Safety Evaluations Under the Proposed US Safe Cosmetics and Personal Care Products Act of 2013: Animal Use and Cost Estimates

Jean Knight [1] *and Costanza Rovida* [2]

[1]Environmental Consultant, Half Moon Bay, California, USA; [2]Center for Alternatives to Animal Testing (CAAT) – Europe, University of Konstanz, Germany

Summary

The proposed Safe Cosmetics and Personal Care Products Act of 2013 calls for a new evaluation program for cosmetic ingredients in the US, with the new assessments initially dependent on expanded animal testing. This paper considers possible testing scenarios under the proposed Act and estimates the number of test animals and cost under each scenario. It focuses on the impact for the first 10 years of testing, the period of greatest impact on animals and costs. The analysis suggests the first 10 years of testing under the Act could evaluate, at most, about 50% of ingredients used in cosmetics. Testing during this period would cost about $ 1.7-$ 9 billion and 1-11.5 million animals. By test year 10, alternative, high-throughput test methods under development are expected to be available, replacing animal testing and allowing rapid evaluation of all ingredients. Given the high cost in dollars and animal lives of the first 10 years for only about half of ingredients, a better choice may be to accelerate development of high-throughput methods. This would allow evaluation of 100% of cosmetic ingredients before year 10 at lower cost and without animal testing.

Keywords: Safe Cosmetics and Personal Care Products Act of 2013/H.R. 1385, animal testing, alternative test methods, cosmetics safety

1 Introduction

This article examines the proposed Safe Cosmetics and Personal Care Products Act (H.R. 1385, 2013) as it relates to cosmetic safety evaluations. In particular, it examines the potential new evaluation requirements, and it estimates the costs and animal use under those proposed new requirements. It also looks at the Act in the context of sweeping changes now under way in the area of chemical safety testing (which includes cosmetic safety testing). For decades, chemical safety testing has relied on fairly standard laboratory tests, including animal tests, but efforts are now under way to develop high-throughput *in vitro* methods combined with computer models that can rapidly and more accurately predict human response. How the timing of any new test requirements dovetails with the development of these new test methods is key to assessing potential impacts of the Act.

1.1 Proposed safety standard under the Act

In the Safe Cosmetics and Personal Care Products Act, the proposed safety standard for ingredients and finished cosmetic products is *reasonable certainty of no harm* (section 614[a][1]). The Act requires the standard to ensure "*not more than a one in a million risk of any adverse health effect*" or "*shown to produce no adverse health effects, incorporating a margin of safety of at least 1,000*" (sections 614[a][2][A] and [B]). It further defines *reasonable certainty of no harm* as "*no harm will be caused to members of the general population or any vulnerable population by aggregate exposure to the cosmetic or ingredient, taking into account possible harmful effects from – (a) low-dose exposures to the cosmetic or ingredient; (B) additive effects resulting from repeated exposure to the cosmetic or ingredient over time; or (C) cumulative exposure resulting from all sources, including both the cosmetic or ingredient and environmental sources*" (section 611[9]).

Under the proposed Act, the US Secretary of Health and Human Services ("the Secretary") would determine for each ingredient an allowable exposure that meets the safety standard. In determining this, the Secretary must consider whether the substance is "*a known or suspected neurological or immunological toxicant, respiratory asthmagen, carcinogen, teratogen, or endocrine disruptor, or have other toxicity concerns, including reproductive or developmental toxicity*" (section 616[a][2]

Received September 27, 2013; accepted in revised form January 14, 2014; Epub January 24, 2014;
http://dx.doi.org/10.14573/altex.1309271

[D]). The Act defines reproductive or developmental toxicity as "*…can contribute to biologically adverse effects on the development of humans or animals, including effects on the female or male reproductive system, the endocrine system, fertility, pregnancy, pregnancy outcomes, or modifications in other functions of the body that are dependent on the integrity of the reproductive system as well normal fetal development*" (section 611[10]).

The Health and Human Services agency designated by the Act to implement the new program is the Food and Drug Administration (FDA), which is the agency that currently oversees cosmetics safety in the US. In the US, cosmetics currently are regulated under the Federal Food, Drug, and Cosmetic Act and the Fair Packaging and Labeling Act. Neither act requires testing or pre-market approval of finished cosmetic products or cosmetic ingredients, except for color additives. UV filters and some finished products (for example, anti-perspirants) are regulated as drugs and require pre-market approval.

Although pre-market approval currently is not required for most cosmetics, manufacturers are still legally responsible for the safety of their products and the ingredients in them. Current law does not mandate specific tests to substantiate the safety of ingredients or finished products, but the FDA issued an advisory in 1975 that remains its advice today: "*the safety of a product can be adequately substantiated through (a) reliance on already available toxicological test data on individual ingredients and on product formulations that are similar in composition to the particular cosmetic, and (b) performance of any additional toxicological and other tests that are appropriate in light of such existing data and information*" (Federal Register, March 3, 1975, page 8916).

The safety standard proposed in the Act follows the traditional approach to toxicology that applies standard protocols to measure no observed adverse effect levels (NOAEL) and extrapolates to levels that presumably provide a specified certainty, by applying factors that consider individual variability, age, interspecies correlation, and other variables.

Where the proposed standard differs from current practice is its application of full toxicological evaluations to all ingredients. Traditionally for cosmetics, the NOAEL is derived only for general systemic toxicity, and normally only for ingredients intended to be biologically active, a small percentage of ingredients (see Section 2.1.1). Many ingredients have such a low toxicological profile that determining the NOAEL normally would not be considered necessary. At face value, the proposed Act could be interpreted to require NOAEL for carcinogenicity and reproductive toxicity, as well as for systemic toxicity. However, a NOAEL does not apply for carcinogenicity or reproductive toxicity for cosmetics, because no level of these risks is considered acceptable in a cosmetic. If any such risk is detected, the substance is not used as a cosmetic ingredient.

1.2 Potential for animal testing under the proposed Act

The proposed Act, if passed in the next couple years, would involve increased animal testing for cosmetic ingredients. This is because it would require more long-term safety evaluations, which are animal methods, than under current recommended practice. Current recommended practice is described in Section 2.1.1, and evaluations under the proposed Act are described in Sections 3.1 and 3.2. The Act (section 624, "Animal Testing Alternatives") includes language that requires alternative test methods "*where practicable*," but alternative test methods for long term studies are not expected to be available for about 10 years (see Section 3.7).

In contrast with the proposed US bill, the new European Union (EU) Cosmetics Regulation, Regulation EC 1223/2009, mandates the use of alternative methods for the toxicological assessment of new cosmetic ingredients (EC, 2009a). Article 18(1a, b) of that Regulation prohibits marketing cosmetics products containing ingredients that "*have been subjected to animal testing using a method other than an alternative method*." This animal testing ban was implemented in three phases that were completed in 2013.

The ultimate impact of the proposed US Act on animals depends on the proposed Act's timeline (see Section 3.7) and when this timeline intersects the final development of non-animal methods, including high-throughput test methods. The Act specifies a timeline for evaluating cosmetic ingredients that would stretch over decades if today's methods are used. This is because of the sheer number of cosmetic ingredients used in finished products: thousands of existing cosmetic ingredients, with hundreds of new ingredients introduced every year, as discussed in Section 3. Within about 10 years, however, high-throughput computer/*in vitro* test methods, which have been under development for several years now, are expected to be available for even the most complex toxicology tests. Toxicologists have likened the development effort to the scale of the Human Genome Project, which took 13 years to complete.

In the US, the new frontier for toxicology is framed in *Toxicity Testing in the 21st Century, A Vision and a Strategy* (NRC, 2007). That document identifies limitations of the current risk assessment process, in which few chemicals can be tested, with *in vivo* procedures that are expensive and of questionable relevance to heterogeneous human populations. The document proposes "*a new toxicity-testing system that evaluates biologically significant perturbations in key toxicity pathways by using new methods in computational biology and a comprehensive array of in vitro tests based on human biology*" (NRC, 2007).

To further this vision, the US Environmental Protection Agency (EPA), the National Institute of Health Chemical Genomics Center (NCGC), the National Toxicology Program, and the FDA have joined in a partnership called Tox21 (http://www.ncats.nih.gov/research/reengineering/tox21/tox21.html). Important advances are the NCGC's robotic screening and informatics platform, which can screen thousands of chemicals per day for toxicological activity in cells, and EPA's ToxCast program, which uses high-throughput screening tests with the aim of both understanding the mechanism of toxicity and selecting the most concerning substances for further testing. ToxCast details, including the list of substances screened to date and early phase results, are available at http://www.epa.gov/ncct/toxcast/.

Because of their low toxicological profile, most cosmetic ingredients could be expected to pass the high-throughput screen-

ing tests, with only a small percentage triggering a red flag due to potential for biological activity or reactivity. About 2% of cosmetic ingredients are reactive or biologically active, with most of those being hair dyes, preservatives, colorants, and UV filters (see Section 2.1.1). High-throughput methods, once available, could rapidly advance cosmetics safety testing.

The proposed Act, at face value, departs from the approach presented in *Toxicity Testing in the 21st Century*, treating all ingredients as substances of concern and requiring full evaluations for all ingredients. Effectively, it drops the concept of screening chemicals and goes directly to full evaluations. If that approach prevails for cosmetics, the advances in high-throughput screening could have less effect on cosmetics safety testing.

1.3 Rationale for the Act

Cosmetic ingredients have a history of generally safe use, although individual adverse reactions to cosmetics and personal care products do occur. Cosmetic ingredients are used precisely because they have a low toxicological profile, so are less likely than many substances to be a problem. The rationale for the Act is not stated in the Act, but it likely is the "precautionary principle," which holds that, in the face of uncertainty, assume the worst case and act accordingly.

The precautionary principle does not imply "by any means." The EU's Cosmetics Regulation, for example, specifically invokes the precautionary principle as a guiding principle, and it also bans animal testing as a means to follow it. The proposed Act would initially involve expanded animal testing. A consideration beyond the scope of this article is whether such expanded testing is justified for cosmetics.

1.4 Article organization and methodology

To estimate possible impacts, we must first understand current conditions. This analysis begins, therefore, by reviewing current cosmetics safety testing in the US, including estimates for current costs and animal use. This information is presented in Section 2.

The proposed Act does not specify a testing program; rather, it specifies a safety standard and requires the FDA to determine for each ingredient the allowable exposure that meets this standard. Section 3 looks at a range of testing programs that might fulfill this requirement and estimates the costs and animal use under each. As part of Section 3, we look at the proposed Act's timeline in detail to see how that may affect costs and the use of animals.

The integrity of this analysis depends on the integrity of the data sources that are the basis for the animal and cost estimates. For this reason, we used primary data sources wherever possible. To obtain estimates of animals used in tests, for example, we examined test dossiers for studies reported in 2012. As reality checks, we cross-checked these numbers with the minimum number of animals specified in the test protocols, with published secondary reports, and with interviews with persons having direct knowledge. For test costs, we obtained current cost estimates directly from laboratories. We compared those with previously published studies of costs, adjusted for inflation to 2013. They compared well, giving confidence that the cost estimates are reasonable. Details on the methodology for each analysis are discussed in the section for that analysis.

Test data is the proprietary data of the company conducting the test and, historically, it has been unavailable to the public. Two pieces of legislation from the EU have created new public databases, making the analyses in this article possible:

– *REACH:* In 2007, the EU implemented the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH, Regulation EC 1907/2006) legislation (EC, 2006). REACH requires safety evaluations of all chemicals produced in quantities greater than 1 ton/year, with the required evaluations dependent on production volume. The agency overseeing REACH is the European Chemicals Agency (ECHA). As part of the REACH program, ECHA has created a public database of chemicals registered under REACH, including the safety dossiers, with test details, for each chemical. This database is available at http://echa.europa.eu/web/guest/information-on-chemicals/registered-substances
– *EU Cosmetics Regulation (Regulation EC 1223/2009, which replaces the previous Cosmetic Directive 76/768/EEC and amendments):* As part of this legislation (EC, 2009a), the European Commission (EC) established the CosIng database of cosmetic ingredients. This is the most comprehensive public database of potential cosmetic ingredients. The CosIng database is available at http://ec.europa.eu/consumers/cosmetics/cosing/

In addition to the raw data available in the EU databases, the experience gained by the EU through these laws gives valuable insight into potential testing under the proposed Safe Cosmetics Act. Where this experience may be relevant, it is included in this article.

1.5 Constraints

This article focuses on proposed new evaluations for cosmetic *ingredients*. The proposed Act also gives the FDA discretion to require evaluations for finished cosmetic products under certain circumstances (section 617[b][2] of the Act), and the Act also briefly mentions ecotoxicity testing (section 615[a][2][C] of the Act). The language in the Act regarding finished cosmetic product testing and ecotoxicity testing is not sufficient to allow analysis here, so these are not considered further.

As noted previously, any evaluations would depend on the implementing regulations. In this paper, we have assumed that implementing regulations would minimize the use of the most cost-intensive, animal-intensive tests (for example, the two-generation reproductive toxicity test). We note, however, that the stipulation of "*reasonable certainty of no harm*" and the requirement to prove "*not more than one in a million risk of any adverse health effect*" could be interpreted much more stringently than assumed in this paper, leading to higher animal use and costs than presented here.

2 Current testing practice

The first step in this analysis is to understand the current practice of cosmetics testing, including estimates of the current

number of animals and current costs. This provides the baseline for comparing the potential impacts of the proposed Act.

2.1 Current cosmetic ingredients testing

In the US, current animal testing for cosmetics is almost exclusively for new ingredients. Safety testing guidelines and safety dossiers for individual cosmetic ingredients reveal details about typical safety testing, as described in this section. This information, paired with estimates of the number of new ingredients each year, allows estimates of the current annual cost and number of animals for cosmetics testing in the US.

Because the US currently does not specify safety tests for cosmetics ingredients, its laws do not determine safety testing for ingredients developed outside the US and then imported into the US. Under the proposed Act, however, expanded safety evaluations would apply to imported ingredients as well as to those developed in the US, increasing animal use and costs worldwide. For this reason, this section estimates current animal use and cost not just in the US, but also worldwide for testing new ingredients, since that baseline will be important when comparing the effects of the proposed Act in Section 3.8. Providing a worldwide estimate also acknowledges that cosmetics in the US use ingredients developed globally, and testing on those ingredients should be included, regardless of where the testing occurs.

2.1.1 Typical safety testing

Cosmetics companies are legally responsible for the safety of their products marketed in the US. For most ingredients, safety tests are not stipulated by law or regulation, but industry and regulatory guidance documents give recommended evaluations, summarized in Tables 1 and 2. Guidance documents generally recommend that companies evaluate all cosmetic ingredients for at least the following health effects, called *health endpoints* (Tab. 1):

– Eye irritation/corrosion
– Skin irritation/corrosion
– Skin sensitization
– Skin penetration
– Mutagenicity/genotoxicity
– Acute toxicity (the effect from a single dose)

As noted in the EU's primary guidance document for cosmetics testing, prepared by the Scientific Committee on Consumer Safety (SCCS), "*A scientifically sound safety evaluation, based on less data than those mentioned above [same as bullets above], becomes quite impossible*" (SCCS, 2010). Although written for the EU, the SCCS guidance document is also recommended by the FDA as a guidance document for cosmetics testing in the US (FDA, 2012).

Beyond these basic tests, considered a first tier of tests, guidance documents emphasize that the need for further testing can only be determined on a case-by-case basis. The SCCS and the EU's Standing Committee on Cosmetic Products (SCCP) give general guidelines for when it may be appropriate to consider a second tier of tests. Summarized in Table 2, second tier tests may be appropriate, for example, for ingredients for which skin penetration tests indicate potentially significant systemic absorp-

Tab. 1: Basic (first tier) data recommended for all cosmetic ingredients by current guidance documents

	Guidance Document Recommendations: X = recommended		
Health Endpoint	**CIR[1]**	**SCCS (2010)[2]**	**SCCP (2012)[3]**
Skin irritation	X	X	X
Skin penetration	X	X	X
Skin sensitization	X	X	X
Mucous membrane (eye) irritation		X	X
Genotoxicity/mutagenicity	X[4]	X[5]	X
Acute toxicity		X	X
Subchronic (90-day repeated dose) toxicity			X

[1] Cosmetic Ingredient Review (CIR), minimum data requirements for reviewing ingredient safety, as reported in Corbett et al. (1999).

[2] SCCS was formerly the SCCP (Scientific Committee on Consumer Products), but is not identical with the Standing Committee on Cosmetic Products (SCCP) in the next reference.

