Food and Chemical Toxicology 109 (2017) 170–193

Contents lists available at ScienceDirect

Food and Chemical Toxicology

journal homepage: www.elsevier.com/locate/foodchemtox

Thresholds of Toxicological Concern for cosmetics-related substances: New database, thresholds, and enrichment of chemical space

Chihae Yang [a, b], Susan M. Barlow [c], Kristi L. Muldoon Jacobs [d, 1], Vessela Vitcheva [a, b, e], Alan R. Boobis [f], Susan P. Felter [g], Kirk B. Arvidson [d], Detlef Keller [h], Mark T.D. Cronin [i], Steven Enoch [i], Andrew Worth [j], Heli M. Hollnagel [k, *]

[a] Altamira LLC, Columbus, OH 43235, USA
[b] Molecular Networks GmbH, Nürnberg, Germany
[c] Harrington House, Brighton BN1 6RE, United Kingdom
[d] US Food and Drug Administration, Center for Food Safety and Applied Nutrition, Office of Food Additive Safety, College Park, MD 20740, USA
[e] Faculty of Pharmacy, Medical University of Sofia, 2 "Dunav" Str., Sofia 1000, Bulgaria
[f] Imperial College London, Department of Medicine, London W12 0NN, United Kingdom
[g] The Procter & Gamble Company, Cincinnati, OH, USA
[h] Henkel AG & CO KGaA, Henkelstr. 67, 40191 Düsseldorf, Germany
[i] School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool L3 3AF, United Kingdom
[j] European Commission, Joint Research Centre, Ispra, Italy
[k] Dow Europe GmbH, 8810 Horgen, Switzerland

ARTICLE INFO

Article history:
Received 15 May 2017
Received in revised form
23 August 2017
Accepted 28 August 2017
Available online 1 September 2017

Keywords:
Threshold of Toxicological Concern
TTC
Cosmetics
Cheminformatics
Cramer classification

ABSTRACT

A new dataset of cosmetics-related chemicals for the Threshold of Toxicological Concern (TTC) approach has been compiled, comprising 552 chemicals with 219, 40, and 293 chemicals in Cramer Classes I, II, and III, respectively. Data were integrated and curated to create a database of No-/Lowest-Observed-Adverse-Effect Level (NOAEL/LOAEL) values, from which the final COSMOS TTC dataset was developed. Criteria for study inclusion and NOAEL decisions were defined, and rigorous quality control was performed for study details and assignment of Cramer classes. From the final COSMOS TTC dataset, human exposure thresholds of 42 and 7.9 μg/kg-bw/day were derived for Cramer Classes I and III, respectively. The size of Cramer Class II was insufficient for derivation of a TTC value. The COSMOS TTC dataset was then federated with the dataset of Munro and colleagues, previously published in 1996, after updating the latter using the quality control processes for this project. This federated dataset expands the chemical space and provides more robust thresholds. The 966 substances in the federated database comprise 245, 49 and 672 chemicals in Cramer Classes I, II and III, respectively. The corresponding TTC values of 46, 6.2 and 2.3 μg/kg-bw/day are broadly similar to those of the original Munro dataset.

1. Introduction

The Threshold of Toxicological Concern (TTC) is a risk assessment approach that can be used to screen substances with few or no toxicological data for which human exposures are likely to be low. The TTC approach utilizes generic human exposure threshold values (TTC values) that have been derived from oral experimental data on cancer and non-cancer toxicity endpoints. If human exposure to a substance is below the relevant TTC value, it can be judged "with reasonable confidence, to present a low probability of a risk" (Munro et al., 1996). The work presented here was undertaken in order to underpin and facilitate the use of the TTC approach for substances found in cosmetics.

The TTC approach was inspired by, and can be considered an extension of, the Threshold Of Regulation (TOR) that was adopted by the US Food and Drug Administration (FDA) for substances used in food-contact articles (US FDA, 1993; 1995). The original TOR concept used a single threshold for all chemicals, based on the

* Corresponding author. ILSI Europe a.i.s.b.l., Avenue E. Mounier 83, Box 6, BE-1200 Brussels, Belgium.
E-mail address: publications@ilsieurope.be (H.M. Hollnagel).
[1] Present address: US Pharmacopeial Convention (USP), 12601 Twinbrook Parkway, Rockville, MD 20852, USA.

http://dx.doi.org/10.1016/j.fct.2017.08.043

Abbreviations

ADI	Acceptable Daily Intake
BMD	Benchmark Dose
BMDL	Benchmark Dose Lower 95% confidence limit
bw	body weight
CAS RN	Chemical Abstract Services Registry Number
CFSAN	Center for Food Safety and Nutrition
DB	Database
DFG	Deutsche Forschungsgemeinschaft
ECHA	European Chemicals Agency
EDTA	Ethylenediaminetetraacetate
EFSA	European Food Safety Authority
EINECS	European INventory of Existing Commercial Substances
EMA	European Medicines Agency
EMEA	European Agency for the Evaluation of Medicinal Products
EPA	Environmental Protection Agency
EU	European Union
FCN	Food Contact Notification
FDA	Food and Drug Administration
HNEL	Highest No-Effect Levels
INCI	International Nomenclature for Cosmetics Ingredients
IRIS	Integrated Risk Information System
JECFA	Joint FAO/WHO Expert Committee on Food Additives
LEL	Lowest Effect Level
LOAEL	Lowest-Observed-Adverse-Effect Level
LOEL	Lowest-Observed-Effect Level
MINIS	MINImum Study
MoS	Margin of Safety
NEL	No-Effect Levels
NOAEL	No-Observed-Adverse-Effect Level
NOEL	No-Observed-Effect Level
NTP	National Toxicology Program
oRepeaTox DB	Oral repeated-dose toxicity database
PAFA	Priority-based Assessment of Food Additives
POD	Point of departure
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
RfD	Reference Dose
SCCS	Scientific Committee on Consumer Safety
SCENIHR	Scientific Committee on Emerging and Newly Identified Health Risks
SCHER	Scientific Committee on Health and Environmental Risks
SEURAT	Safety Evaluation Ultimately Replacing Animal Testing
TOR	Threshold Of Regulation
TTC	Threshold of Toxicological Concern
USA	United States of America
UVCB	substance of Unknown or Variable composition, Complex reaction products or Biological materials
WHO	World Health Organisation

conservative assumption that an untested chemical could pose a cancer risk, even though it was not intended to be used for chemicals with structural alerts or other reason for concern for genotoxicity. Tetra sodium ethylenediaminetetraacetate (EDTA) (Chemical Abstract Services Registry Number [CAS RN]: 64-02-8) was the first chemical to which TOR was applied in 1996 at US FDA Center for Food Safety and Nutrition (CFSAN).[2] It was subsequently expanded into the TTC concept to include non-cancer endpoints by Munro et al. (1996) and further elaborated by Kroes et al. (2004), who proposed the addition of another tier intended to be protective for DNA-reactive carcinogens.

The TTC approach was originally developed for substances present at low levels in the diet and consumed orally (Barlow, 2005) and was used by JECFA for evaluating flavouring substances. It was subsequently evaluated in detail for use in food safety by the European Food Safety Authority (EFSA) (EFSA, 2012). Improvement and expansion of the TTC approach were also discussed in an Expert Workshop convened by EFSA and the World Health Organisation (WHO) in 2014 (EFSA/WHO, 2016). Application of the TTC approach has also been proposed for, or extended to, the risk assessment of other types of substances. These include substances present in consumer products (Antignac et al., 2011; Blackburn et al., 2005; SCCS, SCHER and SCENIHR, 2012; SCCS NfG, 2016): micropollutants, drug residues, pesticide metabolites and other impurities in drinking water (Brüschweiler, 2010; EFSA, 2016; Houeto et al., 2012; Laabs et al., 2015; Melching-Kollmuß et al., 2010; Mons et al., 2013); genotoxic impurities in human pharmaceuticals (EMEA, 2006); herbal preparations (EMEA, 2008); homeopathic medicines (Buchholzer et al., 2014); and human pharmaceutical substances carried over in multiproduct manufacturing facilities (Bercu and Dolan, 2013; Stanard et al., 2015). It has also been used as a first-level screening tool to prioritize for review a large number of substances identified as needing an assessment under the Canadian Environmental Protection Act (Health Canada, 2016). Consideration has also been given to whether the TTC approach could be applied to human biomonitoring data (Becker et al., 2012) and to human exposures by non-oral routes (Carthew et al., 2009; Escher et al., 2010; Hennes, 2012; Kroes et al., 2007; Partosch et al., 2015).

The original reference dataset (Munro et al., 1996) consisted of 613 organic substances representing a “range of industrial chemicals, pharmaceuticals, food substances and environmental, agricultural and consumer chemicals likely to be encountered in commerce”. Although the intent was to cover a broad chemical domain, the dataset is now over 20 years old, and questions have been raised as to whether it is adequately representative of chemicals and structures used in contexts other than its original application in food (Dewhurst and Renwick, 2013). This issue was first raised in relation to cosmetics by Blackburn et al. (2005) and was an important consideration for the use of TTC for chemicals in cosmetics and consumer products in the opinion of the European Commission's non-food Scientific Committees (SCCS, SCHER and SCENIHR, 2012). The Scientific Committees stated that the TTC approach is scientifically acceptable, whilst noting some concerns, including that all risk assessment approaches have some degree of uncertainty, that many complex chemical structures are not adequately represented in currently available databases, and that there is limited knowledge of effects due to dermal and inhalational exposure routes that are more common for consumer products (SCCS, SCHER and SCENIHR, 2012).

Better understanding of the applicability of the TTC concept to substances present in cosmetic products would be particularly

[2] Information provided by Kirk Arvidson at the Office of Food Additive Safety of US FDA CFSAN https://www.accessdata.fda.gov/scripts/fdcc/index.cfm?set=TOR&id=1996-001.

valuable because of the impact of the European Union (EU) Regulation that prohibits the marketing of cosmetic products and ingredients that have been tested on animals after 2009 or 2013 (European Commission, 2009). To this end, the COSMOS project developed a Cosmetics Inventory for substances used or potentially found (e.g., as a contaminant or packaging migrant) in cosmetics as a reference look-up table. A search was then conducted across publicly available databases for toxicity data on all the substances in the Inventory. Only about 10% of the substances in the Inventory had toxicity data with NO(A)EL/LO(A)EL (No/Lowest-Observed-Adverse-Effect Level) values from regulatory submissions and the scientific literature. These substances forming the intersection of the toxicity database and the Cosmetics Inventory were then identified as initial candidates (Fig. 1) and further developed into a COSMOS TTC dataset in this project. The oral TTC values relevant to cosmetics have been derived and can be compared with the previously established generic human exposure thresholds (Kroes et al., 2004; Munro et al., 1996).

The COSMOS project was part of the European research initiative with the long-term goal of achieving "Safety Evaluation Ultimately Replacing Animal Testing" (SEURAT-1), co-funded by the European Commission and Cosmetics Europe. The overall aim of the COSMOS project was to develop computational methods that can serve as viable alternatives to toxicity testing in animals for cosmetic ingredients. Derivation of TTC values from a cosmetics toxicity dataset would provide higher confidence in the use of the TTC approach in that context.

COSMOS project set up two collaborative working groups coordinated by ILSI Europe. One group addressed dermal-to-oral extrapolation, using a flux decision-tree approach, to derive dermal systemic exposures for comparison with oral TTC values (Williams et al., 2016). The other group addressed whether the chemical space of cosmetics ingredients was adequately reflected by the chemicals in the current TTC database. First, it was necessary to define the chemical space of cosmetics-related chemicals; hence the COSMOS Cosmetics Inventory was developed by the COSMOS project and is described in detail elsewhere (Cronin et al., 2012). In this publication, we focus on the use of the curated COSMOS TTC dataset of non-cancer endpoints for derivation of TTC values for cosmetics-related chemicals.

2. Materials and methods

The workflow for this project was complex and is summarized here to orient the reader. First, it was necessary to define what chemicals can be considered as "cosmetics-related" by establishing a look-up inventory. At the start of the project (2011), the EU (CosIng) database was still being developed and did not provide a public resource for a complete inventory. The COSMOS Cosmetics Inventory, containing 20,974 substances, was therefore developed. Extensive searches were then conducted across publicly available sources for toxicity data on all the substances in the Inventory. To be usable, the toxicity data was constrained to that which had numeric endpoints, i.e. NO(A)ELs/LO(A)ELs. From these searches on the 20,974 substances in the Inventory, just over 2000 substances with NO(A)EL/L(O)AEL values were identified (see Fig. 1). From this initial data compilation, it was evident that certain chemical classes needed to be enriched. Therefore, a new database was built, called "oRepeaTOX", which added 228 cosmetics-related chemicals, including ones from new chemical classes such as hair dyes, preservatives, UV filters, and relevant impurities. The oRepeaTOX database was merged with the initial toxicity data compilation into the COSMOS database, containing over 2300 substances. From this COSMOS database, a new NO(A)EL/L(O)AEL database, containing 1059 chemicals was built by applying study selection criteria. A further set of rules was then applied to determine a point of departure for each chemical, to be used for calculation of TTC values. This resulted in a final COSMOS TTC dataset of 552 chemicals. The workflow is outlined in Fig. 2 and each part of the process is described in more detail below.

2.1. COSMOS Cosmetics Inventory

The development of the new TTC database for cosmetics-related chemicals begins with the ability to identify the substances as such, as depicted in Fig. 1. Due to the complexity of use and product categories and differing regulatory or reporting systems for cosmetics ingredients in the EU and the United States of America (USA), a centralized inventory was needed as a reference library to define the "cosmetics-related" chemical space. The COSMOS Cosmetics Inventory is a listing of cosmetics ingredients (although they are not all intentionally used in cosmetics) and other substances that have been reported to be present in cosmetics products in the EU and the USA. The Inventory was prepared by merging the substance lists from the European Union CosIng (Cosmetic Ingredients) (European Commission, 2012) and the US Personal Care Products Council (Bailey, 2011) Databases. The Inventory includes the International Nomenclature for Cosmetics Ingredients (INCI) name, the CAS RN, the European INventory of Existing Commercial Substances (EINECS) number, function (according to EU CosIng), and product category (USA). The COSMOS Cosmetics Inventory contains 9876 unique CAS RN and 19,473 unique INCI names. Approximately 50% of the inventory comprises botanicals, animal fats, polymers, resins and UVCBs (substances of Unknown or Variable composition, Complex reaction products or Biological materials), which are not amenable for TTC or computational approaches due to their poorly defined chemical structure. Based on this substance inventory, a set of 5270 test substances (4740 unique chemical structures) were

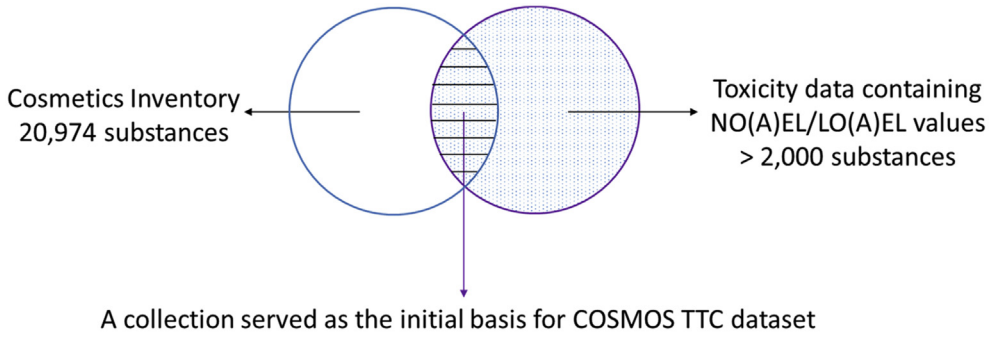


Fig. 1. Identification of candidate collection to be developed for COSMOS TTC dataset.

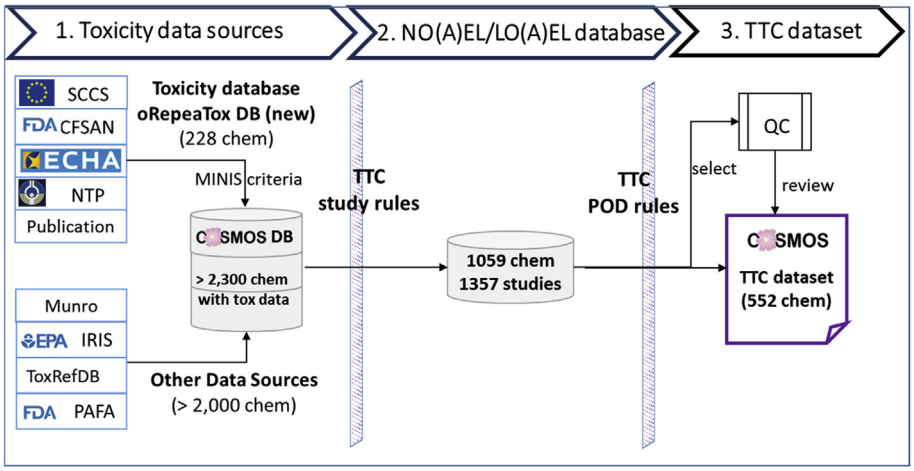


Fig. 2. The curation process for the COSMOS TTC dataset.

identified and were used to define the chemical domain for TTC analysis. Further information on the compilation of the inventory can be found elsewhere (Cronin et al., 2012). The Inventory is freely available within the COSMOS Database (COSMOS DB) v2.0 (Molecular Networks, 2017).

