molecular informatics

DOI: 10.1002/minf.200((full DOI will be filled in by the editorial staff))

Towards a Fuzzy Expert System on Toxicological Data Quality Assessment

Longzhi Yang,[a] Daniel Neagu,*[a] Mark T. D. Cronin,[b] Mark Hewitt,[b] Steven J. Enoch,[b] Judith C. Madden,*[b] and Katarzyna Przybylak[b]

Abstract: Quality assessment (QA) requires high levels of domain-specific experience and knowledge. QA tasks for toxicological data are usually performed by human experts manually, although a number of quality evaluation schemes have been proposed in the literature. For instance, the most widely utilised Klimisch scheme[1] defines four data quality categories in order to tag data instances with respect to their qualities; ToxRTool[2] is an extension of the Klimisch approach aiming to increase the transparency and harmonisation of the approach. Note that the processes of QA in many other areas have been automatised by employing expert systems. Briefly, an expert system is a computer program that uses a knowledge base built upon human expertise, and an inference engine that mimics the reasoning processes of human experts to infer new statements from incoming data. In particular, expert systems have been extended to deal with the uncertainty of information by representing uncertain information (such as linguistic terms) as fuzzy sets under the framework of fuzzy set theory and performing inferences upon fuzzy sets according to fuzzy arithmetic. This paper presents an experimental fuzzy expert system for toxicological data QA which is developed on the basis of the Klimisch approach and the ToxRTool in an effort to illustrate the power of expert systems to toxicologists, and to examine if fuzzy expert systems are a viable solution for QA of toxicological data. Such direction still faces great difficulties due to the well-known common challenge of toxicological data QA that "five toxicologists may have six opinions". In the meantime, this challenge may offer an opportunity for expert systems because the construction and refinement of the knowledge base could be a converging process of different opinions which is of significant importance for regulatory policy making under the regulation of REACH, though a consensus may never be reached. Also, in order to facilitate the implementation of Weight of Evidence approaches and *in silico* modelling proposed by REACH, there is a higher appeal of numerical quality values than nominal (categorical) ones, where the proposed fuzzy expert system could help. Most importantly, the deriving processes of quality values generated in this way are fully transparent, and thus comprehensible, for final users, which is another vital point for policy making specified in REACH. Case studies have been conducted and this report not only shows the promise of the approach, but also demonstrates the difficulties of the approach and thus indicates areas for future development.

Keywords: Toxicological data, quality (reliability) assessment, The Klimisch approach, the ToxRTool, *in silico* modelling, fuzzy expert systems, REACH

1 Introduction

Data quality is one of the central issues of databases. A database could be useless if the data stored within the database do not have "adequate" quality. In general, data quality refers to its fitness for serving its purpose in a given context.[3] Usually, data quality needs to be considered from multiple dimensions, including accuracy, validity, reliability, completeness, adequacy and relevance. Accuracy means that data should be sufficiently accurate for their intended purposes. Data reliability means that data should reflect stable and consistent data recording and collection processes across collection points and over time; this includes ensuring data are recorded in compliance with relevant requirements and good practices. Relevance means that data should be relevant to the purposes for which they are used; this entails periodic review of requirements to reflect changing needs. Adequacy defines the usefulness of data for risk assessment purposes; where the greatest weight is attached to the most reliable and relevant record if there is more than one piece of data for each effect. Completeness means that data requirements should be clearly specified based on the information needs of the body and data collection processes matching these requirements.

The purposes of a database can be multiple and sometimes not all such purposes are specified explicitly *a prior.* The dimensions of data quality can be categorised into two groups by considering if these dimensions are relevant to the context of its application: "inherent quality" and "extraneous quality". Inherent quality of a piece of toxicological data concerns the

[a] *School of Computing, Informatics and Media*
University of Bradford, Bradford, BD7 1DP, UK
**e-mail: D.Neagu@Bradford.ac.uk*

[b] *The School of Pharmacy & Biomolecular Sciences, Faculty of Science*
Liverpool John Moores University, Liverpool, L3 3AF, UK
**e-mail: J.C.Madden@ljmu.ac.uk*

Supporting Information for this article is available on the WWW under www.molinf.com

aspects that come with the condition of the tested chemical compound, the assay, and the recording process of the data. Inherent quality is not relevant to the context of where and how the data are to be used and thus it includes the dimensions of accuracy, reliability and completeness. In contrast, extraneous quality concerns if the data instance is suitable for a particular task. In other words, extraneous quality is related to its specific application domain. The relevance and adequacy of data depend upon the specific context in which they are to be used, and thus these two dimensions belong to extraneous quality.

Toxicological databases not only inherit the properties of general databases, but also have their own properties. Firstly, the accuracy of the data usually cannot be given along with the collection of the data or checked afterwards because it is difficult, if not possible, to know the exact truth. An assay is only one sample of the prediction problem, and it is not practical to have a big number of tests in order to check the accuracy of the data. Secondly, the reliability of toxicological data can be defined by chemical data reliability and biological data reliability. Chemical data reliability refers to the fact that data should contain sufficient information to uniquely identify and characterise the tested chemical compound and the structures used for computational activities. Biological data reliability refers to the fact that the design and execution of biological study should be compliant with related regulations and requirements. It is a common practice in toxicology to use the reliability of a piece of data to represent its inherent quality, which is the main focus of this paper.

A number of formal systems have been proposed for rating the quality of toxicological data in terms of their inherent quality, such as the Klimisch scoring system[1] and Przybylak criteria.[4] In particular, the Klimisch approach has gained most attention and been applied widely. For instance, it has been employed by the Organisation of Economic Cooperation and Development (OECD) Screening Information Data Set (SIDS).[5] This method aims to categorise toxicological data into one of four reliability categories: Reliable Without Restrictions, Reliable With Restrictions, Not Reliable and Not Assignable. Because the approach is very general and it is difficult to distinguish between categories, the Toxicological data Reliability Assessment Tool (ToxRTool)[2] was developed in an effort to improve the Klimisch approach.[1] The tool increases the specificity and transparency by explicitly giving a list of evaluation criteria for scoring.

Notice that the criteria utilised in the ToxRTool are only “YES” or “NO” questions. Also, it is difficult sometimes, and against the natural judging processes of human experts, to categorise a given data instance into any of these crisp categories with respect to a particular criterion. In other words, a piece of data may only partially satisfy a certain criterion. This paper introduces a fuzzy expert system in an effort to “soften” the original crisp assessment criteria given in the ToxRTool in two ways. Firstly, it allows the scoring of partial satisfactions of criteria. Secondly, by noticing that the score itself may not be fully certain based on the currently available information, the system also takes this uncertainty factor into consideration during the quality evaluation processes. Compared to neural networks,[6] a distinguishing advantage of fuzzy logic[7] is that it can preserve comprehensibility and transparency during reasoning processes,[8] which is also crucial for *in vitro*-based hazard assessment, including the necessary quality evaluation process for toxicological data, as stated in OECD guidance document[9] and the ECHA guidance document.[10] The proposed approach mimics well the process of quality assessment by human experts and thus is expected to be able to achieve good results.

The remainder of the paper is organised as follows. Section 2 reviews the Klimisch approach and its extension, the ToxRTool. Section 3 describes the proposed fuzzy expert system. Realistic applications are given in Section 4 for illustration and evaluation purposes. Conclusion is drawn in Section 5, where possible future research directions are also suggested.

2 Background

The development of systems for data quality evaluation (including the Klimisch approach) is initialised by the risk assessment processes for “existing substances”, as described in the Europe Council Regulation 793/93 and Commission Regulation 1488/94 (EEC, 1994). According to these regulations, all the existing data and metadata for those substances included in a published list must be submitted by the manufacturer or importer. As a result of this, multiple data instances with different qualities may be available for the same chemical compound with respect to one single end point. Therefore, the supplied data must be properly evaluated. Then, the data instance with the best quality can be utilised for toxicological prediction.[11]

The growing interest in the evaluation of data quality is also driven by the development of Integrated Testing Strategies (ITS),[12] which has been considered to be utilised to “speeding up” of hazard and risk assessment of chemicals whilst reducing testing costs and animal use.[13] This approach aims to integrate different types of data and information, including existing *in vivo* data, data from *in vitro* assays, *in silico* models and expert systems[12] into the decision-making process so that a conclusion based on all the existing available information can be drawn. Crucially, an ITS can also incorporate approaches such as weight-of-evidence and exposure/population data into the final risk assessment for a substance. Importantly, if low quality data are entered into an ITS, the output of the ITS will, of course, be of low quality and of little regulatory use. In order to maximise the confidence in a correct ITS decision, it is crucial that the quality of data is assessed before the employing of an ITS scheme.[14]

Besides the aforementioned motivations, there is another important driving force for the development of quality assessment systems. Along with the rapid development of data storing and sharing techniques in terms of both hardware and software, multiple data instances scattered across multiple databases may be available to support one single task, and then making choices of data are necessary from time to time. The assessment of data quality is thus a prerequisite to make use of available supporting data. In particular, a fuzzy approach to integrate heterogeneous uncertain data for toxicological prediction has been recently reported.[15]

General guidelines on data evaluation have been available in the literature, such as the "Technical Guidance Document"[16] based on general principles for data evaluation of the International Coordination of Criteria Document Production,[17] and the guidelines described by OECD.[18] In comparison to these general guidelines, the Klimisch approach, especially its extension ToxRTool, is a specific tool, which is able to be applied directly to a toxicological quality assessment task. Based on whether or not the test is conducted and reported according to internationally accepted test guidelines (EU, EPA, FDA, OECD) and in compliance with the principles of Good Laboratory Practice (GLP), the Klimisch approach assigns toxicological data into one of the four reliability categories (referred as the Klimisch categories hereafter) as shown in Table 1.