[3] Unlike the other guidelines, SCCP (2012) does not identify a basic data set. Rather, SCCP (2012) lists all relevant endpoints (from irritation studies to reproductive toxicity studies) and notes that the safety assessor decides which endpoints are relevant on a case-by-case basis. However, SCCP (2012) refers readers to SCCS (2010) for guidance, so we can infer the SCCS (2010) guidelines are recommended guidelines for SCCP (2012), too. One difference is that SCCP (2012) specifically states "*The calculation of a MoS (margin of safety) based only on LD_{50} data derived from single dose tests (instead of a NOAEL from at least subacute tests) cannot be used to justify safe use*" (sec 3.8.4). SCCP (2012) further notes in footnote 23 that the preferred test for the NOAEL is the subchronic (90-day repeated dose) study. This suggests that SCCP (2012) recommends the subchronic toxicity study as part of a basic (first tier) data set.

[4] Two studies required, with one using a mammalian system (e.g., an *in vitro* mammalian cell test).

[5] Recommends using three tests: Bacterial Reverse Mutation Test (OECD TG 471, also called the Ames test), *In Vitro* Mammalian Cell Gene Mutation Test (OECD TG 476), and *In Vitro* Micronucleus Test (OECD TG 487).

tion or for ingredients that may be swallowed in a significant amount. Guidance documents recommend that all available data be considered, in a weight of evidence approach, before resorting to new animal tests, and note that decisions about further testing should take into account exposure, use of the product, physico-chemical structure of the ingredient, experience with similar ingredients, and other known information (SCCS, 2010; SCCP, 2012).

A small percentage of ingredients are specifically regulated in either the US or EU, and these have official testing guidelines:

– Ingredients intended to have biological activity: Most ingredients in this category are colorants, preservatives, and UV filters. The EU regulates colorants, preservatives, and UV filters; the US regulates colorants and UV filters. The EU's CosIng database of cosmetic ingredients lists a total of 152 colorants, 56 preservatives, and 26 UV filters, out of about 20,000 total ingredients, indicating about 1% of cosmetic ingredients fall into the category of biologically active ingredients.
– Hair dyes: Nohynek et al. (2010) note these are the most reactive of cosmetic ingredients. They are regulated in the EU, with safety tests specified (SCCS, 2010). The CosIng database lists 220 hair dyes, or about 1% of ingredients.

For these special cases, further recommended testing always includes a repeated dose toxicity study to identify a no observed adverse effect level (NOAEL). The preferred repeated dose study is the subchronic (90-day) study (SCCS, 2010; SCCP, 2012).

Other tests identified as potentially relevant for these special cases are the prenatal developmental toxicity study (OECD

Tab. 2: Further (second tier) testing that may be appropriate, on case-by-case basis

	Guidance Document Recommendations: X = recommended		
Health Endpoint	**CIR[1]**	**SCCS (2010)**	**SCCP (2012)[2]**
Photo-induced toxicity *When:* For UV absorbing ingredients.	X	X	X
Subacute (28-day repeated dose) toxicity *When:* If skin penetration is significant.	X		
Subchronic (90-day repeated dose) toxicity *When:* For colorants, UV filters, preservatives, and other ingredients intended to be biologically active.		X	X[2]
Further genotoxicity *When:* If considerable oral uptake can be expected or if skin penetration is significant, taking into account the toxicological profile of the substance and its chemical structure.		X	X
Prenatal developmental toxicity *When:* If skin penetration is significant (CIR). If considerable oral uptake can be expected or if skin penetration is significant, taking into account the toxicological profile of the substance and its chemical structure (SCCS, 2010).	X	X	X
Two-generation reproductive toxicity *When:* If considerable oral uptake can be expected or if skin penetration is significant, taking into account the toxicological profile of the substance and its chemical structure.		X	X
Carcinogenicity *When:* If a mutagenicity test is positive (CIR). If considerable oral uptake can be expected or if skin penetration is significant, taking into account the toxicological profile of the substance and its chemical structure (SCCS, 2010).	X	X	X
Toxicokinetics *When:* If skin penetration is significant or if considerable oral uptake can be expected, taking into account the toxicological profile of the substance and its chemical structure (SCCS, 2010).		X	X

[1] Cosmetic Ingredient Review (CIR), minimum data requirements for reviewing ingredient safety, as reported in Corbett et al. (1999).

[2] SCCP (2012) refers readers to SCCS (2010) for guidance, so we can infer that the SCCS (2010) guidelines are recommended guidelines for SCCP (2012), too. One difference is that SCCP (2012) specifically states "*The calculation of a MoS (margin of safety) based only on LD_{50} data derived from single dose tests (instead of a NOAEL from at least subacute tests) cannot be used to justify safe use*" (sec 3.8.4). SCCP (2012) further notes in footnote 23 that the preferred test for the NOAEL is the subchronic (90-day repeated dose) study. This suggests that SCCP (2012) recommends the subchronic toxicity study as part of a basic (first tier) data set. For SCCP (2012), therefore, the subchronic toxicity study is also listed in Table 1, as one of the first tier tests.

TG 414), two-generation reproductive toxicity study (OECD TG 416), carcinogenicity study (OECD TG 451 or 453), and toxicokinetic study (OECD TG refers to Organisation for Economic Co-operation and Development Test Guidelines, which are international standard methods for chemical safety testing). Again, guidance documents emphasize the need for additional evaluations to be considered on a case-by-case basis.

Guidelines do not necessarily represent actual practice. A 2011 review of safety dossiers for EU-regulated cosmetic ingredients gives insight into actual practice (Vinken et al., 2011). This review looked at repeated dose studies in dossiers for 220 cosmetic ingredients regulated under the EU Cosmetics Directive (now the EU Cosmetics Regulation) – primarily ingredients in Annexes IV (colorants), V (preservatives), and VI (UV filters). Because these are regulated ingredients, their chemical safety dossiers should include repeated dose studies according to SCCS (2010) guidelines, including the specification for subchronic (90-day) toxicity studies. Table 3 shows the actual repeated dose studies performed. The most frequently run repeated dose tests were the subchronic study, included in 54% of the dossiers, and the prenatal developmental toxicity study, included in 56% of the dossiers; 30% of the dossiers had no repeated dose study information.

Going forward in this analysis, the percentages in Table 3 are used to help estimate the number of tests currently run on new cosmetic ingredients in special categories, because the percentages reflect real practice. These categories represent a small percentage of cosmetic ingredients, but the additional tests are animal intensive, so their contribution to the total is significant.

2.1.2 Techniques to reduce testing

An *evaluation* for a health endpoint does not necessarily mean *testing*. If a new ingredient is in a chemical group with known properties, the new ingredient's properties sometimes can be estimated based on the known properties of that group, a technique called *read-across*. Estimation methods are accepted as scientifically sound and are encouraged by US regulatory agencies.

An evaluation can also be done based on weight of evidence, in which the safety assessor reviews all existing data and experience to assess the safety of an ingredient. The available data must be sufficient for such an approach, and the approach must be thoroughly justified. The weight of evidence approach, too, is encouraged by regulatory agencies to reduce the need for new testing.

The EU's REACH testing program gives us insight into the frequency of use of estimation techniques and the weight of evidence approach. As part of the REACH program, the European Chemicals Agency (ECHA) compiled these data for studies contained in 1,504 chemical safety dossiers (ECHA, 2011). Table 4 shows their frequency of use for eye irritation, skin irritation, skin sensitization, acute toxicity, and mutagenicity (based on Table 4 of ECHA, 2011). The data show that the read-across technique was used in 20% to 23% of studies for most health endpoints. The exceptions were the *in vitro* skin and eye irritation tests, for which the percentages were 7% and 11.9%, respectively. The weight of evidence approach was used in 6.6% to 13.7% of studies, depending on the health endpoint. Since the first requirement of REACH is "*gathering all existing information*," the dossiers for many substances contain tens of studies for each endpoint (Rovida et al., 2011). The percentages in Table 4 are calculated on all studies that have been presented, rather than on the number of substances.

The REACH dossiers were for high-volume chemicals (>1,000 tons/year). Less than 3% of cosmetic ingredients are in this category (see Section 3.3, later in this article), so these REACH data may not be representative of cosmetics ingredients, in particular for plant ingredients. Plant ingredients may be especially suited to read-across estimates and to comparative analysis to plant extracts with a traditional history of use as food (Nohynek et al., 2012). The CosIng database indicates that plant ingredients comprise about 30% of cosmetic ingredients, so the percentages of estimation techniques for cosmetic ingredients overall may be higher than that for REACH high-volume chemicals.

Tab. 3: Incidence of repeated dose studies for 220 EU-regulated cosmetics ingredients[1]

Repeated Dose Study	No. of Ingredients with Study	% Ingredients with Study
Subchronic (90-day) repeated dose toxicity study	118	54%
Subacute (28-day) repeated dose toxicity study	542	25%[2]
Chronic (>1 year) repeated dose toxicity study	17	8%
Carcinogenicity study	15	7%
Prenatal developmental toxicity study	124	56%
Two-generation reproductive toxicity study	10	5%
No repeated dose studies	66	30%

[1] Data from Vinken et al. (2011), a study that examined dossiers for 220 cosmetics ingredients regulated by the EU. The regulated ingredients are intended to be biologically active, and typically are colorants, preservatives, and UV filters.

[2] 44 of the subacute tests were in addition to the subchronic study for the ingredient; only 10 ingredients (5% of ingredients) were tested only in the subacute study.

Offsetting this, ECHA (2011) reports that estimation techniques may not have been appropriately applied to some REACH chemicals, so the percentages in Table 4 may be high, including invalid use of the techniques. Despite the limitations of the ECHA percentages, they are the best available data currently, and are used in this analysis to estimate the use of read-across and weight of evidence techniques for eye irritation, skin irritation, skin sensitization, acute toxicity, and mutagenicity.

ECHA also reports the percentages of read-across and weight of evidence for repeated dose studies. For new cosmetic ingredients, however, Adler et al. (2011) report that estimation methods are rarely used for repeated dose tests, "*partly due to the limited ability to read across to very novel ingredients and partly because a robust evaluation requires a large amount of animal data on structurally similar materials, which is rarely available*." The EU's ban on animal testing for cosmetic ingredients precludes the use of new repeated dose tests for any ingredient marketed in the EU. The number of very novel ingredients may decline, therefore, until alternative methods are available to replace the repeated dose tests. This may mean that new ingredients are more likely to be variations of compounds whose toxicity history is already understood, allowing use of read-across and weight of evidence techniques to substantiate safety. Offsetting this, cosmetics and ingredients not marketed in the EU may still undergo these tests; in fact, some countries require animal testing. Consistent with Adler et al. (2011), this analysis assumes estimation techniques are not used for repeated dose tests for new cosmetic ingredients.

As noted in Section 1.2, major efforts are under way to develop faster, more predictive chemical safety tests that may eliminate the need for animal testing. In the US, the framing document for this effort is *Toxicity Testing in the 21st Century* (NRC, 2007), which envisions the use of high-throughput *in vitro* screening tests on all chemicals with the aim of both understanding the mechanism for toxicity and immediately selecting the most concerning substances for further testing. Recently, internationally recognized experts met in a workshop to examine the state of the science for alternative methods for cosmetics, as the basis for future testing without the use of living vertebrate animals (Basketter et al., 2012). Once high-throughput and other alternative methods are available, expected in about 10 years, cosmetics ingredients can be evaluated rapidly, likely without the need for animals.

2.1.3 Cost per test

Table 5 shows the cost estimate for each OECD test method. Cost estimates are based on the average of prices in Fleischer (2007), OECD (2012), ECHA (2012), and 2013 prices for two US laboratories and three EU laboratories. All prices were adjusted for inflation to May 2013. If prices were in Euros (€) initially, they were first adjusted for EU inflation to May 2013, and then converted to US$ using the currency conversion rate for May 2013 (1 € = $ 1.30 US). Note that converting € prices to $ US first and then adjusting for US inflation to May 2013 does not change the result significantly, because the inflation rates for the US and Europe were about the same for the period spanning these price quotes.

The Fleischer (2007) prices, which are from June 2004, are the oldest data. They were adjusted for the EU inflation rate from June 2004 to May 2013, which was about 20% using either the HICP (Harmonized Index of Consumer Prices) inflation index or the individual CPI (Consumer Price Index) indices for major countries. For ECHA (2012) and OECD (2012), 2012 prices were adjusted for 2% inflation from 2012 to 2013. The Fleischer report quoted multiple (2 to 12) laboratory cost estimates for each test and included the average of these. For the analysis here, we use the Fleischer average as a single data point, rather than including the individual cost estimates for each test. This weights our cost average toward the newer, 2012 and 2013, data.

Comparing known 2013 prices with the 2004 prices adjusted for inflation, we can see that test prices have remained steady, and possibly even declined slightly, since 2004. This suggests that laboratory capacity was not strained by REACH. One possibility is that laboratories in China and India in particular may

Tab. 4: Use of read-across and weight of evidence techniques in REACH dossiers

Health Endpoint	Use of Read-Across	Use of Weight of Evidence	Total
Eye irritation – *in vivo*	20.9%	6.6%	27.5%
Eye irritation – *in vitro*	7.0%	2.9%	9.9%
Skin irritation – *in vivo*	21.3%	7.7%	29.0%
Skin irritation – *in vitro*	11.9%	10.6%	22.5%
Skin sensitization – *in vivo*	20.8%	13.7%	34.5%
Genotoxicity/mutagenicity – *in vitro*	22.0%	12.1%	34.1%
Acute toxicity – oral	21.2%	8.6%	29.8%
Acute toxicity – dermal	23.1%	7.6%	30.7%
Acute toxicity – inhalation	20.5%	9.6%	30.1%

Data are from ECHA (2011). *The use of alternatives to testing on animals for the REACH regulation.* ECHA-11-R-004.2-EN, Table 4.

have provided sufficient capacity to keep costs from rising. For our later analysis of potential costs under the proposed Act, we will assume a similar pattern: that is, that costs will remain steady, except for inflation, for the 10 year period of the analysis. For this reason, we will not need a present value analysis to account for price changes other than inflation.

Table 5 shows the costs that an average contract laboratory charges for the test only. The following additional services are typically required in conjunction with the test:

– *Study management/administration and preparation of the toxicology advisory:* The cost for each test is increased by at least 50% to account for administrative/management costs plus the toxicology advisory that is always added to come to a practical conclusion in the toxicological assessment of the test item.
– *Pre-dose finding study for long-term studies:* Long-term studies are preceded by a preliminary study, which is used to find the subtoxic doses. These pre-dose finding studies (also called range-finding studies) are usually 15-day repeated dose studies that are performed on the same number of animals at three doses plus the negative control. The added cost from the preliminary dose-finding study is about $ 10,000 per test item. One pre-dose finding study can be used for all long-term studies for the same substance and therefore it should be counted only once per substance.
– *Analytical determinations:* Good Laboratory Practice (GLP) studies require that concentrations be measured by proper analytical methods. Extra costs include the validation of the analytical method and the measure of each solution that is prepared and used. This cost strongly depends on the analytical technique.

