Achar, J, Firman, JW, Tran, C, Kim, D, Cronin, MTD and Öberg, G

Analysis of implicit and explicit uncertainties in QSAR prediction of chemical toxicity: A case study of neurotoxicity

Regulatory Toxicology and Pharmacology 154 (2024) 105716

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

Analysis of implicit and explicit uncertainties in QSAR prediction of chemical toxicity: A case study of neurotoxicity

Jerry Achar [a,*], James W. Firman [b], Chantelle Tran [c], Daniella Kim [d], Mark T.D. Cronin [b], Gunilla Öberg [a]

[a] Institute for Resources Environment, and Sustainability, The University of British Columbia, 2202 Main Mall, Vancouver, BC, V6T 1Z4, Canada
[b] School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, UK
[c] Department of Microbiology and Immunology, The University of British Columbia, 2350 Health Sciences Mall, Vancouver, BC, V6T 1Z4, Canada
[d] Department of Earth, Ocean, and Atmospheric Sciences, The University of British Columbia, 2207 Main Mall, Vancouver, BC, V6T 1Z4, Canada

ARTICLE INFO

Handling Editor: Dr. Martin Van den berg

Keywords:
Quantitative structure-activity relationship
QSAR
Uncertainty
Toxicity
Implicit uncertainty
Explicit uncertainty
Neurotoxicity

ABSTRACT

Although uncertainties expressed in texts within QSAR studies can guide quantitative uncertainty estimations, they are often overlooked during uncertainty analysis. Using neurotoxicity as an example, this study developed a method to support analysis of implicitly and explicitly expressed uncertainties in QSAR modeling studies. Text content analysis was employed to identify implicit and explicit uncertainty indicators, whereafter uncertainties within the indicator-containing sentences were identified and systematically categorized according to 20 uncertainty sources. Our results show that implicit uncertainty was more frequent within most uncertainty sources (13/20), while explicit uncertainty was more frequent in only three sources, indicating that uncertainty is predominantly expressed implicitly in the field. The most highly cited sources included Mechanistic plausibility, Model relevance and Model performance, suggesting they constitute sources of most concern. The fact that other sources like Data balance were not mentioned, although it is recognized in the broader QSAR literature as an area of concern, demonstrates that the output from the type of analysis conducted here must be interpreted in the context of the broader QSAR literature before conclusions are drawn. Overall, the method established here can be applied in other QSAR modeling contexts and ultimately guide efforts targeted towards addressing the identified uncertainty sources.

1. Introduction

Quantitative structure-activity relationships (QSARs) are an *in silico* toxicology approach that aims to establish relationships between descriptors of chemical structure and biological activities (e.g., toxicity) or properties. The implicit assumption is that structurally similar chemicals should have similar activities/properties and the trends can be identified and modeled within groups of molecules (Cronin and Madden, 2010). QSARs have the potential to support the reduction in the use of animal testing in different assessment contexts aimed toward characterizing and/or predicting chemical toxicity (Belfield et al., 2021; Cronin et al., 2019b; Patlewicz et al., 2013). There are increasing calls to incorporate QSAR predictions in the assessment of chemical toxicity (e.g., within the European Food Safety Authority (EFSA), 2010), Health Canada (2023), and the US National Research Council (2007)). However, it is recognized that it will be difficult to address complex endpoints with QSAR alone. An example is the prediction of neurotoxicity, given the unreliability of animal models in assessing this endpoint for reasons such as interspecies differences in brain morphology or differences in biological functions between humans and animals (EFSA, 2010; EFSA et al., 2021; Fritsche et al., 2018). Worth et al. (2011a,b) state that no single QSAR model (or in combination) seems adequate to predict the neurotoxic potentials of chemical compounds. The shortcomings of QSAR models have led to increasing demands to analyze and communicate uncertainties in QSAR models and model predictions of complex endpoints such as neurotoxicity to support efforts aimed at addressing the uncertainties (Belfield et al., 2023; Cronin et al., 2019a,b; Piir et al., 2018; Sahlin et al., 2011; Schultz et al., 2019; Vighi et al., 2019).

Researchers express uncertainties in various ways, including the use of words that implicitly or explicitly qualify or represent the author's confidence in the content of the information communicated or alter precision implied in a measured numerical value (Flari and Wilkinson,

* Corresponding author.
E-mail address: jerry.achar@ubc.ca (J. Achar).

https://doi.org/10.1016/j.yrtph.2024.105716
Received 27 July 2024; Received in revised form 24 September 2024; Accepted 8 October 2024
Available online 10 October 2024

2011; Levin et al., 2004). As defined in Table S1, explicit uncertainties are expressed directly as gaps, unknowns, or quantitative confidence measures regarding, for example, model predictivity (e.g., "it is uncertain" and "more data are needed") and are consequently easily detected (Flari and Wilkinson, 2011; Sahlin et al., 2011). In contrast, implicit uncertainties are expressed indirectly in a subtle manner (or unintentionally) – e.g., through words such as "probably", "maybe", and "might". While understanding the implicitly and explicitly expressed uncertainties can provide valuable nuance and guide quantitative uncertainty estimations, these uncertainties are not easily discerned and can, therefore, be easily overlooked during analysis or interpretation of QSAR predictions (Flari and Wilkinson, 2011; Levin et al., 2004; Zerva, 2019). In other words, at present, it is difficult to elucidate whether modelers use linguistic expressions to indicate how certain they regard the current state of knowledge of QSAR models and the information regarding, for example, input data, parameters, and prediction outputs (Flari and Wilkinson, 2011). To support systematic and transparent accounting of uncertainties in QSAR predictions, we here develop a method that makes it possible to identify and, based on our previously developed framework (Achar et al., 2024a), systematically categorize implicit and explicit uncertainties expressed in texts in studies applying QSARs for chemical toxicity predictions.

Research on uncertainty indicators and how they can influence the perceived certainty of information in written statements exists across different research areas. For example, Markkanen and Schröder (1997), Varttala (2001), and Zerva (2019) explore the use and interpretation of uncertainty expressions that qualify confidence in statements in linguistic and behavioral studies, Stortenbeker et al. (2019) analyze the influence of the use of uncertainty expressions by doctors on patient anxiety during doctor-patient communication, while Ferson et al. (2015), Rubin (2007), and Zerva (2019) explore the use of machine learning to identify probability phrases and numerical uncertainty expressions. Scholars such as Flari and Wilkinson (2011) and Levin et al. (2004) highlight the utility of implicit and explicit uncertainty indicators as markers when identifying uncertain information in scientific texts or written statements. In this study, we follow what is held to be best practice when identifying uncertainty communicated within a statement, which is to first identify uncertainty indicators and then analyze the uncertainty in information in a statement based on how the indicators relate to and affect the conveyed information (Flari and Wilkinson, 2011; Hillen et al., 2017; Levin et al., 2004; Stortenbeker et al., 2019; Zerva, 2019).

Uncertainty estimation in QSAR modeling falls into two major categories: aleatoric uncertainty and epistemic uncertainty (see the definitions in Table S1). While the former cannot be eliminated through additional data, the former can (at least in theory) be eliminated through additional data or knowledge (Gajewicz et al., 2015; Scalia et al., 2020; Wang et al., 2021; Zhong et al., 2022). In this study, we focus on epistemic uncertainty, as it is regarded to be more problematic in QSAR modeling exercises – e.g., with respect to understanding whether a model is fit-for-purpose, availability of relevant and reliable data to provide more insights into chemical structure-activity relationships, or interpretation of the mechanism of action of chemical compounds (Cronin et al., 2019a,b; Sahlin et al., 2013). Methods for analyzing epistemic uncertainties in chemical risk assessment fall into three broad tiers: qualitative, deterministic, and probabilistic methods, for which qualitative analysis of uncertainties is considered the first step in any uncertainty analysis exercise (EFSA, 2006; Sahlin et al., 2013; World Health Organization and International Programme on Chemical Safety (WHO/IPCS), 2018).

While qualitative analysis of epistemic uncertainty is an important first step in uncertainty analysis, QSAR studies predicting the toxicity of chemical compounds have hitherto focused on quantitative aspects. For example, in QSAR mutagenicity prediction, Hung and Gini (2021) applied Bayesian reasoning to quantify epistemic uncertainty in model input data – i.e., uncertainty related to the amount of data used for modeling and the availability of specific chemical information in the modeling data. Zhong et al. (2022), during QSAR development, applied the Gaussian process to quantify epistemic uncertainty with respect to the inclusion or exclusion of specific chemicals in the model training set. Similarly, Wang et al. (2021) and Zhang and Lee (2019) estimated epistemic uncertainty using Bayesian statistics – i.e., distributional uncertainty emanating from QSAR models' lack of recognition of the information in test sets, and uncertainty due to sparse or imbalanced distribution of data in model training sets. However, to our knowledge, little (if anything) has been done to qualitatively identify or categorize uncertainties that are expressed in statements in QSAR modeling studies. This is problematic as systematic and transparent accounting of uncertainties in QSAR modeling requires analyzing and addressing both quantitative and qualitative uncertainties (EFSA, 2006; Sahlin et al., 2013; WHO/IPCS, 2018).

