NOEC or EC_{10} for Training Data Sets in QSAR Model Building – Does the Difference Matter?

R. Arno Wess, Thomas Schmidt, Matthias Loetscher
IES Ltd, Benkenstrasse 260, 4108 Witterswil, Switzerland

SETAC North America 41st Annual Meeting,
Held virtually, November 15th to 19th

Introduction

As early as 1998 the OECD (Series of Testing and Assessment (STA) No. 10) concluded that the NOEC (No-Observed-Effect-Concentration), as the main summary parameter of aquatic ecotoxicity tests, is inappropriate and should be phased out. It was recommended that the OECD should move towards a regression-based estimation procedure. German authorities have evaluated the subject comprehensively (Kalberlah & Hassauer 2003) and OECD repeated this statement in 2006 (STA 54). Interestingly the first European Medicines Agency (EMA) guidance for the Environmental Risk Assessment (ERA) of human pharmaceuticals, in force since 2006, mentions only the use of NOECs, but the guideline update from 2018 states explicitly that an existing 10%-Effective Concentration (EC_{10}) has to be preferred.

Nonetheless, both, EC_{10} and NOEC data, have been used in training data sets for the construction of Quantitative Structure-Activity Relationship (QSAR) models, e.g. in the US EPA EPI Suite™ ECOSAR™ tools (https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface). While they do often not differ very much, there are cases showing one of even two orders of magnitude between the two toxicological threshold values. Jung et al. (2013) and Loetscher (2019) encountered cases with significantly different NOEC and EC_{10} results from the same studies (Table 1). Such differences of effect threshold levels might have an impact in QSAR models built on the basis of a limited number of training data and are critical for the appropriateness of the model.

Only limited data on ecotoxicological effect data are available for active pharmaceutical ingredients marketed before the submission ecotoxicological data in the ERA was mandatory. As pharmaceuticals in the environment are of concern due to their intended bioactivity, it seems desirable to assess data gaps by employing QSAR models. One example for such approach in the EU is the iPiE project (https://www.imi.europa.eu/projects-results/project-factsheets/ipie). The question arises whether the use of NOEC or EC_{10} values from the same study might impact QSAR training sets.

Table 1. Algal toxicity EC_{10} and NOEC values determined according to OECD testing guideline 201

Substance	Species	Exposure [h]	EC_{10} [mg/L]	NOEC [mg/L]	Difference [Factor; Orders of magnitude]	References
Enrofloxacin	*Microcystis aeruginosa*	72	0.0070	0.0090	0.8; 0	Various, secondary source: Wess et al. 2020
Linezolid	*Raphidocelis subcapitata*	72	0.1800	0.1600	1.1; 0	
	Synechococcus leopoliensis	96	0.0900	0.1000	0.9; 0	
Trimethoprim	*Navicula pelliculosa*	96	1.32	1.20	1.1; 0	
Tylosin	*Phaeodactylum tricornutum*	96	0.1924	0.2565	0.8; 0	
	Synechococcus leopoliensis	96	0.0101	0.0082	1.2; 0	
Clarithromycin	*Raphidocelis subcapitata*	72	0.0070	0.0040	1.8; 0	
Clindamycin	*Raphidocelis subcapitata*	72	0.00056	0.00029	1.9; 0	
	Synechococcus leopoliensis	72	0.00239	0.00029	**8.2; 1**	
Metronidazole	*Anabaena flos-aquae*	72	19.76	0.32	**61.8; 2**	Loetscher 2019
Flucloxacillin	*Raphidocelis subcapitata*	72	>95.2	0.9500	**>100.2; 2**	
	Anabaena flos-aquae	72	0.11	0.03	**3.7; 0**	
Gemigliptin tartrate	*Raphidocelis subcapitata*	72	>100 (ca. 109)	0.32	**340.6; 2**	Jung et al. 2013