Not all ingredients would require testing. Some could be evaluated through read-across and weight of evidence techniques. These require the use of existing studies. If an existing study is published and does not contain a disclaimer prohibiting its use without permission, it probably can be used free of charge. For other studies, the safety assessor must obtain permission to use the studies from the owners of those studies. Called Letters of Access, these can be more than half the price of the study itself. The cost estimate for a read-across or weight of evidence evalu-

Tab. 5: Cost estimates for OECD test methods

		Fleischer (2007): Test costs; 2 to 12 labs responded per endpoint (>5 typical)		ECHA (2012b): Costs for TG 416 and 43; 13 labs responded				
Health Endpoint	**OECD Test Method**	**June 2004 avg cost in €**	**adjusted to May 2013 & US $[1]**	**2012 avg cost in €**	**adjusted to May 2013 & US $[2]**	**OECD (2012) prices, adjusted to May 2013[2]**	**Avg CRO 2013 prices/ 2 to 5 labs[3]**	**May 2013 avg cost**
Eye irritation/ corrosion	437, 438	€ 1,615	$ 2,519				$ 3,800	$ 3,480
	405	€ 1,343	$ 2,095				$ 2,455	$ 2,335
Skin irritation/ corrosion	430, 431, 439	€ 1,645	$ 2,566					$ 2,566
Skin irritation/ corrosion	404	€ 1,194	$ 1,863				$ 2,530	$ 2,308
Skin sensitization	429	€ 3,959	$ 6,176				$ 6,626	$ 6,571
Skin penetration	428						$ 24,700	$ 24,700
Mutagenicity/ genotoxicity	471 (Ames), 472	€ 3,174	$ 4,951				$ 4,505	$ 4,590
	473CA	€ 19,161	$ 29,891					$ 29,891
	473MNT	€ 11,000	$ 17,160					$ 17,160
	473 unspecified						$ 17,800	$ 17,800
	476 MLA	€ 16,603	$ 25,901					$ 25,901
	476 HPRT	€ 17,283	$ 26,961					$ 26,961
	476 unspecified						$ 19,400	$ 19,400
Acute toxicity – oral								
	420 (fixed dose)						$ 3,200	$ 3,200

		Fleischer (2007): Test costs; 2 to 12 labs responded per endpoint (>5 typical)		ECHA (2012b): Costs for TG 416 and 43; 13 labs responded				
Health Endpoint	**OECD Test Method**	**June 2004 avg cost in €**	**adjusted to May 2013 & US $[1]**	**2012 avg cost in €**	**adjusted to May 2013 & US $[2]**	**OECD (2012) prices, adjusted to May 2013[2]**	**Avg CRO 2013 prices/ 2 to 5 labs[3]**	**May 2013 avg cost**
	423	€ 1,474	$ 2,299					$ 2,299
	425 (up or down test)						$ 1,640	$ 1,640
Acute toxicity – dermal	402	€ 2,011	$ 3,137				$ 2,654	$ 2,718
Acute toxicity – inhalation	403	€ 11,734	$ 18,305				$ 13,975	$ 15,418
Subacute (28-day repeated dose) toxicity – oral	407	€ 49,390	$ 77,048			$ 63,369	$ 75,575	$ 73,769
Subacute toxicity – dermal	410	€ 49,550	$ 77,298				$ 66,500	$ 71,899
Subacute toxicity – inhalation	412	€ 105,455	$ 164,510					$ 164,510
Phototoxicity	432						$ 5,200	$ 5,200
Subchronic oral toxicity (90-day repeated dose)	408	€ 115,656	$ 180,423				$ 161,400	$ 170,912
Subchronic dermal toxicity (90-day repeated dose)	411						$ 169,200	$ 169,200
Subchronic inhalation toxicity (90-day repeated dose)	413							
Chronic toxicity (>1 year)	452	€ 372,000	$ 580,320					$ 580,320
Carcinogenicity	451	€ 780,357	$ 1,217,357				$ 1,209,000	$ 1,213,178
Extended One-generation Reproductive Toxicity	443							
- Basic study				€ 414,273	$ 646,266			$ 646,266
- Basic study with optional second generation				€ 469,778	$ 732,854			$ 732,854
- Basic study with neurotox (DNT) module				€ 507,444	$ 791,613			$ 791,613

		Fleischer (2007): Test costs; 2 to 12 labs responded per endpoint (>5 typical)		ECHA (2012b): Costs for TG 416 and 43; 13 labs responded				
Health Endpoint	**OECD Test Method**	**June 2004 avg cost in €**	**adjusted to May 2013 & US $[1]**	**2012 avg cost in €**	**adjusted to May 2013 & US $[2]**	**OECD (2012) prices, adjusted to May 2013[2]**	**Avg CRO 2013 prices/ 2 to 5 labs[3]**	**May 2013 avg cost**
- Basic study with immunotox (DIT) module				€ 440,414	$ 687,046			$ 687,046
- Basic study with both modules				€ 567,964	$ 886,024			$ 886,024
- Basic study with two gens and both modules				€ 655,195	$ 1,022,104			$ 1,022,104
Two-generation reproductive toxicity	416	€ 327,975	$ 511,641	€ 285,842	$ 445,914	$ 324,666	$ 429,000	$ 427,805
Reproductive/ developmental toxicity screening	421	€ 54,597	$ 85,171				$ 101,800	$ 93,486
Reproductive/ developmental toxicity screening	422						$ 145,800	$ 145,800
Prenatal developmental toxicity	414	€ 63,100	$ 98,436				$ 113,200	$ 105,818
Prenatal developmental toxicity, second species	414	€ 92,500	$ 144,300					$ 144,300
Toxicokinetics	method not specified	€ 33,041	$ 51,544				$ 650,000	

Currency conversion obtained from http://finance.yahoo.com/currency-converter/?amt=1&from=EUR&to=USD&submit=Convert#from=EUR;to=USD;amt=1

US Consumer Price Index inflation rates obtained from http://www.usinflationcalculator.com/inflation/current-inflation-rates/

EU inflation rates obtained from http://global-rates.com/economic-indicators/inflation/consumer-prices/hicp/eurozone.aspx

[1] Fleischer prices are from 2004. They are adjusted for inflation to May 2013 (inflation about 20% for that period) and then converted to $US (1 € = $ 1.30 on May 31, 2013).

[2] Adjusted for inflation to May 2013 (~2%) and converted to US $ (1 € = $ 1.30 on May 31, 2013).

[3] These prices are from two US labs (Contract Research Organizations, or CROs) and 3 EU labs (CROs). To protect the confidentiality of the labs, their data are shown only as the average of the labs here. For the calculation, however, each lab's prices are considered as a single data point. Not all labs provided prices for all endpoints; the sample size for most averages is 2 to 5 labs. Subacute dermal and subchronic oral and dermal are from one lab. The original data for the EU labs are in Euros, converted to US dollars here (1 € = $ 1.30 on May 31, 2013).

ation must include costs for the following: the demonstration that such a technique is acceptable for the substance; an evaluation of the substance using the estimation techniques; and the possible purchase of Letters of Access. The analysis here assumes that an evaluation through estimation techniques is one-half the cost of an evaluation involving laboratory testing.

2.1.4 Animal use per test

Historically, the number of animals per test has been calculated using official documents such as Hofer et al. (2004) and van der Jagt et al. (2004). Now, a more precise evaluation is possible through ECHA's public database of REACH registration dossiers. This database is quite exhaustive, especially for studies

that have been performed after 2010. In particular, there is good standardization in the protocols that strictly follow the OECD guidelines.

To evaluate the typical number of animals per test, studies were selected through eChemPortal (http://www.echemportal.org), where search by OECD guideline is possible. For each endpoint, the search included the corresponding OECD guideline and the report year of 2012, assuming that the most recent year may provide higher reliability that the study was performed according to the latest Good Laboratory Practice (GLP) requirements. The search returns a list of chemicals with their CAS number. To obtain details on animal use, CAS numbers were randomly selected and then searched in the ECHA database of registered REACH substances (http://echa.europa.eu/web/guest/information-on-chemicals/registered-substances). This provided links to the REACH registration dossiers for those chemicals. The dossiers include the study details, making it possible to get the species and number of animals used for that specific study.

Table 6 shows the animal numbers reported for studies for each OECD guideline. For comparison, it also shows the minimum number of animals specified in the OECD guidelines for each method. Long-term studies generally use more than the minimum to assure enough animals to meet the minimum requirement through to the end of the experiment. Also, OECD guidelines usually call for a negative control group as a minimum control; in practice, some studies include both a positive and a negative control group or two negative control groups. For OECD methods where the REACH dossiers did not provide animal counts, this analysis used the minimum specified in the OECD guidelines.

Most, if not all, long-term studies use the following additional animals, which normally are not reported:

– *Animals in pre-dose finding studies:* As noted in the costs section, long-term studies are preceded by a preliminary study, which is used to find the subtoxic doses. These pre-dose finding studies (also called range-finding studies) are usually 15-day repeated dose studies that are performed on 15 animals at three doses plus the negative control, using up to 80 animals. One pre-dose finding study can be used for all long-term studies for the same substance and therefore it should be counted only once per substance.
– *Satellite animal groups:* For long-term studies using small animals such as rats (which is the case for most cosmetics safety tests), limited blood samples can be drawn. There are humane limitations to the amount of blood taken from an animal each time; significant blood loss and trauma also can affect the study results. In consideration of this, studies use *satellite* groups, which are dosed with the test substance in the same manner as the main group (usually only the highest and control doses), but they have no additional investigation that may cause stress, like withdrawal of blood samples. They are used for additional investigation in case of doubtful outcomes and to replace animals in the main study that unexpectedly die.

Logically, these animals should count, because they undergo testing and are euthanized in the end, too. However, their numbers usually are not included in final study reports. Without reported numbers, we cannot include them in this analysis, because the number may be very variable from study to study. Estimating the numbers by adding a fixed percentage to the final calculation may lead to a higher level of error than not considering it at all, so we do not include such an adjustment.

The reproductive toxicity studies (OECD TG 414 and 416) involve parent animals and their offspring. OECD TG 416 involves the parent generation (P) and their children (F1) and grandchildren (F2). Parent (P) and children (F1) undergo the same dosing and are euthanized once their pups are weaned. The grandchildren (F2) are euthanized after weaning. OECD TG 414 involves pregnant females that are dosed and then euthanized just before delivery, and the fetuses are then examined for toxic effects. To count animals in these tests, investigators historically have used two methods: some have counted only the parent animals, and some have counted both the parents and the offspring. In van der Jagt et al. (2004), for example, only parents are counted for both OECD TG 414 and OECD TG 416; in Hofer et al. (2004), parents and offspring are counted for both 414 and 416; and in Cooper et al. (2006), parents and offspring are counted for OECD TG 416, and only parents are counted for OECD TG 414. Counting offspring is now the official mode in the EU, adopted in EU Directive 63/2010, which considers as experimental animals all animals that are bred for the purposes of scientific experiments plus fetuses that are in the last third of their normal development. The US has no official counting method.

This analysis follows the convention of Hofer et al. (2004) and EU Directive 63/2010 and counts offspring, given that F1 offspring undergo the same test protocol and conditions as parents in OECD TG 416 and all offspring are euthanized in both OECD TG 414 and OECD TG 416. For OECD TG 414, only the female parent is counted, because the study begins with pregnant females.

The extended one generation reproductive toxicity test, OECD TG 443, is recent and its applicability as a replacement for the full two-generation study (OECD TG 416) is still under discussion. The number of animals in Table 6 is estimated from the OECD guideline, as there is still not enough data to evaluate the real number.

Rat is considered the species of choice for all studies but eye irritation and studies that are based on dermal exposure. To calculate the total number of rats used for a reproductive toxicity study, this analysis assumes a litter size of 12 pups per pregnant female. The average of 12 accounts for possible reduction caused by an effect of the tested chemical or the accidental death of any pregnant rats. Note that an average litter size of 12 likely underestimates the real litter size, which is typically about 15 in REACH dossiers. In previous estimations, a lower number was used (Rovida and Hartung, 2009), which derived from current practice at that time. Probably, compared with the past, new studies are generally performed at subtoxic doses and therefore delivery usually is not affected in terms of reduced pups per litter. Sometimes, a higher number of resorptions is recorded at the highest dose, but usually nothing more.

Tab. 6: Animal use for OECD test methods

Health Endpoint	OECD Test Method	Animal Species in Test	Min. Animal No. Specified in Test Method	Typical No. of Animals per Test in ECHA DB	Min-Max in ECHA DB	Number Used in this Analysis[1]
Eye irritation/corrosion	437, 438	none	0			0
Eye irritation/corrosion	405	rabbits	3	6	6	6
Skin irritation/corrosion	430, 431, 439	none	0			0
Skin irritation/corrosion	404	rabbits	3	6	3-12	6
Skin sensitization	429	mice	16	25	20-30[2]	25
Skin penetration	428	none	0			0
Phototoxicity	432	none	0			0
Mutagenicity/genotoxicity						
- Bacterial Reverse Mutation (Ames)	471	none	0			0
- Mammalian Cell Gene Mutation	476	none	0			0
- Mammalian Chromosomal Aberration Test	473	none	0			0
- Mammalian Cell Micronucleus Test	487	none	0			0
Acute toxicity – oral	420 (fixed dose) 425 (up or down)	rats	8	no data		8
Acute toxicity – dermal	402	rabbits, rats, guinea pigs	10	10	10	10
Acute toxicity – inhalation	403	rats	20	15	10-20	15
Subacute (28-day repeated dose) toxicity – oral	407 (422, below, is becoming standard)	rats	40	40	40-60	40
Subacute (28-day repeated dose) toxicity – dermal	410	rabbits, rats, guinea pigs	40	120	50-120	120
Subacute toxicity – inhalation	412	rats	40	80	24-100	80
Subchronic (90-day repeated dose) toxicity - oral	408	rats	80	100	80-120	100
Subchronic (90-day repeated dose) toxicity dermal	411	rabbits, rats, guinea pigs	80	no data		80
Subchronic (90-day repeated dose) toxicity – inhalation	413	rats	80	no data		80
Chronic toxicity (>1 year)	452	rats	160	no data		160
Carcinogenicity	451	rats	400	416	400-616	416
Reproductive/developmental toxicity screening (this is being supplanted by the combined screening/subacute toxicity test, OECD TG 422)	421	rats	480	no data		480
Reproductive/developmental toxicity screening with subacute toxicity	422	rats	480	520	464-680	520

Health Endpoint	OECD Test Method	Animal Species in Test	Min. Animal No. Specified in Test Method	Typical No. of Animals per Test in ECHA DB	Min-Max in ECHA DB	Number Used in this Analysis[1]
Prenatal developmental toxicity	414	rats, rabbits	1,040 (rats) 560 (rabbits)[3]	768	466-1,536	768
Two-generation reproductive toxicity	416	rats	2,080[4]	3,025	2,850-3,200	3,025
Extended One-generation reproductive toxicity	443	rats	1,100	no data		1,100
Toxicokinetics	no standardized protocol			highly variable		not considered

[1] *Number Used in this Analysis* has been selected as either the average or the most representative number among the considered studies.

[2] Rovida (2011).

[3] 80 pregnant females required. Animals arrive pregnant. Assumes average litter size of 12 for rats and 6 for rabbits. Fetuses are euthanized just before delivery, so are counted, consistent with the counting method of Hofer et al. (2004) and EU Directive 63/2010.

[4] The minimum number is calculated as follows: 4 groups (3 dose groups plus control group), with each group having a minimum of 20 males and 20 females (20 pregnancies per group required). This gives a total of 160 parents. If have of 80 pregnant females in parent generation giving birth to avg litter of 12, results in 960 children (F1); then keep 1 male and 1 female from each litter for mating for F2 generation, results in total of 960 grandchildren (F2). Totaling all gives 160+960+960 = 2,080. Minimum assumes 100% pregnancy rate among mated couples.

Eye irritation *in vivo* and skin sensitization according to OECD TG 429 (Local Lymph Node Assay, or LLNA) soon can be fully replaced by *in vitro* alternatives at about the same price.

2.1.5 Number of new cosmetic ingredients annually

A primary source for the number of new ingredients each year is the Personal Care Products Council (PCPC), which assigns the official INCI (International Nomenclature Cosmetic Ingredient) names to new ingredients. About 500 to 700 new cosmetic ingredients receive INCI names each year (Personal Care Products Council, 2012). The PCPC assigns INCI names for new ingredients developed worldwide, not just those developed in the US. The European Union (EU) and Japan develop most new ingredients. In 2005, for example, the EU received 2,599 cosmetic patents, including for new ingredients, Japan received 2,976, and the US received 685 (Global Insight, 2007). Overall in 2005, the US received about 10% of new cosmetic patents, including for new ingredients. Assuming this percentage approximately applies to new ingredient patents, this suggests about 10% of new ingredients, or about 50 to 70 new ingredients, are developed in the US each year.

2.1.6 Pulling it all together: Estimate of annual testing for new cosmetic ingredients

This section pulls together the different threads developed in the previous sections to calculate the estimated current testing in the US. The estimate uses the following assumptions:

– About 60 new ingredients are tested annually in the US each year (see Section 2.1.5).
– The tests are those suggested in the guidance documents (see Tab. 1 and 2 in Section 2.1.1).
– All new ingredients (100%) undergo the first tier tests in Table 1. The percentage of ingredients that involve further testing (Tab. 2 tests) is about 3%, based on the current percentage of regulated ingredients in the EU, which is 2% (see Section 2.1.1), and allowing for another 1% that may fall into the categories typically requiring more rigorous scrutiny. For the phototoxicity test, the percentage used is 0.1%, which is the percentage of UV filters among total ingredients.
– For second tier tests (Tab. 2 tests), the 3% is multiplied by the percentage of dossiers that actually contained that test for regulated ingredients (see Tab. 3 in Section 2.1.1), reflecting that actual practice may differ from the guidelines. For subchronic toxicity, for example, the 3% of ingredients for which the test may be recommended is multiplied by 54%, which is the percentage of dossiers for that category of ingredient that actually contained the subchronic toxicity study. The result is 3% x 54% = 1.6% of ingredients in that category estimated to undergo the subchronic toxicity study.
– Read-across and weight of evidence techniques are used at the percentages indicated in the REACH dossiers (see Tab. 4 in Section 2.1.2). The cost for a read-across or weight of evidence evaluation is about one-half the cost of the corresponding laboratory test (see Section 2.1.3, "Cost Per Test").

