Downloaded from orbit.dtu.dk on: Oct 21, 2019

DTU Library

Prioritization before risk assessment: The viability of uncertain data on food contact materials

Pieke, Eelco Nicolaas; Granby, Kit; Teste, Bruno; Smedsgaard, Jørn; Rivière, Gilles

Published in:
Regulatory Toxicology and Pharmacology

Link to article, DOI:
10.1016/j.yrtph.2018.06.012

Publication date:
2018

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
Pieke, E. N., Granby, K., Teste, B., Smedsgaard, J., & Rivière, G. (2018). Prioritization before risk assessment: The viability of uncertain data on food contact materials. *Regulatory Toxicology and Pharmacology*, *97*, 134-143. https://doi.org/10.1016/j.yrtph.2018.06.012

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Accepted Manuscript

Prioritization before risk assessment: The viability of uncertain data on food contact materials

Eelco N. Pieke, Kit Granby, Bruno Teste, Jørn Smedsgaard, Gilles Rivière

PII: S0273-2300(18)30168-5

DOI: 10.1016/j.yrtph.2018.06.012

Reference: YRTPH 4153

To appear in: *Regulatory Toxicology and Pharmacology*

Received Date: 6 April 2018

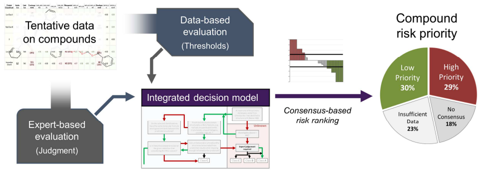
Revised Date: 13 June 2018

Accepted Date: 16 June 2018

Please cite this article as: Pieke, E.N., Granby, K., Teste, B., Smedsgaard, Jø., Rivière, G., Prioritization before risk assessment: The viability of uncertain data on food contact materials, *Regulatory Toxicology and Pharmacology* (2018), doi: 10.1016/j.yrtph.2018.06.012.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Prioritization before Risk Assessment: the viability of uncertain data on food contact materials

Eelco N. Pieke[*a], Kit Granby[a], Bruno Teste[b], Jørn Smedsgaard[a], Gilles Rivière[b]

[a] Technical University of Denmark, National Food Institute, Research Group for Analytical Food Chemistry, Kemitorvet Building 202, 2800 Kgs. Lyngby, Denmark

[b] French Agency for Food, Environmental and Occupational Health & Safety (ANSES), 14 rue Pierre et Marie Curie, 94701 Maisons-Alfort Cedex, France

The views and opinions expressed in this article are those of the authors.

ACCEPTED MANUSCRIPT

[*] Corresponding Author: Eelco Nicolaas Pieke; eelco.pieke@gmail.com; +31 657 151 292

Correspondence for other authors: Kit Granby (kgra@food.dtu.dk), Bruno Teste (bruno.teste@anses.fr), Jørn Smedsgaard (smeds@dtu.dk), Gilles Rivière (gilles.riviere@anses.fr)

Abbreviations: ADI: applicability domain index; CMR: carcinogenicity, mutagenicity, reproductive toxicity; EDA: effect directed analysis; FCM: food contact material; (N)IAS: (non-)intentionally added substances; QSAR: quantitative structure-activity relationship; RA: risk assessment; TMC: total migratable content; TTC: threshold of toxicological concern.

Abstract

The shortage of data on non-intentionally added substances (NIAS) present in food contact material (FCM) limits the ability to ensure food safety. Recent strategies in analytical method development allow investigating NIAS by using chemical exploration; but this has not been sufficiently investigated in risk assessment context. Here, exploration is applied on two paperboard FCM samples followed by risk prioritization for chemicals that can potentially migrate to food. Concentration estimates from exploration are converted into a tentative exposure assessment, while predicted chemical structures are assessed using quantitative structure-activity relationships (QSAR) models for carcinogenicity, mutagenicity, and reproductive toxicity. A selection of 60 chemical compounds from two FCMs is assessed by four risk assessors to classify chemical compounds based on probable risk. For 60% of cases, the assessors classified compounds as either high priority or low priority. Unclassified compounds are due to disagreements between experts or due to a lack of data. Among the high priority substances were high concentration compounds, benzophenone derivatives, and dyes. The low priority compounds contained e.g. oligomers from plasticizers and linear alkane amides. The classification scheme was demonstrated to provide valuable information based on tentative data, able to prioritize discovered chemical compounds for pending risk assessment.

***Keywords**: risk prioritization; FCM; structure assessment; semi-quantification; exposure assessment; hazard assessment*

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

1. Introduction

An all-time debated source of human health risk is the chronic long-term exposure to chemical compounds due to presence in food. One important source of chemicals in food is due their migration from food packaging materials (Arvanitoyannis and Bosnea, 2004; Castle, 2006; Grob, 2014; Jickells, 2007). Investigations into the safety of food contact materials, especially those non-harmonized in legislations affirm that thousands of possible chemicals may be present in paper and board packaging alone (Bengtström et al., 2016; Biedermann et al., 2011; Biedermann and Grob, 2013; Binderup et al., 2002; Ozaki et al., 2005; Triantafyllou et al., 2007), while only a minor fraction of these chemicals have been successfully identified and risk assessed (Geueke et al., 2014). In addition, some chemical compounds originating from paper and board have been shown to have biological activity and therefore are of concern to human health (Bengtström et al., 2016; Honkalampi-Hämäläinen et al., 2010; Rosenmai et al., 2017). As a result, packaging contaminates food with uncharacterized chemicals that may exert significant adverse effects (Gallart-Ayala et al., 2013), yet the extent or nature of the chemical migration is not well-defined because it depends on many parameters, e.g., the packaging material, contact type, temperature, and food type (Barnes et al., 2007; Hauder et al., 2013; Poças et al., 2011).

The regular approach to chemicals in food is to perform a specific risk assessment (RA) for each individual chemical, see Figure 1. However, determining the risk character is convoluted when there is a shortage of available data on migrating compounds (Skjevrak et al., 2005). For the commonly investigated Intentionally Added Substances (IAS), there is often data available from prior research or via accredited methods, but for the more elusive Non-Intentionally Added Substances (NIAS), there is rarely relevant data (Driffield et al., 2016; Grob, 2014; Koster et al., 2015; Pivnenko et al., 2015). In fact, most NIAS do not have assigned chemical structures, concentration data, or characterization of hazards, and few methods are capable to obtain these data for such a large group of chemicals. The sheer amount of possible compounds prohibits

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

performing a dedicated safety evaluation on each compound, and it significantly challenges analytical methods to provide adequate data to perform RA. Consequently, some researchers recently concluded that the existing frameworks RA are inadequate to ensure food safety (Muncke et al., 2017).

The knowledge gap for NIAS and other chemical compounds needs to be reduced in order to incorporate them in legislation. We recently investigated the use of explorative methods to discover chemical compounds in FCM and concluded that untargeted analytical strategies are useful and efficient to estimate the concentration and chemical structure of unknowns (Pieke et al., 2018, 2017). However, it is unrealistic to perform comprehensive analysis on all compounds discovered via exploration (Biedermann and Grob, 2013), so some sort of risk prioritization is required to ensure resources are dedicated to compounds most likely to introduce adverse health effects (Barlow, 2009). One of the core requirements of risk prioritization is to determine a risk character of a chemical compound that is in line with common risk assessment (Guillén et al., 2012; Schymanski et al., 2014).

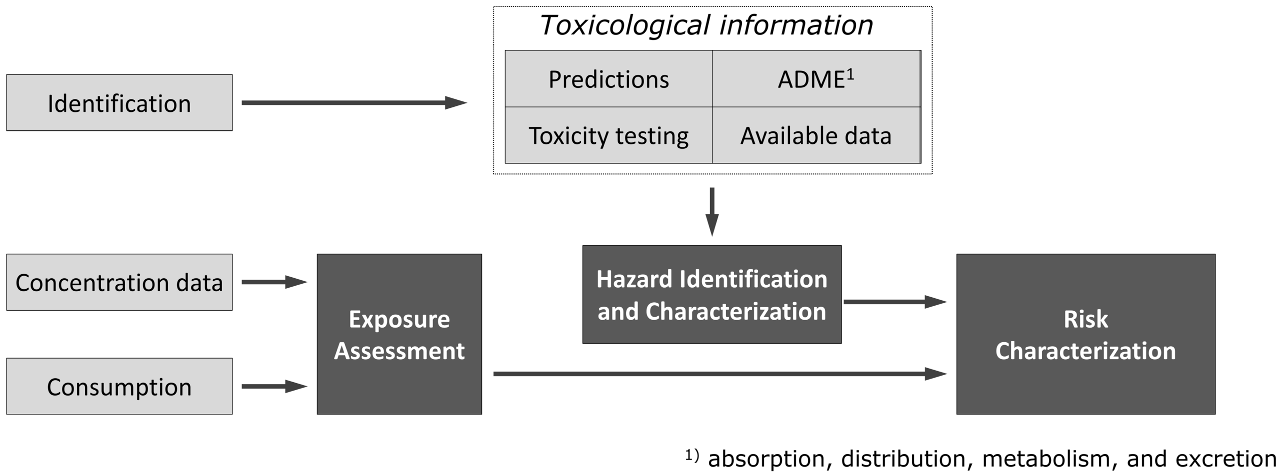


Figure 1: The characterization of risk is a result from highly specific data, which are combined into exposure assessment and hazard identification and characterization. Obtaining the data needed for these assessments is resource-intensive, especially for larger number of compounds with existing data gaps.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

