalidation of three cell transformation assays. *ALTEX* 27, 267-270. doi:10.14573/altex.2011.1.056
- van Ravenzwaay, B. (2010). Initiatives to decrease redundancy in animal testing of pesticides. *ALTEX* 27, 212-214. http://www.altex.ch/resources/raltex_2010_3_212_214_vanRavenzwaay.pdf
- Wang, B. and Gray, G. (2015). Concordance of noncarcinogenic endpoints in rodent chemical bioassays. *Risk Anal* 35, 1154-1166. doi:10.1111/risa.12314
- Williams, G. M., Kobets, T., Iatropoulos, M. J. et al. (2016). GRAS determination scientific procedures and possible alternatives. *Regul Toxicol Pharmacol* 79, Suppl 2, S105-S111. doi:10.1016/j.yrtph.2016.06.015
- Zhu, H., Bouhifd, M., Kleinstreuer, N. et al. (2016). Supporting read-across using biological data. *ALTEX* 33, 167-182. doi:10.14573/altex.1601252

Conflict of interest

The author consults Underwriters Laboratories (UL) on computational toxicology, especially read-across, and has a share of their respective sales.

Acknowledgements

Both FDA's Center for Food Safety and Applied Nutrition and the Grocery Manufacturer Association (GMA) are represented on the CAAT board. The author is indebted to the continuing discussion with their representatives, which helped to understand this sector and its specific needs for safety assessments. I would like to thank also our CAAT-Europe board member Dr. Didier Verloo, EFSA, for discussions and provision of background materials.

Correspondence to

Thomas Hartung, MD PhD
Johns Hopkins Bloomberg School of Public Health
Center for Alternatives to Animal Testing (CAAT)
615 N Wolfe St.
Baltimore, MD, 21205, USA
e-mail: THartun1@jhu.edu