EnviroInfo 2007 (Warschau)
Environmental Informatics and Systems Research
 ISBN: 978-3-8322-6397-3

Data-Availability of Pharmaceuticals Detected in Water: An Evaluation Study by Order Theory (METEOR)

Kristina Voigt[1], Rainer Brüggemann[2]

Abstract

The topic of pharmaceuticals in the environment is of increasing interest both to the scientific community as well as to the public. Although extensive research is currently under way the availability of data on ecological parameters is largely missing. In our research approach we evaluate the data-availability of 16 widely spread pharmaceuticals in 17 publicly available Internet databases. We apply the order theoretical software METEOR which allows the participation of stakeholders in the evaluation process and the software provides the stepwise introduction of weight. We take a closer look at the chosen four point source pharmaceuticals, two cytostatic agents Cyclophoshamide (CYC), 5-Fluorouracil (FLU) and two contrast media Diatrizoate (DIT), Iopromide (IOP). Even for these pharmaceuticals the data-situation is extremely bad. Only the large databases ChemExper Catalog of Chemical Suppliers, Physical Characteristics (CEX) and RXList The International Drug Index (RXL) comprise all of the chosen pharmaceuticals. It has to be mentioned however that neither of these databases contain data on ecotoxicity and/or degradation and accumulation.

1. Introduction

The limited quantity of unpolluted water available for future use as a resource for food production and drinking water supply is one of the major challenges faced around the world, including Europe. Low levels of pharmaceuticals have been detected in many countries in sewage treatment plant (STP) effluents, surface waters, seawaters, groundwater and some drinking waters. The excretion of incompletely metabolized pharmaceuticals by humans and animals is the primary source of antibiotics in the environment (Brown, 2006). Pharmaceuticals are ubiquitous and persistent in urban receiving waters reflecting input from both point and non-point sources (Ellis, 2006). The consumption of human pharmaceuticals in Germany was 28 878 351 kg in 2001 (Huschek, 2005).

Targeted ecotoxicological studies are lacking almost entirely and such investigations are needed focusing on subtle environmental effects. Such studies will allow a better and comprehensive risk assessment of pharmaceuticals in the future (Fent, 2006).

As an initial step it is therefore indicated to take a close look on the data-availability of pharmaceuticals. Several approaches have been performed by the authors. First of all an intensive literature study on the occurrence of pharmaceuticals in the environment was set-up and evaluated with environmetrical and chemometrical methods (Voigt, 2005). Furthermore the availability of data on pharmaceuticals in environmental and chemical databases was scrutinized by mathematical methods (Voigt, 2006).

2. Data-matrix of 17 Databases (Objects) and 16 Pharmaceuticals (Attributes)

In our current research approach we evaluate the data-availability of pharmaceuticals in Internet databases. In this approach we evaluate 17 Internet databases (objects) by 16 pharmaceuticals (attributes). The database is evaluated according to the availability of information on the pharmaceutical x: If information

[1] GSF-National Research Center for Environment and Health, Institute of Biomathematics and Biometry, Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany, e-mail: kvoigt@gsf.de
[2] Leibniz Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, 12587 Berlin, Germany, e-mail: brg_home@web.de

is available: code = 1, if not: code = 0). The chosen Internet databases are: **CIV** (Chemicals Information System for Consumer-relevant Substances), **CEX** (ChemExper Catalog of Chemical Suppliers, Physical Characteristics), **CHF** (Chemfinder), **ECO** (ECOTOX), **ENV** (Envirofacts), **ESI** (ESIS – European Chemical Substances Information System), **GES** (GESTIS – Dangerous Substances Database), **GSB** (GSBL Public), **HSB** (Hazardous Substances Database), **IAR** (IARC), **ICS** (International Chemical Safety Cards), **INT** (INTOX), **OEK** (Oekopro), **OIH** (OECD Integrated HPV Database), **RXL** (RXList The International Drug Index), **SIR** (SIRI Material Safety Data Sheets), **SRC** (SRC PhysProp Database).

The selected pharmaceuticals with their acronyms as well as their belonging to an indicated pharmaceutical class are listed in Table 1.

