Evaluating the alignment of alternative assessment methods with REACH Article 57

Rachel L. London, Juliane Glüge,* and Martin Scheringer

Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, 8092 Zürich, Switzerland

E-mail: juliane.gluege@usys.ethz.ch

Abstract

The assessment of chemical alternatives for hazardous substances is an important prerequisite for avoiding regrettable substitution, and several methods have been developed in the past to perform such an Assessment-of-Alternatives. We investigate whether GreenScreen®, Cradle to Cradle®, Multi Criteria Decision Analysis (MCDA), Cost Benefit Analysis, the Pollution Prevention Options Analysis System, the U.S. EPA Safer Choice Standard and Criteria, and the GHS column model 2020 from IFA use similar criteria for the evaluation of substances as Article 57 of REACH and how suitable these methods are for assessing per- and polyfluoroalkyl substances. MCDA and GreenScreen® were identified as the most promising methods and were analyzed in detail using two different datasets. The results of the assessments show that none of the investigated alternative assessment methods use the same hazard assessment criteria as described in Article 57 of REACH. It was also not possible to parameterize MCDA (using the multi-attribute utility theory approach) without redundant variables so that the results align with an assessment according to Article 57 of REACH. There is therefore an urgent need for a modified/new method that can be used in the future to assess organic substances that are used within the European Economic Area.

Synopsis

To avoid regrettable substitution, effectual assessment of chemical alternatives (AoA) is crucial. Therefore, the assessment of the alignment of current AoA methods with scientific progress, and regulatory priorities, is required.

Introduction

In February 2023, the European Chemicals Agency (ECHA) published a proposal for the restriction of per- and polyfluoroalkyl substances (PFAS)[1] under the Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH) regulation. PFAS are persistent, and many of the studied PFAS show effects that are problematic to human health and the environment. However, they also have unique properties[1] and these properties have led to the use of PFAS in many different applications[2]. Alternatives are now being introduced for a range of PFAS uses and functions[1,3–6]. However, whilst several non-hazardous alternatives have been found for some consumer uses of PFAS, it can be more complex for industrial uses, where the alternatives may have a different hazard profile, rather than being completely non-hazardous[3]. For these alternatives, it is important to evaluate whether they represent a real improvement or whether they have just replaced one hazardous substance with another (regrettable substitution).

An example of regrettable substitution of PFAS was when perfluorooctanoic acid was replaced as a processing aid by 2,3,3,3-tetrafluoro-2-(heptafluoropropoxy)propionic acid (HFPO-DA) (the ammonium salt of which is sometimes referred to as "GenX"). In 2019, the EU also classified HFPO-DA as a Substance of Very High Concern (SVHC)[7]. By transitioning to a less regulated PFAS, rather than a less hazardous alternative, these industries did not solve the problem at hand. Having to perform a search for an alternative twice is an inefficient use of time, money, and expertise. The problem the present article seeks to address is how to minimize regrettable substitutions in the upcoming EU restriction of PFAS, but also for

other chemical substitutions in the future.

To minimize regrettable substitution, all relevant hazards should be considered, and a suitable Assessment-of-Alternative (AoA) method applied. Care should be taken to ensure that all PFAS-relevant hazards are considered, including persistence[8], mobility[9,10], and global warming potential[9]. A lot of different AoA methods have been developed[11] and it is important to evaluate them in terms of both the substances of interest (e.g., PFAS) and within the regulation of interest (e.g., REACH).

The present article addresses the question of how assessments of chemical alternatives to PFAS should be carried out in the context of the EU regulation. To do so, the paper investigates various available AoA methods, including GreenScreen®[12], Cradle to Cradle®[13], Multi Criteria Decision Analysis (MCDA)[14], Cost Benefit Analysis (CBA)[15], the Pollution Prevention Options Analysis System (P2OSys)[16], the U.S. EPA Safer Choice Standard and Criteria[17], and the Institute for Occupational Safety and Health of the German Social Accident Insurance's (IFA's) GHS column model 2020[18] for their suitability for assessing chemicals in the same way as Article 57 under REACH.

Article 57 of REACH is part of the identification of SVHCs[19]. SVHCs may be included in Annex XIV of REACH, which is the list of substances subject to authorization. The complete text of Article 57 is given elsewhere[19], but in short, it addresses substances that are carcinogenic, mutagenic, or toxic for reproduction (CMR), or persistent, bioaccumulative and toxic (PBT) or very persistent and very bioaccumulative (vPvB) or endocrine disruptive. If a substance has been officially identified under REACH as an SVHC, it will be added to the so-called Candidate List of SVHCs for authorization. In the present article, we investigate whether the selected hazard assessment methods are able to identify potential SVHCs and if not, whether they are more or less precautionary than Article 57 of REACH.

Methods

Overview of AoA methods

The hazard assessment methods selected here are those that were also discussed by Jacobs et al.[11]. Specifically, these are Cradle to Cradle®, GreenScreen®, CBA, MCDA, the GHS column model 2020 from IFA, P2OSys, and the U.S. EPA Safer Choice Standard and Criteria. Other methods do exist (e.g., described in Faludi et al.[20] and Beaudrie et al.[21]), however it was decided to concentrate on the most common ones. Method summaries were created by identifying the scope, data requirements, flexibility, weighting method, aggregation method, and whether the method uses the same criteria as described in Article 57 of REACH. MCDA and GreenScreen were identified as the most promising methods and were investigated in-depth.

Data sets used to investigate the methods in-depth

To investigate MCDA and GreenScreen® in detail, two different datasets were used, a 'hypothetical substances dataset' and a 'real substances dataset'.

Hypothetical substances dataset

To be able to use data in the assessment of the two methods, five general data treatment steps are necessary: data gathering, data normalization, data conversion, data aggregation, and data classification. Data gathering means that relevant hazard data (e.g., the degradation half-life of a substance) need to be collected. Data normalization describes the step where the gathered data are categorized (normalized) into hazard levels using external thresholds. For example, a half-life can describe the persistence of a substance as low, moderate, or high. The allocation to the hazard levels depends in this case on the half-life itself, but also on the thresholds used. The third step is the conversion of the hazard levels into a quantitative hazard value. This is followed by the data aggregation step where the nor-

malized hazard values of the different endpoints (persistence, bioaccumulation, etc.) are combined into one final score. The final step is the classification of the substance according to the final score. For the hypothetical substances dataset, no actual data were gathered, instead 256 different combinations of four hazards and four hazard severities were used to investigate the conversion of the qualitative hazard data into hazard values as well as the data aggregation and data classification step in detail. The four hazards included were those relevant to both GreenScreen® and Article 57 of REACH: persistence (P), bioaccumulation (B), human toxicity (T_{hu}) and ecotoxicity (T_{eco}). Mobility (M) was not considered explicitly for the hypothetical substances. The four hazard severities were based on those used within the GreenScreen® method: very high, high, moderate, and low. By using hypothetical substances, it was possible to systematically investigate how the two methods, MCDA and GreenScreen®, respond to each combination of hazard and hazard severity within the scope of the study.