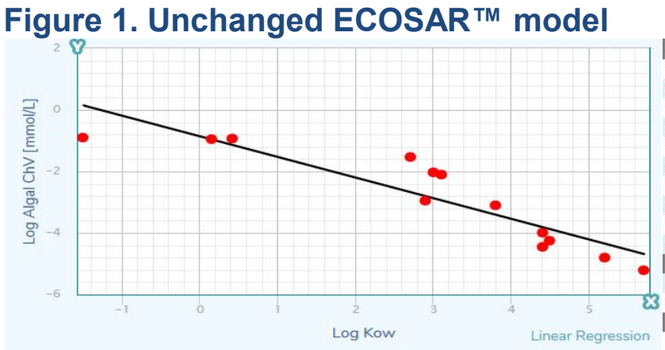
Materials and methods

In order to get an idea on the impact of using NOEC or EC_{10} in the training set for a QSAR model, some hypothetical calculations were made using the methodology applied in the widely used ECOSAR™ models (Mayo-Beana et al. 2012, Clements et al. 1996): Basically the substances were assigned to classes, i.e. chemical categories on the basis of functional groups (chemical structure characteristics), in which the decadic logarithm of the molar toxicity concentration becomes a linear function of the decadic logarithm of the lipophilicity (K_{ow} or D_{ow} at environmentally relevant pH). Linear regression is used to calculate the function from the logarithmized experimental data points. The model data for the Aliphatic Amines class were employed because the model has an acceptable training set, consisting of data from 13 substances over a lipophilicity (Log K_{ow}) range from -1.5 to 5.7 and Clindamycin (difference factor 8.2, Table 1) as well as Gemigliptin (difference factor 340.6, Table 1) belong to it. To study impact scenarios, one data point of the original training set was manipulated (at several positions in the lipophilicity scale) by multiplying the Algal Chronic Value (ChV) [mmol/L] with the factors 8.2 or 340.6. Then the linear regression was calculated with the manipulated data set and the changes of graph and in the goodness measures of the fit were compared. Also an hypothetical substance prediction was made, presuming a MW of 425 g/mol and Log K_{ow}'s of -1.5, 2.1 and 5.7.

Results

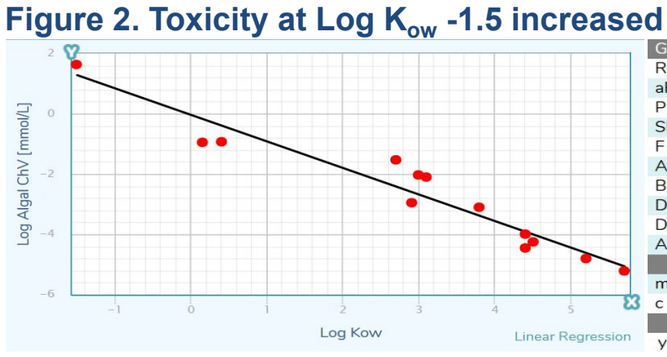
Representative regression curves are shown in Figure 1 to 4. The impact of the model manipulation as alteration of the prediction for a hypothetical substance in shown in Table 2.

Figure 1. Unchanged ECOSAR™ model



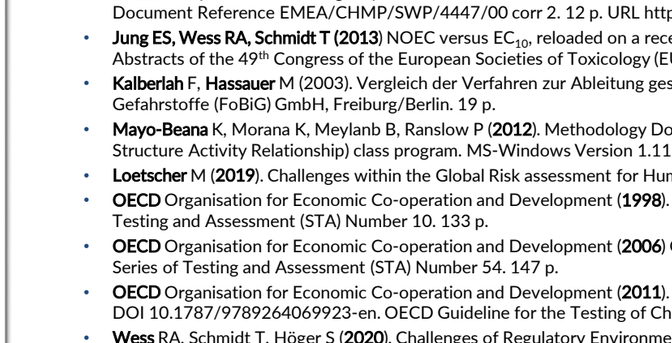
Goodness Measures	
R^2	0.8286
aR^2	0.813
P	0.00001558
SE	0.6761
F	53.18
AIC	28.54
BIC	29.67
DWS	1.088
DoF	11
AICc	29.74
Coefficients	
m	−0.670182 ± 0.0919
c	−0.8625958 ± 0.332
Equation	
y = −0.670182x − 0.8625958	