2.2. Development of databases

The curation strategy for obtaining the final COSMOS TTC dataset required rigorous database constructions across three phases, summarized in Fig. 2. The first phase (1) was the construction of a new oral toxicity database, the oRepeaTOX DB, to enrich the COSMOS database with detailed study result information from 228 cosmetics-related chemicals (i.e. cosmetics ingredients and unintentionally added chemicals found in cosmetics product formulations, such as packaging migrants). This new oRepeaTox DB was then added to the existing collection of toxicity data from sources providing rather NO(A)EL/LO(A)EL values than detailed study result information, described in Fig. 1 as the initial candidates and labelled in Fig. 2 as "Other Data Sources (>2000 chem)". Together these two sources provided toxicity data for more than 2300 chemicals. The second phase (2) was to filter studies appropriate for TTC to compile a database with the NO(A)EL and LO(A)EL values. There were 1059 chemicals covering 1357 studies in this NO(A)EL/LO(A)EL database. The third phase (3) was then to further refine the compilation to establish a new COSMOS TTC dataset of 552 chemicals and point of departure (POD) values.

2.2.1. Oral repeated-dose toxicity database (oRepeaTox DB)

The oRepeaTox DB was developed to address the issue of the relative lack of readily available data on toxicity of chemicals related to cosmetics. It contains oral, repeated-dose, non-cancer toxicity data for cosmetics-related chemicals from subchronic, chronic, carcinogenicity (non-neoplastic findings only), reproductive, and developmental studies. Other study types such as local irritation studies or sensitisation studies did not meet the study inclusion criteria and were not considered. The data were compiled from the following publicly available sources: opinions of the European Commission's Scientific Committee on Consumer Safety (EC SCCS) (European Commission, 2017), opinions of the European Food Safety Authority (EFSA, 2017), Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) Registered Substance Database of the European Chemicals Agency (ECHA) (ECHA, 2017), documents from the US FDA CFSAN[3] and the US FDA Priority-based Assessment of Food Additives (PAFA) (Benz and Irausquin, 1991), documents from the US Environmental Protection Agency (US EPA, 2016), the National Toxicology Program (NTP) database (US Department of Health and Human Services, 2017), and open literature publications. Data were compiled from these sources by manual harvesting. The new oRepeaTox DB provided 341 studies for 228 cosmetics-related chemicals (Gocht and Schwarz, 2014). The study counts from the above sources are: EU SCCS/SCCP/SCCNFP (118), US FDA CFSAN (107), REACH ECHA (42), open literature publications (39), US NTP (21), US FDA PAFA (9), and US EPA (5).

Data were compiled into a data entry tool prepared by the COSMOS DB team. The COSMOS consortium provided three groups for this activity: 1) the content group manually curated the data from the above sources; 2) the software group implemented the database technology; 3) a data expert/master curator profiled the data map, enforced the standards/criteria, and reviewed the quality of the data content. To create this new toxicity database, data from several existing databases were consolidated following minimum study inclusion criteria, termed COSMOS MINIS (*MINI*mum Study) criteria. The COSMOS MINIS criteria for toxicity studies are described in Appendix 1. For many cases where regulatory data sources only pointed to literature publications or study reports, original papers or documents were obtained when possible to capture the detailed dose-level data. For example, some of the studies from ECHA, US EPA, US FDA, or US FDA PAFA were re-harvested from the full toxicity data available in FDA internal documents, open literature publications, or NTP technical reports. All of the test substances from these sources were confirmed as cosmetics-related chemicals by using the COSMOS Cosmetics Inventory as a reference list. The only exception to this rule was the inclusion of some Food Contact Substances and impurities from the US FDA CFSAN's Food Contact Notification (FCN) program in an attempt to include potential impurities from packaging materials; 34 out of the 85 such chemicals from US FDA CFSAN's FCN program were found in the Cosmetics Inventory.

It should be noted that, in this paper, a test substance is distinguished from a chemical structure. A test substance is a particular form of a chemical that has been used in testing (e.g., in vivo or in vitro assays) and that can be further differentiated by

[3] Data from the US FDA CFSAN internal documents were made available by the Office of Food Additive Safety. The QC work was conducted at FDA.

attributes such as synthetic routes or manufacturing processes (reagent vs. technical grade), which can result in different impurity profiles. For example, trichloroethylenes with and without a trace of epichlorohydrin from different manufacturing processes are considered as two different test substances although represented as the same chemical structure. In addition, a chemical can also be differentiated by either well- or ill-defined compositions. Sodium dodecyl (lauryl) sulfate has a well-defined composition and is distinguished from the sodium coco-sulfate, which is ill-defined due to the range of coco chain length (C8-C18 centered around C12). However, both substances can still be represented by the same structure of dodecyl chain. The COSMOS TTC dataset is therefore test substance-centric.

The toxicity information in the oRepeaTox DB is structured such that a particular effect for a site at a given dose level is represented for each study for each chemical as accurately as possible. Study designs are described in detail for species, sex, route and duration of exposure, dose group (levels and number of animals), control information, and references. The effects are described by a set of controlled vocabulary and qualified by time of findings, severity, statistical significance, and treatment-relatedness. The target sites are further differentiated for organ/system, tissue/segment, and cells/organelles. The oRepeaTox DB is available from COSMOS DB v2.0 (Molecular Networks, 2017).

There has also been discussion of whether it would be preferable, from a scientific perspective, to use molar quantities of chemical entities and convert NOAELs from mg/kg-bw into mmol/kg-bw (Escher et al., 2010; Dewhurst and Renwick, 2013). This was not pursued in the present work for the COSMOS TTC dataset since the scientific community utilising the TTC approach is mostly working on a mg/kg-bw basis, and using that basis allows easier comparisons with other published TTC values.

2.2.2. Munro TTC dataset

The current TTC approach for non-cancer endpoints is based on the dataset published by Munro et al. (1996). The Munro dataset contains 613 diverse substances from 609 unique chemicals. The difference is because the Munro dataset sometimes listed the same chemical under different substance names or as duplicate records but with different study types and No-Observed-Effect Level (NOEL) values; these include 5,5-diphenylhydantoin, ascorbic acid, and azorubine (Carmoisine, C.I. ACID RED 14). The Munro et al. (1996) database includes aggregated data of study design parameters (study type, species, route and duration of exposure, doses), NOEL/LOEL values, critical effects, and references. The dataset cited 200 chronic, 233 subchronic, 89 reproductive, and 91 teratogenicity studies.

Munro calculated points of departure (PODs) based on “NOEL” values. A factor of three was used to adjust the NOEL from studies of shorter than chronic duration and designated with an asterisk in the original publication (note that whilst the adjusted values were used in the derivation of the TTC values, they were not explicitly cited in the published tables but indicated with asterisks). The dataset of 613 substances was also divided into the three structural classes defined by Cramer et al. (1978); 137 substances to Class I, 28 to Class II, and 448 to Class III. The three Cramer Classes became the basis of grouping chemicals in the current TTC paradigm.

For this study, the Munro dataset was first downloaded from the EFSA website (Bassan et al., 2011), then the records were corrected back to reflect exactly the same as the original Munro et al. (1996) publication. This version is referred to as “Munro-1996” in this present publication and was used verbatim for analyses where historical comparisons were important. The content was further corrected by COSMOS TTC quality control (QC) as well as additional database QC before importing to the COSMOS DB, where the “the updated “Munro-1996”, is now downloadable (Molecular Networks, 2017). The 190 substances from the Munro dataset that appear in the Cosmetics Inventory were considered as cosmetics-related chemicals. A large number of studies for these 190 cosmetics-related chemicals was reviewed by the COSMOS ILSI Europe Expert Group. This dataset in general has been also checked for record reliability, including study design, results, and references by the COSMOS team. More in depth QC of the Munro dataset is described later (Section 2.3.2).

2.2.3. NO(A)EL/LO(A)EL database

To establish a new database of non-cancer oral data that would be suitable for derivation of TTC values, additional data from existing sources were included. The outcome of this compilation resulted in a new NO(A)EL/LO(A)EL database (see Figs. 1 and 2), whose values can be used to determine PODs. Here, we distinguish between study NO(A)ELs/LO(A)ELs and PODs that were derived from NO(A)ELs/LO(A)ELs by the application of extrapolation factors for study duration and/or LO(A)EL to NO(A)EL extrapolation. With that, PODs reflect actual or estimated (i.e. extrapolated) chronic NO(A)ELs and allow for comparison of substances. To select appropriate studies with NO(A)EL/LO(A)EL values, a strict set of TTC study selection criteria has been established, as listed in Table 1, and applied throughout the curation process for both PODs and the final TTC dataset.

The study design parameters required for inclusion of the studies in the NO(A)EL/LO(A)EL Database (inclusion rules) are quite similar to those of the COSMOS MINIS criteria for the oRepeaTOX DB. Therefore, all data on NO(A)EL and LO(A)EL values from this new toxicity database were merged with the data from other regulatory or risk assessment sources. Various other POD values for non-cancer data were also based on the following: NOELs and LOELs from the Munro dataset, i.e. Munro chemicals found in the Cosmetics Inventory; NOAEL, NOEL, LOAEL, LOEL, BMD (Benchmark Dose), and BMDL (Benchmark dose lower 95% confidence limit) from the US EPA Integrated Risk Information System (IRIS); highest no-effect levels (HNELs) and lowest effect levels (LELs) from the US FDA PAFA database; no-effect levels (NELs) and LELs from the US EPA ToxRefDB. This integrated collection of test substances along with their NO(A)EL values served as a pool for the candidates for the COSMOS TTC dataset, which would then only list one selected POD per chemical.

The combined compilation is stored in the COSMOS DB v2.0 as a Safety Assessment Database for more than 1000 test substances, of which 660 were initially identified from the COSMOS Cosmetics Inventory as unique chemical structures. It should also be noted that percentages of chemical impurities and the active ingredient in the test substances used in toxicity experiments can vary widely, depending on methods of analysis. This aspect has not been uniformly considered in the development of TTC databases because such information is not consistently available for all studies, so that correcting some values but not others would lead to distortion of the database. Accordingly, no such corrections were made. For these 660 chemicals, preliminary NO(A)EL/LO(A)EL values were available for 558 chemical structures. These datasets were used to assess the chemical space and served as a basis for the first preliminary dataset for the initial TTC analysis for cosmetics-related chemicals (EFSA, 2012; European Commission, 2012; Worth et al., 2012).

2.2.4. COSMOS TTC dataset

2.2.4.1. Creation of the dataset. The candidate studies in the NO(A)EL/LO(A)EL Database, which provided a first round of initial NOAEL/LOAEL values, were selected from thousands of studies by applying a set of rules, as described in Table 1. The COSMOS TTC dataset was

Table 1
TTC study selection criteria in defining databases.

Parameters	NO(A)EL/LO(A)EL Database	COSMOS TTC dataset
Study type	Subchronic, chronic, carcinogenicity (non-neoplastic data only), reproductive, developmental, neurotoxicity, immunotoxicity.	Same criteria as in NO(A)EL/LO(A)EL database
Species	Rat and mouse (all studies), monkey and dog (all studies), rabbit (reproductive, developmental).	Same criteria as in NO(A)EL/LO(A)EL database
Duration	Greater than or equal to 28 days for subacute (short-term) and subchronic studies. For reproductive, developmental or multigeneration studies, requirement of "duration days" is not applied.	Same criteria as in NO(A)EL/LO(A)EL database
Route of exposure	Dietary, drinking water, gavage (or intubation)	Same criteria as in NO(A)EL/LO(A)EL database
Dose levels and range	All studies with dose level and regimen information are included. At least one control group is required.	Single dose studies not used. Separations between dose levels (low, mid, high) are reasonable.
Effects	All effects are recorded using controlled vocabulary.	Systemic effects.
Reference	Regulatory submissions, study reports, database sources, published literature (traceable citations).	Regulatory sources with guideline (GLP) studies preferred.

created by applying a further set of rules for POD selections in addition to the TTC study inclusion rules related to data interpretation for the candidate chemicals in the NO(A)EL/LO(A)EL database.

This new non-cancer COSMOS TTC dataset contains 552 structures that are mostly cosmetics-related chemicals found in the COSMOS Cosmetics Inventory (85%), with the rest being food contact substances and impurities. The largest sources for the substances in the COSMOS TTC dataset are US FDA PAFA and CFSAN documents, cosmetics chemicals in the Munro dataset, the EU Scientific Committee on Consumer Safety (EU SCCS/SCCP/SCCNFP), and EPA ToxRefDB.

Consideration was given to the inclusion of prohibited and restricted substances in the databases. Currently, 1379 substances are listed as prohibited or restricted in the use of cosmetic products by Annexes II and III of the EU regulation on cosmetic products (European Commission, 2009). Over 30 of these substances are still found in the CosIng database, many of them are botanicals and petrochemicals. Others include butane, isobutane, C21-C28 alkanes, dibutyl phthalate, diethylhexyl phthalate, butyl benzyl phthalate, oxyquinoline/sulfate, ergocalciferol, and retinoic acid. The prohibited and restricted list identifies a substance in conjunction with specific use category such that a substance can be prohibited for one use category, but still allowed to be used in another. In other cases, a substance might be banned from use in any cosmetic in the EU but not necessarily in all geographies. For this reason, 27 substances that are in the list of prohibited substances for use in cosmetic products in the EU remain included in the COSMOS TTC dataset.

The COSMOS TTC dataset consists of two domains. The first is a test substance-centered chemistry domain, containing substances used in the study, chemical structures, identifiers, physicochemical properties, CAS RN, and Cramer Class designation. For TTC, only those chemicals that are representable by structures and hence classifiable by Cramer Classes are considered. The second is a toxicity study domain, containing the background information, study design parameters and study references linked to the aggregated study results of NOAEL and LOAEL (or equivalent) values along with critical effects. The COSMOS TTC dataset is available from COSMOS DB v2.0 (Molecular Networks, 2017) and is also presented in the Supplementary Material to this paper.

2.2.4.2. Selection of the PODs. To select the POD for a given chemical for TTC derivation from the multiple NOAELs/LOAELs or equivalent data from various sources, the following procedure was systematically applied:

1. NOAEL decisions stated in the EC Scientific Committee opinions were in general accepted. In particular, the NOAEL/LOAEL value identified by the EC SCCS/SCCP/SCCNFP in their calculation of the Margin of Safety (MoS) was accepted "as is" and selected as the COSMOS POD with the highest priority. When questions arose due to large discrepancies between values from different data sources, careful reviews by the Expert Group were conducted.
2. The NOAEL/LOAEL or equivalent POD value (e.g. BMDL) used to derive an Acceptable Daily Intake (ADI) by EFSA, or the Joint FAO/WHO Expert Committee on Food Additives (JECFA), or used by the US EPA IRIS to derive a Reference Dose (RfD), was taken and used 'as is' unless conflicts were found with the decisions from step #1.
3. The NOEL/LOEL value determined by US FDA CFSAN or reported in the Munro dataset was also used "as is" if this was the only data source. When conflicts arose with other data sources, studies were reviewed as part of the QC process (2.3.2).
4. Substance entries in US FDA PAFA, US EPA ToxRefDB, or REACH (from the Registered Substance Database at ECHA) are associated with many studies with varying HNEL/NEL and LEL or NOAEL/LOAEL values for the chemical. If the data were of equal quality, then NOAEL/LOAEL values were determined by selecting values algorithmically according to the following rules. The data quality is defined in detail in 2.3.1.
   a. First the study with the lowest no effect level (HNEL, NOEL or NOAEL) that also had a clear lowest effect level (LEL, LOEL or LOAEL) was taken.
   b. If the minimum no effect level (HNEL, NOEL or NOAEL) was free standing (i.e., the highest dose tested), then the priority was given to an alternative pair with a clearly defined lowest effect level (LEL, LOEL, or LOAEL) value and a no effect level (HNEL, NOEL or NOAEL) value (as shown in Fig. 3).
5. The data from the literature or NTP technical reports were evaluated by the COSMOS TTC group and NOAEL/LOAEL values were extracted when necessary.
6. Where possible, NOAEL values were taken from chronic studies as the TTC values are intended to cover lifetime exposure. In cases where a shorter-term study was preferred over a chronic study, the database clearly lists the rationale for the choice of study.
7. NOAELs were adjusted for study duration by applying adjustment factors, as follows:

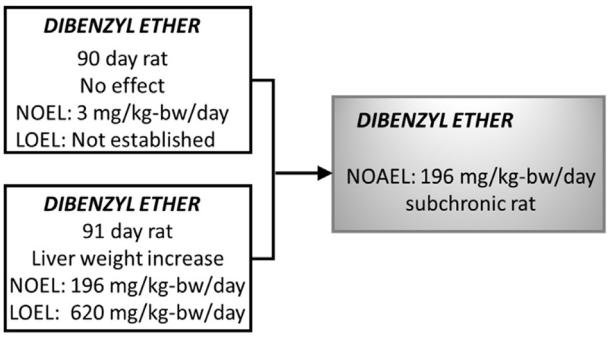


Fig. 3. Selection of NOAEL from a NOEL/LOEL pair with the lowest NOEL value.

a. For subchronic studies (84 days–179 days), an adjustment factor of 3 was applied to allow for chronic effects, as was used by Munro et al. (1996);
b. For short-term studies (28–83 days), an adjustment factor of 6 was applied to allow for chronic effects, as recommended in REACH guidance (ECHA, 2012);
c. For reproductive and developmental studies, no duration adjustment factor was applied, regardless of whether the effects were systemic or reproductive/developmental in nature. The effect of performing such an adjustment was evaluated; see Section 2.5.1).

8. If no NOAELs were found in the above steps, NOAELs were derived from the lowest available LOAEL by applying an adjustment factor of 3.
9. Where possible, studies conforming to internationally accepted guidelines/protocols were preferred. When such studies were not used, the database lists the rationale for the choice of study.
10. After all the NOAEL/LOAEL values were assigned for each chemical from each data source, an overlap profile of the PODs was prepared. For each chemical, other than the decisions from the EU SCCS/SCCP/SCCNFP, EFSA, US EPA IRIS and JECFA, the lowest NOAEL values were selected, except for a few cases where weight of evidence was applied, as documented in the Supplementary Material. The final NOAEL value for each chemical was then used to derive the final POD (POD = NOAEL adjusted for less-than-chronic study duration) for calculating TTC values for the COSMOS TTC dataset.