As shown in Table 7, this yields an estimate of 2,700 animals annually for ingredient evaluations in the US. The cost estimate is about $ 3.2 million for laboratory tests and $ 600,000 for read-across/weight of evidence evaluations. As noted in Section 2.1.3, laboratory test costs are for only the test itself. The cost for management/administration/analysis of the test and prepara-

Tab. 7: Estimated animal use and cost for new US cosmetic ingredient testing (60 new ingredients/year)

Health Endpoint	% Ingredients Evaluated[1]	% Ingredients Evaluated by RA/ WOE[2]	No. of Ingredients Undergoing Test[3]	OECD Test Method	No. of Animals Per Test[4]	Animals Tested	Cost per Test (US $)[4]	Extended Test Cost (US $)	Cost for RA/WOE Evaluations (US $)[5]
Eye irritation – *in vivo*	50%	27.5%	21.8	405	6	131	$ 2,300	$ 50,025	$ 9,488
Eye irritation – *in vitro*	50%	9.9%	27.0	437/438	0	0	$ 3,500	$ 94,605	$ 5,198
Skin irritation – *in vitro*	100%	22.5%	46.5	431	0	0	$ 2,500	$ 116,000	$ 16,875
Skin sensitization	100%	34.5%	39.3	429	25	983	$ 6,500	$ 255,450	$ 67,275
Skin penetration	100%	30%[2]	42.0	428	0	0	$ 25,000	$ 1,050,000	$ 225,000
Mutagenicity – Ames test	100%	34.1%	39.5	471	0	0	$ 4,600	$ 181,884	$ 47,058
Mutagenicity – mammalian cell test	100%	34.1%	39.5	473/476	0	0	$ 20,000	$ 790,800	$ 204,600
Acute toxicity (oral – fixed dose)	100%	30%	42.0	420	8	336	$ 3,200	$ 134,400	$ 28,800
Subacute toxicity (oral)	0.75%	0%	0.44	407	40	18	$ 74,000	$ 32,695	$ 0
Subchronic toxicity (oral)	1.6%	0%	0.97	408	100	97	$ 170,000	$ 164,127	$ 0
Chronic toxicity (oral)	0.24%	0%	0.14	452	160	22	$ 580,000	$ 80,673	$ 0
Prenatal developmental toxicity	1.7%	0%	1.01	414	768	779	$ 105,000	$ 106,527	$ 0
Two-generation reproductive toxicity	0.15%	0%	0.08	416	3025	248	$ 430,000	$ 35,182	$ 0
Carcinogenicity	0.21%	0%	0.12	451	416	51	$ 1,200,000	$ 147,273	$ 0
Toxicokinetics	no data; assume 0.1%	0%	0.06	417	9	1	$ 50,000- $ 650,000	$ 6,000	$ 0
Phototoxicity	0.1%	30%[2]	0.04	432	0	0	$ 5,200	$ 218	$ 47
					Total	2,664		$ 3,246,109	$ 604,340

[1] For the derivation of these percentages, see Section 2.1.6. For eye irritation, 100% of ingredients are assumed to have this test, split 50-50 between the *in vivo* test and *in vitro* test. A recent update of the OECD TG 437 (July 26, 2013) was expanded to accept also negative results and consequently this 50/50 ratio may move in the future towards a larger applicability of the *in vitro* procedure.

[2] Percentages for eye and skin irritation, skin sensitization, mutagenicity, and acute toxicity are from REACH dossiers (ECHA, 2011). See Table 4 in this article for details. No REACH data were available for skin penetration and phototoxicity. For these, an estimate of 30% is used, which is in the midrange of known percentages for the other endpoints. For repeated dose studies, estimation techniques are rarely used for new ingredients (Adler et al., 2011), so 0% is used here.

[3] Calculated as (60 new ingredients) x (% ingredients evaluated for this endpoint) x (1 - % evaluated by RA/WOE). The result is less than 1 for tests that are typically not concluded in one calendar year or that are rarely conducted.

[4] For the derivation of the costs and animal numbers, see Section 2.1.3 and 2.1.4, respectively.

[5] Calculated as (60 new ingredients) x (% ingredients evaluated for this endpoint) x (% evaluated by RA/WOE) x (50% of lab test cost). As noted in Section 2.1.3, this analysis assumes the unit cost of a RA/WOE evaluation is about one-half the cost of the corresponding laboratory test.

tion of the toxicology report adds about another 50%, or about another $ 1.6 million. This results in a total annual cost estimate of $ 5.4 million.

An ingredient developed in the US is not necessarily tested in the US. Similarly, US testing laboratories serve foreign companies, too. There is no easy way to account for this, but it is also not critical to account for it. Note that even if you take the most extreme assumption, that all 500 to 700 new ingredients worldwide are tested in the US, this results in an estimate of 27,000 animals and $ 54 million annually. As will be seen later in this analysis, these numbers are low compared with the estimate of animals and cost under the proposed Act, so further refining the estimate for current testing in the US is not critical.

Also, as noted previously, cosmetics in the US use ingredients developed globally. Currently, US law does not mandate specific safety tests for ingredients, so its laws are not a primary factor in the choice of safety tests worldwide. The safety evaluations mandated by the proposed Act, however, would affect safety testing worldwide. The worldwide number, therefore, is a baseline for comparing the effects of the proposed Act. If we consider testing worldwide, we can use the cost and animal estimates just given for 500 to 700 new ingredients: 27,000 animals and $ 54 million annually. Although EU law prohibits cosmetic ingredients tested on animals, those ingredients may still undergo animal tests when exported to other countries; therefore, the numbers probably need not be adjusted to deduct EU-developed ingredients.

2.1.7 Reality check of estimate

Under US law, laboratories that use animals for testing are inspected by the US Department of Agriculture (USDA), and the USDA inspection reports include the number of animals; however, the reports exclude mice and rats, and they do not break out the number of animals by testing purpose (for example, for cosmetics testing). The USDA reports, therefore, cannot provide supporting information for this estimate of cosmetics testing.

The European Union, however, does include mice and rats in reports and does break out animals by testing purpose. Table 8 summarizes the EU reports for 2005 and 2008, which were the last two reports before the first stage of the EU's ban on animal testing for cosmetic ingredients took effect in 2009 (the full ban was completed in 2013). We can know these tests were for ingredients, rather than for finished cosmetics, because the EU banned animal testing of finished cosmetic products in 2004, so all cosmetic-related animal tests between 2004 and 2009 can only be for ingredients. The total number of animals for mammalian toxicity tests in the 2005 EU report is 5,571, and in the 2008 EU report is 1,245 (EC, 2007, 2010). Note that the 2008 report lists a total of 1,967 animals, but 722 are fish, presumably for the LC_{50} fish toxicity test. Since we are not considering ecotoxicity tests in this article, the 722 fish are excluded here, resulting in the total of 1,245.

Much of the difference between the 2005 and 2008 totals likely relates to more ingredients having been tested in 2005. Also, in 2005, some ingredients underwent repeated dose tests, whereas the ingredients in 2008 did not. Such year to year differences are expected: Since relatively few ingredients fall into the categories that undergo the second tier of testing, not all years would include the more intensive repeated dose tests. The average of these two years is 3,408 animals.

Tab. 8: EU reports of animals used for cosmetic ingredient testing

Health Endpoint	2005[1]	2008[2]
Eye irritation	300	54
Skin irritation	469	87
Skin sensitization	2,222	699
Mutagenicity	213	0
Acute/subacute toxicity	1,033	405[3]
Subchronic/chronic toxicity	966	0
Prenatal developmental toxicity	368	0
Reproductive toxicity	0	0
Carcinogenicity	0	0
Total	5,571	1,245

[1] From EC (2007), Table 8.1.

[2] From EC (2010), Table 8.1.

[3] The reported total number of animals for this category is 1127; however, 722 are fish most likely used for the LC_{50} ecotoxicity test. We have not counted the fish in the total here, since we are not including ecotoxicity tests in our analysis.

These EU numbers suggest that the US estimate in this analysis (2,700) is reasonable. Remember that the EU published more than triple the cosmetic patents of the US in 2005; there have been no market shifts since 2005 that would significantly change this pattern. At the least, we can see that the scale of animal testing for new ingredients in the US is on the order of thousands per year, rather than tens of thousands.

As a further reality check, we directly asked people with first-hand observations – those working in testing laboratories and the USDA staff who inspect the laboratories. In an informal phone survey of three US testing laboratories and a regional supervisor of USDA inspectors (who in turn polled his inspectors), no one had seen animal testing for cosmetics for years. When asked if an annual estimate of 0 to the low thousands was a reasonable estimate, the laboratory employees and USDA supervisor answered yes (the inspectors themselves were not asked that particular question). Although not a scientific survey, these anecdotes are useful as another reality check on the estimate.

2.2 Current finished cosmetic product testing: Non-animal methods are the norm

For finished cosmetic products, current routine practice in the US involves the following safety tests, all of which are alternative, non-animal test methods:

– *Eye irritation:* Usually tested *in vitro* through OECD TG 437 or 438.
– *Skin irritation/corrosion:* Can be tested through several cell-based methods, for example, EpiSkin, or through the Cumula-

tive (21-day) Irritation Test, which uses humans. Testing on humans is possible, because a finished cosmetic has ingredients within limits already determined to be safe through the ingredient testing.

– *Acute toxicity:* Assumed safe based on the ingredient results. An acute toxicity test is not conducted, because the ingredients have typically already undergone an acute toxicity test. The assumption is that if the individual ingredients have been found safe, then the overall product is safe. This is consistent with the FDA's guidance (FDA, 2012): "*…the safety of a cosmetic product should be evaluated by analyzing the physicochemical properties and the relevant toxicological endpoints of each ingredient in relation to the expected exposure levels resulting from the intended use of the finished product*." This is also consistent with the language of the proposed Act, which says that the Secretary shall presume that a finished cosmetic meets the safety standard if it consists solely of ingredients that meet the safety standard (Safe Cosmetics and Personal Care Products Act of 2013, section 617 [b][1]).
– *Skin sensitization:* Similarly to acute toxicity, skin sensitization of a finished product may be assumed safe based on ingredient results, even though skin absorption may have an effect. The Human Repeated Insult Patch Test may be used to confirm the safe use of a potentially sensitizing substance. The test dose is normally at the upper end of the suggested use range and is below any dose giving positive results in animal tests (Basketter et al., 2005).

The absence of animal testing for finished cosmetics can be at least partly attributed to the EU's ban on animal testing for any finished cosmetic products marketed in the EU. The ban, which took effect in 2004, spurred the development of alternative, non-animal methods. Given that alternative methods are now widely accepted in the US for finished cosmetic products, are cost competitive with animal tests, and are the only option for companies selling into the EU, no incentive exists for using animal tests for finished cosmetics. In fact, the disincentives are enough that use of animals in finished cosmetic product testing would be atypical.

3 Testing under the proposed Act

Under the proposed Safe Cosmetics and Personal Care Products Act of 2013, safety evaluations would be expanded for all existing and new ingredients. The FDA would determine for each ingredient an allowable exposure that meets the proposed safety standard: "*not more than a one in a million risk of any adverse health effect*" or "*shown to produce no adverse health effects, incorporating a margin of safety of at least 1,000*" (sections 614[a][2][A] and [B]).

As noted in the Introduction, the Act requires the FDA to consider whether the substance is "*a known or suspected neurological or immunological toxicant, respiratory asthmagen, carcinogen, teratogen, or endocrine disruptor, or have other toxicity concerns, including reproductive or developmental toxicity*" (section 616[a][2][D]). The Act defines *reproductive or developmental toxicity* as "*… can contribute to biologically adverse effects on the development of humans or animals, including effects on the female or male reproductive system, the endocrine system, fertility, pregnancy, pregnancy outcomes, or modifications in other functions of the body that are dependent on the integrity of the reproductive system as well normal fetal development*" (section 611[10]).

The Act clearly identifies a requirement to determine, for each ingredient, an exposure scenario that will "*produce no adverse health effects*." Currently, repeated dose studies are used for the derivation of the *no observed adverse effect level* (NOAEL), a key parameter for this determination. The preferred study is the subchronic (90-day) toxicity study (SCCS, 2010; SCCP, 2012), because it allows a more precise estimate of the NOAEL than can be obtained from a subacute (28-day) study. In fact, any adverse effect recorded in other long term studies, for example, studies to assess carcinogenicity potential or reproductive toxicity, would prevent the ingredient from being used in cosmetic products. Under the Act, therefore, we can assume that all ingredients would need the subchronic toxicity study in order to calculate a more precise NOAEL. Currently, the subchronic toxicity study is typical only for regulated ingredients, which are about 2% of all ingredients, so new subchronic toxicity studies likely would be needed for most ingredients.

The proposed Act does not specify other tests by name; it specifies only the health endpoints that must be considered when determining the NOAEL, as quoted above. The following section looks at other toxicity tests that may be considered in order to meet this requirement.

3.1 Potential new tests under the proposed Act

OECD test methods are internationally accepted (including by the US) standard methods for chemical safety testing. All OECD test guidelines can be downloaded for free at the OECD website (http://www.oecd.org/env/ehs/testing). If additional evaluations are needed for the health endpoints specified in the Act, they would likely involve one or more of the following OECD test methods:

– *Genotoxicity:* Genotoxicity is usually performed *in vitro*. In some cases, further, *in vivo* studies are requested if results are suspicious. Considering that cosmetic ingredients are chemicals with a very low toxicological profile, in depth investigation is usually not necessary. The European Commission Services' working group to review the status of genotoxicity testing for cosmetics concluded that *in vivo* tests for cosmetics are not relevant (Maurici et al., 2005).
– *Repeated Dose (Subacute/28-day or Subchronic/90-day) Toxicity Studies:* These evaluate toxic effects on organs and organ systems, including liver, kidneys, lungs, reproductive organs, central nervous system, hematopoietic system (lymph nodes and bone marrow), immune system, and endocrine system. As noted previously, repeated dose studies also allow the development of NOAELs, which would be needed under the proposed Act in order to establish the allowable exposures. For cosmetics, guidelines recommend the subchronic (90-day) study, rather than the subacute (28-day) study, because the subchronic study is better for determining a NOAEL. The Vinken et al. (2011) review of 220

dossiers for EU-regulated cosmetic ingredients found 54% had the subchronic study and often included a subacute study as well. Only 5% had just a subacute study. This suggests the subchronic study is the norm.

– *Prenatal Developmental Toxicity Study, sometimes called Teratology Study (OECD TG 414):* This study evaluates implantation/resorption, embryonic development, fetal growth, and morphological variations and malformations. According to the Vinken et al. (2011) report, 56% of EU-regulated cosmetic ingredients had this study. For general toxicological assessments, some negative results from OECD TG 414 performed on rats are confirmed in a second non-rodent species, usually rabbit. This is not the case for cosmetic ingredients, which are discarded if there is a suspicion of a possible toxic effect on the reproductive system. For cosmetic ingredients, the negative result from OECD TG 414 is supported by a mild toxicological profile of the test item and confirmation in a second species is deemed redundant.
– *Two-generation Reproduction Toxicity Study (OECD TG 416) or Extended One-generation Reproductive Toxicity Study (OECD TG 443):* Both tests evaluate the male and female reproductive systems, fertility, pregnancy outcomes, and growth and development of the offspring. These tests may also be expanded to include segments for neurotoxicity, and OECD TG 443 may also include a segment for immunotoxicity. OECD TG 443 is a relatively recent adaptation of OECD TG 416 that uses fewer animals, but it is about 50% more expensive than OECD TG 416 and requires more lab capacity (CEHTRA, 2012), so the more established OECD TG 416 may be preferred. This study has typically been performed only as part of a second tier of testing, on a case-by-case basis. In the Vinken et al. (2011) study, 5% of EU-regulated cosmetic ingredients had this study. OECD TG 416 and 443 are the most animal-intensive of the potential tests, using about 1,000 animals for OECD TG 443 and about 3,000 animals for OECD TG 416 (see Section 2.1.4). They are also among the most time-consuming and expensive of tests, requiring 1 to 2 years and costing about $ 430,000 for OECD TG 416 and $ 650,000-$ 1,000,000 (depending on modules included) for OECD TG 443.
– *Reproduction/Developmental Toxicity Screening Test, OECD TG 421 or OECD TG 422:* Possible alternatives to OECD TG 416 and 443 are the reproduction/developmental toxicity screening tests, OECD TG 421 or OECD TG 422. OECD TG 422 combines the screening test with a subacute (28-day) repeated dose toxicity test; OECD TG 421, which is just the screening test, is becoming obsolete. These tests use about 520 animals, including pups (760 if satellite animals are counted). OECD TG 421 and 422 do not provide all information specified in the Act, nor do they provide evidence for definite claims of no effects. In the EU's REACH program, the screening tests are not considered sufficient: "*DNEL derived from a screening test for reproductive/developmental toxicity shall not be considered appropriate to omit a prenatal developmental toxicity study or a two-generation reproductive toxicity study*" (EC, 2009b). Still, these have been considered adequate for assessing reproductive toxicity for the OECD Screening Information Data Set (SIDS) for chemical safety and also for low-production volume chemicals under REACH. In fact, even high-volume chemicals under REACH sometimes substituted OECD TG 422 for OECD TG 414 and 416, combining OECD TG 422 with existing data in a weight of evidence approach, even though OECD TG 414 and 416 are explicitly required for these chemicals under REACH. This analysis, therefore, will consider the screening tests as possible alternatives when evaluating different scenarios.
– *Combined Chronic Toxicity (>1 year)/Carcinogenicity Studies (OECD TG 453):* This test evaluates carcinogenicity and other possible health hazards likely to arise from repeated exposure for a period lasting up to the entire lifespan of the species used. There is a stand-alone carcinogenicity study (OECD TG 451), but the combined chronic toxicity/carcinogenicity study may be preferred (for example, in the EU REACH Regulations) to reduce the overall number of animals used for testing. In the Vinken et al. (2011) study, 7% of EU-regulated ingredients had the carcinogenicity study and 8% had the chronic toxicity study. However, it should be considered whether those tests were performed for cosmetic purposes only. Some cosmetic ingredients may also have application in the pharma industry, which requires extensive testing. This study is the most expensive of the tests, at about $ 1.2 million and 416 animals for the carcinogenicity test (480 animals for the combined test), and requiring 1 to 2 years per test.
– *Developmental Neurotoxicity Study (OECD TG 424 and 426):* This is a stand-alone test, but similar information can be obtained from OECD TG 416 and 443 if they include a neurotoxicity segment. This analysis assumes that, if neurotoxicity information is needed, it would be included as part of OECD TG 416 or 443 to minimize the number of animals used. OECD TG 424 or 426, therefore, is not included in any of the scenarios for this analysis.