Using the neurotoxicity endpoint as an example, the aim of this study was to develop a method that allows for systematic and transparent accounting for implicit and explicit uncertainties in QSAR modeling of chemical toxicity. The choice for neurotoxicity was based on the fact that, similar to other complex toxicological endpoints, it suffers from limited experimental data, which leads to higher epistemic uncertainty (Madden et al., 2020; Worth et al., 2011a). It is also well recognized that, presently, models such as QSARs struggle to accurately predict neurotoxicity given the complex nature of the underlying biological mechanisms (Bal-Price et al., 2018; Fritsche et al., 2018; Gadaleta et al., 2022; Madden et al., 2020; Worth et al., 2011a); thus, making uncertainty analysis important for QSAR modeling of neurotoxicity. Accordingly, implicit and explicit uncertainty indicators, expressed in peer-reviewed papers on the QSAR modeling of neurotoxicity, were first identified. The indicators were then used to identify implicit and explicit uncertainties. The identified uncertainties were then systematically categorized according to the uncertainty sources proposed by Achar et al. (2024a). This allowed us to identify uncertainty sources that were most commonly highlighted by researchers. By identifying and categorizing the implicit and explicit uncertainties, we contend that this method can be used to draw attention to epistemic uncertainties in QSAR modeling of specified endpoints, here illustrated by neurotoxicity. Our hope is that the information gained from our study can be used to inform decision-support initiatives by modelers and regulatory authorities when identifying research needs and the type of data required to reduce or eliminate uncertainties.

2. Methodology

2.1. Uncertainty indicators

Levin et al. (2004) proposed four categories of implicit uncertainty indicators in chemical risk assessment: epistemic, inferential, contentual and conditionalizing implicit uncertainty indicators (definitions of the indicators are provided in Fig. S1). We considered the concepts described for the 'implicit epistemic uncertainty indicators' to be directly relevant to our study, as our aim is to identify epistemic uncertainties. However, our initial analysis suggested that researchers commonly do not distinguish between epistemic uncertainty (e.g., "it is presumed that …") and inferential uncertainty indicators (e.g., "on this basis, it is presumed that …"). We therefore decided to also include inferential indicators as part of epistemic indicators (a detailed explanation and definition is given in Table S1, Text S1 and Fig. S1 of the Supplementary material). Furthermore, the concept of explicit epistemic uncertainty indicators (e.g., qualitative or quantitative statements such as "we don't know" and it is uncertain") described by Sahlin et al. (2011) and (Stortenbeker et al., 2019) was adopted to identify explicit uncertainty indicators (see detailed explanation in Text S1 of the Supplementary material).

Fig. 1 outlines the process used in the present study to identify and categorize implicit and explicit uncertainties. This started with the

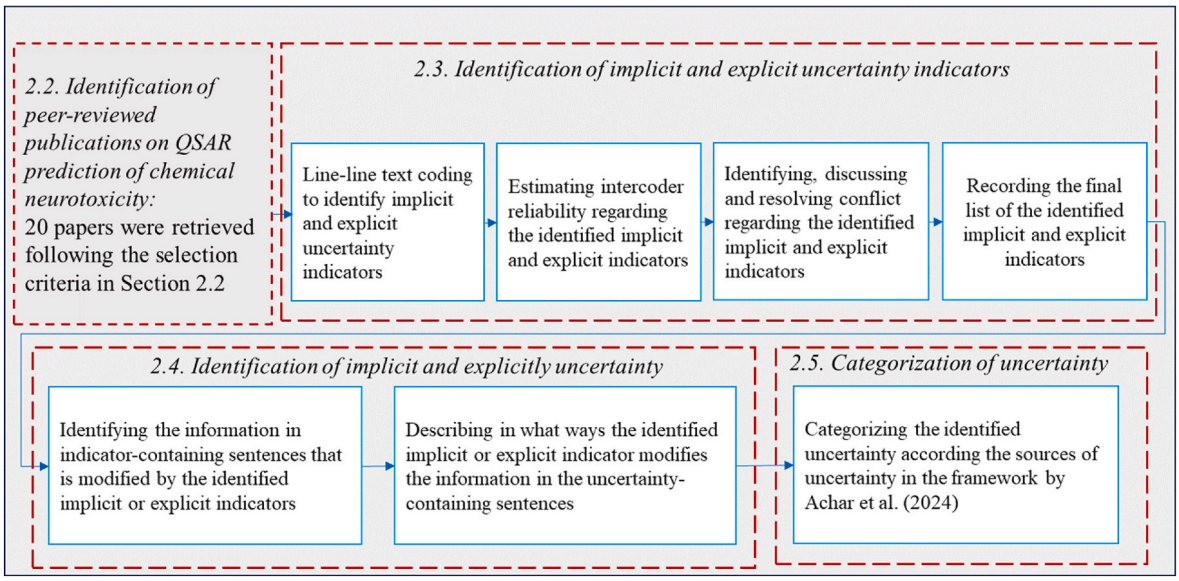

Fig. 1. The steps undertaken to identify and categorize implicit and explicit uncertainties in peer-reviewed papers that apply QSAR to predict the neurotoxicity of chemical compounds.

sourcing of peer-reviewed papers (section 2.2), followed by the identification of epistemic explicit and implicit uncertainty indicators (section 2.3), whereafter these indicators were used to highlight uncertainty in QSAR neurotoxicity papers (section 2.4). Finally, with the goal of supporting systematic and transparent accounting of uncertainties, we used the uncertainty sources we recently proposed (Achar et al., 2024a) (see Section 2.5) to categorize the identified implicit and explicit uncertainties.

2.2. Selecting peer-reviewed papers for analysis

A search was conducted in the Web of Science Clarivate database using the following keywords and Booleans: neurotox* (TOPIC) AND QSAR (TOPIC), while excluding review papers. This led to the identification of 75 papers. Duplicates were excluded whereafter titles and abstracts were skimmed, and the following inclusion criteria were used: the paper had to (1) be published in a peer-reviewed journal, (2) be an original research article, (3) apply QSAR model(s), (4) address neurotoxicity assessment of a chemical (including drugs), and (5) be in the English language. After reading the papers in their entirety, 20 papers that met the selection criteria were selected for analysis (see the Supplementary material for the list of 20 papers). These articles were added to the Zotero reference management tool (version 6.0.36).

2.3. Identification of implicit and explicit uncertainty indicators

Implicit epistemic uncertainty indicators were identified through line-by-line text coding of the methods, results, discussion, and conclusion sections of the papers identified under section 2.2. These sections were selected as they are the locations in a paper where one would expect authors to describe the model(s), modeling data, parameters and variables applied in their study, report and discuss the study findings/modeling output, and draw conclusions about the study. Three coders (first, third and fourth authors) independently coded each of the 20 papers (Table S2), identifying, color-marking, and recording implicit and explicit indicators. Each coder reviewed each paper at least twice until no new indicators could be discovered. Thereafter, the intercoder agreement measure, calculated as the percent agreement, was estimated using Krippendorff's Alpha (α_k) (Krippendorff, 2004) on the web-based K-Alpha Calculator developed by Marzi et al. (2024). The measure relates to whether the coders agreed if the coded texts were indicators and, if yes, whether the indicators qualified to be categorized as explicit or implicit epistemic indicators. All conflicting issues were identified, discussed, and resolved before the final list of indicators was recorded (Table S2). We describe the uncertainty indicator identification procedure in more detail in the Supplementary materials (Text S2). The risk of double coding was reduced by checking whether an indicator-containing sentence was repeated in different sections of a paper (e.g., identical phrases could be used in the Results and Discussion and the Conclusion sections).

2.4. Identification of implicit and explicit uncertainties

The indicator-containing sentences and the context in which the indicators appear were used to interpret and describe the implicit or explicit uncertainties communicated in these sentences; this was collaboratively performed by the first and second authors. Inspired by the process proposed by Zerva (2019), we started by first noting the location of uncertainty indicators in the indicator-containing sentences, followed by identifying the central piece of information communicated in the sentences. This information was identified by reading each paper in its entirety to get a general idea about the study, and then each indicator-containing sentence and, when necessary, an entire paragraph containing the indicator-containing sentence. Next, we interpreted how each indicator modified the central piece of information. The assumption is that if the central piece of information is modified by the indicator, then the information becomes uncertain (Zerva, 2019). An example of how the descriptions of the uncertainties were developed is illustrated in Table 1 (Step 1–3) using two of the indicator-containing sentences from two of the 20 studies (i.e., Estrada et al., 2001; Zhang and Lee, 2019).

2.5. Categorization of the identified uncertainties

This study systematically categorized each of the uncertainties identified under Section 2.4 guided by the 20 uncertainty sources (Table 2) established within a framework we recently developed to systematically categorize general sources of uncertainty across different *in silico* toxicology methods (Achar et al., 2024a; in press). The framework conceptualizes uncertainty as a multi-source phenomenon that is associated with recognized QSAR components and modeling processes (see Fig. S2); thus making it a valuable tool in this study to facilitate

Table 1
Steps followed to identify uncertainties implicitly and explicitly expressed in the indicator-containing sentences.

	Implicit uncertainty	Explicit uncertainty
Indicator-containing sentence	"It demonstrates that a smaller molecule is more likely to cause adverse effect to the nervous system" (Zhang and Lee, 2019; p 4).	"Unfortunately, there is not enough experimental data to corroborate these findings" (Estrada et al., 2001; p 457).
Step 1: Noting the indicator	"more likely"	"there is not enough experimental data"
Step 2: Identifying the information communicated	In this context, smaller molecules are more likely to cause adverse effects to the nervous systems than larger molecules	A group of chemicals were found to induce neurotoxicity in the test animal (i.e., mouse)
Step 3: Interpreting how the indicator modifies the information	The indicator modifies information by implicitly clarifying that it cannot be guaranteed that smaller molecules will cause more adverse effects than larger molecules.	The indicator modifies information by explicitly clarifying that "there is not sufficient experimental data to corroborate" that the studied chemicals will induce neurotoxicity in mice.

mapping out of diverse sources of uncertainty in QSAR modeling exercises. In applying the framework, however, we note that Data accuracy was considered not just to entail the measure of correctness of data in relation to a "true value" (as defined in the framework – Table 2) but also such measure in relation a "distribution of true values". This adjustment was necessary to accommodate the different (implicit or explicit) descriptions of data accuracy in the 20 analyzed studies. For example, Turabekova et al. (2008), in "The bad fitting to correlation line for those compounds can be explained by possible errors […]" (Table S2), implicitly mention data accuracy with respect to "distribution of true values". In contrast, Cronin (1996) in "[…] most confidence intervals are between 10% and 30% of the original value" (Table S2) explicitly mention it with respect to a "true value".