Table 1. Klimisch et al.[1] scoring system to assess data reliability

Score	Description
I	**Reliable Without Restrictions**
	"Studies or data from the literature or reports which were carried out or generated according to generally valid and/or internationally accepted testing guidelines (preferably performed according to GLP) or in which the test parameters documented are based on a specific (national) testing guideline (preferably performed according to GLP) or in which all parameters described are closely related/comparable to a method."
II	**Reliable With Restrictions**
	"Studies or data from the literature, reports (mostly not performed according to GLP), in which the test parameters documented do not totally comply with the specific testing guideline, but are sufficient to accept the data or in which investigations are described which cannot be subsumed under a testing guideline, but which are nevertheless well documented and scientifically acceptable."
III	**Not Reliable**
	"Studies or data from the literature/reports in which there were interferences between the measuring system and the test substance or in which organisms/test systems were used which are not relevant in relation to the exposure (e.g., unphysiologic pathways of application) or which were carried out or generated according to a method which is not acceptable, the documentation of which is not sufficient for assessment and which is not convincing for an expert judgement."
IV	**Not Assignable**
	"Studies or data from the literature, which do not give sufficient experimental details and which are only listed in short abstracts or secondary literature (books, reviews, etc.)."

Although the Klimisch scheme has become the most widely used data quality measure approach, the description of the four data quality categories in the Klimisch approach is very general and thus it is difficult to be practically used to distinguish between categories. Toxicological data Reliability assessment Tool (ToxRTool)[2] was developed in an effort to improve the Klimisch approach. The tool increases the transparency by explicitly giving a list of evaluation criteria for scoring. It is applicable to various types of experimental data, endpoints and studies. The tool consists of two parts which can be utilised to evaluate *in vivo* and *in vitro* data, respectively. In particular, *in vivo* data is evaluated based on 21 detailed criteria (referred as the ToxRTool criteria hereafter), as shown in Table 2; and *in vitro* data are evaluated based on 18 detailed criteria. For simplicity, only the part of tool which is used for *in vivo* data is discussed in this paper.

Table 2. The 21 criteria utilised in the ToxRTool[2] for *in vivo* data

No.	Criteria Group I: Test substance identification
1	*Was the test substance identified?*
2	Is the purity of the substance given?
3	Is information on the source/origin of the substance given?
4	Is all information on the nature and/or physico-chemical properties of the test item given, which you deem indispensable for judging the data (see explanation for examples)?
	Criteria Group II: Test organism characterisation
5	*Is the species given?*
6	Is the sex of the test organism given?
7	Is information given on the strain of test animals plus, if considered necessary to judge the study, other specifications (see explanation for examples)?
8	Is age or body weight of the test organisms at the start of the study given?
9	For repeated dose toxicity studies only (give point for other study types): Is information given on the housing or feeding conditions?
	Criteria Group III: Study design description
10	*Is the administration route given?*
11	*Are doses administered or concentrations in application media given?*
12	*Are frequency and duration of exposure as well as time-points of observations explained?*
13	*Were negative (where required) and positive controls (where required) included (give point also, when absent but not required, see explanations for study types and their respective requirements on controls)?*
14	*Is the number of animals (in case of experimental human studies: number of test persons) per group given?*
15	Are sufficient details of the administration scheme given to judge the study (see explanation for examples)?
16	For inhalation studies and repeated dose toxicity studies only (give point for other study types): Were achieved concentrations analytically verified or was stability of the test substance otherwise ensured or made plausible?
	Criteria Group IV: Study results documentation
17	Are the study endpoint(s) and their method(s) of determination clearly described?
18	Is the description of the study results for all endpoints investigated transparent and complete?
19	Are the statistical methods for data analysis given and applied in a transparent manner (give also point, if not necessary/applicable, see explanations)?
	Criteria Group V: Plausibility of study design and data

20	*Is the study design chosen appropriate for obtaining the substance-specific data aimed at (see explanations for details)?*
21	Are the quantitative study results reliable (see explanations for arguments)?

This multi-criteria evaluation method aims to assess the quality of a piece of data by examining the global score with respect to the ToxRTool criteria. In particular, each ToxRTool criterion incorporated in the tool is awarded one point when the criterion is met. Having evaluated all the 21 criteria, the points are summed up and utilised to attribute Klimisch categories I, II or III to the data via the use of a scoring function (only Klimisch categories I, II and III are considered in the ToxRTool). If the summation is less than 15, the data is categorised as Klimisch category III (Not Reliable); if the total score is between 15 and 18, the data is categorised as Klimisch category II (Reliable With Restrictions); otherwise, the data is categorised as Klimisch category I (Reliable Without Restrictions). Denote the score for ToxRTool criteria as $s_1, s_2, s_3, \ldots, s_{21}$, then the total score calculation approach described above can be concisely represented as a mathematical equation:

$$Q = s_1 + s_2 + s_3 + s_4 + s_5 + s_6 + s_7 + s_8 + s_9 + s_{10} + s_{11} + s_{12} + s_{13} + s_{14} + s_{15} + s_{16} + s_{17} + s_{18} + s_{19} + s_{20} + s_{21}. \quad (1)$$

Due to the cruciality of some criteria, the ToxRTool also provides a revised classification by defining those crucial criteria as "red criteria" (indicated by italic in Table 2). In particular, for *in vivo* data, there are 8 red criteria within the total of 21. If any of these red criteria fails, the categorisation of the piece of data is revised to Klimisch category III (Not Reliable) no matter how high the summation of the score is. The total score calculation approach for the revised score can be concisely represented as a mathematical equation:

$$Q' = (s_1 * s_5 * s_{10} * s_{11} * s_{12} * s_{13} * s_{14} * s_{20}) * (s_1 + s_2 + s_3 + s_4 + s_5 + s_6 + s_7 + s_8 + s_9 + s_{10} + s_{11} + s_{12} + s_{13} + s_{14} + s_{15} + s_{16} + s_{17} + s_{18} + s_{19} + s_{20} + s_{21}). \quad (2)$$

3 Construction of the Expert System

An expert system is a computer program that simulates the decision making process of human experts. The system is able to assist human experts during problem-solving processes for those problems which are too complex to be solved by human experts. It is also able to act in the place of the experts in the situations where expertise is lacking. Typically, an expert system contains two fundamental parts: a knowledge base which represents the captured expertise (or experience) of human experts, and an inference engine which infers results by consulting the knowledge base. For a given set of inputs, the inference engine first pre-processes the inputs, then consults the knowledge base with respect to the inputs, and finally generates an output.

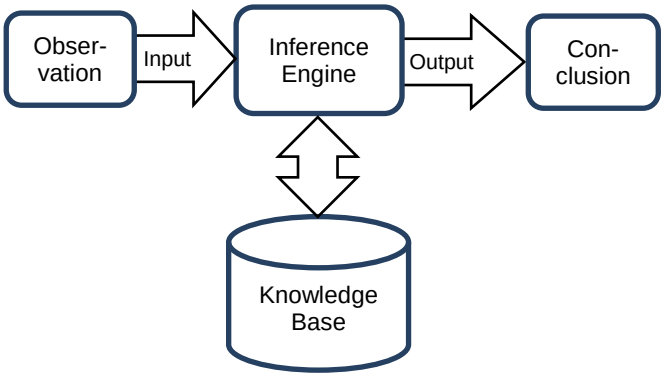

Figure 1. The structure of an expert system

For toxicologists, quality evaluation is a time-consuming and error-prone task. What is more, the rationales behind the decisions generated by human experts vary from toxicologist to toxicologist. A fuzzy expert system may help by: 1) assisting human experts in complex situations; 2) helping human experts to reduce the number of errors; and 3) making decisions based on a set of transparent standards with detailed explanations. In order to construct an expert system for toxicological data quality evaluation, of course a knowledge base and an inference engine need to be built. Prior to these, it is important to properly represent the information and knowledge.