Current practice is to conduct these tests only on a case-by-case basis, taking into consideration the physico-chemical properties of the ingredient and experience. Under the proposed Act, at least a repeated dose toxicity evaluation would become necessary for all ingredients in order to determine the NOAEL. How other tests would be considered cannot be known without implementing regulations, but this analysis looks at a range of possibilities, from a minimum case of adding just the subchronic toxicity study to cases that would include more intensive studies, such as the prenatal developmental toxicity study.

As noted, the two-generation reproductive toxicity test, extended one-generation reproductive toxicity test, and carcinogenicity test are much costlier than other tests, in terms of monetary cost, time, and animal lives. Assumptions about the frequency of these tests, therefore, are the most critical assumptions in this analysis. The choices related to these tests, both in this analysis and in any legislation and regulations, will dominate the impacts.

For this analysis, we have assumed that implementing regulations would minimize the use of the most cost-intensive, animal-intensive tests, such as the two-generation reproductive toxicity test. We note, however, that the Act could be interpreted much more stringently than assumed in this paper, leading to higher animal use and costs than presented here.

3.2 Potential testing scenarios

This section explores a range of possible testing scenarios under the proposed Act, from a minimum case in Case 1 that would involve the current first tier evaluations plus the subchronic toxicity study, to a maximum case in Case 4, in which ingredients undergo further second-tier testing. The potential scenarios are based on current evaluation programs in the US and EU.

The proposed Act would require all ingredients to have the same safety evaluation. It does not make distinctions based on toxicological profile, biological activity, volume, or other factors. The following potential testing scenarios are analyzed here:

Case 1: Subchronic toxicity study

In this case, all ingredients would be evaluated for the first tier endpoints that are typical practice currently (see Tab. 1). In addition, all ingredients would be evaluated for subchronic toxicity to evaluate systemic toxicity and the NOAEL.

Case 2: SIDS with screening studies

In this case, all ingredients would be evaluated for the Screening Information Data Set (SIDS). This is an internationally accepted (including by the US), minimum test set for evaluating chemical properties, including mammalian toxicology. The reproductive/developmental screening study (OECD TG 421 or 422) would provide the only data on reproductive and developmental toxicity. A full reproductive toxicity study (OECD TG 416 or 443) would not be required; nor would the FDA require the full prenatal developmental toxicity study (OECD TG 414). This case is based on the guidelines in *OECD Manual for the Assessment of Chemicals,* Section 2.2, "The Screening Information Data Set (SIDS)" (OECD, 2004a).

Case 3: SIDS with prenatal study

The SIDS requirement for the reproductive/developmental toxicity health endpoint can also be met with the full prenatal developmental toxicity study (OECD TG 414), along with the subchronic (90-day repeated dose) toxicity test if the subchronic test examines the reproductive organs and sufficiently documents the examination. Because this case includes the full prenatal developmental toxicity study, it may better meet the requirements of the proposed Act than would Case 2. Like Case 2, Case 3 is based on the guidelines in OECD *Manual for the Assessment of Chemicals,* Section 2.2, "The Screening Information Data Set (SIDS)" (OECD, 2004a).

Case 4: SCCS guidelines

All ingredients would be evaluated to the same degree as the 2% of cosmetic ingredients that currently have specific evaluation requirements in the SCCS guidelines (SCCS, 2010). For this case, the analysis uses these guidance documents:

– *The SCCS's Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation* (SCCS, 2010). This gives the EU guidelines for evaluating ingredients that are specifically regulated (for example, UV filters and colorants). The FDA recommends this document as guidance for safety testing of ingredients (FDA, 2012).
– *Draft Decision on Guidelines Annex I: Commission Implementing Decision on Guidelines on Annex I to Regulation (EC) No. 1223/2009 on cosmetic products* (SCCP, 2012). This is the EU's latest document on cosmetics safety testing.

Other guidelines considered in this analysis for all cases were:

– *Draft Guidance for Industry: Safety of Nanomaterials in Cosmetic Products* (FDA, 2012). This gives insight into the FDA's current thinking on appropriate tests for cosmetics safety.
– *Guidance on Information Requirements and Chemical Safety Assessment,* Chapter R.7a-c, "Endpoint Specific Guidance" (ECHA, 2012). Guidance for the implementation of the EU's Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) law, which took effect June 2007.
– *OECD Guidelines for Testing of Chemicals,* Section 4: Health Effects Guidelines. These are internationally agreed-upon testing methods for characterizing potential health effects of chemicals.
– *OECD Guidance Document for Mammalian Reproductive Toxicity Testing and Assessment* (OECD, 2008) and *OECD Guidance Document for Neurotoxicity Testing* (OECD, 2004b). These documents give detailed testing protocols and recommendations for these health endpoints.

3.3 Estimated number of cosmetic ingredients

The European Union has compiled the Cosmetic Ingredient (CosIng) database, a comprehensive list of known cosmetic ingredients and their functions. The URL for this database is http://ec.europa.eu/consumers/cosmetics/cosing/.

This analysis used two different types of CosIng searches to find the total number of cosmetic ingredients: an "Inventory" search and an "All" search.

– *Inventory search*
  For this search, the search parameters were Version: *Cosmetics Directive*, Scope: *Inventory* (ingredient and fragrance), and Status: *Active*
  This search returns cosmetic ingredients by their standard INCI names, including fragrances, colorants, and preservatives. It excludes ingredients prohibited under Annex II of the EU's Cosmetics Regulation, the legislation governing cosmetics in the EU. If an ingredient does not have an INCI name, it does not appear to be returned by this search. (Ingredients without an INCI name are called *substances* in CosIng. *Ingredient* is reserved for ingredients that have an INCI name.) The total here should slightly underestimate the total number of cosmetic ingredients, because not all cosmetic ingredients have INCI names (most do). The "All" search described next catches all ingredients, even those without INCI names.
– *All search*
  For this search, the search parameters were Version: *Cosmetics Directive*, Scope: *All*, and Status: *Active*
  This search returns all cosmetic ingredients, including

those without INCI names and those that are prohibited under Annex II of the EU's Cosmetics Regulation. To obtain the number of non-prohibited ingredients, the prohibited ingredients are subtracted.

In both searches, the Search field was left blank to display the full database. Duplicates were identified by sorting the database, looking for duplicate names or ID numbers (CAS or EC numbers). Not all duplicate ID numbers denote duplicate substances, since the same ID number may apply to related but different substances. Each set of duplicate IDs was evaluated case-by-case to determine duplicate ingredients.

The results of the searches are shown in Table 9. They are similar and place the number of cosmetic ingredients at about 20,000 in October 2012, when the searches were performed. A more recent Inventory search, in September 2013, shows this number is now 20,790.

The proposed Safe Cosmetics Act requires evaluations of ingredients that are in use in cosmetics. Not all ingredients listed in CosIng are used in cosmetics at a given time, and those that are used are constantly in flux given the frequency with which new products are created and existing ones reformulated. The number of ingredients in use will be less than the total number, but probably not by much. One online database identified about 9,000 ingredients among about 70,000 products. Given that Amazon alone lists more than 500,000 different cosmetic products, most of the CosIng ingredients are likely in use in some product over the 10-year period of this analysis. What we can say with some certainty is that the percentage in use is more than 50% and less than 100%. For the analyses here, we use a conservative 70%.

Chemicals often are categorized by production volume in regulatory programs. The EU's REACH regulations catego-

Tab. 9: Number of ingredients returned by CosIng searches

CosIng Database Search (accessed Oct 18, 2012)	No. of Cosmetic Ingredients	Duplicates	Prohibited Annex II Ingredients	Total Unique Cosmetic Ingredients
"Inventory" search	19,817	0	0	19,817
"All" search	21,950	306	1478	20,166[1]

[1] calculated as: Search Returns - Duplicates - Prohibited Annex II Ingredients

Tab. 10: Estimated percentages of cosmetic ingredients by production volume

	Number of REACH Substances, Excluding Intermediates[1]		Cosmetic Ingredients Evaluated under REACH[2]		
Tonnage per Manufacturer/ Importer	No. of REACH Substances[3,4]	% of Total	Number of Cosmetic Ingredients	Source	% of Total
>1,000 tons/yr	2147	10%	447	CosIng/REACH	2.2%
100-1,000 tons/yr	1954	9%	270	CosIng/REACH	1.3%
10-100 tons/yr	3952	18%	5353	back-calculated	25.75%
1-10 tons/yr	13,895	63%	14,720	back-calculated	70.75%
Total	21,948	100%	20,790	CosIng	100%

[1] From Pedersen et al. (2003). The European Union used these estimates of numbers of REACH substances per tonnage range to estimate testing requirements under REACH. Substances >1,000 tons/yr and 100-1,000 tons/yr have now been registered under REACH, and the numbers correspond reasonably well with the original estimates. For example, the number of non-intermediate substances registered >1,000 tons/yr is about 2400; Pedersen et al. estimated those to be 2147. The number of non-intermediates registered in the 100-1,000 tons/yr range as of August 2013 is about 1600; Pedersen et al.'s estimate is 1954. The REACH registration deadline for this range was May 2013, and not all substances have been entered in the ECHA database yet. The reasonable correspondence of the Pedersen et al. estimates for these ranges gives confidence in the Pedersen et al. estimates for the other tonnage ranges. The deadline for registering tonnage ranges less than 100 tons/yr is later, so the actual numbers for those ranges are not yet available.

[2] For the derivation of these percentages, see Section 3.3.

[3] Pedersen et al.'s numbers for each tonnage range include REACH type 4 intermediates. Pedersen et al. calculated from existing chemical databases that intermediates were 20.6% of total substances, and they assumed this percentage applied to each tonnage range. Because the Safe Cosmetics and Personal Care Products Act would not apply to intermediates, this table excludes the 20.6% attributed to intermediates from each of the tonnage ranges; therefore, the numbers, here are 20.6% less than the numbers in Pedersen et al.'s original table.

[4] The table here does not include Pedersen et al.'s estimate for REACH type 3 intermediates, since the Safe Cosmetics and Personal Care Products Act would not apply to these intermediates either. Pedersen et al. included these as a separate line item, so that line item was deleted here.

rize chemicals by the following production volumes: >1,000 tons/year, 100 to 1,000 tons/year, 10 to 100 tons/year, and 1 to 10 tons/year, with different test requirements for each range. Higher production volume chemicals typically have more safety data, so it is helpful to estimate the number of cosmetic ingredients by production volume. The EU's REACH and CosIng databases let us do that:

– *>1,000 tons/year*
  The CosIng ingredients were checked against the REACH substance database, which gives the production volume range for each substance. For tonnage >1,000/year, 447 cosmetic ingredients were found in the REACH database. This is 2.2% of total cosmetic ingredients. All substances produced in volumes greater than 1,000 tons/year were required to be registered under REACH by November 30, 2010, so this is a near complete inventory.
– *100 to 1,000 tons/year*
  This tonnage range was required to be registered under REACH by May 31, 2013. The REACH database has not yet been updated with all of the new dossiers. The method was to check the CosIng ingredients against the substances registered to date under REACH, and then to extrapolate to the total, assuming the ultimate number of REACH substances in this range is near the original estimate of 1,954 (Pedersen et al., 2003). This resulted in an estimate of 270 ingredients, or 1.3% of total cosmetic ingredients.
– *10 to 100 tons/year and 1 to10 tons/year*
  In a 2003 study evaluating REACH, Pedersen et al. (2003) estimated the percentages for these REACH tonnage ranges were 18% and 63%, respectively. Adding these to the now-known percentages of the other tonnage ranges totals to about 84.5%, leaving 15.5% of ingredients in these lower ranges unallocated. For this analysis, the unallocated 15.5% are evenly assigned to the two lower tonnage ranges. For 10 to 100 tons/year, this gives 18% + 7.75% = 25.75%. For 1 to 10 tons/year, this gives 63% + 7.75% = 70.75%. The estimated number of ingredients can then be back-calculated by multiplying the percentage by 20,790, the known total number of ingredients in September 2013. This gives 5,353 and 14,720 ingredients, respectively, for these tonnage ranges.

Table 10 shows the results, indicating that perhaps over 95% of cosmetic ingredients have relatively low production volume, less than 100 tons/year.

At the moment (August 2013), about 9,900 unique substances have been registered according to the REACH regulation for substances >1,000 tons/year and 100-1,000 tons/year. Even though this number includes intermediates and some new chemicals, it clearly exceeds Pedersen et al. (2003)'s original estimate for these tonnage ranges. However, for non-intermediate substances, a category that includes all cosmetic ingredients, Pedersen et al.'s estimates are closer to the real number of REACH registrations, giving more confidence in these numbers.

Note that REACH does not include substances with a production volume less than 1 ton/year, categories that may include many cosmetics ingredients (consider that 30% of ingredients in the CosIng database are plant ingredients). For this analysis, ingredients less than 1 ton/year are included in the 1-10 ton/year range for simplicity; it has no effect on the calculations.

Also, REACH does not include substances of natural origin, defined in REACH as naturally occurring substances as such or extracted through specified means. Only room temperature water is allowed as a solvent. All other substances, including those extracted by solvents or by vapor must be registered if their production volume is more than 1 ton/year. Cosmetics may include food additives, which are evaluated separately from REACH but with a very similar approach.

The production volume is also important because low-volume ingredients would be disproportionately affected by the cost of additional evaluations and testing. For very low volume ingredients, the ingredient manufacturer may choose to discontinue the ingredient rather than incur the cost of the evaluation. Estimating this effect is not possible, except to say that the number of ingredients evaluated could be less than the total number shown in the CosIng database, because marginally profitable ingredients could be discontinued in response to the Act. Since most cosmetic ingredients have small production volumes, this effect could be significant.

3.4 Impact of existing data for cosmetic ingredients

Most ingredients should have existing test results for the first tier tests that are routinely run on cosmetic ingredients: eye and skin irritation, skin sensitization, mutagenicity, and acute toxicity. To estimate existing test results for second tier evaluations, which are mainly repeated-dose tests, the following sources provide primary data:

– RPA & Statistics (2002) estimated the amount of existing chemical data based on questionnaires sent to chemical manufacturers/importers and to their associations. The questionnaire asked about availability of complete data sets for the tests required under REACH. Because the questionnaire asked about complete data sets only, it does not account for chemicals that have partial data sets. As RPA & Statistics notes, this means the results underestimate the amount of existing test data. Still, these results are useful as a lower bound for the amount of existing test data (see Tab. 11).
– ECHA (2011), responsible for implementing the EU's REACH program, reported statistics for existing tests submitted to meet REACH requirements. ECHA analyzed dossiers for 1,504 substances with production volumes over 1,000 tons/year and for 218 substances with production volumes of 100 to 1,000 tons/year. The analysis indicates the amount of existing data is higher than estimated in RPA & Statistics, as expected. The ECHA numbers, however, likely overestimate the amount of existing data, because the initial ECHA review of the data indicated *"…the experimental data provided in the dossiers are in some cases also not sufficient to meet information requirements under REACH*" (ECHA, 2011). The ECHA statistics, therefore, can be taken as an upper bound on the amount of existing data (see Tab. 11).

– Bishop et al. (2012) reviewed the US EPA's High Production Volume (HPV) Challenge Program, which was a voluntary program for industry to provide a SIDS data set for high production volume chemicals. The data set included acute, subacute, and subchronic toxicity tests; the prenatal developmental toxicity study (OECD TG 414) and one-generation reproductive toxicity study (OECD TG 415); and the reproductive/developmental toxicity screening tests (OECD TG 421 and 422). Bishop et al. (2012) found that industry had existing data for 27% of health endpoints in that program. Their report did not break out the percentages of existing data for each endpoint, but this percentage lies between the lower and upper bounds from the other two reports and generally supports this as a reasonable range for ingredients produced in volumes >1,000 tons/year.