The Threshold of Toxicological Concern (TTC) concept has been adopted within European Union (EU) legislation as a tool to better deal with NIAS and other unknown chemical compounds (EFSA and WHO, 2016; Kroes et al., 2004). The TTC concept uses tentative exposure data to assess if intake is below an accepted threshold of no concern, defined by assigning a Cramer class based on the chemical structure. Hence, TTC is an preliminary assessment tool. It has been applied in a strategy for NIAS discovery by Koster et al. (2014), and may be viable for the exploration approaches shown recently by Pieke et al. (2018, 2017). However, TTC requires compounds to show no genotoxicity (e.g., mutagenicity) or do not exceed an exposure of 0.15 µg $person^{-1}$ day^{-1}. Hence, if the exposure exceeds 0.15 µg $person^{-1}$ day^{-1} genotoxicity testing is required, which is problematic for the large number of compounds that may exceed this threshold. Quantitative Structure-Activity Relationship (QSAR) modeling of chemical hazard may provide substitute toxicity data if testing is prohibitive, which has successfully been applied to FCM for hazard-based assessment and prioritization by van Bossuyt et al. (2017). However, a limitation in hazard-based approaches is that these generally do not always consider occurrence, migration, and exposure.

In present article, we aim to develop a strategy for risk prioritization of chemical compounds in FCM following their prior discovery by exploration strategies. For this, we aim to establish the link between tentative data, e.g., semi-quantification and tentative identification, and existing hazard-based and exposure-based assessment tools, e.g., TTC and QSAR, to perform qualitative risk prioritization. The risk prioritization tool is designed to mimic conventional risk assessment, identically obtaining exposure assessment and hazard assessment, followed by an expertise decision on risk. The tool proposed here is not suggested as a definite method for performing qualitative risk prioritizations, but emphasizes the need and possibility for using tentative data in a risk assessment perspective.

ACCEPTED MANUSCRIPT

2. Methods

2.1. Analysis

Analysis is performed as reported in two previous studies (Pieke et al., 2018, 2017). In brief: UHPLC-MS was performed on an Agilent 1290 system (Agilent Technologies, Santa Clara, CA, USA). Two UHPLC columns were used serially (Phenomenex Luna Omega Polar C18 100 Å, 1.6 µm, 100 x 2.1 mm (Phenomenex, Denmark) and Waters ACQUITY UPLC CSH C18 130 Å, 1.7 µm, 100 x 2.1 mm (Waters, Denmark)). Mass analysis post-UHPLC was performed using an Agilent 6550 Quadrupole-Time of Flight (Q-TOF) mass spectrometer (Agilent Technologies, Santa Clara, CA, USA) equipped with Agilent JetStream electrospray ionization (ESI) interface. The optimization, operating conditions, data collection, and data interpretation are discussed in previous studies (Pieke et al., 2018, 2017).

Semi-quantification was used to determine estimated concentration of chromatographically eluting chemical substances within a threefold error (Pieke et al., 2017). The semi-quantification was limited to the 1200 largest eluting peaks and to detectable analytes in the sample. The chemical structures of compounds in the extract of the sample were tentatively identified by recording fragmentation spectra and using structure correlations to propose a best matching chemical compound (Pieke et al., 2018). The tentative identification results (five predicted chemical structures) were later combined with the semi-quantification results by comparing exact mass and retention times.

2.2. Construction and evaluation of a decision unit

The decision unit for risk prioritization and risk profile classification boundaries was designed by discussing with various interdisciplinary experts at the “Risk assessment for substances and processes submitted to human food regulation” panel at the French Agency for Food, Environmental and Occupational Health & Safety (ANSES). Based on the feedback of the expert

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

panel, the decision unit was designed to involve automation (data-based decisions) and manual assessing (expertise-based decisions).

To test the classification scheme the semi-quantification and tentative identification results of two paper and board FCM samples were used. 30 identified compounds were selected per sample, of which 20 from ESI+ and 10 from ESI−, resulting in a total test set of 60 chemical compounds. The selected entries were evaluated to avoid including chemicals which would not produce a meaningful classification, e.g., no predicted structures. The chemical compounds were gathered in a single Excel-based program available as Supplementary Information. The file contained the predicted structure(s), QSAR consensus and individual prediction by the QSAR models, estimated intake compared to a defined threshold (TTC Excess), absolute estimated intake, and finally the predicted Cramer class from the TTC methodology (Cramer et al., 1976).

The 60 entries were assessed by four individual assessors using the decision unit. Each assessor was tasked with classifying 60 compounds via the decision unit into one of the three risk profile classes: high expected risk ([A]), low expected risk ([B]), or insufficient data ([C]). Prior to classification, each assessor obtained documented instructions on how to work with the Excel program and decision unit. Following, each assessor individually classified the chemical subset. The assessors were specifically instructed to use the decision unit as much as possible, but also to deviate from the decision unit in case their opinion would conflict with the decision unit result.

2.3. Quantitative Structure-Activity Relationship (QSAR)

Possible adverse health effects of tentatively identified chemicals were predicted using Quantitative Structure-Activity Relationship (QSAR) models and software. Three endpoints were defined: Carcinogenicity, Mutagenicity, and Reproductive Toxicity; abbreviated as CMR. To predict CMR activity, the VEGA-QSAR platform (Benfenati et al., 2013) was employed using the included four models for carcinogenicity (CAESAR, ISS, IRFMN/Antares, IRFMN/ISSCAN-CGX), four Ames models for mutagenicity (CAESAR, SarPy/IRFMN, ISS, KNN/Read-Across), and two models for

ACCEPTED MANUSCRIPT

reproductive toxicity (PG Toxicity Library, CAESAR). The VEGA-QSAR platform predicted only the likely activity of the chemical compound, not a dose-response relationship. The Applicability Domain Index (ADI) was used as performance criterion to define the quality of prediction (Istituto di Ricerche Farmacologiche Mario Negri Milano, 2017).

An in-house solution was applied to integrate the QSAR results from VEGA-QSAR into the decision unit. For each prediction the result, active (+) or inactive (−), and the prediction applicability domain index (ADI) were extracted. A QSAR consensus score was calculated from each endpoint results and accompanying ADI score. Each QSAR model applied contributed a fraction to the total consensus score, e.g., for carcinogenicity four models were used, so each model contributed a maximum ±0.25 score. The score was corrected for lower ADI (i.e. prediction certainty) so that less certain prediction had lower weight in the consensus score. The consensus was calculated by the biggest sum for either positive values (active effect) or negative values (no active effect). Hence, the result of the consensus calculation was a value between −1 and +1, in which a negative number indicated a predicted non-active effect and a positive number indicated an active effect. Values closer to the extremes −1 or +1 were results of good agreement between model predictions on the same endpoint; values close to zero indicated a poor agreement.

2.4. Sample selection and preparation

Two paper and board samples were analyzed. The first sample was a recycled unused carton pizza box, similar to the sample used in (Pieke et al., 2018), because these are known to contain many extractable compounds (Bengtström et al., 2016). The second sample was a carton sheet part of the packaging of luxury chocolates. The sheet was folded in a way to compartment separate chocolates, thereby being in contact with the chocolates on multiple sides. The sheet was unfolded before preparing the sample. From each sample, a 10 cm x 10 cm (1 dm^2) sample was prepared using a clean knife.

Each of the 10 cm x 10 cm samples were cut into four identically sized pieces of 2.5 cm x 10 cm and inserted into a glass vessel. The Total Migratable Content (TMC, see 3.1) was recovered by adding 100 mL of warm (40–50 °C) Food Simulant D1 (50 v/v ethanol water) to the vessel. The vessel was closed and sealed into a calibrated thermostat compartment at 40°C. The setup was left to soak for 24 hours, after which the food simulant was removed from the vessel and allowed to cool to room temperature. Proceeding, the food simulant was filtered and prepared for semi-quantification and identification as described in recent work (Pieke et al., 2017).

3. Results and discussion

3.1. The total migratable content (TMC)

TENAX is frequently used for paper and board migration testing, but shows different behaviour for polar and non-polar compounds depending on vapour pressure (Poças et al., 2011). In addition, the use of TENAX implies limited direct contact transfer, but it has been shown that migration from direct contact is not negligible for paper and board and migration can occur even for non-volatile compounds (Biedermann-Brem et al., 2012; Triantafyllou et al., 2007). In addition, there are examples of food contact by paperboard that question the assumption of exclusively dry indirect migration like TENAX simulates, e.g., pizza boxes, snacks, fast food, or fruits (Binderup et al., 2002; Bradley, 2006). Some of the test methods presented in Commission Regulation no. 10/2011 regarding plastic FCM might be used for paper FCM (European Parliament and Council of the European Union, 2011). However, the usage pattern of paper and plastic is different: plastic is often used in longer term storage of a wide array of products, whereas paper is used for shorter contact times or for freezing boxes, e.g., fast food or prepared foods. To use the plastic migration test conditions (10 days at 40°C) on paper materials may not be representative, and this perception is supported by U.S. FDA recommendations proposing migration studies on uncoated paper at 40°C for 24 hours (FDA, 2007). Therefore, here a smaller testing window of 24 hours was used for paper and board FCM.

