Statistical Methods & Applications
https://doi.org/10.1007/s10260-025-00794-0

REVIEW PAPER

From animal testing to *in Silico* models: a systematic review and practical guide to cosmetic assessment

Tamara G. Vasiljev[1] · Lucia Salvioni[2] · Miriam Colombo[2] · Paolo Galli[3] · Francesca Greselin[4]

Accepted: 13 May 2025

Abstract

Assessing the safety and efficacy of skincare products has become increasingly important, with the rise of alternative methods to animal testing, due to ethical and regulatory demands. We reviewed the integration of classical statistical techniques with modern *in Silico* approaches, providing a structured guide for researchers. Following PRISMA 2020 guidelines, we conducted a systematic PubMed search for studies applying statistical or computational methodologies to assess cosmetic product safety and efficacy, and published between 2013 and 2023. Papers lacking methodological rigor or clear application to cosmetics were excluded. Two independent reviewers screened studies to minimize bias, and only peer-reviewed articles were included. Tables and figures were prepared to synthesize the main results. A total of 195 studies met the inclusion criteria. Our findings highlight the increasing role of *in Silico* approaches and machine learning techniques in cosmetic safety evaluation, alongside traditional statistical methods such as regression analysis, hypothesis testing, and multivariate techniques. The review provides practical guidance on selecting methodologies based on data availability and research objectives, along with a critical analysis of their strengths and limitations. The findings are shaped by search keywords, which may have excluded some related studies. *In Silico* methods emerge as promising alternatives to animal testing, though their reliability depends on robust validation and high-quality datasets. Standardization and regulatory integration are crucial for broader adoption in cosmetic science.

Keywords *In Silico* methods · Cosmetics · QSAR · Read-across · Structural alerts · Rule-based methods

Extended author information available on the last page of the article

Published online: 16 June 2025

1 Introduction

In recent years, the skincare industry has experienced a surge in the development and marketing of innovative products to improve skin health and appearance. We have witnessed a shift towards more sustainable, plant-based, nature-inspired, and animal-free testing cosmetics. Cosmetic companies are embracing innovation and exploring nature's wealth of resources to create eco-friendly products that promote skin health and environmental well-being. As the demand for sustainable cosmetics continues to grow, further advancements and exciting developments in the field are emerging, aiming to strike a balance between beauty, science, and environmental consciousness.

The second significant consideration that motivates our analysis, is the full implementation of the animal testing bans (Regulation No 1223/2009) in European countries, forbidding *in Vivo* testing of finished products and cosmetic ingredients in 2004 and 2009, respectively. Therefore, classical statistical methodologies, serving as indispensable tools for assessing the safety and efficacy of skincare products in a reliable and evidence-based manner, now encounter new challenges arising from the described emerging perspectives.

This review aims to guide cosmetic researchers who conduct experimental studies to develop innovative green and eco-friendly cosmetic formulations. Our focus will be on the crucial aspect of the choice of statistical tools for robust and reproducible skincare studies. Specifically, the discussion will include the selection and application of appropriate statistical tests.

Combined with proven efficacy, cosmetics products need to have a comprehensive toxicological assessment. An increasing number of alternatives to animal testing have been developed and validated for the safety and efficacy assessment of cosmetic products and ingredients. These advances led to the growth of New Approach Methodologies (NAMs) that aim to replace the old tests. NAMs are validated tools that predict a substance's systemic toxicity in the absence of animal testing. Many of them include new *in Vitro* Methods, such as the 3D human skin equivalent models used to evaluate skin irritation potential. NAMs have also provided substantial progress in the *in Silico* methodologies: new computational methods mimic *in Vitro* and *in Vivo* assays, are based on mathematical and statistical models, and incorporate the use of 'big data' provided by previous studies.

In Silico approaches hold great promise in accurately predicting different safety parameters without resorting to traditional *in Vivo* testing. By leveraging sophisticated statistical tools and models, these methods offer a cost-effective and environmentally friendly alternative, allowing researchers to bypass resource-intensive laboratory testing.

The aim is to explore the most important emerging trends together with the traditional statistical testing employed in skincare research. This work emphasises the importance of employing appropriate statistical approaches in the safety and efficacy assessment of new green skincare products, according to the new shift in the market and the current regulations. By combining *in Silico* methods with conventional assays and appropriate statistical testing, we want to provide up-to-date

guidelines to the cosmetics industry for developing effective, eco-friendly products based on natural and sustainable ingredients while adhering to rigorous safety standards.

2 Methods

This systematic review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines for reporting systematic reviews (Page et al. 2021). To ensure methodological rigor and transparency, we implemented a structured framework characterized by a clearly defined search strategy, predefined inclusion and exclusion criteria, and a systematic data extraction process (see Appendix for the detailed PRISMA 2020 item checklist). In line with recommendations from the Cochrane Collaboration, two independent reviewers conducted both the study screening and data extraction processes to minimize bias and enhance reliability. The following sections detail each step of this methodology, outlining how relevant studies were identified, selected, and analyzed.

2.1 Databases and research strings

Studies containing the safety and efficacy assessment of cosmetic products were identified via the scientific database PubMed, which comprises more than 36 million citations for biomedical literature from MEDLINE, life science journals, and online books. Keywords and search terms related to skincare products, statistical approaches, *in Silico* methods, and relevant toxicity parameters, such as skin corrosion and irritation, eye irritation, skin sensitization, mutagenicity, genotoxicity, carcinogenicity, dermal absorption, and phototoxicity, were used to retrieve significant papers. The PubMed Advanced Search Builder was used to write the following clauses:

(i) (cosmetic) AND (phototoxicity) AND (assessment) AND (model);
(ii) (cosmetic) AND (skin irritation) AND (assessment) AND (model);
(iii) (cosmetic) AND (eye irritation) AND (assessment) AND (model);
(iv) (cosmetic) AND (skin corrosion) AND (assessment) AND (model);
(v) (cosmetic) AND (skin sensitization) AND (assessment) AND (model);
(vi) ((dermal absorption[Title/Abstract]) OR (dermal penetration[Title/Abstract])) AND (cosmetic[Title/Abstract]);
(vii) (cosmetic) AND (assessment) AND (model) AND((mutagenicity) OR (genotoxicity));
(viii) (cosmetic) AND (carcinogenicity) AND (assessment) AND (model);
(ix) (cosmetic) AND (assessment) AND (*in Silico*) AND ((acute toxicity) OR (repeated dose toxicity) OR (reproductive toxicity));
(x) (efficacy[Title/Abstract]) AND (cosmetic[Title/Abstract]) AND (*in Silico*).

As the marketing bans on cosmetic products have evolved over time, we chose not to take into account the studies published before 2013, which is the date on which the main animal testing bans came into effect. All articles published between January 2013 and December 2023 were taken into account.

2.2 Selection process

Papers are included in the review based on the following eligibility criteria:

(i) Relevance: The paper should focus on statistical approaches for the safety and efficacy assessment of skincare products, specifically discussing one or more of the recommended safety parameters, including skin corrosion and irritation, eye irritation, skin sensitization, mutagenicity, genotoxicity, carcinogenicity, dermal absorption, and phototoxicity;
(ii) Scientific rigor: only peer-reviewed articles and reputable sources, such as reports from regulatory agencies (e.g., FDA, ECHA, OECD) and recognized scientific organizations, are considered to ensure the quality and reliability of the information;
(iii) Publication date: the specific date range was set to 2013–2023 to encompass the most recent and relevant research up to the present;
(iv) Language of publication: only articles published in the English language were considered for inclusion in this comprehensive review.

Exclusion criteria are applied as follows:

(i) articles that do not mention any statistical or *in Silico* method are excluded;
(ii) studies that do not focus on hazard assessment but rather on exposure assessment are excluded;
(iii) research that focuses on environmental risk assessment rather than specific risk assessment related to human health are excluded.

The titles and abstracts of all identified PubMed records were screened to determine their relevance. The selection process was conducted in three stages, in accordance with Page et al. (2021):

1. **Initial screening** - Titles and abstracts were reviewed to exclude non-relevant studies.
2. **Full-text assessment** - The remaining studies with available full text were assessed for eligibility based on the predefined inclusion and exclusion criteria.
3. **Final selection** - Studies meeting all criteria were included in the review.

Any discrepancies between reviewers were resolved through discussion or consultation with a third reviewer when necessary. To visually represent the search process, we employed a PRISMA-recommended flow diagram (see Fig. 1), constructed using the tool described in Haddaway et al. (2022).

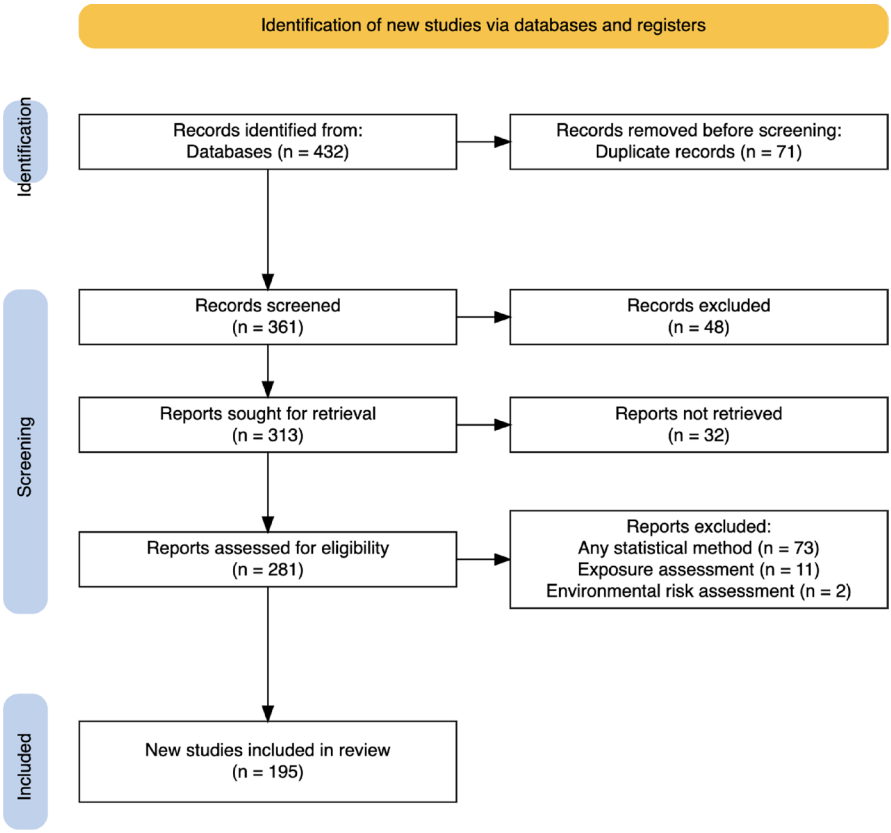

Fig. 1 PRISMA Flow Diagram describing the selection process of the articles admitted in the review

2.3 Extraction of relevant information

A meticulous analysis of the selected articles has been conducted to derive a structured framework of statistical approaches used to evaluate the safety and efficacy of skincare products. The extracted information included whether the study focused on efficacy or safety. For safety-related studies, the specific parameters assessed (e.g., skin corrosion, irritation, sensitization) were recorded. Studies were further categorized as experimental (*in Vivo*/*in Vitro*) or *in Silico*.

For experimental studies, we documented the statistical methodologies employed, considering the type of data and research objectives.

For each *in Silico* study, we extracted details on the software used, its accessibility (free or proprietary), the safety parameters it predicts, and the models it implements (QSAR, read across, rule-based method, or hybrid models).

The obtained information is summarized through explanatory tables and figures, presented in the following section, which provides a structured overview of the methodologies and tools reported across the included studies.

More generally, the classification of extracted information was not always unique. A study could investigate both efficacy and safety, assess multiple safety parameters, or employ a combination of laboratory and computational tests. Additionally, studies often utilized multiple statistical or computational methods, which were all documented to reflect the full scope of their approach.

Regarding missing or unclear data, we only included outcomes that were explicitly specified in the study reports.

3 Results

The systematic search yielded a careful selection of studies investigating the safety and efficacy of skincare products. These studies vary in study design, research objectives, and methodological approaches, reflecting the multidisciplinary nature of cosmetic science. We included experimental studies, observational studies, reviews, reports, and computational studies, the latter introducing and validating new *in Silico* methods, software tools, and predictive models. While traditional experimental methods remain predominant, an increasing number of studies incorporate *in Silico* models, highlighting their growing role in cosmetic product assessment. The following sections provide a detailed breakdown of the literature search process and the key methodological approaches identified in this review.

3.1 Literature search

The results of the selection process for the papers included in the review are described in the flowchart represented in Fig. 1.

The systematic review began by identifying 432 records, of which 71 duplicates were removed. The high number of duplicates is due to the fact that many studies assess multiple safety parameters simultaneously (e.g., skin sensitization together with skin irritation and genotoxicity). After de-duplication, the titles and abstracts of 361 papers were carefully examined, resulting in the exclusion of 48 records that were not relevant to the primary focus of the study. The remaining 313 articles were attempted to be retrieved in full text. However, 32 of them were found to be unretrievable. As a result, a thorough screening of 281 reports was conducted to ensure that the selected studies aligned with the review's predefined objectives. This rigorous process revealed that 86 reports did not address the core themes of the review; in fact, they did not include any statistical or computational method relevant to cosmetic safety assessment, focused solely on exposure assessment, or addressed environmental risk assessment instead of human safety, as detailed in Fig. 1. Consequently, they were excluded from further consideration. In the end, the systematic review included a total of 195 studies.

3.2 Statistical methods for assessing cosmetic products

The main findings highlight the widespread use of different statistical methods in the skincare field. In particular, a wide variety of statistical tests were met,

demonstrating their key role in assessing the efficacy and safety profile of skincare formulations in *in Vivo* and *in Vitro* studies.

Furthermore, an increasing proportion of *in Silico* studies was observed, reflecting a contemporary trend in skincare research. *in Silico* studies explored a spectrum of statistical modelling approaches and probabilistic models, including QSAR models and machine learning techniques.

The following paragraphs will provide a detailed explanation of each of these two main approaches.

3.2.1 Statistical testing

When assessing the safety and effectiveness of skincare products, it is crucial to select the appropriate statistical tests. This directly affects the reliability and interpretability of study findings. To address the multifaceted nature of these evaluations, we have compiled a comprehensive table (Table 1) that lists various statistical approaches tailored for cosmetic product analysis. It is important to note that the choice of a statistical test depends on the research objectives and the characteristics of the data being investigated. Table 1 provides a practical guide to clarify the relationship between research goals and appropriate statistical methods. It also includes references to articles, that passed the inclusion criteria, for each statistical test and goal. This feature aims to offer researchers insight into both theoretical aspects and the real-world implementation of statistical methodologies in the evaluation of skincare products.

From the reviewed studies, we observed that the majority of experimental studies analyzed continuous (Gaussian) outcomes. Most of these studies reported their results using mean and standard deviation, with the one-sample t-test and ANOVA being the most frequently used statistical methods for evaluating efficacy and safety parameters. Further details on these tests, along with practical examples to guide the selection of appropriate statistical methods, are provided in Sect. 4.1.

