Effects of global change on the emission, fate, effects and risks of chemicals in aquatic ecosystems

Marie Curie Actions

SHORT TITLE:	ECORISK2050
COORDINATOR:	Prof Dr Paul van den Brink
ORGANISATION:	Wageningen University
TOPIC:	H2020-MSC_ITN 2018
PROJECT NUMBER:	813124

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813124

Deliverable: D6.2 Prioritisation of pharmaceutical pollutants by environmental risk

Contributing:

Samuel A. Welch (ESR12), **NIVA** (Norwegian Institute for Water Research)

Jannicke Moe (beneficiary, supervisor), **NIVA**

Merete Grung (beneficiary, supervisor), **NIVA**

Mohammad Nouri Sharikabad (partner organisation, supervisor) **FHI/NIPH** (Norwegian Institute of Public Health

Knut Erik Tollefsen (beneficiary, supervisor), **NIVA**

Kristine Olsen (partner organisation), **FHI**

1. Introduction

Pharmaceuticals play a pivotal role in both the human and environmental risk landscapes, and careful management of their use is required to mitigate environmental risk as far as possible without impacting human health.

Presently, in the European Union, pharmaceuticals are regulated by the European Medicines Agency (EMA), being assessed against a panel of human health metrics to determine whether a product will be granted marketing authorisation. An Environmental Risk Assessment (ERA) is nominally required for all pharmaceutical substances, but exemptions are common and environmental risk is not considered in final authorisation decisions (EMA 2006).

This has created a system with few incentives to environmentally risk assess pharmaceuticals, and data on both the environmental exposure and effects of these substances can be hard to acquire. Furthermore, where environmental risk assessment is conducted, focus usually lies on toxicity-driven risk, leaving bioaccumulation and especially persistence data even less available.

Deliverable D.6.2 of the ECORISK 2050 project is described in the original proposal document as a "*Prioritisation of pharmaceuticals and PHCP* (Personal Health Care Products) *which pose highest risk*". Based on our work with the Norwegian Institute of Public Health (NIPH or FHI), we have created a prioritisation list, using our wholesale-derived predicted exposure data and publicly available toxicity data to calculate a risk quotient, and supplementing this with persistence and bioaccumulation data where available. As no public record of pharmaceuticals *and* personal care product sales in Norway exists, and due to the generally lower toxicity posed by the latter group of substances, we elected to focus this exercise entirely on pharmaceutical substances.

We present, then, below, a prioritisation list of pharmaceuticals sold in Norway between the years 2016-19, ordered by the toxicity-based risk quotient and with persistence, mobility and bioaccumulation data appended where available.

2. Methods

Sales data for years 2016-2019 w extracted from the Norwegian Drugs Wholesale Database (Figure 2, Sales data), covering all sales to pharmacies, hospitals, nursing homes, and non-pharmacy outlets licensed to sell drugs within Norway, including prescriptions, over-the-counter sales, and procurement by medical establishments (NIPH 2019). In its raw form this data consisted of per-product sales, such as a packet containing multiple sheets of pills, or a suspension of liquid medicine.

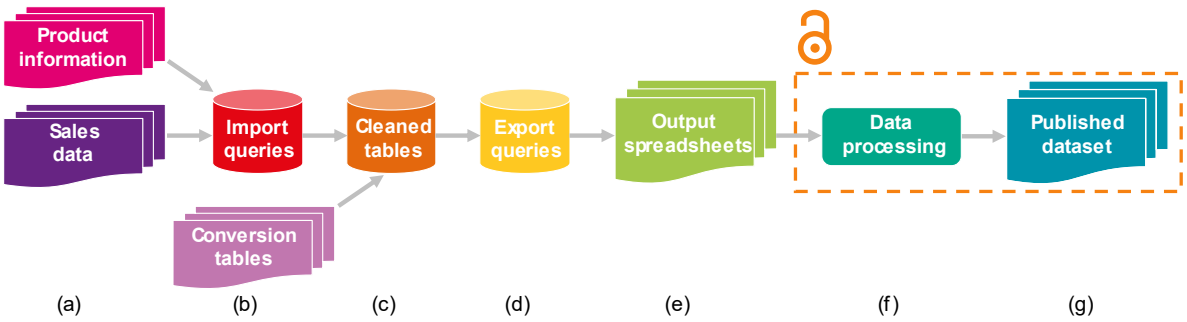

Figure 1: Simplified diagram of data extraction and management pipeline. Data and code to be made publicly available denoted by the dashed orange box.