The above process is illustrated in Fig. 4. It is worthwhile mentioning that the COSMOS TTC preferred to use NOAEL/LOAELs rather than NOEL/LOELs for all chemicals which were subject to QC reviews. However, not all chemicals were reviewed in detail, and the above description of the different data sources demonstrates that multiple chemicals in the dataset are designated as having NOEL/LOELs. Furthermore, many older data sources, from a terminology perspective, do not distinguish clearly between NOELs and NOAELs. In practice, many of the NOAELs in the COSMOS TTC dataset are the same as those reported in the original sources, e.g. EC SCCS/SCCP/SCCNFP, US EPA IRIS, US FDA or key studies reported in REACH. The remaining NOAELs in the COSMOS dataset were decided on the "most appropriate" basis as described above.

2.2.4.3. Exclusions and inclusions from the database. Although it was ascertained that lipid soluble vitamins (A, D, E and K) and amino acids did not drive the TTC values (they did not affect the 5th percentile value of the NOAELs), they were nevertheless excluded from the database. For nutrients, the magnitude of the differences between intakes that are essential for normal physiological function and intakes that may be toxic can be relatively small. Hence, it is widely recognized that conventional risk assessment approaches for such substances are not appropriate since the application of default Uncertainty Factors of 100 to the PODs for toxic effects can give rise to values that would result in nutrient deficiencies. Nutrients have therefore been excluded from the COSMOS TTC dataset.

Proteins, inorganic substances, organometallic substances, coordination complexes and metals were also mostly excluded from the database, as they were in the Munro analysis. However, the COSMOS TTC dataset included the organosilicones (part of the organometallic class) to cover siloxane-based chemicals. Also included were oligomers and lower molecular weight polymeric surfactants whose repeating unit ranges are known (e.g. poly(-ethylene glycol) or alcohol ethoxylates).

2.3. Quality control of the databases

The three databases illustrated in Fig. 2, the oRepeaTOX, the NOAEL/LOAEL database, and the COSMOS TTC dataset were compiled from the outputs of many different regulatory and advisory agencies and data sources. Hence, it was critically important to review the data so that not only the factual records were standardized, but also the underlying information was evaluated to obtain coherence and the best possible NOAEL/LOAEL decisions. Thus, quality control (QC) included two step-wise components.

2.3.1. Data record reliability

QC for data record reliability entailed the checking of records in the database in comparison with the original sources to ensure that the database records truly represent the original sources. At this stage, the NOEL/NOAEL values (if specified by the document source) were not questioned *per se* – the QC evaluated only whether they were correctly recorded according to the inclusion criteria (see Table 1 and Appendix 1).

Data record reliability was also assessed to classify or filter out unacceptable studies. For data from an existing database, the quality scores from the data source were adopted. For example, the US FDA PAFA database and the US EPA ToxRefDB classify studies for "completeness" and "data usability", respectively. These standards use regulatory guidelines either from the US FDA Redbook (US FDA, 2000) or the US EPA Office of Chemical Safety and Pollution Prevention (OCSPP) (US EPA, 2003), respectively. For the two defining databases (oRepeaTOX DB and COSMOS TTC dataset), if the study design satisfied the respective agency's guideline, the study was deemed to "meet the current standards" or to be "acceptable". When the study was not compliant with the guideline but acceptable according to the COSMOS database inclusion criteria in Table 1 and Appendix 1, the study was classified as "not meeting the current standards, but meeting the core standards" or as "non-guideline, but acceptable". When the study did not meet the minimum standard for the COSMOS database, it was considered "unacceptable by not meeting the core standards" or "not usable".

New public literature studies were also harvested by COSMOS for the oRepeaTox DB. For these, data record completeness was assessed by establishing minimum study inclusion criteria and a scoring system. All studies conducted according to guidelines from the Organisation for Economic Co-operation and Development (OECD), US FDA Redbook (US FDA, 2000), and US EPA OCSPP (US EPA, 2003) were accepted. The COSMOS MINIS criteria, described in Appendix 1, are less stringent than the requirements of the OECD test guidelines but are similar to the US FDA PAFA core standards. The US FDA PAFA core standards were established for subchronic, chronic, and reproductive/developmental studies. The parameters necessary to be reported include study duration, animal species,

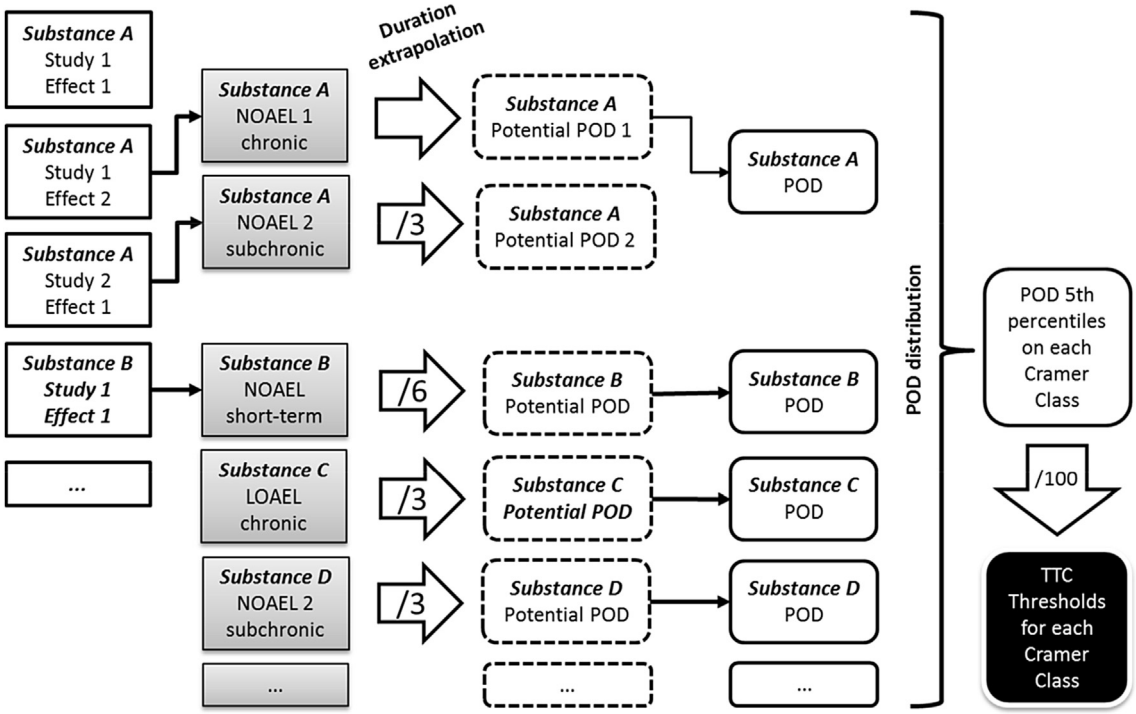


Fig. 4. Process for deriving TTC values from multiple studies.

route of exposure, animal age/weight, number of animal/dose/sex, control, number of the doses used, dosage regimen, clinical signs, water/food consumption, hematology, clinical chemistry, urinalysis, organ weight, general necropsy/macro pathology, and micro/histopathology.

2.3.2. QC for results interpretation

Further QC on a defined proportion of the substances in the COSMOS TTC dataset was undertaken by the COSMOS ILSI Europe Expert Group to define the reliability and relevance of the harvested studies to TTC for both chemistry and toxicity data. To ensure that NOAEL values were not simply driven by rote application of the algorithms described under Section 2.2.4.2, NOAEL selection was reviewed in detail and a consensus decision reached on the final NOAEL values to be assigned. A primary consideration was that effects used as the basis of the final NOAEL values should represent toxicologically relevant systemic effects and also be relevant to humans. A definitive decision on human relevance can typically only be reached in cases where there is detailed mode of action information. In the absence of such information, it was assumed that an effect could be relevant to humans. For instance, if there were histopathological changes in rat kidney, in the absence of any other information, it was assumed that this effect is relevant to humans. On the other hand, if there was information that the effect was due to a rat-specific pathway (e.g., binding to the male rat specific protein alpha 2μ-globulin), it was excluded (US EPA, 1991).

As a detailed toxicological review of all data on each chemical was resource prohibitive, two approaches were taken to prioritize substances for review. First, the most potent substances, those with the lowest 10% of NOAEL values in the entire dataset, were reviewed. These substances were considered "high-impact" since their potency can markedly affect the 5th percentiles of the NOAEL distributions. QC1 was conducted for all studies for which NOAEL values of the substances were found in the lowest 10% of the entire dataset; in subsequent QCs, records giving NOAELs in the lowest 10% of the values for each Cramer Class were reviewed. Consequently, all studies with NOAEL values under 5 mg/kg bodyweight per day (mg/kg-bw/day) were reviewed. Secondly, the studies on chemicals for which there were large conflicts (high variability) in NOAEL values across different data sources were reviewed. This group comprised chemicals with NOAEL values in the range of 5–50 mg/kg-bw/day if the maximum/minimum ratio was >5, and in the range of 50–500 mg/kg-bw/day if the maximum/minimum ratio was >10.

Also considered were the results of the EFSA (2012) QC of the Munro et al. (1996) dataset for the lowest 10% of NOEL values for 16 Cramer Class I and 50 Class III substances whose NOELs were in the lowest 10th percentile. The three NOEL values of the Class I substances rejected by EFSA were replaced by values assigned by the COSMOS ILSI Europe Expert Group during the QC process since they are found in the COSMOS Cosmetics Inventory; these substances were ethyl acrylate, methyl methacrylate, and triethylene glycol. NOEL values for non-cosmetics chemicals such as phenyl-1-propanol-2 (Class I), azinphos methyl (Class III) and coumaphos (Class III) were also rejected by EFSA, and hence were not included the Munro dataset used by COSMOS. The NOEL for ascorbic acid (Class I) deemed "not verifiable" by EFSA was removed from the Munro dataset. Of the 190 substances in common (overlap) in both Munro and COSMOS TTC datasets, 40 NOEL values were scrutinised under QC by the COSMOS ILSI Europe Expert Group and these values were used for both COSMOS TTC and Munro datasets. Most of the rest of these overlapping 190 substances were also reviewed for record reliability including study design, results, and references. At the end, based on various QC results from EFSA and the COSMOS ILSI Europe Expert Group, 91 studies from the Munro dataset were included in the COSMOS TTC dataset as the basis of POD. All 190 substances common to both Munro and COSMOS TTC datasets are

now represented with the same NO(A)EL values and studies in the COSMOS compilations. This new Munro dataset, after QC of the data records and study QC by the Expert Groups (both from EFSA and COSMOS ILSI Europe), is here referred to as "Munro-2016". The Munro dataset file downloadable from the COSMSO DB web site contains both Munro-1996 (Section 2.2.2) and Munro-2016 (COSMOS QC version) with documentation of QC status and rationales.

The US FDA PAFA database covered a chemical space that was very close to that for cosmetics due to similar substance use types; in many cases, it was the only data source for many cosmetics-related chemicals. The US FDA PAFA database was put together to assess chemical safety with a hazard identification perspective (Benz and Irausquin, 1991). Of the 552 chemicals in the COSMOS TTC dataset, 220 chemicals appear in the US FDA PAFA database and 91 PODs were derived from US FDA PAFA data. More than 50% of the FDA PAFA chemicals used in this TTC dataset were Cramer Class I. Nearly 25% of the US FDA PAFA chemicals in the TTC dataset had POD values greater than 500 mg/kg-bw/day, whereas only 10% of the rest of the TTC dataset was found within the same potency range. All of the US FDA PAFA studies resulting in a POD equal to or less than 5 mg/kg-bw/day were reviewed under expert QC; more than 50% of the US FDA PAFA studies resulting in a POD of between 5 and 50 mg/kg-bw/day were also reviewed by experts. The ILSI Europe Expert Group also revisited some of the substances in the Munro dataset and replaced the data for isopropyl alcohol and ethanol with that from more reliable studies. Data for dinocap and linamarin (Cramer Class III) were removed as the toxicity data providing the lowest NOEL values were from hamster studies, which were excluded by the COSMOS MINIS criteria (Table 1 and Appendix 1). In addition, in the COSMOS TTC dataset, the studies which used only one dose level of the cosmetics-related chemicals in the Munro dataset were also not included. In cases in which the overall NOAEL was changed as a result of the QC work and the same substance was also present in the Munro dataset, the original NOEL was replaced by the new, revised NOAEL value in the Munro-2016 dataset.

Some phthalates and parabens are included in the COSMOS TTC and Munro datasets, but NOAELs for some of these types of substances have changed (reduced) considerably over time and some of the toxicity endpoints studied more recently have not resulted in consensus regarding repeatability and relevance to human health. Accordingly, these substances were scrutinised in the QC process and some of the NOAELs were reassigned, including lowering of some of the NOELs in the Munro dataset.

Although the QC process scrutinised the data for the COSMOS TTC dataset, we have not reviewed all of the data, putting more emphasis on the most potent chemicals since they are the ones that affect the human exposure thresholds. Overall, 91 POD values were determined by the ILSI Europe Expert Group and included in the final COSMOS TTC dataset along with the additional 223 reviewed by the master curator of the COSMOS DB team. A chronological summary of the COSMOS QC process is shown in Table 2.

In addition, during the QC process, new supporting studies, where available, were added. In making final consensus decisions on NOAEL values to be assigned, the following general criteria were applied to these studies:

- NOAELs should be based on systemic effects;
- the studies should have Klimisch scores (Klimisch et al., 1997), assigned by the ILSI Europe Expert Group, of "reliable without restriction" (score 1) or "reliable with restriction" (score 2); studies with Klimisch scores higher than this (score 3 – not reliable, or score 4 – not assignable) not to be used;
- only effects with relevance to humans should be included (default assumption is that effects are relevant unless there are convincing data to demonstrate otherwise);
- NOAELs from regulatory sources should be preferred, when available.

The above criteria were applied to any relevant adverse effects reported in the available studies regardless of potential mode of action, e.g. adverse effects by cytotoxicity were handled the same as adverse effects by endocrine mechanisms. On completion of the QC process, the NOAEL decisions were documented and the resulting COSMOS TTC dataset was finalised.

2.4. POD distribution and threshold development

2.4.1. The 5th percentile POD values

The 5th percentile value for each Cramer Class was determined from the cumulative distribution function (CDF) of the POD values, derived as described in Section 2.2.4.2 and Fig. 4. The names of the substances within the 5th percentile group for each Cramer Class, together with their POD values, are listed in the Supplementary Material under "Quantiles". To derive robust threshold values for TTC, the 5th percentiles were determined from either parametric fitting by assuming a lognormal distribution or by non-parametric estimation of empirical values. The parametric curve fitting of lognormal distribution requires only the estimates for mean and the dispersion parameters, e.g. standard deviation of a sufficiently large dataset, obtained from fitting each data set to a lognormal distribution. This method provides a common standard that does not depend on interpretations or interpolations as long as the log-transformed data can be assumed to be normally distributed. Non-parametric evaluations do not assume that the data are normally distributed, but often apply smooth interpolation techniques ranging from simple to empirical smoothed quantiles. In this study, non-parametric estimations based on smoothed empirical likelihood quantiles using a kernel density estimation were performed (Silverman, 1998). Both parametric and non-parametric estimations were calculated from MatLab R2013b (MathWorks (2017), JMP Pro 11.2.1 (SAS institute) (JMP, 2017), and R-3.3.2 (R-Project) (The R Project for Statistical Computing, 2017).

To establish the baseline for the method employed in this study, the Munro thresholds were estimated using the same data provided in the appendix of the Munro et al. (1996) publication. Since the objective was to confirm that the published values can be reproduced, data printed in the 1996 paper were used verbatim even in those cases where records were clearly erroneous (e.g. triethylene glycol with incorrect NOEL due to dose unit error). Under these constraints, the parametric estimations of the 5th percentile NOEL values for Cramer Class I, II, and III of the Munro-1996 dataset were 2.90, 0.90, and 0.15 mg/kg-bw/day, respectively. These values are for all practical purposes the same as the published values of 3.0, 0.91, 0.15 mg/kg-bw/day, estimated by a parametric method, for the 5th percentile of Cramer Class I, II, III, respectively. Hence, all parametric estimation in this study was based on fitted lognormal distribution, uncentered and unscaled quantiles.

Munro et al. (1996) also reported non-parametric estimations of 5th percentile NOEL values of 3.3, 1.6 and 0.12 mg/kg-bw/day for Cramer Class I, II, and III, respectively; however, the nature of the non-parametric method used was not indicated in their paper. In this COSMOS study, based on the non-parametric method described above, the 5th percentiles of the distributions of the Munro-1996 dataset were 2.93, 0.91, and 0.13 mg/kg-bw/day for Cramer Class I, II, and III, respectively. However, as Munro et al. (1996) did not describe the method used, comparison of the

Table 2
Summary of the study QC process of the COSMOS TTC dataset by the ILSI Europe Expert Group.

QC	QC description	Results
Initial preliminary dataset (2011)	• Data record QC	• 660 test substances; 558 structures (v1.3) • 385 structures (v1.2) (Worth et al., 2012)
QC1 (2012–2013)	• Potent chemicals: the lowest 10% of the whole dataset • Data variability (NOAELs): greatly differ across the data sources	• 68 unique test chemicals were evaluated (v1.4, v1.5) • Result: 460 (v1.6)
QC2 (2013–2014)	• Potent chemicals: the lowest 10% of each Cramer Class • Data variability (NOAELs): chemicals greatly differ across the data sources	• 57 compounds were evaluated • Result: 562 (v1.7)
QC 2a (2014)	• Compound classes (phthalates, parabens) • Cosmetics-related chemicals (4) for which data were deemed questionable by EFSA's QC of the Munro DB • Cramer Class evaluation by COSMOS experts	• 5 parabens, 9 phthalates, and 4 other unreliable data • Result: 558
QC3 (2014–2015)	• Potent chemicals: the lowest 10% of each Cramer Class • Data variability (NOAELs): chemicals greatly differ across the data sources	• 10 compounds reviewed • Result: 560 (v1.8, candidate for final)
COSMOS DB QC	• Data variability (NOAELs): chemicals greatly differ across the data sources • Remove intractable data or replace with more reliable data	• Additional 92 compounds reviewed • Result: 552 (final)

respective values is not informative.