No migration tests exist for paper and board FCM, so the intended food simulant should ideally have a broad extractable range and compatibility for further analysis by LC-MS. From the analytical and investigative perception in this study these two criteria are met using water/ethanol mixtures. However, using water/ethanol mixtures in contact with paper or board for 24 hours is not a migration test. Noticeably compared to plastics, paper is porous, inhomogeneous, and poorly resistant to liquids, which lead to large numbers of extractives (FDA, 2007). Hence, we consider these testing conditions to be somewhat more severe than a migration test, yet less severe than a complete extraction, as the material integrity is preserved. Instead, we defined the tests performed here as Total Migratable Content (TMC). The TMC contains chemical compounds from the FCM that can reasonably be expected to migrate into food, but is an overestimation of actual-use migration levels. Consequently, TMC implies a thorough screening of extractable chemical compounds, which when observed in the simulant can – but not necessarily will – migrate.

3.2. Risk characterization of tentative data

3.2.1. Tentative hazard identification

A recently published identification strategy allows high-throughput tentative elucidation of the chemical structure of a potentially unknown compound, but does not provide an unambiguous chemical structure, instead presenting several chemical candidate structures (Pieke et al., 2018). Finding existing toxicity data on multiple structures is convoluted. Here, we applied predictive hazard modeling by QSAR. Because QSAR assumes that similar molecules likely have similar effects (Raies and Bajic, 2016), it is compatible with the concept of tentative identification: if the structure prediction closely resembles the actual molecule, the QSAR prediction results will likely be similar. A precaution in using QSAR is that the application of different models can produce different and sometimes conflicting results. To minimize the leverage of a single model in cases where the model performed inadequate, several models are used in parallel for the same endpoint on the same molecule. This presented a battery of results for each prediction, of which the average prediction can cancel the effect of single outliers or false predictions (Benfenati et al., 2013). Consequently, the average prediction of these models (the QSAR consensus) is more likely to contain accurate information than any model alone.

The in-house consensus model closely mimics those presented by the VEGA software (Benfenati et al., 2013). To evaluate thresholds for consensus relevance, the consensus approach was applied on chemical compounds of IARC's Group 1, 2A, and 2B of known, probable, and possible carcinogens list (International Agency for Research on Cancer, 2017). To ensure a strict consensus, the VEGA QSAR results were compared to the assumption chemicals on the extracted carcinogen list (n = 204) are active carcinogens. The threshold for false negative prediction results was set to 2.5%. The results indicated that a consensus score of at least +0.40 was required to minimize the chance of a false negative prediction. This value can be logically evaluated to make sense: +0.40 only be obtained by two or more models predicting the same results, considering the best-case predictions can only contribute +0.25 per model.

Characterizing the hazard as demonstrated here is limited to interpretation of the QSAR evaluation on Carcinogenicity, Mutagenicity, and Reproductive Toxicity (CMR) prediction models. However, there are other toxicity endpoints that influence the probable risk of a substance, e.g., hepatotoxicity, neurotoxicity, or endocrine disruption, but these are not well-studied and few broad range QSAR models exist for these. In addition, CMR is already incorporated in the TTC approach, and a CMR substance has the most strict exposure limit (0.15 µg person^{-1} day^{-1}). Consequently, a reliable CMR alert from QSAR is sufficient to assign the hazard characterization of the substance as a high priority substance.

ACCEPTED MANUSCRIPT

3.2.2. Tentative exposure assessment

Semi-quantification reports a concentration per volume or per surface with a maximum uncertainty of threefold (Pieke et al., 2017). The content per surface area cannot be directly used for assessing exposure, because the contact factor of the FCM is usually unknown. According to European regulations, it is usually considered that an average person has a body weight of 60 kg and consumes 1 kg of food containing the substance daily in contact with a plastic FCM with 6 dm^2 packaging (European Parliament and Council of the European Union, 2011). However, other studies have shown that actual food contact is likely in the range of 10–14 dm^2 (Bouma et al., 2003; Duffy et al., 2007; ILSI Europe Packaging Material Task Force, 1996), and in some cases even higher at 30–40 dm^2 (Bouma et al., 2003). However, paper and board FCM constitute only a limited fraction of 10–20% of the total used packaging materials (Duffy et al., 2007; FDA, 2007). Hence, by applying a usage reduction factor of 10–20% on the worst-case estimate of packaging results in an estimated contact range of 3–8 dm^2 $person^{-1}$ day^{-1}, which is close to EFSA assumptions of 6 dm^2 $person^{-1}$ day^{-1} and likely to be sufficiently conservative. Hence, by adopting the standard used by EFSA, the semi-quantitative concentration data in μg dm^{-2} can be converted to μg $person^{-1}$ day^{-1} by multiplying with 6 dm^2 $person^{-1}$ day^{-1}.

Due to the similarities of the data in this study with that needed in the Threshold of Toxicological Concern (TTC) approach (EFSA and WHO, 2016; Kroes et al., 2004), parts of the TTC strategy are applicable here. Notably, the division of chemical compounds into Cramer classes is useful, because it provides an exposure limit below which likelihood of adverse effect is considered to be very low: for Class I compounds max. 1800 μg $person^{-1}$ day^{-1}; Class II compounds max. 540 μg $person^{-1}$ day^{-1}; and for Class III compounds max. 90 μg $person^{-1}$ day^{-1} (Cramer et al., 1976; Kroes et al., 2004).

It should be noted that the use of the TTC approach for risk assessment is not without criticism (Bschir, 2017). A large number of uncertainties are propagated throughout the TTC approach, where in this study these uncertainties are potentially larger. Hence, as the aim of the study is provide a qualitative human risk ranking of discovered chemical substances, here the TTC is not applied as a tool for preliminary risk assessment. Instead, the TTC approach is applied as means to derive an exposure limit for a tentatively identified chemical compound rather than as risk assessment method. This effectively makes use of the Cramer Class approach, which could be debated as taking into account chronic low dose exposure insufficiently (Bschir, 2017), but provides an estimated exposure limit in case where full identification is not available, as is the case with the results used here. Essentially, other methods that provide exposure limits based on structure may be used if these are found more appropriate, but currently few of these methods exist and are used at the scale at which the TTC is.

Here, the exposure (in µg $person^{-1}$ day^{-1}) from semi-quantification is compared to the limit imposed by the Cramer class assignment calculated from the tentative identification. The result is the TTC Excess factor, which is the fraction of exposure compared to the threshold, i.e., TTC Excess of 100% means the predicted intake is equal the threshold from the TTC approach. However, not every structural prediction was successful where the structure of the chemical compound was unresolved or largely uncertain. For those cases, we considered the worst-case scenario excluding carcinogenicity by assigning Class III. Considering uncertainty of the concentration estimate, worst case ± 3-fold, TTC Excess above 300% would most probably indicate that the TTC would be exceeded, whereas below 33% indicates that most probably the TTC would not be exceeded. Values within this range are to be decided on a case by case basis.

3.3. Risk Prioritization based on tentative data

3.3.1. Risk profile classification

Prioritization based on semi-quantification and structure predictions is convoluted: even if a complete “picture” of exposure and hazard is available, these still contain considerable uncertainty. Consequently, it is not recommended to perform a quantitative risk assessment (RA) on these

ACCEPTED MANUSCRIPT

data, and it should be more feasible to use qualitative risk prioritization, where all variables are evaluated stepwise in order to determine a likely risk profile of the chemical compound. While it would be convenient to classify chemical compounds into subgroups with well-described risk profiles and priority, it is practically less achievable. Here, prioritization is likely to produce only broad risk profiles of chemical compounds, because uncertainties in the estimated hazard and estimated exposure assessment do not support clear boundaries for risk profile classes. The concept for broad classification with uncertain data is not new, as the TTC approach effectively only uses two Cramer classes: Low (Class I) and High (Class III), supplemented by the highly specific Intermediate (Class II).

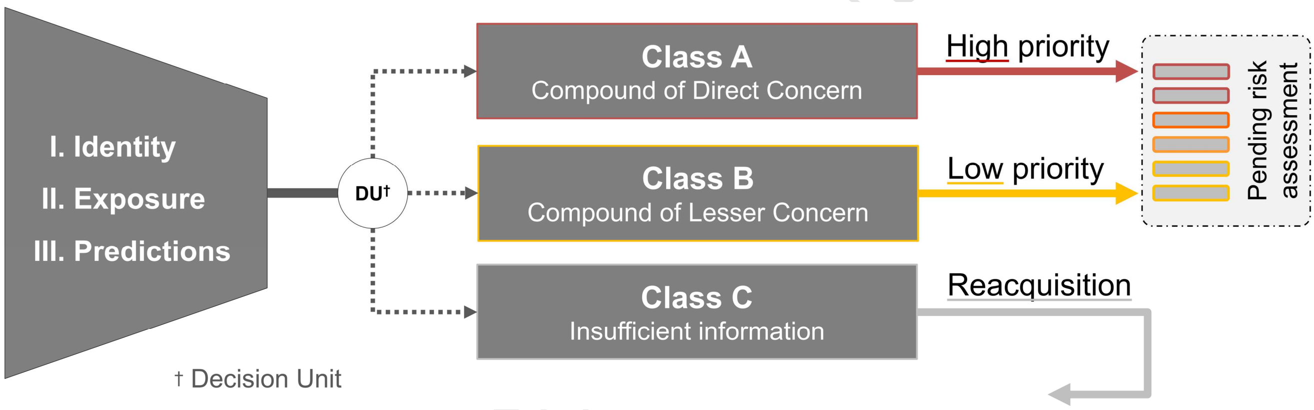


Figure 2: Framework representing one of the possible approaches to incorporate tentative data from exploration in risk assessment principles. The chemical compounds are subdivided into three priority classes following a decision unit (DU), which is an expertise-driven decision tool. The resulting risk profile classes can be used to prioritize further risk assessments.