Table 1:
Drug Groups with their corresponding Pharmaceuticals

Class of Pharmaceutical	**Acc.**	**Pharmaceuticals including accronyms**	**Number**
Analgesics	ANAL	Diclofenac (DIC), Ibuprofen (IBU), Phenazone (PHE)	3
Antibiotics	ANTI	Roxithomycin (ROX), Sulfamethoxazole (SUL)	2
Antiepileptics	AEPI	Carbamazepine (CAR)	1
Beta blocker	BETA	Metoprolol (MET)	1
Contrast media	CONT	Diatrizoate (DIT), Iopromide (IOP)	2
Cytostatic agents	CYTO	Cyclophoshamide (CYC), 5-Fluorouracil (FLU)	2
Lipid regulators	LIPI	Bezafibrate (BEZ), Clofibric acid (CLO), Fenofibrate (FEN)	3
Psychiatric drugs	PHYS	Diazepam (DAP),	1
Steroids	STER	Ethinyl Estradiol (EES)	1

3. METEOR – Method of Evaluation by Order Theory

Partial order in general and partially ordered sets (posets) have a long standing in mathematics with Birkhoff saying "The world around us abounds with examples of partly ordered sets" (Birkhoff, 1967). Partial order theory provides many concepts to derive linear orders without any additional introduction of external knowledge (Lerche, 2003). As no subjective weighting is involved the linear order obtained from a partial order is called a "canonical order" (Brüggemann, 2005). In contrast to derive canonical linear orders, METEOR (Method of Evaluation by Order Theory) attempts to resolve the incomparabilities among objects by inclusion of external knowledge. It intends to obtain a clear decision (one best solution), maintaining transparency and allowing participation. METEOR is based on the well-known and often used concept of a hierarchy of criteria in multi-criteria decision aids. Consequently METEOR allows a step-by-step aggregation of indicators by forming weighted sums about subsets of indicators. One may first aggregate similar indicators then proceed to higher levels of the hierarchy of criteria. The possibility of a step-by-step aggregation of indicators provides the freedom to thoroughly analyse the effects of indicator weights and compensation. Furthermore, preferences (indicator weights) which are most sensitive to the evaluation result can easily be identified. For further reading on the role of METEOR in the context of the Hasse Diagram Technique (HDT) see the text-book on Partial Order in Environmental Sciences edited by Brüggemann and Carlsen (2006).

3.1 Aggregation Strategies under METEOR

From a logical point of view one should start with indicators expressing similar kinds of effects. If for example chemicals are to be evaluated, one may consider exposure indicators on the one side as candidates for an aggregation and effect indicators as candidates for another aggregation, obtaining two super-indicators "Exposure" and "Effects". This point of view is comfortable for stakeholders as it allows them first to consider general aspects and then –perhaps - to go into details. From the point of evaluation we might call this procedure a bottom - up procedure. However similar indicators are often well correlated (indeed one may even define similarity by the correlation behaviour) and their aggregation has little effect on the poset and is hence of little use for decision making. More efficient but less convincing from an evaluation point of view is to aggregate those indicators which have a high degree of conflicting potential. Those indicators are often anti - correlated. Hence their aggregation will rather efficiently reduce the incomparabilities. This kind of procedure one may call a top-down procedure: First reduce the most conflicting indicator subsets and then analyze the results by applying partial order.

Even if we have decided to follow a top-down procedure it is not clear how the aggregation functions should look like: If some indicators q(i) are linear combined, taking weights as scalars, then any resulting "**superattribute**", φ(k) is calculated as:

$$\varphi(k) := \sum_{i=1}^{n(k)} g(i) * q(i)$$

together with the normalization:

$1 = \sum_{i=1}^{n(k)} g(i)$ n(k) being the number of indicators, actually combined in order to calculate φ(k),

then any superattribute has the "freedom" of n(k)-1 of freely varying the scalars g(i) ($\in [0,1]^{n(k)-1}$). We call $[0,1]^{n(k)-1}$ the g-space of the kth superattribute. Therefore we associate to any superattribute a space of weights with the dimension n(k) -1 and any aggregation step in METEOR is accompanied by the product of all g-spaces (k=1,..m), which we call the G-space. In general n(k) may vary and may depend on the intuition of the researcher, applying METEOR. Here, however, we restrict ourselves on aggregation schemes with freedom 1, i.e. we analyze in the subsequent parts of the paper for any superattribute a linear space. If we combine for example four indicators pairwise to two superattributes, the two linear g-spaces are combined, forming a two-dimensional space [0,1]*[0,1]. As we will see later in the text, the restriction to freedoms = 1 considerably simplifies the procedure and we call a procedure, based on a purely pairwise combination of attributes the "orthogonal-METEOR" (abbr.: o-METEOR).