3.1 Domain Knowledge Representation

Effective knowledge representation is essential for the development of any artificial intelligent system, and expert systems are no exception. As stated previously, an explicit drawback of the Klimisch approach is that all the domain knowledge is only represented (and thus processed) as Boolean values which is not sufficient. In order to ease or eliminate the drawback, this paper extends the approach by considering the satisfication of ToxRTool criteria as fuzzy sets[19] where not only partial satisfication is allowed, but also the uncertainty of the partial satisfication is taken into consideration.

Briefly, a fuzzy set is a class of objects with a continuum of grades of membership. Such a set is characterised by a membership (characteristic) function which assigns to each object a grade of membership ranging between 0 and 1.[19] The set of objects whose membership is 1 is called the core, and the set of objects whose membership is greater than 0 is called the support. For instance, for the fuzzy set A shown in Figure 2, the membership of object 0.7 is 0.5, denoted as $\mu_A(0.7) = 0.5$. The core of fuzzy set A is 0.6, and the support of A is [0.3, 0.8]. A fuzzy set is said to be normal if there is at least one member of the set whose membership is 1. Given any two member points within a fuzzy set, if the membership of any member point between the two given points is greater than the minimum of the memberships of the two given member poinits, this fuzzy set is said to be convex. The fuzzy set given in Figure 2 is normal and convex.

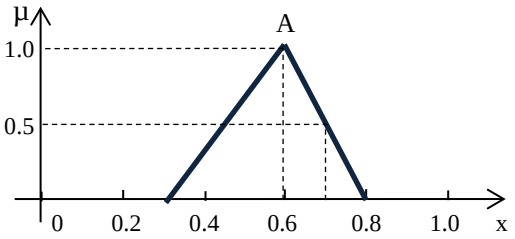

Figure 2. An example fuzzy number representing satisfaction degree

Fuzzy numbers are normal and convex fuzzy sets of the real line, which extends regular numbers in the sense that each fuzzy number refers to a connected set of possible values rather than one single value. The closed interval between the minimum of support and the minimum of the core is called the left support, and the closed interval form the maximum of the core to the maximum of the support is called the right support. For instance, the fuzzy set A shown in Figure 2 may be interpreted as "about 0.6", where the left support is [0.3, 0.6], and the right support is [0.6, 0.8]. The fuzziness of a fuzzy set describes to what degree the concerned concept is not exact, which is characterised by its membership function. Intuitively, the wider the support of a fuzzy number is, the more uncertain the core of the fuzzy number is.

Usually, fuzzy numbers with simpler shapes of membership functions have more intuitive and more natural interpretation. Also, such fuzzy numbers can be easily processed.[20] In this work, triangular fuzzy numbers are utilised to represent the satisfication degree of the ToxRTool critia for efficiency and transparency purposes. Without losing generality, a triangular fuzzy number can be represented as a triple (a_1, a_2, a_3), where (a_1, a_3) and a_2 are the support and core of the fuzzy number, respectively. The core of a satisfication fuzzy number indicates the most likely satisfication degree for a given chemical compound with respect to a particular ToxRTool criterion. The fuzziness of a satisfication fuzzy number indicates the uncertainty of that satisfication degree. For instance, fuzzy set A shown in Figure 2 is a triangular fuzzy number, which can be interpreted as "the satisfaction degree is about 0.6 with certainty degree 0.5".

Reversely, given the most likely satisfaction degree s ($0 \le s \le 1$), and its corresponding certainty degree c ($0 < c \le 1$), a fuzzy number can be generated accordingly as follows:

$$\begin{cases} a_1 = \mathrm{s} - \mathrm{s} * (1 - \mathrm{c}) \\ a_2 = \mathrm{s} \\ a_3 = \mathrm{s} + (1 - \mathrm{s})(1 - \mathrm{c}) \end{cases} \quad (3$$

where the core of the fuzzy number is the most likely satisfaction degree s. Since certainty degree c represents to what extent the most likely satisfaction degree s is true, the uncertainty degree uc then is $(1 - c)$. If $c = 1$, that is $uc = 0$, the satisfaction degree is then definitely s and not possible to be anything else. In this case, fuzzy number (s, s, s) is used to represent the satisfaction situation (where a crisp number is viewed as a special case of triangular fuzzy number. In contrast, if $c = 0$, that is $uc = 1$, the satisfaction degree s is not assignable because there is no relevant information at all. In order to represent this special situation, a Boolean mark is also attached to each ToxRTool criterion besides the most likely score and its certainty degree. Of course, if any criterion is marked, its corresponding s and c are artificially fixed as 0 and 1, respectively, during the calculation of the total score to represent the total ignorance.

Notice that the score of some of the ToxRTool criteria only can be fully certain or fully uncertain, and cannot fall in between. Therefore, the certainties of such ToxRTool criteria are fixed as 1. For instance, criterion 6 is about the sex of animal, and only yes (score 1) or no (score 0) can be answered with full certainty degree. This also applies to criteria 10.

3.2 Knowledge Base

The knowledge base of an expert system contains the expert experience or knowledge in the concerned application domain. There are basically two main ways to construct a knowledge base for a given problem. The first way is directly translating expert knowledge into rules[21] Because rules are representations of expert knowledge, the knowledge base built in this way offers a high semantic level and a good generalisation capability. However, for complex systems, it is difficult to build a knowledge base in such a way due to the sophistication in understanding the mechanism, which has led to another approach of knowledge base construction. This approach is driven by data, which is the knowledge represented in the knowledge base is obtained from data by machine learning techniques rather than human expert knowledge.[22] In contrast to the knowledge base generated from expert knowledge, the knowledge bases built in this way lack comprehensibility and transparency.

The knowledge base in the ToxRTool is built by representing the knowledge of human experts. Suppose that the aggregation result from Equations 1 and 2 are Q and Q' respectively, then the expert knowledge for the ToxRTool can be summarised as follows:

- **IF** $Q \le 12$ ($Q' \le 12$ when red criteria are considered), **THEN** the category is "Not Reliable";
- **IF** $13 \le Q \le 17$ ($13 \le Q' \le 17$ when red criteria are considered), **THEN** the category is "Reliable With Restrictions";
- **IF** $Q \ge 18$ ($Q' \ge 18$ when red criteria are considered), **THEN** the category is "Reliable Without Restrictions".

In order to better mimic the toxicological data evaluation processes conducted by human experts, Equations 1, 2 and these rules (knowledge) need to be extended in order to represent and consider "uncertain" or "fuzzy" satisfaction degrees of ToxRTool criteria. Notice that the Klimisch categories given in Table 1 are linguistic-based which usually do not have crisp boundaries as defined in the ToxRTool. Also, fuzzy sets have widely been utilised to represent linguistic-based values, and the calculi for fuzzy sets have been well developed to compute on linguistic words. The Klimisch categories in this work are represented as trapezoidal fuzzy sets, as illustrated in Figure 3. That is, $A = (0, 0, 12, 13)$, $B = (12, 13, 17, 18)$, and $\mathrm{C} = (17, 18, 21, 21)$.

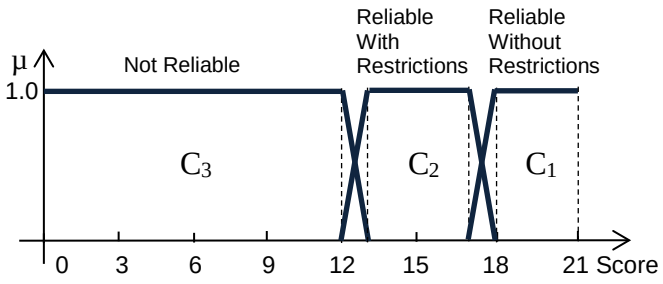

Figure 3. Klimisch categories represented in fuzzy sets

In this figure, the "Score" axis represents the overall marks calculated based on Equation 1 or 2 (which will be fuzzified in the next section); and the "μ" axis indicates the membership of any particular crisp score to the Klimisch categories. For a given piece of data, suppose the score calculated according to Equation 1 or 2 is $\boldsymbol{Q}$ and $\boldsymbol{Q'}$, respectively**.** The extended knowledge base then can be interpreted as follows:

- **IF** Q (or Q' when red criteria are considered) matches fuzzy set C_1, **THEN** the category is "Reliable Without Restrictions";
- **IF** Q (or Q' when red criteria are considered) matches fuzzy set C_2, **THEN** the category is "Reliable With Restrictions";
- **IF** Q (or Q' when red criteria are considered) matches fuzzy set C_3, **THEN** the category is "Not Reliable".

