Evaluating the environmental hazard of industrial chemicals from data collected during the REACH registration process

Mikael B. Gustavsson[a*], Andreas Hellohf[a], Thomas Backhaus[a]

[a]University of Gothenburg – Department of Biological and Environmental Sciences,
PO Box 461
SE 405 30 Göteborg
Visiting address: Carl Skottsbergs gata 22 B, 413 19 Göteborg

*Corresponding Author.
E-mail address: mikael.gustavsson@bioenv.gu.se

Keywords: PNECs, Assessment Factors, Daphnia magna, biocides, pharmaceuticals, priority pollutants

Abstract

Registration dossiers for 11678 industrial chemicals were retrieved from the database of the European Chemicals Agency, of which 3566 provided a numerical entry for the corresponding predicted no effect concentration for the freshwater environment (PNEC). A distribution-based examination of 2244 of these entries reveals that the average PNEC of an industrial chemical in Europe is 238 nmol/L, covering a span of 9 orders of magnitude. A comparison with biocides, pesticides, pharmaceuticals and WFD-priority pollutants reveals that, in average, industrial chemicals are least hazardous (hazard ranking: industrial chemicals << pharmaceuticals < pesticides < Water Framework Directive priority pollutants < biocides). However, 280 industrial chemicals have a lower environmental threshold than the median pesticide and 73 have a lower environmental threshold than even the median biocide. Industrial chemicals produced and/or imported in higher tonnages have, on average, higher PNECs which most likely is due to the lower assessment factors used for the PNEC determination. This pattern indicates that the initial AF of 1000 comprises a measure of conservatism. The vast majority of PNEC values are driven by EC50 and NOEC data from tests with *Daphnia magna*. Tests with marine species are rarely provided for the hazard characterization of industrial chemicals.

1 Introduction

REACH, Regulation No 1907/2006 on Registration, Evaluation and Authorization of CHemicals, is the European Union's (EU) main legislative framework for the environmental hazard assessment of industrial chemicals. REACH requires the submission of a registration dossier in order to allow for the production or import of an industrial chemical into the EU at volumes exceeding 1 tonnes per year. The amount of ecotoxicological data requested depends on the production or import tonnage: higher tonnages require the provision of more extensive datasets. For instance, compounds produced or imported at 1-10 tonnes per year only require the provision of results from a short term test with aquatic invertebrates and an algae growth inhibition test (REACH, Annex VII), while compounds produced or imported at 100-1000 tonnes per year require the addition of information from acute and chronic tests with fish and aquatic invertebrates (REACH Annex IX, see also (Tarazona et al., 2014)). All this information is collected within substance-specific dossiers hosted in a database maintained by the European Chemicals Agency (ECHA) (ECHA, Database). Currently ecotoxicological data for freshwater are more comprehensive than data for soil and sediments (Sobanska et al., 2014).

Two features of the REACH registration process should be noted: i) the information in the dossiers is supplied by the registrant itself and ii) the registrant is allowed to use data from closely related compounds in order to minimize testing. Thus, not all ecotoxicological data in a single dossier is for the actual dossier compound. Such data will be referred to as "read across" data in the rest of the text, following the REACH terminology.

Ecotoxicological data from the dossiers are used to determine the Predicted No Effect Concentration (PNEC) separately for each environmental compartment (i.e. freshwater, marine water, soil etc.). The PNEC represents an environmental threshold and is defined as a concentration "below which adverse effects in the environmental sphere of concern are not expected to occur". (REACH, Annex I Article 3.0.1). The PNEC is required information for compounds produced or imported at more than 10 tonnes per year, as well as for compounds classified as PBT (persistent, bioaccumulative or toxic) or vPvB (very persistent or very bioaccumulative). The majority of PNECs currently found in the REACH dossier database are determined by dividing the lowest EC50 or NOEC with an assessment factor (AF). These AFs are determined by the type and amount of ecotoxicity data available and range between 1 and 1000 for the freshwater environment. An AF of 1000 is used for the so-called base-set of data, which comprises acute EC50 values for algae, aquatic invertebrates and fish. The AF can be lowered by conducting additional ecotoxicological tests (ECHA, 2008).

Other groups of chemicals for which an environmental hazard assessment is mandatory include biocides (Regulation EU No 528/2012), plant protection products (Regulation EC 1107/2009), pharmaceuticals (Directive 2001/82/EC; Directive 2001/83/EC) and WFD priority pollutants (Directive 2000/60/EC). The hazard assessment is in principle carried out in a manner similar to REACH, but the resulting environmental thresholds are labelled differently in some of the different regulatory frameworks (see

Methods section for specifics). Additionally the use of a tonnage trigger for requesting more ecotoxicological data is a unique feature of REACH.

The data collected since REACH entered into force in 2007 offer a unique opportunity to provide an overview of the environmental hazards of industrial chemicals on the European market and to compare it to other environmentally relevant chemical classes. Therefore, the present paper presents i) a distribution-based summary of the hazard of REACH-registered chemicals to the freshwater environment; ii) a comparison with five other chemical classes (biocides, personal care products, pesticides, pharmaceuticals and WFD-priority pollutants); iii) an analysis of production volumes and ecotoxicological input data (species and taxonomic groups) as determinants for hazard estimates. Finally, we discuss these findings in the broader context of chemical hazard and risk assessment.

2 Methods

In the following we provide the details on data sources, filtering and merging.

2.1 REACH dossiers

ECHA hosts a database of the dossiers in which all chemicals currently registered under REACH are documented (ECHA, Database). Several dossiers might be present for a given chemical substance, if it is produced and/or imported by different companies. We retrieved the following information in March 2014 from this database: Substance name, CAS-number, molecular weight, information on production tonnage class, all ecotoxicological data, all PNEC data, and the AF used to determine each PNEC.