Real substances dataset

The second dataset used to investigate MCDA and GreenScreen® in detail was a dataset published in a previous article[22] for the comparison of 17 substances (the brominated flame retardant decabromodiphenyl ether and 16 alternatives) using MCDA. The hazards considered in this dataset were P, B, M and toxicity (which included both T_{eco} and T_{hu}). Some of the data points had been determined experimentally; others were derived from quantitative structure-activity relationships (QSARs) alone. For our assessment, we used the raw data from Zheng et al.[22] in the data gathering step. However, the data normalization was not adopted from Zheng et al.[22], but done by using the thresholds from the Guidance on Information Requirements and Chemical Safety Assessment Chapter R.11: PBT/vPvB assessment[23,24]. The normalized data were then used to investigate the data aggregation and data classification in the two methods, MCDA and GreenScreen®. Important to note is here that the dataset of Zheng et al.[22] did not include uncertainty ranges of the hazard data.

However, as the primary purpose of the present article is an investigation into the methods, these data limitations were considered acceptable.

Investigations into GreenScreen®

A spreadsheet (SI-2) was created to categorize all substances in both datasets according to the GreenScreen® decision tree, as described in Annex 3 of the Guidance[12] and shown in SI-1 Figure S1. This was done by transforming the GreenScreen® decision tree into AND, IF, and OR statements, as shown in Figure 1 for Benchmark 1 ("Avoid: chemical of high concern").

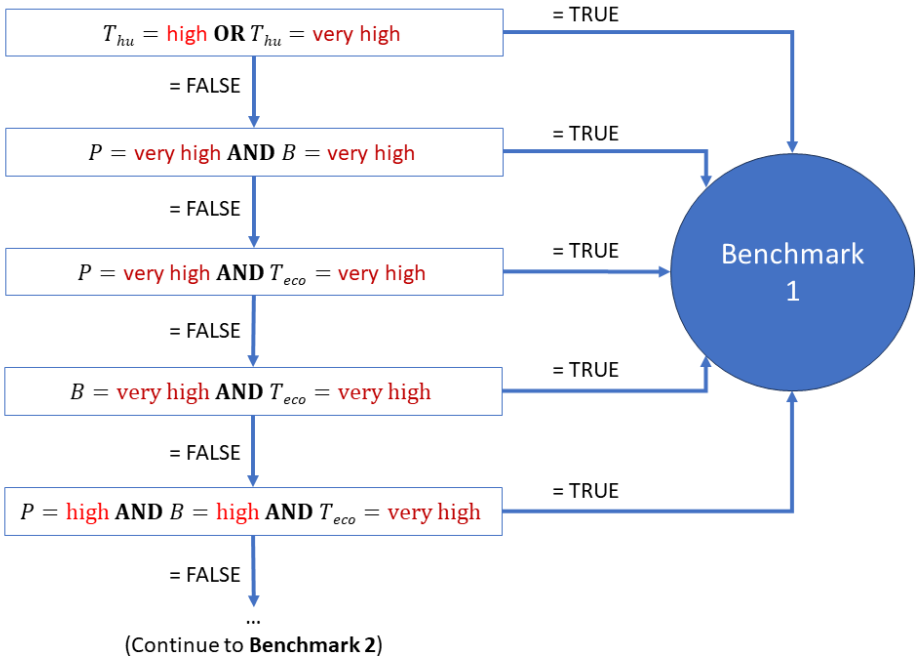

Figure 1: Section of the decision tree used to replicate GreenScreen® (Benchmark 1)
P = Persistence, B = Bioaccumulation, T_{eco} = Ecotoxicity, T_{hu} = Human Toxicity

If Benchmark 1 gave a FALSE for all decision points, the hazard scores were assessed using an equivalent process for Benchmark 2. This process was repeated for subsequent benchmarks, until the substance was categorized. Physical hazards were excluded from the assessment, as these data are not relevant for Article 57 of REACH. For the real substances dataset, it should be noted that approximate assessments were made for the GreenScreen®

endpoints using solely the data already provided by Zheng et al.[22]. These data were not always exactly of the type or format required by the GreenScreen® guidance (see SI-1 Section S2.1); however, they were sufficient for the investigation of the method. It is still important to note that the GreenScreen® assessments made here should not be used outside the present article.

Investigations into the variability of MCDA

MCDA, specifically multi-attribute utility theory (MAUT), was also performed on the two datasets, and the outcome then contrasted with the outcome of GreenScreen®. MAUT calculates a value, f, which is the MCDA hazard score in the present article. The MCDA hazard score can range from zero (worst) to one (best) for each alternative by aggregating the performance in objectives (e.g., low persistence) so that the alternatives can then be ranked from most to least desirable. The MCDA-MAUT method parameters most appropriate for an assessment of chemical alternatives had not yet been investigated in the literature. Therefore, to determine these parameters, a bounding analysis was done on different combinations of the parameters aggregation, curvature of the value function, and weighting of objectives.[25]

For these investigations, the MCDA-MAUT tool, ValueDecisions[25,26], was used. Five different aggregations were investigated: additive (Eq. 1), geometric (Eq. 2), geometric-additive (Eq. 3), minimum (Eq. 4), and maximum aggregation (Eq. 5).

$$f_{\text{add}} = \sum_{i=1}^{n} w_i \cdot \nu_i \tag{1}$$

$$f_{\text{geo}} = \prod_{i=1}^{n} \nu_i^{w_i} \tag{2}$$

$$f_{\text{geo-add}} = \left(\sum_{i=1}^{n} w_i \cdot \nu_i^{0.5} \right)^2 \tag{3}$$

$$f_{\min} = \min(\nu_i) \tag{4}$$

$$f_{\max} = \max(\nu_i) \tag{5}$$

The equation for the five different aggregations were originally developed by Langhans et al.[27] and are included in ValueDecisions. For each alternative considered, the MCDA hazard score, f, is calculated using one of the aggregation equations above. The subscript i denotes different objectives in the MCDA. The variable ν_i ($0.0 \leq \nu_i \leq 1.0$) represents the alternative's performance in terms of the corresponding objective (i). ν_i is determined by transforming and combining attributes (e.g., half-life in water) into a single representative value. This transformation is done using value functions, mathematical expressions that convert the attribute data into comparable values, the output of which is then combined into ν_i. A weight of an objective, w_i, is also required in some of the aggregation equations ($0 \leq w_i \leq 1$ where $\sum w_i = 1$).