The influence of a number of factors on the distributions in this study was evaluated without assuming normality. The normality test was performed using the Shapiro-Wilk method (Shapiro and Wilk, 1965); pairwise comparisons of the distribution of the Cramer Classes were also performed using the non-parametric pair-wise Kolmogorov-Smirnov (pair-wise K-S) test (Conover, 1999). It is worthwhile to note that the 28 NOEL values of Cramer Class II failed to meet the normality test, possibly reflecting the small number of values available. The NOEL distributions of the Cramer Class I and II pair, and the Cramer Class I and III pair were found to be significantly different. This is discussed in more in detail in Section 3.3.2.

2.4.2. Human exposure threshold values

Munro et al. (1996) developed human exposure threshold values (TTC values) based on the parametric estimation of the 5th percentile NOELs for each Cramer Class after applying a 100-fold safety factor to the 5th percentile POD values (as illustrated in Fig. 3). The TTC values derived by Munro et al. (1996) were 1800, 540 and 90 μg/person/day (person per day) for a 60 kg person, equivalent to 30, 9 and 1.5 μg/kg-bw/day, for Cramer Class I, II and III respectively. The same method was applied in this project to derive TTC values.

2.5. Sensitivity analysis

2.5.1. Duration extrapolation factors for developmental and reproduction studies

For some substances, the lowest reported NOAELs originated from systemic toxicity effects on parental animals in developmental and reproductive toxicity (DART) studies. This raised the question of whether it was necessary to adjust the NOAEL values of non-DART effects for the shorter than chronic exposure duration in most DART studies. Non-DART effects included body weight changes (parent, weanling), organ weight other than reproductive organs, mortality and clinical signs in adult, food/water consumption, and maternal toxicity. For such findings, applying duration adjustment factors for non-DART effects is hampered by the lack of a precise description of the exposure duration for parental animals in many studies. The DART effects include reproductive effects, reproductive organ effects, delayed/retarded ossification, teratogenic/malformation effects, embryotoxicity, and embryo-fetal development. No duration extrapolation factor was applied to DART effects unless specified by Munro publications when dealing with the Munro-1996 analysis.

The impact of treatment duration in DART studies on non-DART effects was systematically evaluated for both the COSMOS and Munro-2016 TTC datasets by investigating the NOAEL distribution and hence the 5th percentile value for the distributions. The following duration adjustments were chosen by COSMOS and applied to the NOAEL values:

- 1-generation studies: a factor of 6 was applied for short-term duration (approximately 56 days for mouse, 70 days for rat).
- 2-generation studies: a factor of 3 was applied for duration equivalent to subchronic studies.
- ≥3-generation studies: considered as chronic duration, hence not adjusted.
- Maternal effects in (pre-natal) developmental studies were not adjusted since they arose from dosing of the dams during the already more sensitive period of gravidity.

For the COSMOS TTC dataset, 48 NOAEL values were derived from such studies; only one case was reported without the specific study duration. For the Munro-1996 dataset, 91 reproductive and multigeneration studies were cited with 63 records assigned to duration of Not Given (NG); during the QC (Section 2.4), 13 duration records were entered.

Analyses of the impact of duration extrapolation factors for non-DART effects in reproductive/multigeneration studies were conducted for all COSMOS TTC, Munro-1996, and Munro-2016 datasets. Comparisons were made for the changes in distributions; statistical inference was made by applying the pairwise K-S test.

2.5.2. Substance types and chemical classes

The impact of various substance types and chemical classes on the cumulative distribution and the resulting 5th percentile POD values were also evaluated. These substances include possible nutrients (see Section 2.2.4.3), hair dyeing agents, and the chemicals in the list of substances prohibited (European Commission, 2009) in cosmetic products in EU as well as organophosphates and carbamates.

The database was not evaluated for the presence of chemicals with potential for bioaccumulation.

2.6. Cramer class evaluation

2.6.1. Cramer classifications by Toxtree and Munro

Cramer Classes are given in 1996 publication for chemicals listed

Table 3
Comparison of Cramer Classifications between Toxtree v2.6.13 and Munro-1996.

	Toxtree v2.6.13 Class I	Toxtree v2.6.13 Class II	Toxtree v2.6.13 Class III
Munro Class I	106	7	24
Munro Class II	3	15	10
Munro Class III	4	2	438

in the Munro dataset. In this study, Cramer Classes were assigned using various versions of Toxtree (2017). To compare the classifications between Munro and Toxtree, the Structured-Data (SD) file was batch-processed within Toxtree. Metal ions (Na+, Ca2+, or Fe3+, etc.) were not removed from the connection table due to the nature of the questions in the decision tree (Cramer et al., 1978). The comparison of assignments for 609 unique structures between Munro and Toxtree v2.6.13 are summarized in Table 3.

The patterns of discrepancies are originated from the interpretations of metal ions for salt forms and azo dyes or recognition of easy-to-hydrolyze esters and metabolically-active functional groups. Classifications of these fifty chemicals have been manually reviewed and resolved by the COSMOS Chemistry QC. The results of this comparative analysis are captured in the Munro-2016 dataset and the rationales are also documented in the export file available from the COSMOS DB v2 TTC export site (Molecular Networks, 2017).

2.6.2. Cramer Class QC for the COSMOS TTC and Munro overlap

During the course of this 5-year project, several versions of Toxtree v2.5 and v2.6 were released. A number of conflicts in the Cramer classifications assigned by the various versions of Toxtree v2.5 and v2.6 were also identified and resolved by COSMOS Chemistry QC. In addition, there were 33 common chemicals between COSMOS TTC and Munro-1996 dataset, whose Cramer classifications were in conflict. These discrepancies between Toxtree and Munro-1996 have been resolved as part of COSMOS chemistry QC. In addition, the patterns that emerged from the comparative analysis mentioned in 2.6.1 were applied to the rest of COSMOS TTC dataset. The Cramer Classes used for the COSMOS TTC dataset are documented in detail in the export file available from the COSMOS DB v2 (Molecular Networks, 2017). The results listed in the export file were obtained using Toxtree v2.6.0. The final comparisons in this manuscript were made employing Toxtree v2.6.13. No major differences in assignments were found between the two versions.

2.7. Construction of a federated dataset based on COSMOS and Munro TTC datasets

Although the new TTC dataset enriched with cosmetics-related chemicals is much needed, it is also desirable to have one master TTC dataset for non-cancer endpoints. To this end, a larger set encompassing greater chemical space based on both COSMOS and Munro datasets was established. The federated approach does not force integration or merging of all the records as one physical entity, but allows the construction of this virtual entity for searching and analysis. Three datasets can be identified: COSMOS TTC, Munro, and the overlap.

There are numerous practical issues that present challenges in joining datasets to build one federated set These include consistent study inclusion criteria, regulatory perspectives, and enforcing the same decision making process. These issues lead to study selection issues when the two PODs from COSMOS and Munro datasets are based on different studies or conflict even when from the same study. Fortunately, in this project due to our data curation approach (Section 2.3), approximately 30% of the studies cited in Munro-1996 had already been subject to expert QC, and many more studies were subject to database QC for record reliability to yield the Munro-2016 dataset. Another problem also recognized earlier in the project was that there are discrepancies in Cramer Class assignments between Munro et al. (1996) and those obtained using cheminformatics tools such as Toxtree or the OECD Toolbox. To further support the analysis of a federated set of COSMOS and Munro, an additional 45 Munro structures that are not part of the COSMOS TTC dataset were reviewed to resolve the conflicts between the classifications by Munro and Toxtree v2.6.0/v2.6.13 (see Appendix 2). The resulting dataset is downloadable from the COSMOS DB TTC workflow (Molecular Networks, 2017).

2.8. Characterisation method for the cosmetics chemical space

2.8.1. Molecular properties

The chemical space of the COSMOS TTC dataset was characterized from the perspectives of both structural features and physicochemical properties. The structural feature space was described by ToxPrint chemotypes (Toxprint, 2017). The use of this method to profile the chemical space of inventories and databases has been reported previously (Richard et al., 2016; Yang et al., 2015). The Structure-Data (SD) files of the Munro and COSMOS TTC datasets were prepared based on structures in the COSMOS DB v2. The fingerprint files based on 729 ToxPrint chemotypes were generated using the ChemoTyper software tool (Chemotyper, 2017).

The property space of the datasets was explored by publicly available CORINA Symphony Descriptors Community Edition web service provided by Molecular Networks GmbH, Nürnberg, Germany. The whole molecule properties employed to profile the Munro and COSMOS TTC datasets included logP, topological polar surface area (TPSA), complexity (computed based on paths, branching, atom types), dipole moment, water solubility, and molar volume.

2.8.2. Visualization methods

The chemical space of datasets was compared by principal components projections (C Yang et al., 2008) and hierarchical clustering methods. These techniques can use both structural chemotypes and properties.

Principal component analysis (PCA) is a multivariate data analysis technique that reduces the high dimensionality of domain (such as chemical space) and helps to represent the variations with a few latent variables. In this study, PCA has been applied using both structural feature and molecular property space. For structural feature space, 179 chemotypes matching 4 or more structures were pre-selected and linearly combined to form principal components (PC). The scores of a few of these PCs were then plotted to visualize the grouping of structures. More detailed methods using PC projections based on structural features have been published elsewhere (C Yang et al., 2008). For property space, the PCs were extracted using a set of 13 molecular properties (molecular weight, number of H donors and acceptors, XlogP, TPSA, polarizability, dipole moment, aqueous solubility, number of Lipinski rule-of-five

Table 4
Data sources of substances in the COSMOS TTC dataset.

EU SCCS	ECHA	EFSA	FDA PAFA	FDA CFSAN	EPA IRIS	EPA TOXREF	NTP	JECFA	MUNRO	
153	5	4	3	2	1	6	5	2	3	EU SCCS
	36	4	2	2	0	6	4	2	0	ECHA
		25	4	1	0	3	1	4	2	EFSA
			220	83	5	24	28	21	36	FDA PAFA
				131	3	4	3	5	5	FDA CFSAN
					43	8	7	0	25	EPA IRIS
						140	35	4	24	EPA TOXREF
							79	2	38	NTP
								98	93	JECFA
									190	MUNRO

violations, molecular complexity, ring complexity, and diameter).

Two-dimensional clustering against both molecular properties and ToxPrint chemotypes was performed. Based on their presence in more than 4 structures, 241 ToxPrint chemotypes were used for hierarchical clustering of the Munro and COSMOS TTC datasets. The structures were also clustered using the same set of 13 molecular properties as in the case of PCA. When clustering with structural features such as ToxPrint chemotypes, average linkage method with Jaccard distance was employed. Ward linkage method with Euclidean distance was applied for molecular properties. Against the two-dimensional dendrogram (the first for structural features using ToxPrint chemotypes and the second for molecular properties), each compound was plotted in a scatterplot. In addition, the compounds in each of the Cramer Classes can be clustered separately using either ToxPrint chemotypes or molecular properties. This analysis can be used to illustrate the structural similarities and differences between Cramer Classes and between the TTC datasets.

3. Results

3.1. Chemistry characterisation of the COSMOS TTC dataset

3.1.1. Profile by data sources

The final COSMOS TTC dataset consists of 552 substances and NOAEL values, which originated from over 1000 studies from 10 different sources. The number of chemicals from each data source and their overlapping chemical coverage are compared in Table 4. Although there are 613 substances in the Munro dataset, only 190 of these substances (178 unique chemical structures) are considered as cosmetics-related chemicals by the COSMOS Cosmetics Inventory and are included in the COSMOS TTC dataset. The initial sources, in the order of where most of the data came, were US FDA PAFA database, EU SCCS opinions, Munro dataset, US FDA CFSAN public documents, and US EPA ToxREFDB. Most of the data on Munro substances came from JECFA, US EPA IRIS, and NTP reports. Other minor data sources include the European Medicines Agency/European Agency for the Evaluation of Medicinal Products (EMA/EMEA), Deutsche Forschungsgemeinschaft (DFG), US EPA public documents, Report to the US Consumer Product Safety Commission by the Chronic Hazard Advisory Panel (CHAP), and open literature articles.

Although initially a large number of studies were from US FDA PAFA, Munro, and ToxREFDB, at the end their contributions to the NOAEL values in the COSMOS TTC set was much reduced after the QC process. For example, only 91 NOEL values were used out of Munro's 190 values. Likewise, only 91 HNELs from the PAFA database were selected out of the initial count of 220 candidates. Fig. 5 depicts the contributing NOAEL sources in the COSMOS TTC dataset.

3.1.2. Profile by chemical space

3.1.2.1. Structure space. The chemical space is characterized by ToxPrint chemotypes and physicochemical properties using various categorization methods described in Section 2.8.

The COSMOS TTC dataset differs from the Munro dataset in that it is enriched with substances used as skin and hair conditioners, humectants, hair dyes, perfumes/fragrances, antimicrobials, emulsifiers, surfactants, and plasticizers. The resulting differences are compared in Fig. 6 using the ToxPrint chemotypes. Both Munro and COSMOS TTC sets are compared for each chemotype; the longer the bar, the higher the frequency of the chemotype in the dataset. Chemical groups with little or no representation in the Munro dataset include non-ionic and cationic surfactants as well as organosilicone and siloxane compounds. The Munro dataset contains higher numbers of organohalides, steroids (none in COSMOS TTC set), and ureas. There were 44 organophosphorus (OP) chemicals found in Munro, which were all considered OPs involved in

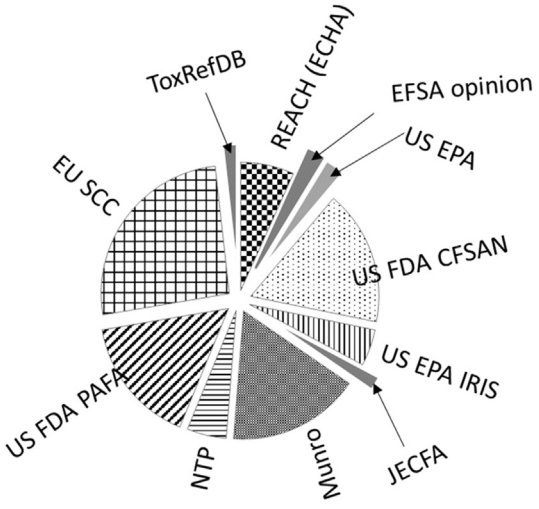


Fig. 5. NOAEL sources in the COSMOS TTC dataset.

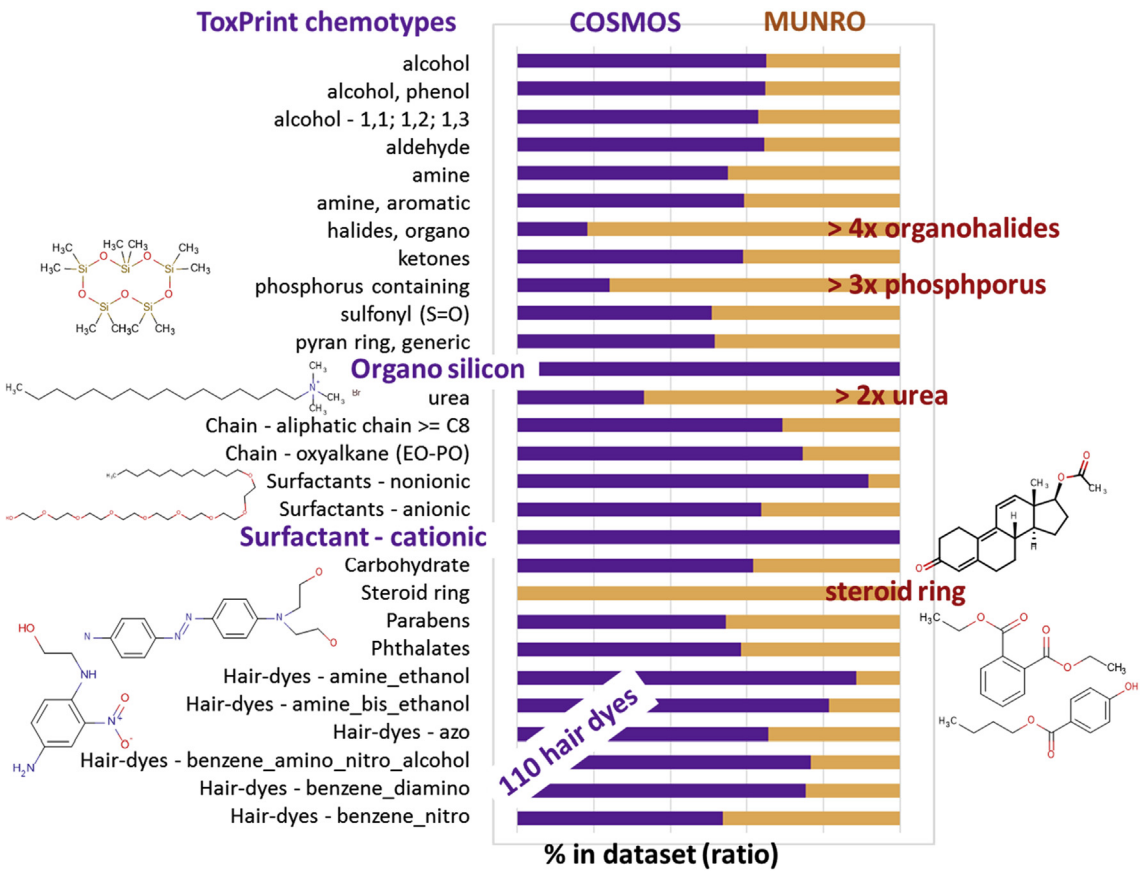


Fig. 6. Histogram of ToxPrint chemotypes of the chemicals in COSMOS and Munro TTC datasets.

acetylcholinesterase inhibition, except for two, inosinic acid and its salt. In the COSMOS TTC dataset only one OP is included; all the other five were phosphorus-containing flavouring agents, e.g. inosinates and guanylates. In the Munro dataset 32 carbamates were found, whilst only 3 were found in the COSMOS TTC dataset.