The following dossiers were excluded from further analysis:

1. Dossiers that assess chemical mixtures instead of individual compounds. This includes compounds labelled as UVCBs (“unknown or variable composition, complex reaction products or biological materials”), MCSs (“multi-constituent substances”) and any other dossier relating to mixtures rather than individual compounds.
2. Dossiers that document intermediates, i.e. compounds that are isolated as pure chemicals, but only used within a production chain (on-site or transported).
3. All notifications of new substances (NONS) as those do not comprise any tonnage information.
4. Dossiers in which the production / import tonnage class is kept confidential.
5. Entries without a CAS number.
6. Duplicates PNEC entries.
7. Dossiers lacking PNEC values for freshwater.
8. Compounds for which, even after manual search, no reliable molecular weight information could be found.

‘Dossiers lacking PNEC values for freshwater’ was the most important criterion excluding 8046 dossiers from further analysis. 1053 dossiers were then identified as UVCB/MCS compounds (criterion 1). In total, the filtering steps reduced the dataset from 11678 dossiers found in ECHA’s data to 2222 different dossiers, yielding 2244 unique PNECs that could be analyzed further (S.I. Table 1).

If no upper limit was provided for the tonnage class, the production/import volume was assumed to fall into the closest tonnage band. For example, compounds with a production/import volume of "10+ tonnes per year" were reclassified into the 10 – 100 tonnes per year class.

Harmonization of the ecotoxicity data was performed by streamlining entries for concentration endpoints (e.g. changing ECr50 to EC50), recalculating all test-durations to hours and all concentrations into µg/L and nmol/L. Any value provided as a range was assumed to equal the arithmetic mean of the range. Finally, all species names were spell-checked and run through NCBI's taxonomy database (NCBI, Taxonomy) in order to ensure that up-to-date species names were consistently used in the final database. If a species was not found in the taxonomy database a manual check using the primary literature was performed. This information was then used to check and update the ECHA-provided grouping of test species into fish, aquatic invertebrates and algae as it turned out that in a number of dossiers the test species were misclassified, this was corrected accordingly'

No specific information is provided in the dossiers regarding which test that is finally used to calculate the PNEC. Therefore, the following algorithm was used to determine the species that "drives" the PNEC (i.e. to determine the species that was exposed in the biotest whose result was used to calculate the numerical values of the PNECs):

1) All EC50 and NOEC values that are lower than 0.9*PNEC*AF or higher than 1.1*PNEC*AF were discarded.
2) All semi-quantitative data ("greater than" and "less than") were discarded.
3) The remaining dataset now contains only similar concentration-values for each compound (90% to 110% of the PNEC*AF). However, the data might still comprise a mixture of EC50 and NOEC values. For example, the following data situation might be encountered; PNEC = 1, AF = 100, EC50, species A: 100, NOEC, species B: 100. In such instances the NOEC value was disregarded, assuming that an EC50 value indicates a higher effect than the NOEC and thus indicates the more sensitive species / bioassay

All species that remained after these filtering steps were identified as "PNEC drivers" for a given compound.

All PNECs were recalculated from mg/L and µg/L into nmol/L, in order to avoid a bias in the comparative hazard characterizations. Finally, the whole suite of retrieval and filtering steps was manually checked using a set of 50 randomly selected chemicals.

2.2 Pesticides

Data were gathered by collecting the "conclusions on pesticides" reports from the European Food Safety Authority (EFSA) (EFSA, webpage). Each of these reports details EFSA's conclusions for a given active substance, based on the initial risk assessment carried out by the competent authority of a selected member state (so-called "rapporteur"). If more than one report with aquatic data was found for a given compound, only the report with the most recent set of aquatic data was used. Data were compiled by first discarding all formulation data so that only the data on active ingredient were considered in the following. If NOEC and EC50 values were available for aquatic invertebrates or fish the NOEC data were used; if NOEC and EC50 values for algae and higher plants/macrophytes were available, the EC50 data were used. This selection follows the EFSA guidance document (EFSA, 2013). In order to estimate the environmental threshold, the EC50 or NOEC respectively, from the most sensitive bioassay was divided with the corresponding trigger value (10 for algal and macrophyte EC50 data, 100 for fish and aquatic invertebrates EC50 data, 10 for fish and aquatic invertebrates NOEC data). These trigger values correspond to the AF's used in the context of REACH and the biocide regulation.

Finally, for 13 compounds mesocoms or SSD data was reported for the most sensitive taxa. In those instances the corresponding measurement and trigger value was used to determine the environmental threshold, in place of the single species assays..

This yielded a dataset of initially 403 documents, which was reduced to a final 298 compounds by filtering out datasets that i) did not contain any aquatic toxicity data (46), ii) were duplicates (42), iii) were considered as describing chemical mixtures or only contained formulation data (9), iv) did contain only semi-quantitative data (greater than, smaller than) (8).

2.2.1 Pharmaceuticals

Hazard data for pharmaceuticals were retrieved from a report published by the Norwegian Pollution Control Authority (Grung et al., 2007). The selection of pharmaceuticals included in the report is based on sales-information from Sweden and Norway. The PNECs have been determined using assessment factors which closely resemble the corresponding assessment factors as reported in the REACH guidance (assessment factors range between 10 and 1000, less guidance on which species that should be used for chronic testing in order to lower the assessment factor, ECHA, 2008; Grung et al., 2007).