For the value functions, three different curvatures were investigated: linear, exponential–concave, exponential-convex. The respective equations are provided in the SI-1 Section S3.1. For the weights, three different weights (w_{P}) were investigated by varying the weight of the objective of persistence relative to the three other objectives: equal ($w_{\mathrm{P}} = 0.25$), high ($w_{\mathrm{P}} = 0.55$), or low ($w_{\mathrm{P}} = 0.15$). These weights were chosen to ensure that they summed to one in each scenario, therefore for the equal weight $w_{\mathrm{P}} = 1/4 = 0.25$. Marttunen et al.[28] cautioned against utilizing "very low weights" (≤ 0.05); thus, the low weight was established equidistant from this at 0.15. The high weight was derived once the low weight had been established. If all four weights must sum to one, and there are three "low" weights of 0.15, the remaining single "high" weight must be 0.55 ($w_{\mathrm{P}} = 1 - 3 \cdot 0.15 = 0.55$).

From the hypothetical substances dataset, a subset of 17 substances with low (4), medium (7), and high variation (6) in hazard severity were selected (SI-2). Substance 86 is an ex-

ample of a substance with low variation of hazard severity (all hazards = "high"), whilst Substance 64 is an example of a substance with high variation of hazard severity (persistence = "very high", all other hazards = "low"). In the first investigation using the hypothetical dataset, 15 MCDAs were conducted each with a unique parameter combination of the five different aggregations and the three different curvatures, whilst the weights were kept constant (equal, $w_i = 0.25$). In the second experiment of the hypothetical dataset, 15 MCDAs were conducted each with a unique parameter combination of the five different aggregations and the three different weights, whilst the curvature was kept constant (linear). For the real substance's dataset, also 15 MCDAs were conducted with aggregation and curvature varied, whilst weights kept the same (equal, $w_i = 0.25$). The objective hierarchy used for each of these datasets is shown in Figure 2.

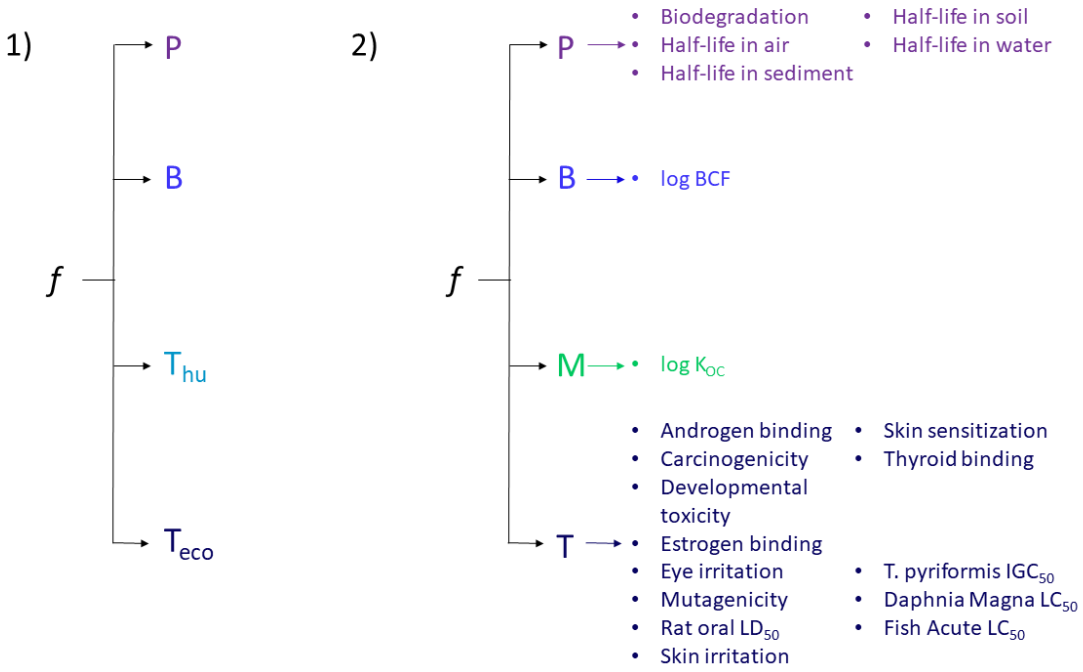

Figure 2: MCDA objective hierarchy showing the hazards that contribute to the MCDA hazard score (f) of a substance for 1) the hypothetical substances dataset and 2) the real substances dataset. P = persistence, B = bioaccumulation, M = mobility, T = toxicity, T_{eco} = ecotoxicity, T_{hu} = human toxicity

Selecting MCDA method parameters to align with Article 57 of REACH

The ability of MCDA to classify substances as potential SVHCs was investigated by using the hypothetical substances dataset. In a first step, all hypothetical substances were manually assigned a hazard category, which was either "potential SVHC" or "not an SVHC" according to the criteria laid down in Article 57 of REACH.

In a second step, nine different MCDAs, each with a different set of parameters, were applied to the hypothetical substances dataset. The parameters were a combination of three different aggregations (additive, geometric mean, and minimum) with three different curvatures (linear, convex, and concave) with equal weights throughout ($w_i = 0.25$) (SI-2).

Finally, a classification threshold for f was defined below which a substance would be classified as equivalent to a "potential SVHC". For all nine MCDA parameter combinations described above, the classification threshold for f was optimized so that the MCDA outcome had the highest agreement with the evaluation according to Article 57 of REACH. This was done by initially setting the classification threshold to 0.1 and increasing it incrementally by 0.1 to 0.9. Subsequently, the range was determined in which the highest match occurred. However, only one explicit threshold was selected for the results section.

Results

Overview of AoA Methods

The assessment of the seven AoA methods is summarized in Table 1. Detailed descriptions of the methods and additional information are provided in SI-1 Section S1.1.

GreenScreen® and MCDA were identified as the most suitable of the methods investigated: the flexibility of MCDA made it an effectual method if the right parameters could be identified, and GreenScreen® was the most nuanced of the decision tree methods, allowing for

differentiation between imperfect alternatives.

The other methods were excluded from further investigation for various reasons. Cradle to Cradle® and U.S. EPA Design for Environment are certification standards that can be used to identify consumer products with exceptionally low hazard. However, they have limited flexibility and may not be suitable where one must differentiate between imperfect alternatives, rather than identify a hazard-free one. CBA does not accurately account for the long-term effects of persistent chemicals in the environment. The GHS column model 2020 from IFA is a decision-making system with no flexibility and a scope built around the current Globally Harmonized System of Classification, Labelling and Packaging of Chemicals (GHS) criteria. P2OSys does not align with Article 57 of REACH as the aggregation used in P2OSys allows poor performance in one hazard category to be compensated for by good performance in another hazard category.