Only three classes are used in this prioritization approach, shown in Figure 2: [A] — Compound of Direct Concern; [B] — Compound of Lesser Concern; and [C] — insufficient information available. It could be argued that a class between [A] and [B] is needed that defines moderate concern. However, more than two risk profile classes require the capability to define a clear distinction between classes. This is not straightforward due to uncertainty in the data, and a large number of substances might not be classified properly when too many classes are present, thereby making the decision process much more complicated for the assessor. The ultimate goal of risk prioritization is to categorize chemical compounds for probable risk, so a limited decision choice of two risk profile classes was thought to be sufficient for this purpose, whereas for actual risk assessment more classes would be desirable.

3.3.2. Design of a decision unit

To facilitate the assignment of a risk profile class to a chemical compound, a decision unit was designed to incorporate all available data from the tentative exposure assessment and the tentative hazard identification, as shown in Figure 2. The goal of the decision unit is to provide a simple, unified, and reproducible workflow for risk assessors to evaluate input data from exploration experiments into a risk profile classification. Input data for the decision unit consisted of tentative exposure, i.e., estimated intake, Cramer Class exposure limit, and resulting TTC Excess; and tentative hazard identification, i.e., predicted structures, structure correlation scores, QSAR CMR predictions, and QSAR consensus. Due to the tentative nature of the data, the input data can contain variations especially in structure predictions, which affect hazard predictions and intake limit by Cramer class. Hence, it is important to note that small changes in chemical structure may affect different Cramer classifications and exposure limits, so these values should always be seen in context, e.g., evaluation of the actual intake in addition to the TTC Excess.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

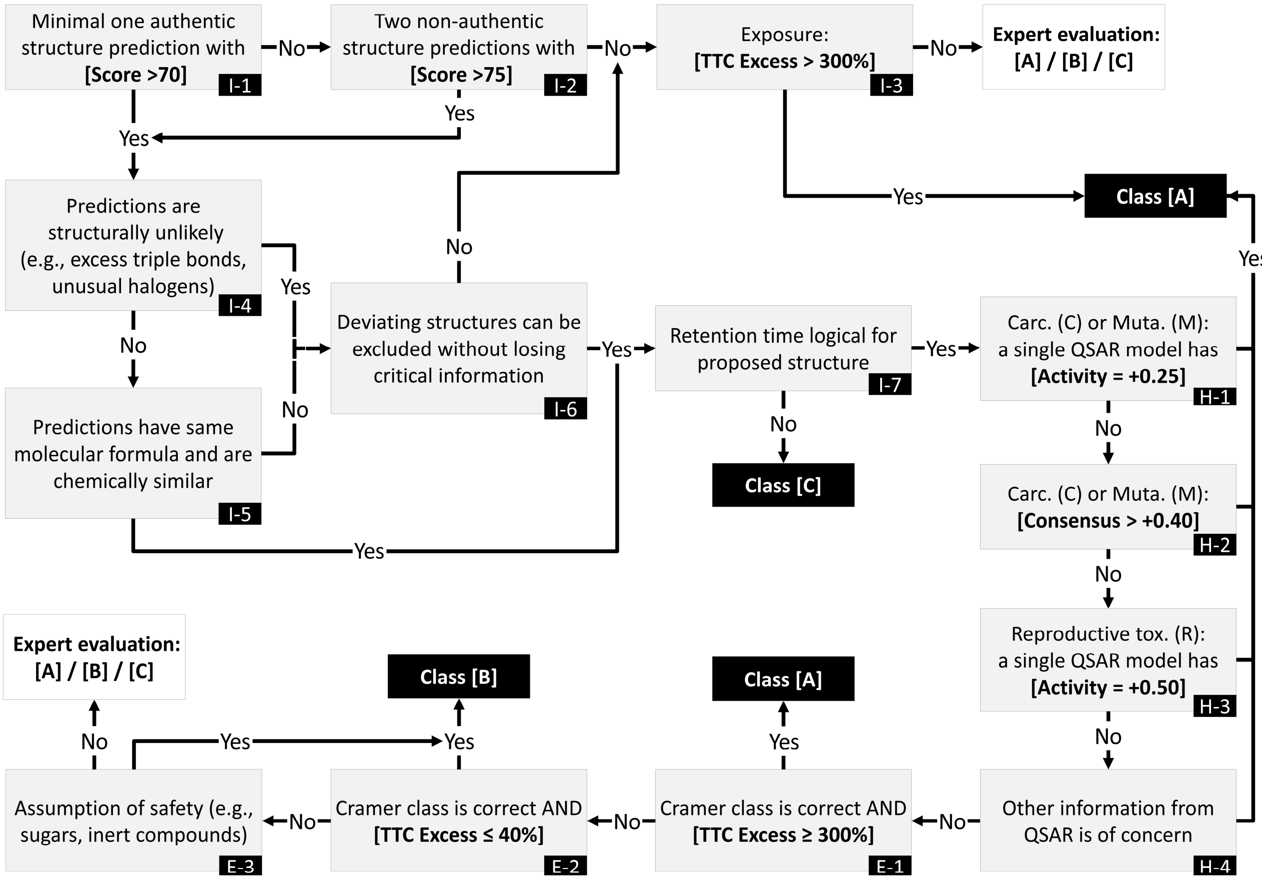


Figure 3: Implementation of a decision unit for risk prioritization. The decision unit is designed as a decision tree that is evaluated by an expert for each node. The result from the decision unit is risk profile class: [A] high priority, [B] low priority, or [C] insufficient data. The risk profile can be determined either data-driven or via expert decision, in which an experienced assessor decides the class based on all available data.

The decision unit was constructed like a decision tree, as shown in Figure 3, built up from the structure prediction by tentative identification (I), the hazard prediction by QSAR prediction (H), and the TTC Excess exposure prediction (E). The decision unit consists of 14 decision-nodes and 6 risk profile classification end-nodes. Decision nodes systematically evaluate all input data and provide a path to the most appropriate end-node. End-nodes within the decision unit result in the assignment of a risk profile to a compound (discussed in 3.3.1), but are always expert judgements. In some cases the end-node provides a non-binding advice for the most likely risk profile class

ACCEPTED MANUSCRIPT

considering the data. In case no such advice is attainable, i.e., where data interpretation cannot unambiguously result in classification, the final risk profile is a decision that needs to be made by the assessor. Consequently, the decision unit is not designed as an automated data evaluation tool although it contains decisions based purely on data, but more as a guide for assessors to stepwise evaluate all available data.

3.3.3. Design of data-driven decision modules

The first nodes in the decision unit are to assess the quality and appropriateness of the predicted structure(s) by tentative identification. The nodes in the tree (see Figure 3) evaluate structure and quality parameters, but require assessor feedback and insight. As discussed in by Pieke et al. (2018) at least one authentic prediction (I–1) or two non-authentic predictions with sufficient prediction score are required (I–2). When there is insufficient chemical structure information, the exposure (TTC Excess) should be evaluated for exceeding the TTC threshold (I–3). When there are sufficient structural predictions, the predicted structures should be evaluated for chemically unlikely features (I–4), molecular mass and similar chemical structure (I–5), and sufficient chemical information (I–6). Finally, the polarity and molecular weight of the predictions should be proportional to the chromatographic retention time (I–7).

Following, the predicted chemical structures are evaluated for exerting possible CMR activity. If there is sufficient QSAR data that suggests CMR activity, the compound is immediately classified as [A]. The QSAR results are checked for experimental data on possible carcinogenicity (H–1), mutagenicity (H–2), or reproductive toxicity (H–3) evident from a maximum reliability score. In addition, the prediction consensus for C and M — but not for R, only limited to two models — is evaluated for exceeding the threshold >0.40 (H–2). The final node is an expert assessment on concerns with the chemical structure regarding hazard, or below-threshold QSAR alerts that promotes concern for safety and should therefore be classified as [A] (H–4).

Finally, the Exposure Module is evaluated by means of the Cramer class and TTC Excess. By comparing the estimated intake with the intake threshold the acceptability of exposure can be decided. First, the intake is assessed compared to the threshold beyond the uncertainty of the exposure measurement, i.e., more than 300% TTC Excess, in which case it will be risk profile [A] (E–1), or below the uncertainty, i.e., less than 40% TTC Excess, in which case it is risk profile [B] (E–2). Next, there is a final expert evaluation node that confirms that the given substance is not known for likely safety, like sugars or inert materials, because the derived TTC limit may be too strict for these, especially since the Cramer classification is often Class III (High) if the structure deviates slightly from a well-known Class I (Low) structure (E–3).