3.3 Inference Engine

An inference engine is a computer program that infers, which is driven by a set of inputs based on a knowledge base. Particularly in the proposed toxicological prediction expert system, the inference engine first pre-processes the inputs and then draws a conclusion based on the input domain values by consulting the knowledge base.

3.3.1 Score Aggregation

In order to make use of the knowledge base, the inputs (fuzzy scores for the ToxRTool criteria) need to be aggregated based on Equation 1 or 2 (if red criteria are taken into consideration). However, the aggregation approach given in Equations 1 and 2 is only suitable for crisp domain values, rather than fuzzy ones. Therefore, Equations 1 and 2 require extensions. Fortunately, this can be readily achieved because arithmetic on fuzzy numbers has been well developed[23] principally based on the extension principle of Zadeh.[19b] When fuzzy arithmetic is operated on fuzzy numbers, the result of the calculations greatly depends on the membership functions (or the shapes) of the operating numbers. Less regular membership functions usually lead to more complicated calculations and thus high computational complexity. This offers another reason why the domain knowledge is represented as triangular fuzzy sets in Section 3.1. Because only multiplication and summation operations are involved in Equations 1 and 2, for simplicity, only these operations on triangular fuzzy numbers are discussed here. Formally, suppose the two fuzzy number operators are A and B, where $A = (a_1, a_2, a_3)$ and $B = (b_1, b_2, b_3)$, then the sum of A and B, denoted as $C = (c_1, c_2, c_3)$, is defined as follows:

$$\mu_C(z) = \bigvee_{z=x+y} \wedge \{\mu_A(x), \mu_B(y)\} \quad \textbf{(4}$$

where ∨ and ∧ represent triangular-conorm and triangular-norm operators, respectively. In particular, ∨ and ∧ are usually implemented by maximum and minimum operators, respectively, which are also the case for this paper. The product of A and B, denoted as D, is calculated by:

$$\mu_D(z) = \bigvee_{z=x*y} \wedge \{\mu_A(x), \mu_B(y)\} \quad \textbf{(5}$$

Based on the extension principle of Zadeh, the adding of two triangular fuzzy numbers leads to another triangular fuzzy number:

$$\begin{cases} c_1 = a_1 + b_1 \\ c_2 = a_2 + b_2 \\ c_3 = a_3 + b_3. \end{cases} \quad \textbf{(6}$$

However, the product of two triangular fuzzy numbers is not with the same kind, that is the resulted fuzzy number from multiplication is not of a triangular fuzzy membership function. Nevertheless, it is a common practice in the real world applications that the product of two triangular fuzzy numbers is also approximated as a triangular fuzzy number. Let $D = (d_1, d_2, d_3)$ be the product of A and B, which can be approximated as follows:

$$\begin{cases} d_1 = \min(a_1 * b_1, a_1 * b_3, a_3 * b_1, a_3 * b_3) \\ d_2 = a_2 * b_2 \\ d_3 = \max(a_1 * b_1, a_1 * b_3, a_3 * b_1, a_3 * b_3). \end{cases} \quad \textbf{(7}$$

If all the fuzzy numbers are greater than 0, that is $x \geq 0, \forall \mu(x) > 0$, Equation 7 can be simplified as:

$$\begin{cases} d_1 = a_1 * b_1 \\ d_2 = a_2 * b_2 \\ d_3 = a_3 * b_3. \end{cases} \quad \textbf{(8}$$

Based on the definition of addition and multiplication operations for triangular fuzzy numbers, Equations 1 and 2 can be readily extended. Suppose that the satisfaction fuzzy numbers for the ToxRtool criteria are $S_1, S_2, S_3, \ldots, S_{21}$, then, if red criteria are not taken into consideration, the aggregated result S is calculated as follows:

$$S = \sum_{j=1}^{21} S_j \quad \textbf{(9}$$

If red criteria are considered, the aggregated result S is calculated as follows:

$$S' = \prod_{i \in \{1,5,10,11,12,13,14,20\}} S_i * \sum_{j=1}^{21} S_j \quad \textbf{(10}$$

Note that, Equations 9 and 10 will degenerate to Equations 1 and 2 respectively, if all the fuzzy inputs $\boldsymbol{S_j}$ degenerate to crisp real values. In other words, if all the score values are

assigned as 0 or 1, and all the certainty values are assigned as 1, the classification results generated by the ToxRTool and the proposed approach will be the same. Then, the sensitivity problem of red criteria in the ToxRTool remains in the proposed approach. However, rather than an unsatisfied red criteria in ToxRTool just leading to Klimisch category III (Not Reliable), the proposed approach also identifies where the problem is and flags it for the user to decide how important the information is. In other words, the new tool allows the user to be more discreet. For certain situations, the approach may be too strict such that every red criterion is flagged, even if the criterion is largely satisfied. In order to address this, the approach may be modified to only flag those red criteria which are not satisfied to a certain degree. This soon provokes a disputation of what is the proper threshold for a given situation, which usually cannot be easily solved. For simplicity, the most cautious approach is applied in this paper, that is any dissatisfaction of red criteria is flagged no matter how small the dissatisfaction degree is.

3.3.2 Consequence Generation

In order to facilitate the discussion, the concept of matching degree is introduced first. When two real numbers are compared in classical mathematics, the two are said to be equal if they exactly match each other; otherwise, they are unequal. However, when two fuzzy numbers are compared, the result usually is not limited to absolute "equal" or "unequal", but could be "equal" to some extent, usually represented as matching degree. A number of approaches have been proposed to calculate fuzzy matching degrees in the literature,[24] which can be typically categorised into two classes: geometric distance-based measures[25] and set theory-based measures.[26] The former are the extensions of the classical concept of metric space and the associated distance function. One example case of this is to extend the Euclidean distance between two points to a fuzzy distance between two fuzzy sets where fuzzy sets can be artificially viewed as points with vague edges. In contrast, the latter are built on the basis of set operations. This approach is rooted in the assertion that the assessment of similarity may be better described as a comparison of features rather than a computation of metric distance between points.[26b] In this case, similarity among objects is expressed as a linear combination of the measures of their common and distinct features, which degenerates to set operations when special parameters are chosen. For simplicity, set theory-based matching degree is utilised in this work and the employment of the distance-based measure remains as an active research area. Particularly, suppose the universe of discourse is X; the matching degree between two fuzzy sets H and I, denoted as $M(H, I)$, is defined as:

$$\mathrm{M(H, I)} = \sup_{x \in X} \{ min\big(\mu_H(x), \mu_I(x)\big)\}, \qquad \textbf{(11}$$

where $\mu_H(x)$ and $\mu_I(x)$ are the memberships of x to fuzzy set H and I, respectively.

Having generated the knowledge base in Section 3.2, and been able to calculate the overall fuzzy score to the ToxRTool criteria as described in Section 3.3.1, the next step is to draw a conclusion based on the overall score by consulting the knowledge base. In conventional classification, an object can either fully belong to a category if the features of the object completely match the features of the category, or not belong to the category at all if the features of the object do not match all the features of the category, but not in between. The situation is different for fuzzy classification where an object could belong to several categories to different degrees, as the concept of matching is a matter of degree which represents to which extent the features of the category are matched. In other words, the fuzzy classification of a piece of given toxicological data with respect to its reliability (or inherent quality) can be achieved by computing the matching degree between the calculated fuzzy number representing the overall score and the fuzzy number representing the Klimisch categories. Then, the knowledge base given in Section 3.2 can be rewritten as:

- **IF** S (or S' when red criteria are considered) matches fuzzy set C_1 to the degree d, **THEN** the category is "Reliable Without Restrictions" to the degree d;
- **IF** S (or S' when red criteria are considered) matches fuzzy set C_2 to the degree d, **THEN** the category is "Reliable With Restrictions" to the degree d;
- **IF** S (or S' when red criteria are considered) matches fuzzy set C_3 to the degree d, **THEN** the category is "Not Reliable" to the degree d.

Based on Equation 11, the classification process is straightforward. For a given piece of data, if the overall score is a crisp number 12.2, the data can be categorised as "Not Reliable" to the degree of 0.8; can be categorised as "Reliable With Restrictions" to the degree of 0.2; and can be categorised as "Reliable Without Restrictions" to the degree of 0, as shown in Figure 4. If the overall score is a triangular fuzzy number (16.2, 18.2, 19.3), the data can be categorised as "Not Reliable" to the degree of 0; can be categorised as "Reliable With Restrictions" to the degree of 0.6; and can be categorised as "Reliable Without Restrictions" to the degree of 1, as shown in Figure 4.