Table 1: Overview of AoA methods investigated (extended version in SI-1, Section S1.2)

Hazard assessment method	Similar to the criteria in Article 57 of REACH?	How are PFAS treated?
Cradle to Cradle®	Stricter than the criteria in Article 57. Failure in a single endpoint sufficient for exclusion. Therefore, persistence alone is sufficient to classify substance as unacceptable. Some endpoint thresholds (PURPLE and RED) are the same as those used in REACH for equivalent attributes (e.g., half-life in freshwater). See SI-1 Section S.1.3 for comparison of persistence and bioaccumulation thresholds.	Presence of organohalogen bonds renders a chemical unacceptable. All PFAS are therefore considered unacceptable under this method.
GreenScreen®	Classifies some substances different than Article 57. Also, endpoint thresholds differ from those used in REACH. See SI-1 Section S1.3 for comparison of persistence and bioaccumulation thresholds.	Current method not appropriate for PFAS as it does not evaluate all relevant hazards of concern (does not cover mobility, global warming potential (GWP), and ozone depletion potential (ODP)). Additionally, PFAS classified as (only) vP would be classified as '3 = use, but still opportunity for improvement'
Cost Benefit Analysis (CBA)	Similar criteria could be used, but it may be difficult to account for the long-term effects of persistent substances	Long-term effects are not well addressed by this method. Therefore, CBA may not be appropriate for persistent chemicals whose effects can continue beyond the scope of any project, and whose effects may not be well-defined at the point when the decision is made.
Multi Criteria Decision Analysis (MCDA)	Normally not, because Article 57 of REACH considers combinations of hazards where a few objectives (e.g., low persistence) are included in more than one combination (e.g., low persistence occurs in PBT and vPvB); in MCDA objectives should not be redundant. However, a modified MCDA method that allows the inclusion of specific endpoints several times could allow this.	Yes, potentially. It is imperative that suitable method parameters are selected.

Continued on the next page

Table 1 – continued from previous page

Hazard assessment method	Similar to the criteria in Article 57 of REACH?	How are PFAS treated?
The GHS column model 2020 from IFA (the Institute for Occupational Safety and Health of the German Social Accident Insurance)	Differences to REACH Article 57 include: • Endpoints covered (e.g., endocrine disruption not addressed) • Severity with which endpoints are treated (e.g., Reprotoxic not equivalent to Carcinogenicity or Mutagenicity) • How endpoints are aggregated (Trade-offs between hazards possible)	Endpoints covered do not currently capture all hazards of PFAS (e.g., mobility missing). Problematic degradation products not addressed.
MA TURIs 'Pollution Prevention Options Analysis System (P2OSys)	Environmental hazards are combined in a different way than in Article 57 of REACH. Thresholds differ from those used under REACH	PFAS have a high variation of hazard severity. By taking an average of the category scores, this method allows for poor performance in one area (e.g., environmental fate) to be compensated for by good performance in another (e.g., acute human effects). The final scores given to substances with high variation of hazard severity are "good to average". This means this method may be misleading if applied to PFAS
U.S. EPA Safer Choice Standard and Criteria	Environmental endpoints (P, B and T) are combined in both methods, but the implications of these combinations differ. US EPA penalises additional hazard combinations (PT and P). Thresholds of P and B are stricter than those in REACH. Threshold of P varies depending on the acute aquatic toxicity. Additionally, assessment criteria can be modified for specific functional classes (e.g., surfactants, solvents).	Does not consider all PFAS-relevant hazards, e.g., mobility is missing. Persistent degradation products would exclude PFAS from being considered a "safer choice" substance. Proposed revisions to the guidance explicitly prohibit the intentional addition of PFAS to packaging.[29]

Investigations into GreenScreen®

The hypothetical substances and the substances from the real substances dataset were all categorized into one of four Benchmarks: 1 = "Avoid: chemical of high concern", 2 = "Use but search for safer substitutes", 3 = "Use but still opportunity for improvement", 4 = "Prefer: Safer chemical". For the real substances dataset, GreenScreen® categorized all substances into Benchmark 1. For the hypothetical substances dataset, there were substances in all four groups. Table 2 displays how the 256 hypothetical substances were categorized by GreenScreen® and compares this to their classification when the criteria from Article 57 of REACH are used. The comparison shows that eight substances were categorized as GreenScreen® Benchmark 1 although they would not be considered as potential SVHCs under REACH. Six substances are on the other hand potential SVHCs under REACH but received a GreenScreen® Benchmark 2. Table 3 presents more detailed information for a subset of these substances. Detailed information for all other substances are available in the SI-2.

Table 2: Comparison of the assessment of the hypothetical substances using GreenScreen® and the hazard criteria laid down in Article 57 of REACH

GreenScreen® Benchmark	**Potential SVHC under REACH?**	
	No	**Yes**
1 ("Avoid: chemical of high concern")	8	142
2 ("Use but search for safer substitutes")	77	6
3 ("Use but still opportunity for improvement")	22	0
4 ("Prefer: Safer chemical")	1	0

Table 3 shows that GreenScreen® categorizes substances with a high or very high hazard in P, a high or very high hazard in B and a high hazard in T_{eco} as Benchmark 2 while these substances are potential SVHCs under REACH. Substances with a hazard combination of very high P and very high T_{eco} or very high B and very high T_{eco} are otherwise Benchmark 1

Table 3: Selected examples of GreenScreen® benchmarks from the hypothetical dataset, where 1 = "Avoid – chemical of high concern", 2 = "Use but search for safer substitutes", 3 = "Use but still opportunity for improvement", 4 = "Prefer – Safer chemical"

Substance No.	P	B	T_{eco}	T_{hu}	Potential SVHC under REACH?	GreenScreen® Benchmark
24	very high	high	high	low	Yes	2
72	high	very high	high	low	Yes	2
88	high	high	high	low	Yes	2
52	very high	low	very high	low	No	1
196	Low	very high	very high	low	No	1
63	Very high	Low	Low	Middle	No[a]	2
64	Very high	Low	Low	Low	No[a]	3

[a] if only the endpoints P, B, T_{eco}, and T_{hu} are considered

in GreenScreen® while these substances are not explicitly addressed in Article 57 of REACH. This means that GreenScreen® is less strict than Article 57 of REACH with the categorization for some substances and more strict than Article 57 for others.