In adherence with NIPH's commercial confidentiality requirements, all data corresponding to monetary values, product names and quantities of packet sold were excluded from the dataset to be published.

Additional information on individual products (Figure 1, (a) Product information) including number of items per package, quantity of active pharmaceutical ingredients (API) per item, and associated unit

were obtained separately from the centralised NIPH sales database and matched to sales data using internal product codes. In cases where no additional data was available for given products or where automatic matching failed, records were checked manually against product contents records online, principally the Norwegian pharmaceuticals specialities site Felleskatalogen, the UK Electronic Medicines Compendium, and the US resource Drugs.com. Cases where one product contained two or more APIs (combination drugs) were split into separate entries for each API to ensure substances were fully accounted for.

Data processing in Access

The two data sources were imported into a Microsoft Access database (Figure 1, b, c, d, e) and organised into a related set of tables. The main table types were data tables, conversion tables, and code lists. The main data tables are described below and summarised in Figure 2.

1) **t_Product:** the description of each pharmaceutical product (defined by product number), including information on the product type and the product amount per package
2) **t_Product_API:** the concentration of each API per product
3) **t_Sales_Product:** the number of packages sold per product per year

The main code lists, which contain the ATC and API codes, summarised in Figure 3:

1) **t_Code_ATC:** definition of ATC codes found in the sales data
2) **t_Code_API:** a complete list of APIs found in the sales data

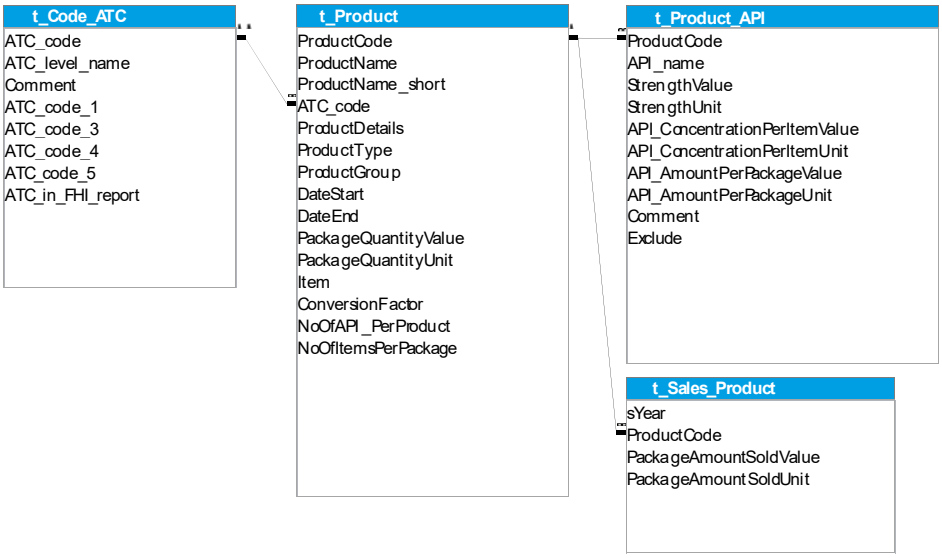

Figure 2: Simplified diagram of database structure: the main data tables.

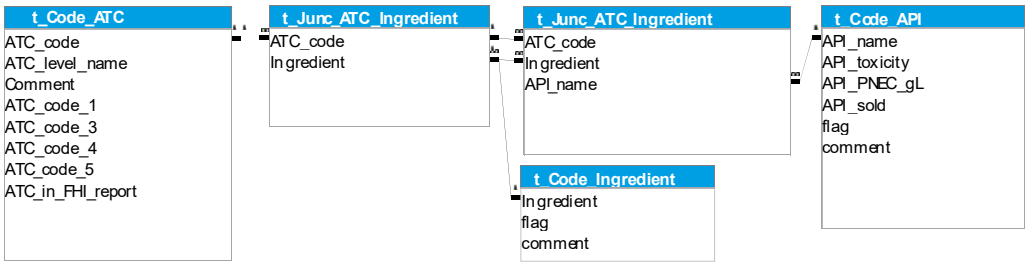

Figure 3: Diagram of code lists and conversion tables, defining the many-to-many relationships between ATC, ingredient, and API in the database.

Information on APIs contained in each product was not available in the original data sources, and instead had to be extracted from the ATC codes associated with the sales data. The many-to-many relationship between ATC and API is represented by the code lists and junction tables shown in Figure 3. Ingredient was added as an intermediate step, accounting for ingredients considered distinct pharmaceutical agents under the ATC code system that were not strictly APIs. Subsequently, each product, the associated API names associated were extracted from the full ATC name and entered in the table t_Product_API.