2.5.1. *The categorization process*

We illustrate how the categorization was performed by referring to the examples in Table 1. For the implicit uncertainty category, the interpretation of uncertainty was not just based on the understanding of the message communicated in the indicator-containing sentence but the overall message in the paragraph containing the sentence or adjacent paragraphs as well. Accordingly, we interpreted the uncertainty (Step 3; Table 1) as: the effect of smaller molecules, relative to larger molecules, is uncertain (where molecular size is characterized using molecular weight descriptor), and (2) there is also uncertainty in knowledge about mechanistic interaction of the smaller molecule (relative to larger molecules) in the nervous system to inform judgments about their effects. In other words, as implicitly expressed by Zhang and Lee (2019), uncertainty in the sentence is not only about the effect of the neurotoxicant molecules (due to exposure), but also the understanding of the mechanistic complexity of the molecules. These uncertainties fit well under the description of two of the uncertainty sources in Table 2: "Activity/potency", which is described as the "Measure of elicited toxicological effect or adverse effect, degree of the effect, or the ability of a chemical to exert an effect"; and "Mechanistic plausibility", which is described as "Toxic causal pathways of chemicals, involving the identification of molecular initiating events/key events linked causally to a target endpoint".

In the explicit uncertainty category in Table 1, the uncertainty noted in Step 3 relates to the evidence supporting the claim that the studied chemicals caused neurotoxic effects in mice. This fits well under the description of the uncertainty source "Activity/potency evidence", which is described as "Available evidence to support the predicted activity/potency" (Table 2); thus, this uncertainty was categorized as "Activity/potency evidence". The categorization process was undertaken for all uncertainties identified under section 2.4. The distribution and frequencies of the categorized implicit and explicit uncertainties were quantified afterward.

Table 2
Sources of uncertainty (arranged in alphabetical order) relatable to practices and features common to *in silico* toxicology modeling (adopted from Achar et al. (2024a) – under review – with permission from the authors).

Uncertainty sources	Definition of the uncertainty sources
Activity/potency	Measure of elicited toxicological effect or adverse effect, degree of the effect, or ability of a chemical to exert an effect on a receptor
Activity/potency evidence	Available evidence to support the predicted activity/potency
Applicability domain	Boundaries within which a model can be applied and provide reliable and accurate predictions (e.g., adequacy of chemical structure space or category to predict effects of similar chemicals)
Chemical similarity	Resemblance or commonality between chemical compounds, e.g., in terms of functional groups and chemical structure
Chemical structure	Quality (e.g., in terms of relevance) of chemical structures or substructures with respect to a set prediction
Coverage of ADME activities	Consideration of ADME activities in biological systems, including effects of metabolites
Data accuracy	The extent to which measured data deviates from its true value
Data balance	Ratio between the number of chemicals in categories in training dataset – chemical categories with known activities (toxicants) and known non-activities (non-toxicants)
Data relevance	Data contain target information (e.g., kinetics and metabolic property) suitable for modeling or adequate for the interpretation of model predictions
Data reliability	Reproducibility of data between test approaches/sources, or reproducibility of the methodology used in generating the data
Data quantity	Amount of data - whether data is sufficiently available
Data validity	Acceptability of the method used to generate data relative to set guidelines or whether the method measured what it was intended to measure
Descriptor concordance	Degree of agreement between descriptors and other chemical features or chemical toxicokinetic or toxicodynamic properties
Descriptor relevance	Extent to which physicochemical or molecular descriptors are considered toxicologically relevant, or suitable for deriving chemical properties or for a specific prediction task
Extrapolation	Making predictions beyond the range of the observed/known data (e.g., toxicity data) in attempts to obtain new unknown data
Mechanistic plausibility	Toxic causal pathways of chemicals, involving the identification of molecular initiating events/key events linked causally to a target endpoint
Metabolic domain	Consideration of production or presence of metabolites as part of chemical interaction with biological systems
Model performance	Predictivity of a model or how well a model can predict outcomes of interest, which can be evaluated through, for example, an internal/external validation or quantitatively using the measure of statistical fit
Model relevance	Transferability of a model or model prediction to a different prediction context (e.g., regulatory application or prediction of new compounds)
Model structure	A model endogenous representation, such as mathematical formulations (e.g., equations or graphs), choice of algorithms, precision of numerical approximations, and relationships between variables

3. Results

3.1. *Intercoder agreement*

From the two papers (i.e., Amnerkar and Bhusari (2010) and Schmidt

et al. (2004)) used for coding practice (see Supplementary Text S2 for details about the practice session), the three coders agreed 90% of the time on whether a statement, phrase, or word identified was an indicator of epistemic uncertainty and, if so, whether the indicator qualified to be categorized as implicit or explicit ($\alpha_k = 0.87$, 95% CI; N = 3). The indicators identified during the practice session (Text S2) and the coder agreement scores are shown in Table S3. Similar results were obtained in the coding of the 20 papers, where the three coders, on average, agreed 87% (82–94%; account for each paper) of the time that the identified indicators represented implicit epistemic uncertainty ($\alpha_k = 0.86$, 95% CI; N = 3), which, according to Krippendorff (2004), is satisfactory.

3.2. Occurrence of uncertainty indicators

Table S2 (second and fifth columns) shows the full list of implicit and explicit indicators (bolded in the sentences) identified from the 20 analyzed studies. A summary of their frequency of occurrence is provided in Table 3. A total of 406 indicators were identified: The majority (75.6%; 307) of these were implicit, and the rest (24.6%; 99) were explicit. A number of words/phrases implicitly expressing uncertainty were repeated across the 20 studies, with "suggest(s/ed/ing)", "may", and "may be" being the most common, while words/phrases like "implies" and "unlikely to" being among the least common (Table S2).

3.3. Variation in the occurrence of implicit and explicit uncertainty sources between publications

The use of the indicators identified under section 2.3, combined with the process described in section 2.4, allowed for the identification of implicitly and explicitly expressed uncertainties within indicator-containing statements. Each of the identified uncertainties was aligned (i.e., grouped according to) at least one of the uncertainty sources in Table 2 (each is henceforth identified by the uncertainty source it is categorized under) (Tables S2 and 4th and 7th columns). All but two of the uncertainty categories were not represented among the recorded uncertainty sources (Data validity and Metabolic domain). A summary of the distribution of the categorized implicit and explicit uncertainty sources across the 20 studies is provided in Table S4. The calculated total number of occurrences of uncertainty sources in both implicit and explicit categories was 162, of which 104 (64%) were implicit and 58 (36%) explicit. None of the uncertainty sources occurred in all the 20 analyzed studies nor was any pattern observed as related to co-occurrence. For example, although uncertainties related to Coverage of ADME effects and Extrapolation occurred in 5 of the 20 studies, they only co-occurred in two of them (Table S4). The most commonly occurring implicit uncertainty sources were Mechanistic plausibility and Model relevance, while Data balance and Data accuracy recorded the

Table 3
Examples of the implicit and explicit uncertainty indicators (bolded in the sentences) identified in the 20 analyzed studies (see Table S1 for the raw data). The data is arranged in the order of the publication numbers presented in Table S2.

Study #	Implicit uncertainty indicator (bolded in the sentence)	Page #	Frequency	Explicit uncertainty indicator (bolded in the sentence)	Page #	Frequency
1	These results **may indicate a certain probability** that compound 11 is a multitarget ligand.	1398	9	This mx-QSAR has excellent goodness-of-fit statistics […] with **sensitivity (Sn), specificity (Sp), and accuracy (Ac) > 80%.**	1394	1
2	The proposed QSAR model **can be a possible** supporting tool […]	50	20			
3	[…] group that **potentially** explains the high number of incorrect predictions.	429	9	These results indicate that **it is currently difficult** to predict […]	429	7
4	Surprisingly, the above equation **suggests** a lack of relationship with hydrophobicity.	105	46	[…] **many of these data are not available** at 25 °C […]	106	11
5	In the case of dioxane, that is outlier for models (2) and (3), **we can think** that neurotoxicities […]	454	5	Unfortunately, **there is not enough experimental data** […]	457	5
6	[…] cochlear development and **potentially** resulting in permanent auditory loss.	7	16			
7	**Probably**, the less unfavorable contacts of the ketal group inside the sub-pocket […].	4	15	Unfortunately, **none of these models succeeded** in finding compounds more potent […]	3	2
8	**To a certain extent, this indicates** that the structural diversity of our compounds is high […]	167	7	Although **the predictive power of our model is not the best** […]	168	2
9	[…] toxicokinetic properties of the chemical **may** play an important role in the neurotoxicity […]	7	21	[…] **the imbalance dataset** was further adjusted […]	4	12
10	[…] (62.5%) were **inferred to be associated with** Parkinson's disease […].	3309	18	[…] **cannot be easily explained** by reduction of dopaminergic neuronal cells	3312	10
11	One compound **may** lead to 1 or more statistical cases because it **may** give different outcomes […]	1872	7	This linear equation presented good results […] with **overall Accuracy in training series above 90%.**	1872	1
12	The use of enzymes from different tissues and species is a **potential** limitation of the study.	232	6	However, the whole picture of influence is **rather complicated.**	236	1
13	The results obtained from the predicted model **could be attributed to** the experimental verifications.	313	12	[…] – MnAChE **still remain unexplored.**	309	2
14	**One tentative explanation** for this event **could be** related to increased hydrogen […]	3800	11	The anticonvulsant mechanism of the semicarbazones is **not clearly defined.**	3399	2
15	The width of the REP range **can be roughly** interpreted as the lowest possible value […]	19	10	[…] **experimental data are lacking adds another layer of uncertainty** to the NEF predictions.	14	7
16	*It appears that* no high-potency PCB congeners with EC2x values ≪0.2 μM exist.	359	23	Because of the **poor predictivity** of the pEC50 QSAR, and concerns […]	358	9
17	The bad fitting to correlation line for those compounds **can be explained by possible** errors […]	11	23	[…] **no substantial features have been identified that would help** to distinguish […]	5	10
18	This is **suggestive of the potential** for increased potency […]	277	5	The source of the IC50 values […] **may provide some uncertainty**.	230	5
19	[…] structural fragments has a *high possibility* to be neurotoxicant.	5	25	The ECFP_10 and eight molecular descriptors **were not able to better describe the property** […]	4	7
20	The results *suggest* that nHAcc and nHDon *may be* obviously *associated* with drug-induced […]	6042	17	**Admittedly, these methods are not perfect** because they […]	6043	7
			Total = 305			Total = 99

lowest number (each 1/104). Among the explicit uncertainty sources, Model performance was most common while seven of the sources were only mentioned once (Table S4).