Substances 63 and 64 correspond to the hazard profile of PFAS, such as HFPO-DA, and were categorized as Benchmark 2 and 3. However, Benchmark 1 would have been more appropriate for HFPO-DA, as it has been classified as an SVHC.[7,30] Even if the current criteria under Article 57 of REACH do not (yet) include mobility, substances that are very persistent and very mobile may be considered as substances with equivalent level of concern to CMR, PBT, and vPvB substances.[31] In this respect, GreenScreen® is less strict than Article 57 of REACH. To assess HFPO-DA and other PFAS with a similar hazard profile, the scope of GreenScreen® would need to be expanded to include mobility, although mobility was not the only criteria that let to the decision to classify HFPO-DA as a SVHC.

Dependency of the MCDA hazard scores on the method parameters

Hypothetical substances dataset

To investigate the variability of MCDA outcomes, MCDA hazard scores (f) were determined for a subset of hypothetical substances, with 1 representing the best possible score and 0 the worst possible score. Figure 3 shows how f changes when the parameters of 1) aggregation and curvature and 2) aggregation and weighting are changed.

Substances 171 and 86 show that for substances with low variation of hazard severity (meaning all endpoints have the same hazard severity), f remains the same when aggregation and weight is varied, but changes in most cases when the curvature is varied (Figure 3). f is sensitive to changes in curvature because the three value functions – exponential-concave, linear and exponential-convex – only have the same value $\nu(x)$ at the beginning and end of the functions, where x represents the minimum or maximum value (e.g., at the lowest and highest possible half-life values.) At all other points along the x-axis, the values of $\nu(x)$ diverge among the three functions. Figure S2 in SI-1 shows this again graphically, alongside a more detailed explanation.

Substance 64 shows that for a substance with high variation of hazard severity (meaning some hazards are very high and some are low), f is sensitive to the aggregation, curvature, and sometimes sensitive to the weighting. Whether f is sensitive to weights depends on the aggregation used because whilst weights influence the geometric, additive, and geometric-additive aggregation (Eqs. 1, 2, and 3), weights are not considered in minimum and maximum aggregation (Eqs. 4 and 5). This can also be seen in Figure S3 in the SI-1.

Substance 64 has the same hazard profile as some PFAS, such as HFPO-DA. So particular care should be taken in selecting the method parameters to be used for an MCDA involving PFAS, as they are often substances with a high variation of hazard severity, and so f determined for these substances will be sensitive to aggregation, curvature, and sometimes weights.

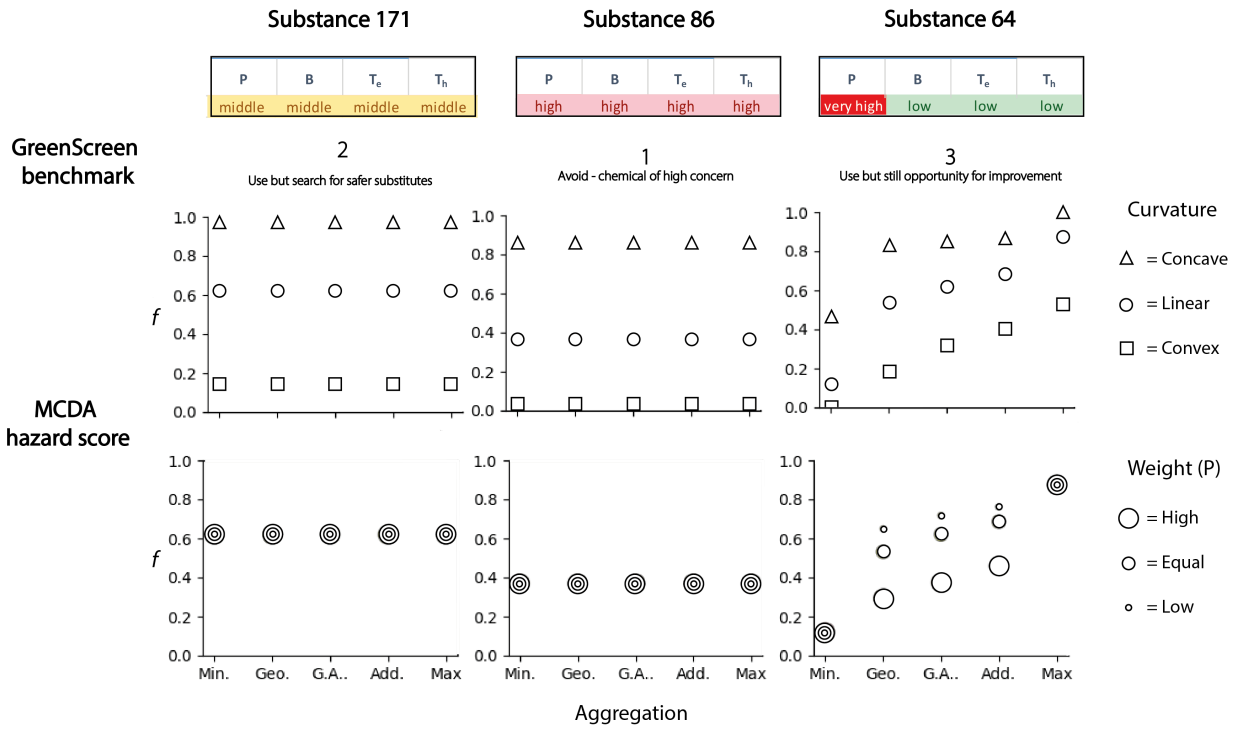

Figure 3: Variability of MCDA hazard scores (f) for three of the hypothetical substances (two with low variation of hazard severity, one with high variation of hazard severity). Top row: f as a function of curvature and aggregation, weights kept constant (equal, $w_i = 0.25$ for each hazard). Bottom row: f as a function of weights and aggregation, curvature kept constant (linear). (Min. = minimum aggregation, Add. = additive aggregation, Max. = maximum aggregation, Geo. = weighted geometric mean aggregation, G.A. = weighted geometric mean – additive aggregation)

Real substances dataset

For the real substances dataset, MCDA hazard scores (f) were calculated 15 times for each of the 17 substances, using different combinations of aggregation and curvature (weights were kept equal and constant). f was then used to rank the 17 substances, with the best performing alternative (i.e., highest median of f) being given a rank of 1 and the worst performing alternative being given a rank of 17. In Figure 4, the variability of the rank assigned to each substance as the method parameters were changed is shown by the bars. Thus, the bars reflect different MCDA ranks, not uncertainty of input data.

Figure 4 shows that, for most substances in this dataset, the ranks assigned to each substance overlap with the ranks achieved by at least one other substance. This shows that the ranking also depends on the parameters chosen. A few substances consistently outperformed or underperformed relative to the other alternatives, regardless of the parameter combination applied. For example, melamine was consistently ranked as one of the best alternatives, whereas tetrabromobisphenol A bis (2,3-dibromopropyl) ether (TBBPA-BDBPE) was consistently ranked as one of the worst alternatives. However, the rank of many of the other substances depended on the parameters.