The dataset was refined by excluding all duplicate entries , prioritizing experimental over modeled PNEC data (however, for 16 out of the final 142 entries only modelled PNECs were available) and removing all data on illicit drugs. One compound (Metacain) was excluded, as no PNEC was given. The final dataset contains 142 human and veterinary pharmaceuticals.

2.2.2 Water Framework Directive Priority Pollutants

Data were also collected for those chemicals flagged as "priority substances in the field of water policy" according to the Directive on environmental quality standards in the field of water policy (2008/105/EC, Annex II). The ecotoxicological data were retrieved from the individual WFD background documents stored in the European Communication and Information Resource Centre for Administrations, Businesses and Citizens. (CIRCABC). These documents describe the environmental threshold of each priority substance as quality standards (QS) for a number of different environmental compartments and for the present study the QS for the freshwater pelagic environment was used.

All compounds and compound groups as listed were recorded (yielding 39 entries) some of which were excluded as describing a group of compounds with variable molecular weight (PAH-group, recorded as individual compounds instead), and Chloroalkanes C10-C13) (2), as describing a group with variable molecular weight and QS (PBDE) (1), as describing measurements of metals and their compounds (5). The final dataset thus contains 31 different WFD priority pollutants.

2.2.3 Biocides

The biocide PNECs were gathered from ECHA (ECHA Biocides, webpage) incorporating all entries provided in June 2016. 161 reports were retrieved of which 85 were excluded: because they were identified as duplicates (68), concern gaseous compounds (5), provide no PNEC value (5), concern UVCBs (3), the PNEC was entered in relation to a background concentration (3), concern only formulations (1). The final dataset therefore contains 76 different active substances used in biocidal products on the European market.

2.3 Distribution Fitting & Statistics

Data distributions were characterized by providing minimum, maximum and median values. Additionally, all datasets were fitted to three different non-linear models (log-normal, log-logistic and Weibull) and then selecting the model with the lowest residual sum as the best fit. The fits were performed in R vers. 3.2.2 (R Core Team, 2016) using the 'drc' package vers. 2.5.12 (Ritz, 2016). Tukey's range test from the R package 'car' vers. 2.1-0 (John Fox and Sanford Weisberg, 2011) was used to identify differences between groups (See S.I. Tables 2-4).

3 Results

In the following we present the average environmental hazard of industrial chemicals in Europe, based on the information retrieved from ECHA. Afterwards we compare these compounds to biocides, pesticides, pharmaceuticals and WFD-priority pollutants. Finally we analyze production volumes and ecotoxicological input data as determinants for the PNEC estimates.

3.1 Hazard of European industrial chemicals to the aquatic environment

We retrieved dossiers for 11678 compounds from the ECHA database, of which 3566 compound-dossiers had a numerical entry for the freshwater PNEC. In the end 2244 PNECs fulfilled all selection criteria (see Methods section) and were further analyzed. Their PNEC's cover a span of more than 9 orders of magnitude ($2.4*10^{-3}$ to $4.2*10^{6}$ nmol/L). As the values were not normally distributed (Shapiro-Wilk normality test, $p < 2.2*10^{-16}$) the average PNEC was calculated as the median value (238 nmol/L) (Table 1, Figure 1). A nonlinear fit to the data results in a very similar estimate for the mid-point (203 nmol/L, see S.I. Table 2).

Additionally the corresponding $PNEC_{marine}$ was retrieved for 2141 of the 2244 freshwater PNEC entries. With 27.1 nmol/L its median value is 8.8 times lower than the median PNEC for freshwater. This is a direct consequences of the additional assessment factor of 10 by which a freshwater PNEC is divided in order to account for the greater biodiversity in marine water ecosystems (ECHA, 2008). In order words, testing an identical set of species groups would lead to a PNEC for the marine environment which is 10 times lower than the PNEC for freshwater. A ratio of 8.8 between the median $PNEC_{marine}$ and $PNEC_{freshwater}$ therefore indicates that ecotoxicological tests with marine species are rarely performed (S.I. Figure 1).

3.2 Comparison with other regulatory classes of chemicals

In addition to industrial chemicals figure 1 also presents the cumulative distribution of environmental thresholds for compounds from other regulatory classes. The median environmental thresholds follow the order: Industrial chemicals >> Pharmaceuticals > Pesticides > WFD-priority pollutants > Biocides (Table 1, for significance testing see S.I. Table 3). A similar pattern is present when looking at the lower 5% percentile as estimated by non-linear regression (S.I. Table 2). The ratio between the median hazard of the industrial chemicals and the pharmaceuticals, being the second least hazardous class, is 34, with the other classes being, on average, only slightly more hazardous. This reflects that all groups except the industrial chemicals are partly composed of compounds which are designed to be biologically active, often even intended to kill specific target organisms.

Despite the industrial chemicals being less hazardous on average, several of them have an environmental hazard in the same order of magnitude as pesticides and biocides. 280 industrial chemicals have a lower environmental threshold than the median of the pesticide group and 73 have a lower environmental threshold than the median biocide, the most hazardous group evaluated.

It should be pointed out that the environmental thresholds for pesticides are derived using a maximum assessment factor of 100 (EFSA, 2013), in contrast to all other groups where the maximum assessment factor is 1000 (EC, 2011; ECHA, 2008; ECHA, 2015; Grung et al., 2007). These differences might reflect the different protection goals in the different regulatory frameworks. Pesticides are intended to be used so that they "*do not have any unacceptable effects on the environment*" (EC No 1107/2009, Article 4), which implies that a certain effect magnitude and duration is deemed acceptable, in order to allow for industrial farming. In contrast, the PNEC for industrial chemicals and biocides is defined as a concentration "below which adverse effects in the environmental sphere of concern are not expected to occur". (EC 1907/2006, Annex 1 Article 3.0.1) implying that basically no adverse effect is deemed acceptable. The WFD defines environmental QSs more generally as a concentration "*which should not be exceeded in order to protect human health and the environment*" (Art 2 paragraph 35).