3.3.4. Incorporation of expert decisions

Several nodes in the decision unit are based on human evaluation by requiring expert input. Expert-based decisions are included in the decision unit for two reasons: First, they are a result of discussions with risk assessment expert panels, which summarized that the need for an expert to control the final decision is critical. Second, a simple decision tree is not able to assess the multiplicative effects of several parameters, or capable to assess the data as a whole instead of individually. Hence, expert judgment is required for cases where data obtained by QSAR and/or quantitative methods are inconclusive (Lester et al., 2018). The use of a human assessor within the decision unit fulfills the need for control, but also mitigates the limitations of simplistically-designed decision units, and can thereby help improve decisions. However, it also requires the full attention of a trained risk assessor throughout the entire decision unit, which is problematic with a very large number of substances. Advances in computer sciences, such as advanced machine learning neural networks, may provide an outcome for this in the future (Ru et al., 2017). Consequently, the outcome from the decision unit is codependent on assessor expertise, which in fact closely resembles the methods for traditional RA.

Within the decision unit, the expert assessments are generally called upon in situations where simple data evaluation did not result in a classification. In other words, most expert decisions are needed when no immediate hazard or exposure of concern is detected. In those cases, a comprehensive picture of all available data followed by an expertise decision is required. For example, there may be stacked evidence for classification without exceeding any of the defined thresholds in prior nodes, e.g., a QSAR consensus of 0.39 for both carcinogenicity and mutagenicity. None of the nodes H–1 to H–3 will have marked this compound as a possible risk, but the expertise decision node H–4 likely will via human evaluation. The expert decision nodes at the end of the tree are needed in case iterating through single descriptors such as exposure or hazard identity did not lead to a proper classification. It is impossible to model every likely scenario into the decision unit and retain its accessibility. In addition, an automated decision unit cannot effectively decide whether the available data is sufficient for classification. Expert decisions are consequently the only decisions that can result into a [C] classification for a lack of information.

3.4. Applying risk prioritization explorative data

3.4.1. Application and results of the decision unit

The decision unit (Figure 3) was applied to a set of data obtained from exploration experiments on two different paperboard FCM samples described in the Experimental section. Assessment results of the 60 discovered chemical compounds are summarized in Table 1. The full dataset, which includes all predicted chemical structures, QSAR predictions, and estimated exposure of these 60 compounds are given in the Supplementary Information. Note that the total number of chemical compounds per sample targeted for structural elucidation was 249 for the pizza box sample and 161 for the chocolate box sample, 410 in total, so the 60 compounds represented here are only a fraction of the total number of discovered compounds.

To convert the assessment into risk ranking, a score was calculated based on the assessors' answers. The score is on a scale from −100, low priority, to +100, high priority. Scores near zero were those either that showed no consensus between assessors or where data was inadequate for classification. Calculation of the score is performed according to Equation 1, where n_X represents the occurrence count of each classification x = [A], [B], or [C] per compound. The formula has deliberately not been simplified for clarification: the first part penalizes differences between [A] and [B], while the second part penalizes a lack of consensus. Hence, more contrast in the classification results in a ranking score closer to zero.

$$\mathrm{Rank} = \frac{n_A - n_B}{n_A + n_B + n_C} * \frac{\max[n_A, n_B, n_C]}{n_A + n_B + n_C} \qquad \text{Equation 1}$$

The threshold of priority and no consensus was set at a score of ±30. This marked the point where above which at least three assessors assigned the same risk profile, but one assessor assigned a conflicting profile or indicated insufficient data, e.g., AAAB or AAAC. If an assessment contained two or more entries of [C] these were marked as uncertain, since at least 50% of assessors indicated that available data was not sufficient to take an appropriate decision.

The overall results from the assessment in Table C.1 reveal that approximately 60% of the chemical compounds were eligible for prioritization as a result of the evaluation, while 40% of the substances either have insufficient data for prioritization, or displayed conflicts in assignments by different assessors. The results show an almost even distribution of cases between high priority (29%), insufficient data (23%), no consensus (18%), and low priority (30%). A number of compounds were unanimously ranked by all assessors as high risk or low risk for 13% and 13% of the cases, respectively.

Some illustrative examples of each consensus result are discussed in order to understand some of the choices behind the classification. For a visualization of the chemical structures discussed, the reader is referred to the Supplementary Information.

ACCEPTED MANUSCRIPT

***Table 1**: Assessments results of four assessors on 60 different chemical compounds from two different samples. Assessors were tasked to assign one of three risk profiles to the chemical substance. The ranking score is calculated from the ratio of the risk profiles reaching four different consensus results, where a score of at least ±30 was considered consensus. When two or more assessors assigned [C], the entry was considered to be deficit in information.*

ID.	Sample	ESI Polarity	Ret. time (min)	1	2	3	4	Ranking score	**Consensus**
10	Pizza	ESI−	18.582	A	A	A	A	100	High priority
13	Choc	ESI−	18.572	A	A	A	A	100	High priority
15	Pizza	ESI−	20.385	A	A	A	A	100	High priority
16	Pizza	ESI+	35.247	A	A	A	A	100	High priority
19	Pizza	ESI+	34.313	A	A	A	A	100	High priority
24	Pizza	ESI+	13.270	A	A	A	A	100	High priority
36	Pizza	ESI+	23.521	A	A	A	A	100	High priority
49	Choc	ESI+	2.088	A	A	A	A	100	High priority
1	Choc	ESI+	3.652	C	A	A	A	56.25	High priority
30	Choc	ESI+	24.146	A	A	C	A	56.25	High priority
32	Pizza	ESI+	8.989	A	A	A	C	56.25	High priority
34	Pizza	ESI−	22.801	A	A	C	A	56.25	High priority
45	Choc	ESI+	26.474	A	A	C	A	56.25	High priority
54	Choc	ESI+	15.071	A	A	C	A	56.25	High priority
5	Pizza	ESI+	34.689	A	B	A	A	37.5	High priority
53	Choc	ESI+	8.988	A	A	A	B	37.5	High priority
57	Choc	ESI+	19.585	A	A	B	A	37.5	High priority

ACCEPTED MANUSCRIPT

ID.	Sample	ESI Polarity	Ret. time (min)	1	2	3	4	Ranking score	**Consensus**
39	Pizza	ESI−	27.602	C	A	C	A	25	Insufficient data
20	Pizza	ESI−	11.742	B	A	C	A	12.5	No consensus
40	Choc	ESI−	17.711	B	A	C	A	12.5	No consensus
6	Pizza	ESI+	27.600	B	C	C	A	0	Insufficient data
11	Choc	ESI+	27.018	C	C	B	A	0	Insufficient data
21	Pizza	ESI+	19.644	C	A	C	B	0	Insufficient data
23	Choc	ESI+	10.820	B	A	B	A	0	No consensus
27	Choc	ESI+	24.373	B	A	B	A	0	No consensus
35	Pizza	ESI−	17.712	B	A	B	A	0	No consensus
38	Choc	ESI−	3.473	C	C	C	C	0	Insufficient data
42	Choc	ESI+	30.173	C	C	C	C	0	Insufficient data
43	Choc	ESI+	33.504	C	C	C	C	0	Insufficient data
46	Choc	ESI−	30.398	C	C	B	A	0	Insufficient data
56	Pizza	ESI+	28.307	B	A	B	A	0	No consensus
58	Pizza	ESI+	10.831	B	A	B	A	0	No consensus
25	Choc	ESI+	13.095	C	A	B	B	−12.5	No consensus
26	Pizza	ESI+	3.283	B	B	C	A	−12.5	No consensus
50	Pizza	ESI−	24.446	B	B	C	A	−12.5	No consensus
52	Pizza	ESI+	36.131	B	B	C	A	−12.5	No consensus
2	Choc	ESI+	31.825	C	C	C	B	−18.75	Insufficient data
51	Choc	ESI−	29.098	C	C	C	B	−18.75	Insufficient data
3	Choc	ESI+	16.438	B	C	C	B	−25	Insufficient data

ACCEPTED MANUSCRIPT

ID.	Sample	ESI Polarity	Ret. time (min)	1	2	3	4	Ranking score	**Consensus**
7	Choc	ESI+	14.272	B	C	C	B	−25	Insufficient data
12	Choc	ESI+	27.434	C	C	B	B	−25	Insufficient data
17	Pizza	ESI−	28.550	C	C	B	B	−25	Insufficient data
14	Choc	ESI+	17.416	B	A	B	B	−37.5	Low priority
22	Choc	ESI−	20.588	B	A	B	B	−37.5	Low priority
31	Choc	ESI+	13.335	B	B	B	A	−37.5	Low priority
44	Pizza	ESI−	2.790	B	A	B	B	−37.5	Low priority
60	Choc	ESI−	18.167	B	A	B	B	−37.5	Low priority
4	Choc	ESI−	19.438	B	B	C	B	−56.25	Low priority
28	Choc	ESI−	14.359	B	C	B	B	−56.25	Low priority
33	Pizza	ESI+	15.083	B	B	B	C	−56.25	Low priority
37	Pizza	ESI+	31.396	B	B	B	C	−56.25	Low priority
59	Pizza	ESI−	14.359	B	C	B	B	−56.25	Low priority
8	Pizza	ESI+	27.289	B	B	B	B	−100	Low priority
9	Choc	ESI−	20.881	B	B	B	B	−100	Low priority
18	Pizza	ESI+	26.490	B	B	B	B	−100	Low priority
29	Choc	ESI+	24.943	B	B	B	B	−100	Low priority
41	Pizza	ESI+	28.868	B	B	B	B	−100	Low priority
47	Pizza	ESI+	33.264	B	B	B	B	−100	Low priority
48	Pizza	ESI+	15.083	B	B	B	B	−100	Low priority
55	Pizza	ESI+	34.327	B	B	B	B	−100	Low priority