3.2.2 *in Silico* methods

A thorough literature review has shown that *in Silico* methods play an essential role in the assessment of safety endpoints and efficacy of skincare products. To evaluate efficacy endpoints, our research has identified a specific application of *in Silico* tools, namely molecular docking (Ledwoń et al. 2023; Masjedi and Solhjoo 2022; Hazarika et al. 2022; Ouedraogo et al. 2022). This computational technique is particularly useful for predicting the orientation of one molecule (ligand) when bound to another (target), and it is implemented by several software such as Discovery Studio Visualiser (Dassault Systemes BIOVIA, U.S.A.), AutoDock and Endocrine Disruptome. Molecular docking will be further explored during the discussion. Table 2 presents the results of the literature search and categorises them according to the different endpoints. From the selected articles, it was possible to identify the most commonly used *in Silico* software employed in cosmetic safety research. A first classification of these tools is illustrated by the circular dendrogram shown in Fig. 2, which provides researchers with insights for

Table 1 Statistical testing for the safety and efficacy assessment of cosmetic products

Goal	Type of Outcome		
	Gaussian	Rank or Continuous non-Gaussian	Binomial
Describe one sample	Mean, standard deviation [1]	Median, interquartile range (Šidlovská et al. 2017; Leskur et al. 2019; Dorier et al. 2019; Gomes et al. 2023)	Proportion or percentage [11]
Compare one sample to a hypothetical value	One-sample t test (Nepalia et al. 2021)	Wilcoxon's rank-sum test	Chi square or Binomial test (Demir et al. 2013; Onyango 2023)
Compare two unpaired samples	Unpaired t test [2]	Mann–Whitney test [8]	Fisher's exact test [12]
Compare two paired samples	Paired t test [3]	Wilcoxon's signed rank test (Savic et al. 2021; Cui et al. 2023; Narda et al. 2019)	McNermar's test
Compare three or more unmatched samples	ANOVA (Analysis of Variance) [4]	Kruskall-Wallis test [9]	Chi square test (Efthimiou et al. 2020; Onyango 2023; Bar-Meir et al. 2023)
Compare three or more matched samples	Repeated-measures ANOVA (Leskur et al. 2019; Shyr and Ou-Yang 2016; Lyons et al. 2020)	Friedman test	Cochrane Q
Quantify associations between two variables	Pearson correlation [5]	Spearman correlation (Estrela et al. 2021)	Phi coefficients
Predict value from an other variable	Simple linear (or non linear) regression [6]	Simple non-parametric model (Chen and Chang 2023; Petit et al. 2017; Kaluzhny et al. 2015)	Simple logistic regression (Onyango 2023; Bar-Meir et al. 2023)
Predict value from several variables	Multiple linear (or non linear) regression [7]	Multiple non-parametric model [10]	Multiple logistic regression [13]

[1]Šidlovská et al. (2017); Leskur et al. (2019); Estrela et al. (2021); Lee et al. (2022); da Silva et al. (2019); Heylings et al. (2018); Keck et al. (2022); Duracher et al. (2015); Kraeling et al. (2018); Pažoureková et al. (2013); Gnonsoro et al. (2022); Iliopoulos et al. (2021); Besrour et al. (2022); Ziemlewska et al. (2022); Walters et al. (2016); Wei et al. (2020); Bengalli et al. (2021); Zerbinati et al. (2021); Golembo et al. (2022); Bamford et al. (2021); Zlabiene et al. (2021); Brathwaite et al. (2022); Sipahi et al. (2022); Navabhatra et al. (2022); Chansriniyom et al. (2021); Imai et al. (2022); Johansson et al. (2017); De Silva and Pathiratne (2023); Kizhedath et al. (2019); Gentile (2019); Mottola et al. (2022); Kobets et al. (2018); Zijno et al. (2015); Andreoli et al. (2018); Jalili et al. (2022); Mottola et al. (2021); Wang et al. (2020); Suriyaprabha et al. (2019); Hadrup et al. (2019); Ahmed et al. (2017); Sahu et al. (2016b); Anand et al. (2017); Han et al. (2013); Reisinger et al. (2018); Sahu et al.

Table 1 (continued)

(2016a); Yuki et al. (2013); Vila et al. (2018); Rajnochová Svobodová et al. (2017); Tang et al. (2018); Maciel et al. (2019); Campodoni et al. (2023); Brown et al. (2023); Leelawattanachai et al. (2023); Saewan et al. (2023); Cui et al. (2023); Parveen et al. (2023); Varet et al. (2024); Gomes et al. (2023); Smeriglio et al. (2023); Pop et al. (2016); da Penha et al. (2022); Zhong et al. (2022); Petit et al. (2017); Pfannenbecker et al. (2013); Eberle et al. (2020)

[2]Heylings et al. (2018); Keck et al. (2022); Duracher et al. (2015); Kraeling et al. (2018); Wei et al. (2020); Bengalli et al. (2021); Golembo et al. (2022); Bamford et al. (2021); Gentile (2019); Zijno et al. (2015); Jalili et al. (2022); Wang et al. (2020); Suriyaprabha et al. (2019); Reisinger et al. (2018); Yuki et al. (2013); Vila et al. (2018); Rajnochová Svobodová et al. (2017); Tang et al. (2018); Raak et al. (2017); Park et al. (2018); Ilves et al. (2014); Raja et al. (2021); Tarnowska et al. (2020); Zheng et al. (2023); Bar-Meir et al. (2023); Gomes et al. (2023); Zhong et al. (2022); Kim et al. (2016); Leoni et al. (2023); Chen et al. (2021a); Choudhury et al. (2017)

[3]Shyr and Ou-Yang (2016); Savic et al. (2021); Leelawattanachai et al. (2023); Saewan et al. (2023); Cui et al. (2023)

[4]Nepalia et al. (2021); Savic et al. (2021); Estrela et al. (2021); da Silva et al. (2019); Keck et al. (2022); Kraeling et al. (2018); Pažoureková et al. (2013); Iliopoulos et al. (2021); Besrour et al. (2022); Zerbinati et al. (2021); Zlabiene et al. (2021); Brathwaite et al. (2022); Sipahi et al. (2022); Navabhatra et al. (2022); Chansriniyom et al. (2021); De Silva and Pathiratne (2023); Kizhedath et al. (2019); Gentile (2019); Mottola et al. (2022); Kobets et al. (2018); Zijno et al. (2015); Andreoli et al. (2018); Jalili et al. (2022); Mottola et al. (2021); Wang et al. (2020); Suriyaprabha et al. (2019); Hadrup et al. (2019); Ahmed et al. (2017); Sahu et al. (2016b); Anand et al. (2017); Reisinger et al. (2018); Sahu et al. (2016a); Vila et al. (2018); Tang et al. (2018); Maciel et al. (2019); Raak et al. (2017); Raja et al. (2021); Sahu et al. (2015); Senapati et al. (2015); Germano-Costa et al. (2022); Sommer et al. (2018); Neubert et al. (2018); Alonso et al. (2015); Bora et al. (2017); Silveira et al. (2016); Kao et al. (2020); Turkez et al. (2022); Qin et al. (2020); Garcia et al. (2013); Abe et al. (2017); Onoue et al. (2013); Campodoni et al. (2023); Brown et al. (2023); Zheng et al. (2023); Leelawattanachai et al. (2023); Jordão et al. (2024); Dvořáková et al. (2023); Wang et al. (2023); Parveen et al. (2023); Varet et al. (2024); Gomes et al. (2023); Smeriglio et al. (2023); Pop et al. (2016); da Penha et al. (2022); Zhong et al. (2022); da Silva et al. (2018); Petit et al. (2017); Kim et al. (2016); Pfannenbecker et al. (2013); Leoni et al. (2023)

[5]Estrela et al. (2021); Duracher et al. (2015); Gnonsoro et al. (2022); Walters et al. (2016); Dimitrov et al. (2016); Chapman et al. (2014); Pop et al. (2016); Lyons et al. (2020)

[6]Lee et al. (2022); Słoczyńska et al. (2022); Parveen et al. (2023); Willhite et al. (2014); Kobets et al. (2018)

[7]Savic et al. (2021); da Silva et al. (2019); Andreoli et al. (2018); Hadrup et al. (2019); Chapman et al. (2014); Bar-Meir et al. (2023); Willhite et al. (2014)

[8]Šidlovská et al. (2017); Dorier et al. (2019); Leskur et al. (2019); Savic et al. (2021); Estrela et al. (2021); da Silva et al. (2019); Iliopoulos et al. (2021); Hadrup et al. (2019); Ilves et al. (2014); Diniz et al. (2019); Gomes et al. (2023); Kalčíková et al. (2023); Pop et al. (2016); Leoni et al. (2023)

[9]Leskur et al. (2019); Dorier et al. (2019); Savic et al. (2021); Estrela et al. (2021); da Silva et al. (2019); Keck et al. (2022); Kraeling et al. (2018); Iliopoulos et al. (2021); Ziemlewska et al. (2022); Hadrup et al. (2019); Alonso et al. (2015); Diniz et al. (2019); Ho et al. (2017); Gomes et al. (2023); Kalčíková et al. (2023); Miller et al. (2024); Leoni et al. (2023)

[10]Navabhatra et al. (2022); Johansson et al. (2017); Williams et al. (2016); Liu et al. (2020); Cho et al. (2019); Ambe et al. (2021); Johansson et al. (2014); Chen et al. (2021b); Frombach et al. (2018); Strickland et al. (2022)

[11]Yuki et al. (2013); Sahu et al. (2015); Silveira et al. (2016); Chen et al. (2021b); Kidd et al. (2021); Sahu et al. (2014); Onyango (2023); Heikkinen et al. (2015)

[12]Yuki et al. (2013); Sahu et al. (2015); Silveira et al. (2016); Chapman et al. (2014); Chen et al. (2021b); Kidd et al. (2021); Sahu et al. (2014); Bar-Meir et al. (2023)

[13]Matsumoto et al. (2016); Onyango (2023); Pop et al. (2016); Reeves et al. (2019); Heikkinen et al. (2015)

Table 2 Results of the in Silico literature search, classified by endpoint

Endpoint	Articles
Dermal absorption	Petry et al. (2017); Canavez et al. (2021); Ates et al. (2016); Selvestrel et al. (2021); Juncan et al. (2023); Hatanaka et al. (2015)
Skin sensitization	Canavez et al. (2021); Selvestrel et al. (2022b); Reynolds et al. (2021); Hirota et al. (2018); Silveira et al. (2016); Strickland et al. (2022); Borba et al. (2022); Johansson et al. (2013); Skare et al. (2015); Jeon et al. (2022); Kleinstreuer et al. (2018)
	Wilm et al. (2018); Fung et al. (2020); Li et al. (2019); Tourneix et al. (2019, 2020); Wilm et al. (2019); Canipa et al. (2017); Hirota et al. (2015); Gilmour et al. (2022); Leist et al. (2014); Borba et al. (2020)
	Vandecasteele et al. (2021); Selvestrel et al. (2021); Asturiol et al. (2016); Gautier et al. (2020); Toropov and Toropova (2021); Sarath Kumar et al. (2016); Leontaridou et al. (2016); Toropov et al. (2021); Dvořáková et al. (2023); Im et al. (2023); Juncan et al. (2023)
Skin corrosion	Silveira et al. (2016); Borba et al. (2022)
Skin irritation	Canavez et al. (2021); Selvestrel et al. (2022b); Golembo et al. (2022); Silveira et al. (2016); Borba et al. (2022); Skare et al. (2015); Popiół et al. (2019)
Eye irritation	Canavez et al. (2021); Silveira et al. (2016); Borba et al. (2022); Skare et al. (2015); Silva et al. (2021); Verma and Matthews (2015a, 2015b)
Phototoxicity	Canavez et al. (2021); Xiong et al. (2019)
Mutagenecity	Canavez et al. (2021); Ates et al. (2016); Silveira et al. (2016); Skare et al. (2015); Toropov et al. (2021); Baderna et al. (2020); Popiół et al. (2019); Juncan et al. (2023)
Genotoxicity	Canavez et al. (2021); Ates et al. (2016); Silveira et al. (2016); Skare et al. (2015); Selvestrel et al. (2021); Lebre et al. (2022); Baltazar et al. (2020); Baderna et al. (2020); Dvořáková et al. (2023); Rioux et al. (2023); Yang et al. (2021)
Carcinogenicity	Canavez et al. (2021); Silveira et al. (2016); Leist et al. (2014); Gao et al. (2019); Rioux et al. (2023)
Acute toxicity	Canavez et al. (2021); Borba et al. (2022); Selvestrel et al. (2021); Yang and Ni (2023); Erhirhie et al. (2018); Hisaki et al. (2015); Rioux et al. (2023)
Repeated dose toxicity	Ates et al. (2016); Leist et al. (2014); Selvestrel et al. (2022a); Hisaki et al. (2015); Rioux et al. (2023); Qian et al. (2022); Batke et al. (2016); Api et al. (2019); Paini et al. (2017)
Reproductive toxicity	Skare et al. (2015); Leist et al. (2014); Hisaki et al. (2015); Ouedraogo et al. (2022); Popiół et al. (2019)
Efficacy	Hazarika et al. (2022); Ledwoń et al. (2023); Masjedi and Solhjoo (2022)

selecting the appropriate *in Silico* tool to detect a specific endpoint. Secondly, the software are grouped in Table 3 according to the predictive model they implement. This complementary classification framework is consistent with the categorisation provided in Wilm et al. (2018) and divides the computational approaches into five categories: linear QSAR, non-linear QSAR, rule-based methods, read-across approaches, and hybrid models. The latter organisation facilitates comparison and selection based on specific research needs and objectives. In the discussion section, we analyze each approach individually, referencing the relevant software packages listed in Table 3 as examples.

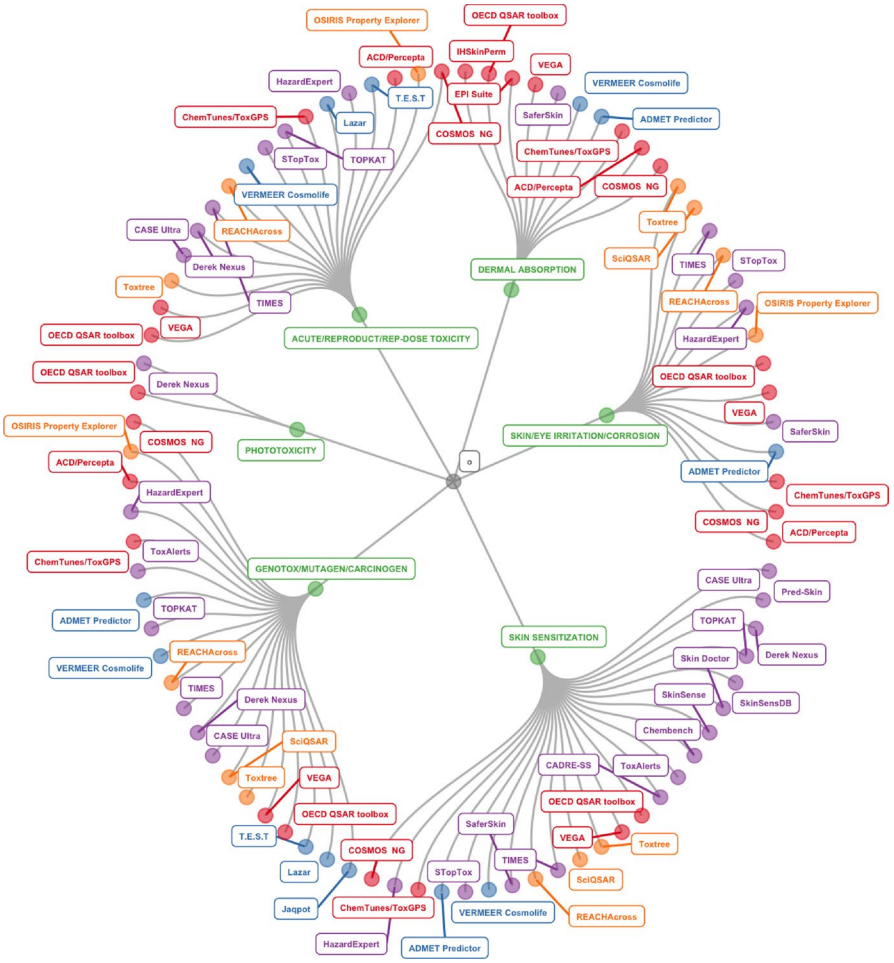

Fig. 2 Circular dendrogram of *in Silico* software

4 Discussion

4.1 Empirical data collection protocols for safety and efficacy

An ever-growing awareness of the importance of consumer protection has led to stricter regulatory requirements on cosmetics. The need to provide more and more evidence of product safety and effectiveness encourages the progress and evolution of the cosmetic testing field. Before discussing the methodologies and basics of experimental data collection, a brief insight into safety and efficacy in cosmetics is essential. Although the contents of this paragraph predominantly refer to the legal framework of the EU, most principles can be applied globally.