In total 63 compounds belong to more than one regulatory class with the largest overlap between the biocide and the pesticide group (21 compounds). The estimated environmental thresholds differ by no more than a median factor of 4.0 between all cases. The environmental thresholds estimated in the separate regulatory frameworks differ by more than a factor 100 for 3 chemicals. Two substances (fipronil and zeta-cypermethrin) are estimated to be less hazardous to the environment by factors of 70 000 and 1230, respectively, if used as pharmaceuticals, compared to a use as pesticides. In contrast lithium is considered 400 times less environmentally hazardous when used as an industrial chemical. For a full list of threshold-data and overlaps see S.I. Tables 5-6.

3.3 Relation between production/import tonnages and estimated environmental hazards

The REACH-dossiers provide the estimated total tonnage put on the European market (production plus import tonnage) in orders of magnitude (1-10 tonnes/year, 10-100 tonnes/year, etc). For higher tonnage classes more ecotoxicological data are requested, and consequently, lower AFs are used (EC 1907/2006, Article 3.3.1; EC 1907/2006, Annex VI-Annex X).

Figure 2 shows an increase in the median PNEC with increasing tonnages for all volumes above 10 tonnes per year, from 0.08 to 5.62 µmol/L (Table 2, for significance analysis see S.I. Table 4). This trend can in principle be due to either a decrease in the actual ecotoxicity of compounds from higher tonnage-classes, or can be caused by the use of smaller AF's. Indeed, smaller AF's are increasingly applied for assessing compounds from higher tonnage classes (Figure 3), with the exception of the highest tonnage class (10 000 000 to 100 000 000 tonnes per year), which comprises only nine compounds, of which three are evaluated using an AF of 1000 (chlorine dioxide, urea and ammonium). The trend towards increasing PNECs therefore does not seem to be caused by a decrease in the compound's ecotoxicity, but rather is a result of using lower AF's. In order words, providing additional ecotoxicity data and consequently using a lower AF typically leads to a lower estimated hazard.

3.4 Which species drive the PNEC?

In order to characterize the relative importance of the various test species for the hazard assessment, we determined how often the EC50 or NOEC of a particular species has been used to derive the numerical value of the PNEC. Those species are termed "PNEC drivers" in the following.

For 212 dossiers several numerically identical E50 and NOEC values were retrieved (each corresponding to 90-110% of PNEC*AF, see material and methods). As there is no further information given in the data made available by ECHA, it is not possible to determine with absolute certainty which value was used to determine the PNEC (the "PNEC driver"). Under these circumstances, all NOEC values were discarded from further analysis. This strategy is based on the assumption that species whose growth, reproduction or physiology is affected by 50% at the given concentration are more sensitive than species for which the same concentration only corresponds to a NOEC. If numerically identical EC50 values were retrieved for two or more species, all of them were retained for the following analysis. In total it was possible to identify the PNEC drivers for 1666 out of the 2244 PNECs initially collected from ECHA's database, (see S.I. Table 7).

Table 3 presents the results of this analysis for the three most commonly used species, considering data from algae, aquatic invertebrates and fish. It shows that *Daphnia magna* is the most commonly used species (used for testing 1 609 chemicals), followed by *Selenastrum capricornutum* (used for testing 1013 chemicals) and *Desmodesmus subspicatus* (used for testing 770 chemicals). *Daphnia magna* is also the PNEC driver for almost half of the compounds (701 of 1 666), completely dominating the group of aquatic invertebrates and making it the most important group of test species overall. This corresponds well with the pattern previously identified by Tarazona (Tarazona et al., 2014).

4 Discussion

In 2014 the 28 member states of the European Union produced 140.0 million tonnes of chemicals classified as hazardous to the aquatic environment (acutely and/or chronically toxic), *http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=env_chmhaz* (Eurostat, 2016). This huge toxic potential emphasizes the need for a reliable and robust system for chemical risk assessment and management, for which high quality data are the key prerequisite. The data collected and made publically available in the ECHA database also allow a characterization of broader patterns, as presented in this paper. However, the reliability of all these estimates is obviously strictly dependent on the quality of the data collected in the REACH dossiers.

80% of the initially retrieved data could not be analyzed further, mainly because freshwater PNEC's were missing in ECHA's database. The reason for these datagaps are currently unclear. A check, database completion and follow-up study including these compounds would therefore certainly be valuable. However, there are currently no indications that the analysis of the 20% of compounds for which we were able to retrieve PNEC values resulted in biased hazard estimates.

The 5 year update of the REACH baseline study highlighted that the regulation led to a marked increase in the quality of the toxicological, ecotoxicological and exposure-related information available (Eurostat, 2012), based on a sub-sample of 62 chemicals. In contrast, a recent in-depth evaluation of the REACH dossiers of 1814 high production volume chemicals by the German Environment Agency (UBA, 2015) revealed that the submitted ecotoxicological data were fully REACH-compliant for only 26% of the chemicals. 9% of the datasets were identified as non-compliant and a full 65% were classified as undecidable, i.e. containing substantial data gaps. Such a systematic quality check was beyond the scope of the present study, we took the data in the REACH dossiers at face value. However, the fact that we frequently encountered misspelled species names and missing or inconsistent data entries give reason for concern as it hampers the comparison of hazard profiles of different chemicals. It should also be noted that all data presented and evaluated in the dossiers are collected by industry. This results in a clear conflict of interest, i.e. the desired outcome is to demonstrate the safe use of the assessed chemical according to the REACH criteria. It has already been demonstrated in other areas of chemical assessment that such situations might bias data compilation and evaluation (e.g. Lundh et al., 2012).