441

3.4.2. Compounds with high priority

One notable entry unanimously marked as risk profile [A] is ID 10 and ID 13, which is in fact the same chemical compound observed in different samples. The chemical structure suggests a benzophenone-like compound at relatively high exposure levels of 360−445 μg person^{-1} day^{-1}, excluding the three-fold semi-quantitative uncertainty, compared to the 90 μg person^{-1} day^{-1} TTC limit of Cramer Class III compounds, with QSAR alerts that indicate carcinogenicity and mutagenicity. There is another instance of a benzophenone-like compound among the high priority compounds: ID 15, which has a less unambiguous predicted structure which occurs at nearby retention time. The presence of benzophenone compounds in paper and board is known mostly due to recycling of printed board (Anderson and Castle, 2003), so detection of benzophenone-like substances is not unexpected; however, the concentration estimates indicate a relatively high exposure potential. This potential can also be limited by the overestimation in the TMC, but it is nevertheless a compound of concern.

Another entry clearly marked as risk profile [A] is ID 19, which strongly represents an azo dye Pigment Red 2. The chemical compound could exceed TTC limits with an estimated intake of 50 μg person^{-1} day^{-1}, excluding uncertainty, compared to 90 μg person^{-1} day^{-1} defined by Cramer Class III. However, QSAR results clearly indicate a possible carcinogenicity and mutagenicity, which would exempt the compound from Class III limits and instead impose the stricter limit of 0.15 μg person^{-1} day^{-1}. The presence of pigments, especially pigment red, has been observed before (Bengtström, 2014). Azo dyes are capable of breaking down into carcinogenic substances like amines and aromatic amines, which can be cause for concern, e.g., in cosmetic products (SCCNFP, 2002).

Compound ID 32 and ID 53, both marked as risk profile [A], represent an isothiazolinone fungicide compound present in both samples at similar retention times. For the pizza box, it exceeds the maximum exposure significantly: 700 μg person^{-1} day^{-1} excluding uncertainty, but for the chocolate

ACCEPTED MANUSCRIPT

box semi-quantification was unsuccessful. When strictly following the decision unit, the presence in the chocolate box would likely be marked as risk profile [B] or [C] because none of the thresholds are explicitly exceeded, but since it was classified as risk profile [A] by most experts this demonstrates the added value of an expert decision. Here, the expert decision rightly classified the same chemical compound with the same priority, despite differences in available data. This substance has been discovered and more extensively discussed in previous work (Pieke et al., 2018).

3.4.3. Compounds with low priority

ID 47 and 55 represent two compounds marked as risk profile [B], and are chemically similar long-chain amides originating from the pizza box. Estimated intake of these substances is significantly below a TTC Class III compound at 33–38 µg person^{-1} day^{-1}, excluding uncertainty, which is unlikely to exceed 90 µg person^{-1} day^{-1} when including uncertainties. In addition, there are no QSAR alerts for these compounds. These substances have previously been identified (Pieke et al., 2018) in a similar sample, where these were also considered unlikely to pose a risk. The consensus of the risk prioritization here emphasises that the previous assessment is probably correct, and this type of compound is not anticipated to be at risk by different evaluators.

ID 31, ID 33, and ID 48 represent polyethylene glycol (PEG) oligomers, while ID 18 represents dipropylene glycol dibenzoate. These are all commonly used plasticizers. The intake for these compounds is relatively high compared to other compounds listed here, but these compounds are not commonly associated with any hazardous effects. It was shown here that the expert decisions play a critical role in ensuring the proper class assignment, e.g., ID 31 has large TTC Excess values because the compound is marked as Cramer Class III. Despite that, three out of four assessors marked the compound as risk profile [B] because the chemical structure was known to them as a PEG oligomer, for which a Cramer Class I is more likely appropriate. All of these compounds are relatively inert plasticizers with no QSAR alerts, and especially PEG oligomers are unlike to pose a risk at these concentrations.

ID 4, ID 8, ID 37, ID 41, and ID 60 are a number of diverse, yet chemically similar and simple structures that are each marked as risk profile [B] by most assessors, indicating a low priority. These compounds are characterized by a generally low exposure estimate, simple chemical structures composed predominantly of C, H, and O, and few carbon-rings, and rarely contain any QSAR alerts for CMR. A number of the predicted chemical structures are classified as Cramer Class I, which increases the exposure limit significantly to 1800 μg person^{-1} day^{-1}, but for most of these compounds the 90 μg person^{-1} day^{-1} is not exceeded.

3.4.4. Compounds with no assigned priority

The compounds that did not have a prioritization can be separated in two main groups: compounds with insufficient data, or compounds with mixed information containing both elements of high priority and low priority, which prevented consensus. Compounds with insufficient data are marked if at least half of the assessors indicated that the available data is insufficient to assign a risk profile [A] or [B], e.g., ID 51, ID 2, ID 38, and ID 43. These cases are not discussed extensively, but reassessment should only occur upon obtaining additional or improved data.

Interesting cases of non-consensus compounds are ID 23 and ID 58. Some structure predictions seem to indicate a polyethylene glycol (PEG) oligomer, similar to ID 31, ID 33, ID 48 previously discussed. However, the exposure is significantly higher: 1240 μg person^{-1} day^{-1} for ID 58 and 520 μg person^{-1} day^{-1} for ID 23. In addition, some of the predicted structures seem to be PEG derivatives or unrelated structures, which have more severe QSAR alerts and TTC Excess due to being Cramer Class III. Different assessors interpreted this information differently: two considered this a high priority substance and two considered this a low priority substance. Based on the exposure, it is sensible to consider these substances as high priority, but on the other hand the knowledge of PEG oligomers can render these compounds as low priority. Consequently, the lack

ACCEPTED MANUSCRIPT

of consensus can warrant the need for a discussion on the substances to clarify where differences in opinion are originating from.

Another case where no consensus could be achieved is for ID 56. The predicted structures greatly varied between the authentic and non-authentic databases, where structure predictions of the latter appeared unlikely, but the predictions from the first were of relatively low confidence. The estimated exposure was 83 μg person^{-1} day^{-1}, which is close to the limit of 90 μg person^{-1} day^{-1} of a Cramer Class III compound. There are no obvious QSAR alerts that indicated direct concern. Here, the assessment of the compound was primarily based on expert decision, and assessors are unable to agree. ID 56 is an example of a group of compounds that have very little information, or where the information shows conflict between different predictions, so a decision for low- or high-priority is not straightforward. Some other examples are ID 20, ID 26, ID 27, and ID 35. The proper classification of these compounds may require additional information, a stricter decision unit, or more assessors.

In a number of cases some assessors considered the information to be not sufficiently informative, but others tried to give a classification. These cases are characterized by an equal distribution in assessors indicating [C] and [A] or [B]. An illustrative case is ID 3, which initially does not appear to lack information. However, the structure predictions are varying greatly and are accompanied by low confidence, so no good structural image can be obtained. Because there is no structural image, QSAR alerts cannot be considered reliable. Yet, the exposure to this compound is low: 12 μg person^{-1} day^{-1}, which is including uncertainty well below the TTC limit for a Class III compound. As a result, half the assessors indicated [B] for no likely risk due to the low exposure, but the other half indicated [C] likely due to the poor quality of structure predictions. There are some other examples where this occurred, e.g., ID 59, ID 6.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

3.5. The implementation of risk prioritization tools

Based on the results in 3.4, a different course of action is required for each assigned priority and rank score. Compounds that show the maximum rank score do not require much discussion, as these are classified similarly by all assessors, so their priority for risk assessment is fairly unanimous. For compounds that do not score maximally, but are still classified as low- or high-priority (e.g., AAAC), it is suggested to discuss these entries briefly to understand the reason for reaching a less than maximum consensus. Unless there is a good reason to deviate from the advice of the consensus, it should be maintained. Results with rank scores close to zero need to be investigated: either the available data is insufficient or has too many uncertainties, or the assessors disagree on the risk profile. In the first case, insufficient data, more data will need to be gathered or, as discussed in the next paragraph, the quality of results needs to be improved. The latter case, no consensus, requires discussion and is cause for concern. In some cases, the differences occur as a result of data weight: some assessors weigh the exposure heavier than hazards. Disagreements between assessors will need to be better understood in order to improve the decision unit.