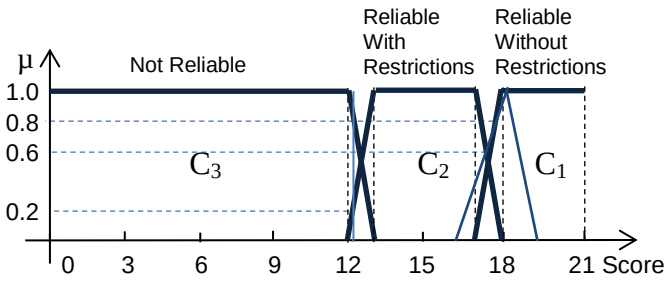

Figure 4. The degrees to Klimisch categories for crisp overall score 12.2 and fuzzy overall score (16.2, 18.2, 19.3)

3.4 The Assignment of Category IV

It is possible that there is no available information at all for some criteria for a given piece of toxicological data as discussed in Section 3.1. Notice that partial certainty also indicates the lack of information. These considerations then can be utilised to implement the assignment of Klimisch Category IV, that is Not Assignable. Formally, suppose that the certainty factors for every ToxRTool criteria are $c_1, c_2, \ldots, c_{21}$, where the certainty factor is artificially assigned as 0 if there is no information at all for the corresponding criterion. The satisfaction degree of the piece of data in question to the

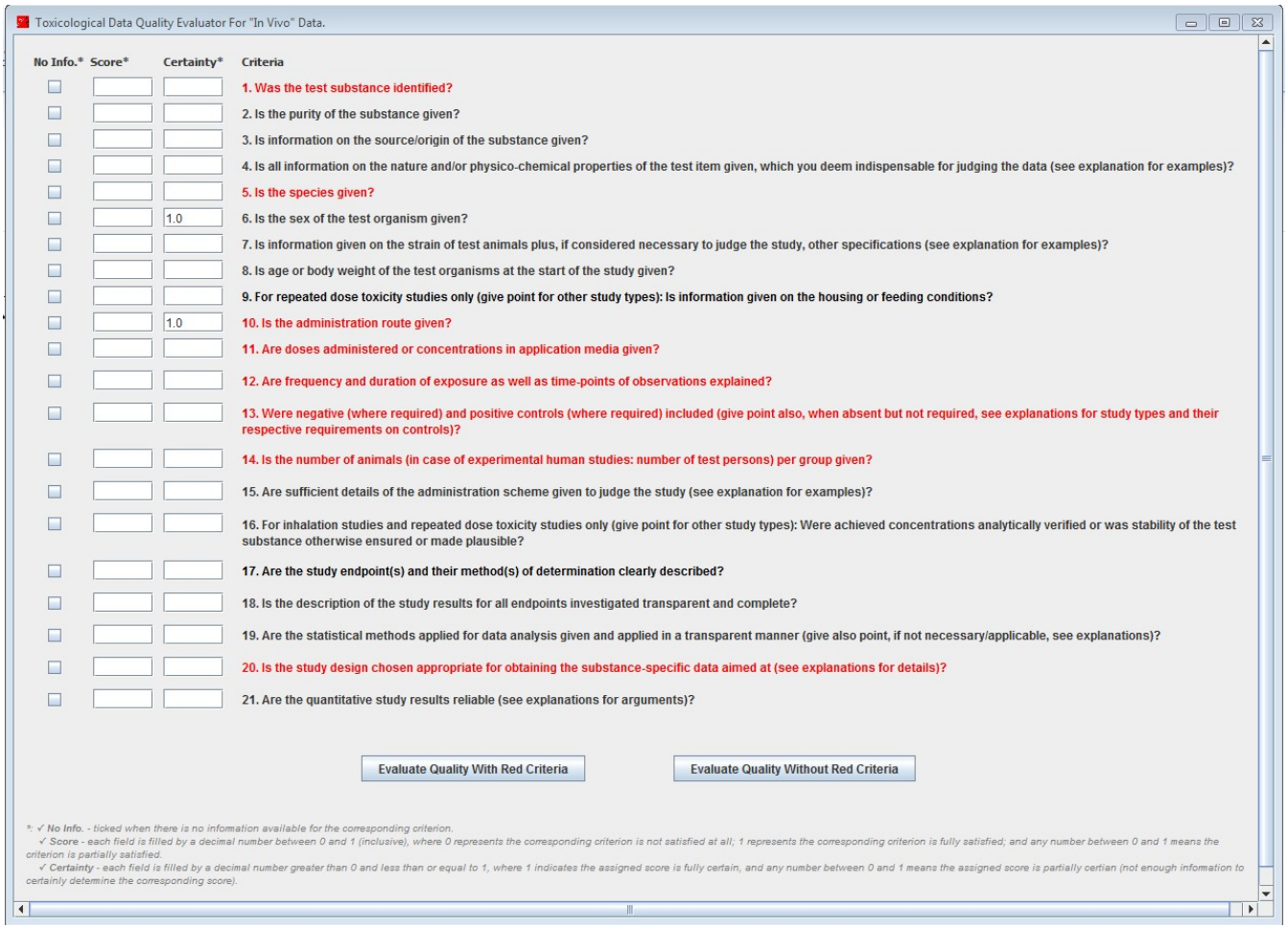

Figure 5. Screen shot of the Toxicological Data Quality Evaluator

Klimisch category IV (Not Assignable) is calculated as:

$$\mathrm{d} = \frac{21 - \sum_{i=1}^{21} c_i}{21}. \qquad (12$$

Because there does not exist an explicit threshold about the missing information related to Klimisch Category IV, the provision of the satisfaction degree to Klimisch category IV (Not Assignable) only aims to inform the users to what extent the data is subject to be classified as Klimisch category IV. Based on this, the final users could then decide whether the missing information is really relevant to the data in question or the purpose for which it is to be used. If the user decides to ignore the missing information, the result given based on Section 3.3 then can be utilised; otherwise, if the user decides the information that is missing is too important, then Klimisch category IV can be assigned.

3.4 Discussions

In comparison to the Klimisch or the ToxRTool approaches, for a given piece of toxicological data, the proposed approach may result in more than one Klimisch category being assigned, each assisted by a satisfaction degree. This is useful as it may help to more accurately represent the true level of confidence in the data. For example, if a piece of data is categorised into multiple Klimisch categories with similar satisfaction degrees for each category , then it is not clear to which category the data should be assigned. In this case, the data is of great variability. However, if the piece of data is categorised into only one Klimisch category (or into multiple categories but only one with a high degree of satisfaction), then there is greater confidence that the data has been assigned to the most appropriate category. The usefulness of the data can then be determined based on which category the data predominantly belongs to.

4 Case Studies Based on the Expert System

The above proposed approach for *in vivo* data has been implemented in Java, named as Toxicological Data Quality Evaluator. A snapshot of the main window is shown in Figure 5. All the criteria are directly adapted from ToxRTool and the explanations of these criteria (as given in the ToxRTool program) are also suitable for the current program. The first column of tick boxes are used to indicate if the information with respect to certain criteria is totally missing. If any of these tick boxes is selected, the corresponding score and certainty fields are then greyed out (not able to be filled) and marked as “N/A” (not applicable), because the related information is not stated at all. Otherwise, each score field needs to be filled by a decimal number between 0 and 1, which represents the most likely satisfaction degree of the corresponding score; and each certainty field is filled by a decimal number which is greater than 0, and less than or equal to 1, indicating the certainty degrees of the assigned scores based on the available information.