3.2 Crucial Weights

Imagine that four indicators are pairwise aggregated as follows:

$$\varphi(1) = g(1)*q(1)+(1-g(1))*q(2) \text{ and} \quad (1a)$$

$$\varphi(2) = g(2)*q(3)+(1-g(2))*q(4) \quad (1b)$$

Assume that the database x is incomparable with database y due to:
q(1, x) > q(1,y) and q(2,x) < q(2,y).
For this case we write: $x \parallel_{(q1,q2)} y$.
If $x \parallel_{(q1,q2)} y$ then the result of aggregation (1a) depends on the one weight g(1), whether or not φ(x) > φ(y).

Obviously the equation

$$\varphi(1,x) = \varphi(1,y) \qquad (2)$$

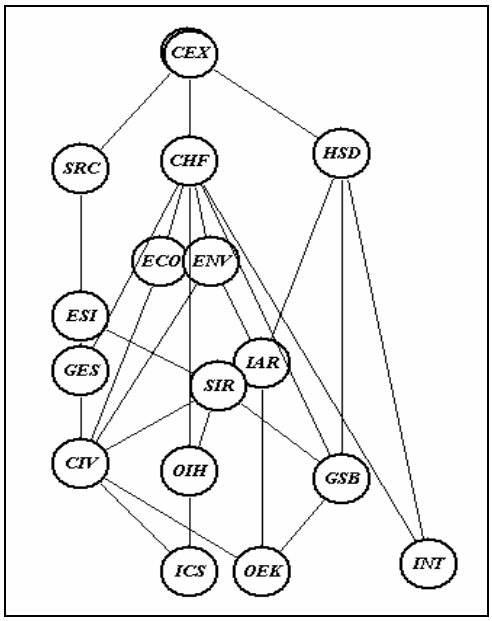
determines that g(1) value where the character of order relation between x and y changes.
For further details of the concept of crucial weights we refer to a paper by Brüggemann et al. (submitted to J. Env. Modelling & Software).
Stability fields are subspaces of the G-space where a change of weights does not change the relative positions of any two incomparable objects.

4. Evaluation of the Complete Data-Matrix

First we calculate the Hasse Diagram of the complete (17x16) data-matrix.

In Figure 1 we can see that there is only one maximal object, the equivalence class {CEX;RXL} which means that these two databases are better with respect to their availability on data of the 16 selected pharmaceuticals than all the other 15 databases. The minimal objects are ICS, OEK and INT. They are the worst databases in this evaluation approach.

This initial evaluation step is called the "let first the data speak" step. In the following procedure we include subjective preferences.



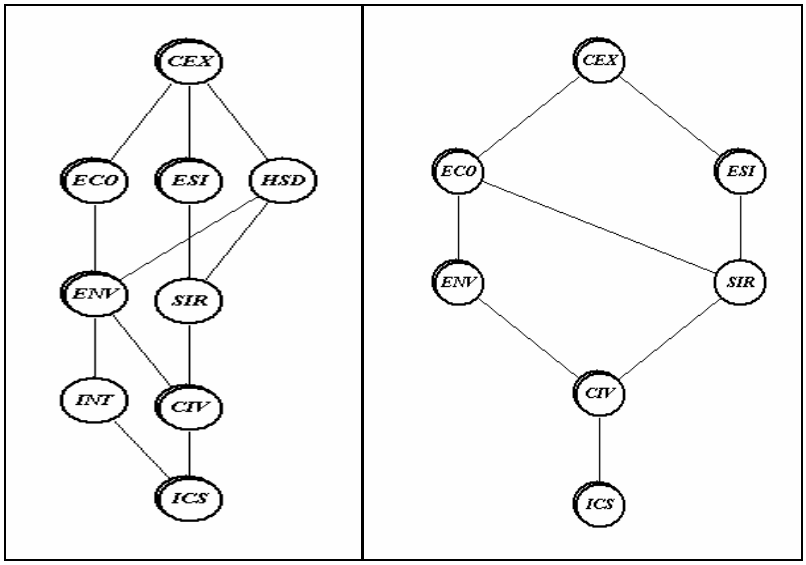
equivalent objects: {CEX;RXL}

Figure 1: Hasse Diagram 17 Databases x 16 Pharmaceuticals (complete data-matrix)