In most cases the information needed for calculating the amount of API per package (i.e., the concentration of API in the product and the amount of the product per package) was available in the original data source (the product information table). Where this information was not provided, we extracted the information manually from the Product name.

For products where API information could not be found in the included data, it was instead sourced for each individual product from the Norwegian pharmaceutical specialties website Felleskatalogen or Summaries of Product Characteristics (SPCs) from the pharmaceutical specialties websites of other nations. This was also the case for combination products containing two or more APIs.

Finally, the information on yearly sales (number of packets) per product was stored in the table t_Sales_Product. This information was combined with the calculated amount of API per product package during data extraction to obtain the total amount of API per year from the sales data. Data were exported into flat files for calculation of predicted environmental concentrations and analysis in the future.

Data Processing in R

The exported dataset (Figure 1, e, f), was prepared for analysis and publication in R version 4.1.1 "*Kick Things*" (R Core Team 2021).

Sales weights per product per year were filtered to remove any zero values, and values for which no units were assigned, representing records for which the API amount could not be calculated (Figure 4). Sales weights were then summed by API by year, and APIs were filtered according to a list of exemptions from risk assessment on the basis of non-toxicity (as applies to vitamins, vaccines, antibodies, etc. (EMA 2006)). The final dataset will be published as a comma-separated values (.csv) file.

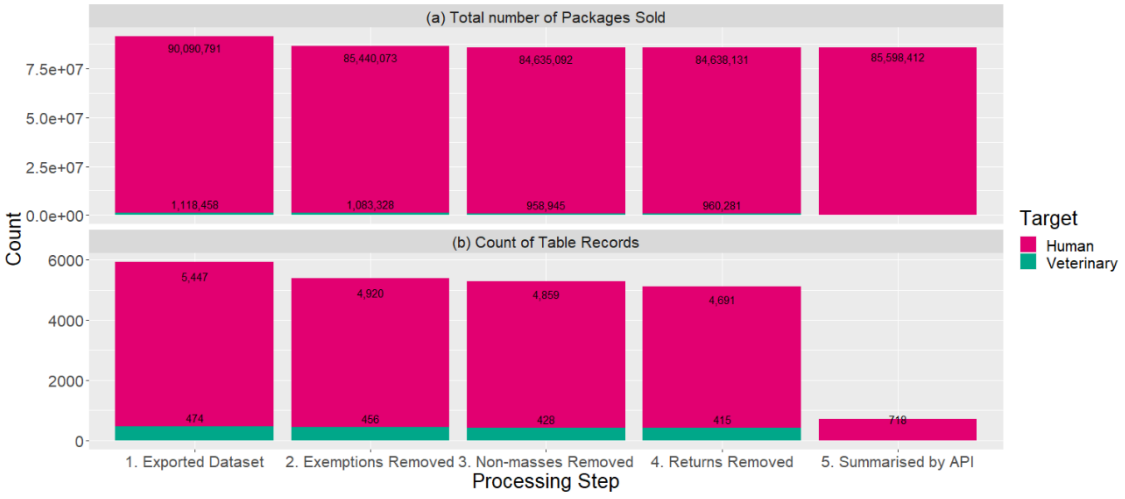

Figure 4: Count of overall packages (a) sold in 2019, and unique table records (b) retained at each step of data processing, categorised as human or veterinary. Aggregation by API condenses the dataset to 718 records, but all component packages are still accounted for (no loss of records (a) between steps 4 and 5. =

3. Risk Assessment

PMBT Concepts

Environmental risk assessment of chemicals is primarily driven by comparing measured or predicted exposure concentrations to toxicity (T) expressed as predicted no-effect concentrations. However, EMA risk assessment also considers the persistence (P) of chemicals – measured as their half-life in a variety of aquatic or terrestrial environments – and their potential to bioaccumulate (B) – typically based on the chemical's affinity for the fatty alcohol octanol in an octanol-water system. These three properties are often abbreviated to PBT.

Beyond this, in recent years growing attention has been paid to the mobility (M) of substances in the environment as another important driver of risk (Rüdel et al. 2020). Substance mobility is determined by affinity for organic carbon in a carbon-water system. Physical test data for PMB is frequently unavailable, and consequently where possible estimated values are used for these parameters; see the section QSARs below for more details.