Table 3. Assignment of scores and certainties for studied cases

	Case 1			Case 2.1			Case 2.2			Case 2.3		
Criteria	No Info.	Score	Certainty	No Info.	Score	Certainty	No Info.	Score	Certainty	No Info.	Score	Certainty
1		1.0	1.0		1.0	1.0		1.0	1.0		1.0	1.0
2		1.0	1.0		1.0	1.0	√	N/A	N/A	√	N/A	N/A
3		1.0	1.0		1.0	1.0		1.0	1.0	√	N/A	N/A
4		1.0	0.5		1.0	0.5		1.0	0.5		1.0	0.5
5		1.0	1.0		1.0	1.0		1.0	1.0		1.0	1.0
6		1.0	1.0		1.0	1.0		1.0	1.0	√	N/A	N/A
7		1.0	1.0		1.0	1.0		1.0	1.0		1.0	1.0
8		1.0	1.0		1.0	1.0		1.0	1.0		1.0	1.0
9		1.0	1.0		1.0	1.0		1.0	1.0		1.0	1.0
10		1.0	1.0		1.0	1.0		1.0	1.0		1.0	1.0
11	√	N/A	N/A		1.0	1.0		1.0	1.0		1.0	1.0
12		1.0	1.0		1.0	0.8		1.0	0.8		1.0	0.8
13		0.5	0.8		0.5	0.8		0.6	1.0		0.6	1.0
14		1.0	1.0		1.0	1.0		1.0	1.0		0.0	0.5
15		1.0	1.0		1.0	1.0		1.0	1.0		1.0	1.0
16		1.0	1.0		1.0	1.0		1.0	1.0		1.0	1.0
17		1.0	1.0		1.0	1.0		1.0	1.0		1.0	1.0
18		1.0	1.0		1.0	0.9		1.0	0.9		1.0	0.9
19		1.0	1.0		1.0	0.8		1.0	0.8		1.0	0.8
20		1.0	1.0		1.0	1.0		1.0	1.0		1.0	1.0
21		1.0	1.0		1.0	1.0		1.0	1.0		1.0	1.0

Two case studies were conducted: in the first case, all the data instances in a dataset were evaluated by a single run of the software because all of the assay conditions (for the toxicological data used) were the same, while in the second case, toxicological data for three separate chemicals were considered individually, as the assay conditions were not the same for each of the compounds. The former represents the case where one laboratory has reported data on a series of compounds and the latter represents the (common) situation where a dataset reported in the literature is a compilation of data from many sources.

4.1 Case Study 1

The case studies were based on those analysed in the paper by Przybylak et al.[4] Two murine skin sensitisation local lymph node assay (LLNA) datasets were chosen. One of them, the Patlewicz LLNA dataset highlights the simplest situation with a reasonably small dataset which is the output of a single assay and laboratory.[27] This dataset consists of 40 high production volume (HPV) chemicals each with LLNA data. The study endpoint and methodology of determination are clearly described. The results are presented as effective concentration (EC3%) values if they were calculable and the binary classification is given (positive and negative). In the paper there is also basic information about the conduct of experiments. The LLNA assay was carried out according to OECD test guideline 429. There is some information with regard to the selection of three dose levels, but no details about the exact concentrations. Notice that there is also no explicit information about the weight of animals used and also their strain, but the study was conducted according to the OECD guidance, which states the range of the weight as well as the strain; thus, we believe that a standard strain was used. Similarly, there is also no information about positive and negative controls, however according to the OECD guideline (429) for Skin Sensitisation: Local Lymph Node Assay the positive control should be used (although there may be situations, when laboratories can omit testing positive controls and use historic positive control data). This is also the case for experimental design, which is specified by OECD guidance, albeit there is little explicit additional information about it. These details are necessary to assess the data quality according schemes like the proposed approach and its prototype ToxRTool.

Because of the homogeneity of the Patlewicz dataset, all chemicals can be assessed together by giving overall reliabilities and scores for the whole dataset. The tick box conditions and the assigned values for scores and certainties are shown in the second, third and fourth rows in Table 3, respectively. The assessment of this dataset with the Toxicological Data Quality Evaluator gave two different results (Figures 6 and 7) depending on what option of the evaluator was used in terms of red criteria. With red criteria option on, the Patlewicz dataset is assessed as not reliable because of the lack of information about exact dosage (the red criterion) and lack of clear details about the positive control. Conversely, the option without red criteria evaluated the data as reliable without restrictions. This gives an excellent example of the oversensitivity of the red criteria option.

4.2 Case Study 2

In the second case study the dataset used was the LLNA dataset.[28] These data were carefully chosen and the list of chemicals contained in the dataset represents both the

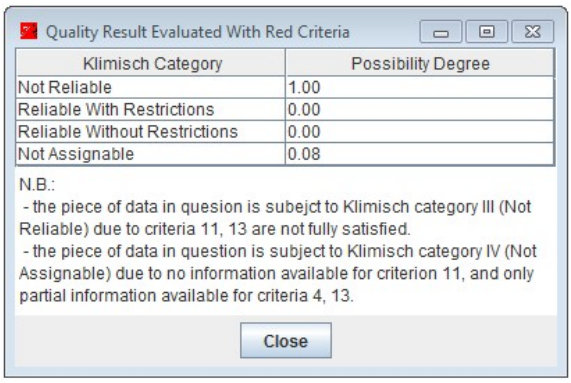

Figure 6. Assessment result of Patlewicz LLNA dataset using the Toxicological Data Quality Evaluator with red criteria

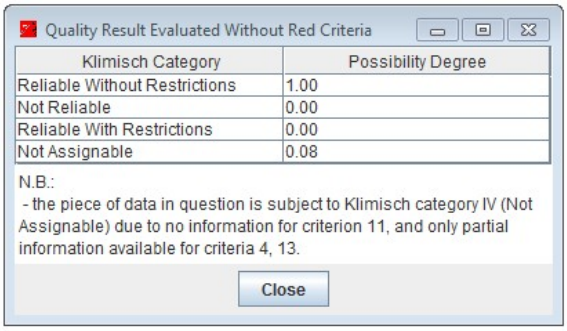

Figure 7. Assessment result of Patlewicz LLNA dataset using the Toxicological Data Quality Evaluator without red criteria

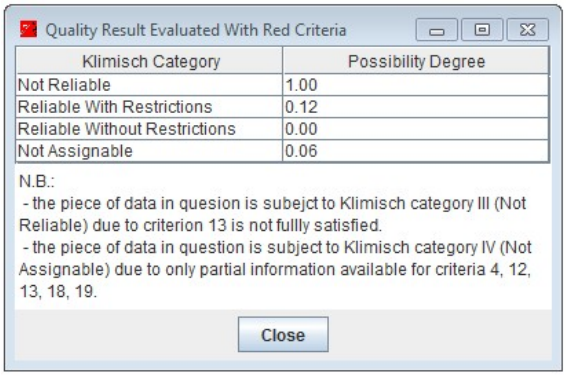

Figure 8. Assessment result of ethyl acrylate using the Toxicological Data Quality Evaluator with red criteria

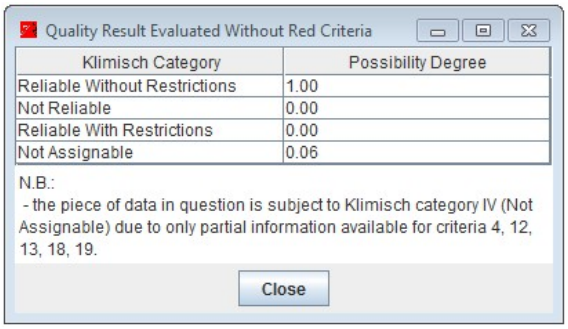

Figure 9. Assessment result of ethyl acrylate using Toxicological Data Quality Evaluator without red criteria

chemical and biological diversity that is known to exist for chemical allergens and non-allergens. The Gerberick dataset is a compilation of 210 chemicals (originally 211, but one compound (3-phenyl-propenal) was removed due to duplication) from 32 original sources. These consist of 29 publicly available papers, a book, two unpublished sources (Unilever and Procter and Gamble) and RIFM (The Research Institute for Fragrance Materials) database. 37 chemicals come from unpublished sources (16 from Unilever, 18 from Procter and Gamble, 3 from RIFM database), making these chemicals impossible to evaluate. A large number of chemicals (93) are retrieved from four publications, whereas eleven publications provide single entries only. For every chemical the following information are recorded: name, chemical structure, CAS number (only three chemicals: bis-1,3-(2',5'-dimethylphenyl)-propane-1,3-dione, C19 Azlactone and methyl 2-sulphophenyl octadecanoate do not have this identification), fundamental physicochemical properties (hydrophobicity, molecular weight, etc.), vehicle, the concentrations, the stimulation indexes (SI) for every concentration, EC3(%), potency category (non-sensitiser, weak, moderate, strong, extreme) and reference.

This dataset poses a much greater challenge as it contains data compiled from numerous sources. That means each compound should be assessed separately as the assessment is likely to differ depending on available information, experimental details and the scheme being used. Therefore, three chemicals from the Gerberick dataset, taken from different sources, were assessed using the proposed approach.

4.2.1 Assessment of ethyl acrylate[29]

In the case of ethyl acrylate, detailed experimental information is available (with the exception of GLP compliance). The necessary information extracted from the source is shown in the fifth, sixth and seventh rows in Table 3. Similarly to the first case study, two options of the Toxicological Data Quality Evaluator gave different results, red criteria assessed the data as not reliable with satisfaction degree of 1 and also as reliable with restrictions with satisfaction degree of 0.12; whereas the option without red criteria evaluated the data as reliable without restrictions (Figures 8 and 9). During the evaluation with red criteria it was noticed that a small change in score for the red criterion, for example from 0.5 to 0.3 can lead to a significant change in the final assessment in terms of satisfactory degree. This shows again that the red criteria are very sensitive and can change the outcome significantly.