Presently designed decision unit was found to be suitable for assessing a small to moderate number of chemical compounds with tentative data. The decision unit is currently an expert-based model in which the decision tree is a helpful tool for the experts to reach classification. The value of the expert decision was shown throughout the data, e.g., in classifying ID 32 and ID 53. However, for a larger number of compounds the workload on the assessors increases similarly, so the current design may not be suitable for a very large number of chemical compounds. For this, automation may be a solution, but discussions with risk assessment experts indicated caution to changing the decision of risk to a fully automated process. In addition, automated decisions for tentative data are complex since they require a multivariate approach that can incorporate multiple uncertainties, whereas it also must be able to derive decisions from experience as humans do.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Hence, while automated decisions are desirable for many compounds, these should be developed with caution to the human expertise needed to classify compounds.

Further needs are to improve the input data on which decisions are based, which will reduce the number of non-consensus and insufficient data prioritizations. For example, the inclusion of more and improved *in silico* models (e.g., Expert Model, QSAR, or hybrids) may allow a better decision process, as more hazards can be included in the assessment possibly with higher prediction certainties. Moreover, the current strategy assesses on a compound-to-compound basis without including mixture effects. Mixture effects are highly complex and the assessment thereof not strongly developed, which make them and this strategy currently incompatible, but may be an interesting addition for future research. To enable the assessment of mixture effects, it could be a possibility to incorporate Effect Directed Analysis (EDA) into the strategy, which could provide toxicity data on mixtures based on chromatographic fractions. This would significantly improve the toxicological basis of risk priority, but it would require substantial pre-decision work, which can negate the speed of the strategy as currently presented.

In addition, a reduction in uncertainty originating from tentative data is beneficial, e.g., lower error in concentration estimation or improved structure predictions. Suggestions for improving the strategies for semi-quantification and tentative identification have been provided in the respective research (Pieke et al., 2018, 2017). Both, however, highlight that these explorative methods are relatively novel in applications, and will need substantial further developments. Finally, the inclusion of more assessors can improve the classification results. More assessors permit more combinations of risk profile assessments, which will improve the amount of ranks that are available, as well as allowing a better investigation and discussion of compounds that did not reach a risk prioritization consensus.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

4. Conclusion

A strategy for risk prioritization based on tentative data is demonstrated for ranking the tentative risk of discovered compounds. This tool is based on a simple and low cost approach. The classification/prioritization of 60 substances was performed in a short time (less than 1h). The strategy is demonstrated to be capable to discriminate sufficiently (>60%) within a test set of 60 compounds between low priority compounds expected not to be of concern, and high priority substances expected to be of concern or demonstrating indications of concern. The tool is validated on compounds previously reported in literature as being of concern, so the strategy is able to sort relevant results. Consequently, the tool can easily be transposed on the total set of 400 compounds discovered by exploration to greatly improve the chemical knowledge on complex samples from a risk assessment perspective.

Currently, the strategy is demonstrated with a limited number of hazard endpoints and assessment is limited to the intake of a single compound at a time. A critical reason for this is the tools needed to assess mixture effects or more advanced toxicity endpoints are currently not sufficiently developed; therefore, these would likely be supported to a lesser degree by risk assessors. If assessors do not trust the predictive models to be accurate, it would limit the effectiveness of the decision tree model. As a result, the demonstrated strategy uses a limited set of QSAR models known to be relatively reliable and focuses on single compounds; yet, this strategy is adaptable to include more predictive models (e.g., endocrine disruption models, mixture toxicity models) and experimental techniques (e.g., EDA) in the decision process, which makes it robust to future developments in the field of structure-based hazard predictions.

Automation of part of the decision process may be needed to ensure more rapid decisions for larger sets of data. However, implementation of automated processes is complicated because the current presentation of data is reliant on interpretation and experience, for which dedicated *in silico* models would be required. However, improving the quality of the tentative data, e.g., by reducing

ACCEPTED MANUSCRIPT

uncertainties, will be helpful in reducing the number of compounds that remain unclassified after prioritization, and will also assist in improving the quality of the decisions.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors confirm that there are no known conflicts of interest associated with this publication. The authors are grateful to the members of the ANSES panel “Risk assessment for substances and processes submitted to human food regulation” for scientific discussions, namely: Fabien Bolle, Jalloul Bouajila, Nicolas Cabaton, Marie-Christine Chagnon, Dany Chevalier, Véronique Coma, Luc Fillaudeau, Angel Gil-Izquierdo, Florence Lacoste, Claude Lambre, Michel Laurentie, Jean-Michel Maixent, Anne Platel, Philippe Saillard, Patrick Sauvegrain, and François Zuber.

Supplementary Material

The quantification, identification, expert assessment, and risk prioritization data results are available as supplementary Microsoft Office Excel file (*.xlsx*).

References

Anderson, W.A.C., Castle, L., 2003. Benzophenone in cartonboard packaging materials and the factors that influence its migration into food. Food Addit. Contam. 20, 607–618. https://doi.org/10.1080/0265203031000109486

Arvanitoyannis, I.S., Bosnea, L., 2004. Migration of Substances from Food Packaging Materials to Foods. Crit. Rev. Food Sci. Nutr. 44, 63–76. https://doi.org/10.1080/10408690490424621

Barlow, S.M., 2009. Risk assessment of food-contact materials: past experience and future challenges. Food Addit. Contam. Part A 26, 1526–1533. https://doi.org/10.1080/02652030903233231

Barnes, K., Sinclair, R., Watson, D. (Eds.), 2007. Chemical Migration and Food Contact Materials.

ACCEPTED MANUSCRIPT

Benfenati, E., Manganaro, A., Gini, G., 2013. VEGA-QSAR: AI inside a platform for predictive toxicology, in: CEUR Workshop Proceedings. pp. 21–28.

Bengtström, L., 2014. Chemical identification of contaminants in paper and board food contact materials. Technical University of Denmark, Kongens Lyngby, Denmark.

Bengtström, L., Rosenmai, A.K., Trier, X., Jensen, L.K., Granby, K., Vinggaard, A.M., Driffield, M., Højslev Petersen, J., 2016. Non-targeted screening for contaminants in paper and board food-contact materials using effect-directed analysis and accurate mass spectrometry. Food Addit. Contam. Part A 33, 1080–1093. https://doi.org/10.1080/19440049.2016.1184941

Biedermann-Brem, S., Kasprick, N., Simat, T., Grob, K., 2012. Migration of polyolefin oligomeric saturated hydrocarbons (POSH) into food. Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess. 29, 449–460. https://doi.org/10.1080/19440049.2011.641164

Biedermann, M., Grob, K., 2013. Assurance of safety of recycled paperboard for food packaging through comprehensive analysis of potential migrants is unrealistic. J. Chromatogr. A 1293, 107–119. https://doi.org/10.1016/j.chroma.2013.04.009

Biedermann, M., Uematsu, Y., Grob, K., 2011. Mineral oil contents in paper and board recycled to paperboard for food packaging. Packag. Technol. Sci. 24, 61–73. https://doi.org/10.1002/pts.914

Binderup, M.L., Pedersen, G.A., Vinggaard, A.M., Rasmussen, E.S., Rosenquist, H., Cederberg, T., 2002. Toxicity testing and chemical analyses of recycled fibre-based paper for food contact. Food Addit. Contam. 19, 13–28. https://doi.org/10.1080/02652030110089878

Bouma, K., Stavenga, K., Draaijer, A., 2003. Domestic Use of Food Packaging Materials in the Netherlands.

Bradley, E., 2006. Case study: Chemical migration from snack and take-away food packaging, Chemical Migration and Food Contact Materials. Woodhead Publishing Limited. https://doi.org/10.1533/9781845692094.3.416

Bschir, K., 2017. Risk, Uncertainty and Precaution in Science: The Threshold of the Toxicological Concern Approach in Food Toxicology. Sci. Eng. Ethics 23, 489–508. https://doi.org/10.1007/s11948-016-9773-2

Castle, L., 2006. Chemical migration into food: An overview, Chemical Migration and Food Contact Materials. Woodhead Publishing Limited. https://doi.org/10.1533/9781845692094.1

Cramer, G.M., Ford, R.A., Hall, R.L., 1976. Estimation of toxic hazard-A decision tree approach. Food Cosmet. Toxicol. 16, 255–276. https://doi.org/10.1016/S0015-6264(76)80522-6

Driffield, M., Bradley, E.L., Castle, L., 2016. Safety Assessment of Food Contact Materials : The Role of High-resolution Mass Spectrometry in the Comprehensive Analysis of the Total Migrate.

Duffy, E., Hearty, A.P., McCarthy, S., Gibney, M.J., 2007. Estimation of exposure to food packaging materials. 3: Development of consumption factors and food-type distribution factors from data collected on Irish children. Food Addit. Contam. 24, 63–74. https://doi.org/10.1080/02652030600865475

EFSA, WHO, 2016. Review of the Threshold of Toxicological Concern (TTC) approach and

ACCEPTED MANUSCRIPT

development of new TTC decision tree. EFSA Support. Publ. 2016EN-1006 1–50. https://doi.org/10.2903/SP.EFSA.2016.EN-1006

European Parliament and Council of the European Union, 2011. Commission Regulation (EU) No 10/2011 of 14 January 2011 on plastic materials and articles intended to come into contact with food. Off. J. Eur. Union 50, 12–88.