Under REACH, all chemicals produced in volumes greater than 100 tons/year must be evaluated for subchronic toxicity and prenatal developmental toxicity; and those greater than 1,000 tons/year must undergo evaluations for two-generation reproductive toxicity and carcinogenicity as well. REACH

Tab. 11: Estimate of existing data for health endpoints

		Chemicals with Existing Data	
Health Endpoint	**Production Volume of Chemical**	**RPA & Statistics (2002)[1]**	**ECHA (2011)[2]**
Subchronic Oral Toxicity (90-day Repeated Dose)[3]	>1,000 tons/yr	7%	43%
	100-1,000 tons/yr	7%	45%
	1-10 tons/yr, 10-100 tons/yr	3%	N/A
Subchronic Dermal Toxicity (90-day Repeated Dose)[3]	>1,000 tons/yr	–	9%
	100-1,000 tons/yr	–	3%
	1-10 tons/yr, 10-100 tons/yr	–	N/A
Subchronic Inhalation Toxicity (90-day Repeated Dose)[3]	>1,000 tons/yr	7%	45%
	100-1,000 tons/yr	7%	30%
	1-10 tons/yr, 10-100 tons/yr	3%	N/A
Prenatal Developmental Toxicity	>1,000 tons/yr	7%	42%
	100-1,000 tons/yr	7%	44%
	1-10 tons/yr, 10-100 tons/yr	3%	N/A
Two-generation Reproduction Toxicity	>1,000 tons/yr	5%	32%
	100-1,000 tons/yr	5%	N/A
	1-10 tons/yr, 10-100 tons/yr	2%	N/A
Reproductive/Developmental Toxicity Screening	>1,000 tons/yr	22%	N/A
	100-1,000 tons/yr	22%	30%
	10-100 tons/yr	17%	Not yet available
	1-10 tons/yr	17%	N/A
Chronic Toxicity/Carcinogenicity	>1,000 tons/yr	5%	39%
	100-1,000 tons/yr	5%	56%
	1-10 tons/yr, 10-100 tons/yr	2%	N/A

[1] Derived from RPA & Statistics (2002), Table 5.2. The percentages above can be inferred from that table.

[2] Derived from ECHA (2011), Table 4. That table includes the number of existing experimental studies for each endpoint. To derive the percentages here, the number of existing studies is divided by the total number of substances analyzed; for example, the subchronic dermal endpoint had 129 existing test results for 1504 substances, or 9%. Some substances had multiple studies for the same health endpoint. For these, the percentage above is the percentage of existing studies divided by the total number of records for substances in that tonnage range; for example, the subchronic oral toxicity endpoint had 2365 records for 1504 substances; among these records were 1025 existing test studies. The percentage was calculated as 1025/2365, or 43%.

[3] RPA & Statistics (2002) does not specify an exposure route, but usually oral with perhaps inhalation as a second test are run, so the percentages are entered only for the oral and inhalation tests.

N/A: Not applicable. REACH does not require these evaluations for these production volumes.

Tab. 12: Estimated percentages of existing data, including if REACH data are also available

	>1,000 tons/yr (450 cosmetic ingredients)		100-1,000 tons/yr (270 cosmetic ingredients)		10-100 tons/yr (5,350 cosmetic ingredients)		<10 tons/yr (14,720 cosmetic ingredients)	
Health Endpoint	**Existing Data[1]**	**REACH Data (includes Existing)**	**Existing Data[1]**	**REACH Data (includes Existing)**	**Existing Data[1]**	**REACH Data (includes Existing)**	**Existing Data[1]**	**REACH Data[2] (includes Existing)**
Subchronic oral toxicity (90-day repeated dose)	7% - 43%	100%	7% - 45%	100%	3% - 27%	3% - 27%	3% - 27%	3% - 27%
Subchronic dermal toxicity (90-day repeated dose)	0% - 9%	100%	0% - 3%	100%	0% - 3%	0% - 3%	0% - 3%	0% - 3%
Subchronic inhalation toxicity (90-day repeated dose)	7% - 45%	100%	7% - 30%	100%	3% - 27%	3% - 27%	3% - 27%	3% - 27%
Chronic toxicity (>1 year)/carcinogenicity	5% - 39%	100%	5% - 56%	5% - 56%	2% - 27%	2% - 27%	2% - 27%	2% - 27%
Two-generation reproductive toxicity	5% - 32%	100%	5% - 27%	5% - 27%	2% - 27%	2% - 27%	2% - 27%	2% - 27%
Prenatal development toxicity	7% - 42%	100%	7% - 44%	100%	3% - 27%	3% - 27%	3% - 27%	3% - 27%
Reproductive/ developmental screen	22% - 27%	22% - 27%	22% - 30%	100%	17% - 27%	100%	17% -27%	17% -27%
Developmental neurotoxicity	unknown	100%	unknown	100%	unknown	unknown	unknown	unknown

[1] Lower value is from RPA & Statistics (2002); upper value is from ECHA (2011). Where data from ECHA (2011) do not exist for the upper value, the more general estimate of 27% from Bishop et al. (2012) is used. See the text and Table 11 for details. The exception is for dermal toxicity, which ECHA (2011) indicates has only 3% existing data for the 100-1,000 ton/yr range; this 3% is used here as the upper value for ranges less than 100 tons/yr, too.

[2] REACH does not require these tests for this production volume range; therefore, it will provide no additional data for this range.

requires a comprehensive data set only for the highest production volume substances, those over 1,000 tons/year. An important difference between the scope of REACH and the proposed Safe Cosmetics Act is that REACH addresses all chemicals, and a carcinogenic substance may still be used provided that the risk is sufficiently managed and no contact with consumers will occur. On the other hand, cosmetics are by definition used in direct contact with consumers. Consequently, some in depth investigations that are required by REACH for chemicals of concern do not apply to cosmetic ingredients simply because when a higher risk is suspected, that ingredient is not used.

Theoretically, REACH studies could provide good data sets for ingredients in the 100-1,000 ton/year and >1,000 ton/year ranges. However, most of those existing studies are proprietary, and the owner usually charges for a Letter of Access at a cost that is often more than half the cost for repeating it. If REACH data are available for cosmetic safety assessments, the REACH data could be sufficient for the higher production volume cosmetic ingredients, about 3.5% of total cosmetic ingredients (see Tab. 10 in the preceding section). For lower production volume ingredients, the REACH data could provide partial data sets. Table 12 indicates the estimated percentages of existing data without and with REACH.

For existing ingredients, this analysis will use the simple assumption that 15% of ingredients have existing data for repeated dose tests. A 15% estimate is the midpoint of the range in Table 12 for ingredients in the 0-10 tons/year and 10-100 tons/year ranges, which is 95% of cosmetic ingredients. The main exception is the dermal exposure route, for which data appear to be almost non-existent (0%-3%). Oral tests can be extrapolated to the dermal route, however, so for simplicity's sake, we use 15% for the dermal exposure route, too. Keeping the assumption simple also acknowledges that choosing any number is a guess, but at least a guess within an expected range.

REACH data could potentially provide 100% of data for some endpoints, but only for about 3.5% of ingredients, so we do not consider it further here. The exception is for the reproductive toxicity screening test, for which REACH could provide data for ingredients >10 tons/year, which is about 30% of ingredients.

For new ingredients, the analysis also assumes that 15% will have available data, because some new ingredients will originate from the pharmaceutical and food industries and be fully characterized.

For phototoxicity, we assume 0.1% of existing and new ingredients have available data, because that is the percentage of ingredients that are UV absorbers. UV absorbers are regulated ingredients in the US and EU and generally undergo this testing. Other ingredients normally are not tested for phototoxicity.

In summary, we assume the following percentages for available data:

– Existing ingredients: 100% for all first tier endpoints, 30% for the reproductive toxicity screening study, 0.1% for the phototoxicity study, and 15% for other endpoints.
– New ingredients: 15% for all endpoints except phototoxicity, which is 0.1%.

3.5 Estimation methods and waivers that can reduce testing

As described in Section 2.1.2, read-across and weight of evidence methods can reduce the need for testing. Now that REACH has been in force several years, statistics are available for the use of read-across and weight of evidence techniques in actual practice (ECHA, 2011). Table 13 shows the percentages reported to ECHA under REACH for the >1,000 tons/year range and 100-1,000 tons/year range. Note that ECHA has stated that estimation techniques may not have been appropriately applied to some REACH chemicals, so the percentages in Table 13 may be high, including invalid use of the techniques (ECHA, 2011).

For existing ingredients in this analysis, we use the average (midpoint) of the ECHA percentages in Table 13 for the two tonnage ranges, but we note that this likely overestimates the use of estimation methods. For new ingredients, we assume no estimation techniques are used for repeated dose tests, based on Adler et al. (2010) (see Section 2.1.2).

Waivers may also reduce the number of tests needed. For cosmetics ingredients, the only basis for a waiver probably would be that an exposure route does not apply. *Exposure routes* are the ways by which humans may be exposed to a substance: by swallowing it (the oral route), by skin contact (the dermal route), or by inhaling it (the inhalation route). The main exposure route for most cosmetic ingredients is dermal, although oral and inhalation are significant routes for some ingredients.

The Vinken et al. (2011) review of 220 cosmetic ingredient dossiers found the only exposure route tested was the oral exposure route. This reflects current standard practice in the EU, where the oral route is the primary exposure route tested for cosmetic ingredients. In the US, the choice of exposure route is less documented. US regulators generally suggest that toxicity tests use the exposure route that corresponds to the human exposure route (Nohynek et al., 2010), but test data are usually not public, so whether this suggestion translates into practice is unknown.

Rovida and Hartung (2009) estimate waivers for acute oral, inhalation, and dermal toxicity to be 10%, 50%, and 60%, respectively, assuming that oral is the main route tested. Each route includes waivers for 10% of substances for which no human exposure is likely or which are already known to be toxic and so require no further testing. For cosmetics and

Tab. 13: Use of read-across and weight of evidence to reduce testing under REACH

Health Endpoint	% Substances Evaluated through Read-Across[1]	% Substances Evaluated through Weight of Evidence[1]	Total
Subchronic oral toxicity (90-day repeated dose)	45%, 43%	8%, 5%	53%, 48% (51% avg)
Subchronic dermal toxicity (90-day repeated dose)	37%*, 58%*	4%, 8%	41%, 66% (53% avg)
Subchronic inhalation toxicity (90-day repeated dose)	40%, 39%	6%, 6%	46%, 45% (45% avg)
Chronic toxicity (>1 year)	33%, 33%	15%, 16%	48%, 49% (49% avg)
Carcinogenicity	28%, 22%	10%, 6%	38%, 28% (33% avg)
Two-generation reproductive toxicity	24%, N/A[2]	12%, N/A[2]	36%
Prenatal developmental toxicity	30%, 30%	11%, 5%	41%, 35% (38% avg)
Reproductive/developmental toxicity screening test	N/A, 24%[2]	N/A, 10%[2]	34%

[1] The first percentage in the pair is for substances >1,000 tons/yr; the second percentage is for substances 100-1,000 tons/yr. An asterisk (*) means the sample size was so small that the percentage may not reflect the true situation. N/A, not applicable: REACH does not require this test for this range.

[2] ECHA (2011) does not indicate to which reproductive toxicity tests these percentages pertain. The ECHA text refers to both the two-generation test and the screening test, but does not break these out separately in the ECHA tables. For substances >1,000 tons/year, the two-generation study is required, and for substances 100-1,000 tons/year, the screening test is required, so for this analysis, we have assumed the percentages for substances >1,000 tons/year are for the two-generation study, and the percentages for substances 100-1,000 tons/year are for the screening study. The percentages for both ranges are nearly identical, so the assumption does not skew the results.

cosmetic ingredients, this 10% for waivers does not apply, because (1) human exposure is a given for cosmetics and (2) cosmetics do not contain substances already known to be toxic. The ingredients are available precisely because they are already believed to be safe, having already undergone some toxicity testing. Subtracting this 10% from the waivers leaves 0%, 40%, and 50% waivers, respectively, for the oral, inhalation, and dermal routes. For this analysis, we assume 0% for oral and 50% for dermal. For the inhalation route, however, we assume 95% waivers. Inhalation tests for cosmetics typically are performed only in exceptional cases, due to their complexity, cost, and more intensive animal use. In spite of the relevance of this administration route, we assume inhalation evaluations would be extrapolated from other exposure routes in the majority of cases.

Regarding repeated dose studies, Pedersen et al. (2003) estimate that 10% of repeated dose tests may be waived, again, for substances for which no human exposure is likely. That estimate has been recently revised upward for reproductive and developmental toxicity only (Rovida et al., 2011), confirming that use of waivers was significant in the REACH registration process. As previously noted, this type of waiver does not apply to cosmetic ingredients, because human exposure is a given. Waivers may be allowed based on exposure route. This analysis assumes the same percentage of waivers for repeated dose toxicity tests as for acute toxicity tests: 0% for oral, 95% for inhalation, and 50% for dermal.

Note that oral studies require techniques to extrapolate the oral results to the dermal route (see, for example, IGHRC, 2006). The reliability of such extrapolations has been questioned (Nohynek et al., 2010); however, IGHRC notes that extrapolations must be used despite their limitations, given the predominance of oral data and practical and ethical constraints on conducting new tests for different exposure routes (IGHRC, 2006). There is general agreement that the situation is not ideal, and that it introduces another extrapolation in addition to the extrapolation of the animal result to a human result.

3.6 No non-animal alternatives for the newly proposed evaluations

The proposed Act includes the following language regarding alternative, non-animal tests: "… *the Secretary shall (1) require, where practicable, alternative testing methods that (A) do not involve the use of an animal to test the chemical substance; …(section 624[a][1])*".

The only tests for which alternative, non-animal methods are available are the tests that have already been completed for most substances: eye and skin irritation/corrosion, skin sensitization, and mutagenicity. Eye irritation *in vivo* and skin sensitization according to OECD TG 429 (LLNA) soon can be fully replaced by *in vitro* alternatives. Recently, ECVAM endorsed the NT3 cytotoxicity method for the evaluation of non-toxic substances in the oral acute endpoint (EURL ECVAM, 2013). This endorsement may have a strong impact on acute toxicity testing for cosmetics as it is fully accepted for the recognition of non-toxic substances. For repeated dose evaluations, which use the most animals, no validated alternative methods are available.

3.7 Limits to testing: The Act's timeline and agency capacity

The proposed Act specifies a sequence for implementing the new evaluations (Safe Cosmetics and Personal Care Products Act of 2013, section 616):

Years 0 to 2: The FDA would conduct an initial review of all ingredients and place them in one of the following categories:

– Ingredients that meet the safety standard without any restrictions
– Ingredients that meet the safety standard, but with restrictions
– Ingredients that are prohibited
– Ingredients for which data are insufficient to make a determination

Nearly all ingredients would initially fall into the last category, because the expanded evaluations would be new for all but about 2.2%-3.5% of cosmetic ingredients: those that are already regulated by the EU (colorants, preservatives, UV filters, and hair dyes) or those for which REACH may have sufficient data.

Year 2: From the pool of ingredients without sufficient data, the FDA would create a priority list of 300 ingredients minimum and request additional information for those ingredients.

Year 3 and each year thereafter: The FDA would add another 100 ingredients minimum to the list for further evaluation each year, until all ingredients were placed on the list and evaluated.

This timeline would result in the evaluation of 1,200 ingredients by year 10. Currently, the number of ingredients in the CosIng database is about 20,800. By year 10, the number of CosIng ingredients will be about 26,800 if it continues at the pace of adding 500-700 new ingredients per year. If we assume only 70% of CosIng ingredients are used in cosmetics over the 10-year period (see Section 3.3), the total number of ingredients in use will be about 18,800. If the FDA evaluates the ingredients at the rate specified in the timeline, 1,200 over 10 years, this would be about 4% of total ingredients in year 10, or about 6% of ingredients in use.

Note that the priority list would never catch up with the number of ingredients. Instead, it would fall further behind each year, because the number of new ingredients would outpace the evaluations. In year 2, for example, the list would have about 21,700 ingredients remaining to be evaluated (21,700 = 20,800 existing ingredients + 1,200 new ingredients over 2 years – 300 ingredients added to the priority list). In year 10, it would have about 25,600 ingredients remaining to be evaluated (26,800 existing ingredients in year 10 – 1,200 ingredients added to the priority list in 10 years).

Presumably, the authors of the bill do not intend this. It is unclear what the timeframe should be, however, given key constraints:

– The capacity of existing laboratories to conduct the expanded testing.

– The capacity of the FDA to evaluate large numbers of ingredients each year.

Let us assume the laboratories can expand capacity to meet the demand, leaving FDA capacity as the main constraint. To evaluate all ingredients in use within 10 years (again assuming only 70% of potential ingredients are used), the FDA would need to evaluate about 1,900 ingredients each year. For comparison, the US EPA's intensive effort to reassess the safety tolerances for pesticide products evaluated 9,700 products within 10 years, or about 1,000 per year. As the EPA notes: "*This degree of success for such an ambitious, controversial and complex undertaking is unprecedented*" (US EPA, 2012). For the FDA to meet the 10-year goal would require about double that EPA effort.