Figure 2. Toxicity at Log K_{ow} -1.5 increased by Factor 340.6



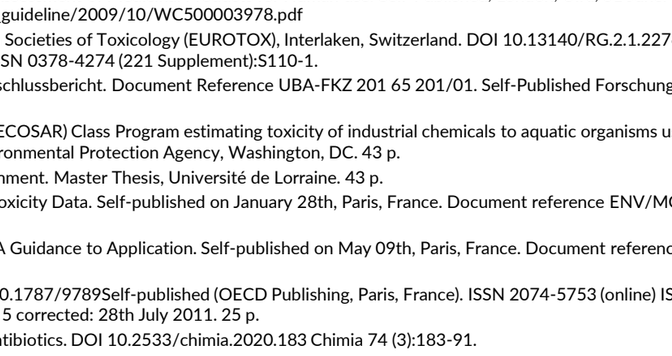
Goodness Measures	
R^2	0.9273
aR^2	0.9206
P	$1.334*10^{-7}$
SE	0.5466
F	140.2
AIC	23.01
BIC	24.14
DWS	1.972
DoF	11
AICc	24.21
Coefficients	
m	−0.8796609 ± 0.07429
c	−0.04341092 ± 0.2684
Equation	
y = −0.8796609x − 0.04341092	

Figure 3. Toxicity at Log K_{ow} 3.1 increased by Factor 8.2



Goodness Measures	
R^2	0.7675
aR^2	0.7463
P	0.00008605
SE	0.8159
F	36.31
AIC	33.43
BIC	34.56
DWS	1.103
DoF	11
AICc	34.63
Coefficients	
m	−0.6681904 ± 0.1109
c	−0.798534 ± 0.4006
Equation	
y = −0.6681904x − 0.798534	

Figure 4. Toxicity at Log K_{ow} -1.5 increased by Factor 8.2



Goodness Measures	
R^2	0.8934
aR^2	0.8838
P	0.000001106
SE	0.5711
F	92.23
AIC	24.16
BIC	25.29
DWS	1.3
DoF	11
AICc	25.36
Coefficients	
m	−0.7455282 ± 0.07763
c	−0.5679483 ± 0.2805
Equation	
y = −0.7455282x − 0.5679483	

Table 2. Algal toxicity ChV [mg/L] prediction for a hypothetical substance according to the models represented in Figures 1 to 4

Log K_{ow}	Figure 1	Figure 2	Figure 3	Figure 4
-1.5	590.291	8025 (Δ +1260%)	679.423 (Δ +15%)	1509 (Δ +1569%)
2.1	2.283	5.466 (Δ +139%)	2.671 (Δ +17%)	3.125 (Δ +37%)
5.7	0.009	0.004 (Δ -56%)	0.01 (Δ +11%)	0.006 (Δ -33%)

Discussion

According to the regression equations in Figures 1-4, the algal ChVs decrease with increasing lipophilicity of the respective molecule. If only one of the 13 trainings set toxicity data (at the minimum Log K_{ow} -1.5) is arbitrarily changed by a factor for 340.6, the goodness of the fit as R^2 is significantly improved (Figure 2) if compared with the actual data based model (Figure 1), but the model predictions are different by up to 1260 % (Table 2: ChV at Log K_{ow} -1.5). Accordingly a high goodness of fit can be fallacious as it may pretend a reliable training set. Outliners, even with a difference factor of two orders of magnitude, can be hardly discernible even if they are at an extreme value (Figure 2 and 4), while they may look suspect in the mid-range of a value table (Figure 3). Therefore, (at least) data points at the extreme x-values in a training set for QSAR model calculation according to the ECOSAR™ methodology should be considered carefully and preferably regression based, if possible.