FDA, 2007. Guidance for Industry : Preparation of Premarket Submissions for Food Contact Substances : Chemistry Recommendations 1–18.

Gallart-Ayala, H., Núñez, O., Lucci, P., 2013. Recent advances in LC-MS analysis of food-packaging contaminants. TrAC Trends Anal. Chem. 42, 99–124. https://doi.org/10.1016/j.trac.2012.09.017

Geueke, B., Wagner, C.C., Muncke, J., 2014. Food contact substances and chemicals of concern: A comparison of inventories. Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess. 31, 1438–1450. https://doi.org/10.1080/19440049.2014.931600

Grob, K., 2014. Work plans to get out of the deadlock for the safety assurance of migration from food contact materials? A proposal. Food Control 46, 312–318. https://doi.org/10.1016/j.foodcont.2014.05.044

Guillén, D., Ginebreda, A., Farré, M., Darbra, R.M., Petrovic, M., Gros, M., Barceló, D., 2012. Prioritization of chemicals in the aquatic environment based on risk assessment: Analytical, modeling and regulatory perspective. Sci. Total Environ. 440, 236–252. https://doi.org/10.1016/j.scitotenv.2012.06.064

Hauder, J., Benz, H., Rüter, M., Piringer, O.-G., 2013. The specific diffusion behaviour in paper and migration modelling from recycled board into dry foodstuffs. Food Addit. Contam. Part A 30, 599–611. https://doi.org/10.1080/19440049.2012.762605

Honkalampi-Hämäläinen, U., Bradley, E.L., Castle, L., Severin, I., Dahbi, L., Dahlman, O., Lhuguenot, J.-C., Andersson, M.A., Hakulinen, P., Hoornstra, D., Mäki-Paakkanen, J., Salkinoja-Salonen, M., Turco, L., Stammati, A., Zucco, F., Weber, A., von Wright, A., 2010. Safety evaluation of food contact paper and board using chemical tests and in vitro bioassays: role of known and unknown substances. Food Addit. Contam. Part A 27, 406–415. https://doi.org/10.1080/19440040903401358

ILSI Europe Packaging Material Task Force, 1996. SUMMARY OF A WORKSHOP HELD 15 JULY 1996, BRUSSELS, in: Food Consumption and Packaging Usage Factors.

International Agency for Research on Cancer, 2017. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, years 1972 - 2017.

Istituto di Ricerche Farmacologiche Mario Negri Milano, 2017. VEGA interpretation [WWW Document]. URL https://www.vegahub.eu/download/vega-interpretation/

Jickells, S., 2007. Chemical migration from secondary packaging into foods, in: Chemical Migration and Food Contact Materials. Elsevier, pp. 395–415. https://doi.org/10.1533/9781845692094.3.395

Koster, S., Bani-Estivals, M.-H., Bonuomo, M., Bradley, E., Chagnon, M.-C., Garcia, L.M., Godts, F., Gude, T., Helling, R., Paseiro-Losada, P., Pieper, G., Rennen, M., Simat, T., Spack, L., 2015. Guidance of Best Practices on the Risk Assessment of NIAS in Food Contact Materials

ACCEPTED MANUSCRIPT

and Articles, International Life Science Institute. https://doi.org/10.1146/annurev.phyto.40.120301.093728

Koster, S., Rennen, M., Leeman, W., Houben, G., Muilwijk, B., van Acker, F., Krul, L., 2014. A novel safety assessment strategy for non-intentionally added substances (NIAS) in carton food contact materials. Food Addit. Contam. Part A 31, 422–443. https://doi.org/10.1080/19440049.2013.866718

Kroes, R., Renwick, A.G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., Schilter, B., Schlatter, J., Van Schothorst, F., Vos, J.G., Würtzen, G., 2004. Structure-based thresholds of toxicological concern (TTC): Guidance for application to substances present at low levels in the diet. Food Chem. Toxicol. 42, 65–83. https://doi.org/10.1016/j.fct.2003.08.006

Lester, C., Reis, A., Laufersweiler, M., Wu, S., Blackburn, K., 2018. Structure activity relationship (SAR) toxicological assessments: The role of expert judgment. Regul. Toxicol. Pharmacol. 92, 390–406. https://doi.org/10.1016/j.yrtph.2017.12.026

Muncke, J., Backhaus, T., Geueke, B., Maffini, M. V, Martin, O.V., Myers, J.P., Soto, A.M., Trasande, L., Trier, X., Scheringer, M., 2017. Scientific Challenges in the Risk Assessment of Food Contact Materials. Environ. Health Perspect. 125, 1–9. https://doi.org/10.1289/EHP644

Ozaki, A., Yamaguchi, Y., Fujita, T., Kuroda, K., Endo, G., 2005. Safety assessment of paper and board food packaging: Chemical analysis and genotoxicity of possible contaminants in packaging. Food Addit. Contam. 22, 1053–1060. https://doi.org/10.1080/02652030500090885

Pieke, E.N., Granby, K., Trier, X., Smedsgaard, J., 2017. A framework to estimate concentrations of potentially unknown substances by semi-quantification in liquid chromatography electrospray ionization mass spectrometry. Anal. Chim. Acta 975, 30–41. https://doi.org/10.1016/j.aca.2017.03.054

Pieke, E.N., Smedsgaard, J., Granby, K., 2018. Exploring the chemistry of complex samples by tentative identification and semiquantification: A food contact material case. J. Mass Spectrom. 53, 323–335. https://doi.org/10.1002/jms.4052

Pivnenko, K., Eriksson, E., Astrup, T.F., 2015. Waste paper for recycling: Overview and identification of potentially critical substances. Waste Manag. 45, 134–142. https://doi.org/10.1016/j.wasman.2015.02.028

Poças, M. de F., Oliveira, J.C., Pereira, J.R., Brandsch, R., Hogg, T., 2011. Modelling migration from paper into a food simulant. Food Control 22, 303–312. https://doi.org/10.1016/j.foodcont.2010.07.028

Raies, A.B., Bajic, V.B., 2016. In silico toxicology: computational methods for the prediction of chemical toxicity. Wiley Interdiscip. Rev. Comput. Mol. Sci. 6, 147–172. https://doi.org/10.1002/wcms.1240

Rosenmai, A.K., Bengtström, L., Taxvig, C., Trier, X., Petersen, J.H., Svingen, T., Binderup, M.-L.L., van Vugt-Lussenburg, B.M.A., Dybdahl, M., Granby, K., Vinggaard, A.M., Barbara Medea Alice, van V.L., Dybdahl, M., Granby, K., Vinggaard, A.M., 2017. An effect-directed strategy for characterizing emerging chemicals in food contact materials made from paper and board. Food Chem. Toxicol. 106, 250–259. https://doi.org/10.1016/j.fct.2017.05.061

Ru, G., Crescio, M.I., Ingravalle, F., Maurella, C., Gregori, D., Lanera, C., Azzolina, D., Lorenzoni, G., Soriani, N., Zec, S., Berchialla, P., Mercadante, S., Zobec, F., Ghidina, M., Baldas, S.,

ACCEPTED MANUSCRIPT

Bonifacio, B., Kinkopf, A., Kozina, D., Nicolandi, L., Rosat, L., 2017. Machine Learning Techniques applied in risk assessment related to food safety. EFSA Support. Publ. 14. https://doi.org/10.2903/sp.efsa.2017.EN-1254

SCCNFP, 2002. Safety review of the use of certain azo-dyes in cosmetic products.

Schymanski, E.L., Singer, H.P., Longrée, P., Loos, M., Ruff, M., Stravs, M.A., Ripollés Vidal, C., Hollender, J., 2014. Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry. Environ. Sci. Technol. 48, 1811–8. https://doi.org/10.1021/es4044374

Skjevrak, I., Brede, C., Steffensen, I.L., Mikalsen, A., Alexander, J., Fjeldal, P., Herikstad, H., 2005. Non-targeted multi-component analytical surveillance of plastic food contact materials: Identification of substances not included in EU positive lists and their risk assessment. Food Addit. Contam. 22, 1012–1022. https://doi.org/10.1080/02652030500090877

Triantafyllou, V.I., Akrida-Demertzi, K., Demertzis, P.G., 2007. A study on the migration of organic pollutants from recycled paperboard packaging materials to solid food matrices. Food Chem. 101, 1759–1768. https://doi.org/10.1016/j.foodchem.2006.02.023

van Bossuyt, M., van Hoeck, E., Raitano, G., Manganelli, S., Braeken, E., Ates, G., Vanhaecke, T., Van Miert, S., Benfenati, E., Mertens, B., Rogiers, V., 2017. (Q)SAR tools for priority setting: A case study with printed paper and board food contact material substances. Food Chem. Toxicol. 102, 109–119. https://doi.org/10.1016/j.fct.2017.02.002

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

- A strategy for risk prioritization of FCM-borne chemical compounds is shown
- Application of a decision model utilizing both expert judgment and tentative data
- Non-target scope enables prioritization of NIAS and newly discovered compounds
- Compound priority constructed from expert-assigned risk profile consensus
- Strategy demonstrated on a subset 60 compounds from paper and board FCM

ACCEPTED MANUSCRIPT