The COSMOS TTC dataset is enriched with chemicals used as hair dyeing agents. These chemicals were intentionally included to provide a realistic coverage of the full spectrum from low to potentially high safety concerns for cosmetics-related chemicals, which also enhanced the structural diversity and coverage, particularly in the important category of Cramer Class III. The hair dyeing agents are represented in Fig. 6 as nitro benzene, diamino benzene, amino nitro phenol, azo, and ethanol amines.

The comparison of chemical space can also be visualized as multivariate, as illustrated in Fig. 7, using the PC-score projections based on the selected ToxPrint chemotypes (method described in Section 2.8). The Munro (non-cosmetics) and COSMOS TTC datasets are quite well separated in the latent variable space. The separation of the two groups was close to 90-degrees to each other, which means that their chemical space share little common chemotype profiles. The Munro chemicals that are heavily loaded on the PC3 are also mostly Cramer Class III structures. Only a handful of COSMOS structures appear in this part of the chemical space of Munro Class III. They are Red 28, deltamethrin, tetrabromophenol blue, and triclosan. The overlaps (blue diamond in Fig. 7) are Munro chemicals appearing in the Cosmetics Inventory defined in Section 2.1, which tend to mostly cluster with the COSMOS dataset.

Both analyses, depicted in the distribution bar chart and the PC projection plot based on the ToxPrint chemotypes, provide assurance that concerns about the cosmetics-relevant chemical space of the current TTC approach can be resolved using this new COSMOS TTC dataset. Furthermore, the analyses confirm that the extension of chemical space by combining the two datasets is significant.

3.1.2.2. Property space. The chemotypes that are unique in cosmetics collections such as the COSMOS TTC dataset include surfactants (hydrophobic tails and hydrophilic heads), silicones and siloxanes. These surface-active chemicals give rise to physicochemical and molecular surface properties that are distinctively different from those of the chemicals in the Munro dataset. Fig. 8 illustrates how the set of molecular properties defined in Section 2.8 describes the chemical space of COSMOS and Munro TTC datasets through a PC projection scores plot. In general, these molecular properties do not clearly differentiate the COSMOS and Munro TTC datasets. While showing the loadings mostly on PC3, there are almost two separate clusters with positive and negative scores for compounds in both COSMOS and Munro datasets (Fig. 8).

Cosmetics ingredients tend to have more extreme values in polarity and diameter whereas the Munro dataset has more of smaller more non-polar structures. It is notable that chemicals used in cosmetics formulations are scattered much more widely. Areas of non-ionic and cationic surfactants as well as long alkyl chain carboxylic esters can be easily identified.

3.1.2.3. Combined structure-properties space. Chemical space can be also characterized by both structure and properties at the same time using a 2-D clustering technique. The multivariate hierarchical

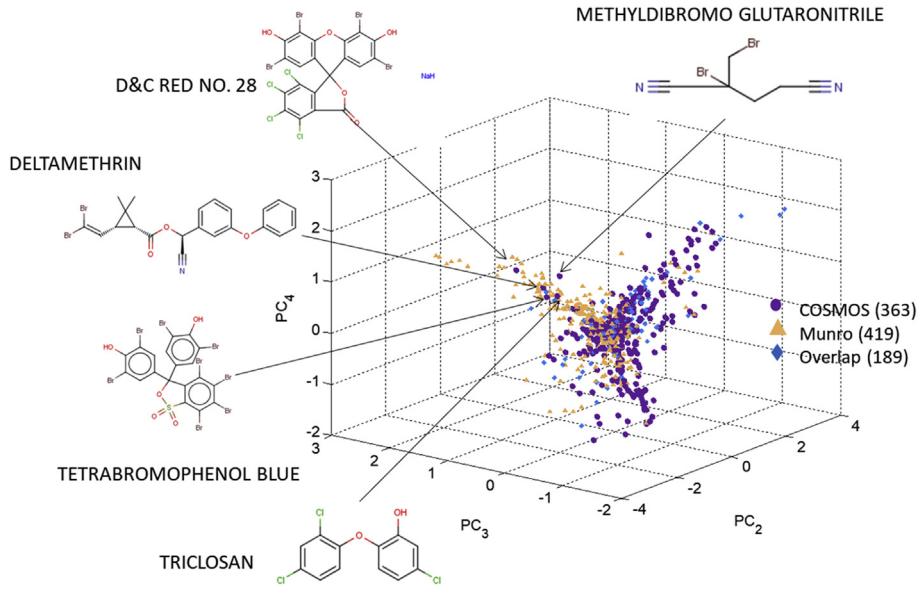


Fig. 7. Principal Component Scores projection for COSMOS and Munro TTC datasets.

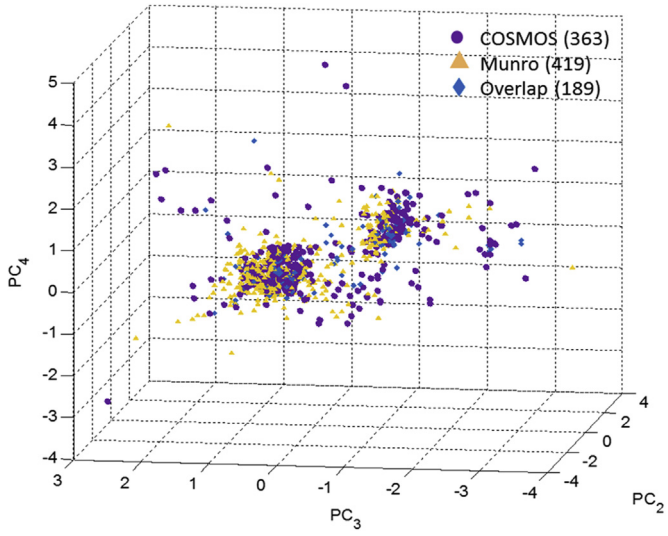


Fig. 8. Properties space of COSMOS and Munro TTC datasets.

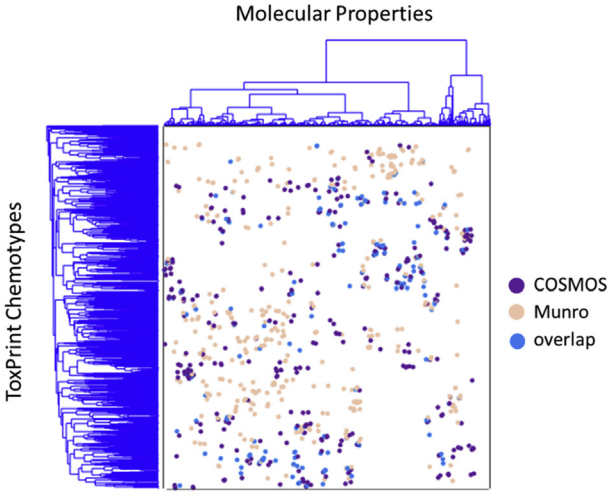


Fig. 9. 2-D clustering of Datasets by ToxPrint Chemotypes and Molecular Properties.

clustering method was applied to ToxPrint chemotypes and the whole molecule properties.

The pattern in Fig. 9 shows that Munro chemicals are smeared throughout the 2-D clustering map of structure and properties, indicating that the collection exhibits very diverse characteristics both in chemotypes and molecular properties.

On the other hand, the COSMOS chemicals tend to group more tightly in clusters, indicating that there are more local areas where structures are more highly correlated with properties in the COSMOS dataset. This observation is consistent with the fact that physicochemical properties are important in determining the uses of cosmetics.

3.1.3. Profile by Cramer Classification

3.1.3.1. Results of evaluation of Cramer Classification by Toxtree. Cramer classification is one of the central paradigms of the current TTC approach, where the toxicological potency is correlated to structural classes. Therefore, the Cramer Classes have a large impact on the 5th percentile NO(A)EL values and TTC values. Shortcomings of Cramer classifications using the OECD toolbox or Toxtree have already been well documented in previous publications (Bhatia et al., 2015; Roberts et al., 2015).

The comparative analysis of Cramer Class assignments from Munro et al. (1996) and Toxtree gave some additional insights. The source of the discrepancies seems to be mostly due to the knowledge of chemical reactivity and metabolism. Munro seemed to have applied the knowledge implicitly to such classes as aliphatic (alkenyl and allyl) esters, sucrose esters, and 2-butanol. Some examples of the conflicts between Munro and Toxtree are summarized in Table 3 and Appendix 2. For example, sucrose esters such as sucrose palmitate or stearate are easily hydrolysed to sugar and fatty acids, which are natural constituents of the human body and diet, so that the esters would be allocated to Cramer Class I based on the original rules while Toxtree placed them into Cramer Class III. Another example deals with tautomers (e.g., inosine ring where the oxo or oxy forms of the purine-ring are the tautomers). Knowledge

Table 5
Distribution of Cramer Classes in the COSMOS and Munro TTC datasets.

	Cramer Class I	Cramer Class II	Cramer Class III	Total
COSMOS[a]	219	40	293	552
Munro-1996[b]	137	28	448	613
Munro-2016[c]	141	30	435	608
COSMOS/Munro overlap[a]	112	21	57	190
Federated set	243	49	671	963

[a] These counts are the results of COSMOS reviews of the Cramer Classes as described inSection 2.6.
[b] For Munro-1996 dataset, the assignments presented in the article (Munro et al., 1996) were strictly followed.
[c] For Munro-2016 dataset, COSMOS reviews were followed as described in Section 2.6.

Table 6
Profile of studies and species in COSMOS TTC dataset.

	Median NOAEL (mg/kg-bw/day) and counts		
	Chronic/Carc/Combined[a]	Short-term/Subchronic	DART[b]
Rat	212.5 (N = 96)	95.7 (N = 271)	100 (N = 103)
Mouse	168 (N = 14)	100 (N = 11)	563 (N = 5)
Dog	37.2 (N = 12)	125 (N = 16)	15 (N = 1)
Monkey	0.2 (N = 1)	None	4.1 (N = 1)
Rabbit	None	None	23 (N = 18)

[a] Chronic, carcinogenicity, and chronic/carcinogenicity combined studies.
[b] Reproductive/developmental including multigeneration reproductive studies.

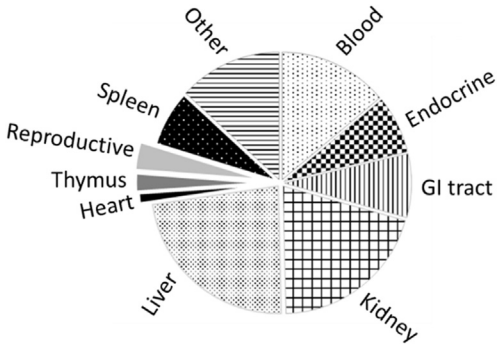


Fig. 10. Target organ profile of critical studies in the database.

of tautomers and mesomers becomes important in evaluating certain structures of multiple ring system colorants or hair dyes, which Toxtree does not handle well. Furthermore, the assignments of sodium or calcium salts to Class III by Toxtree has been already documented elsewhere for its interpretation of Rule 4 (Lapenna and Worth, 2011; Patlewicz et al., 2008). Although converting the salts to neutral species when preparing structure files is common practice, caution is recommended since the original Cramer rules are related to the metal salts. The same patterns were also observed in the COSMOS TTC dataset, which was accordingly corrected.

3.1.3.2. Cramer Class distribution. In comparison to the Munro dataset, the COSMOS TTC dataset is enriched with Cramer Class I chemicals and is well balanced between Class I and III. Table 5 compares the numbers of chemicals in each Cramer Class of the two TTC databases. As described previously (Section 2.6), conflicts between the Cramer classifications of Munro and COSMOS TTC datasets were evaluated manually by COSMOS Chemistry QC.

Approximately 75% (103 out of 137) of the Munro Cramer Class I chemicals are cosmetics-related substances. In the overlap between the Munro and COSMOS TTC datasets, nearly 60% are Cramer Class I chemicals, whereas only 30% are assigned to Cramer Class III.

3.2. Study profile of the COSMOS TTC dataset

Although the COSMOS TTC approach preferred chronic toxicity NOAELs, the most abundant studies in the resulting dataset turned out to be subchronic/short-term studies (54%), in particular, rat subchronic studies (49%), as listed in Table 6. The frequency of chronic, carcinogenicity, and combined chronic/carcinogenicity studies in rats (17%) was similar to that of DART studies in rats (19%). As described in section 2.2.2, the Munro dataset contains 27% subchronic rat, 31% chronic/combined carcinogenicity rat, and 21% DART rat studies. The COSMOS TTC dataset included 103 DART studies in rats. The high frequency of DART studies in both datasets demonstrates that DART effects are well covered by the TTC approach. The profile of the COSMOS TTC dataset in terms of study types, species and potency of critical effects is also illustrated in Table 6. Chemicals tested in subchronic rat studies were in general of higher toxicity than those tested in chronic studies. For POD derivation, the subchronic and short-term NOAELs were further divided by study duration factors of 3 or 6, respectively.

All other combinations of studies and species did not provide statistically large enough sampling size to make comparisons. The small number of studies in species such as mice, dogs, monkeys and rabbits compared to rats imposes limitations on statistical analysis of the influence of species. It should be noted that the Munro-1996 dataset does not include any dog or monkey studies. Furthermore, dog and monkey studies have a limitation compared to rodent studies in that much smaller numbers of animals per dose group are generally used. The most common target organs for these cosmetics-related chemicals are liver, kidney, endocrine system (e.g. adrenal, thyroid, pituitary), spleen, and gastrointestinal tract as shown in Fig. 10. Target organ effects are mostly represented by organ weight changes and pathology changes (macroscopic and microscopic). Most common general signs of toxicity include body weight changes, and food/water consumption changes.

There were 91 chemicals with critical effects in rat liver, of which 61 originated from subchronic rat studies. For kidney, 61 chemicals were identified with critical effects in rats and 59 were from rats in subchronic/short-term studies.

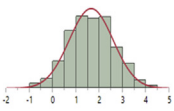
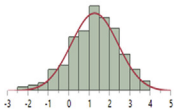
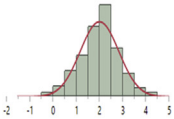
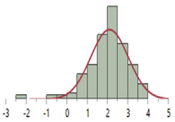
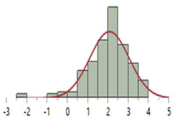
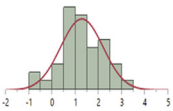
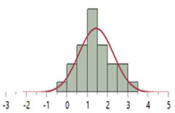
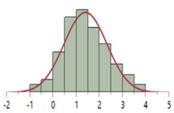
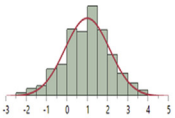
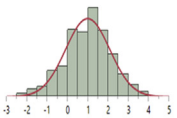
3.3. POD distribution of the COSMOS TTC dataset

3.3.1. General comparisons of POD distribution

The logPOD distribution of the whole dataset as well as that of each Cramer Class was compared for COSMOS, Munro-1996, and Munro-2016 datasets in Table 7. The QC results from COSMOS and EFSA shifted the Munro dataset towards less potent, although the difference was not statistically significant.

The median and geometric mean of POD values of the whole COSMOS TTC dataset are 42.2 and 43.2 mg/kg-bw/day, respectively. The median and geometric mean of POD values for the whole Munro-2016 dataset (N = 606) are 20.7 and 18.3 mg/kg-bw/day, respectively. Interestingly, the median and geometric mean of the Class III of the COSMOS TTC dataset were higher than those for the

Table 7
logPOD distribution of TTC datasets.

Cramer Class	COSMOS TTC	Munro-2016	Munro-1996
	Stat description	Stat description	Stat description
All	 Median: 42.2 Geometric Mean: 43.2 N: 552	 Median: 20.8 Geometric Mean: 18.4 N: 611	 Median: 18.0 Geometric Mean: 17.2 N: 613
I	 Median: 100 Geometric Mean: 104 N: 219	 Median: 170 Geometric Mean: 142 N: 141	 Median: 156 Geometric Mean: 112 N: 137
II	 Median: 18.3 Geometric Mean: 18.5 N: 40	 Median: 27.4 Geometric Mean: 28.0 N: 30	 Median: 26.3 Geometric Mean: 24.4 N: 28
III	 Median: 20.7 Geometric Mean: 25.1 N: 293	 Median: 10.0 Geometric Mean: 9.23 N: 440	 Median: 10.0 Geometric Mean: 9.46 N: 448

Munro dataset, i.e. on average COSMOS Class III was less potent than Munro Class III, but the distribution for Class I in the COSMOS TTC dataset was shifted to lower median and geometric mean of POD values.

3.3.2. *Cumulative distribution functions*

The empirical cumulative distribution function (CDF) curves for each Cramer Class are presented in Fig. 11, where the abscissa represents the log(POD) values and the ordinate gives the cumulative fraction, F(x).