Table 3 In Silico models for the predictions of safety endpoints

In Silico method	Definition	Main Computational Approaches	Existing Softwares
Linear QSAR	A subset of QSAR models that assume a direct and proportional relationship between the molecular descriptors of chemical compounds and their toxicity.	Linear regression, multiple linear regression, Partial Least Squares (PLS) regression, Principal Component Regression (PCR).	SKINPERM DERMWIN (EPI Suite) KowWin (EPI Suite) OECD QSAR Toolbox SciQSAR VERMEER Cosmolife TopKat T.E.S.T ACD/Percepta
Non-linear QSAR	A subset of QSAR models that allow for more complex relationships between molecular descriptors and chemical toxicity.	Decision Tree (DT), Random Forest (RF), Support Vector Machines (SVM), Neural Networks (NN).	OECD QSAR Toolbox VEGA Pred-Skin SciQSAR VERMEER Cosmolife STopTox TopKat Skin Doctor ADMET Predictor SkinSense ACD/Percepta
Rule-based methods	Models that rely on expert-derived criteria to classify compounds based on their structural features.	Knowledge-based model, Structural Alerts (SAs) identification, DT, Bayesian networks.	OECD QSAR Toolbox VEGA Toxtree SaferSkin ToxAlerts (OChem) HazardExpert OSIRIS Property Explorer
Read-across	Method to predict the toxicity of a chemical compound based on data from similar compounds with known properties.	Tanimoto coefficient, cosine similarity, euclidean distance, k-nearest neighbors (KNN).	OECD QSAR Toolbox Chembench COSMOS NG

Table 3 (continued)

In Silico method	Definition	Main Computational Approaches	Existing Softwares
Hybrid models	Computational models that integrate multiple modelling techniques or data sources.	Read-across + SAs, Read-across + QSAR, SAs + QSAR.	OECD QSAR Toolbox VEGA CASE Ultra Derek Nexus TIMES REACHAcross CADRE-SS SkinSensDB Jaqpot ChemTunes/ToxGPS Lazar

All regulations agree that safety is the most important feature to be checked before marketing (Barthe et al. 2021; Applebaum 2021). This is mandatory for minimizing adverse health effects and continues as surveillance until the product is sold. The safety assessment aims to quantify product-related risks (i.e. the probability that a specific hazard occurs) and hence it should consider both its composition (i.e. ingredients, contaminants, and their relative amounts) and the conditions of use Kim et al. (2021).

Properties on which data should be available are acute and repeated dose toxicity, skin corrosion and irritation, eye irritation, sensitizing/photosensitizing potential, mutagenicity, genotoxicity, carcinogenicity, reproductive toxicity, dermal/percutaneous absorption, phototoxicity (SCCS, 2021). These parameters can be evaluated by considering both individual ingredients and the entire formulation. The assessment is performed on data collected by own tests or/and derived from literature. Generally, testing the final product is the best way to prove its safety because complex mixtures may exhibit effects resulting from ingredient interaction. However, single ingredient evaluations are widely accepted because most products derive from a limited number of substances (SCCS, 2021).

In addition to a comprehensive toxicological evaluation, it is of paramount importance that cosmetics have proven efficacy (Nobile 2016). This is generally achieved by providing evidence that supports effect-related claims. Defining claims as any public information declared about the product, the aim of this procedure is to protect the end-users from misleading messages (Applebaum 2021). Product claims that have to be substantiated include performance and sensory attributes; relevant data can be obtained from nonclinical and/or clinical studies. The tests could be carried out either on a single ingredient or on the whole formulation to examine a consumer perception or a condition that can be instrumentally analysed (Cosmetics Europe 2008; Applebaum 2021).

Cosmetic testing based on empirical data collection could be conducted *in Vitro*, *ex Vivo* and *in Vivo*. *In Vitro* experiments rely on simple cellular models (i.e. cell lines or reconstructed tissues) and are generally used for screening and/or to evaluate the impact of product/ingredient on a specific mechanism, whereas *ex Vivo* tests employ isolated living tissues increasing the complexity of procedure and data interpretation. After the ban of animal testing in the EU and the progressive alignment of the other countries/regions (Sreedhar et al. 2020), *in Vivo* studies are mainly conducted on human volunteers. Although being more time-consuming compared with the former approaches, *in Vivo* tests provide more relevant results. In this context, ethical issues should be carefully considered: as a rule, safety assessment should precede efficacy testing, while strong preliminary evidence should be collected before testing safety *in Vivo* (Nobile 2016).

Considering the diversity of the endpoints to be addressed, it is beyond the scope of this review to describe the specific experimental approaches. In this regard, it is worth mentioning that few guidelines are available (e.g., OECD guidelines for *in Vitro* safety assessment or EEMCO guidelines for clinical efficacy testing) and they do not cover all the possible cases (Kim et al. 2021). Every scientific methodology that can furnish reliable and reproducible results could be considered for empirical studies (Cosmetics Europe 2008). This justifies the ongoing research in the field.

Moreover, as a consequence of animal testing bans or restrictions, alternative methods that provide representative toxicological data of new ingredients are urgently demanded (Arnesdotter et al. 2021; Nobile 2016). Besides the technical point of view, data interpretation and, therefore, the selection of appropriate statistical analysis are also critical to achieving consistent conclusions.

4.2 Deciding a statistical test

Once empirical data collection through *in Vitro*, *ex Vivo* and *in Vivo* studies has been completed, researchers need to perform statistical analyses to determine whether the new formulation meets the desired efficacy or safety endpoints. Statistical tests serve as indispensable tools for rigorously assessing the performance and potential risks associated with cosmetic products, providing insights that inform further development and regulatory decisions in the cosmetics industry. However, with a wide range of available tests, it can be challenging for researchers to determine which test is appropriate for a given situation. Several considerations need to be made when selecting a statistical test, including elements such as the study design, the number of groups compared, and the nature of the data. Our study makes a contribution to the field by providing a practical guide in the form of a clear and organised table (Table 1). This resource is valuable as it helps researchers match their scientific questions to the structure of their data.

When selecting a statistical test, the user should first consider the nature of the outcome, that is, the result obtained from the experiment and whether it is *qualitative* or *quantitative*. Qualitative variables, also called *categorical*, represent qualitative characteristics or attributes that can be divided into distinct categories or groups. If the categories have a natural order or ranking, the variable is also called *ordinal* or *ranked*, otherwise it is called *nominal*. Quantitative variables are divided into *continuous* when used for collecting measurements, and *discrete* for counts. Binary variables are particular cases of discrete variables. They are often used to represent yes/no or success/failure scenarios, such as the presence or absence of a characteristic. In contrast, continuous variables are measurements that can take any value within a range of real numbers. Continuous variables can also be divided into two families, Gaussian and non-Gaussian. Gaussian variables follow a Gaussian (or normal) distribution, that is a symmetrical and bell-shaped distribution, with the majority of observations clustered around the mean and fewer observations at the extremes. Many natural phenomena and biological measurements tend to follow a Gaussian distribution. On the other hand, non-Gaussian variables may have skewed or asymmetric distributions, heavy tails, or other departures from normality.

Statistical tests are classified as parametric or non-parametric, depending on whether they assume an underlying parametric model, for instance, the Gaussian case, or not. Therefore, non-parametric tests apply to all kinds of data. In determining whether a variable follows a normal distribution, statistical tests such as the Shapiro-Wilk test can provide valuable insights.

Some examples from the selected papers are provided below to enhance comprehension of variables in the cosmetic field.

An example of a continuous variable may be the half-maximal inhibitory concentration (IC_{50}). This endpoint was considered by Maciel et al. (2019) to test whether a particular cosmetic formulation is phototoxic. Firstly, the Shapiro-Wilk and Levene tests were used to confirm the normality and homogeneity of variance of the continuous variable, afterwards, the ANOVA test (for comparison to negative control cells with solvent) was performed.

However, when the Photo Irritation Factor (PIF) index is taken into account, as categorized by the OECD 432 guideline (i.e a $PIF < 2$ predicts the absence of a phototoxic effect; a $2 \leq PIF \leq 5$ predicts a probable phototoxic effect; and a $PIF > 5$ predicts a phototoxic effect (OECD 2019)), the nature of the variable shifts to an ordinal response. Herein lies the importance of considering the ordinal nature of the variable when choosing the test. Finally, a safety evaluation endpoint, such as skin sensitization, may be considered as an example of a binary variable. In Silveira et al. (2016) the skin sensitization potential is categorized into two possible responses, 'sensitized' or 'non-sensitized', producing a binary outcome.

Statistical tests can also be classified based on the number of groups they are intended to analyze. The most straightforward case consists of comparing one group to a hypothetical value. For example, in Nepalia et al. (2021) the authors wanted to test the mutagenicity of some skincare products. In particular, they were interested in testing whether the mean mutagenicity rate (MR) exceeded a predefined value of 2. As the MR is a continuous value, approximately Gaussian, they chose the one-sample t-test to identify the potentially mutagenic products. On the other hand, one can be interested in comparing a specific endpoint in two or more different groups. This is the case, for example, of Savic et al. (2021), where parametric (t-test and ANOVA) or non-parametric tests (Mann–Whitney test and Kruskall-Wallis) are chosen depending on the normality or non-normality of the data. In the former case, the t-test compares the means between two groups to determine if the observed differences are statistically significant. ANOVA extends this comparison to more than two groups, providing an overall assessment of group differences. If significant differences exist, post-hoc tests, such as Tukey's, employed in Savic et al. (2021), can identify specific group differences.

When dealing with ordinal data or when assumptions of normality are not met, non-parametric versions of the t-test and ANOVA are provided by the Mann–Whitney and Kruskal-Wallis tests, respectively. Moreover, in cases where the Kruskal-Wallis test produces a significant result indicating differences between groups, post-hoc pairwise comparisons can be made using Dunn's test, as shown in Estrela et al. (2021), to identify specific group differences.

Finally, another critical factor to consider when selecting an appropriate statistical test is the study design, in particular the nature of the relationship between the groups being compared - whether they are paired (dependent) or unpaired (independent). In scenarios where groups are paired, such as in a longitudinal study or when measurements are taken on the same individuals under different conditions, paired tests become applicable. For example, the paired t-test is valuable when comparing hydration levels before and after sunscreen treatments on the same individual, as in Shyr and Ou-Yang (2016). This allows a nuanced examination of within-subject changes.

Conversely, unpaired tests are appropriate when dealing with independent groups. For example, the unpaired t-test was used in Golembo et al. (2022) to evaluate the effectiveness of an anti-acne product, by comparing a treated arm with an untreated one.

In skincare studies, quantifying the association between two variables is another critical task that researchers often face. This task is essential for understanding the relationships and dependencies between different factors involved in cosmetic formulations and their effects on the skin. An example is given in Gnonsoro et al. (2022), where the authors wanted to investigate the relationship between heavy metal concentrations and specific physicochemical properties (pH and density) of some hydroalcoholic gels. In this case, the variables were Gaussian, so the Pearson correlation coefficient was used. The Spearman correlation coefficient must instead be employed when dealing with ordinal variables, as in Estrela et al. (2021). On the other hand, when two binary variables are considered, the phi coefficient is the appropriate measure. The common aim of these different indices is to describe the association between two variables. They take values between −1.0 and 1.0, with 0.0 representing complete independence and −1.0 or 1.0 representing a perfect association. In the previous instance of Gnonsoro et al. (2022), the correlation coefficient between cadmium concentration and pH was determined to be 0.19, indicating a lack of significant association between these two variables.

In many cosmetic research studies, there is a need to compare observed values and to predict values using known information. The predictive aspect emerged in da Silva et al. (2019), where a linear model was used to predict the Sun Protection Factor (SPF) based on the percentage of some active ingredients in a sunscreen formulation. The results of this regression allowed the authors to identify the best active ingredient to provide a high level of UV protection. When regression models are applied and compared, it is common practice to perform sensitivity analysis and assess accuracy, as shown by Liu et al. (2020). These steps are used to evaluate the models' predictive power and to determine their reliability in capturing the underlying relationships between variables. On the other hand, non-parametric models can be used, for example, to classify a skin irritant, as illustrated in Liu et al. (2020), where the prediction is based on the viability measured after exposure to certain chemicals over time.

4.3 Adopting an *in Silico* method

In skincare research, *in Silico* methods are of great importance as they provide a cost-effective and time-efficient approach to assess the safety and efficacy of skincare products. Computational models and algorithms are employed by *in Silico* tools to predict the potential effects of ingredients on various endpoints. They are used in the field of efficacy, for example, to predict the optimal match between two molecules to achieve a more stable and effective formulation. This predictive capability enables researchers and cosmetic companies to screen many compounds and formulations *in Silico* before conducting expensive and time-consuming *in Vitro* or *in Vivo* studies. Of course, the effectiveness and applicability of *in Silico* tools depend

on the quality and quantity of available data for model development (Luechtefeld et al. 2018). Recent results (Asturiol et al. 2016) suggest that *in Silico* tools have the potential to outperform traditional *in Vitro* and *in Chemico* methods. High-quality, accurate, and comprehensive datasets enable these tools to produce reliable predictions, while larger datasets enhance statistical power and generalizability. Insufficient or biased data can result in unreliable outcomes, underscoring the importance of maintaining rigorous standards for data quality and promoting diversity in datasets.

A critical perspective is essential when selecting which *in Silico* method to use, as each comes with distinct advantages and limitations, depending on the research context. A thorough evaluation of these aspects can guide the choice of the most appropriate method.

4.3.1 Computational approaches for predicting toxicity

Various computational methods have been used to develop *in Silico* tools for safety assessment. These methods range from simple similarity measures, which compare a molecule to known toxic or non-toxic compounds, to more advanced techniques such as multivariate QSAR models. We will focus on several key approaches: QSAR, rule-based methods, read-across, and hybrid models. We will introduce these methods and explore their specific implementations to assist scholars and practitioners.

4.3.1.1 QSAR models

QSAR models are computational tools that predict how a compound's chemical structure influences its toxicological effects. They have been developed using data on compounds with known toxicities, either from public databases or through direct experimental investigation. The first step of data pre-processing (data curation) ensures accuracy and relevance before calculating molecular descriptors. Descriptors, such as molecular weight, size, shape, polarity, and electronic properties, provide quantitative representations of a compound's structural features. They are then analysed in relation to the collected toxicity data to uncover a statistical relationship and construct the model. Finally, rigorous statistical measures are applied to assess the model's predictive power and reliability. Through this comprehensive workflow, QSAR plays a pivotal role in expediting the safety assessment process by offering a nuanced understanding of how new compounds with similar structural features may exhibit toxicological responses, thereby streamlining the safety assessment process. QSAR methods can be divided into two types: linear models, which assume a linear relationship between descriptors and response values, and non-linear methods, which do not make this assumption.

These models have several advantages that enhance their usefulness and effectiveness: they are capable of predicting a sound value for a specific toxicity endpoint, categorizing compounds into toxicity classes, or calculating the likelihood of belonging to a particular class (Gleeson et al. 2012).

However, QSAR models may not always be applicable. One limitation is the requirement for a substantial number of chemicals at the learning stage of the model to achieve statistical significance (Deeb and Goodarzi 2012). Additionally, they

perform feature selection to identify the most significant and independent molecular descriptors. Otherwise, using a large number of descriptors would lead to a complex and fragmented multidimensional space, posing challenges in model interpretation and implementation (Weaver and Gleeson 2008). In some cases, QSAR models may lack biological interpretability and do not take into account critical factors such as dose, duration, or metabolites, which are essential considerations in toxicity assessment (Raies and Bajic 2016). It is crucial to address these limitations by improving the quality of the data and enhancing model development techniques to enhance the reliability and applicability of QSAR approaches in toxicology research.