In evaluating timelines, an important consideration is when alternative methods will be available for the remaining animal tests. Once alternative methods are available, the pace of evaluations can increase rapidly, because alternative methods for these endpoints are expected to involve high-throughput *in vitro* screening tests combined with computer models capable of assessing thousands of substances simultaneously. Bottini and Hartung (2009) and Adler et al. (2010) estimated that development and acceptance of alternative test methods would take at least 10 years. The scale of effort has also been compared with the Human Genome Project, which took 13 years. Efforts to develop alternative tests are already under way; therefore, this article assumes 13 years is a reasonable expectation for when the alternative methods could be available and validated. If the proposed Act becomes law, new testing likely would not begin until year 3 or year 4, leaving about 10 years until the alternative tests could replace the animal tests.

With this in mind, consider again the timeline as currently proposed in the Act. If ingredients are evaluated at the minimum rate specified in the Act, about 4% of total ingredients (6% of ingredients in use) would be evaluated within those 10 years. If alternative high-throughput methods are available near year 10, the remaining 96% could be evaluated rapidly through the new methods.

If the FDA achieves the high rate of evaluation that the US EPA accomplished, approximately 1,000 evaluations per year, this would enable the evaluation by year 10 of about 10,000 ingredients, which is about 37% of total ingredients in year 10 (or about half of all ingredients in use in year 10). Again, about year 10, the remaining 63% could be quickly evaluated with new methods.

3.8 Pulling it all together: Animal use and evaluation costs under the proposed Act

This section pulls together the information from Section 3 to estimate the animal use and cost for each safety evaluation scenario described in Section 3.2. The analysis uses the following assumptions, summarized in Table 14:

– The minimum number of ingredients evaluated over 10 years is 1,200, based on the Act's minimum requirement of 300 evaluations in the first year and 100 each year thereafter. The maximum number of ingredients evaluated is 10,000, based on expected maximum FDA capacity of 1,000 evaluations each year.
– New ingredients would be tested according to the new standard as part of their pre-market testing. Assuming 600 new ingredients per year, this is 6,000 new ingredients tested over 10 years. Note that although test results could be available for these ingredients, the FDA's capacity to evaluate the results and the required public comment periods would still be limiting factors, so this affects the calculation of the number of ingredients tested, but not the number of ingredients evaluated. The number evaluated is a maximum of 10,000 over 10 years, regardless of the number of ingredients with available test data.
– The potential testing scenarios are those described in Section 3.2.
– Existing data are available for 15% of new and existing ingredients for all endpoints except phototoxicity and the reproductive toxicity screening test. For phototoxicity, we assume 0.1% existing data. For the reproductive toxicity screening test, we assume 30% for existing ingredients and 15% for new ingredients (see Section 3.4 for details).
– Read-across and weight of evidence techniques are used at the percentages reported by ECHA (2011). The exception is repeated dose tests for new ingredients, for which we assume no estimation techniques would be used. Section 3.5 provides more detailed information. The cost for a read-across or weight of evidence evaluation is assumed to be about one-half the cost of the corresponding laboratory test (see Section 2.1.3, "Cost Per Test").
– Waivers for the oral and dermal exposure routes are used at the percentages discussed in Section 3.5. For the inhalation exposure route and for the two-generation reproductive toxicity test and the carcinogenicity test, a waiver of 95% is used, reflecting the rarity with which these are normally requested.
– The new evaluations would not necessarily be performed in the US; however, they are new evaluations that otherwise would not be performed. This calculation considers all new evaluations that would be triggered by the proposed Act, regardless of the location at which the evaluation is conducted.

Table 15 summarizes the tests for each scenario. Table 16 summarizes the calculations of total cost and animal use for each scenario.

Minimum

In our analysis, the minimum number of ingredients tested by year 10 is 7,200, which is the minimum 1,200 ingredients undergoing evaluation by the FDA plus 6,000 new ingredients undergoing the testing as part of their pre-market testing. For simplicity's sake, we assume that the FDA evaluations will be for existing ingredients (rather than new ingredients), but we recognize that some new ingredients could make the priority list, too, so there may be overlap between the 1,200 evaluations and the 6,000 new ingredients tested. The summaries in Table 16 break out these data to allow readers to test different assumptions.

Tab. 14: Assumed percentages for waivers, existing data, and estimation methods

		Global Parameters		Existing Ingredients				New Ingredients			
Health Endpoint	**OECD Test Method**	**% Waivers for this Endpoint[1]**	**% RA/WOE for this Endpoint[2]**	**% Ingredients with Existing Data[3]**	**% Ingredients with Waivers[4]**	**% Ingredients undergoing RA/WOE[5]**	**% Ingredients undergoing Test[6]**	**% Ingredients with Existing Data[3]**	**% Ingredients with Waivers[4]**	**% Ingredients undergoing RA/WOE[5]**	**% Ingredients undergoing Test[6]**
Eye irritation/corrosion	437, 438	0%	10%	100%	0%	0%	0%	15%	0%	9%	77%
Skin irritation/corrosion	430, 431, 439	0%	23%	100%	0%	0%	0%	15%	0%	20%	65%
Skin sensitization	429	0%	35%	100%	0%	0%	0%	15%	0%	30%	55%
Skin penetration	428	0%	30%	100%	0%	0%	0%	15%	0%	26%	60%
Phototoxicity	432	99.9%	30%	0.1%	99.8%	0.0%	0.1%	0.1%	100%	0.0%	0.1%
Mutagenicity/genotoxicity											
– Bacterial Reverse Mutation (Ames)	471	0%	34%	100%	0%	0%	0%	15%	0%	29%	56%
– Mammalian Cell Gene Mutation	476	0%	34%	15%	0%	29%	56%	15%	0%	29%	56%
– Mammalian Chromosomal Aberration Test	473	0%	34%	15%	0%	29%	56%	15%	0%	29%	56%
– Mammalian Cell Micronucleus Test	487	0%	34%	15%	0%	29%	56%	15%	0%	29%	56%
Acute toxicity – oral	420, 425	0%	30%	100%	0%	0%	0%	15%	0%	26%	60%
Acute toxicity – dermal	402	50%	31%	15%	43%	13%	29%	15%	43%	13%	29%
Acute toxicity – inhalation	403	95%	30%	15%	81%	1%	3%	15%	81%	1.3%	3%
Subchronic (90-day repeated dose) toxicity – oral	408	0%	50%	15%	0%	43%	43%	15%	0%	0%	85%
Subchronic (90-day repeated dose) toxicity – dermal	411	50%	50%	15%	43%	21%	21%	15%	43%	0%	43%
Subchronic (90-day repeated dose) toxicity – inhalation	413	95%	50%	15%	81%	2%	2%	15%	81%	0%	4%
Chronic toxicity (>1 year)	452	95%	50%	15%	81%	2%	2%	15%	81%	0%	4%
Carcinogenicity	451	95%	35%	15%	81%	1%	3%	15%	81%	0%	4%
Reproductive/developmental toxicity screening with subacute toxicity	422	0%	35%	30%	0%	25%	46%	15%	0%	0%	85%
Prenatal Developmental Toxicity	414	0%	35%	15%	0%	30%	55%	15%	0%	0%	85%
Two-generation Reproductive Toxicity	416	95%	35%	15%	81%	1%	3%	15%	81%	0%	4%
Toxicokinetics	no standard protocol										

Global parameters apply to both existing and new ingredients

[1] See Section 3.5 for waiver percentages. For the reproductive/developmental toxicity screening study (422) and the prenatal study (414), waiver is set to 0% because it is assumed if these are required, they would be required for all ingredients.

[2] RA/WOE, read-across or weight of evidence. See Table 4 in Section 2.1.2 for RA/WOE percentages for first-tier studies (irritation through acute toxicity) and Section 3.5 for percentages for second-tier studies.

[3] See Section 3.4 for existing data percentages.

[4] Calculated as (%Waivers for This Endpoint) X (1-%Ingredients with Existing Data).

[5] Calculated as (%RA/WOE for This Endpoint) X (1-%Ingredients with Existing Data) X (1-%Waivers for This Endpoint). For new ingredients, this is set to 0% for all repeated dose tests, because new ingredients rarely have enough studies from similar substances to allow RA or WOE approaches (see Section 2.1.2).

[6] Calculated as (1-%Ingredients with Existing Data) X (1-%Waivers for This Endpoint) X (1- %RA/WOE for This Endpoint).

The cost for the minimum number of ingredients, shown in Table 16, ranges from about $ 1.7 billion (case 2) to about $ 4.9 billion (case 4) over 10 years. Animal use ranges from about 900,000 (case 1) to about 6.4 million (case 4) over 10 years.

Note that the case with the lowest animal use (case 1) is not the case with the lowest cost (case 2). Case 1 uses about 2 million fewer animals than case 2, but costs about $ 1 billion more. This is because OECD TG 422, at a cost of about $ 145,000 per test, is less expensive than the subchronic toxicity test, at about $ 170,000 per test. However, OECD TG 422 uses about five times as many animals as the subchronic toxicity test: about 520 for OECD TG 422, and about 100 for the subchronic toxicity test. Testing guidance prefers the 90-day test to the 28-day test for determining an NOAEL (SCCS, 2010), so it is questionable whether the less expensive OECD TG 422 would be an acceptable substitute for the subchronic (90-day) toxicity test, especially given the animal welfare considerations.

Tab. 15: Summary of evaluation scenarios
Case 1: Subchronic Toxicity Study; Case 2: SIDS with OECD TG 422; Case 3: SIDS with Subchronic Toxicity Study and OECD TG 414; Case 4: SCCS with Tier 2 Tests

				Evaluations (X = endpoint is evaluated)			
Health Endpoint	**OECD Test Method**	**No. of Animals/Test**	**Test Cost in $ US**	**Case 1**	**Case 2**	**Case 3**	**Case 4**
Eye irritation/corrosion	437, 438	0	$ 3,500	X	X	X	X
Skin irritation/corrosion	430, 431, 439	0	$ 2,500	X	X	X	X
Skin sensitization	429	25	$ 6,500	X	X	X	X
Skin penetration	428	0	$ 25,000	X	X	X	X
Phototoxicity	432	0	$ 5,200	X	X	X	X
Mutagenicity/genotoxicity							
– Bacterial Reverse Mutation (Ames)	471	0	$ 4,600	X	X	X	X
– Mammalian Cell Gene Mutation	476	0	$ 22,000				X
– Mammalian Chromosomal Aberration Test	473	0	$ 20,000	X	X	X	
– Mammalian Cell Micronucleus Test	487	0	$ 20,000				X
Acute toxicity – oral	420, 425	8	$ 2,500	X	X	X	X
Acute toxicity – dermal	402	10	$ 2,700	X	X	X	X
Acute toxicity – inhalation	403	15	$ 15,000	X	X	X	X
Subchronic (90-day repeated dose) toxicity – oral	408	100	$ 170,000	X		X	X
Subchronic (90-day repeated dose) toxicity – dermal	411	80	$ 170,000	X		X	X
Subchronic (90-day repeated dose) toxicity – inhalation	413	80	$ 370,000	X		X	X
Chronic toxicity (>1 year)	452	160	$ 580,000				X
Carcinogenicity	451	416	$ 1,200,000				X
Reproductive/developmental toxicity screening with subacute toxicity	422	520	$ 145,000		X		
Prenatal Developmental Toxicity	414	768	$ 105,000			X	X
Two-generation Reproductive Toxicity	416	3,025	$ 430,000				X
Toxicokinetics	no standardized protocol		highly variable	not considered for any case			

Tab. 16: Summary of costs and animal use for evaluation scenarios over 10 years[1]

	Case 1: Subchronic Toxicity		Case 2: SIDS with OECD 422		Case 3: SIDS with Subchronic & OECD 414		Case 4: SCCS with Tier 2 Tests	
	No. of Animals	Cost (US $)	No. of Animals	Cost (US $)	No. of Animals	Cost (US $)	No. of Animals	Cost (US $)
MINIMUM: 1,200 Existing + 6,000 New								
Existing Ingredients:								
Laboratory test totals	77,495	$ 154,478,266	287,975	$ 94,163,266	586,679	$ 224,093,266	704,828	$ 307,728,166
Study management/ report preparation (add 50%)		$ 77,239,133		$ 47,081,633		$ 112,046,633		$ 153,864,083
Cost for RA/WOE[2] evaluations		$ 73,548,036		$ 25,120,536		$ 92,290,536		$ 139,363,086
Total for Existing Ingredients		$ 305,265,435		$ 166,365,435		$ 428,430,435		$ 600,955,335
New Ingredients:								
Laboratory test totals	866,108	$ 1,630,904,932	2,783,708	$ 975,554,932	4,782,908	$ 2,166,404,932	5,701,163	$ 2,797,274,932
Study management/ report preparation (add 50%)		$ 815,452,466		$ 487,777,466		$ 1,083,202,466		$ 1,398,637,466
Cost for RA/WOE evaluations		$ 52,213,378		$ 52,213,378		$ 52,213,378		$ 71,287,378
Total for New Ingredients		$ 2,498,570,776		$ 1,515,545,776		$ 3,301,820,776		$ 4,267,199,776
Total for Existing + New	**943,602**	**$ 2,803,836,211**	**3,071,682**	**$ 1,681,911,211**	**5,369,586**	**$ 3,730,251,211**	**6,405,990**	**$ 4,868,155,111**
MAXIMUM: 10,000 Existing + 6,000 New								
Existing Ingredients:								
Laboratory test totals	645,788	$ 1,287,318,886	2,399,788	$ 784,693,886	4,888,988	$ 1,867,443,886	5,873,564	$ 2,564,401,386
Study management/ report preparation (add 50%)		$ 643,659,443		$ 392,346,943		$ 933,721,943		$ 1,282,200,693
Cost for RA/WOE evaluations		$ 612,900,297		$ 209,337,797		$ 769,087,797		$ 1,161,359,047
Total for Existing Ingredients		$ 2,543,878,626		$ 1,386,378,626		$ 3,570,253,626		$ 5,007,961,126
New Ingredients:								
Laboratory test totals	866,108	$ 1,630,904,932	2,783,708	$ 975,554,932	4,782,908	$ 2,166,404,932	5,701,163	$ 2,797,274,932
Study management/ report preparation (add 50%)		$ 815,452,466		$ 487,777,466		$ 1,083,202,466		$ 1,398,637,466
Cost for RA/WOE evaluations		$ 52,213,378		$ 52,213,378		$ 52,213,378		$ 71,287,378
Total for New Ingredients		$2,498,570,776		$1,515,545,776		$3,301,820,776		$ 4,267,199,776
Total for Existing + New	**1,511,895**	**$5,042,449,402**	**5,183,495**	**$2,901,924,402**	**9,671,895**	**$6,872,074,402**	**11,574,726**	**$ 9,275,160,902**

[1] See Table 15 for the tests included in each scenario. [2] RA/WOE, read-across or weight of evidence.

Maximum
The maximum number of ingredients tested by year 10 is 16,000 ingredients, which is 10,000 existing ingredients plus 6,000 new ingredients. This is about 60% of the total number of ingredients by year 10. The cost ranges from about $ 2.9 billion (case 2) to about $ 9.3 billion (case 4). Animal use ranges from about 1.5 million (case 1) to about 11.5 million (case 4).

Considerations for new ingredients
For both the minimum and maximum analyses, the testing for new ingredients is a dominant factor. It accounts for about 90% of the cost and animal use for the minimum analysis and about 50% for the maximum analysis. The analyses assume manufacturers will test new ingredients according to the Act's standards as part of pre-market testing, even if the ingredient is not scheduled for evaluation by the FDA. Liability concerns, for example, could warrant this. From a financial perspective, however, a present value analysis could indicate that waiting makes more sense, at least for the most cost-intensive tests. If manufacturers choose not to conduct the safety tests pre-market, and instead wait until the FDA requests the data according to the timeline, then the animal and cost numbers shift down significantly. In that case, only 1,200-10,000 ingredients total are tested over 10 years, a maximum of 37% of total ingredients (about 50% of those in use).

Considerations for reproductive toxicity tests
The numbers are sensitive to the choice of reproductive toxicity tests. For example, the prenatal developmental toxicity test in cases 3 and 4 accounts for 8.1 million animals, or 85% of total animals in case 3 and 70% of total animals in case 4, and it accounts for 20% to 30% of the total testing costs.

This analysis assumes that the two-generation reproductive toxicity test (OECD TG 416) or extended one-generation test (OECD TG 443) would be used rarely, if at all. The cost in dollars and animal lives is so large for these tests that they would be difficult to justify for cosmetics ingredients. For example, consider this comparison if OECD 443 or 416 is substituted for the screening test OECD 422 in case 2:
- OECD TG 422: 3.1 million to 5.2 million animals; $ 1.7 billion to $ 2.9 billion.
- OECD TG 443: 6.3 million to 10.8 million animals; $ 6.0 billion to $ 10.8 billion.
- OECD TG 416: 17.2 million to 30 million animals; $ 4.1 billion to $ 7.4 billion.