3.4. Frequency of uncertainty sources

EFSA et al. (2021) recommend the adoption of three tiers to describe analyzed uncertainty: (1) describe uncertainties collectively (i.e., combined uncertainty for an assessment as a whole), (2) describe uncertainties separately with regards to the main parts of the assessment, and, where possible, (3) describe uncertainties with regards to the smaller parts of the assessment. In this study, we followed these steps to describe the frequencies of the uncertainty sources.

3.4.1. General distribution of uncertainty sources

The total number of times a specific uncertainty source (i.e. implicit and explicit) was expressed in the 20 studies is given in Fig. 2. Mechanistic plausibility was by far the most common, with over 90 occurrences, followed by a group of four sources that were referred to in about 50 occurrences (Descriptor concordance, Model relevance, Model performance, and Activity/potency. These results suggest that a large number of uncertainties in QSAR prediction of neurotoxicity fall within these frequently cited sources. In contrast, Data balance, Data accuracy, and Chemical similarity were among the least frequently cited sources, not to mention Data validity and Metabolic domain, which were not expressed in the analyzed papers, suggesting that these uncertainty sources are not of great concern to researchers in the field.

3.4.2. Comparison of frequencies relating to implicit and explicit uncertainty

The frequencies of uncertainty sources expressed implicitly and explicitly were analyzed in order to determine the extent of any variation (Fig. 3). The most frequent uncertainty source that was implicitly expressed was Mechanistic plausibility (73/310). Other high-frequency sources included Descriptor concordance (51/310), Model relevance (39/310), Activity/potency (32/310), and Model performance (27/310). As noted earlier, Data validity and Metabolic domain were not cited in any of the analyzed studies. Among sources that were explicitly referenced, Model performance was the most frequent one (20/102), followed by other sources such as Data quantity (18/102), Activity/potency (16/102), and Mechanistic plausibility (12/102). In contrast, Data accuracy, Chemical structure, and Model structure (each occurring only once) were among the least frequent. Similar to the implicitly expressed uncertainty sources, Data validity and Metabolic domain were not referred to at all. Implicit and explicit mentions were equally common for 2/20 of the sources (i.e., Data accuracy and Extrapolation), while explicit mentions were more common for only 4/20 sources (i.e., Data quantity, Data balance, and Activity/potency evidence) (see Fig. 3). Taken together, the analysis shows that implicit mentions were more common for the majority of the uncertainty sources (13/20; 65%), which indicates that uncertainty is more commonly expressed implicitly in QSAR studies predicting neurotoxicity of chemicals.

4. Discussion

QSARs for toxicity prediction have become ubiquitous in chemical safety assessment. Understanding the uncertainties in QSAR models and their predictions is a vital and fundamental part of assessing the quality and robustness, or otherwise, of a model for their successful use, for instance, in regulatory contexts. This study evaluated 20 papers relating to QSARs for neurotoxicity prediction to assess the occurrence of implicit and explicit uncertainties expressed in them. The study results are discussed below.

4.1. Contribution of implicit versus explicit uncertainty sources to the overall uncertainty sources

Figs. 2 and 3 indicate a variety of uncertainty sources and the different frequencies in which they are mentioned, ranging from relatively frequent to no mention at all. These results raise three questions: Why are the uncertainties variably expressed? What is the contribution of the implicit versus explicit uncertainty sources to the overall

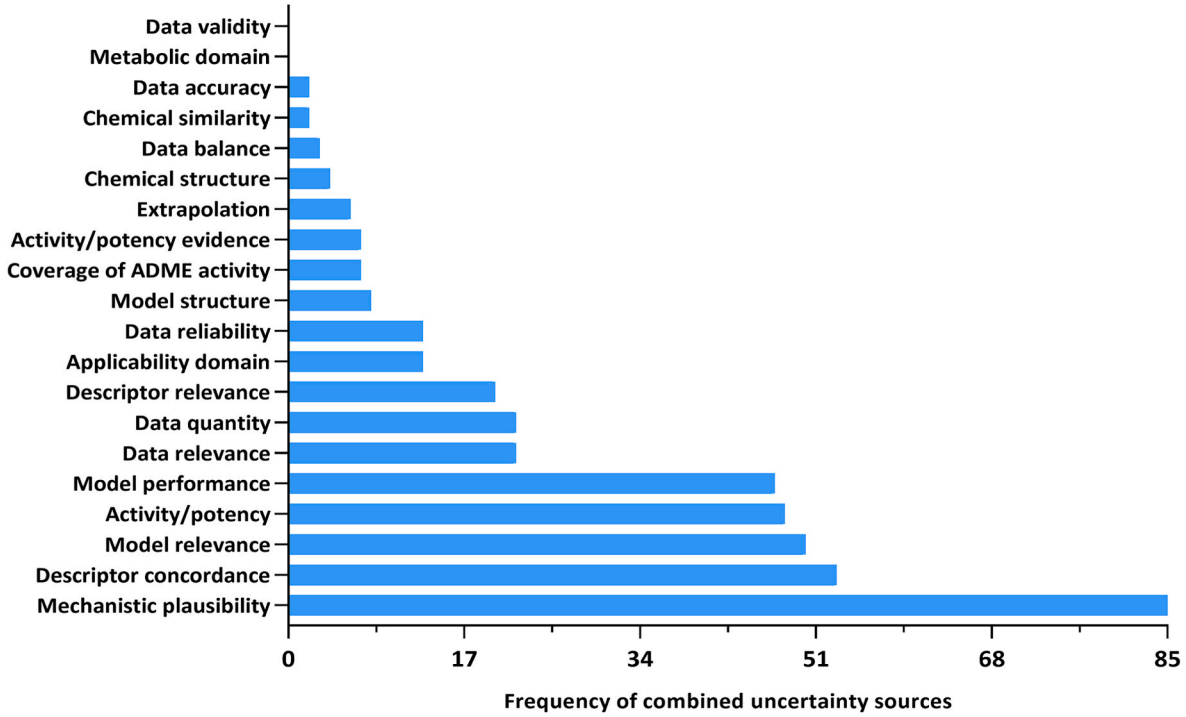

Fig. 2. Frequencies of the combined (implicit + explicit) uncertainty sources. The data are arranged from the highest to the lowest value, according to the magnitude of the combined frequencies.

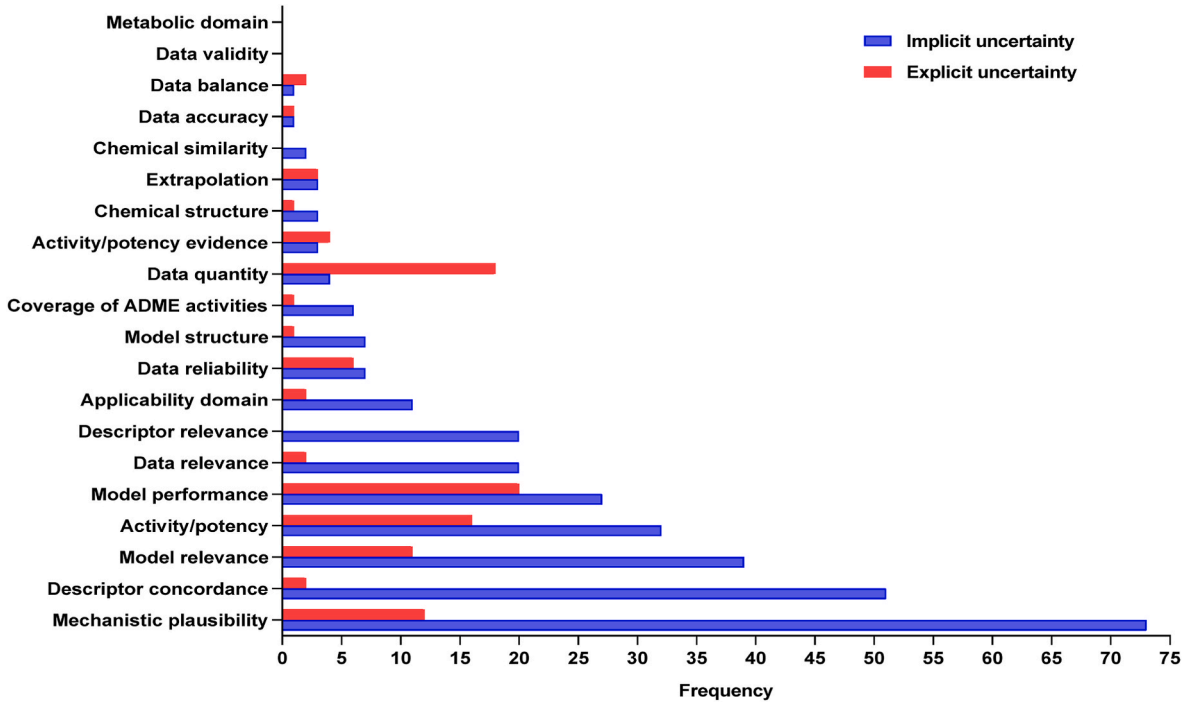

Fig. 3. Frequency of the uncertainty sources in the implicit and explicit uncertainty categories (arranged according to the magnitude of frequency of implicit sources).

uncertainty? With respect to uncertainty communication in QSAR studies predicting neurotoxicity, what is the possible explanation for why implicit uncertainty is more frequently expressed than explicit uncertainty?