Taken together this indicates the need for a continuous, impartial (as far as reasonably possible) quality-control of the REACH registration dossiers. In this context it might be argued that fulfilling the legal obligation of ECHA to conduct a compliance check of 5% of the dossiers (a minimum that is set in REACH Article 41) is insufficient.

Very few studies (Austin et al., 2015; Igos et al., 2014; Müller et al., 2016) have started to explore the usefulness of the public data compilation provided by ECHA as a source for detailed retrospective hazard and risk analyses. None, to the best of our knowledge, has provided a comparative hazard characterization across regulatory silos as presented in this paper. Efforts to identify broad patterns in the hazard and risk profiles of chemicals on the European market might be especially hampered by the interface to the ECHA database that focusses on manual dossier retrievals, substance-by-substance, but does not support an automated data collection. Additionally, given the complexity of the data and the current reproducibility crisis in empirical sciences (Baker, 2016; Dekant, 2016), we feel that all data that form the basis of a paper should be available for independent scrutiny and critique. We therefore provide the data collection from the present study on Github, at https://github.com/ThomasBackhausLab/Environmental-Thresholds.git

The appropriate sizing of assessment factors used to account for uncertainties in hazard assessments is subject to a continuous evaluation and debate, e.g. Chapman et al. (1998), Falk-Filipsson et al., (2007), Malkiewicz et al. (2009). Our results indicate that extended datasets result, in average, in higher PNECs. This indicates that, as intended, the initial AF of 1 000 comprises a measure of conservatism and additional data, in conjunction with lowered AF values, therefore generate higher PNECs. Further evaluations might target the question which factors (intra-laboratory variability, acute to chronic, lab to field extrapolations, species sensitivity distributions) are important components of the overall uncertainty.

Very few species govern the collected hazard assessments, with *Daphnia magna* being both the most dominant, with respect to the sheer number of dossiers that provide test data from this species, and in terms of the number of chemicals for which this species is the PNEC driver. This species has a wide geographical distribution, which is advantageous for European-wide hazard assessments. However, none of the three most commonly used fish species is native to Europe. The geographic origin of the tested species, fortunately, does not appear to impact hazard estimates based on species-sensitivity distributions (Hagen et al., 2014; Maltby et al., 2005), but similar analyses for hazard estimates based on point estimates seem to be currently missing. Data on marine species are scarce and hence no conclusions can be derived from the available data on whether the recommended assessment factor of 10 for the extrapolation to marine life is sufficient. More data, especially on exclusively marine organism groups such echinoderms or brachiopods, would be needed for this evaluation.

Industrial chemicals are by a factor of 34 less hazardous, in average, than any other evaluated chemical class. However, the distribution of hazard estimates covers 9 orders of magnitude, and almost 300 industrial chemicals have a hazard exceeding that of an average biocide. It might be worth to further analyze whether and to what extend those chemicals have common chemical structures, for example in order to guide future developments along the principles of green chemistry, i.e. to design future chemicals with minimum toxicity.

Finally, it should be noted that a risk-analysis of the evaluated chemicals is currently not possible, as predicted environmental concentrations (PEC's) are not provided in the dossiers. Consequently, it seems to be accepted that chemicals co-occur in the same environment, which obviously also reflects actual exposure situations in which the various environment compartments, as well as humans, are exposed to complex chemical mixtures. However, it is well established, that mixture risks might substantially exceed the risk of each individual component (see reviews in e.g. Kortenkamp et al. 2009). It has to be concluded therefore, that the assessment under REACH might systematically underestimate actual environmental risks. Given that exposure estimates are not available, it is currently not possible to evaluate whether the environmental risk due to chemical exposure is actually on an acceptable level.

Conflict of interest

None

Funding

This work was implemented in the context of the Centre for Future Risk Assessment and Management at the University of Gothenburg (FRAM). Parts of this work were also funded by the European Commission in the context of the 7th Framework Programme, R&D project "SOLUTIONS", under grant agreement No. 603437.

References

Austin, T., Denoyelle, M., Chaudry, A., Stradling, S. and Eadsforth, C., 2015. European Chemicals Agency dossier submissions as an experimental data source: Refinement of a fish toxicity model for predicting acute LC50 values. Environmental Toxicology and Chemistry, 34(2), pp.369-378. doi:10.1002/etc.2817

Baker, M., 2016. 1,500 scientists lift the lid on reproducibility. *Nature* 533.7604 pp. 452-454.

CIRCABC. Communication and Information Resource Centre for Administrations, Businesses and Citizens. https://circabc.europa.eu/w/browse/b55f4c81-d664-43db-8b27-264b26a7424b (available 10.24.16)

Chapman, P.M., Fairbrother, A. and Brown, D., 1998. A critical evaluation of safety (uncertainty) factors for ecological risk assessment. Environmental Toxicology and Chemistry, 17(1), pp.99-108. doi: 10.1002/etc.5620170112Deker, W. 2016. Toxicology and the reproducibility crisis: Scientific publishing, hazard assessment and risk characterization. Toxicology Letters. 30, 236 pp. 76-77. doi:10.1016/j.toxlet.2016.09.001

Falk-Filipsson, A., Hanberg, A., Victorin, K., Warholm, M. and Wallén, M., 2007. Assessment factors—applications in health risk assessment of chemicals. Environmental research, 104(1), pp.108-127. doi:10.1016/j.envres.2006.10.004

EC, 2011. Guidance Document No. 27, Technical guidance for deriving environmental quality standards. ISBN 978-92-79-16228-2. doi: 10.2779/43816

ECHA, Biocide. Biocidal active substances. http://echa.europa.eu/information-on-chemicals/biocidal-active-substances (available 10.24.16)

ECHA, Database. Search for chemicals. URL https://echa.europa.eu/search-for-chemicals (available 10.24.16).