The Applicability Domain (AD) is a cornerstone in QSAR modeling. It defines the range of chemical, structural, or biological characteristics covered by the data used to build the model. This definition sets the boundaries within which the model can make reliable predictions for new compounds, thereby ensuring the validity of predictions for chemicals similar to those in the original dataset. In Weaver and Gleeson (2008) the authors demonstrated how monitoring the AD is a valuable tool for tracking QSAR performance, estimating prediction accuracy, and identifying when models need recalibration. To further validate these models, it is essential to incorporate experimental data from new compound sets that differ structurally from those used in the original training data (Kar and Roy 2010). Without this, models may show high accuracy in the training set, but their predictive performance often decreases when tested on new chemicals (Valerio Jr 2009). Recent efforts have focused on developing automated procedures to determine the AD of the model. Two primary approaches for AD estimation have been proposed: one based on the interpolation region of the training set in the descriptor space and the other relying on similarity analysis (Nikolova-Jeliazkova and Jaworska 2005). The latter highlights that a compound's similarity is relative and should align with features pertinent to the modeled endpoint (Nikolova and Jaworska 2003; Bender and Glen 2004). For high-dimensional models, which utilize numerous fragments as descriptors, estimating the AD can be particularly complex. Range-based methods, which define the AD by specifying the minimum and maximum values of descriptors, offer a practical approach when combined with PCA for data pre-processing. Nikolova-Jeliazkova and Jaworska (2005) applied this integrated method to the KOWWIN model, used for predicting bioavailability through n-octanol/water partition coefficients (Kow). However, this approach has limitations whenever the number of dimensions increases.

The validation of QSAR models is not just a step, but a crucial checkpoint before their integration into regulatory and commercial applications (OECD 2004). Model accuracy is typically assessed using validation sets, which consist of chemicals with known target values. Creating these sets is not just a task, but a critical process, as well-designed validation datasets lead to accurate evaluations. An effective validation set should reflect the intended application of the model, which requires careful selection of chemicals within the AD of the model (Luechtefeld et al. 2018). The size and diversity of validation datasets remain significant obstacles. Building robust models often requires large datasets with a wide range of compounds to cover the chemical space adequately. However, the limitations of proprietary data and the high costs associated with data collection can hinder the availability of such data sets

(Luechtefeld et al. 2018). To improve validation methods, future approaches could incorporate balanced sampling techniques that ensure a diverse representation of chemical categories, leading to more accurate measurements of model performance.

From a practical standpoint, QSAR models excel in high-throughput screening scenarios, especially when working with well-characterized compounds and large datasets. Their speed and cost-efficiency make them invaluable in early-stage evaluations (Neves et al. 2018). When precision and mechanistic insights are required, particularly for complex or novel chemical entities, QSAR models may not perform as effectively. In data-poor situations, hybrid approaches or alternative methods like read-across, where experts draw on prior knowledge to predict the behavior of new chemicals, are often more pragmatic and contextually suitable (Shah et al. 2016). Thus, while QSAR offers robust early-stage predictions, alternative methods should be preferred, particularly when biological interpretability or nuanced toxicological predictions are needed.

Linear QSAR A linear QSAR model represents the relationship between the structural features of compounds and their biological activity through a linear equation. The equation usually takes the form:

$$y = \beta_0 + \sum_{i=1}^{M} \beta_i x_i \tag{1}$$

where y represents the predicted toxicity endpoint (measured on a continuous scale), $x_1, \ldots x_M$ are the molecular descriptors, β_i is the regression coefficient related to the i-th molecular descriptor and β_0 is the intercept term. Linear QSAR models assume a direct proportionality between changes in molecular descriptors and biological activity, where the coefficients β_i are estimated from a training set of compounds with known toxicities. Once developed, linear models can predict the activity or toxicity of new compounds with similar structural features.

For example, the Potts and Guy model (Potts and Guy 1992) is a linear QSAR used to predict the percutaneous flux of pharmacological compounds. It estimates the permeability coefficient through the skin by incorporating molecular descriptors such as molecular weight and lipophilicity. It is implemented by numerous *in Silico* software packages, such as IHSkinPerm, VERMEER Cosmolife (formerly SpheraCosmolife Selvestrel et al. 2021) and the Dermwin program of the US EPA's EPISuite software.

Linear QSAR models offer significant advantages, particularly in their simplicity and interpretability. When these models perform well, they provide a set of predicted values and clear insights into structure-activity relationships in the dataset, making them valuable tools for researchers. Naturally, a critical point of view arises when considering that some studies based on linear QSAR often rely solely on regression coefficients to draw broad conclusions (Guha 2008). This tendency can oversimplify the complexities inherent in the data, potentially overlooking nuanced relationships. Therefore, while linear QSAR models can be very effective, it is essential to approach their findings with due caution and avoid overgeneralization.

Non-linear QSAR Non-linear QSAR methods, with their ability to explore non-linear relationships between the structural features of compounds and their

biological responses, offer a powerful tool. Their adaptability, particularly advantageous for large datasets, allows for the identification of complex relationships between chemical structures and biological activities. However, it's important to note that for smaller datasets, these methods may be susceptible to overfitting (Hewitt et al. 2009). This occurs when the model captures noise in the data rather than actual underlying relationships and hence has bad performance on new datasets. Furthermore, concerns have also been raised regarding the need for more interpretability of such models (Worth et al. 2011).

The IRFMN/JRC model, implemented in VEGA, represents one of the non-linear QSAR approaches for the prediction of skin sensitization. The model uses a decision tree algorithm to classify chemical compounds as sensitizers or non-sensitizers, based on their structural features and physicochemical properties. By recursive partitioning of the data, it identifies informative patterns and relationships that effectively distinguish between the two categories. Each node of the algorithm represents a molecular descriptor, and each branch embodies a decision rule based on the value of that descriptor. Another commonly used non-linear model is the random forest (RF), an ensemble method that combines multiple decision trees to improve their accuracy and robustness. STopTox (Borba et al. 2022) is an *in Silico* tool that implements this approach, providing a user-friendly interface for researchers to apply the random forest algorithm to their datasets. Artificial neural networks (ANNs) also play an important role in *in Silico* modelling. They consist of interconnected nodes, or neurons, organised into layers. Through iterative training, they learn to map input data to output predictions and are adept at capturing complex patterns. For example, Hirota et al. (2018) developed an ANN model for the risk assessment of skin sensitisation.

Non-linear QSAR models vary significantly in interpretability. Decision trees, for example, offer a clear and intuitive representation of decision-making processes. They create a tree-like structure that allows users to easily trace the path from input features to the final prediction. This transparency makes it straightforward to understand how specific characteristics of a compound influence its classification. In contrast, other non-linear methods, such as RF and ANNs, offer significantly less interpretability (Luechtefeld et al. 2018). In RF, individual decision paths are obscured, making it difficult to identify which features contribute the most to the final output. Similarly, ANNs have a "black box" nature that complicates the interpretation of the decision-making process. The intricate interactions between nodes and the non-linear transformations applied to the data can hide the model's inner workings, hindering efforts to explain how specific input variables lead to certain predictions. However, it is noteworthy that several recent advancements have been reported to improve the interpretability of these complex models, offering new ways to better understand how predictions are generated (Ribeiro et al. 2016; Rioux et al. 2023).

4.3.1.2 Rule-based approaches

Rule-based methods, unlike statistical models, rely on established rules and expert-derived knowledge to predict safety endpoints. They are based on the understanding that toxicity data and chemical structures can be analyzed to identify recurring patterns and establish rules that correlate specific structural features, known as structural alerts (SAs), with toxicological effects. These rules, typically expressed as

if-then statements, are a product of human expertise and provide a simple and accessible way to interpret and implement toxicity prediction models (Raies and Bajic 2016). A variety of commercial (SaferSkin, HazardExpert) and freely available tools (OECD QSAR Toolbox, VEGA, Toxtree, ToxAlerts, OSIRIS Property Explorer) are accessible for virtual screening of structural features that may lead to toxicity. Tox-tree, a notable example of the latter category, employs a set of structural rules curated by experts from industry, academia, and government to provide a qualitative assessment of the likelihood that a compound will exhibit toxicity across various endpoints (Gleeson et al. 2012) as depicted in Fig. 2.

Rule-based methods are the best choice in cases where toxicological mechanisms are complex or poorly understood. Additionally, they can enhance other modelling techniques by offering context-specific understanding of chemical toxicity. However, despite their advantages, these approaches have several limitations. Firstly, they rely solely on binary features, categorising chemical structures as either present or absent, without considering variations in concentration or structural complexity. Moreover, they are restricted to qualitative endpoints and cannot provide insight into the underlying biological pathways of toxicity (Raies and Bajic 2016). One of the significant drawbacks is their potential incompleteness. The list of predefined SAs and associated rules may not cover all toxicological mechanisms, resulting in a high risk of false negatives (Venkatapathy and Wang 2013; Milan et al. 2011; Roncaglioni et al. 2013). This means that certain toxic chemicals may be misclassified as non-toxic if they do not match any of the existing rules. To enhance the predictive accuracy and reliability of these models, it is crucial to ensure that the list of SAs and rules is comprehensive and refined as more experimental data becomes available. Therefore, this approach is most effective when qualitative screening is sufficient. For broader applications requiring deeper mechanistic insights, methods such as QSAR or read-across may provide a more comprehensive understanding (Cronin and Madden 2010).

To address the limitations of Human-Based Rules (HBRs), which rely on predefined expert knowledge that may be incomplete or difficult to update, Induction-Based Rules (IBRs) offer a data-driven alternative (Valerio Jr 2009; Venkatapathy and Wang 2013). IBRs generate rules computationally from large datasets, identifying patterns between chemical structures and toxicity endpoints that may escape human detection (Valerio Jr 2009; Lepailleur et al. 2013). This also allows for the development of hybrid rule systems that may combine the strengths of both HBRs and IBRs, offering broader coverage and adaptability in toxicity predictions (Venkatapathy and Wang 2013). In addition, deep learning algorithms could assist in identifying new structural alerts by guiding decisions and suggesting patterns, moving toward an AI-assisted approach in toxicity prediction (Goh et al. 2017).

4.3.1.3 Read-across methods

In the field of *in Silico* toxicology, read-across (RAx) is a frequently used approach for predicting endpoint information by utilizing available data on the same endpoint of related substances (Patlewicz et al. 2013; OECD 2014; Schultz et al. 2015). This involves evaluating the toxic potential of molecules based on their similarity. Similarity is typically computed using fingerprints, which represent molecules as bit strings defining their fragments, atom paths, or pharmacophoric points.

Metrics such as Tanimoto or Tversky coefficients and the Euclidean distance are commonly used to assess the similarity between molecules (Leach 2001). However, it is important to note that even highly similar molecules may not exhibit similar activity levels. This is because minor structural differences can lead to significant variations in response to a particular process or activity at a specific receptor (Wassermann and Bajorath 2010). Hence, the suitability of the general similarity concept for predicting toxicity must be assessed on a case-by-case basis, considering both the toxicity endpoint and the molecular series being studied (Gleeson et al. 2012). An example of a read-across tool has been implemented in VEGA (Enoch et al. 2008). The tool is designed to assess the skin sensitization potential of alkenes, a group of chemicals known for undergoing a specific chemical reaction, called *Michael addition*. It relies on a database of similar substances tested for skin sensitization to gauge the strength of this reaction. VEGA compares a property called *electrophilicity* among these chemicals. When evaluating a new chemical, the model estimates its skin sensitization potential by comparing its electrophilicity to those in the database.

RAx, with its several advantages, including its ease of interpretation and implementation, is a method that can be readily adopted in the field of *in Silico* toxicology (Enoch 2010). RAx can model quantitative and qualitative toxicity end-points, accommodating various descriptors (Dimitrov and Mekenyan 2010). This adaptability makes it a powerful tool for toxicity prediction.

For example, Enoch et al. (2008) demonstrated a quantitative approach to RAx, extending its application within a known mechanism of action. One key strength of RAx is that it requires minimal data, with only a few high-similarity compounds needed to extrapolate activity. However, finding the correct type of high-similarity data can be particularly challenging, especially for novel structural classes, which may limit its applicability in some cases (Hemmerich and Ecker 2020). However, there are also limitations. Statistical similarity measures allow for comparisons between chemicals, but they do not provide direct biological insight into the toxicity mechanisms. Moreover, complex similarity measures can make model interpretation difficult, potentially complicating the understanding of predictive outcomes (Dimitrov and Mekenyan 2010). Dimitrov and Mekenyan (2010). Conflicting toxicity profiles or insufficient availability of similar compounds may also compromise the validity of read-across methods (Modi et al. 2012). In such cases, the QSAR approach offers a viable alternative, especially when a large data set is available and a quantitative evaluation is needed (Enoch 2010; Modi et al. 2012; Venkatapathy and Wang 2013; Jeliazkova et al. 2010).

Several studies propose workflows for conducting a proper read-across study. Patlewicz et al. (2013) suggest a stepwise approach that includes identifying potential analogues, collecting and evaluating data, and comparing physicochemical properties and toxicokinetics to ensure reliable predictions. Schultz et al. (2015) emphasize the importance of chemical category membership, bioavailability, and mechanistic plausibility to justify read-across predictions for regulatory purposes. Additionally, hybrid approaches that integrate RAx are common. For instance, Modi et al. (2012) present a method combining data-mining, QSAR, and read-across, where QSAR serves as a final step when no toxicological data on analogues is available.

4.3.1.4 Hybrid models

Hybrid *in Silico* models combine two or more of the above-mentioned components to improve prediction accuracy and the applicability of computational methods (Wilm et al. 2018). They combine elements such as QSAR, rule-based approaches, and read-across. For instance, the OECD QSAR Toolbox combines rule-based profilers with RAx to identify appropriate analogs or construct chemical categories. This is achieved by utilizing experimental data on various endpoints for a wide range of substances. Another hybrid approach for predicting skin sensitization potency, exemplified by CADRE-SS (Kostal and Voutchkova-Kostal 2016), integrates a linear QSAR method with expert-curated rules. This model is a three-class categorical hybrid that describes various stages of the sensitization process. It incorporates predictions from Monte Carlo simulations, rule-based assignment of reaction domains, and chemical reactivity calculations based on physicochemical and quantum mechanical descriptors.

Hybrid methods can also integrate machine learning algorithms. For example, Luechtefeld et al. (2018) developed an approach that predicts the skin sensitization potential of compounds by combining fingerprint similarity analysis with RF. It considers dependencies between 19 different endpoints and uses an RF algorithm for prediction, it can handle missing data (enhancing its applicability in various scenarios) and is accessible through the proprietary platform REACHAcross (https://www.ulreachacross.com). REACHAcross is a web-based, automated read-across QSAR tool that combines several public databases. In March 2017, its underlying dataset included more than 70 million structures, 300,000 of which contained biological data and 20,000 linked to animal data. Luechtefeld et al. (2018) assessed the predictive capacity for acute and topical endpoints using leave-one-out cross-validation, revealing that more than 80% of toxic chemicals were correctly identified (sensitivity= 80), with specificities (ability to detect true non-toxic molecules) ranging from 54% to 71%.

Integrating read-across and QSAR principles significantly enhances the interpretability of model predictions and provides a clear framework for understanding the relationships between structurally similar compounds (Luechtefeld et al. 2016). This synergy allows for more intuitive insight into how chemical structures influence biological activity, making it easier for researchers to identify potential toxicants. However, it is essential to remain vigilant, as this integration can sometimes lead to oversimplified conclusions if the intricate complexities of chemical interactions are not fully explored (Luechtefeld et al. 2018).

While hybrid models offer valuable advantages, they also present challenges due to the complexities of interpreting their results. For instance, correlations between predictions from different components may not necessarily indicate high reliability. Misinterpretation can occur when overlapping training data or shared modeling methods create a false impression of agreement (Rorije et al. 2013; Fitzpatrick et al. 2018). Nonetheless, hybrid models have emerged as promising first-tier screening tools for various endpoints in toxicological assessments, balancing the need for interpretability with the intricacies of chemical interactions (Kostal and Voutchkova-Kostal 2016).