In the Munro-1996 dataset, although Class I and II overlap at lower POD values, the separation of each Cramer Class is clearer than that in the COSMOS TTC dataset, where the distributions of the Cramer Class II and III are very similar. For the COSMOS TTC dataset, 10% of the data (fraction of 0.1) is below 5.44, 1.67, and 1.67 mg/kg-bw/day for Cramer Class I, II, and III, respectively. In the Munro

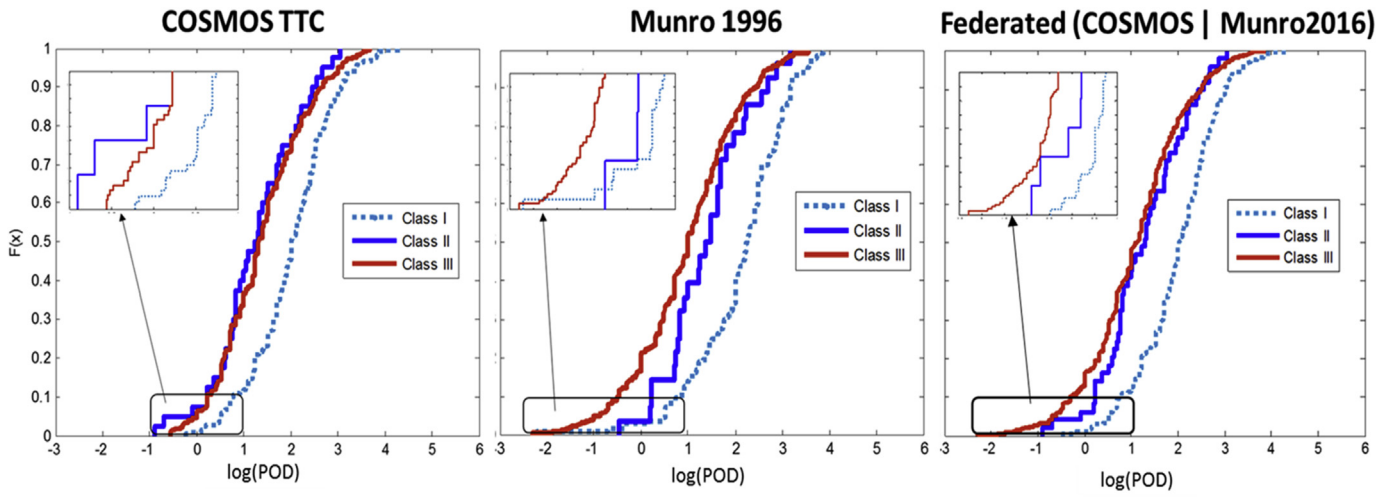


Fig. 11. Comparison of the empirical cumulative distribution function in the COSMOS TTC, Munro et al., 1996 and Federated datasets.

Table 8
Effect of adjusting for DART study duration on POD distribution.[a,b]

Cramer Class	COSMOS 5th percentile POD values		Munro-2016 5th percentile POD values		Munro-1996 5th percentile POD values	
	without adjustment	with adjustment	without adjustment	with adjustment	without adjustment	with adjustment
Cramer I	3.42 (N = 219)	3.40 (N = 217)	3.78 (N = 141)	3.78 (N = 141)	2.91 (N = 137)	2.93 (N = 137)
Cramer II	0.41 (N = 40)	0.41 (N = 40)	0.91 (N = 30)	0.96 (N = 30)	0.85 (N = 28)	0.91 (N = 28)
Cramer III	0.93 (N = 293)	0.86 (N = 291)	0.13 (N = 435)	0.12 (N = 435)	0.14 (N = 448)	0.13 (N = 448)

[a] Non-parametric estimation method was used as described in Section 2.4.
[b] Chemicals associated with DART studies whose duration is not clear were not included in this analysis.

dataset, these 10% quantiles (fraction of 0.1) are 8.24, 1.67, and 0.33 mg/kg-bw/day for Cramer Class I, II, III, respectively.

The regions up to 10% quantiles are depicted in more detail in Fig. 11. The lower end of the CDF near 5th and 10th percentiles illustrates that the Class III of the COSMOS TTC dataset may be less potent than that of Munro. The plots also indicate that Class II and III of the COSMOS TTC dataset do not separate well and that there may be a few more potent chemicals for Class II than in Class III in that region, even though Cramer Class II is generally intended to capture chemicals of lower potency compared to Cramer Class III. For example, the two Class II chemicals, allyl heptanoate and canthaxanthin, are the two most potent chemicals of the COSMOS TTC dataset. In the COSMOS TTC dataset, the 10% quantile for Cramer Class I is also lower than that of Munro-2016. To test whether the distributions of each of the Cramer Classes are significantly different, the pair-wise K-S test was performed. For the COSMOS TTC dataset, the differences between Class I & III (p-value = 0.0001, N = 512) as well as between Class I & II were significant (p-value = 0.0001, N = 259); however, the difference between Class II & III distributions was not significant. Even without any chemical insights, this simple statistical test further suggests that there is not a solid basis to distinguish between Cramer Class II and Class III in the COSMOS TTC dataset. Similar observations were found in the Munro-1996 dataset regardless of the Cramer classification methods used (COSMOS experts, Munro, Toxtree). However, in their 1996 publication Munro et al. stated that the distribution difference between Cramer II and III was statistically significant. This statement could only be reproduced if the raw NOEL values of the original dataset were compared (p-value = 0.0309, N = 476) without applying the study duration factors. Since the approach chosen for this project was to analyse distributions and 5th percentiles after all the adjustments are made, the use of raw NOEL values for the significance test was not appropriate.

3.3.3. Effect of adjustment for DART study duration

In DART studies, the reliability of recording treatment duration can be challenging and hence making the duration adjustment of non-DART POD values from DART studies difficult. To this end, the effect of duration adjustment for DART studies on the POD distributions and 5th percentile values were evaluated for both COSMOS and Munro TTC datasets. Since the Cramer Classifications affect the distribution significantly, the criteria used to establish the COSMOS TTC dataset were applied also to the Munro dataset for this comparison.

The 5th percentile POD values in Table 8 indicate that that duration adjustments of reproductive studies did not result in appreciable changes of the 5th percentile values using both non-parametric (shown) and parametric (not shown) methods. In addition, the pair-wise comparisons of the observed (non-parametric) distributions using the pair-wise K-S test also confirmed that there is no significant impact on the POD distributions by the duration adjustment of reproductive studies in both COSMOS and the two Munro TTC datasets (Munro-1996 and Munro-2016). Therefore, in the final COSMOS TTC approach, no duration adjustment factors were applied to reproductive studies.

3.3.4. Effect of study QC on 5th percentile

In establishing the databases that contribute to the final estimates of TTC values for substances used in cosmetics, a considerable effort was made during the curation of the chemical and toxicological information contained in the databases to ensure quality. The transparent and rigorous processes used for study selection and QC of the toxicity data have been described. Particular scrutiny was undertaken on the most potent sections of the COSMOS TTC dataset in order to establish robust 5th percentile POD values. Statistical testing of the hypothesis on whether the QC process shifted the datasets toward less potency has been conducted before and after QC1 (dataset version v1.4 and v1.5) applying the pair-wise K-S test. None of the increases in the 5th percentile POD values of each Cramer Class before and after the QC was significantly different.

3.4. Federated TTC dataset of COSMOS and Munro

As described in section 3.1.2, the chemical space of both Munro and COSMOS TTC datasets can be improved when augmented by each other. In addition, it would be beneficial to provide a TTC approach based on one master database rather than separated by the substance use types (e.g., cosmetics, pesticides, antimicrobials, etc.).

In this study, since systematic and thorough QC efforts had been undertaken by the COSMOS partners, the overlap between COSMOS TTC and Munro datasets was simply replaced by COSMOS content in the case of Munro-2016. The final count was 963 substances with 243 for Cramer Class I, 49 for II, and 671 for III. This was an increase of over 75% for Class I and II, and a 50% increase for Class III. To test whether the distributions of each of the Cramer Classes are significantly different, the pair-wise K-S test was performed. For the federated set, as shown in the CDF (Fig. 11), the differences between Class I & III (p-value <0.001, N = 914) as well as between Class I & II were significant (p-value<0.001, N = 292), whilst the difference between Class II & III distributions was not significant. In contrast to the COSMOS TTC dataset, in the federated dataset, the lowest quartiles of the PODs of Cramer Classes II and III do not overlap quite as much, retaining the empirical CDF shape of the Munro dataset.

3.5. TTC analysis

TTC values were derived from the 5th percentiles based on POD results in the COSMOS TTC dataset as demonstrated in Fig. 11 and the previous sections. From this point on, whenever comparisons were made between the COSMOS TTC and Munro datasets, the same adjustment factors were used for both datasets so that meaningful comparisons could be made. In addition, a revised

Table 9
Summary of 5th percentiles POD values for COSMOS TTC/Munro datasets.

	Datasets	5th percentile POD values (mg/kg-bw/day)		
		Cramer I	Cramer II	Cramer III
non-parametric	COSMOS[a]	3.42 (N = 219)	0.41 (N = 40)	0.93 (N = 293)
	Munro-2016[a]	3.78 (N = 141)	0.91 (N = 30)	0.13 (N = 435)
	Munro (published value)[b]	3.30 (N = 137)	1.60 (N = 28)	0.12 (N = 448)
	Federated set	3.54 (N = 243)	0.74 (N = 49)	0.22 (N = 671)
parametric	COSMOS[a]	4.20 (N = 219)	0.58 (N = 40)	0.79 (N = 293)
	Munro-2016[a]	4.90 (N = 141)	1.07 (N = 30)	0.15 (N = 435)
	Munro (published value)[b]	3.0 (N = 137)	0.91 (N = 28)	0.15 (N = 448)
	Federated set	4.57 (N = 243)	0.62 (N = 49)	0.23 (N = 671)

[a] Adjustment factors and Cramer Classifications were applied according to the COSMOS TTC criteria. Analysis methods are described in Section 2.4.1.
[b] Listed are the original Munro published values for the 5th percentile (Munro et al., 1996). Issues related to reproducing these values for Munro-1996 dataset are discussed in Section 2.4.1.

dataset denoted as Munro-2016 was used for analysis after correcting some Cramer classes and other errors of the Munro-1996.

3.5.1. Fifth percentile comparisons

The 5th percentile POD values for each Cramer Class in the COSMOS TTC and Munro datasets are summarized in Table 9. As expected, the 5th percentile of Cramer Class I of the COSMOS TTC dataset was higher than that of Class II or III. For Cramer Class III, the COSMOS TTC dataset gave a higher 5th percentile value than that of Munro. However, the 5th percentile value for Cramer Class II in the COSMOS TTC dataset was lower than those for other Cramer Classes. As explained earlier in section 3.3.2, the Cramer Class I/II and I/III are statistically significantly different, but not III/II based on the pair-wise K-S Test. The possible reasons were presented in previous sections.

The selection of Munro et al. (1996) for Cramer Class II chemicals illustrated that the raw NOEL distribution (without any duration adjustments) was significantly different ($p = 0.0309$) from that of Cramer Class III; however, in the COSMOS TTC dataset even the distributions of the raw NOAEL values of Cramer II and III were not significantly different although the sample size is larger. Therefore, reasons other than just the sample size also need to be considered. Both Munro and COSMOS TTC datasets have 28–40 chemicals in Cramer Class II. The low number of chemicals means in practice that the one or two chemicals with the lowest NOAELs can dramatically shift the threshold for Cramer Class II. Indeed, this was the case for the Class II chemicals allyl heptanoate (NOAEL = 0.125 mg/kg-bw/day) and canthaxanthin (NOAEL = 0.2 mg/kg-bw/day), the two chemicals with the lowest NOAELs of the entire COSMOS TTC dataset. In contrast, the numbers of chemicals are such that the Cramer Class I and III distributions are robust and the thresholds do not change if some PODs are updated.

Table 9 also shows that the POD values do not change statistically significantly between COSMOS, Munro-2016, Munro-1996, and the federated dataset of Munro-2016-with-COSMOS; this demonstrates the robustness of the 5th percentile thresholds for Cramer Class I and III. Thus far, the federation of the two existing datasets has shown that the chemical space can be expanded and complementary and that the human exposure threshold values still broadly support the existing TTC values. This analysis provides powerful utilitarian value by pooling the appropriate data for TTC approaches.

3.5.2. Human exposure threshold values

Three aspects can be summarized for chemicals in the COSMOS TTC dataset of cosmetics-related chemicals: (1) Cramer Class I is still less potent than Cramer Class III; (2) Cramer Class II results in a slightly lower 5th percentile value than Cramer Class III within the constraints of the small sample size for Class II; (3) the overall distribution patterns and the ranges in the COSMOS TTC dataset are broadly similar to those of the Munro datasets (including Munro published values, Munro-1996, Munro-2016). Therefore, based on the Munro approach, the human exposure thresholds for cosmetics-related chemicals have been derived from the COSMOS TTC dataset by applying a 100-fold safety factor to the 5th percentile POD values. No TTC value is proposed for Cramer Class II of the COSMOS TTC dataset since the Cramer II and III distributions overlap in this dataset (Fig. 11). This is consistent with proposals of EFSA and WHO (EFSA/WHO, 2016). If Classes II and III are combined, the parametric estimation of the fifth percentile is 0.76 mg/kg-bw/day, which is practically equivalent to the Cramer Class III value of 0.79 as shown in Table 9. The human exposure threshold values (TTC values) for the four different datasets are listed in Table 10, expressed in both μg/person day and μg/kg-bw/day.

TTC values for cosmetics-related chemicals in the COSMOS TTC dataset can be compared with the currently widely-used TTC values proposed by Munro et al., in 1996 for food-related chemicals. Except that the Cramer Class III is significantly less potent in the COSMOS dataset than in the Munro dataset, the two datasets show similar distribution characteristics. When the two sets are federated, the resulting cumulative distributions for the three Cramer Classes were not significantly different from the Munro dataset based on pair-wise K-S tests.

4. Discussion

4.1. Impact on chemical space enrichment

Until now, there has not been a dataset available that enables the application of the TTC approach to cosmetics products to be addressed specifically. The COSMOS TTC dataset, which has been rigorously curated and collated, fills that gap. It contains cosmetics-related chemicals, including some more complex molecules such as hair dyes. Although there is some overlap between the COSMOS TTC and the Munro datasets, in Cramer Class I, 65% of the chemicals in the COSMOS TTC dataset are different from those in the Munro dataset; in the important Cramer Class III group, 81% of the COSMOS TTC chemicals are different from those in the Munro dataset. The COSMOS TTC dataset also has a more even split between Cramer Class I and Class III than the Munro dataset (both Munro-1996 and Munro-2016).

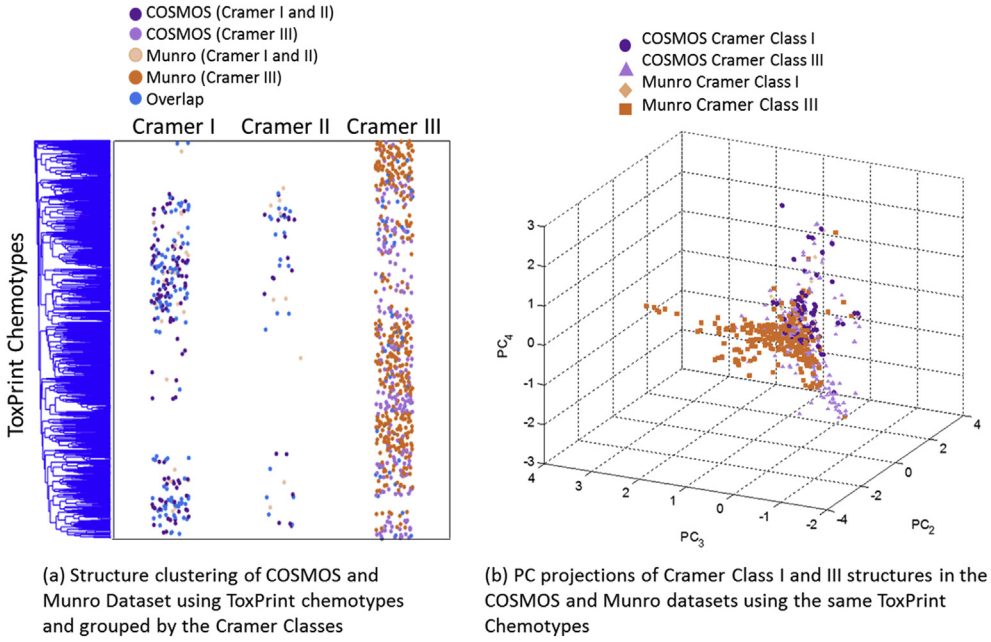
These descriptive differences are also apparent graphically in a clustering map where structures defined by ToxPrint chemotypes are again clustered for Cramer Classes as depicted in Fig. 12.

There are structural clusters showing up in Munro Cramer Class III, but not observed (white space in the vertical bar for each Cramer

Table 10
Comparison of human exposure threshold values.[a]

Datasets (number of chemicals)	Human exposure threshold values (μg/person/day)			Human exposure threshold values (μg/kg-bw/day)		
	Cramer Class I	Cramer Class II	Cramer Class III	Cramer Class I	Cramer Class II	Cramer Class III
COSMOS (552)	2500	NA	470	42	NA	7.9
Munro-1996[b] (613)	1800	540	90	30	9.0	1.5
Munro-2016 (606)	2900	640	90	49	11	1.5
Federated set (963)	2700	370	140	46	6.2	2.3

[a] All threshold values were calculated by parametric estimation of the cumulative distribution. Per person values were calculated based on a default body weight of 60 kg.
[b] These values are verbatim copy of the Munro et al., 1996 publication.



(a) Structure clustering of COSMOS and Munro Dataset using ToxPrint chemotypes and grouped by the Cramer Classes

(b) PC projections of Cramer Class I and III structures in the COSMOS and Munro datasets using the same ToxPrint Chemotypes

Fig. 12. Visualization of chemical space differentiation of Cramer Classes in COSMOS and Munro-2016 TTC datasets.