According to Han et al. (2011) and Maxim (2015), the first question (i.e., Why are the uncertainties variably expressed?) can be answered by considering the tendency of studies to prioritize communicating particular uncertainties over others, or the inherent difficulty of including all relevant uncertainties in a study. A similar observation has been made within modeling contexts. For example, studies have found that there is a higher tendency to communicate uncertainties related to model parameters than those related to, for example, extrapolation of laboratory experimental data to humans or field settings, suitability or robustness of models or model structure for an intended prediction, or inaccuracies in the experimental design used for data generation (Maxim, 2015; Moschandreas and Karuchit, 2002; National Research Council, 2009; Verdonck et al., 2005). According to Kirchner et al. (2021), this commonly results from modelers' assumption that the added value of including particular uncertainty sources is not worth it, as it may push models to unrealistic and extreme boundary solutions, or for reasons such as resource or computational constraints.

The answer to the second question – i.e., What is the contribution of the implicit versus explicit uncertainty sources to the overall uncertainty? – can be deduced from the distribution of the frequency datasets in Figs. 2 and 3, which indicates that the frequencies of the implicit uncertainty sources were generally more than explicit sources. Based on the analyzed studies and with respect to these results, it can thus be concluded that implicit uncertainties contributed more to the overall uncertainties. Indeed, these results are consistent with the frequencies of the indicators, which were more in the implicit uncertainty category than the explicit uncertainty category (Table 2). Previous research reported similar findings. For example, Flari and Wilkinson (2011), in the text analysis of uncertainties expressed in EFSA documents on health risk assessments, found more implicit uncertainties (972/1133) than explicit uncertainties (161/1133). Stortenbeker et al. (2019) also recorded ≈1.54-fold more implicit uncertainties than explicit uncertainties during physician explanations of medical symptoms to patients.

The third posed question is: With respect to uncertainty communication in QSAR studies predicting neurotoxicity, what is the possible explanation for why implicit uncertainty is more frequently expressed than explicit uncertainty? To answer this question, as a starting point, it might be useful to consider possible reasons why modelers would preferentially express uncertainties implicitly. Studies on uncertainty communication in scientific studies offer different explanations that may be relevant to this discussion. For example, according to Dhami and Mandel (2022), the preference to implicitly communicate uncertainties is based on modelers' attempts to maintain the perceived credibility of their research, especially in the event that erroneous predictions are made. For instance, using the phrase "it may not occur" to express uncertainty about unexpected prediction output may lead to less credibility loss than "there is uncertainty about the predicted results". van der Bles et al. (2020) and the National Academies of Sciences, Engineering and Medicine (2017) similarly interpreted the tendency for implicit communication of uncertainty, or reluctance to communicate uncertainty, to stem from researchers' efforts to avoid signalling incompetence or inviting criticism regarding their research based upon the presence of uncertainties. In the context of our study, the reluctance to openly communicate uncertainties could also be taken more broadly to imply uncertainty communication bias among QSAR modelers of neurotoxicity (Steijaert et al., 2021); this is based on their prioritization to communicate implicit sources over explicit sources. Cronin et al. (2019b) noted that such bias can negatively affect QSAR models and their use. For example, implicit recognition of uncertainties in the prediction output can give users (e.g., regulators) a false sense of accuracy in the output, or else make it difficult to identify model inputs that require additional data to reduce propagated uncertainties in the outputs.

The variation in the frequencies of implicit versus explicit uncertainties also reflects a paradox between the preference to implicitly communicate uncertainty for reasons such as the perceived negative consequences of explicit communication of uncertainty (Dhami and Mandel, 2022), and the need to explicitly communicate uncertainty to enhance transparency about risk and uncertainty (EFSA et al., 2021;

Flari and Wilkinson, 2011). Our study aligns with this call to explicitly communicate uncertainty. We argue that implicit communication of uncertainties associated with the QSAR prediction of neurotoxicity undermines transparent assessment of, for example, the validity of the models as well as their accuracy. An explicit expression of uncertainty is particularly important during fitness-for-purpose evaluation of QSAR models, as this can guide explicit characterization of relevant uncertainties to improve defensibility of a predicted output and provide a critical basis for informed decision-making on the need for appropriate measures to reduce potential risk of neurotoxicants and the nature of the measures (Belfield et al., 2021; Cronin et al., 2019b; WHO/IPCS, 2018).

4.2. Level of concerns raised about the uncertainty sources

The data presented in Fig. 2 suggest that Mechanistic plausibility constitutes the area of most concern for uncertainty for QSAR modeling of neurotoxicity. To our knowledge, no study has specifically targeted analysis of uncertainty in QSAR prediction of neurotoxicity of chemicals; nevertheless, scholars have expressed concern about uncertainties in mechanistic characterization of substances, which corroborates our findings. For example, Crofton et al. (2022), in their review of the current status of the application of *in silico* approaches towards developmental neurotoxicity, suggest that the incomplete understanding of the underlying mechanisms behind the emergence of adverse outcomes seems to constitute uncertainties in understanding adverse outcomes pathways related to developmental neurotoxicity. Others have similarly suggested that the inability to identify or interpret the mechanisms of actions of neurotoxicants, to an extent, contributes to uncertainties in neurotoxicity assessment (Mundy et al., 2015; Worth et al., 2011a). This especially seems to be the case for developmental neurotoxicity, where exposure to neurotoxicants during brain development or the developmental window further complicates the understanding of the underlying mechanisms of the adverse effects (Bal-Price et al., 2018; Fritsche et al., 2018; Mundy et al., 2015).

Our study also suggests that other highly cited uncertainty sources (e.g., Descriptor concordance, Model relevance, and Model performance) also constitute major areas of concern for uncertainty in relation to neurotoxicity prediction. For example, a number of QSARs have been developed to support neurotoxicity prediction based on the statistical correlation between blood-brain barrier penetration of compounds, and specific neuronal bioactivities (Worth et al., 2011a). Most of these models are based on descriptors such as *in vivo* Log BB (blood-brain barrier), Log PS (permeability-surface area), unbound brain-to-plasma partitioning coefficient, as well as physicochemical descriptors (e.g., lipophilicity, hydrogen bonding, and polar surface area) (Crofton et al., 2022). However, the predictive performance (and consequently the relevance) of the individual or combined models is considered limited. Such limitations have been attributed to, for example, a lack of (relevant) data to establish robust models and inadequacy of the descriptors to support the interpretation of observed adverse effects (Crofton et al., 2022; Worth et al., 2011a,b). Elsewhere, Bal-Price et al. (2018) note that the fact that there are only a few QSAR studies on the effects of chemical compounds on the peripheral and central nervous systems, suggests uncertainties in the characterization of hazard and risk potential neurotoxicants.

Taken collectively, given the limited research on uncertainty in QSAR prediction of neurotoxicity of chemicals, our findings provide a tentative conclusion that the frequencies with which the four sources (Mechanistic plausibility, Descriptor concordance, Model relevance, and Model performance) are cited reflect the state of uncertainty of most concern in this field, and possibly a reflection of their importance in the OECD Principles for the Validation of QSARs (OECD, 2007) or their role in regulatory application of the models. It is, therefore, reasonable that when setting priorities aimed at addressing uncertainties within the field, especially under conditions of limited resources, these sources should be the primary focus.

A number of uncertainty sources shown in Fig. 2 had relatively low frequencies, with two of them not cited at all. Notably, some of these low-frequency sources are related to elements that are commonly considered in model training and test sets (i.e., data balance, accuracy, validity and chemical similarity). The importance of these uncertainty sources is well recognized, given their direct influence on the level of model predictive accuracy, reliability, and adequacy (Cronin et al., 2019b; Madden et al., 2020; Pham et al., 2019; Worth et al., 2011b). An example is uncertainty related to Data validity, which emanates from the quality of the experimental studies from which QSAR modeling data are obtained (Achar et al., 2024b; Belfield et al., 2021; Cronin et al., 2019a,b; EFSA, 2006; Karmaus et al., 2022; Madden et al., 2020; Nelms et al., 2020; Pham et al., 2019; WHO/IPCS, 2018; Worth et al., 2011b). The fact that this uncertainty source was not mentioned in the 20 studies, even though it is recognized as a high concern uncertainty source (Worth et al., 2011b), shows that its analysis must be reviewed in light of relevant literature before drawing conclusions about how to prioritize research needs within the field of QSAR prediction of neurotoxicity; otherwise, further uncertainty might be introduced into model predictions interpretation. Furthermore, where applicable, we believe it might be most useful for modelers to refer to “ignorance”, or lack of knowledge. For example (in this case of Data validity and Metabolic domain), acknowledge that uncertainties related to these sources are not considered in a study due to modelers’ lack of knowledge around them or, if that is the case, acknowledge that these sources are not accounted for reasons such as to make model prediction less complex or easy to interpret.