4.3.2 Molecular Docking

Molecular docking, an *in Silico* tool widely used in cosmetic research, plays a crucial role in evaluating the efficacy of skincare products, particularly cosmeceuticals - formulations that bridge the gap between cosmetics and pharmaceuticals by containing active ingredients with proven therapeutic effects. This methodology predicts the optimal binding conformation between one molecule (the ligand) and another (the target). In cosmetic formulations, the ligand represents the active ingredient, while the target represents the skin protein with which it interacts. Docking involves a sophisticated combination of computational techniques, including sampling algorithms, such as genetic algorithms, which generate different orientations of the ligand-target complex. Such orientations are then evaluated using scoring functions to select the most stable binding configurations, and to allow researchers to screen and optimise candidate compounds for their ability to modulate skin processes effectively. For example, Ledwoń et al. (2023) investigated the potential of novel thiosemicarbazone (TSC)-peptide conjugates as inhibitors of skin pigmentation enzymes using Discovery Studio Visualizer software, revealing promising tyrosinase inhibitors for skin whitening in cosmeceutical applications. Similarly, Masjedi and Solhjoo (2022) investigated the potential of fenugreek seed-derived trigonelline with the software AutoDock, while (Budama-Kilinc et al. 2022) utilized molecular docking to explore chlorogenic acid's potential as a tyrosinase inhibitor for hyperpigmentation treatment, further supporting its use in cosmetic applications.

Molecular docking is also widely used to predict the efficacy of bioactive compounds, particularly in evaluating their potential as anti-aging agents. For example, Nurkolis et al. (2023) applied molecular docking to assess various compounds' effectiveness in promoting skin regeneration, offering valuable insights into anti-aging therapies. Similarly, Nutho and Tungmunnithum (2024) underscores the potential of flavonoids extracted from Asian water lily to inhibit key skin aging enzymes, including collagenase, elastase, and tyrosinase. These studies highlight the utility of molecular coupling to accelerate the discovery and optimization of cosmetic ingredients, providing a cost-effective and time-efficient approach to the development of skincare products.

One of the primary advantages of this *in Silico* technique is its ability to generate leads rapidly by predicting how small molecules will interact with biological targets. The lock-and-key model, a foundational concept in molecular docking, has demonstrated considerable success in identifying potential drug candidates across various biological systems. Gschwend et al. (1996) indicate that, when applied to enzymes, this technique achieves hit rates in the micromolar range of 2% and 20%, demonstrating its effectiveness in generating selective leads. However, the application of molecular docking has its challenges. Incomplete molecular structures and inherent shortcomings limit current methods in scoring functions, which hinder their accuracy in predicting binding affinities (Fan et al. 2019). These limitations arise from the oversimplification of interactions and the reliance on static models that do not account for the dynamic nature of molecular interactions. The flexibility of both the ligand and the target enzyme is only sometimes accurately captured, leading to potential errors in predicting the most stable binding conformation. This issue is particularly critical in Su et al. (2024), a study on

skin-lightening compounds where the focus is on developing tyrosinase inhibitors and precise binding interactions are crucial for efficacy.

Comprehensive biological data may be integrated into the scoring functions to provide a more nuanced understanding of binding interactions and improve accuracy (Enyedy and Egan 2008). Additionally, combining molecular docking with other computational methods, such as molecular dynamics (MD) simulations, can provide a more detailed and dynamic view of the binding process. MD simulations are a computational technique that is used to model the physical movements of atoms and molecules over time, offering a more realistic representation of molecular systems than static models. For instance, Nutho and Tungmunnithum (2023) combined *in Vitro* enzymatic assays with molecular docking and MD simulations. Their findings demonstrated that these flavonoids exhibited strong inhibitory potential and favorable skin permeability, which made them promising candidates for anti-aging cosmeceutical applications.

In addition, redocking is a validation technique that is often employed to enhance the accuracy of traditional docking methods. It involves removing a known ligand from its binding site and then redocking it to assess the performance of the docking algorithm. This process helps verify the consistency of the docking predictions and bolsters confidence in the reliability of the results. For example, in Su et al. (2024), a multidimensional molecular screening strategy that included redocking and MD simulations significantly improved the accuracy of identifying effective tyrosinase inhibitors.

In summary, while molecular docking is a powerful tool in the initial stages of drug and cosmetic development, its limitations must be addressed through integrated computational approaches and enhanced data utilization to improve accuracy and reliability.

5 Limitations

One limitation of this systematic review is that we relied on PubMed as the database for our search. While PubMed is a comprehensive resource for biomedical and life sciences literature, using a single database may have resulted in the omission of related studies indexed in other databases. Additionally, the search results are strongly dependent on the specific keywords used in the search strategy. Our keyword selection was focused on cosmetic safety and efficacy assessment, which means that studies outside of this defined scope were not included, potentially limiting the generalizability of our findings to broader safety assessment methodologies. Future reviews might consider expanding the database sources and broadening the keyword selection to encompass a wider domain.

6 Conclusion: practice, policy and future research implications

In the realm of cosmetics, ensuring safety is of utmost importance and requires thorough evaluation prior to product launch. Regulatory frameworks governing cosmetics exhibit significant variation across different regions. In the United

States, the Food and Drug Administration (FDA) oversees cosmetic safety under the Federal Food, Drug, and Cosmetic Act, while in Asia, regulatory bodies such as the State Administration for Market Regulation (SAMR) in China and the Dubai Municipality in the United Arab Emirates hold sway. Within the European Union (EU), Regulation (EC) No 1223/2009 sets out strict guidelines for the marketing of cosmetic products, and similar procedures are in place in countries such as the UK, Australia, Canada, and various Asian countries. Thus, understanding and adhering to the specific regulations of the target market is of utmost importance for cosmetic manufacturers (Ferreira et al. 2022).

We have reviewed the entire spectra of statistical methods for cosmetic assessment to guide researchers and practitioners in selecting the appropriate approach according to the research question. We briefly described the inherent nature of each approach, and we discussed their advantages and limitations to support informed decision-making in cosmetic research.

As the cosmetics industry embraces innovation, statistical and *in Silico* methodologies are increasingly integrated into safety and efficacy evaluations. These theoretical approaches offer multiple advantages, mainly reducing animal testing, and also improving the efficiency of product evaluation and aligning with the overarching goal of minimizing traditional testing methods (Silva and Tamburic 2022).

The evolution of *in Silico* toxicology has been marked by significant advancements, using expansive databases, sophisticated models and algorithms, and various computational techniques. This progress has led to various *in Silico* methods covering a wide range of chemicals and toxicological endpoints. Such models are ultimately simulations of reality, and accurately reflecting the intricate physiological processes that lead to toxicity remains a major challenge (Kimber et al. 2011). Confidence in these models can be significantly improved when used alongside other critical information - such as their limitations, applicability domains, as well as data on metabolites and exposure - creating a more comprehensive framework for risk assessment (Stouch et al. 2003).

In a rapid overview, we have seen that when applying QSAR, it is essential to define the applicability domain. Linear models offer simplicity and interpretability, while non-linear models can reveal more intricate relationships but often lack transparency. Rule-based methods aid qualitative screening, and IBRs complement HBRs with AI-assisted structural alerts. Read-across approaches are easy to interpret and work well in integrated workflows. Hybrid models provide effective first-tier screening tools. Molecular coupling predicts molecular interactions rapidly, and the integration of biological data into scoring functions enhances accuracy.

To enhance the reliability of in Silico predictions, it is essential to improve the quality, standardization, and accessibility of datasets.

First, ensuring high data quality is fundamental for computational toxicology models (Cherkasov et al. 2014). AI-driven predictive methods are particularly sensitive to inconsistencies in data annotation, duplication, and poor reproducibility, which can lead to misleading results and inflated performance metrics. Rigorous data curation is essential to mitigate these risks and enhance model accuracy (Alves et al. 2021).

Second, the lack of broadly accepted standards for data integration and model validation presents a major obstacle to the reproducibility and regulatory acceptance of in Silico approaches. The heterogeneity of data sources-spanning different formats, collection methodologies, and annotation conventions- may complicates cross-study comparability and model generalizability. Harmonization efforts, such as those initiated by EU-STANDS4PM, highlight the necessity of establishing common frameworks to facilitate the interpretability of computational models in toxicology and biomedical research (Brunak et al. 2020).

Finally, accessibility remains a key factor in improving the robustness of in Silico assessments. Curated, publicly available repositories would not only enhance validation efforts but also encourage broader collaboration within the scientific community. At the same time, increasing data sharing initiatives will require the implementation of privacy-preserving techniques to address privacy concerns (Luechtefeld et al. 2018).

Overcoming these challenges can make *in Silico* methods more reliable, reproducible and acceptable in toxicology.

As data continues to grow in volume and complexity, integrating machine learning into *in Silico* approaches transforms chemical hazard assessment and offers robust and reliable alternatives to conventional methods (Luechtefeld et al. 2018). However, the complexity of these models often results in a "black-box" nature, posing challenges for transparency and interpretability (Guha 2008). To address transparency, standardized reporting guidelines that clearly document model architectures, training procedures, and evaluation metrics should be developed (Togo et al. 2022). Additionally, prioritizing interpretability in computational models remains crucial. Hybrid approaches that balance predictive accuracy with transparency could provide a viable path forward, ensuring both performance and regulatory acceptance.

A key limitation in current *in Silico* models is their reliance on animal-derived or limited human datasets. This scarcity results in restricted structural diversity and a small pool of available data; in such cases, the creation of a robust predictive model is severely hampered (Cherkasov et al. 2014). To address this issue, data augmentation methods are being explored (Papadopoulos and Karalis 2023). These approaches can partially overcome the limitations of small clinical studies.

To date, most published *in Silico* models have focused on classifying chemicals as safe or unsafe, which limits their applicability, particularly for weak or moderately unsafe substances in commercial products. Introducing multi-class models that distinguish between mild, moderate, and severe toxic compounds would broaden their practical applications. This enhancement is particularly crucial for products like rinse-off cosmetics, which remain on the skin surface for a brief period (Ta et al. 2021).

Additionally, to ensure the safety of cosmetics, an urgent need exists for data-driven approaches to aggregate/mixture exposure assessment that closely reflects real-world scenarios. Unlike individual chemicals, mixtures can exhibit complex, non-additive toxicity profiles, leading to synergistic or antagonistic effects. These interactions significantly complicate the assessment process. In this context, Kar and Leszczynski (2019) demonstrated the effective application of *in Silico* tools for predicting the toxicity of chemical mixtures, highlighting

the potential of computational methods to enhance safety evaluations in cosmetic formulations.

Looking ahead, while *in Silico* methods show promise for specific toxicological endpoints such as skin sensitization and genotoxicity, there remain gaps in their applicability, particularly for endpoints like repeated dose toxicity and reproductive toxicity. This underscores the ongoing necessity for further research and development in this field to enhance its regulatory utility and efficacy.

The landscape of efficacy assessment in cosmetics presents unique challenges, particularly in the realm of *cosmeceuticals* - topical preparations that, while sold as cosmetics, exhibit characteristics akin to pharmaceutical action. Dr. Kligman, who coined the term during his pivotal experimentation with the anti-aging effects of tretinoin, emphasized this dual nature (Kligman 2005). Despite the well-defined regulatory frameworks that ensure the safety of cosmetics through established toxicity endpoints, efficacy claims often do not undergo the same rigorous scrutiny. This gap highlights the importance of *in Silico* techniques like molecular docking, which has emerged as a key method in evaluating the pharmaceutical potential of cosmeceuticals.

Molecular docking, in particular, allows researchers to explore and predict the interactions between bioactive compounds and skin proteins. By optimizing the binding affinity of these molecules, the method enhances both the stability and overall efficacy of cosmeceutical formulations. In the context of anti-aging, docking studies have proven instrumental in assessing the potential of compounds to inhibit key enzymes like collagenase, elastase, and tyrosinase, which are involved in skin aging. This approach offers a modern, data-driven pathway to evaluate and substantiate efficacy claims, ensuring that cosmeceutical products deliver measurable results alongside their cosmetic appeal.

In conclusion, we have seen *in Silico* methods provide valuable tools for cosmetic safety and efficacy assessment, although their optimal utilization requires an integrated, multidisciplinary approach. Future advances are expected to be driven by innovations in *in Vitro* and sensory technologies, Machine Learning (ML), and personalized product development. A key priority will be leveraging ML to develop more tailored cosmetic formulations, ensuring products align with individual skin profiles and sensitivities. At the same time, enhancing algorithm transparency and refining regulatory frameworks will be essential to ensure compliance, safety, and public trust. Ethical considerations, including data privacy and responsible decision-making, must also be addressed to foster acceptance and reliability in ML-driven assessments (Mirakhori and Niazi 2025). Establishing standardized datasets and harmonizing international regulations will play a crucial role in shaping a more rigorous and globally accepted framework for cosmetic testing.

7 Registration and protocol

The protocol for this systematic review was prospectively registered on PROSPERO with registration number CRD420251029213. No amendments were made to the information provided at the time of registration.

Appendix

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Page 1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Page 1
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 1-2, Section 1
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 2, Section 1
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Page 3-5, Sections 2.2-2.3
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 3, Section 2.1
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Page 3, Section 2.1
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Page 4, sections 2.1-2.2
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 4-5, section 2.3
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Page 4-5, section 2.3
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Page 4-5, section 2.3
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Page 3, Section 2
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Not applicable[1]
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Page 4-5, section 2.3
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Page 5, section 2.3
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Page 4, section 2.3
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Page 4, section 2.3
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Not applicable[2]
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Not applicable[2]
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Not applicable[2]
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Not applicable[2]

PRISMA 2020 Checklist (segue)

Section and Topic	Item #	Checklist item	Location where item is reported
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Page 6, Figure 1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Page 5, Section 3.1
Study characteristics	17	Cite each included study and present its characteristics.	Page 8-9, Tables1-2
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Not applicable[2]
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Not applicable[2]
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Not applicable[2]
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Not applicable[2]
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Not applicable[2]
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Not applicable[2]
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Not applicable[2]
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Not applicable[2]
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Page 7-22, Section 4
	23b	Discuss any limitations of the evidence included in the review.	Page 22, Section 5
	23c	Discuss any limitations of the review processes used.	Page 22, Section 5
	23d	Discuss implications of the results for practice, policy, and future research.	Page 23-26, Section 6
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Page 26, Section 8
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Page 26, Section 8
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Page 26, Section 8
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Page 26, Section 9
Competing interests	26	Declare any competing interests of review authors.	Page 26, Section 10
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Page 26, Section 7

Acknowledgements This research has been supported by the MUSA - Multilayered Urban Sustainability Action - project, funded by the European Union - NextGenerationEU, under the National Recovery and Resilience Plan (NRRP) Mission 4 Component 2 Investment Line 1.5: Strengthening of research structures and creation of R&D "innovation ecosystems", set up of "territorial leaders in R&D".

Funding: The research was supported by the Multilayered Urban Sustainability Action project, funded by the European Union NextGenerationEU, National Recovery and Resilience Plan. PROSPERO registration number: CRD420251029213

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest We declare that there are no Conflict of interest regarding the publication of this paper.