Predicted Environmental Concentrations

Predicted Environmental Concentrations of individual APIs in the compartment Surface Water were calculated using a modified form (Eq. 1) of the standard refined PEC_{SW} equation outlined in the EMA's guidelines for pharmaceutical environmental risk assessment (2006). As no specific bodies of water are specified in the guidelines, the model is assumed to apply to all relevant freshwater bodies, i.e., rivers and lakes.

$$PEC_{SW} = \frac{g\ of\ API\ sold \times (1 - WWTP\ Removal)}{365 \times Wastewater\ consumption \times Population \times Dilution\ factor} \quad \text{(Eq. 1)}$$

Table 1: Table of PEC_{SW} equation variables and parameters

Component	Unit	Description
g of API sold	*g*	The total weight (g) of an API sold in a year
WWTP removal	*unitless*	The proportion of the API removed at WWTP (default of 0)
365	*days*	The number of days in a year
Wastewater consumption	*L person^{-1} day^{-1}*	The average wastewater consumption (L) of the population of a given area per day
Population	*persons*	The population of a given area
Dilution factor	*unitless*	The ratio of dilution between WWTP effluent and receiving waters (default of 10)

Whereas the standard equation estimates sales weights from the maximum dose of a given API and the proportion of people in a population taking that API, by using our dataset of pharmaceutical wholesales we can input an exact figure for consumption across the entire population of Norway.

PECs were individually calculated per API, per year, using information on yearly average wastewater consumption and Norwegian population, obtained from Statistics Norway.

Predicted No Effect Concentrations

Toxicity data was obtained in the form of Predicted No Effect Concentrations for 257 APIs from the Norwegian Pharmaceutical Specialties website Felleskatalogen (2020), and converted from the original concentrations (typically µg/L) into g/L for internal consistency. These PNECs were originally calculated by the Swedish Pharmaceutical Specialities website, FASS.se (2019), where full equations and constituent test data were given. However, these data could not easily be converted into a machine-readable format and hence Felleskatalogen's more accessible, but less transparent dataset was used. In any case, a full account of the toxicity data's origin is impossible, as the studies that produced said data are not to the authors' knowledge publicly available.

Risk Quotients

Predicted risks per API per year were calculated as simple Risk Quotients following the standard ecotoxicological method (Equation 2):

Equation 2: Calculation of Risk Quotient (RQ) in surface water from Predicted Environmental Concentration (PEC) and Predicted No Effect Concentration (PNEC). (EMA 2006)

$$RQ_{Surface\ water} = \frac{PEC_{Surface\ water}}{PNEC} \quad \text{(Eq. 2)}$$

QSARs

PMBT information for APIs is not readily available. To canvas data efficiently for 774 substances, the OPERA (Open (Q)SAR) app (EPA 2018) was used to generate estimated parameters through Qualitative Structure-Activity Relationships (QSARs), comparing APIs to substances with similar structures.

The following parameters were generated using OPERA:

Table 2: QSAR PMT Parameters generated for APIs sold 2016-19 in Norway.

Property	**Parameter**	**Comment**	**Thresholds**
Persistence	Biodegradation half life	Medium not specified, but assumed to be freshwater	Half-life > 60 days: Persistent Half-life > 180 days: Very Persistent
Bio-accumulation	Octanol/water adsorption coefficient	Ratio of lipophilicity/hydrophilicity parameter used as a trigger value for BCF/PBT assessment	At a pH of 5-9 Log K_{ow} ≥ 3: BCF assessment trigger Log K_{ow} ≥ 4.5: PBT assessment trigger
	Fish bioconcentration factor	Ratio of concentration in an organism to that in water	BCF > 2000: Bioaccumulative BCF > 5000: Very Bioaccumulative
Mobility	Soil/sediment adsorption coefficient	Ratio of concentration of substance in organic carbon vs in solution at equilibrium	At a pH of 4-9 Log K_{OC} < 4: Mobile Log K_{OC} < 3: Very Mobile

Where an API was not in the Applicability Domain for generating a QSAR parameter, the entry was left blank.

4. Results

Dataset Overview

More than 700 API sales weights (Table 3) were calculated for each year of sales, representing 753 unique APIs across the four year period. Of the total packages recorded sold from 2016-19, across all APIs, a mean of 6% were removed during processing, leaving the final summarised dataset representative of 94% of all input sales.

Table 3: Table of number of unique products input from starting dataset and number of unique API output, by year.

Year	Starting dataset entries	Unique APIs
2016	5,729	701
2017	5,845	711
2018	5,911	712
2019	5,921	718

Comparison with Felleskatalogen Data

The Norwegian Pharmaceutical Specialities website Felleskatalogen maintains a rolling risk assessment on a yearly basis of pharmaceutical risk, using sales data from the market research firm Farmastat AS. In order to benchmark the completeness and accuracy of our dataset, we compared out calculated sales weights to theirs.