Contrasting results from GreenScreen® and MCDA

Contrasting the results of the real substances dataset between GreenScreen® and MCDA (as shown in Figure 4) shows important differences. MCDA ranked melamine as the best performing alternative, whilst GreenScreen® categorized all substances, including melamine, as Benchmark 1: "Avoid chemical of high concern". This contrast demonstrates the effect of using methods that evaluate hazards on either an absolute or relative scale. In MCDA, the value functions of each hazard (ν_{P}, ν_{B}, ν_{eco}, ν_{hu}) were scaled relative to the dataset, not on an absolute scale. MCDA identified f_{MA} (the performance value of melamine) as a maximum, but this is a local, not a global maximum. There are substances outside the real substance data set that would have achieved $f > f_{\mathrm{MA}}$.

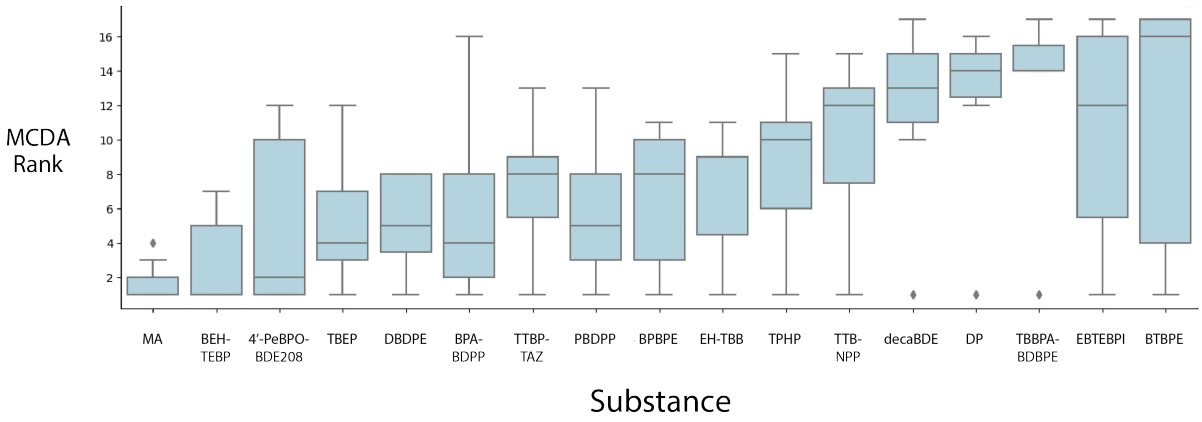

Figure 4: MCDA ranks for the real substances dataset. Rank as a function of curvature and aggregation, weight kept constant.
(MA = melamine, BEH-TEBP = bis(2-ethylhexyl) tetrabromophthalate, 4'-PeBPOBDE208 = tetradecabromodiphenoxybenzene, TBEP = tris(2-bromoethyl) phosphate, DBDPE = decabromodiphenyl ethane, BPA-BDPP = bisphenol A diphenyl phosphate, TTBP-TAZ = tris(tribromophenoxy) triazine, PBDPP = resorcinol bis(diphenyl phosphate), BPBPE = 1,2-bis(pentabromophenoxy) ethane, EH-TBB = 2-ethylhexyl tetrabromobenzoate, TPHP = triphenyl phosphate, TTBNPP = tris(tribromoneopentyl) phosphate, decaBDE = decabromodiphenyl ether , DP = bis(hexachlorocyclopentadieno)cyclooctane, TBBPA-BDBPE = tetrabromobisphenol A bis (2,3-dibromopropyl) ether, EBTEBPI = ethylene bis-tetrabromophthalimide, BTBPE = Bis(tribromophenoxy) ethane)

Contrasting the results of the hypothetical substances dataset by GreenScreen® and MCDA demonstrates the challenges associated with flexible method parameters. For example, the single combination of method parameters permitted by GreenScreen® allows for a single outcome for Substance 64 (the hypothetical substance similar to some PFAS). In contrast, the 15 combinations of MCDA method parameters explored here allow for multiple outcomes for Substance 64. There are MCDA method parameter combinations where:

- $f_{\text{substance\_64}} > f_{\text{substance\_86}}$
  - e.g., w_{P} = low, curvature = linear, aggregation = additive
- $f_{\text{substance\_64}} = f_{\text{substance\_86}}$
  - e.g., w_{P} = high, curvature = linear, aggregation = geometric-additive
- $f_{\text{substance\_64}} < f_{\text{substance\_86}}$

– e.g., w_P = equal, curvature = convex, aggregation = minimum

This range in MCDA outcomes demonstrates the necessity of standardized method parameters for the effective use of MCDA as a chemical alternative assessment method.

Selecting MCDA method parameters to align with REACH

Aggregation and curvature were systematically tested with the hypothetical substances to see whether it is possible to align the MCDA outcome with the identification of potential SVHCs under REACH according to the criteria of Article 57. Figure 5 shows that the agreement between MCDA and REACH Article 57 was for all nine combinations between 67% and 77%.