References

Abe FR, Mendonça JN, Moraes LA et al (2017) Toxicological and behavioral responses as a tool to assess the effects of natural and synthetic dyes on zebrafish early life. Chemosphere 178:282–290

Ahmed B, Dwivedi S, Abdin MZ et al (2017) Mitochondrial and chromosomal damage induced by oxidative stress in zn2+ ions, zno-bulk and zno-nps treated allium cepa roots. Sci Rep 7(1):40685

Alonso C, Lucas R, Barba C et al (2015) Skin delivery of antioxidant surfactants based on gallic acid and hydroxytyrosol. J Pharm Pharmacol 67(7):900–908

Alves VM, Auerbach SS, Kleinstreuer N et al (2021) Curated data in-trustworthy in silico models out: the impact of data quality on the reliability of artificial intelligence models as alternatives to animal testing. Altern Lab Anim 49(3):73–82

Ambe K, Suzuki M, Ashikaga T et al (2021) Development of quantitative model of a local lymph node assay for evaluating skin sensitization potency applying machine learning catboost. Regul Toxicol Pharmacol 125:105019

Anand AS, Prasad DN, Singh SB et al (2017) Chronic exposure of zinc oxide nanoparticles causes deviant phenotype in drosophila melanogaster. J Hazard Mater 327:180–186

Andreoli C, Leter G, De Berardis B et al (2018) Critical issues in genotoxicity assessment of TiO2 nanoparticles by human peripheral blood mononuclear cells. J Appl Toxicol 38(12):1471–1482

Api A, Belsito D, Botelho D et al (2019) Rifm fragrance ingredient safety assessment, propanal diethyl acetal, cas registry number 4744–08-5. Food Chem Toxicol 130:110588

Applebaum SA (2021) Practical aspects of cosmetic testing: How to set up a scientific study in skin physiology. Plast Reconstr Surg 148(3):700–701

Arnesdotter E, Rogiers V, Vanhaecke T et al (2021) An overview of current practices for regulatory risk assessment with lessons learnt from cosmetics in the European union. Crit Rev Toxicol 51(5):395–417

Asturiol D, Casati S, Worth A (2016) Consensus of classification trees for skin sensitisation hazard prediction. Toxicol In Vitro 36:197–209

Ates G, Steinmetz FP, Doktorova TY et al (2016) Linking existing in vitro dermal absorption data to physicochemical properties: contribution to the design of a weight-of-evidence approach for the safety evaluation of cosmetic ingredients with low dermal bioavailability. Regul Toxicol Pharmacol 76:74–78

Baderna D, Gadaleta D, Lostaglio E et al (2020) New in silico models to predict in vitro micronucleus induction as marker of genotoxicity. J Hazard Mater 385:121638

Baltazar MT, Cable S, Carmichael PL et al (2020) A next-generation risk assessment case study for coumarin in cosmetic products. Toxicol Sci 176(1):236–252

Bamford E, Grahn A, Århammar C et al (2021) Mesoporous magnesium carbonate for use in powder cosmetics. Int J Cosmet Sci 43(1):57–67
Bar-Meir M, Bendelac S, Shchors I (2023) Chlorhexidine bathing in a tertiary care neonatal intensive care unit: a pilot study. PLoS ONE 18(3):e0283132
Barthe M, Bavoux C, Finot F et al (2021) Safety testing of cosmetic products: overview of established methods and new approach methodologies (nams). Cosmetics 8(2):50
Batke M, Gütlein M, Partosch F et al (2016) Innovative strategies to develop chemical categories using a combination of structural and toxicological properties. Front Pharmacol 7:321
Bender A, Glen RC (2004) Molecular similarity: a key technique in molecular informatics. Organic Biomol Chem 2(22):3204–3218
Bengalli R, Colantuoni A, Perelshtein I et al (2021) In vitro skin toxicity of cuo and zno nanoparticles: application in the safety assessment of antimicrobial coated textiles. NanoImpact 21:100282
Besrour N, Oludemi T, Mandim F et al (2022) Valorization of Juglans regia leaves as cosmeceutical ingredients: bioactivity evaluation and final formulation development. Antioxidants 11(4):677
Bora NS, Pathak MP, Mandal S et al (2017) Safety assessment and toxicological profiling of a novel combinational sunprotective dermal formulation containing melatonin and pumpkin seed oil. Regul Toxicol Pharmacol 89:1–12
Borba JV, Braga RC, Alves VM et al (2020) Pred-skin: a web portal for accurate prediction of human skin sensitizers. Chem Res Toxicol 34(2):258–267
Borba JV, Alves VM, Braga RC et al (2022) STopTox: an in silico alternative to animal testing for acute systemic and topical toxicity. Environ Health Perspect 130(2):027012
Brathwaite ACN, Alencar-Silva T, Carvalho LA et al (2022) Pouteria macrophylla fruit extract microemulsion for cutaneous depigmentation: evaluation using a 3d pigmented skin model. Molecules 27(18):5982
Brown A, Furmanczyk M, Ramos D et al (2023) Natural retinol analogs potentiate the effects of retinal on aged and photodamaged skin: results from in vitro to clinical studies. Dermatol Therapy 13(10):2299–2317
Brunak S, Bjerre Collin C, Cathaoir Eva Ó, K, et al (2020) Towards standardization guidelines for in silico approaches in personalized medicine. J Integrat Bioinf 17(2–3):20200006
Budama-Kilinc Y, Gok B, Kecel-Gunduz S et al (2022) Development of nanoformulation for hyperpigmentation disorders: experimental evaluations, in vitro efficacy and in silico molecular docking studies. Arab J Chem 15(12):104362
Campodoni E, Montanari M, Artusi C et al (2023) Biomineralization: a new tool for developing ecosustainable ti-doped hydroxyapatite-based hybrid uv filters. Biomater Adv 151:213474
Canavez ADPM, de Oliveira Prado Corrêa G, Isaac VLB, et al (2021) Integrated approaches to testing and assessment as a tool for the hazard assessment and risk characterization of cosmetic preservatives. J Appl Toxicol 41(10):1687–1699
Canipa SJ, Chilton ML, Hemingway R et al (2017) A quantitative in silico model for predicting skin sensitization using a nearest neighbours approach within expert-derived structure-activity alert spaces. J Appl Toxicol 37(8):985–995
Chansriniyom C, Nooin R, Nuengchamnong N et al (2021) Tandem mass spectrometry of aqueous extract from ficus dubia sap and its cell-based assessments for use as a skin antioxidant. Sci Rep 11(1):16899
Chapman K, Thomas A, Wills J et al (2014) Automation and validation of micronucleus detection in the 3d epiderm ™ human reconstructed skin assay and correlation with 2d dose responses. Mutagenesis 29(3):165–175
Chen HY, Cheng KC, Wang HT et al (2021) Extracts of antrodia cinnamomea mycelium as a highly potent tyrosinase inhibitor. J Cosmet Dermatol 20(7):2341–2349
Chen L, Li N, Liu Y et al (2021) A new 3d model for genotoxicity assessment: Episkin ™ micronucleus assay. Mutagenesis 36(1):51–61
Chen T, Chang H (2023) In vitro eye irritation testing models may play pivotal role in effort to pursue mild baby cleansers. Toxicol In Vitro 89:105578
Cherkasov A, Muratov EN, Fourches D et al (2014) Qsar modeling: Where have you been? Where are you going to? J Med Chem 57(12):4977–5010
Cho SA, An S, Park JH (2019) High-throughput screening (hts)-based spectrophotometric direct peptide reactivity assay (spectro-dpra) to predict human skin sensitization potential. Toxicol Lett 314:27–36

Choudhury SR, Ordaz J, Lo CL et al (2017) From the cover: zinc oxide nanoparticles-induced reactive oxygen species promotes multimodal cyto-and epigenetic toxicity. Toxicol Sci 156(1):261–274
Cosmetics Europe (2008) Guidelines for the evaluation of the efficacy of cosmetics products. COLIPA
Cronin M, Madden J (2010) In silico toxicology: principles and applications. Royal Society of Chemistry, Cambridge
Cui H, Feng C, Zhang T et al (2023) Effects of a lotion containing probiotic ferment lysate as the main functional ingredient on enhancing skin barrier: a randomized, self-control study. Sci Rep 13(1):16879
De Silva W, Pathiratne A (2023) Nano-titanium dioxide induced genotoxicity and histological lesions in a tropical fish model, nile tilapia (oreochromis niloticus). Environ Toxicol Pharmacol 98:104043
Deeb O, Goodarzi M (2012) In silico quantitative structure toxicity relationship of chemical compounds: some case studies. Curr Drug Saf 7(4):289–297
Demir E, Turna F, Vales G et al (2013) In vivo genotoxicity assessment of titanium, zirconium and aluminium nanoparticles, and their microparticulated forms, in drosophila. Chemosphere 93(10):2304–2310
Dimitrov S, Mekenyan O (2010) An introduction to read-across for the prediction of the effects of chemicals. In: Cronin M, Madden J (eds) In silico toxicology: principles and applications. Royal Society of Chemistry, Cambridge, pp 372–383
Dimitrov S, Detroyer A, Piroird C et al (2016) Accounting for data variability, a key factor in in vivo/in vitro relationships: application to the skin sensitization potency (in vivo llna versus in vitro dpra) example. J Appl Toxicol 36(12):1568–1578
Diniz RR, Paiva JP, Aquino RM et al (2019) Saccharomyces cerevisiae strains as bioindicators for titanium dioxide sunscreen photoprotective and photomutagenic assessment. J Photochem Photobiol B 198:111584
Dorier M, Tisseyre C, Dussert F et al (2019) Toxicological impact of acute exposure to e171 food additive and tio2 nanoparticles on a co-culture of caco-2 and ht29-mtx intestinal cells. Mutation research/genetic toxicology and environmental mutagenesis 845:402980
Duracher L, Visdal-Johnsen L, Mavon A (2015) In vitro and in vivo dermal absorption assessment of acetyl aspartic acid: a compartmental study. Int J Cosmet Sci 37:34–40
Dvořáková M, Svobodová L, Rucki M et al (2023) The safety assessment of cosmetic perfumes by using in chemico and in vitro methods in combination with gc-ms/ms analysis. Altern Lab Anim 51(4):224–248
Eberle CE, Sandler DP, Taylor KW et al (2020) Hair dye and chemical straightener use and breast cancer risk in a large us population of black and white women. Int J Cancer 147(2):383–391
Efthimiou I, Georgiou Y, Vlastos D et al (2020) Assessing the cyto-genotoxic potential of model zinc oxide nanoparticles in the presence of humic-acid-like-polycondensate (halp) and the leonardite ha (lha). Sci Total Environ 721:137625
Sj Enoch (2010) Chemical category formation and read-across for the prediction of toxicity. In: Puzyn T, Leszczynski J, Cronin MT (eds) Recent advances in QSAR studies: methods and applications. Springer, Netherlands, Dordrecht, pp 209–219
Enoch SJ, Cronin MTD, Schultz TW et al (2008) Quantitative and mechanistic read across for predicting the skin sensitization potential of alkenes acting via Michael addition. Chem Res Toxicol 21(2):513–520
Enyedy IJ, Egan WJ (2008) Can we use docking and scoring for hit-to-lead optimization? J Comput Aided Mol Des 22:161–168
Erhirhie EO, Ihekwereme CP, Ilodigwe EE (2018) Advances in acute toxicity testing: strengths, weaknesses and regulatory acceptance. Interdiscip Toxicol 11(1):5–12
Estrela FN, Guimarães ATB, da Costa Araújo AP et al (2021) Toxicity of polystyrene nanoplastics and zinc oxide to mice. Chemosphere 271:129476
Fan J, Fu A, Zhang L (2019) Progress in molecular docking. Quantitat Biol 7:83–89
Ferreira M, Matos A, Couras A et al (2022) Overview of cosmetic regulatory frameworks around the world. Cosmetics 9(4):72
Fitzpatrick JM, Roberts DW, Patlewicz G (2018) An evaluation of selected (q) sars/expert systems for predicting skin sensitisation potential. SAR QSAR Environ Res 29(6):439–468
Frombach J, Sonnenburg A, Krapohl BD et al (2018) Lymphocyte surface markers and cytokines are suitable for detection and potency assessment of skin-sensitizing chemicals in an in vitro model of allergic contact dermatitis: the lcsa-ly. Arch Toxicol 92:1495–1505

Fung ES, Towle KM, Monnot AD (2020) Devising a tier-based skin sensitisation screening strategy for personal care and cosmetic products. Altern Lab Anim 48(2):70–77
Gao Y, Li G, Qin Y et al (2019) New theoretical insight into indirect photochemical transformation of fragrance nitro-musks: mechanisms, eco-toxicity and health effects. Environ Int 129:68–75
Garcia I, Pouzet C, Brulas M et al (2013) Evaluation of thp-1 cell line as an in vitro model for long-term safety assessment of new molecules. Int J Cosmet Sci 35(6):568–574
Gautier F, Tourneix F, Vandecasteele HA et al (2020) Read-across can increase confidence in the next generation risk assessment for skin sensitisation: a case study with resorcinol. Regul Toxicol Pharmacol 117:104755
Gentile L (2019) Assessment of iron oxide nanoparticle ecotoxicity on regeneration and homeostasis in the replacement model system schmidtea mediterranea. Altex 36(4):583–596
Germano-Costa T, Bilesky-José N, Guilger-Casagrande M et al (2022) Use of 2d and co-culture cell models to assess the toxicity of zein nanoparticles loading insect repellents icaridin and geraniol. Colloids Surf B 216:112564
Gilmour N, Reynolds J, Przybylak K et al (2022) Next generation risk assessment for skin allergy: decision making using new approach methodologies. Regul Toxicol Pharmacol 131:105159
Gleeson MP, Modi S, Bender A et al (2012) The challenges involved in modeling toxicity data in silico: a review. Curr Pharm Des 18(9):1266–1291
Gnonsoro UP, Ake Assi YED, Sangare NS, et al (2022) Health risk assessment of heavy metals (pb, cd, hg) in hydroalcoholic gels of abidjan, côte d'ivoire. Biol Trace Element Res 1–9
Goh GB, Hodas NO, Vishnu A (2017) Deep learning for computational chemistry. J Comput Chem 38(16):1291–1307
Golembo M, Puttagunta S, Rappo U et al (2022) Development of a topical bacteriophage gel targeting cutibacterium acnes for acne prone skin and results of a phase 1 cosmetic randomized clinical trial. Skin Health Disease 2(2):e93
Gomes AR, de Matos LP, Guimarães ATB et al (2023) Plant-zno nanoparticles interaction: An approach to improve guinea grass (panicum maximum) productivity and evaluation of the impacts of its ingestion by freshwater teleost fish. J Hazard Mater 451:131173
Gschwend DA, Good AC, Kuntz ID (1996) Molecular docking towards drug discovery. J Mol Recognit: Interdiscip J 9(2):175–186
Guha R (2008) On the interpretation and interpretability of quantitative structure-activity relationship models. J Comput Aided Mol Des 22:857–871
Haddaway NR, Page MJ, Pritchard CC et al (2022) Prisma 2020: An r package and shiny app for producing prisma 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. Campbell Syst Rev 18(2):e1230
Hadrup N, Rahmani F, Jacobsen NR et al (2019) Acute phase response and inflammation following pulmonary exposure to low doses of zinc oxide nanoparticles in mice. Nanotoxicology 13(9):1275–1292
Han SM, Park KK, Nicholls YM et al (2013) Effects of honeybee (apis mellifera) venom on keratinocyte migration in vitro. Pharmacogn Mag 9(35):220
Hatanaka T, Yoshida S, Kadhum WR et al (2015) In silico estimation of skin concentration following the dermal exposure to chemicals. Pharm Res 32:3965–3974
Hazarika H, Krishnatreyya H, Tyagi V et al (2022) The fabrication and assessment of mosquito repellent cream for outdoor protection. Sci Rep 12(1):2180
Heikkinen S, Pitkäniemi J, Sarkeala T et al (2015) Does hair dye use increase the risk of breast cancer? A population-based case-control study of finnish women. PLoS ONE 10(8):e0135190
Hemmerich J, Ecker GF (2020) In silico toxicology: from structure-activity relationships towards deep learning and adverse outcome pathways. Wiley interdisciplinary reviews: computational molecular science 10(4):e1475
Hewitt M, Cronin MT, Enoch SJ et al (2009) In silico prediction of aqueous solubility: the solubility challenge. J Chem Inf Model 49(11):2572–2587
Heylings JR, Davies DJ, Burton R (2018) Dermal absorption of testosterone in human and pig skin in vitro. Toxicol In Vitro 48:71–77
Hirota M, Fukui S, Okamoto K et al (2015) Evaluation of combinations of in vitro sensitization test descriptors for the artificial neural network-based risk assessment model of skin sensitization. J Appl Toxicol 35(11):1333–1347