Class) in the COSMOS TTC dataset. It is also the case that some structural clusters appear only in COSMOS Cramer Class I (as shown in Fig. 11); the same is true for Munro Class III. Similar observations are also illustrated when scores of the principal components are plotted. The Cramer Class III structures in Munro-2016 (salmon) and COSMOS (lavender) share very few commonalities in the chemical space. It is also the case that the structures of the Cramer Class I and III from both datasets are separated. Therefore, characterisation of the chemical space occupied by the COSMOS TTC dataset has demonstrated that it is different from that of the Munro-2016 dataset, for both Cramer Class I and III. It is also significant from Fig. 12 that the chemical space of Class III is now further expanded beyond that of the Munro-2016 dataset when the cosmetics chemotypes are added.

The new database has also been profiled for study types/species, critical effects and target organs; through this analysis, it has been demonstrated that the studies included are diverse and broadly cover the critical effects that are important in systemic toxicity safety evaluations. In line with the Munro et al. (1996) and Kroes et al. (2004) work on TTC, short term and dermal studies were not included in the database, so that this project did not consider local effects or hypersensitivity effects unless they drove the NOAELs in oral repeated dose studies. Taken together, the above features provide confidence in the applicability of the COSMOS TTC dataset to cosmetics-related chemicals and hence the use of the dataset to evaluate the appropriateness of the TTC values for such chemicals.

The POD values in the new COSMOS TTC dataset span six orders of magnitude, which is similar to that in the Munro dataset. When expressed on a body weight basis, the parametric TTC value derived from the 5th percentile POD value for Cramer Class III in the COSMOS TTC dataset is 7.8 μg/kg-bw per day, which is 5-fold higher than the corresponding TTC value of 1.5 μg/kg-bw per day that was derived by Munro et al. (1996). The TTC values obtained for Cramer Class III of Munro-1996 and Munro-2016 datasets were the same as that reported by Munro et al. (1996). On the other hand, the TTC value for Cramer Class I was slightly increased from 30 μg/kg-bw per day in the Munro-1996 dataset to 49 μg/kg-bw per day in the Munro-2016 dataset in which the QC results from both COSMOS and EFSA, Cramer Class reviews, and COSMOS rules for study duration were applied (described in Section 2.2). This was a numerically higher value than the Cramer Class I value for the COSMOS TTC dataset of 42 μg/kg-bw per day. It was not unexpected that the new database is overall less potent than the Munro dataset since the content was enriched with cosmetics-related chemicals and it would be expected that such chemicals, for use in personal care products, would have generally lower toxicity compared with that of the broader universe of chemicals.

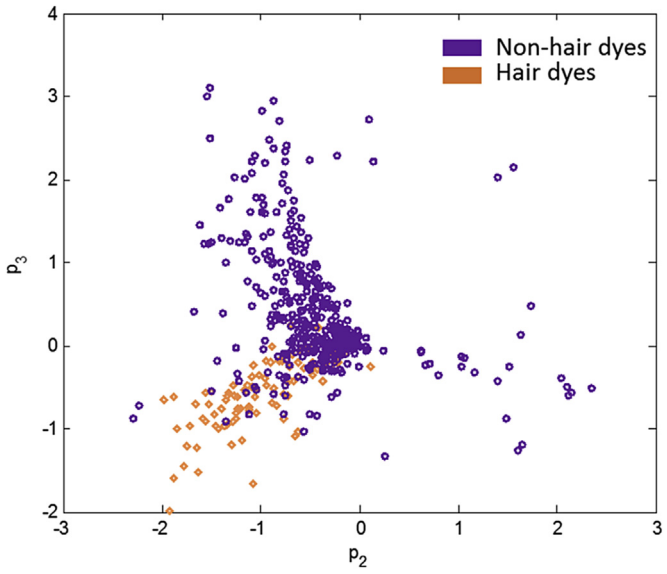


Fig. 13. Principal component projection of hair dyes based on Chemotypes.

4.2. Other factors affecting the 5th percentile POD and TTC values

4.2.1. Effects of certain chemical classes and substances

As discussed in section 3.1.2.1, the chemotypes were applied to the two TTC datasets to distinguish the chemical space (illustrated in Figs. 6, 7, 9 and 12). Using this approach, identification of OPs, carbamates or hair dyeing agents can be easily achieved and the resulting sets were analysed for their influence on the 5th percentile POD values. In the COSMOS TTC dataset, only one OP (POD = 1.67 mg/kg-bw/day) and 3 carbamates (POD = 2.67, 5.0, 100.3 mg/kg-bw/day) were found. Removing these compounds did not have any impact on the 5th percentile POD for Class III of the COSMOS TTC dataset.

Also considered was the addition of 122 hair dyeing agents to reflect the needs of the European cosmetics regulation. As most of these chemicals were in Cramer Class III (109) and only 13 in Class I, the impact of this addition to the dataset needed to be evaluated in both chemical space and POD distributions. These hair dyes are in a structural island, almost exclusively by themselves, as shown in the PCA projections (Fig. 13), occupying chemical classes of ethanolamine, phenolic amines, and aromatic nitro. When using ToxPrint chemotypes that appear in more than 4 structures, the two sets are clearly separated almost exclusive (orthogonal) to each other with the first few principal components.

The impact of hair dyeing agents on the 5th percentile POD values is summarized in Table 11. Also examined were the substances prohibited for use in cosmetics by EU regulation (Section 2.2.4.1).

The exclusion of these hair dyeing agents or the EU-prohibited list did not affect the 5th percentile POD value or the distributions significantly, judging from the pair-wise K-S test. This analysis confirms that hair dyeing agents and substances prohibited for use in cosmetics can be included in the COSMOS TTC dataset to enrich the chemical space without impacting the POD distribution and 5th percentile.

4.2.2. The effect of Cramer classifications

Other than the importance of study quality that drives the POD decisions, one of the most important factors impacting the TTC thresholds is the Cramer classifications. Although Cramer Class II chemicals in the COSMOS TTC dataset have a slightly lower 5th percentile POD value than Cramer Class II chemicals in the Munro dataset, the significance of this difference is unclear due to the small sampling size for Cramer Class II chemicals in both databases (the COSMOS TTC dataset has 40 chemicals in Class II, and Munro has 28). For example, one of the chemical classes below the 5th percentile POD value of the Class II group in the COSMOS TTC database is allyl carboxylic esters, which is known to be quite reactive and toxic. Therefore, in addition to the fact that the POD distributions of Class II and III are not significantly different, the possibility still exists that the chemical space of Class II for cosmetics-related chemicals may be different than that of other chemicals. The difficulty of finding sufficient chemicals to populate Cramer Class II and provide a meaningful analysis has been noted by others using different databases (Batke et al., 2011; EFSA, 2012; Escher et al., 2010; Feigenbaum et al., 2015; Munro et al., 1996; Pinalli et al., 2011; Tluczkiewicz et al., 2011). The present study also found that there were insufficient chemicals in Cramer Class II for a meaningful analysis and derivation of a reliable TTC value for this Cramer Class with the database at hand. This does not automatically imply that other datasets with focus on specific chemical classes cannot lead to more robust Cramer Class II distributions. Hence, for the present project, Cramer Class II chemicals were still analysed separately to enable comparison with previous analyses and also to make the data more easily accessible for potential future research.

Table 11
Effect of hair dyes on 5th percentile POD values.

	Non-parametric estimation of 5th percentile POD value of COSMOS TTC dataset		
	All	hair dyes	prohibited substances
Cramer Class I	3.42 (N = 219)	3.50 (N = 206)	3.54 (N = 210)
Cramer Class II	0.41 (N = 40)	0.41 (N = 40)	0.38 (N = 39)
Cramer Class III	0.93 (N = 293)	0.77 (N = 184)	1.04 (N = 274)

Another factor to consider is the assignment of Cramer Classes, which can have an impact on the TTC thresholds derived. There has been considerable discussion of the Cramer classification system, which was first proposed by Cramer et al. (1978). It has been deemed still fit for purpose by European advisory bodies (EFSA, 2012; SCCS, SCHER and SCENIHR, 2012). However, the fact that it was developed on the basis of toxicological knowledge of several decades ago and because users of the Toxtree software for assignment of Cramer Classes have raised questions about inconsistencies and problems with some of the steps in the Cramer decision tree, the need to revise some aspects of the decision tree has been discussed and certain changes have been proposed (Bhatia et al., 2015; Dewhurst and Renwick, 2013; EFSA/WHO, 2016; Lapenna and Worth, 2011; Patlewicz et al., 2008; Roberts et al., 2015). Some of the proposed changes have been implemented in the Toxtree software. In the development of the COSMOS TTC dataset, a number of substances had to be manually reassigned for Cramer Class. At this present point, we recommend that users of the TTC approach be aware of potential problems and have the opportunity to consult suitable experts that can manually check the assignment of Cramer classifications.

4.2.3. Assessing dermal exposures with oral TTC thresholds

The purpose of this project was to improve the scientific basis of and confidence in applying the TTC concept to assess exposures arising from cosmetics-related chemicals. It is acknowledged that most exposures from cosmetic products will occur via the dermal

route. However, for most substances, many more repeated-dose oral studies have been carried out than repeated-dose dermal studies, even for substances that are dermally applied to humans. Thus, the database of available dermal repeated dose studies is too small to derive meaningful distributions and thresholds. In addition, for many substances, systemic exposure from oral administration is known to be higher than from dermal application (Williams et al., 2016). Therefore, risk assessment based on TTC thresholds will have to make use of route-to-route extrapolation methods, as further elaborated by the other COSMOS ILSI Europe Expert Group (Williams et al., 2016). This situation is in no way different from dermal risk assessment when only oral toxicity studies are available on the chemical in question. Derivation of PBPK modelling-based internal TTC values has been proposed (Partosch et al., 2015) and is being further evaluated, but is also challenging due to the amount of data needed to develop relevant estimates of internal doses arising during oral toxicity studies on one hand, and from dermal exposure on the other hand. The COSMOS project has provided a robust and relevant oral database and TTC thresholds valid for chemicals related to cosmetics, which can be applied in risk assessment in the same way as substance-specific toxicity data from oral, repeated-dose studies.

4.2.4. The new COSMOS TTC dataset as part of the TTC concept

This project developed an enhanced oral non-cancer TTC dataset with larger chemical domain, which is intended to be used as a part of the broader TTC concept as developed by Kroes et al. (2004), including refinements thereof. It is not recommended to use this TTC dataset, or the thresholds derived, in isolation. Hence, the exclusion criteria for chemicals and effects not included in the database (e.g. proteins and sensitisation) and effects of specific concern (e.g. potent carcinogens) also apply to the TTC concept when using this dataset.

5. Conclusion

This study presents a new, transparent and public TTC database of 552 cosmetics-related chemicals. The COSMOS TTC dataset is publicly downloadable at the COSMOS DB v2.0 website (Molecular Networks, 2017). The 5th percentile POD value for each Cramer Class was determined, from which human exposure threshold values (TTC values) have been derived. The number of substances classified in Cramer Class II was insufficient for derivation of a robust TTC value. TTC values of 42 and 7.9 μg/kg-bw/day for Cramer Class I and Cramer Class III, respectively, have been derived for the COSMOS TTC dataset and TTC values of 46, 6.2 and 2.3 μg/kg-bw/day for Cramer Class I, Cramer Class II and Cramer Class III, respectively, of the COSMOS-plus-Munro federated dataset. This study also showed, through federation of datasets, that the TTC approach first proposed by the pioneering work of Munro et al. (1996) still holds in a broad sense even after updating multiple details and when many more cosmetics-related chemicals are added. The small impact of a substantially enlarged database and QC effort demonstrates the power and robustness of the probabilistic approach of TTC.

Conflict of interest

All authors have completed and submitted the journal's 'Form for Disclosure of Potential Conflicts of Interest'.

Funding/support

This work was conducted by an expert group of the European branch of the International Life Sciences Institute, ILSI Europe, within their role in the COSMOS project. This publication was co-ordinated by the Threshold of Toxicological Concern Task Force. Industry members of this task force are listed on the ILSI Europe website at http://ilsi.eu/task-forces/food-safety/threshold-of-toxicological-concern/. Experts are not paid for the time spent on this work; however, the non-industry members within the expert group were offered support for travel and accommodation costs from the COSMOS project to attend meetings to discuss the manuscript. COSMOS was funded by the European Community's Seventh Framework Programme (FP7/2007–2013), under grant agreement number 266835, and the European trade association for the cosmetic, toiletry and perfumery industry, Cosmetics Europe. A small compensatory sum (honorarium) from the Threshold of Toxicological Concern Task Force was offered to the non-industry expert group members, with the option to decline. The expert group carried out the work, i.e. collecting/analysing data/information and writing the scientific paper separate to other activities of the task force. The research reported is the result of a scientific evaluation in line with ILSI Europe's framework to provide a pre-competitive setting for public-private partnership (PPP). ILSI Europe facilitated scientific meetings and coordinated the administrative tasks relating to the completion of this work. For further information about ILSI Europe, please email info@ilsieurope.be or call +3227710014. The opinions expressed herein and the conclusions of this publication are those of the authors and do not necessarily represent the views of ILSI Europe nor those of its member companies or any regulatory authority.

Acknowledgements

Authors wish to acknowledge Bob Safford, Joan Fisher, Karen Blackburn, and Echo Rufer for their participation in this project. The study QC group also express our gratitude to project managers at ILSI Europe during these years including Stéphane Vidry, Tanja Wildeman, Massimo Ambrosio, Alessandro Chiodini, Jackie Whyte, and Cyril Marsaux. CY thanks Ivan Boyer and George Daston who kindly offered many valuable toxicological insights in personal communications. CY also recognizes Aleksey Tarkhov and Aleksandra Mostrag at MN-AM who helped the final database QC for importing the TTC datasets to COSMOS DB v2.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.fct.2017.08.043.

Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.fct.2017.08.043.

Appendix 1. COSMOS MINIS Study Criteria

Study Parameter	Chronic	Subchronic	Reproductive	Developmental	Required?
TEST SUBSTANCE: NAME	text	text	text	text	REQUIRED
TEST SUBSTANCE: NAME SOURCE	controlled vocabulary list	controlled vocabulary list	controlled vocabulary list	controlled vocabulary list	REQUIRED
STUDY BACKGROUND: GUIDELINE	controlled vocabulary list including non-guideline study	controlled vocabulary list including non-guideline study	controlled vocabulary list including non-guideline study	controlled vocabulary list including non-guideline study	REQUIRED
STUDY BACKGROUND: DATA SOURCE GRANULARITY	controlled vocabulary list including summary-only	controlled vocabulary list including summary-only	controlled vocabulary list including summary-only	controlled vocabulary list including summary-only	REQUIRED only for the matched conditions
STUDY BACKGROUND: REFERENCE TYPE	controlled vocabulary list	controlled vocabulary list	controlled vocabulary list	controlled vocabulary list	REQUIRED
STUDY: STUDY TYPE	controlled vocabulary list	controlled vocabulary list	controlled vocabulary list	controlled vocabulary list	REQUIRED
STUDY: DURATION	6 months - 2 years	90 day; >=28 day allowed	Continuous exposure through appropriate generation weaning	Implantation through organogenesis	REQUIRED
TEST SYSTEM: SPECIES	rat/mouse/dog/monkey	rat/mouse/dog/monkey	rat/mouse/dog/ monkey/rabbit	rat/mouse/dog/ monkey/rabbit	REQUIRED
TEST SYSTEM: ROUTE EXPOSURE	diet, drinking water, or gavage	diet, drinking water, or gavage	diet, drinking water, or gavage	diet, drinking water, or gavage	REQUIRED
TEST SYTEM: ANIMAL AGE	broad description - healthy, young adults	broad description - healthy, young adults	broad description - young …		REQUIRED only for the matched conditions
TEST SYTEM: ANIMAL TREATMENT	actual weight	actual age or weight	Male: treated for 10-wk prior to mating; Female: treated for 2 weeks	young, mature, primigravida, untreated	REQUIRED only for the matched conditions
TEST DESIGN: ANIMAL NUMBER	rodent - 10/dose/sex; dog - 3/dose/se	rodent - 10/dose/sex; dog - 3/dose/se	rodent or rabbit - 20/ dose/sex	rodent or rabbit - 20/ dose/sex	REQUIRED only for the matched conditions
TEST DESIGN: CONTROL SUBSTANCE	must be specific	must be specific	must be specific	must be specific	REQUIRED only for the matched conditions
TEST DESIGN: DOSAGE REGIMEN	5 days/week (gavage); 7 day/week; ad lib	5 days/week (gavage); 7 day/week; ad lib	REQUIRED only for the matched conditions		
TEST DESIGN: NUMBER DOSE GROUPS	2	2	3	3	REQUIRED only for the matched conditions
TEST DESIGN: DOSE UNIT	Dose unit	Dose unit	Dose unit	Dose unit	REQUIRED
TEST DESIGN: DOSE VALUE	Dose value	Dose value	Dose value	Dose value	REQUIRED
TEST DESIGN: DOSE TO ANIMAL	Dog/Monkey - dose needs to be in mg/kg-bw/day; Others - PPM or % in diet or in water	Dog/Monkey - dose needs to be in mg/kg-bw/day; Others - PPM or % in diet or in water	Dog/Monkey - dose needs to be in mg/kg-bw/day; Others - PPM or % in diet or in water	Dog/Monkey - dose needs to be in mg/kg-bw/day; Others - PPM or % in diet or in water	REQUIRED only for the matched conditions
RESULTS: BODY WEIGHT	start, final; OR start, mid-way, final	start, final; OR start, mid-way, final	P1 at start + weekly; F1 at birth & d-4,21-weekly thereafter; F2 at birth & d-4,21 (♂ & ♀)	dams: at start, end of dosing and at sacrifice	REQUIRED only for the matched conditions
RESULTS: FODD/ WATER CONSUMPTION	need to be mentioned - normal or any signs	need to be mentioned - normal or any signs	need to be mentioned - normal or any signs	need to be mentioned - normal or any signs	REQUIRED only for the matched conditions
RESULTS: GROSS NECROPSY	all usual organs	all usual organs	PREGNANT FEMALE PARAMETERS - Corpora lutea; fetal deaths; live fetuses, MORPHOLOGY - Visceral, Skeletal, External gross examination	REQUIRED	
RESULTS: HISTOPATHOLOGY	gonads, heart, intestine, kidney, liver, spleen, stomach (high dose minimum); other relevant organs	gonads, heart, intestine, kidney, liver, spleen, stomach (high dose minimum); other relevant organs	REQUIRED only for the matched conditions	REQUIRED only for the matched conditions	
RESULTS: ORGAN WEIGHT	kidney, liver, testes; other relevant organs	kidney, liver, testes; other relevant organs	REQUIRED only for the matched conditions		
RESULTS: CLINICAL SIGNS	daily observation; toxic signs; behavior; mortality	daily observation; toxic signs; behavior; mortality	NOT REQUIRED NOT REQUIRED		

(continued on next page)

(*continued*)

Study Parameter	Chronic	Subchronic	Reproductive	Developmental	Required?
RESULTS: HEMATOLOGY	erythrocytes; leukocytes; other relevant assays	erythrocytes; leukocytes; other relevant assays			
RESULTS: CLINICAL CHEMISTRY	relevant assays	relevant assays	NOT REQUIRED		
RESULTS: URINALYSIS	relevant assays	relevant assays	NOT REQUIRED		

The controlled vocabulary list of the database is available from COSMOS DB v2 (COSMOS Parallel Publication).