4.3. Further consideration of the categorized uncertainty sources

In chemical risk assessment, regulatory authorities such as EFSA (2006) and WHO/IPCS (2018) recommend that each uncertainty be analyzed at one of the three tiers: qualitative, deterministic, or probabilistic. Initially, uncertainties may be analyzed qualitatively to support initial judgements about the extent of the uncertainties or support subsequent steps on quantitative estimation of the uncertainties to the extent that is scientifically feasible (EFSA, 2006; WHO/IPCS, 2018). The emphasis here is that it might not be possible to treat all uncertainties quantitatively and that qualitative consideration of uncertainties can give insights into unquantifiable uncertainties, as well as their impact on the overall confidence associated with the assessment outcomes. The critical question in handling such an impact is whether the level of uncertainty or level of confidence is acceptable. While we did not explore the level of uncertainty in the analyzed studies (based on, for example, the weight, type, and consistency of scientific evidence presented (EFSA, 2010)), for the sake of our discussion here, we assume that the magnitude of the frequencies of the uncertainty sources (Fig. 2) reflects the possible level of uncertainty.

The question above cannot be answered without clearly defining the context in which a decision has to be made. The guidance provided by the OECD constitutes a conceptual basis through which one can judge the acceptability of the characterized uncertainty: the OECD principles for the validation of QSARs describe information that is considered useful for assessing the models and model predictions for regulatory purposes (OECD, 2007). The OECD’s (Q)SAR Assessment Framework (QAF) also provides guidance to assess QSAR results from multiple predictions to facilitate the characterization of levels of uncertainty associated with different models and model prediction elements based on semi-quantitative uncertainty scales (i.e., “low”, “medium”, or “high”), as well as support the determination of whether the characterized levels of uncertainty are acceptable within a given context of regulatory decision-making (Gissi et al., 2024; OECD, 2023). However, considering that these guidance in themselves do not define criteria for characterizing uncertainty, it is challenging to define fixed acceptability criteria, as this depends on the intended regulatory use of a model or model predictions (Achar et al., 2024b; Belfield et al., 2021; Cronin

et al., 2019b). For example, high-level uncertainty might be tolerated in hazard screening to inform risk assessment but not in the mechanistic characterization of a model prediction, which requires high levels of certainty, reliability and model validation (Bal-Price et al., 2018; Belfield et al., 2021). The OECD Handbook (Annex 1) provides guidance regarding areas within neurotoxicity assessment that might tolerate different levels of uncertainty levels (OECD, 2018). The templates proposed in the literature (Belfield et al., 2021; Cronin et al., 2019b) also provide useful guidance to developers and users of QSARs on possible ways of judging the acceptability in regulatory decision-making contexts. Defining criteria to judge the acceptability of the characterized uncertainty sources is, however, beyond the scope of our study – it thus remains for future studies to explore this topic by defining how the different levels of uncertainty can fit into a defined decision context.

4.4. Implication of the proposed method for uncertainty analysis in QSAR modeling

The uncertainty identification and categorization structure proposed in our study provides a method for identifying implicit and explicit uncertainties expressed in QSAR modeling studies. The method points to one major implication for modelers and decision-makers navigating uncertainties expressed in the studies. That is, analysis of implicitly or explicitly expressed uncertainties is a three-step process. When analyzing these uncertainties, assessors should: (1) first identify uncertainty indicators and (2) then analyze the corresponding uncertainty communicated within the context of the indicators, whereafter (3) these uncertainties are categorized in a systematic manner. As discussed in our studies, identifying uncertainties through this process can help identify an important blind spot in the analysis of uncertainty within QSAR modeling – i.e., lack of transparent accounting of uncertainty. Furthermore, when it comes to epistemic uncertainty, Janzwood (2023) emphasizes the need to develop methods that enable analysis of as much uncertainties as possible in order to inform decision-makers not only about the presence of the uncertainties but also regarding their sources. We believe that the three-step process described in our study is rigorous to capture a wide range of uncertainty sources expressed in QSAR modeling studies.

Our study also highlights that quantitative analysis of uncertainty, which is widely considered important and recommended in the literature, can only provide a partial account of uncertainty sources within QSAR modeling. Based on the outcome of our study, it seems reasonable that QSAR modeling studies should accompany quantitative uncertainty analysis (e.g., probabilistic uncertainty analysis) with our proposed uncertainty analysis method in order to account for uncertainties that are not possible to quantify (EFSA, 2006; WHO/IPCS, 2018). To our knowledge, our study is the first to develop such a method; the findings from our study may thus prove useful (especially in regulatory contexts) in further assessment of the level of confidence in a study outcome.

While research is important in addressing areas of uncertainties, the implicit expression of uncertainties still creates difficulties in evaluating or drawing conclusions on the level of confidence in models and their predictions (Flari and Wilkinson, 2011; Levin et al., 2004; Zerva, 2019). A possible way to improve transparency about these uncertainties is to employ systematic ways (as described in our study) of identifying and categorizing the uncertainties and, where possible, quantifying the uncertainties. Where uncertainties cannot be quantified, we recommend that, at minimum, QSAR modeler should acknowledge this uncertainty explicitly – this is in line with the working principles of EFSA, where transparency in communicating uncertainty is inextricably linked to the credibility of any reported risk assessment output (EFSA, 2006).

4.5. Potential limitations of the developed method and future work

While our study presents a promising and reproducible method, a few questions remain unanswered. First, we assume that implicit uncertainty is expressed through hedging words. This means we do not distinguish between expressions where the authors are genuinely uncertain about something and when it is more a matter of convention. Vold (2006) notes that hedging words and expressions in English can indicate cautiousness, tentativeness, or politeness; thus, such expressions do not necessarily signal uncertainty; instead, they can form part of the convention of expressing politeness or a humble attitude in academic writing. Providing this distinction was out of the scope of our study but poses a challenge for future studies intending to automate the identification of uncertainty indicators – this challenge was also highlighted by Shanahan et al. (2006). To address this weakness, like EFSA (2006), we recommend harmonizing uncertainty expressions to minimize inconsistent perceptions of uncertainty in QSAR modeling studies. This can be facilitated by including glossaries of the qualitative expressions used and defining whether each expression implies uncertainty. Additionally, it is important for each study to clearly characterize the possible impact of uncertainties attributed to the expressed uncertainties and where such impact cannot be characterized, it is necessary to report that this is the case and that the conclusions drawn from an assessment are conditional on assumptions about the expressed uncertainties.

Although our study concludes that uncertainties related to Data validity and Metabolic domain sources were not expressed in the studies (see Figs. 2 and 3), it is also possible that the conceptual breadth of uncertainty indicators used in our study was not adequate to capture these uncertainties. A possible way to overcome this challenge is to complement the use of the indicators with other ways of analyzing uncertainty, including a systematic analysis of the quality, type, consistency, and amount of the evidence presented by the researchers in the analyzed studies about particular claims. EFSA et al. (2021) similarly recommended incorporating different methods for uncertainty analysis to improve the quality and robustness of the analysis.

Finally, from the definition of Model performance by Achar et al. (2024a) (see Table 2), we assumed that only epistemic factors influence model performance. In reality, however, it can also be the case that such an influence (in)directly arises from, for example, data variability. While it is important to make this distinction in order to accurately characterize the context in which uncertainty in model performance is considered, this was not done by Achar et al. (2024a); thus presenting a potential limitation of the use of the study. It is also important to note that, as explained under Section 1, we used neurotoxicity because it is a complex endpoint that is presently difficult to predict (especially in mechanistic terms) in QSAR. This suggests that the data distributions in Figs. 2 and 3 might depend on the endpoint used. This potential variation was not explored in the current study but remains for future studies to investigate.

5. Conclusion

This study aimed to identify and categorize implicit and explicit uncertainties expressed in studies that use QSAR models to predict the neurotoxicity of chemical compounds. It was found that most of the identified indicators were implicitly expressed (310), compared to those expressed explicitly (102). This indicates that, within studies on QSAR prediction of neurotoxicity of chemicals, it is considerably more common to express uncertainties implicitly than explicitly. Four uncertainty sources were most commonly referred to: Mechanistic plausibility, Descriptor relevance, Model performance, and Model relevance. This suggests that researchers are concerned about uncertainties related to, for example, predicting the adverse health risks of compounds based on mechanistic knowledge, the applicability of models (e.g., in terms of their relevance), developing models that can adequately predict external data sets or *in vivo* data, and the descriptors used in model development and predictions. It was noted that some of the uncertainty sources that were rarely noted, or not noted at all (e.g. Data validity and Metabolic domain) are in fact flagged as a concern in the broader QSAR literature as areas that researchers commonly overlook. This implies that while

analysis of expressed uncertainty can help identify areas of uncertainties, conclusions can only be drawn after the output has been analyzed in light of the broader QSAR literature. Overall, the findings from our study cannot only be used to guide modelers during modeling processes to prioritize uncertainty sources for uncertainty analysis based on the magnitude of their frequencies but also facilitate a structured dialogue between modelers and decision-makers about the need for more research to improve existing models or develop new ones that can reduce these uncertainties.

Funding information

This study was supported by the Vanier Canada Graduate Scholarship (funded by the Federal Government of Canada) and the Social Science, and Humanities Research Council (SSHRC) grant (Grant No: 435-2019-0465).

CRediT authorship contribution statement

Jerry Achar: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **James W. Firman:** Writing – review & editing, Formal analysis. **Chantelle Tran:** Formal analysis. **Daniella Kim:** Formal analysis. **Mark T.D. Cronin:** Writing – review & editing. **Gunilla Öberg:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yrtph.2024.105716.

Data availability

No data was used for the research described in the article.