OECD TG 443 is more costly than OECD TG 416 (about $ 650,000-$ 1,000,000/test for 443, depending on the test modules, and about $ 430,000/test for 416), but uses fewer animals. Both are far more costly in dollars and animals than the other options considered in this analysis.

3.9 Potential testing not considered: Finished cosmetics and ecotoxicity

The proposed Act focuses on ingredient testing, and that is the focus of this article as well. The Act also notes potential testing for finished cosmetic products and for ecotoxicity. This paper does not consider this potential additional testing, because the many unknowns do not allow meaningful analysis. Rather, the following summaries simply identify the general requirements.

Finished cosmetic products
The Act directs the Secretary to presume that a finished cosmetic product is safe if made solely of ingredients found to be safe or safe within limits. The Act does, however, include a provision for testing of a finished cosmetic product at the FDA's discretion: "*The Secretary may require that a brand owner demonstrate that a cosmetic meets the safety standard under section 614(a) (including by requiring that the brand owner conduct safety testing, or request such safety testing from relevant suppliers and manufacturers) of a cosmetic described under paragraph (1)) if the cosmetic (A) contains penetration enhancers, sensitizers, estrogenic chemicals, or other similar ingredients; (B) contains ingredients that react with each other or with other substances to form harmful by-products;…*" (section 617[b][2]).

Penetration enhancers are common ingredients in facial moisturizers and serums, to make the products more effective. Ingredients with low-level estrogenic activity are also fairly common, especially in plant extracts which are increasingly being incorporated into cosmetics as the natural products segment grows rapidly. How finished cosmetic products testing might be implemented is difficult to predict. With more than 500,000 cosmetic products on Amazon alone, the potential may be significant. Offsetting this, the threat of required testing could cause a business, especially a small business, to withdraw the product from the market rather than test, to avoid the cost of testing and possible bad publicity.

Some tests under the proposed Act would take longer than the typical life of a product, making test results moot. The cosmetics industry is extremely dynamic, characterized by rapid turnover of products. Hartung (2008) notes 25% turnover within 6 months. Products are typically reformulated at least every two years, and many are reformulated every year, some even twice a year. This rapid turnover is across brands and cosmetic types. By the time test results were available for a product, the product might not exist anymore.

Ecotoxicity
The proposed Act requires brand owners to submit "*(C) Exposure and fate information*" for each ingredient in its cosmetics and for each cosmetic owned by the brand (section 615[a][2][C]). This simple sentence can mean significant additional toxicity studies, including animal testing.

4 Conclusions

Currently, cosmetic ingredients typically undergo safety testing for eye and skin irritation, skin sensitization, mutagenicity, and acute toxicity; and for UV filters (about 0.1% of ingredients), testing for phototoxicity, too. About 2% of ingredients – mainly hair dyes, colorants, preservatives, and

UV filters – are intended to be reactive or biologically active and may undergo additional, repeated dose studies to assess systemic and reproductive toxicity.

The annual cost and animal use for ingredient safety evaluations in the US are estimated here at $ 5.4 million and 2,700 animals. The proposed US Safe Cosmetics Act would affect testing worldwide, however, so the worldwide baseline number is also important. The current totals worldwide for cosmetic ingredients are estimated to be $ 54 million and 27,000 animals.

Under the proposed US Safe Cosmetics Act, every ingredient would be evaluated to develop its NOAEL from a repeated dose study. Until alternative methods are available for this (about 10 years), the most likely study, based on current guidance documents, would be the subchronic 90-day repeated dose toxicity study (SCCS, 2010; SCCP, 2012). We evaluated four potential testing scenarios to estimate cost and animal use.

In the first scenario, ingredients would undergo basic tier 1 evaluations plus the subchronic study. Over 10 years, this would cost about $ 2.8-$ 5.0 billion and use about 900,000 to 1.5 million animals. Annually, this would be about $ 280-$ 500 million and 90,000 to 150,000 animals.

Optionally, a reproductive toxicity screening study (OECD TG 422) could replace the subchronic study, with the 28-day repeated dose portion of the screening study used to derive the NOAEL. We consider this scenario less likely, because the subchronic study is preferred for obtaining the NOAEL. In this scenario, the cost over 10 years is $ 1.7-$ 2.9 billion and 3.1 million to 5.2 million animals. Annually, this is $ 170-$ 290 million and 310,000 to 520,000 animals.

A third scenario is the same as the first, but adds the prenatal developmental toxicity study (OECD TG 414). The prenatal study would be in addition to the subchronic study. The resulting total cost is $ 3.7-$ 6.9 billion and 5.4 million to 9.7 million animals. Annually, the estimates are $ 370-$ 690 million and 540,000 to 970,000 animals.

Finally, we considered a scenario in which guidelines currently recommended for reactive and biologically active ingredients (about 2% of ingredients) would be applied to all ingredients. This could involve a cost- and animal-intensive multi-generation test, OECD TG 416 or 443, but we assume that test would be requested rarely, for only 3% of ingredients. For this case, the cost over 10 years is $ 4.9- $ 9.3 billion and 6.4 million to 11.5 million animals. The corresponding annual numbers are $ 490-$ 930 million and 640,000 to 1.2 million animals.

An analysis such as this relies on many assumptions, some with a high degree of uncertainty. Within the reasonable bounding range for each assumption, however, it is clear that, no matter which assumption we use within that range, the cost of the proposed evaluations is much higher compared with current practice, both in terms of dollars and animal use.

A striking mismatch exists between the multi-year evaluations of ingredients under the Act and the rapid product development cycles in the cosmetics industry. These cycles are similar to those in many technology industries, with constant pressure for new products and reformulations of old products, both to take advantage of new science and to meet changing consumer preferences. Each year, the cosmetics industry worldwide generates hundreds of new ingredients and thousands of new products and reformulations. Under the proposed Act, the FDA would begin evaluations for 300 ingredients in the first year of evaluations, and then begin 100 more evaluations annually each year thereafter. Given there are over 20,000 potential ingredients currently and 500-700 new ingredients added worldwide each year that are either developed in the US or may be imported to the US, this is a timeline that can never catch up with the industry.

If the goal is to evaluate all cosmetic ingredients, the only viable approach would seem to be through new methods currently under development. These techniques use high-throughput cellular- and molecular-based screening tests combined with computational biology. Traditional animal-based evaluation methods are time-consuming and would always lag this industry. High-throughput methods are expected to be available in about 10 years. Even with this 10-year development time, employing new methods is the fastest way to evaluate cosmetics ingredients and, as noted, is the only way to keep up with the rapid product cycles in this industry. If the Act's goal is to be achieved, rapid completion of the development of high-throughput methods will be essential.

Ironically, implementation of the Act could postpone the development of the new testing that would facilitate reaching its goals. The infrastructure for toxicology testing is finite, used both for conducting testing and for developing alternative methods. If more resources are used for the proposed testing, fewer will be available for developing alternative methods, postponing their development.

The rationale for the Act is not stated, but it is likely the precautionary principle. The EU Cosmetics Regulation also invokes the precautionary principle; however, it prohibits animal testing for cosmetics. The proposed US Act, in asking for expanded animal testing until alternative methods are available, conflicts with the EU Cosmetics Regulation. If the proposed US Act passes, it is not clear whether cosmetics companies will need two different types of products: one with animal testing for the US market and one without animal testing for the EU market.

In this article, we have focused on numbers, but it is important to ask whether the expanded testing in the Act, and the consequent impacts on the cosmetics industry, the toxicology field, and animal welfare, are justified for cosmetics. We hope others will take up this discussion.

References

Adler, S., Basketter, D., Creton, S., et al. (2011). Alternative (non-animal) methods for cosmetics testing: current and future prospects – 2010. *Arch Toxicol 85*, 367-485. http://dx.doi.org/10.1007/s00204-011-0693-2

Basketter, D., Casati, S., Gerberick, G. F., et al. (2005). 4. Skin sensitisation. *Altern Lab Anim 33, Suppl 1*, 83-103. http://www.frame.org.uk/atla_article.php?art_id=134&pdf=true

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*, 3-91. http://www.altex.ch/All-issues/Issue.50.html?iid=129&aid=1

Bishop, P. L, Manuppello, J. R., Willett, C. E., and Sandler, J. T. (2012). Animal use and lessons learned in the U.S. High Production Volume Chemicals Challenge Program. *Environ Health Perspect 120*, 1631-1639.

Bottini, A. and Hartung, T. (2009). Food for thought … on the economics of animal testing. *ALTEX 26*, 3-16. http://www.altex.ch/All-issues/Issue.50.html?iid=104&aid=1

Cooper, R. L., Lamb, J., Barlow, S., et al. (2006). A tiered approach to life stages testing for agricultural chemical safety assessment. *Crit Rev Toxicol 36*, 69-98. http://dx.doi.org/10.1080/10408440500541367

Corbett, J. F., Sharma, R. K., and Dressler, W. E. (1999). Cosmetic toxicology. In H. Marquardt, S. G. Schäfer, R. O. McClellan, and F. Welsch (ed.). *Toxicology*. Academic Press. 899-918.

EURL ECVAM (2013). EURL ECVAM recommendation on the 3T3 neutral red uptake cytotoxicity assay for acute oral toxicity testing. *JRC79556 EUR 25946 EN*, 1-46.

EC – European Commission (2006). Regulation (EC) of No 1907/2006 of the European parliament and of the council 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC.

EC (2007). Commission Staff Working Document – Annex to the report from the Commission to the Council and the European Parliament. Fifth report on the statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union. SEC. 1455, 1-277.

EC (2009a). Regulation (EC) of No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products.

EC (2009b). Commission Regulation (EC) No 134/2009 of 16 February 2009 amending Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH).

EC (2010). Commission Staff Working Document – Accompanying document to the Report from the Commission to the Council and the European Parliament. Sixth report on the statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union. SEC. 1107, 1-311.

ECHA (2011). The use of alternatives to testing on animals for the REACH regulation. ECHA-11-R-004.2-EN.

ECHA (2012a). Guidance on information requirements and chemical safety assessment, Chapter R.7a, Endpoint specific guidance. ECHA-12-G-11-EN.

ECHA (2012b): Report on Survey of Worldwide CROs: Costs and Practicalities of Two New OECD Guidelines for Testing Chemical Substances: OECD TG 442, Extended One-Generation Reproductive Toxicity Study, and OECD TG 488, Transgenic Rodent Somatic and Germ Cell Mutation Assay. Prepared by CEHTRA (Consultancy for Environment and Human Toxicology and Risk Assessment), Nottingham UK, June 2012.

FDA (2012). Draft guidance for industry: Safety of nanomaterials in cosmetic products. http://www.fda.gov/Cosmetics/GuidanceComplianceRegulatoryInformation/GuidanceDocuments/ucm300886.htm#toxicology

Fleischer, M. (2007). Testing costs and testing capacity according to the REACH requirements – results of a survey of independent and corporate GLP laboratories in the EU and Switzerland. *Journal of Business Chemistry 3*, 96-114.

Global Insight (2007). A study of the European cosmetics industry. Prepared for the European Commission, Directorate General for Enterprise and Industry. http://bit.ly/1jm1k9c

Hartung, T. (2008). Food for thought … on alternative methods for cosmetics safety testing. *ALTEX 25*, 147-162. http://www.altex.ch/All-issues/Issue.50.html?iid=101&aid=1

Hofer, T., Gerner, I., Gundert-Remy, U., et al. (2004). Animal testing and alternative approaches for the human health risk assessment under the proposed new European chemicals regulation. *Arch Toxicol 78*, 549-564. http://dx.doi.org/10.1007/s00204-004-0577-9

H.R.1385 (2013). A bill to amend title VI of the Federal Food, Drug, and Cosmetic Act to ensure the safe use of cosmetics, and for other purposes. 1-59.

IGHRC – Interdepartmental Group on Health Risks from Chemicals (2006). Guidelines on route-to-route extrapolation of toxicity data when assessing health risks of chemicals. Institute of Environment and Health, Cranfield University, Silsoe, Bedfordshire, UK. http://www.silsoe.cranfield.ac.uk/ieh/ighrc/ighrc.html

Maurici, D., Aardema, M., Corvi, R., et al. (2005). Genotoxicity and mutagenicity. *Altern Lab Anim 33*, 117-130.

Nohynek, G. J., Antignac, E., Re, T., and Toutain, H. (2010). Safety assessment of personal care products/cosmetics and their ingredients. *Toxicol Appl Pharmacol 243*, 239-259. http://dx.doi.org/10.1016/j.taap.2009.12.001

NRC – National Research Council (2007). *Toxicity testing in the twenty-first century: A vision and a strategy*. Washington DC, USA: National Academies Press. http://www.nap.edu/catalog.php?record_id=11970

OECD (2004a). Chapter 2. Data gathering. In OECD (ed.). Manual for the Assessment of Chemicals. 1-33.

OECD (2004b). OECD guidance document for neurotoxicity testing. 67 pp.

OECD (2008). Guidance Document for Mammalian Reproductive Toxicity Testing and Assessment. 88 pp.

OECD (2012). Guidance Document on Standardised Test Guidelines for Evaluating Chemicals for Endocrine Disruption. Series on Testing and Assessment. No. 150.

Pedersen, F., de Bruijn, J., Munn, S., and van Leeuwen, K. (2003). Assessment of additional testing needs under REACH – effects of (Q)SARs, risk based testing and voluntary industry initiatives. JRC Report EUR 20863.

Personal Care Products Council (2012). Email communication, including estimate of new ingredients per year, from Glenda Williams, Director of International Regulatory Affairs, Personal Care Products Council. August 2, 2012.

Rovida, C. (2011). Local lymph node assay: How testing laboratories apply OECD TG 429 for REACH purposes. *ALTEX 28*, 117-129. http://dx.doi.org/10.14573/altex.2011.2.117

Rovida, C. and Hartung, T. (2009). Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals – a report by the Transatlantic Think Tank for Toxicology. *ALTEX 26*, 187-208. http://www.altex.ch/All-issues/Issue.50.html?iid=107&aid=4

Rovida, C., Longo, F., and Rabbit, R. R. (2011). How are reproductive toxicity and developmental toxicity addressed in REACH dossiers? *ALTEX 28*, 111-132. http://dx.doi.org/10.14573/altex.2011.4.273

RPA & Statistics (2002). RPA and Statistics Sweden (2002). Assessment of the impact of new regulations in the chemicals sector. Final Report. Prepared for European Commission Enterprise Directorate General. http://ec.europa.eu/enterprise/reach/docs/whitepaper/bia_report-2002_06_en.pdf

SCCS (2010). The SCCS's notes of guidance for the testing of cosmetic ingredients and their safety evaluation.7th Revision. European Commission, Directorate-General Health and Consumer Protection, December 14, 2010.

SCCP (2012). Draft decision on guidelines annex I: Commission implementing decision on guidelines on annex I to Regulation (EC) No. 1223/2009 on cosmetic products.12/SANCO/COS/24. November 5, 2012.

US EPA – Environmental Protection Agency (2012). Excerpt from the US EPA's web site. http://epa.gov/pesticides/regulating/laws/fqpa/fqpa_accomplishments.htm (accessed Oct 2012).

van der Jagt, K., Munn, S., Torslov, J., and de Bruijn, J. (2004). Alternative approaches can reduce the use of test animals under REACH. Addendum to the report: Assessment of additional testing needs under REACH. Effects of (Q)SARS, risk-based testing and voluntary industry initiatives. JRC REPORT EUR 21405. http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/8790/1/EUR%2021405%20EN.pdf (accessed Oct 2012).

Vinken, M., Pauwels, M., Ates, G., et al. (2011). Screening of repeated dose toxicity data present in SCC(NF)P/SCCS safety evaluations of cosmetic ingredients. *Arch Toxicol 86*, 405-412. http://dx.doi.org/10.1007/s00204-011-0769-z

Acknowledgements

The authors gratefully acknowledge Sue Leary (Coalition for Consumer Information on Cosmetics), Marty Stephens (CAAT-Johns Hopkins), Brent Gilbert (CEETOX), John Majnarich (Bio Research Laboratories), and Marie Butler-Knight (consulting editor) for their encouragement and early reviews of the manuscript and to the anonymous peer reviewers for their important insights on this paper. We also thank Thomas Hartung (CAAT-Johns Hopkins) for introducing us, making this collaboration possible.

Neither author has received compensation for writing this paper. Author Knight owns a cosmetics store for brands that are not tested on animals, but outcomes related to the Act would have no material effect on the store.

Correspondence to

Jean Knight
643 Main Street, Ste 6
Half Moon Bay, CA 94019
USA
Phone: +1 650 339 0479
e-mail: jeanknight@alum.mit.edu