ECHA, 2008. Guidance on information requirements and chemical safety assessment. Chapter R.10: Characterisation of dose [concentration]-response for environment.

ECHA, 2008b. Guidance on information requirements and chemical safety assessment Chapter R.16: Environmental Exposure Estimation

ECHA, 2015. Guidance on the Biocidal Products Regulation Volume IV Environment - Part B Risk Assessment (active substances) ISBN 978-92-9247-093-7.

EFSA, webpage. Search results for 'Conclusions on Pesticides'. http://www.efsa.europa.eu/en/search/site/?f[0]=sm_hierarchy_type%3Aconclusion_on_pesticides (available 10.24.16)

EFSA, 2013. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. EFSA Journal 2013;11(7):3290, 268 pp. doi:10.2903/j.efsa.2013.3290.

Eurostat, 2012. The REACH baseline study 5 years update – comprehensive study report. doi: 10.2785/33723

Eurostat, 2016. Energy, transport and environment indicators. doi:10.2785/138586

John Fox and Sanford Weisberg (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion, https://cran.r-project.org/package=car (available 10.24.16)

Grung, M., Heimstad, E., Moe, M., Schlabach, M., Svenson, A., Thomas, K., Woldegiorgis, A., 2007. Human and veterinary pharmaceuticals, narcotics, and personal care products in the environment, Swedish Environmental Research Institute. doi:10.1139/m88-128

Hagen, T.G. and Douglas, R.W., 2014. Comparative chemical sensitivity between marine Australian and Northern Hemisphere ecosystems: Is an uncertainty factor warranted for water-quality–guideline setting?. Environmental toxicology and chemistry, 33(5), pp.1187-1192. doi: 10.1002/etc.2548

Igos, E., Moeller, R., Benetto, E., Biwer, A., Guiton, M. and Dieumegard, P., 2014. Development of USEtox characterisation factors for dishwasher detergents using data made available under REACH. Chemosphere, 100, pp.160-166. doi: 10.1016/j.chemosphere.2013.11.041

Kortenkamp, A., Backhaus, T., & Faust, M. 2009. State of the Art Report on Mixture Toxicity, Report to the European Commission.

Lundh, A., Sismondo, S., Lechin, J., Busuioc, O.A., Bero, L. 2012. Industry sponsorship and research outcome. Cochrane Database of Systematic Reviews. Issue 12. doi: 10.1002/14651858.MR000033.pub2

Malkiewicz, K., Hansson, S.O. and Rudén, C., 2009. Assessment factors for extrapolation from short-time to chronic exposure—Are the REACH guidelines adequate?. Toxicology letters, 190(1), pp.16-22. doi: 10.1016/j.toxlet.2009.06.858

Maltby, L., Blake, N., Brock, T. and Van den Brink, P.J., 2005. Insecticide species sensitivity distributions: importance of test species selection and relevance to aquatic ecosystems. Environmental Toxicology and Chemistry, 24(2), pp.379-388. doi: 10.1897/04-025R.1

Müller, N., de Zwart, D., Hauschild, M., Kijko, G. and Fantke, P., 2016. Exploring REACH as a potential data source for characterizing ecotoxicity in life cycle assessment. Environmental Toxicology and Chemistry. doi:10.1002/etc.3542

NCBI Taxonomy. Taxonomy Database. https://www.ncbi.nlm.nih.gov/taxonomy (available 10.24.16)

R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

Ritz, C., Baty, F., Streibig, J. C., Gerhard, D. (2015) Dose-Response Analysis Using R. PLOS ONE, 10(12). doi:10.1371/journal.pone.0146021. https://cran.r-project.org/package=drc (available 10.24.16)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2785v2 | CC BY 4.0 Open Access | rec: 20 Mar 2017, publ: 20 Mar 2017

Sobanska, M.A., Cesnaitis, R., Sobanski, T., Versonnen, B., Bonnomet, V., Tarazona, J. V, De Coen, W., 2014. Analysis of the ecotoxicity data submitted within the framework of the REACH Regulation. Part 1. General overview and data availability for the first registration deadline. Sci. Total Environ. 470–471, 1225–1232. doi:10.1016/j.scitotenv.2013.10.074

Tarazona, J. V, Sobanska, M.A., Cesnaitis, R., Sobanski, T., Bonnomet, V., Versonnen, B., De Coen, W., 2014. Analysis of the ecotoxicity data submitted within the framework of the REACH Regulation. Part 2. Experimental aquatic toxicity assays. Sci. Total Environ. 472, 137–145. doi:10.1016/j.scitotenv.2013.10.073

UBA, 2015. REACH Compliance: Data Availability of REACH Registrations Part 1: Screening of chemicals > 1000 tpa. ISSN 1862-4804