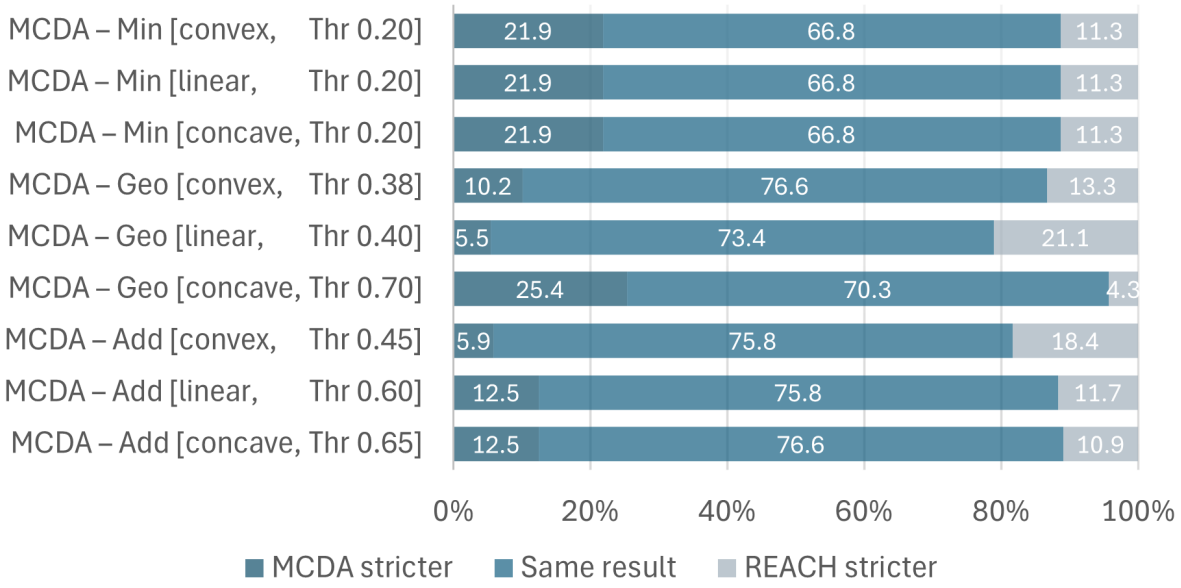

Figure 5: Percentage agreement for the hypothetical substances evaluated according to Article 57 of REACH and MCDA. For each substance, the results from REACH and MCDA could either be the same, MCDA being stricter, or REACH being stricter. For MCDA, nine different method parameter combinations were investigated, with three different aggregations (Add = additive, geometric = Geo, or Min = minimum) and three different value functions (linear, concave or convex). Equal weights were used throughout ($w_i = 0.25$). The classification thresholds are given in the labels under "Thr.".

None of the combinations of MCDA method parameters resulted in 100% agreement with REACH (i.e., same result = 100%). The combination that was previously used in other

publications (additive aggregation with linear value function) resulted in 75.8% agreement. For 11.7% of the hypothetical substances, the evaluation with REACH Article 57 was stricter meaning that the MCDA evaluation would have overlooked regrettable substitutes. This shows that although MCDA is a very flexible method it is not possible to replicate Article 57 of REACH with the objective hierarchy given in Figure 2.

Discussion

Limitations of Study

In terms of methods initially investigated, the present article is not comprehensive, there are other methods available, and it is possible that one of the unreviewed methods may outperform those reviewed in the present article.

In the assessment of GreenScreen®, the decision tree for organic chemicals was used. There are other versions of this decision tree (e.g., for inorganic chemicals), but these were excluded from this study. Within this decision tree, only hazards that were relevant to Article 57 of REACH were considered, i.e. hazards that were considered within the GreenScreen® decision tree, but not amongst the criteria for SVHCs, were excluded (e.g., physical hazards, such as explosiveness).

The MCDA methods used were limited to the MAUT approach; other MCDA methods such as ÉLECTRE were not considered in this study. Additionally, the present article focuses on one stage of the MCDA process, namely how to structure the problem. Other aspects central to the practical application of MCDA, such as gathering hazard data, dealing with data gaps, or accurately quantifying the preferences of different stakeholders, were not considered.

General assessment of the AoA methods

In this study, we assessed seven different AoA methods to determine their suitability for evaluating chemical alternatives to PFAS in the European marketplace under REACH. Two methods, GreenScreen® and MCDA, were selected for detailed investigation. It is worth noting that the five methods excluded from further investigation may well be applicable in different scenarios, such as different regulatory contexts or other specific market requirements (e.g., aiming for market leadership vs. fulfilling the legal requirements).

In the evaluation of the hypothetical substances, GreenScreen® deviated from the criteria in Article 57 of REACH for 9% of the substances. Specifically, it indicated higher hazard for combinations of PT_{eco} and BT_{eco} than REACH, while signaling lower hazard for PBT substances. In addition, mobility was not included in the scope of GreenScreen®, leading to an underestimation of the hazard of certain PFAS such as HFPO-DA.

The effect of varying MCDA method parameters was investigated in detail. It was found that substances with high hazard variability, such as PFAS, were particularly sensitive to changes of the MCDA parameter setting. This highlights the need for a careful selection of MCDA parameters in the assessment of chemical alternatives. The importance of MCDA parameter selection was highlighted further when the outcomes of GreenScreen® and MCDA were compared. The real-substances dataset illustrated that without fixed hazard thresholds, the MCDA outcome may not identify hazardous substances.

Within this study, attempts to find a combination of MCDA method parameters that aligned with the criteria of Article 57 of REACH were unsuccessful (see Figure 5). This may be because REACH relies on related objectives (e.g., PBT and vPvB both include persistence), whilst MCDA typically uses non-redundant objectives (e.g., see Figure 2). It may be possible to modify the objective hierarchy of MCDA so that the results align at the end with the criteria of Article 57 of REACH. The development and discussion of such a modified MCDA method is the subject of our accompanying paper London et al. [32]

Using GreenScreen® in AoA

Many smaller companies with limited resources and technical expertise may use GreenScreen® as it is user-friendly and relatively easy to implement. However, users should be aware of the fact that GreenScreen® does not fully reflect the criteria of Article 57 of REACH, and of the method's limited scope, excluding hazards such as mobility, GWP, and ODP. Finally, it is advisable to reevaluate whether the priorities, as set by the current version of the GreenScreen® decision tree, still align with recent scientific developments. For example, in the current GreenScreen® decision tree, the hazard of persistence alone is evaluated as Benchmark 3, indicating 'Use but with room for improvement.' However, there is increasing support for considering the hazard of persistence alone as being sufficient for regulation.[33]

Using MCDA in AoA

Researchers, larger organizations and governments may use MCDA in AoA as in MCDA it is possible to choose parameter settings, and thus set priorities, in contrast to GreenScreen® where parameters are set by the method's authors. However, care should be taken in the parameter selection as it is not possible to mimic the criteria laid down in Aticle 57 of REACH with MCDA if non-redundant parameters are used. A modified MCDA method is presented in our accompanying paper[32] that we recommend using in the future for the AOA when assessing chemicals under REACH.

Acknowledgement

The authors thank Kevin Kleemann for helpful feedback on the manuscript. This work received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement 860665 (PERFORCE3 Innovative Training Network).

Supporting Information Available

The Supporting Information-1 contains detailed descriptions of the investigated methods, published thresholds that can be used for the data normalization and further details on the investigations into Greenscreen® and MCDA. The Supporting Information-2 is a MS Excel workbook that shows the evaluation of the two methods, MCDA and Greenscreen®, using the hypothetical substances.

References

(1) ECHA, *Annex XV Restriction Report Proposal for Per- and polyfluoroalkyl substances*; Technical Report, 2023.

(2) Glüge, J.; Scheringer, M.; Cousins, I. T.; Dewitt, J. C.; Goldenman, G.; Herzke, D.; Lohmann, R.; Ng, C. A.; Trier, X.; Wang, Z. An overview of the uses of per- And polyfluoroalkyl substances (PFAS). *Environ. Sci.: Processes Impacts* **2020**, *22*, 2345–2373.