4.2.2 Assessment of benzocaine[30]

In the case of benzocaine, much experimental information is available, but GLP compliance and purity are missing. The tick box conditions and the assignment of scores and certainties are listed in the eighth, ninth and tenth rows in Table 3. With red criteria taken into consideration, the data is assessed as not reliable with satisfaction degree of 1; whilst without red criteria taking into account, the data is evaluated as reliable without restrictions also with satisfactory degree of 1 (Figures 10 and 11). The category "not reliable" obtained in the assessment with red criteria is caused by the lack of

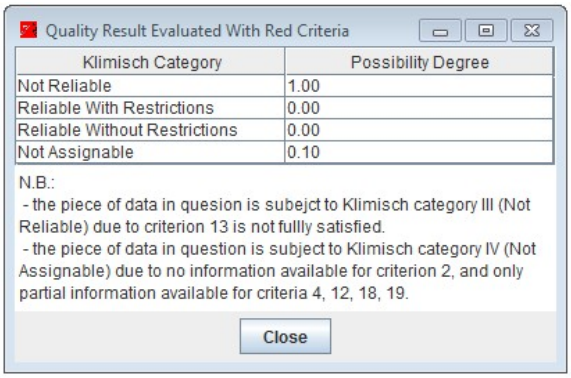

Figure 10. Assessment result of benzocaine using the Toxicological Data Quality Evaluator with red criteria

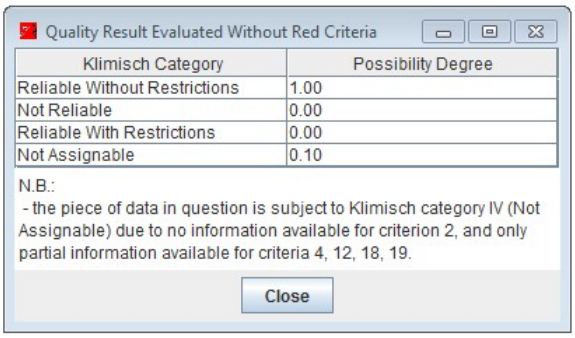

Figure 11. Assessment of benzocaine using the Toxicological Data Quality Evaluator without red criteria

information about negative and positive controls and according to the OECD guidance 429 the details about the concurrent and/or historical positive and negative control data for the testing laboratory should be present.

4.2.3 Assessment of 2-amino-6-chloro-4-nitrophenol[31]

In the case of 2-amino-6-chloro-4-nitrophenol, some experimental information is missing, such as purity, the origin of the substance, the number of animals per group and not certain information about negative and positive controls. The details about the tick boxes and assignment of scores and certainties are listed in the last three rows in Table 3. As a result of the assessment, two different results were obtained, when two options of the Toxicological Data Quality Evaluator were applied. The assessment with red criteria classified the data as not reliable with satisfaction degree of 1, because of the lack of details about positive and negative controls and number of animals per group; whereas the option without red criteria evaluated the data as reliable with restrictions with satisfactory degree of 1 and as reliable without restrictions with satisfactory degree of 0.07 (see Figures 12 and 13).

4.3 Summary and Discussions

We have used the Toxicological Data Quality Evaluator for assessment of skin sensitisation LLNA data. As discussed previously, this tool is an adaptation of the existing ToxRTool. However, the current implementation now benefits from the addition of a Klimisch category IV (Not Assignable), which is flagged when sufficient information is not available to make a quality assessment. In contrast, ToxRTool previously

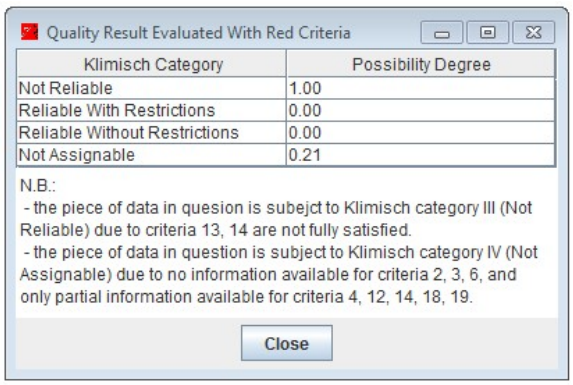

Figure 12. Assessment of 2–amino-6-chloro-4-nitrophenol using the Toxicological Data Quality Evaluator with red criteria

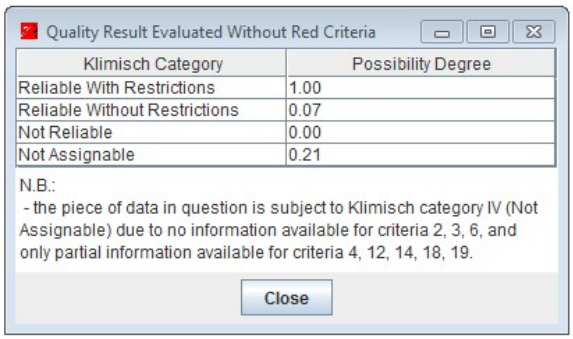

Figure 13. Assessment of 2-amino-6-chloro-4-nitrophenol using the Toxicological Data Quality Evaluator without red criteria

defaulted to Klimisch category III (Not Reliable) in such cases. In addition, the Toxicological Data Quality Evaluator is more transparent and clearly displays to the assessor, which criteria cannot be fully satisfied due to information missing or partial information lacking. This allows the user to closely examine these and decide on their relative importance towards data quality on a case by case basis.

It must be stressed that this tool is still under development. The case studies employed here demonstrate current progress and highlight areas for improvement. We will welcome other groups to use this tool and provide us with feedback in order to streamline future developments.

In this study, we applied the Toxicological Data Quality Evaluator for the assessment of skin sensitisation (LLNA) data using the whole dataset (Patlewicz dataset) as well as three single chemicals taken from Gerberick dataset. The case studies give the examples of two different situations of the data quality assessment: (1) the whole dataset can be assessed together, and (2) the dataset could not be assessed as one; each chemical has to be evaluated separately, as the dataset is a compilation of compounds from different sources. The second example is more common as well as more problematic.

The evaluator consists of 21 questions adapted from ToxRTool. A significant problem in using this tool is that the questions are very specific and require advanced knowledge about the assay of interest. It is quite difficult to properly interpret the questions and give the right score if the assessor does not have a background in toxicology. Thus, the assessment result could be very subjective and of course subject to the background and knowledge of the assessor. In

order to help to minimise the subjectivity, it is worthwhile to associate each score and certainty field a justification to explain why such values are given.

Also, only one option of the Toxicological Data Quality Evaluator should be used - without red criteria. The case studies show that the red criteria are very sensitive and often assign reasonably high quality data to the not reliable category, because of scoring zero for a red criteria question.

What is more, the evaluations of score and certainty can pose problems regarding what decimal values should be assigned. It is very subjective and dependent on the individual interpretation of the criterion. Therefore, instead of the number range (0-1) for certainty, three categories may be more appropriate: high, moderate and low. This will give less variability in the assessment when applied by different users and probably will give more consistent output.

Last but not least, this tool is an adaptation of ToxRTool, which is very "professional" in terms of the content of the questions and therefore poses many problems with the interpretation of the very specific questions. In the ideal scenario, the assessor should have appropriate background to score based on these criteria. People without toxicological and biological knowledge may have difficulties in interpreting and answering these questions, and thus in assigning the correct values of scores and certainties. Therefore, it is evident that detailed guidance on each of the questions and on the assignment of score and certainty values of each question should be very helpful for the assessors. At present, this remains as an area of future work.

5 Conclusions

This paper presented a fuzzy expert system which is able to evaluate the inherent qualities (reliabilities) of toxicological data, based on the currently available metadata. The expert system is developed from the existing data quality evaluation tool ToxRTool, but allowing partial satisfaction degree and also taking the uncertainty of the partial satisfaction degree into consideration, in order to better meet the human experts' appeal. The proposed approach was implemented in Java and evaluated by two case studies. The results not only show the potential of the approach in automatising the quality evaluation processes, but also pose some questions which lead to the directions of future research.

Decision frameworks like ITS require tools to integrate data from heterogeneous sources and to attribute a quantitative measure of the current available data. In this case, Klimisch categories may not serve such purpose well although approaches have been invented to convert the nominal results generated by the Klimisch scheme to numerical numbers.[12, 15] The proposed approach may help in the generation of numerical inherent quality (reliability) values. This can be achieved by: 1) removing the last step of mapping back the total score (represented as fuzzy numbers) to Klimisch category; and 2) defuzzifying the total score represented as fuzzy numbers back to real numbers.[32] Then the resulted real numbers can be directly utilised as the numerical inputs of decision frameworks, such as ITS.