Figure 5 summarises agreement between the two datasets. A mean (blue line) difference extremely close to 0 on the y-axis indicates little average difference between calculations. However, a number of substances below the red line (a) and present in Felleskatalogen records but absent in our records (c) indicate that further work on our part is needed to determine why sales weights for these substances do not appear.

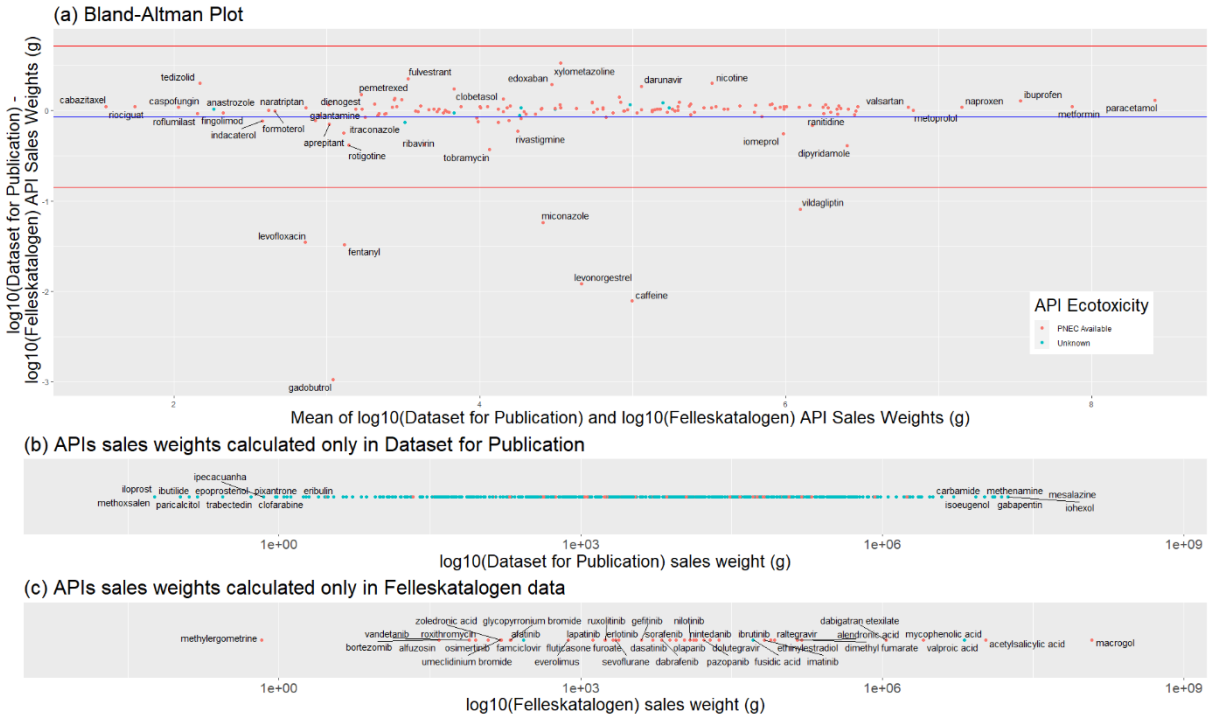

Figure 5: Bland-Altman or Tukey mean-difference plot (a) of difference (y axis) and mean (x axis) of log10-transformed sales weight data from our and Felleskatalogen sources. Blue line marks mean difference, and red 95% Confidence Intervals. Also included are dot plots of APIs only calculated in our data (b) and only by Felleskatalogen (c), graphed across log10 sales weight.

APIs Ordered by Risk Quotient

As a first pass assessment of API risk, risk quotients have been calculated for all APIs where PNECs are available, using the highest PEC over the four years. In total, RQs were calculated for 173 APIs, representing 23% of total APIs sold in the period. Of these RQs, one (levonorgestrel) was extremely high (RQ > 100), five (ciprofloxacin, abiraterone, estradiol, ibuprofen, amoxicillin) were high (RQ > 10), nine were moderate (RQ > 1), 13 low (RQ > 0.1) and 145 negligible (RQ < 0.01). The top 20 highest RQs are summarised below (Table 4).

To provide for broad, intuitive groups, APIs were roughly sorted into classes based on their ATC level 2 codes (for instance, N02 – analgesics). Note that due to the anatomical structure of ATC codes, this has resulted in some splitting of substances such as antiseptics and analgesics that can be applied topically to different parts of the body.