Appendix 2. Cramer Class QC

Name	CAS	Cramer class used in this study	Munro assignment	Toxtree assignment
1,3-BUTYLENE GLYCOL	107-88-0	1	1	2
2-BUTANOL	78-92-2	2	1	2
ALPHA-TOCOPHEROL	59-02-9	1	1	2
C.I. FOOD BLACK 1	2519-30-4	1	1	3
ETHYLENE GLYCOL MONOPHENYL ETHER	122-99-6	2	1	2
GAMMA-NONALACTONE	104-61-0	1	1	2
GAMMA-UNDECALACTONE	104-67-6	1	1	2
HEXYLRESORCINOL	136-77-6	2	1	2
INOSINIC ACID	131-99-7	3	1	3
LITHOCHOLIC ACID	434-13-9	1	1	3
METHYLENEBIS, 2,2'-	22656-77-5	1	1	3
SODIUM ERYTHORBATE	6381-77-7	3	1	3
SUCROSE MONOPALMITATE	26446-38-8	1	1	3
SUCROSE MONOSTEARATE	25168-73-4	1	1	3
CAROTENE, BETA-	7235-40-7	2	2	1
CAFFEINE	58-08-2	2	2	3
DIKETOPIPERAZINE	29990-68-9	3	2	3
FURFURAL	98-01-1	3	2	3
ISOBORNYL ACETATE	125-12-2	2	2	1
METHYL ANTHRANILATE	134-20-3	2	2	3
PIPERONAL	120-57-0	2	2	3
PROPARGYL ALCOHOL	107-19-7	2	2	3
PYRIDINE	110-86-1	3	2	3
THUJONE	546-80-5	2	2	3
ALLYL ISOVALERATE	2835-39-4	2	3	2
ANTHRANILIC ACID	118-92-3	1	3	1
C.I. ACID RED 14	3567-69-9	1	3	1
C.I. ACID RED 18	2611-82-7	1	3	1
CANTHAXANTHIN	514-78-3	2	3	2
FD&C RED NO. 2	915-67-3	1	3	3
FD&C YELLOW NO. 6	2783-94-0	1	3	3
METHYL CARBAMATE	598-55-0	3	3	1
SODIUM CYCLAMATE	139-05-9	1	3	1

References

Antignac, E., Nohynek, G.J., Re, T., Clouzeau, J., Toutain, H., 2011. Safety of botanical ingredients in personal care products/cosmetics. Food Chem. Toxicol. 49 (2), 324–341.

Bailey, J., 2011. In: Compilation of Ingredients Used in Cosmetics in the United States. The Personal Care Products Council, Washington DC (First). Washington, DC: The Personal Care Products Council).

Barlow, S., 2005. Threshold of Toxicological Concern (TTC): a Tool for Assessing Substances of Unknown Toxicity Present at Low Levels in the Diet null, null(-).

Bassan, A., Fioravanzo, E., Pavan, M., Stocchero, M., 2011. Applicability of physico-chemical data, QSARs and read-across in Threshold of Toxicological Concern assessment. EFSA J. 8 (6), 159.

Batke, M., Escher, S., Hoffmann-Doerr, S., Melber, C., Messinger, H., Mangelsdorf, I., 2011. Evaluation of time extrapolation factors based on the database RepDose. Toxicol. Lett. 205 (2), 122–129.

Becker, R.A., Hays, S.M., Robison, S., Aylward, L.L., 2012. Development of screening tools for the interpretation of chemical biomonitoring data. J. Toxicol. 2012, 1–10. https://doi.org/10.1155/2012/941082.

Benz, R.D., Irausquin, H., 1991. Priority-based assessment of food additives database of the US food and drug administration center for food safety and applied nutrition. Environ. Health Perspect. 96, 85.

Bercu, J.P., Dolan, D.G., 2013. Application of the threshold of toxicological concern concept when applied to pharmaceutical manufacturing operations intended for short-term clinical trials. Regul. Toxicol. Pharmacol. 65 (1), 162–167.

Bhatia, S., Schultz, T., Roberts, D., Shen, J., Kromidas, L., Api, A.M., 2015. Comparison of cramer classification between toxtree, the OECD QSAR toolbox and expert judgment. Regul. Toxicol. Pharmacol. 71 (1), 52–62.

Blackburn, K., Stickney, J.A., Carlson-Lynch, H.L., McGinnis, P.M., Chappell, L., Felter, S.P., 2005. Application of the threshold of toxicological concern approach to ingredients in personal and household care products. Regul. Toxicol. Pharmacol. 43 (3), 249–259.

Brüschweiler, B., 2010. Das TTC-Konzept: beurteilungsmethode von Kontaminanten unbekannter Toxizität im Trinkwasser (The TTC concept: method of assessment of contaminants of unknown toxicity in drinking water). Gas-Wasser-Abwasser 90 (4), 295–303.

Buchholzer, M.-L., Werner, C., Knoess, W., 2014. Current concepts on integrative safety assessment of active substances of botanical, mineral or chemical origin

in homeopathic medicinal products within the European regulatory framework. Regul. Toxicol. Pharmacol. 68 (2), 193–200.
Carthew, P., Clapp, C., Gutsell, S., 2009. Exposure based waiving: the application of the toxicological threshold of concern (ttC) to inhalation exposure for aerosol ingredients in consumer products. Food Chem. Toxicol. 47 (6), 1287–1295.
Chemotyper, 2017. https://chemotyper.org/, Accessed 17 July 2017.
Conover, W., 1999 (Third). In: Conover, W.J. (Ed.), Practical Nonparametric Statistics. Wiley, Ann Arbor.
Cramer, G.M., Ford, R.A., Hall, R.L., 1978. Estimation of toxic hazard–a decision tree approach. Food Cosmet. Toxicol. 16 (3), 255–276.
Cronin, M., Richarz, A.-N., Neagu, D., Yang, C., Pavan, M., Zaldívar-Comenges, J.-M., Meinl, T., 2012. Integrated in silico models for the prediction of human repeated dose toxicity of COSMetics to optimise safety. In: Towards the Replacement of Invivo Repeated Dose Systemic Toxicity Testing, pp. 140–173.
Dewhurst, I., Renwick, A.G., 2013. "Workshop report" evaluation of the threshold of toxicological concern (TTC): challenges and approaches. Regul. Toxicol. Pharmacol. 65 (1), 168–177.
European Chemicals Agency (ECHA), 2012. Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.8: Characterisation of Dose [concentration]-response for Human Health, ECHA-2010-g-19-en. European Chemicals Agency, Helsinki.
European Chemicals Agency (ECHA), 2017. Registered Substance.
European Food Safety Authority (EFSA), 2012. Scientific Opinion on Exploring options for providing advice about possible human health risks based on the concept of Threshold of Toxicological Concern (TTC). EFSA J. 10 (7), 2750. https://doi.org/10.2903/j.efsa.2012.2750.
European Food Safety Authority (EFSA), 2016. Guidance on the establishment of the residue definition for dietary risk assessment. EFSA J. 14 (12), 1–129. https://doi.org/10.2903/j.efsa.2016.4549.
European Food Safety Authority (EFSA), & World Health Organization (WHO), 2016. Review of the threshold of toxicological concern (TTC) approach and the development of a new TTC decision tree. European food safety authority and World health organization. EFSA J. 13 (3), 1–50. https://doi.org/10.2903/sp.efsa.2016.EN-1006.
European Commission, 2009. Regulation (EC) No 1223/2009 of the european parliament and of the Council of 30 november 2009 on cosmetic products. Off. J. Eur. Union 59–209.
European Commission, 2012. CosIng.
European Commission, 2017. Scientific Committee on Consumer Safety (SCCS) Opinions.
European Medicines Agency (EMEA), 2006. Guideline on the Limits of Genotoxic Impurities. Committee for Medicinal Products for Human Use (CPMP), London.
European Medicines Agency (EMEA), 2008. Guideline on the Assessment of Genotoxicity of Herbal Substances/Preparations. Adoption by HMPC for Release for Consultation (vol. 21); London.
Escher, S.E., Tluczkiewicz, I., Batke, M., Bitsch, A., Melber, C., Kroese, E.D., et al., 2010. Evaluation of inhalation TTC values with the database RepDose. Regul. Toxicol. Pharmacol. 58 (2), 259–274.
European Food Safety Authority (EFSA), 2017. Scientific Outputs.
Feigenbaum, A., Pinalli, R., Giannetto, M., Barlow, S., 2015. Reliability of the TTC approach: learning from inclusion of pesticide active substances in the supporting database. Food Chem. Toxicol. 75, 24–38.
Gocht, T., Schwarz, M., 2014. Main achievements and challenges in the third year. In: Towards the Replacement of in Vivo Repeated Dose Systemic Toxicity, pp. 189–193.
Health Canada, 2016. Science Approach Document: Threshold of Toxicological Concern (TTC)-based Approach for Certain Substances (Ottawa).
Hennes, E.C., 2012. An overview of values for the threshold of toxicological concern. Toxicol. Lett. 211 (3), 296–303.
Houeto, P., Carton, A., Guerbet, M., Mauclaire, A.-C., Gatignol, C., Lechat, P., Masset, D., 2012. Assessment of the health risks related to the presence of drug residues in water for human consumption: application to carbamazepine. Regul. Toxicol. Pharmacol. 62 (1), 41–48.
JMP, 2017. https://www.jmp.com/en_us/software.html (accessed 7.17.17).
Klimisch, H.J., Andreae, M., Tillmann, U., 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul. Toxicol. Pharmacol. RTP 25 (1), 1–5. https://doi.org/10.1006/rtph.1996.1076.
Kroes, R., Renwick, A.G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., et al., 2004. Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food Chem. Toxicol. 42 (1), 65–83.
Kroes, R., Renwick, A.G., Feron, V., Galli, C.L., Gibney, M., Greim, H., et al., 2007. Application of the threshold of toxicological concern (TTC) to the safety evaluation of cosmetic ingredients. Food Chem. Toxicol. 45 (12), 2533–2562.
Laabs, V., Leake, C., Botham, P., Melching-Kollmuß, S., 2015. Regulation of non-relevant metabolites of plant protection products in drinking and groundwater in the EU: current status and way forward. Regul. Toxicol. Pharmacol. 73 (1), 276–286.
Lapenna, S., Worth, A., 2011. Analysis of the Cramer classification scheme for oral systemic toxicity-implications for its implementation in Toxtree. JRC Sci. Tech. Rep. EUR 24898.
MathWorks, 2017. https://nl.mathworks.com/product/ltc/matlab.html?s_tid=hp_products_matlab, Accessed 17 July 2017.
Melching-Kollmuß, S., Dekant, W., Kalberlah, F., 2010. Application of the "threshold of toxicological concern" to derive tolerable concentrations of "non-relevant metabolites" formed from plant protection products in ground and drinking water. Regul. Toxicol. Pharmacol. 56 (2), 126–134.
Molecular Networks, 2017. COSMOS [WWW Document]. URL https://cosmosdb.eu/cosmosdb.v2/accounts/login/?next=/cosmosdb.v2/(accessed 7.17.17).
Mons, M.N., Heringa, M.B., Van Genderen, J., Puijker, L.M., Brand, W., Van Leeuwen, C.J., et al., 2013. Use of the Threshold of Toxicological Concern (TTC) approach for deriving target values for drinking water contaminants. Water Res. 47 (4), 1666–1678.
Munro, I.C., Ford, R.A., Kennepohl, E., Sprenger, J.G., 1996. Correlation of structural class with no-observed-effect levels: a proposal for establishing a threshold of concern. Food Chem. Toxicol. An Int. J. Publ. Br. Ind. Biol. Res. Assoc. 34 (9), 829–867.
Partosch, F., Mielke, H., Stahlmann, R., Kleuser, B., Barlow, S., Gundert-Remy, U., 2015. Internal threshold of toxicological concern values: enabling route-to-route extrapolation. Arch. Toxicol. 89 (6), 941–948.
Patlewicz, G., Jeliazkova, N., Safford, R.J., Worth, A.P., Aleksiev, B., 2008. An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. SAR QSAR Environ. Res. 19 (5–6), 495–524.
Pinalli, R., Croera, C., Theobald, A., Feigenbaum, A., 2011. Threshold of toxicological concern approach for the risk assessment of substances used for the manufacture of plastic food contact materials. Trends Food Sci. Technol. 22 (9), 523–534.
Richard, A.M., Judson, R.S., Houck, K.A., Grulke, C.M., Volarath, P., Thillainadarajah, I., et al, 2016. ToxCast chemical landscape: paving the road to 21st century toxicology. Chem. Res. Toxicol. 29 (8), 1225–1251.
Roberts, D.W., Aptula, A., Schultz, T.W., Shen, J., Api, A.M., Bhatia, S., Kromidas, L., 2015. A practical guidance for Cramer class determination. Regul. Toxicol. Pharmacol. 73 (3), 971–984.
SCCS, SCHER and SCENIHR, 2012. Opinion on the Use of the Threshold of Toxicological Concern (TTC) Approach for Human Safety Assessment of Chemical Substances with Focus on Cosmetics and Consumer Products. Brussels.
SCCS NfG, 2016. The SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation. SCCS/1564/15.
Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52 (3–4), 591–611.
Silverman, B.W., 1998. Density Estimation for Statistics and Data Analysis. Density Estimation for Statistics and Data Analysis. Taylor & Francis, London.
Stanard, B., Dolan, D.G., Hanneman, W., Legare, M., Bercu, J.P., 2015. Threshold of toxicological concern (TTC) for developmental and reproductive toxicity of anticancer compounds. Regul. Toxicol. Pharmacol. 72 (3), 602–609.
The R Project for Statistical Computing, 2017. https://www.r-project.org/, Accessed 17 July 2017.
Tluczkiewicz, I., Buist, H.E., Martin, M.T., Mangelsdorf, I., Escher, S.E., 2011. Improvement of the Cramer classification for oral exposure using the database TTC RepDose–a strategy description. Regul. Toxicol. Pharmacol. 61 (3), 340–350.
Toxprint, 2017. https://toxprint.org/, Accessed 17 July 2017.
Toxtree, 2017. http://toxtree.sourceforge.net/, Accessed 17 July 2017.
US Department of Health and Human Services, 2017. National Toxicology Program.
US Environmental Protection Agency (EPA), 1991. Alpha 2u-globulin: Association with Chemically Induced Renal Toxicity and Neoplasia in the Male Rat, vol. 20460. Protection Agency, Risk Assessment Forum, Washington, DC. Washington, DC.).
US Environmental Protection Agency (EPA), 2003. OCSPP Harmonized Test Guidelines. Series 870 – Health Effects Guidelines. Last updated March 2003. Washington, DC.
US Environmental Protection Agency (EPA), 2016. Pesticide Reregistration Status.
US Food and Drug Administration (FDA), 1993. Food additives: threshold of regulation for substances used in food-contact articles; proposed rule. Fed. Regist. 58, 52719–52729.
US Food and Drug Administration (FDA), 1995. Food additives: threshold of regulation for substances used in food-contact articles; final rule. Fed. Regist. 60, 36582–36596.
US Food and Drug Administration (FDA), 2000. FDA Redbook. Guidance for Industry and Other Stakeholders: Toxicological Principles for the Safety Assessment of Food Ingredients. Redbook 2000, College Park.
Williams, F.M., Rothe, H., Barrett, G., Chiodini, A., Whyte, J., Cronin, M.T.D., et al., 2016. Assessing the safety of cosmetic chemicals: consideration of a flux decision tree to predict dermally delivered systemic dose for comparison with oral TTC (Threshold of Toxicological Concern). Regul. Toxicol. Pharmacol. 76, 174–186.
Worth, A., Cronin, M., Enoch, S., Fioravanzo, E., Fuart-Gatnik, M., Pavan, M., Yang, C., 2012. Applicability of the Threshold of Toxicological Concern (TTC) Approach to Cosmetics—Preliminary Analysis. JRC Report EUR 25162EN. Publications Office of the European Union, Luxembourg, pp. 1–32.
Yang, C., Hasselgren, C.H., Boyer, S., Arvidson, K., Aveston, S., Dierkes, P., et al., 2008. Understanding genetic toxicity through data mining: the process of building knowledge by integrating multiple genetic toxicity databases. Toxicol. Mech. Methods 18 (2–3), 277–295.
Yang, C., Tarkhov, A., Marusczyk, J., Bienfait, B., Gasteiger, J., Kleinoeder, T., et al., 2015. New publicly available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling. J. Chem. Inf. Model. 55 (3), 510–528. https://doi.org/10.1021/ci500667v.