Although the work is promising, the effectiveness of the system is strongly based on the validity of the knowledge base adapted from ToxRTool, and the compatibility between the ToxRTool questions and the proposed system. Therefore, it is worthwhile to further evaluate and improve the knowledge base by conducting more case studies, and to make the ToxRTool questions more compatible with the fuzzification of the system. Also, it is important to let the users understand how the final result is achieved, especially for regulatory purposes. Thus, an explanatory subsystem of the expert system is needed in order to clearly explain the inference processes of the evaluated quality as well as how the score and certainty values are assigned in the first place. In addition, although the working of the program is demonstrated by two case studies, more research is needed to further validate the proposed tool.

Acknowledgements

This work is supported by the EU 7th Framework Integrated Project "Integrated *In Silico* Models for Prediction of Human Repeated Dose Toxicity of Cosmetics to Optimise Safety" (COSMOS)[33] (contract no. 266835) and by Cosmetics Europe. The authors are grateful to the editor and the anonymous reviewers for their constructive comments during the paper revision stage.

References

[1] H. J. Klimisch, M. Andreae, U. Tillmann, *Regulatory toxicology and pharmacology : RTP* **1997**, *25*, 1-5.
[2] K. Schneider, M. Schwarz, I. Burkholder, A. Kopp-Schneider, L. Edler, A. Kinsner-Ovaskainen, T. Hartung, S. Hoffmann, *Toxicology Letters* **2009**, *189*, 138-144.
[3] a) X. Fu, A. Wojak, D. Neagu, M. Ridley, K. Travis, *Journal of Cheminformatics* **2011**, *3*, 24; b) V. Khatri, C. V. Brown, *Commun. ACM* **2010**, *53*, 148-152.
[4] K. R. Przybylak, J. C. Madden, M. T. D. Cronin, M. Hewitt, *SAR and QSAR in Environmental Research* **2012**, 1-25.
[5] http://www.chem.unep.ch/irptc/sids/oecdsids/sidspub.html.
[6] a) C. M. Bishop, *Neural networks for pattern recognition*, Oxford University Press, Inc., New York, NY, USA, **1995**; b) S. Haikin, *Neural Networks: A Comprehensive Foundation*, Pearson Education, **1998**.
[7] a) L. A. Zadeh, *Computer* **1988**, *21*, 83-93; b) G. J. Klir, B. Yuan, *Fuzzy sets and fuzzy logic: Theory and applications*, **1995**.
[8] L. Yang, The University of Wales **2011**.
[9] OECD, *Guidance Document No.34 on the Validation and International Acceptance of new or Updated Test Methods for Hazard Assessment* **2005**.
[10] ECHA, *Guidance on information requirements and chemical safety assessment Chapter R.4: Evaluation of available information and Control* **2011**.
[11] M. Makhtar, L. Yang, D. Neagu, M. Ridley, in *Computer Modelling and Simulation (UKSim), 2012 UkSim 14th International Conference on*, **2012**, pp. 236-241.
[12] M. Balls, in *In Silico Toxicology*, The Royal Society of Chemistry, **2010**, pp. 584-605.
[13] a) W. Lilienblum, W. Dekant, H. Foth, T. Gebel, J. G. Hengstler, R. Kahl, P. J. Kramer, H. Schweinfurth, K. M. Wollin, *Archives of toxicology* **2008**, *82*, 211-236; b) C. J. a. P. Leeuwen, G.Y. and Worth, A.P., in *Risk Assessment of Chemicals* (Ed.: C. J. v. a. V. Leeuwen, T.G.), **2007**, pp. 467-509.
[14] J. Jaworska, S. Hoffmann, *Altex* **2010**, *27*, 231-242.
[15] L. Yang, D. Neagu, in *The 13th IEEE International Conference on Information Reuse and Integration (IRI 2012)*, Las Vegas, USA, **2012**, pp. 295-302.
[16] a) EU, *European Commission, Directorate-General Evironment, Nuclear Safety and Civil Protection: Risk Assessment of Existing Substances, Technical Guidance Document (XI, 919/94-EN)* **1994**; b) EU, *European Commission: Risk Assessment of New and Existing Substances; Technical Guidance Document Draft October* **1995**.
[17] IPCS, *Meeting Report on "International Co-ordination of Criteria Document Production* **1993**, Annex 5.

[18] OECD, *Revised Draft SIDS Manual (OECD Secretariat) EXCH, Manual 9405 DOC July* **1994**.
[19] a) H. J. Zimmerman, *Fuzzy Set Theory and Its Applications*, Kluwer, **1991**; b) L. A. Zadeh, *Information and Control* **1965**, *8*, 338-353.
[20] J. Fodor, B. Bede, in *in Proceedings of the 4th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence (SAMI 2006)*, Herlany, Slovakia, **2006**, pp. 54-68.
[21] E. H. Mamdani, S. Assilian, *International Journal of Man-Machine Studies* **1975**, *7*, 1-13.
[22] a) C. Batur, V. Kasparian, in *Systems Engineering, 1989., IEEE International Conference on*, **1989**, pp. 411-414; b) J. L. Castro, J. M. Zurita, *Fuzzy Sets and Systems* **1997**, *89*, 193-203; c) S.-M. Chen, S.-H. Lee, C.-H. Lee, *Applied Artificial Intelligence* **2001**, *15*, 645-664; d) L. X. Wang, J. M. Mendel, *Systems, Man and Cybernetics, IEEE Transactions on* **1992**, *22*, 1414-1427; e) Y. Yuan, M. J. Shaw, *Fuzzy Sets and Systems* **1995**, *69*, 125-139.
[23] a) A. Kaufmann, M. M. Gupta, Nueva York, EUA : Van Nostrand Reinhold, **1991**; b) D. Dubois, and Prade, Henri, in *Analysis of Fuzzy Information, Vol. 1*, CRC Press,, Boca Raton, **1987**, pp. 3-39; c) S. Gao, Z. Zhang, C. Cao, *Journal of Software* **2009**, *4*.
[24] a) S. Cho, O. K. Ersoy, M. Lehto, *Fuzzy Sets and Systems* **1992**, *49*, 285-299; b) R. Zwick, E. Carlstein, D. V. Budescu, *International Journal of Approximate Reasoning* **1987**, *1*, 221-242; c) L. Yang, Q. Shen, *Fuzzy Systems, IEEE Transactions on* **2011**, *19*, 1107-1126.
[25] a) C. P. Pappis, N. I. Karacapilidis, *Fuzzy Sets and Systems* **1993**, *56*, 171-174; b) C. Shi-Jay, C. Shyi-Ming, *Fuzzy Systems, IEEE Transactions on* **2003**, *11*, 45-56.
[26] a) R. Ros, J. L. Arcos, R. Lopez de Mantaras, M. Veloso, *Artificial Intelligence* **2009**, *173*, 1014-1039; b) A. Tversky, *Psychological Review* **1977**, *84*, 327-352.
[27] G. Patlewicz, S. D. Dimitrov, L. K. Low, P. S. Kern, G. D. Dimitrova, M. I. H. Comber, A. O. Aptula, R. D. Phillips, J. Niemelä, C. Madsen, E. B. Wedebye, D. W. Roberts, P. T. Bailey, O. G. Mekenyan, *Regulatory Toxicology and Pharmacology* **2007**, *48*, 225-239.
[28] G. F. Gerberick, C. A. Ryan, P. S. Kern, H. Schlatter, R. J. Dearman, I. Kimber, G. Y. Patlewicz, D. A. Basketter, *Dermatitis* **2005**, *16*, 157-202.
[29] E. V. Warbrick, R. J. Dearman, J. Ashby, P. Schmezer, I. Kimber, *Toxicology* **2001**, *163*, 63-69.
[30] R. J. D. E. V. Warbrick, D. A. Basketter and I. Kimber, *Contact Dermatitis* **2000**, *42*, 164-165.
[31] G. P. E. Estrada, M. Chamberlain, D. Basketter, and S. Larbey, *Chem. Res. Toxicol.* **2003**, *16*, 1226–1235.
[32] W. V. Leekwijck, E. E. Kerre, *Fuzzy Sets and Systems* **1999**, *108*, 159-178.
[33] http://www.cosmostox.eu/home/welcome/.

Received: ((will be filled in by the editorial staff))
Accepted: ((will be filled in by the editorial staff))
Published online: ((will be filled in by the editorial staff))