Table 4: APIs, sorted by RQs calculated from highest PECs over the four years and publicly available PNECs. The top 20 of a total of 145 assessed substances are shown. Substances are grouped into classes adapted from ATC level 2 codes.

API	Class	RQ (2 s.f.)
levonorgestrel	sex hormones/genital system	150
ciprofloxacin	ear treatments	56
abiraterone	endocrine therapy	24
estradiol	sex hormones/genital system	17
ibuprofen	painkillers	12
amoxicillin	antibacterials	11
paracetamol	analgesics	9.2
chlorhexidine	antiseptics	7.5
norethisterone	sex hormones/genital system	6.7
naproxen	antiinflammatories	6.6
etonogestrel	sex hormones/genital system	3.8
desogestrel	sex hormones/genital system	3.5
terbinafine	antifungals	3.4
simvastatin	lipid-modifying	3.3
fulvestrant	endocrine therapy	2.7
nicotine	other nervous system	0.56
dronedarone	cardiac therapy	0.49
amiodarone	cardiac therapy	0.41
mometasone	obstructive airway treatments	0.35
propranolol	beta blockers	0.28

Risk and Missing Data by Class

To provide for comparison between classes, RQs were binned into groups based on RQ order of magnitude and ordered by total number of component APIs (Table 5). As can be seen below, even relatively well-documented groups – e.g., sex hormones and diabetes therapies – RQs were unable to be calculated for a sizable proportion of APIs.

Table 5: Risk Quotients per order of magnitude, by API class based on level 2 ATC code. Arranged by total number of APIs per class. NA (%) denotes the proportion of APIs in a category for which no RQ could be calculated due to a lack of readily available toxicity data..

	RQ						
Class	**> 100**	**> 10**	**> 1**	**> 0.1**	**< 0.1**	**NA (%)**	**Total**
antibacterials		1		1	8	81	52
antineoplastics					4	91	47
antiviral				1	22	47	43
psycholeptics (depressants)					8	80	41
eye treatments					6	83	36
psychoanaleptics (stimulants)				2	8	69	32
sex hormones/genital system	1	2	3	1	4	56	25
obstructive airway treatments				1	13	39	23
antiepileptics				1	10	50	22
renin-angiotensin system					9	55	20
analgesics			1		6	63	19
antiinflammatories			1		2	84	19
other nervous system				1	1	89	18
urologicals					3	82	17
cardiac therapy				2	1	81	16
diabetes				1	9	33	15
anaesthetics					0	100	15
antihistamines					4	71	14
immunosuppressants					7	50	14
antihelmintics				1	0	93	14

Persistence, Bioaccumulation, and Mobility QSARs

Information on API bioaccumulation and biodegradation is available from Felleskatalogen only in the form of hazard statements, where substances are classified as being of (translated from the Norwegian) unknown, low, or high bioaccumulative potential, and unknown, biodegradable in the environment, slowly biodegradable, or potentially persistent. However, these statements are only available for 245 APIs, meaning alternative approaches were needed for the remaining substances.

OPERA was used to predict QSAR bioaccumulation factors, biodegradation half-lives and soil/sediment adsorption coefficients for APIs where sufficient data on similar molecules was available. These predicted results were then compared to standard (B, P) and proposed (M) thresholds as discussed in Table 2. Where no prediction was possible, or the API fell outside of the applicability domain of the QSAR model, entries are left blank.

In total, of the 753 unique APIs input to OPERA, 507 applicable bioaccumulation parameters, 54 persistence parameters and 437 mobility parameters were predicted. However, as can be seen below, crossover between substances with valid RQs and valid persistence, mobility and bioaccumulation parameters was limited. Of the 173 APIs with RQs, only 104 bioaccumulation parameters, 6 persistence parameters and 97 mobility parameters could be predicted.

*Table 6: Top 20 highest RQ substances. Where available, empirical Persistence and Bioaccumulation hazard statements (low, moderate, high), in **bold**, were sourced from Felleskatalogen records. Otherwise, OPERA QSARs (nB – not Bioaccumulative/B - Bioaccumulative/vB – very Bioaccumulative, etc.) were appended. A blank space indicates neither test data nor a QSAR within model applicability domain could be found or generated.*