(3) Glüge, J.; London, R.; Cousins, I. T.; Dewitt, J.; Goldenman, G.; Herzke, D.; Lohmann, R.; Miller, M.; Ng, C. A.; Patton, S.; Trier, X.; Wang, Z.; Scheringer, M. Information Requirements under the Essential-Use Concept: PFAS Case Studies. *Environ. Sci. Technol.* **2021**, *56*, 6232–6242.

(4) OECD, *PFASs and Alternatives in Food Packaging (Paper and Paperboard) Report on the Commercial Availability and Current Uses*; Technical Report 58, 2020.

(5) OECD, *Per- and Polyfluoroalkyl Substances and Alternatives in Coatings, Paints and Varnishes (CPVs). Report on the Commercial Availability and Current Uses*; Technical Report 70, 2022.

(6) Cousins, I. T.; Goldenman, G.; Herzke, D.; Lohmann, R.; Miller, M.; Ng, C. A.; Patton, S.; Scheringer, M.; Trier, X.; Vierke, L.; et al., The concept of essential use for determining when uses of PFASs can be phased out. *Environ. Sci.: Processes Impacts* **2019**, *21*, 1803–1815.

(7) ECHA, *SVHC support document – HFPO-DA and its Salts/Acyl Halides*; Technical Report, 2019; pp 1–134.

(8) Cousins, I. T.; Dewitt, J. C.; Gï, J.; Goldenman, G.; Herzke, D.; Lohmann, R.; Ng, C. A.; Scheringer, M.; Wang, Z. The high persistence of PFAS is sufficient for their management as a chemical class. *Environ. Sci.: Processes Impacts* **2020**, *22*, 2307–2312.

(9) Rudin, E.; Glüge, J.; Scheringer, M. Per- and poly fluoroalkyl substances (PFASs) registered under REACH — What can we learn from the submitted data and how important will mobility be in PFASs hazard assessment ? *Sci. Total Environ.* **2023**, *877*, 162618.

(10) Arp, H. P. H.; Hale, S. E.; Schliebner, I.; Neumann, M. *Prioritised PMT/vPvM substances in the REACH registration database*; Technical Report 21/2023, 2023; pp 1–177.

(11) Jacobs, M. M.; Malloy, T. F.; Tickner, J. A.; Edwards, S. Alternatives assessment frameworks: Research needs for the informed substitution of hazardous chemicals. *Environ. Health Perspect.* **2016**, *124*, 265–280.

(12) Clean Production Action, *GreenScreen® for Safer Chemicals - Hazard Assessment Guidance*; Technical Report, 2018; pp 1–154.

(13) Cradle to Cradle Products Innovation Institute, *Material health assessment methodology*; 2022; pp 1–96.

(14) Eisenführ, F.; Weber, M.; Langer, T. *Rational decision making*; Springer: Berlin, 2010.

(15) OECD, *Cost-Benefit Analysis and the Environment*; OECD Publishing: Paris, 2018; p 456.

(16) MA TURI (Massachusetts Toxics Use Reduction Institute), P2OASYS Tool to Compare Materials Updated. 2021; `https://www.turi.org/Our_Work/Alternatives_Assessment/Alternatives_Assessment/Tools_and_Methods/P2OASys_Tool_to_Compare_Materials/P2OASys_Tool_To_Compare_Materials_Updated`, Webpage.

(17) US EPA, *EPA ' s Safer Choice Program Master Criteria for Safer Ingredients*; Technical Report, 2012; pp 1–31.

(18) IFA, *The GHS Column Model 2020*; Technical Report, 2020.

(19) European Parliament and Council, Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH). OJEU, 2006; Article 57.

(20) Faludi, J.; Hoang, T.; Gorman, P.; Mulvihill, M. Aiding alternatives assessment with an uncertainty-tolerant hazard scoring method. *J. Environ. Manage.* **2016**, *182*, 111–125.

(21) Beaudrie, C.; Corbett, C. J.; Lewandowski, T. A.; Malloy, T.; Zhou, X. Evaluating the Application of Decision Analysis Methods in Simulated Alternatives Assessment Case Studies: Potential Benefits and Challenges of Using MCDA. *Integr. Environ. Assess. Manag.* **2021**, *17*, 27–41.

(22) Zheng, Z.; Peters, G. M.; Arp, H. P. H.; Andersson, P. L. Combining in Silico Tools with Multicriteria Analysis for Alternatives Assessment of Hazardous Chemicals: A Case Study of Decabromodiphenyl Ether Alternatives. *Environ. Sci. Technol.* **2019**, *53*, 6341–6351.

(23) European Chemicals Agency, *Guidance on Information Requirements and Chemical Safety Assessment Chapter R.11: PBT/vPvB assessment - version 4.0*; 2023; pp 1–205.

(24) ECHA, *Guidance on the preparation of an Annex XV dossier for the identification of substances of very high concern*; Technical Report, 2018; pp 1–10.

(25) Haag, F.; Aubert, A. H.; Lienert, J. ValueDecisions, a web app to support decisions with conflicting objectives, multiple stakeholders, and uncertainty. *Environ. Model. Softw.* **2022**, *150*, 105361, Eawag, Switzerland: www.eawag.ch/en/DA/ValueDecisions.

(26) EAWAG, ValueDecisions MCDA App. 2022; `https://eawag.shinyapps.io/ValueDecisions/`, Web application.

(27) Langhans, S. D.; Reichert, P.; Schuwirth, N. The method matters: A guide for indicator aggregation in ecological assessments. *Ecol. Indic.* **2014**, *45*, 494–507.

(28) Marttunen, M.; Haag, F.; Belton, V.; Mustajoki, J.; Lienert, J. Methods to inform the development of concise objectives hierarchies in multi-criteria decision analysis. *Eur. J. Oper. Res.* **2019**, *277*, 604–620.

(29) US EPA, *EPA's Safer Choice and Design for the Environment (DfE) Standard*; 2023.

(30) ECHA, MSC unanimously agrees that HFPO-DA is a substance of very high concern. 2019; `https://echa.europa.eu/-/msc-unanimously-agrees-that-hfpo-da-is-a-substance-of-very-high-concern`, European Chemicals Agency (ECHA) website.

(31) UBA, *Protecting the sources of our drinking water: The criteria for identifying persistent, mobile and toxic (PMT) substances and very persistent and very mobile (vPvM) substances under EU Regulation REACH (EC) No 1907/2006*; Technical Report 127/2019, 2019; pp 1–87.

(32) London, R.; Glüge, J.; Scheringer, M. Modifying multi-criteria decision analysis for analysis of chemical alternatives. *in preparation* **2024**,

(33) Cousins, I. T.; Ng, C.; Wang, Z.; Scheringer, M. Why is High Persistence Alone a Major Cause of Concern? *Environ. Sci.: Processes Impacts* **2019**, *21*, 781–792.

TOC Graphic