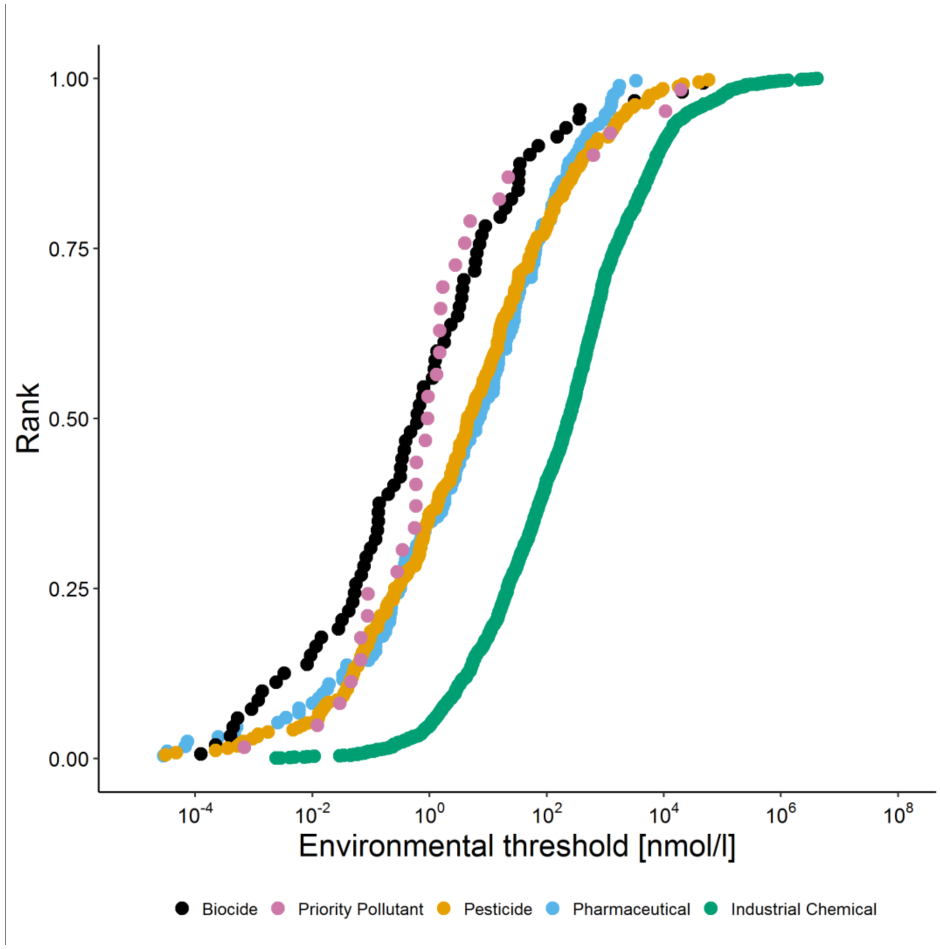

Figure *1*: The cumulative distributions of environmental threshold values for biocides, WFD priority pollutants, pesticides, pharmaceuticals and industrial chemicals.

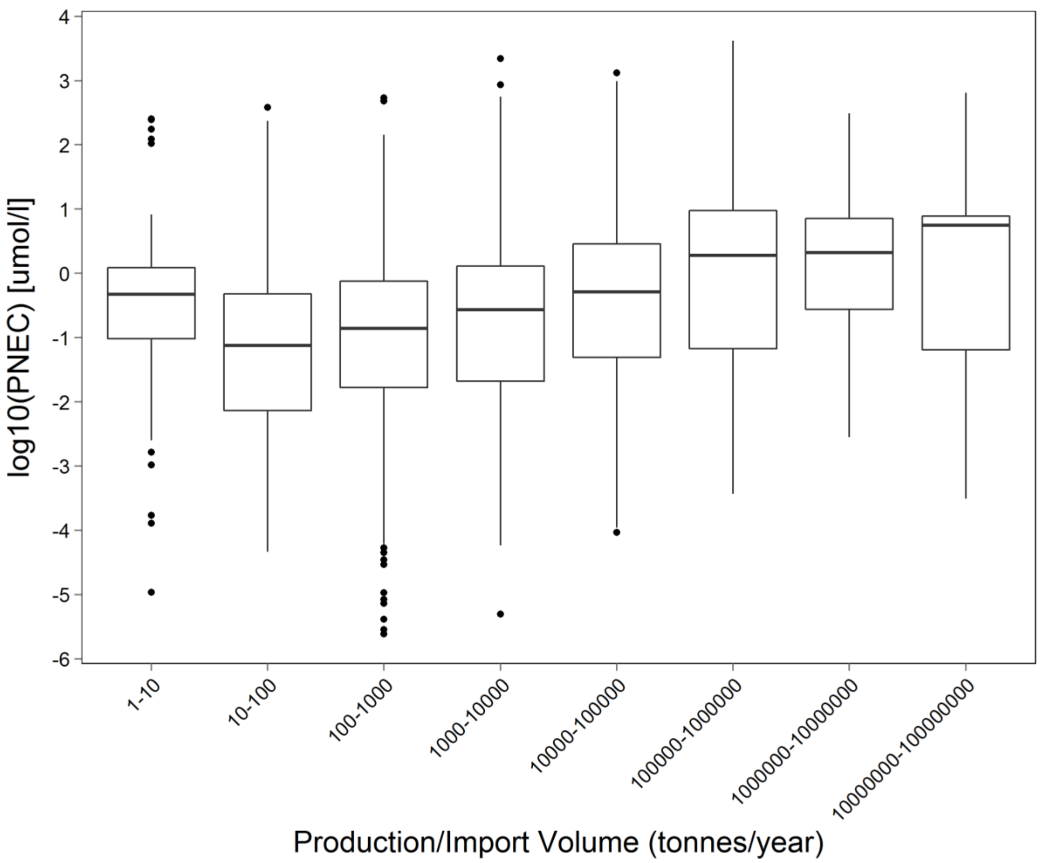

Figure 2: Distribution of PNEC values of industrial chemicals per tonnage class.