			Bioaccumulation, Persistence and Mobility Levels (Felleskatalogen, OPERA QSARs)		
API Name	**Class**	**RQ (2 s.f.)**	**B**	**P**	**M**
levonorgestrel	sex hormones/genital system	150	**low**	**high**	nM
ciprofloxacin	ear treatments	56	nB		vM
abiraterone	endocrine therapy	24	**high**	**low**	nM
ethinylestradiol	sex hormones/genital system	23	nB		nM
estradiol	sex hormones/genital system	17	**low**	**moderate**	nM
ibuprofen	painkillers	12	**low**	**low**	vM
amoxicillin	antibacterials	11	nB		vM
paracetamol	analgesics	9.2	**low**	**moderate**	vM
chlorhexidine	antiseptics	7.5			
norethisterone	sex hormones/genital system	6.7	nB		nM
naproxen	antiinflammatories	6.6	nB		vM
etonogestrel	sex hormones/genital system	3.8	**low**	**moderate**	nM
desogestrel	sex hormones/genital system	3.5	**low**	**moderate**	nM
simvastatin	lipid-modifying	3.3	**low**	**low**	
fulvestrant	endocrine therapy	2.7	**low**	**low**	
vortioxetine	psychoanaleptics (stimulants)	0.96	nB	P	M
nicotine	other nervous system	0.56	nB		vM
dronedarone	cardiac therapy	0.49	**low**	**high**	
drospirenone	sex hormones/genital system	0.47	**low**	**high**	
amiodarone	cardiac therapy	0.41	**high**	**high**	nM

Persistence, Mobility, Bioaccumulation and Toxicity predictions are further summarised in Figure 6 below. Data availability is notably inconsistent between classes, and predictions of persistence are apparent almost only for nervous system pharmaceuticals. Likewise, of the 104 predicted bioaccumulation factors, only mitotane, an antineoplastic that inhibits steroid synthesis and promotes cytostasis, is predicted to pose a high bioaccumulative risk, due to its extensively halogenated and difficult-to-break down structure. Consequently, we elected to include instead a substance's highest predicted log octanol-water partition coefficient (log K_{ow} or Log D) from QSAR modelling at pH 5.5 and 7.4. Here thresholds are taken from the EMA risk assessment triggers for further assessment: Where $K_{ow} \geq 3$ a bioaccumulation threshold is needed, while at $K_{ow} \geq 4.5$ a full PMT assessment is triggered.

By contrast, mobility was able to be predicted for a sizable proportion of studied APIs, likely due to the simplicity of the relevant parameter. Mobile and Very Mobile substances are well represented across the 14 target systems, likely an unintended side-effect of their need for mobility to function effectively as therapeutics.