References:

- **Clements** RG, Nabholz JV, Zeeman M (**1996**). Estimating toxicity of industrial chemicals to aquatic organisms using structure-activity relationships. Self-published Environmental Effects Branch, Health and Environmental Review Division, Office of Pollution Prevention and Toxics, U.S. Environmental Protection Agency, Washington, DC. 404 p.
- **EMA** European Medicines Agency Committee for Medicinal Products for Human Use (CHMP) (**2018**). Guideline on the Environmental Risk Assessment of Medicinal Products for Human use. Draft. Self-Published, London, U.K., 15 November. Document Reference EMEA/CHMP/SWP/4447/00 Rev. 1, 48 p. URL https://www.ema.europa.eu/documents/scientific-guideline/draft-guideline-environmental-risk-assessment-medicinal-products-human-use-revision-1_en.pdf
- **EMEA** European Medicines Agency Committee for Medicinal Products for Human Use (CHMP) (**2006**). Guideline on the Environmental Risk Assessment of Medicinal Products for Human use. Self-Published, London, U.K., 01 June. Document Reference EMEA/CHMP/SWP/4447/00 corr 2. 12 p. URL http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/10/WC500003978.pdf
- **Jung ES, Wess RA, Schmidt T (2013)** NOEC versus EC_{10}, reloaded on a recent show case. Poster P08-38 at the 49th Congress of the European Societies of Toxicology (EUROTOX), Interlaken, Switzerland. DOI 10.13140/RG.2.1.2276.0082. Abstracts of the 49th Congress of the European Societies of Toxicology (EUROTOX). DOI 10.1016/j.toxlet.2013.05.178 Toxicology Letters ISSN 0378-4274 (221 Supplement):S110-1.
- **Kalberlah** F, **Hassauer** M (2003). Vergleich der Verfahren zur Ableitung gesundheitsbezogener Wirkungsschwellen (Benchmark - NOAEL). Abschlussbericht. Document Reference UBA-FKZ 201 65 201/01. Self-Published Forschungs- und Beratungsinstitut Gefahrstoffe (FoBiG) GmbH, Freiburg/Berlin. 19 p.
- **Mayo-Beana** K, Morana K, Meylanb B, Ranslow P (**2012**). Methodology Document for the ECOlogical Structure-Activity Relationship Model (ECOSAR) Class Program estimating toxicity of industrial chemicals to aquatic organisms using the ECOSAR (Ecological Structure Activity Relationship) class program. MS-Windows Version 1.11. Self-published Office of Pollution Prevention and Toxics U.S. Environmental Protection Agency, Washington, DC. 43 p.
- **Loetscher** M (**2019**). Challenges within the Global Risk assessment for Human pharmaceuticals with focus on antibiotics in the aquatic environment. Master Thesis, Université de Lorraine. 43 p.
- **OECD** Organisation for Economic Co-operation and Development (**1998**). Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data. Self-published on January 28th, Paris, France. Document reference ENV/MC/CHEM(98)18. Series of Testing and Assessment (STA) Number 10. 133 p.
- **OECD** Organisation for Economic Co-operation and Development (**2006**) Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application. Self-published on May 09th, Paris, France. Document reference ENV/JM/MONO(2006)18. Series of Testing and Assessment (STA) Number 54. 147 p.
- **OECD** Organisation for Economic Co-operation and Development (**2011**). Freshwater Alga and Cyanobacteria, Growth Inhibition Test. DOI 10.1787/9789Self-published (OECD Publishing, Paris, France). ISSN 2074-5753 (online) ISBN 9789264069923 (PDF) DOI 10.1787/9789264069923-en. OECD Guideline for the Testing of Chemicals, Section 2 Test No. 201, adopted 23rd March 2006, Annex 5 corrected: 28th July 2011. 25 p.
- **Wess** RA, Schmidt T, Höger S (**2020**). Challenges of Regulatory Environmental Risk Assessment for Human Pharmaceuticals with Focus on Antibiotics. DOI 10.2533/chimia.2020.183 Chimia 74 (3):183-91.