References

Achar, J., Firman, J.W., Cronin, M.T.D., Öberg, G., 2024a. A Framework for Categorizing Sources of Uncertainty in *in Silico* Toxicology Methods: Considerations for Chemical Toxicity Predictions (under review).

Achar, J., Cronin, M.T.D., Firman, J.W., Öberg, G., 2024b. A problem formulation framework for the application of in silico toxicology methods in chemical risk assessment. Arch. Toxicol. https://doi.org/10.1007/s00204-024-03721-6.

Amnerkar, N.D., Bhusari, K.P., 2010. Synthesis, anticonvulsant activity and 3D-QSAR study of some prop-2-eneamido and 1-acetyl-pyrazolin derivatives of aminobenzothiazole. Eur. J. Med. Chem. 45 (1), 149–159. https://doi.org/10.1016/j.ejmech.2009.09.037.

Bal-Price, A., Pistollato, F., Sachana, M., Bopp, S.K., Munn, S., Worth, A., 2018. Strategies to improve the regulatory assessment of developmental neurotoxicity (DNT) using in vitro methods. Toxicol. Appl. Pharmacol. 354, 7–18. https://doi.org/10.1016/j.taap.2018.02.008.

Belfield, S.J., Enoch, S.J., Firman, J.W., Madden, J.C., Schultz, T.W., Cronin, M.T.D., 2021. Determination of "fitness-for-purpose" of quantitative structure-activity relationship (QSAR) models to predict (eco-)toxicological endpoints for regulatory use. Regul. Toxicol. Pharmacol. 123, 104956. https://doi.org/10.1016/j.yrtph.2021.104956.

Belfield, S.J., Cronin, M.T.D., Enoch, S.J., Firman, J.W., 2023. Guidance for good practice in the application of machine learning in development of toxicological quantitative structure-activity relationships (QSARs). PLoS One 18 (5), e0282924. https://doi.org/10.1371/journal.pone.0282924.

Crofton, K.M., Bassan, A., Behl, M., Chushak, Y.G., Fritsche, E., Gearhart, J.M., Marty, M.S., Mumtaz, M., Pavan, M., Ruiz, P., Sachana, M., Selvam, R., Shafer, T.J., Stavitskaya, L., Szabo, D.T., Szabo, S.T., Tice, R.R., Wilson, D., Woolley, D., Myatt, G.J., 2022. Current status and future directions for a neurotoxicity hazard assessment framework that integrates in silico approaches. Comput. Toxicol.(Amsterdam, Netherlands) 22, 100223. https://doi.org/10.1016/j.comtox.2022.100223.

Cronin, M.T.D., 1996. Quantitative structure-activity relationship (QSAR) analysis of the acute sublethal neurotoxicity of solvents. Toxicol. Vitro 10 (2), 103–110. https://doi.org/10.1016/0887-2333(95)00109-3.

Cronin, M., Madden, J., 2010. In: In Silico Toxicology: Principles and Applications. Royal Society of Chemistry.

Cronin, M.T.D., Madden, J.C., Yang, C., Worth, A.P., 2019a. Unlocking the potential of in silico chemical safety assessment – a report on a cross-sector symposium on current opportunities and future challenges. Computat. Toxicol. 10, 38–43. https://doi.org/10.1016/j.comtox.2018.12.006.

Cronin, M.T.D., Richarz, A.-N., Schultz, T.W., 2019b. Identification and description of the uncertainty, variability, bias and influence in quantitative structure-activity relationships (QSARs) for toxicity prediction. Regul. Toxicol. Pharmacol. 106, 90–104. https://doi.org/10.1016/j.yrtph.2019.04.007.

Dhami, M.K., Mandel, D.R., 2022. Communicating uncertainty using words and numbers. Trends Cognit. Sci. 26 (6), 514–526. https://doi.org/10.1016/j.tics.2022.03.002.

EFSA, 2006. Opinion of the scientific committee related to uncertainties in dietary exposure assessment. EFSA J. EFSA J. https://doi.org/10.2903/j.efsa.2007.438.

EFSA, 2010. Applicability of QSAR Analysis to the Evaluation of the Toxicological Relevance of Metabolites and Degradates of Pesticide Active Substances for Dietary Risk Assessment | EFSA. https://www.efsa.europa.eu/en/supporting/pub/en-50.

EFSA, Hernández-Jerez, A., Adriaanse, P., Aldrich, A., Berny, P., Coja, T., Duquesne, S., Focks, A., Marinovich, M., Millet, M., Pelkonen, O., Pieper, S., Tiktak, A., Topping, C., Widenfalk, A., Wilks, M., Wolterink, G., Crofton, K., Hougaard Bennekou, S., et al., 2021. Development of Integrated Approaches to Testing and Assessment (IATA) case studies on developmental neurotoxicity (DNT) risk assessment. EFSA J. 19 (6), e06599. https://doi.org/10.2903/j.efsa.2021.6599.

Estrada, E., Molina, E., Uriarte, E., 2001. Quantitative structure-toxicity relationships using tops-mode. 2. Neurotoxicity of a non-congeneric series of solvents. SAR and QSAR in Environ. Res. 12 (5), 445–459. https://doi.org/10.1080/10629360108035384.

Ferson, S., O'Rawe, J., Antonenko, A., Siegrist, J., Mickley, J., Luhmann, C.C., Sentz, K., Finkel, A.M., 2015. Natural language of uncertainty: numeric hedge words. Int. J. Approx. Reason. 57, 19–39. https://doi.org/10.1016/j.ijar.2014.11.003.

Flari, V., Wilkinson, D., 2011. Terminology in risk assessments used by the scientific panels and scientific committee of EFSA. EFSA Support. Pub. 8 (1). https://doi.org/10.2903/sp.efsa.2011.EN-101.

Fritsche, E., Grandjean, P., Crofton, K.M., Aschner, M., Goldberg, A., Heinonen, T., Hessel, E.V.S., Hogberg, H.T., Bennekou, S.H., Lein, P.J., Leist, M., Mundy, W.R., Paparella, M., Piersma, A.H., Sachana, M., Schmuck, G., Solecki, R., Terron, A., Monnet-Tschudi, F., et al., 2018. Consensus statement on the need for innovation, transition and implementation of developmental neurotoxicity (DNT) testing for regulatory purposes. Toxicol. Appl. Pharmacol. 354, 3–6. https://doi.org/10.1016/j.taap.2018.02.004.

Gadaleta, D., Spînu, N., Roncaglioni, A., Cronin, M.T.D., Benfenati, E., 2022. Prediction of the Neurotoxic Potential of Chemicals Based on Modelling of Molecular Initiating Events Upstream of the Adverse Outcome Pathways of (Developmental) Neurotoxicity. Int. J. Mol. Sci. 23 (6), 3053. https://doi.org/10.3390/ijms23063053.

Gajewicz, A., Schaeublin, N., Rasulev, B., Hussain, S., Leszczynska, D., Puzyn, T., Leszczynski, J., 2015. Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: hints from nano-QSAR studies. Nanotoxicology 9 (3), 313–325. https://doi.org/10.3109/17435390.2014.930195.

Gissi, A., Tcheremenskaia, O., Bossa, C., Battistelli, C.L., Browne, P., 2024. The OECD (Q)SAR Assessment Framework: a tool for increasing regulatory uptake of computational approaches. Computat. Toxicol. 31, 100326. https://doi.org/10.1016/j.comtox.2024.100326.

Han, P.K.J., Klein, W.M.P., Arora, N.K., 2011. Varieties of uncertainty in health care: a conceptual taxonomy. Med. Decis. Making: An Int. J. Soc. Med. Decision Making 31 (6), 828–838. https://doi.org/10.1177/0272989x11393976.

Health Canada, 2023. Notice of Intent on the Development of a Strategy to Guide the Replacement, Reduction, or Refinement of Vertebrate Animal Testing under the Canadian Environmental Protection Act, 1999 (CEPA) [Consultations]. https://www.canada.ca/en/health-canada/programs/consultation-strategy-replace-reduce-refine-vertebrate-animal-testing/notice-intent.html.

Hillen, M.A., Gutheil, C.M., Strout, T.D., Smets, E.M.A., Han, P.K.J., 2017. Tolerance of uncertainty: conceptual analysis, integrative model, and implications for healthcare. Soc. Sci. Med. 180, 62–75. https://doi.org/10.1016/j.socscimed.2017.03.024.

Hung, C., Gini, G., 2021. QSAR modeling without descriptors using graph convolutional neural networks: the case of mutagenicity prediction. Mol. Divers. 25 (3), 1283–1299. https://doi.org/10.1007/s11030-021-10250-2.

Janzwood, S., 2023. Confidence deficits and reducibility: toward a coherent conceptualization of uncertainty level. Risk Anal. 43 (10), 2004–2016. https://doi.org/10.1111/risa.14008.

Karmaus, A.L., Mansouri, K., To, K.T., Blake, B., Fitzpatrick, J., Strickland, J., Patlewicz, G., Allen, D., Casey, W., Kleinstreuer, N., 2022. Evaluation of variability across rat acute oral systemic toxicity studies. Toxicol. Sci. 188 (1), 34–47. https://doi.org/10.1093/toxsci/kfac042.

Kirchner, M., Mitter, H., Schneider, U.A., Sommer, M., Falkner, K., Schmid, E., 2021. Uncertainty concepts for integrated modeling—review and application for identifying uncertainties and uncertainty propagation pathways. Environ. Model. Software 135, 104905. https://doi.org/10.1016/j.envsoft.2020.104905.

Krippendorff, K., 2004. Content Analysis: An Introduction to Its Methodology. SAGE.