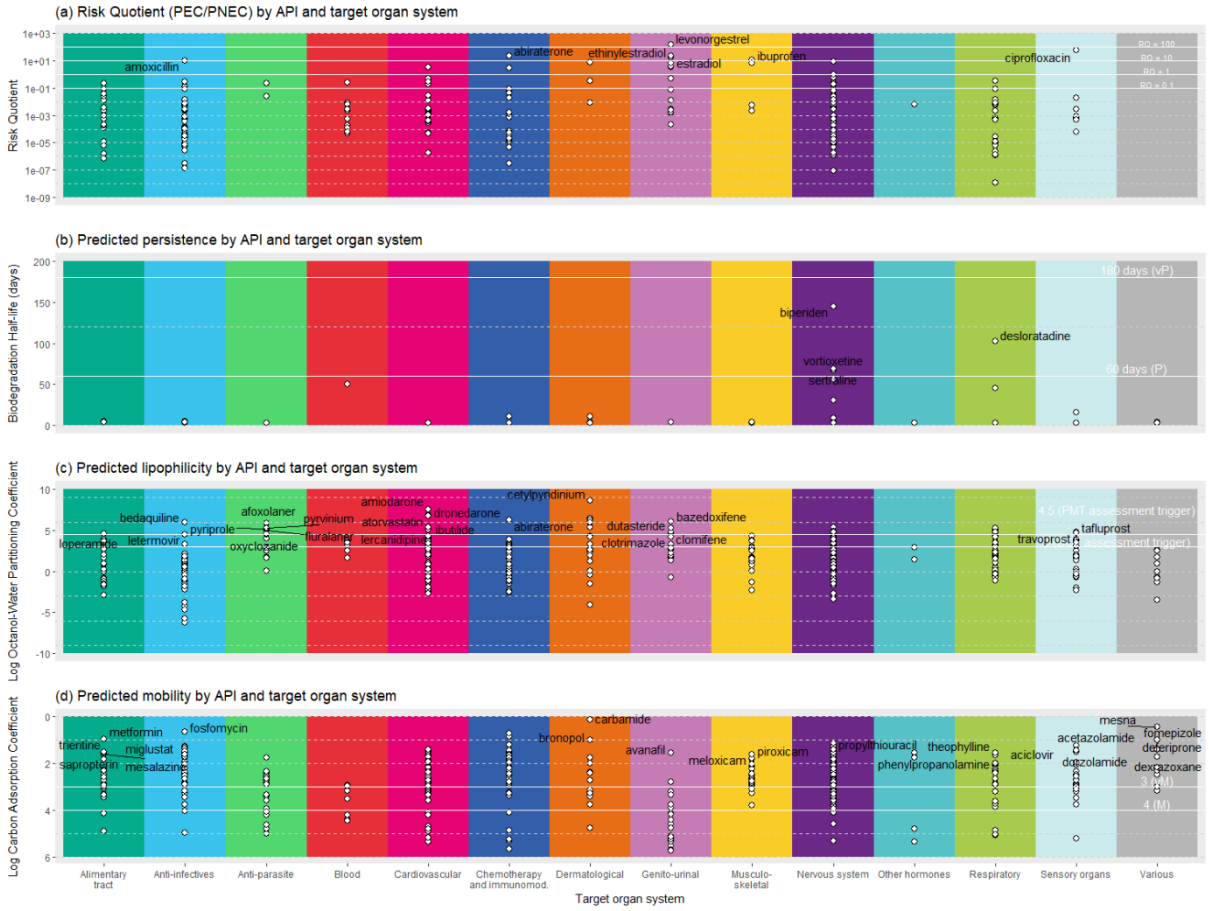

Figure 6: Predicted Risk Quotient, Biodegradation Half Life, low Octanol-Water Coefficient, and log Carbon Adsorption Coefficient (Koc) by API and target organ system. Sold white lines indicate the thresholds of PMBT classification, above which substances are considered to pose a particular hazard. Note that graphs b – c are based entirely on QSARs, and consequently may contradict Table 4. Each white dot represents an API, only those above a threshold are identified by name (where space permits).

5. Conclusions

Through the adaptation of FHI's Wholesale Drug Database it has provided possible to predict exposure year-on-year for over 700 APIs, representing 96% of recorded sales. This represents an extremely useful resource for our studies and others' and will be made publicly accessible as soon as the dataset has been fully checked and verified. However, it only proved possible to predict risk for 173 of these APIs, and persistence, bioaccumulation and mobility data was similarly scarce.

Of these 173 APIs, 15 had an RQ above the conventional cut-off of 1. Across these prioritise high-risk substances, however, bioaccumulation and persistence data were limited, with test data available in 53% of cases and QSARs calculable in a further 17%, while no repository of mobility data is available due to it not being an official parameter.

Across high risk substances, levonorgestrel's toxicity at low levels and high potential for persistence represents cause for concern and should be investigated more thoroughly. Likewise, the moderate or unknown persistence of many other high-risk sex hormones/genital system APIs raises the possibility that over time, concentrations in the environment may build up beyond the ability of the Surface Water Risk Quotient's ability to model, leading to underestimates of environmental risk. Furthermore, the predicted mobility of various common antibacterials and painkillers (ciprofloxacin, ibuprofen, paracetamol, chlorhexidine) also merits study to determine the substances' actual mobility in vitro and in the environment.

In future, we hope to apply the approach explored in this report to further historical sales records, to allow for forecasting of future risk based on sales patterns, demographic, and climate change. Given the unique social role played by pharmaceuticals, a better understanding of their contribution to environmental risk is needed to balance this against their health benefits.

6. References

EMA (2006) Guideline on the Environmental Risk Assessment of Medicinal Products for Human Use. European Medicines Agency

EPA (2018) OPERA: A free and open source QSAR tool for predicting physicochemical properties and environmental fate endpoints. https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=340233&Lab=NCCT. Accessed 20 Dec 2021

FASS (2019) Om Fass - FASS Allmänhet. https://www.fass.se/LIF/menydokument?userType=2&menyrubrikId=40. Accessed 20 Sep 2019

Felleskatalogen (2021) Medisin - Felleskatalogen. https://www.felleskatalogen.no/medisin/. Accessed 27 Nov 2020

NIPH (2019) Wholesaler-based drug statistics. In: Nor. Inst. Public Health. http://www.fhi.no/en/hn/drug/om-den-grossistbaserte-legemiddelforbruksstatistikken/. Accessed 26 Nov 2019

R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria

Rüdel H, Körner W, Letzel T, et al (2020) Persistent, mobile and toxic substances in the environment: a spotlight on current research and regulatory activities. Environ Sci Eur 32:5. https://doi.org/10.1186/s12302-019-0286-x