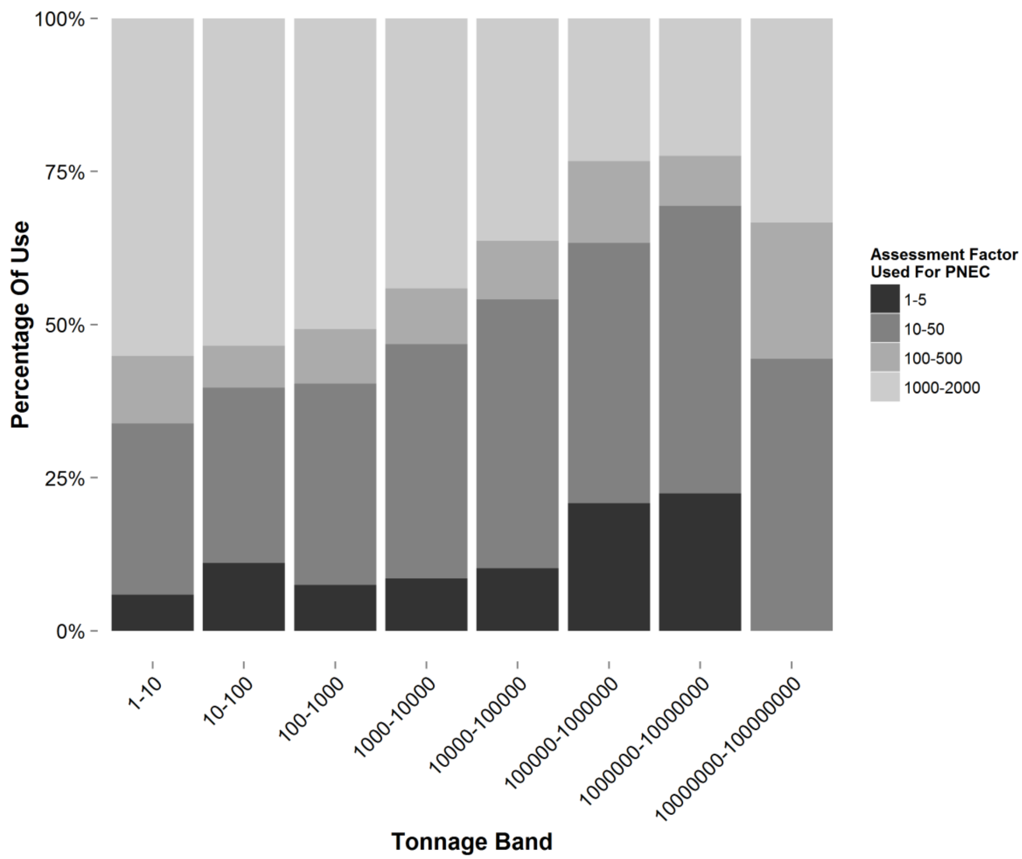

Figure 3: Size of assessment factors used for the PNEC calculation, in dependence of production/import volumes. 14 industrial chemicals from various tonnage classes used an AF of 2000. In comparison 983 PNEC's have an AF of 1000.

Table 1: Summary statistics for the 5 different regulatory chemical classes: maximum, minimum and median environmental threshold per class as. Model fits and further details provided in the supporting information

Group	Number of chemicals	Max [nmol/l]	Min [nmol/l]	Median [nmol/l]
Industrial Chemical	2244	4166667.0	2.4E-03	237.8
Pharmaceutical	142	3332.5	2.9E-05	7.0
Pesticide	298	57921.6	9.9E-06	4.5
Priority Pollutant	33	19427.8	6.8E-04	0.9
Biocide	76	46929.6	1.3E-04	0.6

Table 2: Maximum, minimum and median environmental threshold per tonnage class for industrial chemicals

Tonnage Class [tonnes per year]	Number of compounds	Max [µmol/l]	Min [µmol/l]	Median [µmol/l]
1-10	136	253.3	1.09E-05	0.47
10-100	189	380.9	4.63E-05	0.08
100-1000	771	538.7	2.42E-06	0.14
1000-10000	630	2195.9	4.98E-06	0.27
10000-100000	303	1329.5	9.25E-05	0.52
100000-1000000	120	4166.7	3.71E-04	1.92
1000000-10000000	49	311.7	2.82E-03	2.10
10000000-100000000	9	649.2	3.11E-04	5.62

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2785v2 | CC BY 4.0 Open Access | rec: 20 Mar 2017, publ: 20 Mar 2017

Table 1: Most commonly used test species from each organism group (algae, invertebrates, fish)

Total number of chemicals analyzed is 1 666. The column 'No of chemicals tested' shows how often each species has been tested (in absolute numbers). 'Identification as PNEC Driver' shows how often a species was identified as PNEC driver, in absolute and relative numbers (as percentage of the number of chemicals tested with each species). For further details see text.

Species	Taxa	No of chemicals tested	Identification as PNEC driver	
			Absolute	Percentage
Selenastrum capricornutum	Algae	1013	293	29%
Desmodesmus subspicatus	Algae	770	203	26%
Skeletonema costatum	Algae	159	8	5%
Sum within group	-	1942	504	
Daphnia magna	Aquatic Invertebrate	1609	701	44%
Ceriodaphnia dubia	Aquatic Invertebrate	164	56	34%
Americamysis bahia	Aquatic Invertebrate	129	5	4%
Sum within group	-	1902	762	
Oncorhynchus mykiss	Fish	739	152	21%
Danio rerio	Fish	705	103	15%
Pimephales promelas	Fish	577	101	18%
Sum within group	-	2021	356	