Hirota M, Ashikaga T, Kouzuki H (2018) Development of an artificial neural network model for risk assessment of skin sensitization using human cell line activation test, direct peptide reactivity assay, keratinosens ™ and in silico structure alert parameter. J Appl Toxicol 38(4):514–526
Hisaki T, née Kaneko MA, Yamaguchi M, et al (2015) Development of qsar models using artificial neural network analysis for risk assessment of repeated-dose, reproductive, and developmental toxicities of cosmetic ingredients. J Toxicol Sci 40(2):163–180
Ho YB, Abdullah NH, Hamsan H et al (2017) Mercury contamination in facial skin lightening creams and its health risks to user. Regul Toxicol Pharmacol 88:72–76
Iliopoulos F, Chapman A, Lane ME (2021) A comparison of the in vitro permeation of 3-o-ethyl-l-ascorbic acid in human skin and in a living skin equivalent (labskin ™). Int J Cosmet Sci 43(1):107–112
Ilves M, Palomäki J, Vippola M et al (2014) Topically applied zno nanoparticles suppress allergen induced skin inflammation but induce vigorous ige production in the atopic dermatitis mouse model. Part Fibre Toxicol 11:1–12
Im JE, Lee JD, Kim HY et al (2023) Prediction of skin sensitization using machine learning. Toxicol In Vitro 93:105690
Imai N, Takeyoshi M, Aizawa S et al (2022) Improved performance of the sh test as an in vitro skin sensitization test with a new predictive model and decision tree. J Appl Toxicol 42(6):1029–1043
Jalili P, Huet S, Burel A et al (2022) Genotoxic impact of aluminum-containing nanomaterials in human intestinal and hepatic cells. Toxicol In Vitro 78:105257
Jeliazkova N, Jaworska J, Worth A et al (2010) Open source tools for read-across and category formation. In: Cronin M, Madden J (eds) In silico toxicology: principles and applications. Royal Society of Chemistry, Cambridge, pp 408–445
Jeon B, Lim MH, Choi TH et al (2022) A development of a graph-based ensemble machine learning model for skin sensitization hazard and potency assessment. J Appl Toxicol 42(11):1832–1842
Johansson H, Albrekt AS, Borrebaeck CA et al (2013) The gard assay for assessment of chemical skin sensitizers. Toxicol In Vitro 27(3):1163–1169
Johansson H, Rydnert F, Kühnl J et al (2014) Genomic allergen rapid detection in-house validation-a proof of concept. Toxicol Sci 139(2):362–370
Johansson H, Gradin R, Forreryd A et al (2017) Evaluation of the gard assay in a blind cosmetics Europe study. ALTEX-alternatives to animal experimentation 34(4):515–523
Jordão AC, Dos Santos GS, Teixeira TR et al (2024) Assessment of the photoprotective potential and structural characterization of secondary metabolites of antarctic fungus arthrinium sp. Arch Microbiol 206(1):35
Juncan AM, Morgovan C, Rus LL et al (2023) Development and evaluation of a novel anti-ageing cream based on hyaluronic acid and other innovative cosmetic actives. Polymers 15(20):4134
Kalčíková G, Roy R, Klun B et al (2023) Interactions between titanium dioxide nanoparticles and polyethylene microplastics: adsorption kinetics, photocatalytic properties, and ecotoxicity. Chemosphere 329:138628
Kaluzhny Y, Kandárová H, Handa Y et al (2015) The epiocular eye irritation test (eit) for hazard identification and labelling of eye irritating chemicals: protocol optimisation for solid materials and the results after extended shipment. Altern Lab Anim 43(2):101–127
Kao JK, Hsu TF, Lee MS et al (2020) Subcutaneous injection of recombinant heat shock protein 70 ameliorates atopic dermatitis skin lesions in a mouse model. Kaohsiung J Med Sci 36(3):186–195
Kar S, Leszczynski J (2019) Exploration of computational approaches to predict the toxicity of chemical mixtures. Toxics 7(1):15
Kar S, Roy K (2010) Qsar modeling of toxicity of diverse organic chemicals to daphnia magna using 2d and 3d descriptors. J Hazard Mater 177(1–3):344–351
Keck CM, Eo Chaiprateep, Dietrich H et al (2022) Influence of mechanical skin treatments on dermal penetration efficacy of active ingredients. Pharmaceutics 14(9):1788
Kidd D, Phillips S, Chirom T et al (2021) The 3d reconstructed skin micronucleus assay: considerations for optimal protocol design. Mutagenesis 36(1):37–49
Kim CW, Park GT, Bae ON et al (2016) Application of sv40 t-transformed human corneal epithelial cells to evaluate potential irritant chemicals for in vitro alternative eye toxicity. J Pharmacol Toxicol Methods 80:82–89
Kim KB, Kwack SJ, Lee JY et al (2021) Current opinion on risk assessment of cosmetics. J Toxicol Environ Health Part B 24(4):137–161

Kimber I, Humphris C, Westmoreland C, et al (2011) Computational chemistry, systems biology and toxicology. harnessing the chemistry of life: revolutionizing toxicology. A commentary. J Appl Toxicol 31(3):206–209
Kizhedath A, Wilkinson S, Glassey J (2019) Assessment of hepatotoxicity and dermal toxicity of butyl paraben and methyl paraben using hepg2 and hdfn in vitro models. Toxicol In Vitro 55:108–115
Kleinstreuer NC, Hoffmann S, Alépée N et al (2018) Non-animal methods to predict skin sensitization (ii): an assessment of defined approaches. Crit Rev Toxicol 48(5):359–374
Kligman M (2005) The future of cosmeceuticals: An interview with Albert. Dermatol Surg 31(7 Part 2)
Kobets T, Duan JD, Brunnemann KD et al (2018) In ovo testing of flavor and fragrance materials in turkey egg genotoxicity assay (tega), comparison of results to in vitro and in vivo data. Food Chem Toxicol 115:228–243
Kostal J, Voutchkova-Kostal A (2016) Cadre-ss, an in silico tool for predicting skin sensitization potential based on modeling of molecular interactions. Chem Res Toxicol 29(1):58–64
Kraeling ME, Topping VD, Keltner ZM et al (2018) In vitro percutaneous penetration of silver nanoparticles in pig and human skin. Regul Toxicol Pharmacol 95:314–322
Leach AR (2001) In: Chemistry Central Journal. Pearson education
Lebre F, Chatterjee N, Costa S, et al (2022) Nanosafety: an evolving concept to bring the safest possible nanomaterials to society and environment. Nanomaterials 12(11):1810
Ledwoń P, Goldeman W, Hałdys K et al (2023) Tripeptides conjugated with thiosemicarbazones: new inhibitors of tyrosinase for cosmeceutical use. J Enzyme Inhib Med Chem 38(1):2193676
Lee YJ, Kim HY, Pham QL et al (2022) Pharmacokinetics and the dermal absorption of bromochlorophene, a cosmetic preservative ingredient, in rats. Toxics 10(6):329
Leelawattanachai J, Panyasu K, Prasertsom K et al (2023) Highly stable and fast-dissolving ascorbic acid-loaded microneedles. Int J Cosmet Sci 45(5):612–626
Leist M, Hasiwa N, Rovida C et al (2014) Consensus report on the future of animal-free systemic toxicity testing. Alternatives to Animal Experimentation: ALTEX 31(3):341–356
Leoni C, Majorani C, Cresti R, et al (2023) Determination and risk assessment of phthalates in face masks. An Italian study. J Hazard Mater 443:130176
Leontaridou M, Gabbert S, Van Ierland EC et al (2016) Evaluation of non-animal methods for assessing skin sensitisation hazard: a bayesian value-of-information analysis. Altern Lab Anim 44(3):255–269
Lepailleur A, Poezevara G, Bureau R (2013) Automated detection of structural alerts (chemical fragments) in (eco) toxicology. Comput Struct Biotechnol J 5(6):e201302013
Leskur D, Bukić J, Petrić A et al (2019) Anatomical site differences of sodium lauryl sulfate-induced irritation: randomized controlled trial. Br J Dermatol 181(1):175–185
Li H, Bai J, Zhong G et al (2019) Improved defined approaches for predicting skin sensitization hazard and potency in humans. ALTEX 36(3):363–372
Liu Y, Li N, Chen L et al (2020) A ready-to-use integrated in vitro skin corrosion and irritation testing strategy using episkin ^{TM} model in china. Toxicol In Vitro 65:104778
Luechtefeld T, Maertens A, Russo DP et al (2016) Global analysis of publicly available safety data for 9,801 substances registered under reach from 2008–2014. Altex 33(2):95
Luechtefeld T, Rowlands C, Hartung T (2018) Big-data and machine learning to revamp computational toxicology and its use in risk assessment. Toxicol Res 7(5):732–744
Lyons AB, Kohli I, Nahhas AF et al (2020) Trichloroacetic acid model to accurately capture the efficacy of treatments for postinflammatory hyperpigmentation. Arch Dermatol Res 312:725–730
Maciel B, Moreira P, Carmo H et al (2019) Implementation of an in vitro methodology for phototoxicity evaluation in a human keratinocyte cell line. Toxicol In Vitro 61:104618
Masjedi M, Solhjoo A (2022) Does trigonelline help skin tone? Molecular docking studies of trigonelline on the human tyrosinase, formulation, optimization, and characterization of an emulgel-containing trigonella foenum-graecum l. fenugreek standardized hydroalcoholic extract. J Cosmet Dermatol 21(12):7178–7193
Matsumoto M, Todo H, Akiyama T et al (2016) Risk assessment of skin lightening cosmetics containing hydroquinone. Regul Toxicol Pharmacol 81:128–135
Milan C, Schifanella O, Roncaglioni A et al (2011) Comparison and possible use of in silico tools for carcinogenicity within reach legislation. J Environ Sci Health C 29(4):300–323
Miller E, Beckett E, Roberts B et al (2024) Assessment of worst-case potential airborne asbestos exposure associated with the use of cosmetic talc: application of an exponential decay model. Environ Monit Assess 196(1):39

Mirakhori F, Niazi SK (2025) Harnessing the ai/ml in drug and biological products discovery and development: the regulatory perspective. Pharmaceuticals 18(1):47
Modi S, Hughes M, Garrow A et al (2012) The value of in silico chemistry in the safety assessment of chemicals in the consumer goods and pharmaceutical industries. Drug Discovery Today 17(3–4):135–142
Mottola F, Santonastaso M, Iovine C et al (2021) Adsorption of cd to tio2-nps forms low genotoxic aggregates in Zebrafish cells. Cells 10(2):310
Mottola F, Santonastaso M, Iovine C et al (2022) Tio 2-nps and cadmium co-exposure: in vitro assessment of genetic and genomic dna damage on dicentrarchus labrax embryonic cells. Environ Sci Pollut Res 29:1–11
Narda M, Ramos-Lopez D, Mun G et al (2019) Three-tier testing approach for optimal ocular tolerance sunscreen. Cutan Ocul Toxicol 38(3):212–220
Navabhatra A, Brantner A, Yingngam B (2022) Artificial neural network modeling of nanostructured lipid carriers containing 5-o-caffeoylquinic acid-rich cratoxylum formosum leaf extract for skin application. Adv Pharm Bull 12(4):801
Nepalia A, Singh A, Mathur N et al (2021) Skincare products as sources of mutagenic exposure to infants: an imperative study using a battery of microbial bioassays. Arch Environ Contam Toxicol 80:499–506
Neubert RH, Sommer E, Schölzel M et al (2018) Dermal peptide delivery using enhancer moleculs and colloidal carrier systems. Part ii: Tetrapeptide pkek. Eur J Pharm Biopharm 124:28–33
Neves BJ, Braga RC, Melo-Filho CC et al (2018) Qsar-based virtual screening: advances and applications in drug discovery. Front Pharmacol 9:1275
Nikolova N, Jaworska J (2003) Approaches to measure chemical similarity-a review. QSAR & Combin Sci 22(9–10):1006–1026
Nikolova-Jeliazkova N, Jaworska J (2005) An approach to determining applicability domains for qsar group contribution models: an analysis of src kowwin. Altern Lab Anim 33(5):461–470
Nobile V (2016) Guidelines on cosmetic efficacy testing on humans. Ethical, technical, and regulatory requirements in the main cosmetics markets. J Cosmo Trichol 2:107
Nurkolis F, Purnomo AF, Alisaputra D et al (2023) In silico and in vitro studies reveal a synergistic potential source of novel anti-ageing from two indonesian green algae. J Funct Foods 104:105555
Nutho B, Tungmunnithum D (2023) Exploring major flavonoid phytochemicals from nelumbo nucifera gaertn. as potential skin anti-aging agents: in silico and in vitro evaluations. Int J Mol Sci 24(23):16571
Nutho B, Tungmunnithum D (2024) Anti-aging potential of the two major flavonoids occurring in Asian water lily using in vitro and in silico molecular modeling assessments. Antioxidants 13(5):601
OECD (2004) Report from the expert group on (quantitative) structure-activity relationships [(q)sars] on the principles for the validation of (q)sars. OECD Series on Testing and Assessment, No. 49, OECD Publishing, Paris. https://one.oecd.org/document/env/jm/mono(2004)24/en/pdf
OECD (2014) Guidance on grouping of chemicals. OECD Series on Testing and Assessment, No. 80, OECD Publishing, Paris. https://doi.org/10.1787/9789264085831-en
OECD (2019) Test no. 432: In vitro 3t3 nru phototoxicity test. Tech. rep., Paris. https://doi.org/10.1787/9789264071162-en
Onoue S, Suzuki G, Kato M et al (2013) Non-animal photosafety assessment approaches for cosmetics based on the photochemical and photobiochemical properties. Toxicol In Vitro 27(8):2316–2324
Onyango PO (2023) The cost of beauty: perspectives of salon workers in Kisumu city, Kenya. PLOS Global Public Health 3(11):e0002503
Ouedraogo G, Alexander-White C, Bury D et al (2022) Read-across and new approach methodologies applied in a 10-step framework for cosmetics safety assessment-a case study with parabens. Regul Toxicol Pharmacol 132:105161
Page MJ, McKenzie JE, Bossuyt PM et al (2021) The prisma 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg 88:105906
Paini A, Mennecozzi M, Horvat T et al (2017) Practical use of the virtual cell based assay: simulation of repeated exposure experiments in liver cell lines. Toxicol In Vitro 45:233–240
Papadopoulos D, Karalis VD (2023) Variational autoencoders for data augmentation in clinical studies. Appl Sci 13(15):8793
Park J, Lee H, Park K (2018) Mixture toxicity of methylisothiazolinone and propylene glycol at a maximum concentration for personal care products. Toxicol Res 34:355–361

Parveen N, Akbarsha MA, Wani AL et al (2023) Protective effect of quercetin and thymoquinone against genotoxicity and oxidative stress induced by zno nanoparticles in the wistar rat model. Mutation research/genetic toxicology and environmental mutagenesis 890:503661

Patlewicz G, Ball N, Booth ED et al (2013) Use of category approaches, read-across and (q) sar: general considerations. Regul Toxicol Pharmacol 67(1):1–12

Pažoureková S, Hojerová J, Klimová Z et al (2013) Dermal absorption and hydrolysis of methylparaben in different vehicles through intact and damaged skin: using a pig-ear model in vitro. Food Chem Toxicol 59:754–765

da Penha JR, da Silva ACG, de Ávila RI et al (2022) Development of a novel ex vivo model for chemical ocular toxicity assessment and its applicability for hair straightening products. Food Chem Toxicol 170:113457

Petit JY, Doré V, Marignac G et al (2017) Assessment of ocular discomfort caused by 5 shampoos using the slug mucosal irritation test. Toxicol In Vitro 40:243–247

Petry T, Bury D, Fautz R et al (2017) Review of data on the dermal penetration of mineral oils and waxes used in cosmetic applications. Toxicol Lett 280:70–78

Pfannenbecker U, Bessou-Touya S, Faller C et al (2013) Cosmetics europe multi-laboratory pre-validation of the epiocular™ reconstituted human tissue test method for the prediction of eye irritation. Toxicol In Vitro 27(2):619–626

Pop A, Drugan T, Gutleb AC et al (2016) Individual and combined in vitro (anti) androgenic effects of certain food additives and cosmetic preservatives. Toxicol In Vitro 32:269–277

Popiół J, Piska K, Słoczyńska K et al (2019) Microbial biotransformation of some novel hydantoin derivatives: perspectives for bioremediation of potential sunscreen agents. Chemosphere 234:108–115

Potts RO, Guy RH (1992) Predicting skin permeability. Pharm Res 9:663–669

Qian J, Fl Song, Liang R et al (2022) Predictive and explanatory themes of noael through a systematic comparison of different machine learning methods and descriptors. Food Chem Toxicol 168:113325

Qin H, Zhang J, Yang H et al (2020) Safety assessment of water-extract sericin from silkworm (Bombyx mori) cocoons using different model approaches. Biomed Res Int 2020:9689386

Raak C, Molsberger F, Pittermann W et al (2017) Use of the bovine udder skin model to evaluate the tolerability of mesem cosmetic cream. Altern Lab Anim 45(4):191–200