Levin, R., Hansson, S.O., Rudén, C., 2004. Indicators of uncertainty in chemical risk assessments. Regul. Toxicol. Pharmacol. 39 (1), 33–43. https://doi.org/10.1016/j.yrtph.2003.11.001.

Madden, J.C., Enoch, S.J., Paini, A., Cronin, M.T.D., 2020. A review of in silico tools as alternatives to animal testing: principles, resources and applications. Altern. Lab. Anim. 48 (4), 146–172. https://doi.org/10.1177/0261192920965977.

Markkanen, R., Schröder, H., 1997. Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts. Walter de Gruyter.

Marzi, G., Balzano, M., Marchiori, D., 2024. K-Alpha Calculator–Krippendorff's Alpha Calculator: a user-friendly tool for computing Krippendorff's Alpha inter-rater reliability coefficient. MethodsX 12, 102545. https://doi.org/10.1016/j.mex.2023.102545.

Maxim, L., 2015. A systematic review of methods of uncertainty analysis and their applications in the assessment of chemical exposures, effects, and risks. Int. J. Environ. Health Res. 25 (5), 522–550. https://doi.org/10.1080/09603123.2014.980782.

Moschandreas, D.J., Karuchit, S., 2002. Scenario–model–parameter: a new method of cumulative risk uncertainty analysis. Environ. Int. 28 (4), 247–261. https://doi.org/10.1016/S0160-4120(02)00025-9.

Mundy, W.R., Padilla, S., Breier, J.M., Crofton, K.M., Gilbert, M.E., Herr, D.W., Jensen, K. F., Radio, N.M., Raffaele, K.C., Schumacher, K., Shafer, T.J., Cowden, J., 2015. Expanding the test set: chemicals with potential to disrupt mammalian brain development. Neurotoxicol. Teratol. 52, 25–35. https://doi.org/10.1016/j.ntt.2015.10.001.

National Academies of Sciences, Engineering and Medicine, 2017. Communicating Science Effectively: A Research Agenda. National Academies Press.

National Research Council, 2007. Read "Toxicity Testing in the 21st Century: A Vision and a Strategy" at NAP.Edu. https://doi.org/10.17226/11970.

National Research Council, 2009. Science and Decisions: Advancing Risk Assessment. National Academies Press (US). http://www.ncbi.nlm.nih.gov/books/NBK214630/.

Nelms, M.D., Karmaus, A.L., Patlewicz, G., 2020. An evaluation of the performance of selected (Q)SARs/expert systems for predicting acute oral toxicity. Comput. Toxicol. 16, 100135. https://doi.org/10.1016/j.comtox.2020.100135.

OECD, 2007. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. https://doi.org/10.1787/9789264085442-en.

OECD, 2018. *Users' Handbook Supplement to the Guidance Document For Developing and Assessing Adverse Outcome Pathways* (OECD Series on Adverse Outcome Pathways 1; OECD Series on Adverse Outcome Pathways, vol. 1. https://doi.org/10.1787/5jlv1m9d1g32-en.

OECD, 2023. (Q)SAR Assessment Framework: Guidance for the Regulatory Assessment of (Quantitative) Structure − Activity Relationship Models, Predictions, and Results Based on Multiple Predictions. https://one.oecd.org/document/ENV/CBC/MONO(2023)32/en/pdf.

Patlewicz, G., Ball, N., Booth, E.D., Hulzebos, E., Zvinavashe, E., Hennes, C., 2013. Use of category approaches, read-across and (Q)SAR: general considerations. Regul. Toxicol. Pharmacol. 67 (1), 1–12. https://doi.org/10.1016/j.yrtph.2013.06.002.

Pham, L.L., Sheffield, T.Y., Pradeep, P., Brown, J., Haggard, D.E., Wambaugh, J., Judson, R.S., Paul Friedman, K., 2019. Estimating uncertainty in the context of new approach methodologies for potential use in chemical safety evaluation. Curr. Opinion Toxicol. 15, 40–47. https://doi.org/10.1016/j.cotox.2019.04.001.

Piir, G., Kahn, I., García-Sosa, A.T., Sild, S., Ahte, P., Maran, U., 2018. Best practices for QSAR model reporting: physical and chemical properties, ecotoxicity, environmental fate, human health, and toxicokinetics endpoints. Environ. Health Perspect. 126 (12), 126001. https://doi.org/10.1289/EHP3264.

Rubin, V.L., 2007. Stating with certainty or stating with doubt: intercoder reliability results for manual annotation of epistemically modalized statements. In: Sidner, C., Schultz, T., Stone, M., Zhai, C. (Eds.), Human Language Technologies 2007: the Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers. Association for Computational Linguistics, pp. 141–144. https://aclanthology.org/N07-2036.

Sahlin, U., Filipsson, M., Öberg, T., 2011. A risk assessment perspective of current practice in characterizing uncertainties in QSAR regression predictions. Molecular Informatics 30 (6–7), 551–564. https://doi.org/10.1002/minf.201000177.

Sahlin, U., Golsteijn, L., Iqbal, M.S., Peijnenburg, W., 2013. Arguments for considering uncertainty in QSAR predictions in hazard and risk assessments. Altern. Lab. Anim. 41 (1), 91–110. https://doi.org/10.1177/026119291304100110.

Scalia, G., Grambow, C.A., Pernici, B., Li, Y.-P., Green, W.H., 2020. Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction. J. Chem. Inf. Model. 60 (6), 2697–2717. https://doi.org/10.1021/acs.jcim.9b00975.

Schultz, T.W., Richarz, A.-N., Cronin, M.T.D., 2019. Assessing uncertainty in read-across: questions to evaluate toxicity predictions based on knowledge gained from case studies. Computat. Toxicol. 9, 1–11. https://doi.org/10.1016/j.comtox.2018.10.003.

Shanahan, J.G., Qu, Y., Wiebe, J., 2006. Computing Attitude and Affect in Text: Theory and Applications, Vol. 20. Springer Sci. Business Media, pp. 61–76.

Steijaert, M.J., Schaap, G., Riet, J.V., 2021. Two-sided science: communicating scientific uncertainty increases trust in scientists and donation intention by decreasing attribution of communicator bias. Communications 46 (2), 297–316. https://doi.org/10.1515/commun-2019-0123.

Stortenbeker, I., Houwen, J., van Dulmen, S., olde Hartman, T., Das, E., 2019. Quantifying implicit uncertainty in primary care consultations: a systematic comparison of communication about medically explained versus unexplained symptoms. Patient Educ. Counsel. 102 (12), 2349–2352. https://doi.org/10.1016/j.pec.2019.07.005.

Turabekova, M.A., Rasulev, B.F., Levkovich, M.G., Abdullaev, N.D., Leszczynski, J., 2008. Aconitum and Delphinium sp. alkaloids as antagonist modulators of voltage-gated Na+ channels. AM1/DFT electronic structure investigations and QSAR studies. Comput. Biol. Chem. 32 (2), 88–101. https://doi.org/10.1016/j.compbiolchem.2007.10.003.

van der Bles, A.M., van der Linden, S., Freeman, A.L.J., Spiegelhalter, D.J., 2020. The effects of communicating uncertainty on public trust in facts and numbers. Proc. Natl. Acad. Sci. USA 117 (14), 7672–7683. https://doi.org/10.1073/pnas.1913678117.

Varttala, T., 2001. Hedging in Scientifically Oriented Discourse.

Verdonck, F.a.M., Van Sprang, P.A., Vanrolleghem, P.A., 2005. Uncertainty and precaution in European environmental risk assessment of chemicals. Water Sci. Technol.: A J. Int. Assoc. on Water Pollut. Res. 52 (6), 227–234.

Vighi, M., Barsi, A., Focks, A., Grisoni, F., 2019. Predictive models in ecotoxicology: bridging the gap between scientific progress and regulatory applicability—remarks and research needs. Integrated Environ. Assess. Manag. 15 (3), 345–351. https://doi.org/10.1002/ieam.4136.

Vold, E.T., 2006. Epistemic modality markers in research articles: a cross-linguistic and cross-disciplinary study. Int. J. Appl. Ling. 16 (1), 61–87. https://doi.org/10.1111/j.1473-4192.2006.00106.x.

Wang, D., Yu, J., Chen, L., Li, X., Jiang, H., Chen, K., Zheng, M., Luo, X., 2021. A hybrid framework for improving uncertainty quantification in deep learning-based QSAR regression modeling. J. Cheminf. 13 (1), 69. https://doi.org/10.1186/s13321-021-00551-x.

WHO/IPCS, 2018. Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization, second ed. World Health Organization https://apps.who.int/iris/handle/10665/259858.

Worth, A., Fuart-Gatnik, M., Lapenna, S., Serafimova, R., 2011a. Applicability of QSAR analysis in the evaluation of developmental and neurotoxicity effects for the assessment of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. EFSA Support. Pub. 8 (6). https://doi.org/10.2903/sp.efsa.2011.EN-169.

Worth, A., Lapenna, S., Lo Piparo, E., Mostrag-Szlichtyng, A., Serafimova, R., 2011b. A Framework for assessing in silico toxicity predictions: case studies with selected pesticides. JRC Publications Repository. https://doi.org/10.2788/29048.

Zerva, C., 2019. Automatic Identification of Textual Uncertainty, vol. 422.

Zhang, Y., Lee, A.A., 2019. Bayesian semi-supervised learning for uncertainty-calibrated prediction of molecular properties and active learning. Chem. Sci. 10 (35), 8154–8163. https://doi.org/10.1039/C9SC00616H.

Zhong, S., Lambeth, D.R., Igou, T.K., Chen, Y., 2022. Enlarging applicability domain of quantitative structure–activity relationship models through uncertainty-based active learning. ACS ESandT Eng. 2 (7), 1211–1220. https://doi.org/10.1021/acsestengg.1c00434.