5. Point-source Pharmaceuticals and Their Aggregation

5.1 Evaluation of Point-Source Pharmaceuticals

In our approach we are following the bottom - up strategy. The 4 point source pharmaceuticals (see Table 1) are the two cytostatic agents Cyclophoshamide (CYC), 5-Fluorouracil (FLU) and the two contrast media Diatrizoate (DIT), Iopromide (IOP).
The Hasse Diagram for the 17x4 data-matrix is shown in Figure 2, lhs.



Equivalent Objects: {CIV;GES;OEK;GSB} {CEX;RXL} {ECO;CHF} {ENV;IAR} {ESI;SRC} {ICS;OIH} (17x4), Equivalent Objects: {CIV;GES;OEK;GSB;INT} {CEX;RXL} {ECO;HSD;CHF} {ENV;IAR} {ESI;SRC} {ICS;OIH} (17x2)

Figure 2: 17 Databases x 4 Point-Source Pharmaceuticals (lhs), 17 x 2 Aggregated Point-Source Groups CYTO/CONT Hasse Diagram (rhs)

The 4 point-source pharmaceuticals (see Table 1) are now with aggregated to the following two groups, each comprising two pharmaceuticals.

CYC + FLU = CYTO, DIT + IOP = CONT

The corresponding Hasse Diagram of these two super-attributes is shown in Figure 2 r.h.s. It is demonstrated that the number of incomparabilities is considerably reduced by the aggregation step from 4 point-source pharmaceuticals to two groups of point-source drugs. The maximal objects and minimal objects remain more or less the same.

5.2 Aggregation Procedure of Point-source Pharmaceuticals, Stability Fields and Crucial Weights

Following the aggregation strategy described in section 3.1, the Hasse diagrams of $\varphi(1)$ and $\varphi(2)$ are set up (Figure 3).

φ(1) = g(1)*CYC + (1-g(1))*FLU

φ(2) = g(2)*DIT + (1-g(2))*IOP

φ(1) stands for Cytostatic agents that is to say CYC and FLU, φ(2) stands for Contrast media that is to say DIT and IOP

In the next step we identify the incomparable pairs of objects in the CYC/FLU diagram as well as in the DIT/IOP diagram (see Figure 3).

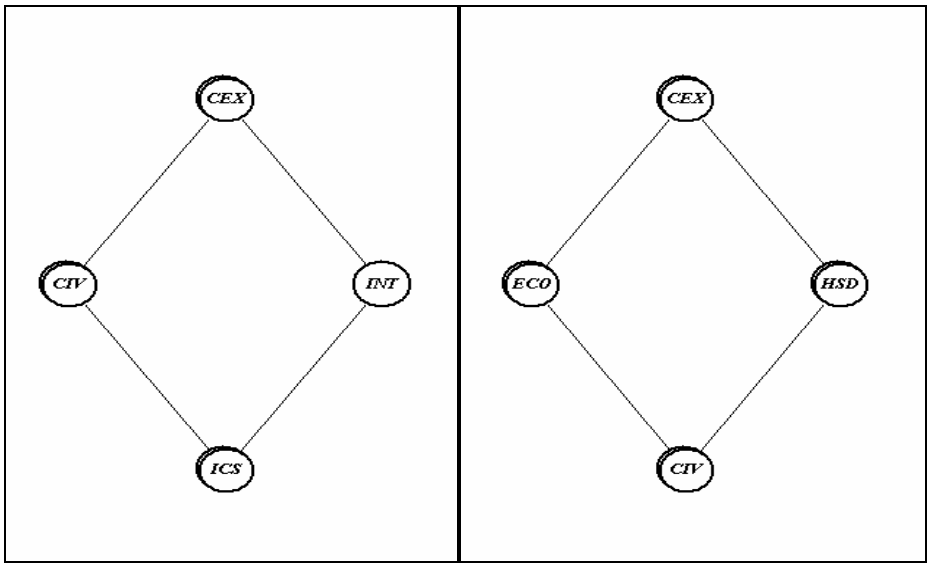


Figure 3: Hasse Diagrams for 17x2: φ(1) CYC, FLU (lhs), φ(2) DIT, IOP (rhs)

Lhs: Equivalent Objects: {CIV;ESI;GES;OEK;GSB;SIR;SRC} {CEX;ECO;ENV;HSD;IAR;CHF;RXL} {ICS;OIH}
Rhs: {CIV;ENV;GES;ICS;OEK;OIH;GSB;IAR;INT} {CEX;ESI;SRC;RXL} {ECO;CHF} {HSD;SIR}

$U_{CYC,FLU}$ = {(CIV;ESI;GES;OEK;GSB;SIR;SRC), (INT) }

$U_{DIT,IOP}$ = {(ECO;CHF)}, {(HSD;SIR)}

This means that we receive **4 stability fields**. The explanation on the calculation of the stability fields and crucial weights is found in Brüggemann et al. (2007 submitted). To assign the resulting order in any of the 4 stability fields, one has to compare the φ(1) with the φ(2) diagram (see Figure 3). In both diagrams (quotient sets) there is only one incomparability U. By aggregation either CIV > INT or CIV < INT from the φ(1) diagram, as both databases are equivalent in the φ(2) diagram, the two orientations do not lead to an incomparability in the resulting order. The same applies to the incomparable databases ECO and HSD in the φ(2) diagram whereas there are equivalent objects in the φ(1) diagram. However, as for

example SIR < ENV in the lhs, whereas SIR > ENV in the rhs of Figure 3 not necessarily linear orders will appear in all four stability fields.

The crucial weights are calculated by the sub-routine Pyhasse7.py which is written in Python by the second author.

The crucial weights for **g(1)** is **0,5** and for **g(2)** is also **0,5.**

We therefore calculate the Hasse diagrams for the stability fields are given in Figure 4.

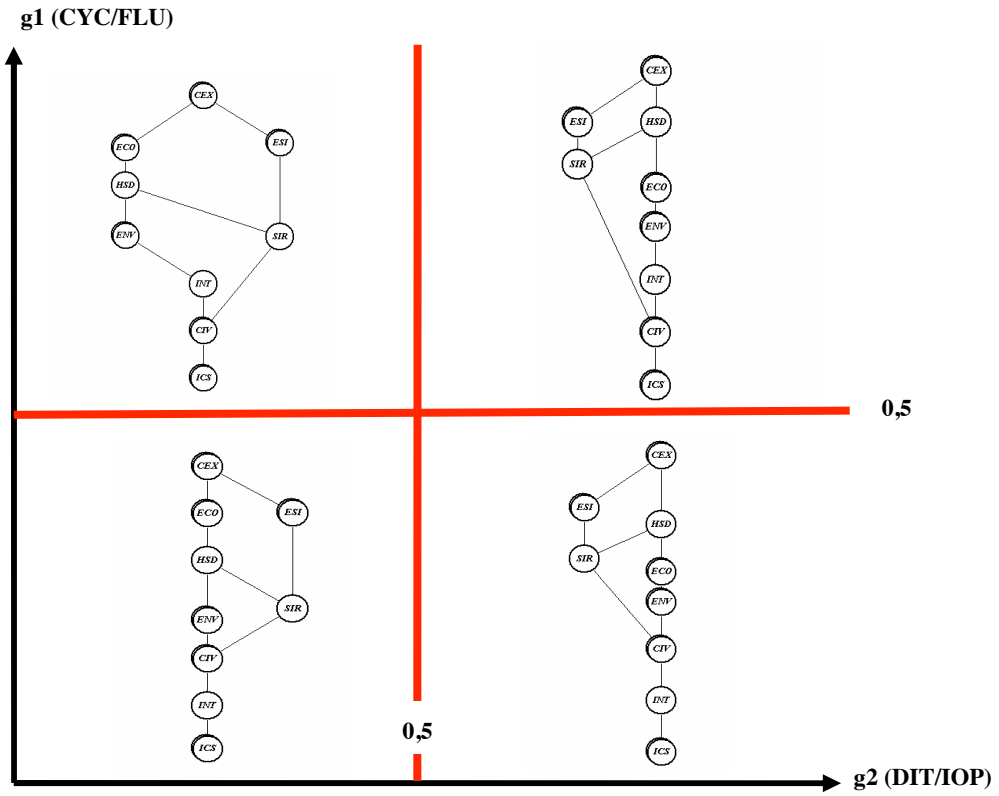


Figure 4: Stability fields and their different Hasse Diagrams

It can be demonstrated that all diagrams in the four stability fields are different from each other. The maximal and minimal objects however are the same in all four Hasse Diagrams. The equivalent objects are not listed for reasons of visibility.

6. Results and Discussion

The applied order theoretical software METEOR allows participation of stakeholders and provides the stepwise introduction of weights. The expectation is that often just some few steps will be helpful for the

decision (here: which database is useful for data availability for pharmaceuticals). For example incomparable databases as shown in Figure 2 are now related to each other in a systematic, i.e. order preserving way. The advantages associated with discrete approaches such as the HDT, which provide high transparency throughout the whole evaluation process is combined with the flexible use of weights, which model the subjective preferences.

From the data-availability view of the approach we have to state a very sad situation concerning data on pharmaceuticals. All conducted approaches show that the data-situation on the chosen test-set 17 publicly available Internet databases concerning their data-availability on 16 well-known and highly produced pharmaceuticals is far from being satisfactory. Only the large databases ChemExper Catalog of Chemical Suppliers, Physical Characteristics (CEX) and RXList The International Drug Index (RXL) comprise all of the chosen pharmaceuticals. It has to be mentioned however that neither of these databases contain data on ecotoxicity and/or degradation and accumulation. The issue of pharmaceuticals in the environment and the unavailability of data necessitate more research and of course closer communication between science and medical healthcare and politicians in the future.

Bibliography

Birkhoff, G. (1967): Lattice Theory, 3rd Edition, American Mathematics Society, Colloquium Publications, Providence, R.I., XXV, p. 418.

Brown, K.D., Kulis, J., Thomson, B., Chapman, T.H., Mawhinney, D.B. (2006): Occurence of antibiotics in hospital, residential, and dairy effluent, municipal wastewater, and the Rio Grande in New Mexico, Science of the Total Environment, 366, 2-3, pp. 772-783.

Brüggemann, R., Simon, U., Mey, S. (2005): Estimation of Average Ranks by Extended Local Partial Order Models, MATCH Communications in Mathematical and Computer Chemistry, 5(3): 489-518.

Brüggemann, R., Carlsen, L. (Eds.) (2006): Partial Orders in Environmental Sciences and Chemistry, Springer-Verlag, Berlin, pp. 1-406.

Brüggemann, R., Voigt, K., Restrepo, G., Simon, U. (2007 submitted): Concept of Stability Fields and Hot Spots in Ranking of Chemicals, submitted to Journal of Environmental Modelling and Software.

Ellis, J.B. (2006): Pharmaceutical and personal care products (PPCPs) in urban receiving waters, Environmental Pollution, 144, 1, pp. 184-189.

Fent, K., Weston, A.A., Caminada, D. (2006): Rewiew Ecotoxicology of human pharmaceuticals, Aquatic Toxicology, 76: 122-159.

Huschek, G., Verbrauchsmengen von Humanarzneimitteln, in: Umweltbundesamt (Ed) (2005): Arzneimittel in der Umwelt – Zu Risiken und Nebenwirkungen fragen Sie das Umweltbundesamt, Texte 29/05, pp. 79-87, Umweltbundesamt, Berlin.

Lerche, D., Sørensen, P.B., Brüggemann, R (2003): Improved Estimation of the Ranking Probabilities in Partial Orders Using Random Linear Extensions by Approximation of the Mutual Probability, J. Chem. Inf. Comp. Sci. 53: 1471-1480.

Voigt, K., Brüggemann, R. (2005): Water contamination with pharmaceuticals: Data availability and evaluation approach with Hasse Diagram Technique and METEOR, MATCH Communications in Mathematical and Computer Chemistry, 54(3): 671-689.

Voigt, K., Brüggemann, R., Pudenz, S. (2006): A multi-criteria evaluation of environmental databases using the Hasse Diagram Technique (ProRank) software, Environmental Modelling & Software, 21: 1587-1597.