Raies AB, Bajic VB (2016) In silico toxicology: computational methods for the prediction of chemical toxicity. Wiley Interdiscip Rev: Comput Mol Sci 6(2):147–172

Raja IS, Lee JH, Hong SW et al (2021) A critical review on genotoxicity potential of low dimensional nanomaterials. J Hazard Mater 409:124915

Rajnochová Svobodová A, Ryšavá A, Psotová M et al (2017) The phototoxic potential of the flavonoids, taxifolin and quercetin. Photochem Photobiol 93(5):1240–1247

Reeves KW, Díaz Santana M, Manson JE, et al (2019) Urinary phthalate biomarker concentrations and postmenopausal breast cancer risk. JNCI: J Natl Cancer Inst 111(10):1059–1067

Reisinger K, Blatz V, Brinkmann J et al (2018) Validation of the 3d skin comet assay using full thickness skin models: Transferability and reproducibility. Mutati Res/Genet Toxicol Environ Mutagen 827:27–41

Reynolds G, Reynolds J, Gilmour N et al (2021) A hypothetical skin sensitisation next generation risk assessment for coumarin in cosmetic products. Regul Toxicol Pharmacol 127:105075

Ribeiro MT, Singh S, Guestrin C (2016) Model-agnostic interpretability of machine learning. Proceedings of the 2016 ICML Workshop on Human Interpretability in Machine Learning. arXiv:1606.05386

Rioux B, Mouterde LM, Alarcan J et al (2023) An expeditive and green chemo-enzymatic route to diester sinapoyl-l-malate analogues: sustainable bioinspired and biosourced uv filters and molecular heaters. Chem Sci 14(47):13962–13978

Roncaglioni A, Toropov AA, Toropova AP et al (2013) In silico methods to predict drug toxicity. Curr Opin Pharmacol 13(5):802–806

Rorije E, Aldenberg T, Buist H et al (2013) The osiris weight of evidence approach: Its for skin sensitisation. Regul Toxicol Pharmacol 67(2):146–156

Saewan N, Jimtaisong A, Panyachariwat N et al (2023) In vitro and in vivo anti-aging effect of coffee berry nanoliposomes. Molecules 28(19):6830

Sahu SC, Roy S, Zheng J et al (2014) Comparative genotoxicity of nanosilver in human liver hepg2 and colon caco2 cells evaluated by fluorescent microscopy of cytochalasin b-blocked micronucleus formation. J Appl Toxicol 34(11):1200–1208

Sahu SC, Zheng J, Yourick JJ et al (2015) Toxicogenomic responses of human liver hepg2 cells to silver nanoparticles. J Appl Toxicol 35(10):1160–1168
Sahu SC, Njoroge J, Bryce SM et al (2016) Flow cytometric evaluation of the contribution of ionic silver to genotoxic potential of nanosilver in human liver hepg2 and colon caco2 cells. J Appl Toxicol 36(4):521–531
Sahu SC, Roy S, Zheng J et al (2016) Contribution of ionic silver to genotoxic potential of nanosilver in human liver hepg2 and colon caco2 cells evaluated by the cytokinesis-block micronucleus assay. J Appl Toxicol 36(4):532–542
Sarath Kumar KL, Tangadpalliwar SR, Desai A et al (2016) Integrated computational solution for predicting skin sensitization potential of molecules. PLoS ONE 11(6):e0155419
Savic SM, Cekic ND, Savic SR et al (2021) 'all-natural'anti-wrinkle emulsion serum with acmella oleracea extract: a design of experiments (doe) formulation approach, rheology and in vivo skin performance/efficacy evaluation. Int J Cosmet Sci 43(5):530–546
Schultz T, Amcoff P, Berggren E et al (2015) A strategy for structuring and reporting a read-across prediction of toxicity. Regul Toxicol Pharmacol 72(3):586–601
Selvestrel G, Robino F, Baderna D et al (2021) Spheracosmolife: a new tool for the risk assessment of cosmetic products. ALTEX 38(4):565–579
Selvestrel G, Lavado GJ, Toropova AP et al (2022) Monte carlo models for sub-chronic repeated-dose toxicity: systemic and organ-specific toxicity. Int J Mol Sci 23(12):6615
Selvestrel G, Robino F, Russo MZ (2022) In silico models for skin sensitization and irritation. Methods Mol Biol 2425:291–354
Senapati VA, Kumar A, Gupta GS et al (2015) Zno nanoparticles induced inflammatory response and genotoxicity in human blood cells: A mechanistic approach. Food Chem Toxicol 85:61–70
Shah I, Liu J, Judson RS et al (2016) Systematically evaluating read-across prediction and performance using a local validity approach characterized by chemical structure and bioactivity information. Regul Toxicol Pharmacol 79:12–24
Shyr T, Ou-Yang H (2016) Sunscreen formulations may serve as additional water barrier on skin surface: a clinical assessment. Int J Cosmet Sci 38(2):164–169
Šidlovská M, Petrovičová I, Kolena B et al (2017) Exposure of children to phthalates and the impact of consumer practices in Slovakia. Rev Environ Health 32(1–2):211–214
da Silva AC, Paiva JP, Diniz RR et al (2019) Photoprotection assessment of olive (olea europaea l.) leaves extract standardized to oleuropein: In vitro and in silico approach for improved sunscreens. J Photochem Photobiol B 193:162–171
Silva AC, Borba JV, Alves VM et al (2021) Novel computational models offer alternatives to animal testing for assessing eye irritation and corrosion potential of chemicals. Artif Intell Life Sci 1:100028
da Silva ACG, Chialchia AR, de Ávila RI et al (2018) Mechanistic-based non-animal assessment of eye toxicity: inflammatory profile of human keratinocytes cells after exposure to eye damage/irritant agents. Chem Biol Interact 292:1–8
Silva RJ, Tamburic S (2022) A state-of-the-art review on the alternatives to animal testing for the safety assessment of cosmetics. Cosmetics 9(5):90
Silveira J, Pereda M, Nogueira C et al (2016) Preliminary safety assessment of c-8 xylitol monoester and xylitol phosphate esters. Int J Cosmet Sci 38(1):41–51
Sipahi H, Orak D, Reis R et al (2022) A comprehensive study to evaluate the wound healing potential of okra (abelmoschus esculentus) fruit. J Ethnopharmacol 287:114843
Skare JA, Blackburn K, Wu S et al (2015) Use of read-across and computer-based predictive analysis for the safety assessment of peg cocamines. Regul Toxicol Pharmacol 71(3):515–528
Słoczyńska K, Popiół J, Gunia-Krzyżak A, et al (2022) Evaluation of two novel hydantoin derivatives using reconstructed human skin model episkintm: perspectives for application as potential sunscreen agents. Molecules 27(6):1850
Smeriglio A, Lionti J, Ingegneri M et al (2023) Xanthophyll-rich extract of phaeodactylum tricornutum bohlin as new photoprotective cosmeceutical agent: Safety and efficacy assessment on in vitro reconstructed human epidermis model. Molecules 28(10):4190
Sommer E, Neubert RH, Mentel M et al (2018) Dermal peptide delivery using enhancer molecules and colloidal carrier systems. part iii: Tetrapeptide gekg. Eur J Pharm Sci 124:137–144
Sreedhar D, Manjula N, Pise A et al (2020) Ban of cosmetic testing on animals: a brief overview. Int J Curr Res Rev 12(14):113
Stouch TR, Kenyon JR, Johnson SR et al (2003) In silico adme/tox: Why models fail. J Comput Aided Mol Des 17:83–92

Strickland J, Allen DG, Germolec D et al (2022) Application of defined approaches to assess skin sensitization potency of isothiazolinone compounds. Appl In Vitro Toxicol 8(4):117–128
Su Y, Xu S, Hu X et al (2024) Rapid discovery of natural skin-lightening ingredients based on an integrated screening strategy based on molecular docking and zebrafish model. J Cosmet Dermatol 23(11):3724–3734
Suriyaprabha R, Balu K, Karthik S et al (2019) A sensitive refining of in vitro and in vivo toxicological behavior of green synthesized zno nanoparticles from the shells of jatropha curcas for multifunctional biomaterials development. Ecotoxicol Environ Saf 184:109621
Ta GH, Weng CF, Leong MK (2021) In silico prediction of skin sensitization: Quo vadis? Front Pharmacol 12:655771
Tang Y, Cai R, Cao D et al (2018) Photocatalytic production of hydroxyl radicals by commercial tio2 nanoparticles and phototoxic hazard identification. Toxicology 406:1–8
Tarnowska M, Briançon S, Resende de Azevedo J et al (2020) The effect of vehicle on skin absorption of mg2+ and ca2+ from thermal spring water. Int J Cosmet Sci 42(3):248–258
Togo MV, Mastrolorito F, Ciriaco F et al (2022) Tiresia: an explainable artificial intelligence platform for predicting developmental toxicity. J Chem Inf Model 63(1):56–66
Toropov AA, Toropova AP (2021) The unreliability of the reliability criteria in the estimation of qsar for skin sensitivity: a pun or a reliable law? Toxicol Lett 340:133–140
Toropov AA, Toropova AP, Marzo M et al (2021) Pesticides, cosmetics, drugs: identical and opposite influences of various molecular features as measures of endpoints similarity and dissimilarity. Mol Diversity 25:1137–1144
Tourneix F, Alépée N, Detroyer A et al (2019) Assessment of a defined approach based on a stacking prediction model to identify skin sensitization hazard. Toxicol In Vitro 60:134–143
Tourneix F, Alépée N, Detroyer A et al (2020) Skin sensitisation testing in practice: applying a stacking meta model to cosmetic ingredients. Toxicol In Vitro 66:104831
Turkez H, Yıldırım S, Sahin E et al (2022) Boron compounds exhibit protective effects against aluminum-induced neurotoxicity and genotoxicity: In vitro and in vivo study. Toxics 10(8):428
Valerio LG Jr (2009) In silico toxicology for the pharmaceutical sciences. Toxicol Appl Pharmacol 241(3):356–370
Vandecasteele HA, Gautier F, Tourneix F et al (2021) Next generation risk assessment for skin sensitisation: a case study with propyl paraben. Regul Toxicol Pharmacol 123:104936
Varet J, Barranger A, Crochet C et al (2024) New methodological developments for testing the in vitro genotoxicity of nanomaterials: comparison of 2d and 3d heparg liver cell models and classical and high throughput comet assay formats. Chemosphere 350:140975
Venkatapathy R, Wang NCY (2013) Developmental toxicity prediction. Comput Toxicol II:305–340
Verma RP, Matthews EJ (2015) Estimation of the chemical-induced eye injury using a weight-of-evidence (woe) battery of 21 artificial neural network (ann) c-qsar models (qsar-21): Part i: Irritation potential. Regul Toxicol Pharmacol 71(2):318–330
Verma RP, Matthews EJ (2015) Estimation of the chemical-induced eye injury using a weight-of-evidence (woe) battery of 21 artificial neural network (ann) c-qsar models (qsar-21): part ii: corrosion potential. Regul Toxicol Pharmacol 71(2):318–330
Vila L, García-Rodríguez A, Marcos R et al (2018) Titanium dioxide nanoparticles translocate through differentiated caco-2 cell monolayers, without disrupting the barrier functionality or inducing genotoxic damage. J Appl Toxicol 38(9):1195–1205
Walters RM, Gandolfi L, Mack MC et al (2016) In vitro assessment of skin irritation potential of surfactant-based formulations by using a 3-d skin reconstructed tissue model and cytokine response. Altern Lab Anim 44(6):523–532
Wang B, Tian L, Tian L et al (2023) Insights into health risks of face paint application to opera performers: the release of heavy metals and stage-light-induced production of reactive oxygen species. Environ Sci Technol 57(9):3703–3712
Wang YF, Chen IW, Subendran S et al (2020) Edible additive effects on zebrafish cardiovascular functionality with hydrodynamic assessment. Sci Rep 10(1):16243
Wassermann AM, Bajorath J (2010) Chemical substitutions that introduce activity cliffs across different compound classes and biological targets. J Chem Inf Model 50(7):1248–1256
Weaver S, Gleeson MP (2008) The importance of the domain of applicability in qsar modeling. J Mol Graph Model 26(8):1315–1326
Wei Z, Liu X, Ooka M et al (2020) Two-dimensional cellular and three-dimensional bio-printed skin models to screen topical-use compounds for irritation potential. Front Bioeng Biotechnol 8:109

Willhite CC, Karyakina NA, Yokel RA et al (2014) Systematic review of potential health risks posed by pharmaceutical, occupational and consumer exposures to metallic and nanoscale aluminum, aluminum oxides, aluminum hydroxide and its soluble salts. Crit Rev Toxicol 44(sup4):1–80
Williams FM, Rothe H, Barrett G 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
Wilm A, Kühnl J, Kirchmair J (2018) Computational approaches for skin sensitization prediction. Crit Rev Toxicol 48(9):738–760
Wilm A, Stork C, Bauer C et al (2019) Skin doctor: machine learning models for skin sensitization prediction that provide estimates and indicators of prediction reliability. Int J Mol Sci 20(19):4833
Worth A, Lapenna S, Lo Piparo E, et al (2011) A framework for assessing in silico toxicity predictions: Case studies with selected pesticides. Technical Report EUR 24705 EN, Publications Office of the European Union, Luxembour. https://doi.org/10.2788/30449
Xiong L, Tang J, Li Y et al (2019) Phototoxic risk assessment on benzophenone uv filters: In vitro assessment and a theoretical model. Toxicol In Vitro 60:180–186
Yang C, Cronin M, Arvidson K et al (2021) Cosmos next generation-a public knowledge base leveraging chemical and biological data to support the regulatory assessment of chemicals. Comput Toxicol 19:100175
Yang YT, Ni HG (2023) Predictive in silico models for aquatic toxicity of cosmetic and personal care additive mixtures. Water Res 236:119981
Yuki K, Ikeda N, Nishiyama N et al (2013) The reconstructed skin micronucleus assay in epiderm ™: reduction of false-positive results-a mechanistic study with epigallocatechin gallate. Mutat Res/Genet Toxicol Environ Mutagen 757(2):148–157
Zerbinati N, Sommatis S, Maccario C et al (2021) A practical approach for the in vitro safety and efficacy assessment of an anti-ageing cosmetic cream enriched with functional compounds. Molecules 26(24):7592
Zheng C, Cao T, Ye C et al (2023) Neutrophil recruitment by cd4 tissue-resident memory t cells induces chronic recurrent inflammation in atopic dermatitis. Clin Immunol 256:109805
Zhong C, Wanasathop A, Shi Z et al (2022) Evaluation of in vitro cornea models for quantifying destructive effects of chemicals. Toxicol In Vitro 85:105462
Ziemlewska A, Wójciak M, Mroziak-Lal K et al (2022) Assessment of cosmetic properties and safety of use of model washing gels with Reishi, maitake and lion's mane extracts. Molecules 27(16):5090
Zijno A, De Angelis I, De Berardis B et al (2015) Different mechanisms are involved in oxidative dna damage and genotoxicity induction by zno and tio2 nanoparticles in human colon carcinoma cells. Toxicol In Vitro 29(7):1503–1512
Zlabiene U, Baranauskaite J, Kopustinskiene DM, et al (2021) In vitro and clinical safety assessment of the multiple w/o/w emulsion based on the active ingredients from rosmarinus officinalis l., avena sativa l. and linum usitatissimum l. Pharmaceutics 13(5):732

Authors and Affiliations

Tamara G. Vasiljev[1] · Lucia Salvioni[2] · Miriam Colombo[2] · Paolo Galli[3] · Francesca Greselin[4]

✉ Tamara G. Vasiljev
t.vasiljev@campus.unimib.it

Lucia Salvioni
lucia.salvioni@unimib.it

Miriam Colombo
miriam.colombo@unimib.it

Paolo Galli
paolo.galli@unimib.it

Francesca Greselin
francesca.greselin@unimib.it

[1] Department of Economics, Management and Statistics, University of Milano-Bicocca, Milan, Italy

[2] Department of Biotechnology and Bioscience, University of Milano-Bicocca, Milan, Italy

[3] Department of Earth and Environmental Sciences, University of Milano-Bicocca, Milan, Italy

[4] Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy