Effect-based trigger values for *in vitro* and *in vivo* bioassays performed on surface water extracts supporting the environmental quality standards (EQS) of the European Water Framework Directive

Authors:

Beate I. Escher[a,b,c,d]*, Selim Aït-Aïssa[e], Peter A. Behnisch[f], Werner Brack[a,g], François Brion[e], Abraham Brouwer[f], Sebastian Buchinger[h], Sarah E. Crawford[g], David Du Pasquier[i], Timo Hamers[j], Karina Hettwer[k], Klára Hilscherová[l], Henner Hollert[g], Robert Kase[m], Cornelia Kienle[m], Andrew J. Tindall[i], Jochen Tuerk[n], Ron van der Oost[o], Etienne Vermeirssen[m] and Peta A. Neale[c,d]

[a]UFZ – Helmholtz Centre for Environmental Research, 04318 Leipzig, Germany

[b]Eberhard Karls University Tübingen, Environmental Toxicology, Centre for Applied Geosciences, 72074 Tübingen, Germany

[c]Australian Rivers Institute, Griffith School of Environment, Griffith University, Southport QLD 4222, Australia

[d]The University of Queensland, Queensland Alliance for Environmental Health Sciences (QAEHS), Brisbane QLD 4108, Australia

[e]Institut National de l'Environnement Industriel et des Risques INERIS, Unité d'Ecotoxicologie, 60550 Verneuil-en-Halatte, France

[f]BDS BioDetection Systems B.V., Amsterdam, The Netherlands,

[g]Department of Ecosystem Analysis, Institute for Environmental Research, RWTH Aachen University, 52074 Aachen, Germany

[h]Bundesanstalt für Gewässerkunde, Am Mainzer Tor 1, 56068 Koblenz, Germany

[i]Laboratoire Watchfrog, 1 Rue Pierre Fontaine 91 000 Evry, France

[j]Vrije Universiteit Amsterdam, Dept. Environment & Health, De Boelelaan 1108, 1081 HZ Amsterdam, The Netherlands

[k]QuoData GmbH, Prellerstr. 14, 01309 Dresden, Germany

[l]Masaryk University, Research Centre for Toxic Compounds in the Environment (RECETOX), Kamenice 753/5, 62500 Brno, Czech Republic

[m]Swiss Centre for Applied Ecotoxicology Eawag-EPFL, Überlandstrasse 133, 8600 Dübendorf, Switzerland

[n]Institut für Energie- und Umwelttechnik e.V. (IUTA, Institute of Energy and Environmental Technology), Bliersheimer Str. 58-60, D-47229 Duisburg, Germany

[o]Waternet Institute for the Urban Water Cycle, Department of Technology, Research and Engineering, Amsterdam, The Netherlands

*corresponding author: beate.escher@ufz.de; Ph: +49 341 235 1244

Abstract

Effect-based methods including cell-based bioassays, reporter gene assays and whole-organism assays have been applied for decades in water quality monitoring and testing of enriched solid-phase extracts. There is no common EU-wide agreement on what level of bioassay response in water extracts is acceptable. At present, bioassay results are only benchmarked against each other but not against a consented measure of chemical water quality. The EU environmental quality standards (EQS) differentiate between acceptable and unacceptable surface water concentrations for individual chemicals but cannot capture the thousands of chemicals in water and their biological action as mixtures. We developed a method that reads across from existing EQS and includes additional mixture considerations with the goal that the derived effect-based trigger values (EBT) indicate acceptable risk for complex mixtures as they occur in surface water. Advantages and limitations of various approaches to read across from EQS are discussed and distilled to an algorithm that translates EQS into their corresponding bioanalytical equivalent

concentrations (BEQ). The proposed EBT derivation method was applied to 48 *in vitro* bioassays with 32 of them having sufficient information to yield preliminary EBTs. To assess the practicability and robustness of the proposed approach, we compared the tentative EBTs with observed environmental effects. The proposed method only gives guidance on how to derive EBTs but does not propose final EBTs for implementation. The EBTs for some bioassays such as those for estrogenicity are already mature and could be implemented into regulation in the near future, while for others it will still take a few iterations until we can be confident of the power of the proposed EBTs to differentiate good from poor water quality with respect to chemical contamination.

Abbreviations

AA-EQS	average annual environmental quality standard
ACR	acute-to-chronic ratio
AhR	arylhydrocarbon receptor
AR	androgen receptor
AWTP	advanced wastewater treatment plant
BEQ	bioanalytical equivalent concentration
DWTP	drinking water treatment plant
EBM	effect-based methods
EBT	effect-based trigger value
EC	effect concentration
EEQ	estradiol equivalent
EF	extrapolation factor
EQS	environmental quality standard
ER	estrogen receptor
FET	fish embryo toxicity
GR	glucocorticoid receptor

GV	guideline value
HC_5	hazardous concentration for 5% of water organisms
LID	lowest ineffective dilution
MAC-EQS	maximum allowable concentration environmental quality standard
MEC	measured environmental concentration
MIE	molecular initiating event
MOA	mode of action
NOAEL	no-observed adverse effect levels
NOEC	no-observed effect concentrations
PAH	polycyclic aromatic hydrocarbons
PCB	polychlorinated biphenyls
POD	point of departure
PPARγ	peroxisome proliferator activated receptor
PR	progesterone receptor
PXR	pregnane X receptor
REF	relative enrichment factor
REP_i	relative effect potency of chemical i
RI	risk index
RQ	risk quotient
SPE	solid-phase extraction
TCDD	2,3,4,7-tetrachlorodibenzo-*p*- dioxin
TH	thyroid hormone
TR	thyroid receptor
TU	toxic unit
WET	whole effluent toxicity
WFD	Water Framework Directive
WWTP	wastewater treatment plant

1 Introduction

1.1 Towards the development of effect-based trigger values

Effect-based methods (EBM), mainly *in vitro* cell-based (often reporter-gene) assays and small-scale *in vivo* whole-organism bioassays (such as algae, daphnids and fish embryos) have been applied for decades to monitor water quality and water treatment processes (Escher and Leusch, 2012; Hamers et al., 2013; Leusch and Snyder, 2015; Prasse et al., 2015; van der Burg et al., 2013; Wernersson et al., 2015). However, currently targeted chemical analysis is still most commonly used for chemical water quality monitoring. This holds true also for the European Water Framework Directive (WFD) (European Parliament and European Council, 2000) although recently the use of EBMs has been recommended for a review of this regulatory framework (Brack et al., 2017). Awareness is increasing that targeted chemical monitoring cannot account for the presence of unknown chemicals or transformation products. Further, chemicals are generally present in the aquatic environment in complex mixtures and, while individual chemicals may be present below guideline values (GV), the mixture effects of many chemicals at low concentrations can be significant (e.g. the “something from nothing” effect (Silva et al., 2002; Walter et al., 2002)). Bioassays provide evidence of the joint biological effect of all active chemicals in a sample (Maletz et al., 2013; Välitalo et al., 2016). Further, they are hazard-scaled, so at similar concentrations more potent chemicals will have a greater contribution to the mixture effect than low-potency chemicals.

EBMs yield quantitative effect measures, e.g. effect concentrations (EC). EC values can be translated into bioanalytical equivalent concentrations (BEQ) to make the effect measure comparable between bioassays targeting the same mode of action (MOA) (Escher and Leusch, 2012). The BEQ of a water sample is the concentration of a reference compound that would elicit the same effect as all compounds in the water sample. By using BEQs for sample characterization before and after treatment, it is possible to quantify treatment efficacy in a wastewater treatment plant (WWTP), an advanced water treatment plant (AWTP) or a drinking water treatment plant (DWTP) (Escher

et al., 2009; Leusch et al., 2005; Neale et al., 2012; Van der Linden et al., 2008). However, since every bioassay has different characteristics, it is not possible to quantitatively compare between bioassays targeting different MOAs or apical endpoints. In addition, combinations of extraction techniques allowing high enrichment factors for organic chemicals (Schulze et al., 2017) and increasingly more sensitive cell lines have allowed effects to be detected even in drinking water and highly treated recycled water (Escher et al., 2014). The fact that an EC can be derived does not always mean that an adverse effect for ecosystem and human health is expected. Many *in vitro* assays, e.g., those indicative of nuclear receptors that trigger enhanced metabolic activity and transcription factors that mediate adaptive stress responses, indicate the activation of defense mechanisms at low doses (Simmons et al., 2009) and thus the presence of contaminants in the sample. Therefore, the limit of detection in an *in vitro* bioassay has no bearing on the adversity of effect related to a given assay and in many cases there is no direct relationship between BEQ and the degree of adversity of *in vivo* effects. Rather, *in vitro* bioassays are used as analytical tools to quantify mixtures of chemicals. Hence EBMs are also often termed bioanalytical tools.

The combination of solid-phase extraction (SPE) and bioassays has led to such low limits of detection that contaminant concentrations in high-quality water are not below the limit of detection any more. Hence, just because an effect is detectable does not mean that this is necessarily unacceptable. For surveillance and monitoring applications, it thus becomes imperative to define thresholds, so called effect-based trigger values (EBT) that differentiate between acceptable and poor water quality with respect to the organic micropollutants, with further testing recommended if a water sample exceeds an EBT. Similar bioassays have been applied across different types of water from drinking water to sewage and even to sediments and biota. Acceptable effect levels will differ depending on the sample type. However, ideally similar methods should be applied for the derivation of EBTs for different matrices and protection targets.

EBTs for surface water need to be in line with consented environmental and human health related quality standards for individual compounds.

Consequently, they need to be protective for ecosystem health and for human health due to the use of surface water for drinking water abstraction or water reuse. The goal of this study is to develop a generic method for the derivation of EBTs that reads across from chemical GVs and can be applied to any set of chemical GVs and to any bioassay. The methods will be applied here specifically in the context of the assessment of water quality for European surface waters as one case study. No final numerical EBTs are proposed, but the focus lies on the evaluation of various derivation methods with the goal to propose a coherent and widely applicable method for future applications. The effect data used to evaluate the various explored methods might still be incomplete or not completely adequate for the purpose. Therefore, the resulting numerical EBTs are preliminary and will need to be refined in a second step by targeted measurement of more effect data for environmentally relevant and regulated chemicals. A commonality in all approaches that we use here is to base EBT values on BEQs. Hence, the EBTs will be defined as an effect-based trigger BEQ (EBT-BEQ).

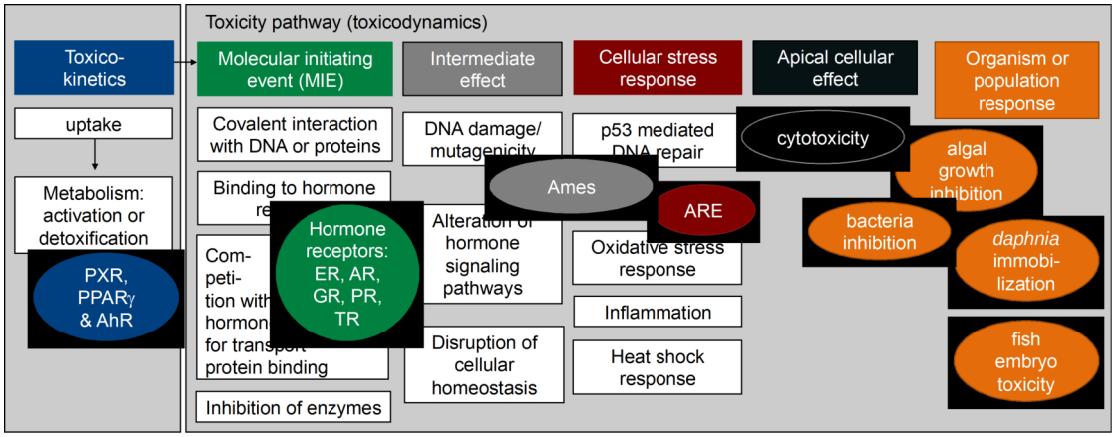
1.2 Environmental Quality Standards

The European WFD aims to integrate biological and chemical information in order to obtain an overall insight into the quality of individual water bodies (European Parliament and European Council, 2000). According to the WFD, the chemical status of a water body is determined by analyzing and assessing the concentrations of 45 (groups of) priority substances. A good chemical status is reached when the concentrations of all priority substances are below the annual average and maximal allowable concentration. Environmental Quality Standards (AA-EQS and MAC-EQS) were defined to protect the environment and human health (EuropeanCommission, 2011). We use the AA-EQS values for substances under the WFD approach as a case study here. Similar GVs for water quality were derived in other jurisdictions and the method introduced here can be applied to those as well.

1.3 Panels of cellular and whole-organism assays for water quality monitoring

A large number of bioassays indicative of different endpoints have been developed over recent decades. Their strength is that they account for mixtures of chemicals acting together - all chemicals in the case of apical endpoints and groups of chemicals with the same MOA for reporter gene assays. By applying a panel of cellular and small-scale whole-organism assays it is possible to obtain a more holistic profile of the effects of all chemicals present in a water sample without identifying the causative compounds individually. To capture effects commonly detected in environmental waters and to protect against missing unexpected effects, it is important to assemble a bioassay test battery that covers different types of effects. Test batteries should ideally include bioassays indicative of different stages of the cellular toxicity pathway, including induction of xenobiotic metabolism, receptor mediated effects, reactive MOA, adaptive stress responses and cell viability, as well as apical effects in whole organisms (Figure 1, adapted from Escher et al. (2014; Neale et al. (2017b).

Test batteries covering these endpoints have recently been applied to surface water, wastewater and recycled water (Jia et al., 2015; Leusch et al., 2014; Neale et al., 2017a). Further, bioassays indicative of reactive toxicity and induction of adaptive stress responses (Neale et al., 2012; Hebert et al., 2018) and hormone-receptor-mediated effects (Brand et al., 2013) have also been applied specifically to drinking water.



238 *Figure 1 Summary of bioassays included in the EBT-derivation (figure adapted from (Neale et al.,*
239 *2017b)).*

240

241 1.4 State of the art

242 There are principally two approaches to derive EBTs: If the point of departure
243 (POD) is an adverse effect, then one needs to translate concentrations of
244 potent reference chemicals that are considered safe *in vivo* to concentrations
245 detectable *in vitro.* Such an approach does not account for mixtures but
246 mixture considerations can be included and bioassays are per definition
247 quantifying mixture effects if they are applied to samples that contain more
248 than one component. An example of this approach is the drinking water EBTs
249 developed by Brand et al. (2013) for hormonal activity. This approach was
250 restricted to health-based EBTs and requires information such as acceptable
251 daily intake values and estimated bioavailability data.

252 The second approach is to base the derivation of EBTs on existing EQS as
253 the POD. The procedure to derive a GV or EQS follows similar principles in
254 many jurisdictions, with no-observed effect concentrations (NOEC) for
255 environmental species or no-adverse effect levels (NOAEL) in test animals as
256 the POD and a series of extrapolation steps employed uncertainty factors or
257 species sensitivity distribution-based estimates to derive a safe concentration,
258 which is then used as the GV.

259 The simplest approach is to translate an EQS directly to its corresponding
260 BEQ and use this value as the EBT. This approach was proposed for
261 estrogenic chemicals (Kunz et al., 2015) in surface water. A similar weighted
262 method using the four most potent estrogens was suggested for wastewater
263 (Jarosova et al., 2014). These EBT options are limited to assays where one or
264 a few compounds with defined EQS dominate effects.

265 Environmental EBTs for apical endpoints were further proposed by van der
266 Oost et al. (2017) in the SIMONI (Smart Integrated Monitoring) strategy. The
267 SIMONI-EBTs for apical endpoints were derived from acute ECs assuming an
268 acute-to-chronic ratio (ACR) of 10 and an additional safety factor of two for
269 extraction recovery. The resulting toxic unit (TU) of 0.05 (TU = 1/EC), which
270 corresponds to a relative enrichment factor (REF) of 20, was then used as the

EBT. The REF is an indicator of concentration and takes into account sample enrichment and dilution in the assay. In addition, ECs from aquatic *in vivo* data were integrated to estimate the safe BEQ (lowest observed chronic effect equivalents in the database), the HC_5-BEQ (hazardous concentration for 5% of water organisms, determined with species sensitivity distribution on all chronic EC_{50}-BEQs) and a background BEQ (bioassay response at eight sites with good ecological status). These three BEQ values were used to derive SIMONI-EBTs for a panel of *in vitro* bioassays.

Mixture considerations were included in the derivation of EBTs for drinking water and recycled water (Escher et al., 2015) but were limited to cell-based assays.

Here we build on all of these earlier approaches to establish a common derivation method that reads across from existing EQS and explicitly addresses mixtures. The method can be applied to any bioassay from reporter gene cell-based assays to whole-organism assays provided sufficient data for the effects of regulated chemicals are available.

2 Materials and Methods

2.1 Point of departure

The POD for the derivation of EBTs is taken from existing GVs. Any coherent set of such GVs will permit the derivation of EBTs but it remains a regulatory decision where and when to implement the EBT. This paper used the EU and Swiss AA-EQS (hereafter just termed EQS) as case studies but the approach is versatile enough to be used for any set of GVs, e.g., drinking water GV (WHO, 2011), GVs for recycled water (NRMMC & EPHC & NHMRC, 2008), or for discharge of wastewater (Federal Ministry for the Environment, Nature Conservation and Nuclear Safety, 2004).

2.2 One algorithm for all bioassays?

We can classify bioassays into two categories: Category 1 (defined mixtures) includes those bioassays that target one highly specific molecular initiating

301 event, such as the binding to a hormone receptor, and for which the majority
302 of active chemicals are known and category 2 (undefined mixtures) are those
303 that are responsive to many if not all chemicals. The category 1 bioassays
304 include typically receptor-mediated effects, e.g., activation of the estrogen
305 (ER), androgen (AR), glucocorticoid (GR), progesterone (PR) or thyroid (TR)
306 receptors (Figure 1) or specific effects on an organism level such as inhibition
307 of photosynthesis. Here iceberg modeling (König et al., 2017; Neale et al.,
308 2017a; Tang and Escher, 2014) and effect-directed analysis (Brack et al.,
309 2016; Hashmi et al., 2018; Muschket et al., 2018) have demonstrated that
310 typically only a few highly bioactive molecules (often natural hormones or
311 synthetic drugs) can explain a high proportion of the mixture effects observed
312 by the cocktail of chemicals in a water sample.

313 In contrast, there are bioassays that register more integrative effects, e.g., the
314 cellular stress responses such as the oxidative stress response, apical cellular
315 effects and the *in vivo* organism responses (Figure 1). For these category 2
316 bioassays, even if dozens or hundreds of chemicals are quantified and the
317 effects of these single chemicals are known, the computed mixture effect of
318 these known chemicals in the concentrations they occur typically explains only
319 a small fraction of the effect, often less than 1% (Escher et al., 2013; Neale et
320 al., 2017a; Tang et al., 2013).

321 We cannot say *a priori* which bioassays fall into which of the two categories
322 and there is also a grey area between the two categories, e.g., an apical
323 endpoint such as algal growth can be very specific for herbicides inhibiting
324 photosynthesis. However, clearly these two categories need to be treated
325 somewhat differently in the EBT derivation because category 1 bioassays are
326 mainly triggered by high-potency chemicals, while in category 2 bioassays
327 many chemicals have low potency but together they may cause effects of
328 concern. As will be shown, there needs to be a specific provision to consider
329 mixture effects for category 2 bioassays. Hence, in principle, there is one
330 algorithm for all bioassays but category 2 bioassays require an additional
331 mixture factor as will be introduced below.

2.3 Translating an EQS into a BEQ associated to this EQS

To translate any chemical concentration, for example an EQS concentration, into an associated BEQ the relative effect potencies (REP_i), i.e. the potencies of the compounds of interest in relation to the potency of the reference compound in a certain bioassay, are needed.

REP_i can be calculated by equation 1 from the EC of a reference compound divided by the EC of the compound of interest i. The effect endpoints must be compatible, such that the effect y in EC_y of the reference compound and all tested chemicals should be matching, e.g., EC_{10}, PC_{10} (Kunz et al., 2017), EC_{50}, or $EC_{IR1.5}$ (Escher et al., 2014). The slopes of the sigmoidal concentration-effect curves must be similar or linear concentration-effect curves must be used to obtain an effect-level independent REP_i.

$$REP_{i,in\ vitro} = \frac{EC(\text{reference compound})}{EC(\text{compound i})} \quad (1)$$

BEQs can be directly measured in a bioassay (BEQ_{bio}) or calculated from chemical measurements by multiplying the measured concentration of an active compound i in the bioassay with its REP_i ($BEQ_{i,chem}$). BEQ_{bio} accounts for effects of all chemicals present in the sample, known or unknown, while the sum of $BEQ_{i,chem}$ only considers the mixture effects of known chemicals.

One can assign a $BEQ_{i,chem}$ to each chemical i at its EQS_i concentration via equation 2.

$$BEQ_{i,chem} = REP_{i,in\text{-}vitro} \cdot EQS_i \quad (2)$$

If $REP_{in\text{-}vivo}$ = $REP_{in\text{-}vitro}$ and the EQS_i did not consider further hazard indicators such as persistence, bioaccumulation and secondary poisoning, then all $BEQ_{i,chem}$ for one given bioassay should theoretically be equal. In practice, $BEQ_{i,chem}$ vary because the EQS is derived to protect the entire aquatic ecosystem, not for one bioassay. Most *in vitro* bioassays are indicative of one specific step in the toxicity pathway. The species applied in *in vivo* whole-

organism assays will not necessarily match with the species that is driving the EQS derivation as the most sensitive species. And even if there were a perfect match between the *in vivo* endpoint driving the EQS derivation and the *in vitro* bioassay, then differences in toxicokinetics would possibly lead to further differences, e.g., if a chemical were only active after metabolic activation and the *in vitro* assay has no capacity for metabolism.

2.4 Accounting for mixture effects

Bioassays intrinsically account for mixture effects because all chemicals acting with the same MOA will result in a concentration additive effect in a given MOA specific bioassay. For whole-organism assays multiple types of interaction in mixtures are possible, including independent action, concentration addition, synergy and antagonism. In general, concentration addition is a robust reference model (Backhaus and Faust, 2012; Tang et al., 2013; Warne and Hawker, 1995) for water samples that contain very large numbers of chemicals and no dominant individual chemicals. In the case of a reporter gene assay targeting a specific nuclear receptor or transcription factor, a mixture of agonists can be assumed to follow concentration addition.

In contrast, EQS are derived for single chemicals. While legally each chemical could be present just below its EQS_i and considered safe on its own, the larger the number of chemicals present, the more probable that the mixture effects could exceed some effect threshold ("something from nothing effect" (Silva et al., 2002)).

A measure of how close the measured environmental concentration (MEC) is to the EQS is the risk quotient RQ, which is defined as the ratio between MEC and the corresponding safe concentration represented by the EQS. To calculate the cumulative risk of a chemical mixture, a risk index (RI) is used. The RI is the sum of the risk quotients RQ_i of n chemicals i. The RI should only be calculated for chemicals with the same MOA because the condition of its derivation is that concentration addition applies. Similar to the RQ, RI = 1 would typically be assigned as the threshold between acceptable RI and risk.

$$RI = \sum_{i=1}^{n} RQ_i = \sum_{i=1}^{n} \frac{MEC_i}{EQS_i} \quad (3)$$

Translated to bioassays the RI would be conceptually equivalent to the ratio between the measured BEQ_{bio} and the EBT-BEQ but this RI also includes the effect from unknown chemicals.

$$RI = \frac{BEQ_{bio}}{EBT\text{-}BEQ} \quad (4)$$

There is a caveat to the approach of adding up the RQs. The RI is dependent on the number of chemicals n, so RI will automatically increase as the number of chemicals n increases. In contrast, if we calculated the RQ_i individually and check if each chemical had a $RQ_i < 1$, then we might underestimate the risk of chemicals acting together in mixtures. Each EQS is derived and specific for one chemical. If an EQS was truly protective for the ecosystem, then it must also be protective for another chemical acting according to the same MOA but just with a different value scaled according to potency. Therefore, a balance must be struck to account for mixture effects without being overprotective, which is discussed in more detail in section 2.6.

2.5 *In vitro* bioassays as integrators of modes of action

Chemicals that act according to the same MOA elicit a concentration-additive mixture effect. An EQS for a single compound should also be protective for the mixture that is equipotent to the single chemical, as long as the GVs have been derived from toxicity data based on the MOA monitored in the *in vitro* assay. For example, an EBT-BEQ for 17β-estradiol based on its carcinogenicity will not be useful when comparing it to an estradiol equivalent

(EEQ) value measured from an estrogenicity bioassay, but an EBT-BEQ for 17β-estradiol based on an estrogenicity assay will be applicable.

Thus, the first step in deriving an EBT-BEQ is to match chemicals with EQS to the appropriate bioassays. This might be a very simple and straightforward endeavor for very well-known MIEs such as binding to the ER, but in many cases the relevant MOAs are not known for chemicals that have an EQS and many chemicals exhibit multiple MOAs with different inherent potencies. Therefore, we have not made any prior assignments based on MOA but have included all available bioassay data.

One way to assign MOAs to chemicals is to test if they are responsive in an *in vitro* assay, e.g. a reporter gene assay that is specific for a given MIE. However, all chemicals will cause cytotoxicity and apical effects (at different concentrations) and close to cytotoxic concentrations, reporter genes are often activated in a non-specific manner, which was termed cytotoxicity burst (Judson et al., 2016). Therefore, not every chemical that has an EC value in a given bioassay exhibits the associated MOA in the whole organism and should be included in the EBT derivation, but rather only those with certain proximity (to be defined) of its EC in the bioassay to the EQS should be included. Low-potency chemicals skew the EBT distributions and therefore need to be excluded by the filtering step described in the next section.

2.6 Evaluated options for EBT derivation

We evaluated various options for the EBT derivation and recommend two final approaches, one for category 1 bioassays and one for category 2 bioassays.

Ideally, if a bioassay were protective for the entire ecosystem, then a chemical's EQS could be directly translated to an associated BEQ and all EBTs derived for different chemicals would be the same. In practice, this is of course not the case. Therefore, a first step in any derivation of EBTs will be to translate all available EQS_i to BEQ_i.

If all chemicals were allowed to be present at their EQS (which would legally be possible) and concentration addition applies as the mixture model, then the EBT would just be the sum of all BEQ_i (Option A). The resulting EBT would then be dependent on the number of chemicals included (n). This would not

be a problem if all REP_i values of EQS compounds were available for all bioassays but this is not the case. While intuitively this option A appears unreasonable, mathematically it would be the correct way of approaching the read across. In theory, all regulated chemicals could be present just below their EQS_i and the water quality would still be acceptable.

Option A: $$EBT = \sum_{i=1}^{n} BEQ_i$$

Option A: (5)

To avoid the dependence on n, the EBT could be defined as average BEQ of all chemicals at their EQS, which is equivalent to the 50th percentile of a normal distribution of BEQ_i (Option B).

Option B: $$EBT = \frac{\sum_{i=1}^{n} BEQ_i}{n}$$

Option B: (6)

As biological data is often log-normally distributed, an alternative option to derive the EBT is to take the mean of the log BEQ_i as a basis for EBT (Option C).

Option C: $$EBT = 10^{\left(\frac{\sum_{i=1}^{n} \log BEQ_i}{n}\right)}$$

Option C: (7)

The method using mean values (Option B) might not be sufficiently protective. An alternative option would be to derive the 5th percentile of a normal distribution (Escher et al., 2015) or to apply an extrapolation factor (EF) (Option D). If the distribution was normal, then the ratio between the 50th and the 5th percentile would be approximately an EF of 10.

Option D: $$EBT = \frac{\sum_{i=1}^{n} BEQ_i}{n} / EF$$

Option D: (8)

Jarosova et al. (2014) proposed for estrogenic compounds in WWTP to choose the minimum of the BEQ_i of the potent estrogens as EBT. This approach (Option E) will be included in the evaluation but it will only be useful if only high potency compounds are included in the derivation, i.e., possibly for a subclass of category 1 bioassays.

Option E: $EBT = \min(BEQ_i)$ (9)

However, it must be noted that the low-potency compounds have associated low BEQ_i, if at the same time the EQS is low, which would in turn mean that low potency compounds would unduly influence the EBT derivation. As the ratio of EC_i/EQS_i increases, the BEQ_i are decreasing, therefore an additional filtering step might be useful to exclude compounds with too low bioanalytical potency to avoid skewing the EBT towards low values. For this reason, only substances with REP_i values > 0.001 were considered for the SIMONI-EBT derivation (van der Oost et al., 2017). However, what counts is not the REP_i alone but the product of the REP_i and EQS_i. Previously, we had proposed to remove chemicals and bioassay combinations with an EC_i / GV_i ratio > 10 in the derivation of EBTs for Australian drinking water (Escher et al., 2015). However, if the bioassay battery was expanded to less specific endpoints and given that EQS_i are often orders of magnitude lower than drinking water guideline values, we propose to use a threshold EC_i / EQS_i > 1000 for filtering. We only explored Option B with the additional filtering step (Option F) but other combinations are included in the Appendix A.

Option F:

$$EBT = \frac{\sum_{i=1}^{n} BEQ_i}{n} \text{ only for data with } \frac{EC_i}{EQS_i} < 1000$$

500 Option F: only for data with (10)

501 If the fraction of each chemical in the mixture were known, instead of using
502 the mean, one could use the exposure-corrected mean by applying the
503 fraction f_i prior to summing up the contribution to the EBT (Option G). This is
504 not realistic for most chemical mixtures because their mixture composition will
505 vary between different sites and scenarios but for estrogens we often observe
506 a typical pattern and the most potent estrogen (EE2) is always present at very
507 low fractions (Kase et al., 2017). Option G was proposed earlier by Jarosova
508 et al. (2014) for safe levels in WWTP effluents.

$$EBT = \sum_{i=1}^{n} f_i \cdot BEQ_i$$

509 Option G:

510

511 (11)

512 In category 2 bioassays a large number of chemicals trigger only a small
513 fraction of effect, which would mean that the BEQ_i are very low and the EBT
514 would be overprotective. Therefore, we must add a mixture factor for these
515 bioassays (Option H). This affects those bioassays where after applying the
516 selection criterion EC_i / EQS_i < 1000 there remain no or less than three
517 chemicals. If an observed effect can be caused by many unknown
518 compounds, then the mixture factor is high. The choice of the mixture factor in
519 equation 12 is difficult but should be dependent on the EC_i / EQS_i ratio not on
520 the REP_i. It should also take into account the fraction of effect typically
521 explained by known chemicals in iceberg modeling. Our initial proposal would
522 be to set the mixture factor to 100 for the receptor-mediated endpoints and
523 1000 for the adaptive stress responses. For the apical endpoints it will be a
524 case-by-case decision that is discussed in more detail below.

525 Option H:

$$EBT = \text{mixture factor} \cdot \frac{\sum_{i=1}^{n} BEQ_i}{n}$$

526

(12)

For some relevant biological endpoints there exists a multitude of different bioassays. This is the case for activation of the ER. Therefore, alternatively to deriving each EBT for each assay for a given biological endpoint, one could also define an average generic EBT per endpoint and adjust this value with a bioassay-specific sensitivity factor. This option is not pursued any further here but could well be applicable to the estrogenicity assays, eleven of which were included here.

As EBT-BEQs cannot be directly compared between the bioassays indicative of different endpoints due to different reference compounds, we also derived an effect threshold with equation 13 using the EBT-BEQ and the EC value of the assay reference compound. The effect threshold is the REF of the water samples above which we can expect effects (10% for EC_{10}, IR1.5 for $EC_{IR1.5}$, 80% for PC_{80} etc.) in the bioassay.

$$\text{effect threshold} = \frac{EC_{\text{reference compound}}}{EBT-BEQ} \quad (13)$$

2.7 Data collection of Environmental Quality Standards

Freshwater AA-EQS from the EU Directive 2013/39/EU and proposed freshwater AA-EQS from the Centre for Applied Ecotoxicology, Switzerland, were collected for 100 chemicals using the ETOX database (ETOX, 2017) (Table A1). AA-EQS values from the EU Directive were prioritized above those from the Centre for Applied Ecotoxicology, Switzerland, if AA-EQS were available in both for a particular chemical. There were no EU Directive or Centre for Applied Ecotoxicology AA-EQS available for two of the chemicals, which appeared to be of high environmental relevance, triclosan and triphenyl phosphate, so AA-EQS proposed by the Umweltbundesamt, Germany, were used. The results from the ETOX database for the studied chemicals can be found in the Supplementary Information, Table A2.

To ensure that the selected POD was protective for both, environmental and human health, we compared a common list of 21 EQS-values and WHO drinking water GVs (WHO, 2011) and in all cases the EQS was more protective (Table A1).

2.8 Data collection of effect data from bioassays

Effect data for the studied chemicals were collected from the peer-reviewed literature or the US EPA ToxCast database (U.S. EPA, 2015) and the BDS database (P. Behnisch, unpublished) and are listed in the Appendix A, Tables A4-A51. Effect data were reported as EC_{10}, PC_{10} (Kunz et al., 2017), EC_{50} or PC_{50} for assays where a maximum effect was reached (e.g. receptor mediated effects, apical effects). Effect data for assays run in antagonist mode were reported as the effect concentration causing a suppression ratio of 0.2 ($EC_{SR0.2}$) (Escher et al., 2014) or 20% suppression of the agonist effect, PC_{80}. For assays where no maximum effect could be reached, such as adaptive stress response assays, the effect data were reported as $EC_{IR1.5}$ (Escher et al., 2014). The same effect endpoint and measure (10%, 50%, etc.) was used for each bioassay. For example, if the available effect data for a particular assay were reported in some cases as EC_{10} and in others as EC_{50}, then the EC_{50} values were converted to EC_{10} by assuming a slope of 1. If multiple EC values were available for the same chemical in an assay, then the arithmetic mean was used after outlier analysis.

Effect data in the US EPA ToxCast database are provided as the 50% activity concentration (AC_{50}). AC_{50} were converted to either $EC_{10\ absolute}$ or $EC_{SR0.2\ absolute}$ as previously discussed in Neale et al. (2017a). Raw fluorescence data were collected from the ToxCast MySQL database and re-evaluated using linear concentration-effect curves to determine $EC_{IR1.5}$ for the ARE GeneBLAzer assay.

2.9 Data collection of effect data from case studies with wastewater and surface water

Case studies that applied bioassays to water extracts were collected from the peer-reviewed literature. Studies were included that reported effects either in

588 EC values in units of REF (Brion et al., 2012; Escher et al., 2017; Escher et
589 al., 2014; Escher et al., 2012; Gerlach et al., 2014; ISO/DIS19040-2, 2017;
590 König et al., 2017; Neale et al., 2015; Neale et al., 2017a; OECD, 2015;
591 Schmitt et al., 2012; U.S. EPA, 2015) or BEQs (Creusot et al., 2010;
592 Gehrmann et al., 2016; Itzel et al., 2017; Leusch et al., 2017; Tousova et al.,
593 2017). If another reference compound was used, the BEQ for the literature's
594 reference compound was translated into the BEQ with the reference
595 compound used here by the ratio of their EC values.

596 3 Results and Discussion

597 3.1 The big picture

598 Table 1 provides a summary of the recommended option for EBT-BEQ
599 derivation for all assays and in Appendix A, Table A3, one can find the
600 numerical values for all options in more detail.

601 Option A, which sums up all available BEQ, is strongly dependent on the
602 number of chemicals included. Since the number of chemicals included is
603 dependent on the data availability, it is not a robust approach to derive EBTs
604 even if it were compliant with single chemical EQS. Legally all chemicals can
605 be present just below the EQS_i but if there were many components they may
606 act together to cause a measurable effect (something-from-nothing effect).
607 However, since the EQS is likely to be no more than 100 to 1000 times lower
608 than the NOEC and LC_{50} used to derive them, this would mean in turn that 100
609 chemicals present at their (accepted) EQS may be lethal to an organism.
610 Therefore, summing up BEQ_i is not protective and was not further considered.

611 Option B leads to EBT values that appear reasonable at first sight but the
612 question is if filtering the data were important (Option F). If both high and low
613 potency compounds were present simultaneously, the low potency
614 compounds that have associated low BEQ reduced the EBT to unrealistically
615 low levels (Table A3). In these cases, it was imperative that the filtering step
616 was applied. In other cases, where all chemicals have similar relative potency,
617 filtering would not be necessary. By implementing the filtering step in all cases
618 there is no harm done and the advantage is that there is no decision point

necessary but the algorithm can be automatically run. Thus, no decisions are needed on which data to include and whether or not to apply the filtering step. Often Option B and F yielded similar EBT values if compounds of similar potency were present and in the individual sections below we will further explore Option B vs. Option F for each bioassay.

Option C is equivalent to a log-normal distribution of the BEQ_i. The resulting EBTs were much lower than those from Option A and B (Table A3), which is caused by the fact that the considered BEQs cover several orders of magnitude. The mean BEQ is influenced more by the higher values, while for logBEQ lower values have more impact on the mean. The differences are less pronounced for the filtered data that cover fewer orders of magnitude. Overall there appears to be no benefit in Option C and it was not further pursued.

Option D applies an extrapolation factor, which is equivalent to including more low-potency chemicals in the derivation of EBTs because it also reduces the EBT (Table A3). Adding an extrapolation factor would be comparable to expanding the filtering band in Option F. We have not included Option D in further discussion because the choice of the extrapolation factor would need to be justified and we tried to limit the number of decisions, to make the derivation as neutral and as data-driven as possible.

Option E, using the minimum BEQ, which had been a useful approach for estrogenicity assays when only the high-potency natural hormones were included, is very dependent on the choice of the compounds included and will be driven by low-potency compounds as is shown in Table A3. Hence, this option is not suitable and was not further considered.

Option F will be equivalent to Option B if the range of EC_i/EQS_i was fairly narrow and only high-potency compounds are included that exhibit the MOA of the given bioassay. However, in practice there were many high EC_i/EQS_i ratios (Tables A4-A51), which make the filtering step imperative.

All of the options A to F were applied to category 1 bioassays, and Option F specifically for estrogenicity assays. However, for category 2 bioassays, we saw large EC_i/EQS_i ratios and a large spread of these ratios and therefore Option H, which is based on Option B with a mixture factor, is warranted. The

choice of the mixture factor is dependent on whether a bioassay leans towards category 1 or 2, which will be discussed below.

3.2 EBTs for bioassays indicative of activation of metabolism

Activation of metabolism is not an adverse effect per se, but it indicates the presence of bioactive chemicals in a water sample. In particular, the arylhydrocarbon receptor (AhR), peroxisome proliferator activated receptor (PPARγ) and pregnane X receptor (PXR) are activated by many WWTP and surface water samples (Escher et al., 2014).

3.2.1 Arylhydrocarbon receptor AhR

For activation of the AhR the dioxin 2,3,4,7-tetrachlorodibenzo-*p*-dioxin (TCDD) is typically used as a reference compound. TCDD is not included in the list of PODs, therefore we selected benzo[a]pyrene (B[a]P) as the reference compound because polycyclic aromatic hydrocarbons (PAH) are also known activators of AhR, although they do not lead to the same toxic syndrome as dioxin-like chemicals. However, for the application as a monitoring tool, the ability to activate the receptor is sufficient, not the final adverse effect. There exists a multitude of AhR reporter gene assays with variable sensitivity (Ghorbanzadeh et al., 2014). A human liver cell line was selected for the Tox21 database (He et al., 2011) but nowadays there are even more sensitive AhR cell lines available (Brennan et al., 2015), one of which (H4L1.1c4 (rat), Table A4) was also recently tested for single chemicals and water quality (Neale et al., 2017b). The H4L1.1c4 AhR assay (Brennan et al., 2015) was about three orders of magnitude more responsive towards PAHs with an EC_{10} of $8.4 \cdot 10^{-10}$ M for B[a]P (Neale et al., 2017b) in comparison to $4.6 \cdot 10^{-7}$ M in ToxCast. Thus, it is suitable for our application because only PAHs are included in the WFD, not polychlorinated biphenyls (PCB) or dioxins. We also included the PAH-CALUX assay (Table A5), which targets specifically PAHs as activators of the AhR, with an exposure time of only 4h and an EC_{10} of B[a]P of $2 \cdot 10^{-10}$ M (Pieterse et al., 2013).

For the H4L1.1c4 AhR assay (Brennan et al., 2015) the EBT-B[a]P-EQ was 64 $pg_{B[a]P}$/L before filtering based on only four chemicals, with filtering removing all chemicals. PAH-CALUX was a similar case with an EBT-B[a]P-

683 EQ of 62 $pg_{B[a]P}$/L based on three chemicals with Option B. The fact that all
684 chemicals are filtered out was consistent with the observation by iceberg
685 modeling indicating that known chemicals can explain only a fraction of effects
686 in water samples. Hence, we needed to invoke a mixture factor in the
687 derivation of the EBT for this endpoint. A mixture factor of 100 appears
688 appropriate given that the EC_{10}/EQS ranged from below 1000 to over 10000,
689 i.e. are smaller than for the adaptive stress responses and cytotoxicity
690 endpoints. The resulting EBT-B[a]P-EQ was 6.4 $ng_{B[a]P}$/L for H4L1.1c4 AhR
691 assay and 6.2 $ng_{B[a]P}$/L for PAH-CALUX (Table 1).

692 The EBT values for AhR activity were approximately twenty times lower than
693 the SIMONI-EBTs for DR- and PAH-CALUX of 150 $ng_{B[a]P}$/L B[a]P-EQ (van der
694 Oost et al., 2017). More experimental data on single chemicals would be
695 required to refine the mixture factor, which would possibly then also improve
696 the comparability with the SIMONI-EBTs.

697 The AhR is an interesting case because to our knowledge, there are hardly
698 ever three order of magnitude differences in the EC of reference compounds
699 between different reporter gene constructs. But the present analysis
700 demonstrates that despite this large difference and the little overlap in the
701 chemicals with available EC_{10} values, our unbiased method yields fairly robust
702 and comparable EBTs.

703 ### 3.2.2 Peroxisome proliferator activated receptor PPARγ

704 The only common tested compound between PPARγ-GeneBLAzer (Table A6)
705 and PPARγ-CALUX (Table A7) was diclofenac. It was not possible to derive a
706 robust EBT with the available literature data for PPARγ-GeneBLAzer. After
707 filtering we were left with three chemicals, but diclofenac was filtered out.
708 Therefore, we had to use rosiglitazone as the reference compound. The
709 resulting EBT-rosiglitazone-EQ was 36 $ng_{rosiglitazone}$/L for PPARγ-GeneBLAzer
710 but this value is highly uncertain because it was based on only three
711 chemicals. The corresponding effect threshold is a REF of 10 (Table 1).

712 There were only two EC_{10} values for PPARγ-CALUX, and all had EC/EQS
713 ratios above 1000 and were removed in the filtering step. We could not yet
714 define an EBT for PPARγ-CALUX. It is interesting to note that the EC_{10} for the

reference compound rosiglitazone is ten times lower for PPARγ-GeneBLAzer (EC_{10} of $9.9 \cdot 10^{-10}$ M) than for PPARγ-CALUX (10^{-8} M) indicating an inherent difference in responsiveness of the two reporter gene cell lines.

We can also not conclude if a mixture factor should be included. Overall more experience must be gained about what types of waterborne contaminants activate PPARγ before a final EBT for PPARγ-GeneBLAzer and PPARγ-CALUX can be recommended.

The SIMONI EBT derived for PPARγ-CALUX was 10 $ng_{rosiglitazone}$/L rosiglitazone-EQ, which corresponded very well to the EBT derived here from a fairly weak database.

3.2.3 Pregnane X receptor PXR

The availability of single chemical data for HG5LN-hPXR (26 chemicals, Table A8) and PXR-CALUX (13 chemicals, Table A9) was excellent. As none of the typical reference compounds have assigned EQS values, di(2-ethylhexyl)-phthalate (DEHP) was chosen as the reference chemical because there was data available for both PXR cell lines. DEHP had a fairly high REP_i in relation to the typically used reference compound nicardipine with REP_i 0.16 for PXR-CALUX and 0.23 for HG5LN-hPXR and was therefore deemed suitable to serve as the reference compound. Unfortunately, after filtering, only 4 out of 26 and 6 out of 13 chemicals were left for HG5LN-hPXR and PXR-CALUX, respectively. We know also from the iceberg modeling that less than 0.1% of BEQ could be explained by analyzed chemicals (Neale et al., 2015), therefore it is necessary to invoke a mixture factor of at least 100 to account for mixture effects. The resulting EBT-DEHP-EQ were 16 μg_{DEHP}/L for HG5LN-hPXR and 272 μg_{DEHP}/L for PXR-CALUX (Table 1).

The SIMONI EBT for PXR CALUX was based on nicarpidine as the reference compound (3 $\mu g_{nicardipine}$/L, equivalent to 15 μg_{DEHP}/L). The SIMONI-EBTs PXR and PPAR were mainly based upon background BEQs that exceeded the HC_5 BEQs. Therefore, these EBTs are used to indicate non-specific chemical stresses, which is consistent with the application of a mixture factor.

3.3 EBTs for hormonal effects

3.3.1 EBTs covering bioassays for estrogenic effects

Estrogenicity provides a good testing ground for exploring the various options for EBT derivation because the effect is relevant for surface water, there is rich data available and the research community has been very active and proposed various EBTs against which the new algorithm can be tested. It must be kept in mind, though, that we know much more about estrogenicity than other biological effects and the algorithm will not make use of all of that knowledge but is the common denominator for data rich and data poor chemicals and bioassays. Eleven different ER assays were included (Table 1), nine of which were ER reporter gene assays, and two were using transgenic fish embryos targeting aromatase activity and estrogen axis activity (Brion et al., 2012; Spirhanzlova et al., 2016).

As the list of EC values contains, both, high and low-potency compounds, i.e., hormones and xenoestrogens, there is a large difference in the proposed EBT between Option B and F (Table A3). When Option B was applied, the EBT-EEQs for the various estrogenicity assays varied from 0.02 ng_{E2}/L for ER-CALUX to 0.50 ng_{E2}/L for SSTA ERα-HeLa-9903. Part of this variability is likely to be caused by true bioassay-specific sensitivity but also by the fact that the derivation was based on different chemicals.

Filtering (Option F) reduced the number of included chemicals to 3 to 11, depending on the bioassay (Tables A10 to A18) but the inclusion of high-potency hormones and low-potency xenoestrogens led to quite variable EBT-EEQs. For the ER-CALUX 11 chemicals remained after filtering and the EBT hardly changed from option B (i.e. 0.02 ng_{E2}/L) to 0.05 ng_{E2}/L, while for others the filtering step excluded many more chemicals including EE2 in case of ISO-LYES (McDonnell) which increased the EBT-EEQ >50-fold.

Option G was applied with experimental fractions of 11 % estradiol (E2), 9 % ethinylestradiol (EE2) and 80% estrone (E1) derived from experimental observations of 33 wastewater and surface water samples across a wide geographic distribution in Europe (Kase et al., 2017). Option G resulted in EBT-EEQ ranging from 0.10 ng_{E2}/L (ER-CALUX) to 1.07 ng_{E2}/L (ISO-LYES

(McDonnell)). The assays for estrogenicity are a somewhat specific case because EE2 is such a highly potent compound and always present at much lower concentrations in surface water, which makes option G necessary.

Jarosova et al. (2014) also derived bioassay-specific EBT-EEQ ranging from 0.1 to 0.4 ng_{E2}/L. Other proposals for EBT-EEQs assumed fixed, bioassay-independent values of 0.3 to 0.5 ng_{E2}/L (Kunz et al., 2017; van der Oost et al., 2017). The generic (bioassay-independent) EBTs were directly derived from estradiol without any mixture considerations (Kunz et al., 2017) or seven substances with REP > 0.001 in ER CALUX (van der Oost et al., 2017). As the responsiveness of the nine reporter gene assays varies by a factor up to ten, one common EBT-EEQ would lead to disfavoring the more responsive bioassays. Also, the reality is that the EEQ of effects of one water sample are dependent on the applied bioassays (see chapter 4.7).

The EASZY assay that applies transgenic (cyp19a1b-GFP) zebrafish embryos had an EBT-EEQ of 2.16 ng_{E2}/L but has the advantage that it is an *in vivo* endpoint taking into account the pharmacodynamics of compounds acting either directly or indirectly with the ER-regulated cyp19a1b gene (Brion et al., 2012). It also provides a true brain-specific response of fish exposed to estrogens, thus adding additional toxicological relevance to the EBT-EEQ.

The REACTIV assay using chgh-gfp transgenic medaka embryos in the presence or absence of testosterone is also an *in vivo* assay and is capable of capturing modulations in estrogen axis activity and alterations in steroidogenesis, in particular aromatase and 5α-reductase activity. The EBT-EEQ of 0.8 ng_{E2}/L for the REACTIV assay shows high consistency with the other assays.

The antagonistic mode of the estrogenicity assays (Tables A21-A23) is not relevant because all regulated chemicals were of low potency and many were also acting as agonists. Since antagonistic ER effects are rare in surface water, no EBT for anti-ER was derived.

3.3.2 EBTs for effects on the androgen receptor

For AR, the agonist mode is not relevant for most surface waters, as is also reflected in the low potency of the regulated chemicals (Tables A24-A27). Here antagonistic effects were frequently observed in surface water and EBTs were derived only for the anti AR.

The anti AR-GeneBLAzer, anti MDA-kb2, anti AR CALUX and anti AR RADAR (spiked) had 16, 18, 25 and 3 data points before and 2, 3, 4 and 2 after filtering, respectively (Tables A28-A31). This indicates that many chemicals have a fairly low specificity in the anti-androgenic assays. It is also possible that some of the relatively high anti-androgenicity observed might be due to cytotoxicity artifacts. In assays run in the agonistic mode 10% cytotoxicity is typically used as the cytotoxicity cut-off, where inducing effects are considered invalid. For the bioassay run in antagonistic mode, cytotoxicity cannot be differentiated from antagonism; therefore, the cytotoxicity cut-off would have to be much stricter than for bioassays run in the agonistic mode, which is not yet common practice. Therefore, we had to use a mixture factor of 100 on top of Option B to accommodate the low specificity of the response. The resulting EBT-Flutamide-EQ with flutamide as the reference compound were 3.3 $\mu g_{flutamide}$/L for anti AR-GeneBLAzer, 3.5 $\mu g_{flutamide}$/L for anti MDA-kb2, 14.4 $\mu g_{flutamide}$/L for the anti AR CALUX and 3.6 $\mu g_{flutamide}$/L for the anti AR RADAR (spiked). Flutamide is not an ideal reference compound. It would be desirable to take a reference compound from the list of EQS but as of now, there are no EQS defined for potent AR antagonist chemicals.

With the SIMONI approach, an EBT-flutamide-EQ of 25 $\mu g_{flutamide}$/L was derived for the anti-AR CALUX (van der Oost et al., 2017) which is less than a factor of two from our independent derivation. The SIMONI EBT for AR inhibition was mainly based on the background BEQ in order to avoid major EBT exceedances at relatively unpolluted sites. High background BEQs (exceeding HC_5 BEQs) were typically observed for the more promiscuous endpoints, such as anti AR, but also for PXR and oxidative stress (van der Oost et al., 2017). This supports our finding that a mixture factor is needed to derive the EBT-flutamide-EQ because of the lack of specificity.

We suggest deeper research into the mechanisms of antagonistic effects on the AR by environmental samples and appropriate quality control of testing. It must be shown that the effects are true competitive antagonism and not just non-specific suppression of the AR signal before an EBT for anti-AR can be adopted.

3.3.3 EBTs for effects on the progesterone receptor

The activation of PR has not been observed in surface water but 28 of the chemicals with an EQS showed an antagonistic effect on PR in the anti PR-CALUX (Table A32). However, after filtering only two chemicals remained, pointing to a similar case as the anti AR where a mixture factor of 100 had to be applied. The resulting EBT-endosulfan-EQ of 1.97 $\mu g_{endosulfan}$/L has to be treated with caution and is too preliminary to derive a final effect threshold (Table 1).

3.3.4 EBTs for effects on the glucocorticoid receptor

The activation of GR and the antagonistic effect in the presence of a GR agonist, e.g., dexamethasone is an important effect observed regularly in wastewater and surface waters (Van der Linden et al., 2008) but no EBT could be derived because there were no single chemical EQS data available for GR-CALUX (Peter Behnisch, unpublished, 2017) and all regulated chemicals were of low potency in the GR-GeneBLAzer (REP $2 \cdot 10^{-4}$ to $4 \cdot 10^{-6}$ in relation to dexamethasone, Table A33) and the anti GR-GeneBLAzer (REP $3 \cdot 10^{-4}$ to $7 \cdot 10^{-6}$ in relation to mifepristone, Table A34), which would lead to exceedingly low EBT-BEQs. Therefore, further investigations are needed to identify and add these not yet included chemicals and pharmaceuticals in future water quality research.

A SIMONI-EBT of 100 $ng_{dexamethasone}$/L DEXA-EQ was derived for the GR-CALUX (van der Oost et al., 2017), which had a fairly good discriminatory power to differentiate between wastewater (11-243 $ng_{dexamethasone}$/L DEXA-EQ) and surface waters (0.39-1.3 $ng_{dexamethasone}$/L DEXA-EQ) (Van der Linden et al., 2008).

3.3.5 EBTs for thyroid hormone-related effects

Environmental contaminants can disrupt the thyroid axis via a range of mechanisms, including altered thyroid hormone (TH) biosynthesis, secretion, plasmatic transport, binding to TH membrane transporters, TH metabolism, excretion and TR activation or inhibition (Wegner et al., 2016). A battery of *in vitro* bioassays and/or *in vivo* whole-organism bioassays is therefore required to cover all the potential MOAs of thyroid disrupters (Leusch et al., 2018).

The TTR-binding assay is an *in vitro* binding assay to measure a compound's potency to compete with thyroid hormone thyroxine (T4) or triiodothyronine (T3) for binding to its plasma transporter protein transthyretin (TTR). The TTR-radioligand binding assay (RLBA) is very sensitive to halogenated phenols (Hamers et al., 2006). In practice, one of the main routes of exposure to such compounds is via metabolism and therefore the test run in presence of S9 would be potentially a more environmentally relevant measure of the TTR-binding activity. Ren et al. (2012) developed a fluorescent variant of the TTR-binding assay, in which TTR is simultaneously incubated with thyroxine coupled to a fluorescent probe (FITC-T4) and the test compound. This variant has recently been applied to water samples (Leusch et al., 2018). Here we used as an example only TTR-binding data without S9 addition, which may be an underestimation of potential effects after metabolic activation. A preliminary EBT of 0.06 μg_{T4}/L thyroxine (T4)-EQ was derived from four available EC values with Option B for the classic TTR-RLBA, though filtering was not possible as it would have reduced the number of active chemicals to one (Table A35). For the TTR (FITC-T4) we derived a preliminary EBT-T4EQ of 0.49 μg_{T4}/L from only four EC values, which also went down to one after the filtering step (Table A36). EQS values are derived for parent compounds, whereas many TTR-binding compounds are only active after metabolism. This requires either a translation of the EQS value into EQS values of the corresponding metabolite profile using REP_i values for each metabolite (i), or the inclusion of a standardized biotransformation step in the bioassay protocol. In principle, the EBT derivation will also work with the assay run in the presence of S9 as long as in one EBT derivation all data are of the same sort and EC values with and without S9 are not mixed.

The Xenopus Embryonic Thyroid Assay (XETA) has been applied to environmental chemicals and water samples for ten years (Castillo et al., 2013; Fini et al., 2017; Leusch et al., 2018; Spirhanzlova et al., 2017; Valitalo et al., 2017). This short term *in vivo* assay, currently under validation by the OECD to become an OECD test guideline, uses transgenic xenopus embryos expressing GFP under the control of thyroid signaling. Any event leading to thyroid disruption causes an increase or a decrease in fluorescence. The test is run in two modes: unspiked and spiked. In spiked mode, T3 is added to reveal chemicals acting on the transport, metabolism or excretion of thyroid hormones or antagonizing the thyroid receptor. Various chemicals commonly found in surface water and wastewater, such as bisphenol A, diclofenac, metoprolol and perfluorooctanoic acid were active in the XETA in unspiked mode (Neale et al., 2017b). From six EC_{20} values we derived an EBT-T3EQ of 0.62 ng_{T3}/L with Option B (Table A37). Filtering reduced the data set to one chemical and as we have no information on mixture interactions, we could not further refine the EBT.

The antagonistic effect on TR was assessed with the anti TR-LUC-GH3 assay. Twenty-seven chemicals were active according to the ToxCast Database (https://comptox.epa.gov/dashboard/) but all appear to act fairly non-specifically and only one was left after the filtering step (Table A38). This is an indication that the assay is not sensitive enough to detect the chemicals at their EQS values.

Overall, the thyroid response is important for water quality assessment, but more work is required to understand mixture interactions, incorporate metabolic activation in the assays and define robust EBT and associated effect thresholds.

3.4 Bioassays for genotoxicity

The Ames test is a popular mutagenicity and genotoxicity assays for chemicals but they have rarely been used for water quality monitoring in conjunction with SPE extracts. We included two popular Ames strains TA98 and TA100 but only two of the tested chemicals were active in each strain (Tables A39-40). Many aquatic micropollutants are only genotoxic after metabolic activation, therefore we recommend a similar approach as for the

TTR binding assays where the bioassay should be run in the presence of rat liver S9 and the EBTs would then be derived for the assay with S9 (e.g., p53-CALUX (van der Linden et al., 2014)).

3.5 Bioassays for adaptive stress response

The oxidative stress response is the most prominent of all adaptive stress responses observable in surface water (Escher et al., 2014). Dichlorvos was used as the reference compound, though B[a]P is more potent but the latter is more bound to particulate matter that freely dissolved in the water phase. The less hydrophobic dichlorvos, which is still the most potent among the freely dissolved chemicals triggering oxidative stress, was therefore preferred as the reference chemical selected from the chemicals that had an overlap between EQS and EC. Eleven chemicals were active in AREc32, 24 in ARE GeneBLAzer and 7 in Nrf2-CALUX but after filtering none were left in any of these ARE assays (Tables A41-43). This is no surprise as it is well established that many chemicals activate the oxidative stress response and most of them have a rather low REP_i and can explain only a small fraction of observed effects in water samples (Escher et al., 2013).

Accordingly, a mixture factor of 1000 was applied, resulting in an EBT-dichlorvos-EQ of 156 $\mu g_{dichlorvos}$/L for AREc32, 392 $\mu g_{dichlorvos}$/L for ARE-GeneBLAzer and 26 $\mu g_{dichlorvos}$/L for Nrf2-CALUX (Table 1). The great similarity between the EBTs for the three different reporter gene constructs and cell lines indicates the robustness of the approach.

The proposed SIMONI-EBT for Nrf2-CALUX of 10 $\mu g_{curcumin}$/L translated to 6.2 $\mu g_{dichlorvos}$/L (van der Oost et al., 2017) is six times lower than the EBT derived in the present study. The SIMONI-EBT was mainly based on the background BEQ of 2.7 $\mu g_{dichlorvos}$/L.

An EBT based on a measured effect of a REF of 6 was proposed for AREc32 applied to recycled water and drinking water (Escher et al., 2013). This means that a water sample that was enriched 6 times and showed an effect causing an induction ratio of 1.5 or less would be compliant. This value compares well with the effect thresholds of the present method ranging from 10 to 34 REF (Table 1).

3.6 Whole-organism *in vivo* bioassays: EBTs meet whole-effluent testing (WET)

The method for derivation of EBT can be extended without problems to whole organism bioassays. Here discussions focus on how to include the mixture considerations. Whole organisms respond to all chemicals present in water they are sensitive to and therefore mixture considerations are warranted. However, some groups of chemicals may dominate the mixture toxicity in specifically susceptible organisms, e.g. herbicides in algae and insecticides in water flea.

Whole effluent toxicity testing is used in many countries to define emission limits of liquid waste streams (den Besten et al., 2005). In the German Wastewater Ordinance, lowest ineffective dilution (LID) is defined for industrial wastewater permits. The LID for direct discharge to receiving waters is 32 for the bacterial Microtox assay (corresponding to 3.1 % wastewater), 16 (6.2 % wastewater) for algal toxicity, 8 (12.5 % wastewater) for *Daphnia magna* and 2 (50% wastewater) for the fish embryo toxicity (FET) assay (Gartiser et al., 2009). We can compare the effect threshold with the acceptable emissions if we convert LID to units of REF (REF = 1/LID) and assume that the dilution of directly discharged wastewater would be one hundred-fold. Note that especially if wastewater effluent goes into any smaller streams or in dryer seasons, the dilution factor in the streams is significantly less than 100 and can be commonly around 10 or even less. The resulting safe enrichment factor in the river $100^{x} EC_{LID}$ is within a factor 2 to 8 from the effect threshold derived with eq. 13 from the EBTs, which is a good agreement (Table 2). This demonstrates that the effect thresholds derived here are indeed consistent with the Wastewater Ordinance.

However, it must be noted that there is a substantial difference between whole effluent testing and bioanalytical assessment of organic micropollutants extracted from water samples. The derived EBTs hold only for mixtures of organic micropollutants, hence they cannot be applied to whole effluent toxicity testing results in case some other components (metals, inorganics, DOC) are actually the causative agent in the whole water sample.

Table 2. Comparison of lowest ineffective dilution (LID) of wastewater and derived EBT for apical endpoints.

	Wastewater		Surface water		
	LID	EC_{LID} (REF)[a]	$100^{x}EC_{LID}$ (REF)[b]	Effect threshold (REF)	EBT-BEQ
Microtox	32	0.031	3.1	10	Baseline-TEQ 1.2 mg/L
Algae	16	0.063	6.3	49-247[c]	DEQ 0.08-0.12 μg_{diuron}/L
Daphnia	8	0.125	13	37	Chlorpyrifos-EQ 15 $\mu g_{chlorpyrifos}$/L
FET	2	0.5	50	119	BPA-EQ 138 μg_{BPA}/L

[a]EC_{LID} (REF) = 1/LID. [b]EC_{LID}*100 for one hundred-fold dilution of wastewater in surface water, [c]Effect threshold of REF 247 for the 72h algal growth inhibition, 70 for 24h synchronous algae reproduction, 331 for 24h combined algae assay (growth) and 49 for the 24h combined algae assay (2h-PSII inhibition).

3.6.1 Bacterial assays: Microtox

One can dispute if it is reasonable to derive EBTs for assays such as the Microtox assay that reacts to most chemicals but most act as baseline toxicants in this assay (Escher et al., 2017). If we did it as part of this exercise, and assumed a mixture factor of 10000, because all chemicals are active in the Microtox assay and all chemicals with EQS are of low potency and therefore were removed in the filtering step (Table A44), then we obtained an EBT-baseline-TEQ of 1.26 mg/L. The mixture factor stems from the low fraction of explained chemicals in iceberg modeling (Tang et al., 2013). Note that the baseline-TEQ does not refer to a specific reference compound (because all baseline toxicants are intrinsically equipotent) but to a virtual baseline toxicant, which is a generic compound of a molecular weight of 300 g/mol and a logK_{ow} of 3 (Escher et al., 2008a). The associated effect threshold is a REF of 9.7 (Table 1).

An EBT based on measured effect of a REF of 3 was proposed for recycled water and drinking water (Escher et al., 2013). The proposed SIMONI-EBT for Microtox and other apical bioassays for surface water is a REF of 20, based upon an acute-to-chronic conversion of 10 and an estimated 50%

concentration recovery (van der Oost et al., 2017). This means that water samples that were enriched 3 or 20 times and showed an effect of 50% or less would be compliant. Both thresholds are consistent with our new approach for derivation of EBT, which is a further confirmation of the robustness of the approach and the need to apply a mixture factor in the read across method.

3.6.2 Algal toxicity

Although algal toxicity is an apical endpoint and, as such, responsive to all chemicals the test organisms are sensitive to, our previous work has shown that in surface water and even in wastewater, the highly specifically acting herbicides dominate the mixture toxicity and the contribution of non-specifically acting compounds can be neglected. Accordingly, there was no need to invoke a mixture factor and filtering hardly reduced the number of included chemicals (Tables A45-48).

The 72h growth rate inhibition test with *Desmodesmus subspicatus* according to the OECD guideline (OECD, 1984) had an EBT expressed as diuron equivalent concentrations, EBT-DEQ, of 0.12 μg_{diuron}/L (Table A46). For the large-volume 24h synchronized algae reproduction assay with *Scenedesmus subspicatus* the EBT-DEQ derived without filtering was 0.08 μg_{diuron}/L and after reducing the number of eligible chemicals from 16 to 12 in the filtering step, the EBT-DEQ was 0.11 μg_{diuron}/L (Table A47). The microtiter plate-based combined algal assay had 12 EC data entries and filtering was not necessary. The resulting EBT-DEQ were 0.13 μg_{diuron}/L for the 24h growth inhibition endpoint and 0.07 μg_{diuron}/L for the photosynthesis inhibition endpoint (Tables A47-48).

The EQS for the single chemical diuron is 0.07 μg_{diuron}/L proposed by the Swiss Ecotox Centre (Ecotox Centre, 2016) and 0.2 μg_{diuron}/L in the WFD. Previous proposals for EBTs for algal toxicity have proposed to read across from diuron (Kienle et al., 2015), which indeed in this case would have been very well possible. While the one-to-one read across appears to work well for herbicides, we cannot assume that all bioassays are that straightforward or the choice of the reference compound is as evident. Therefore, we still propose to use the general algorithm for the derivation of the EBT for algal

toxicity. Compliance with the WFD diuron-EQS might be a reason to adjust the proposed SIMONI-EBT for algal growth inhibition from 0.05 to 0.025 TU, i.e., 0.19 μg_{diuron}/L DEQ (van der Oost, unpublished, 2017)

3.6.3 Acute toxicity towards *Daphnia magna*

While insecticides typically dominate the acute toxicity (48h immobilization, (OECD, 2004)) towards *Daphnia magna,* there are other non-insecticidal active chemicals that were not filtered out, e.g. anthracene, DEET or EDTA (Table A49). Due to those lower potency chemicals, it becomes necessary to apply a mixture factor of 10 to account for both, potent and weakly acting, chemicals. We used chlorpyrifos as the reference chemical and the EBT-chlorpyrifos-EQ was 15 $\mu g_{chorpyrifos}$/L with an associated effect threshold of 37 (Table 1).

For the SIMONI strategy an EBT of 0.05 TU was proposed for the *Daphnia magna* immobilization assay, i.e. an EC_{50} at REF 20 (van der Oost et al., 2017), which is within a factor of two of our proposal.

3.6.4 Fish embryo toxicity

As the mixture toxicity of water samples in the FET were typically not dominated by individual chemicals and iceberg modeling established a large gap between effects triggered by typically quantified chemicals and unknown chemicals (Neale et al., 2015), we applied a mixture factor of 100. The chosen reference chemical was bisphenol A (BPA) and the resulting EBT-BPA-EQ was 276 μg_{BPA}/L for mortality after 48h and 183 μg_{BPA}/L for mortality after 96/120h (Tables 50 and 51), equating to an associated effect threshold of 59 and 31, respectively (Table 1).

3.7 Application of EBT for assessing environmental samples

We applied the newly derived EBT-BEQ to case studies from the literature. Details are given in Appendix B. We included studies that had information on wastewater treatment and on surface water. All studies used SPE to enrich the water samples and remove matrix components, such inorganics, metals and salts, and reduce natural organic matter (Neale and Escher, 2014). The

EBTs cannot be applied for effect data from direct testing of water because they were derived from read across from EQS of organic chemicals with EQS and cannot account for matrix effects. SPE typically has a good recovery for effects for diverse SPE materials (Neale et al., 2018) and therefore SPE-extracted samples are the choice if one is interested in the organic pollutants.

The goal of the comparison of EBT with water quality case studies was to assess if the EBTs have some relationship with water quality. Of course, this analysis is limited in two ways: first, the EBTs are preliminary due to insufficient data for a robust derivation and, second, a discrimination between wastewater and surface water is not necessarily expected as there is surface water that has low quality and there are WWTP that treat water to extremely high qualities. Hence, this comparison is rather to find out if the newly derived EBTs are in a reasonable range rather than to simulate true compliance testing.

Experimental data on water quality monitoring was scarce for the H4L1.1c4 AhR assay, which was developed in 2015 and was only used in one water quality study on WWTP (Nivala et al., 2018). Both WWTP influent and effluent were above the EBT but if a hundred-fold dilution was assumed for the effluent, it would be compliant (Table B1). PPARγ-GeneBLAzer was tested in a case study in Novi Sad (König et al., 2017), where untreated wastewater would have not been compliant but the Danube as the receiving river would be compliant (Table B2). There are more case studies available for HG5LN-hPXR (Creusot et al., 2010; Escher et al., 2014; Neale et al., 2015): mostly highly treated water and surface water would be compliant and effluent would not be compliant (Table B3). The bioassays indicative for activation of metabolism have only very recently been integrated in water quality monitoring and more experience needs to be gained before associated EBTs can be implemented.

The largest number of studies was available for the estrogenicity assays. Here we would expect that untreated wastewater would be non-compliant and treated wastewater compliant or non-compliant depending on the treatment technology and surface water should be compliant. For the MELN assays such a picture was essentially found (Table B4): Danube water would have

been compliant apart from one out of 22 samples (Neale et al., 2015). Several small Swiss streams would have been compliant until WWTP effluent was added (Neale et al., 2017a). Diverse European surface water samples were 72% compliant (Tousova et al., 2017). French WWTP influent and effluent would be non-compliant, while samples taken in the river above and below the effluent discharge would often be compliant (Jugan et al., 2009; Miege et al., 2009). Ozonation in an Italian WWTP treating textile industry effluent did not lead to compliancy with EBT-EEQ (Schiliro et al., 2012). Tunisian surface water was also far above the EBT-EEQ (Mnif et al., 2012). ER-GeneBLAzer showed a similar picture in various case studies (Table B5, review in Leusch et al., 2017; Nivala et al., 2018) and ERα-Luc-BG1 was even able to detect the raw wastewater in the Danube river at Novi Sad (Table B6, König et al., 2017). ERα-HeLa-9903 performed equally well (Table B7) in one case study (Escher et al., 2014), and data was abundant but also consistent for ER-CALUX (Table B8, Bain et al., 2014; Escher et al., 2014; Leusch et al., 2010; Roberts et al., 2015; Scott et al., 2014; Van der Linden et al., 2008). Only few quantitative case studies were found for the relatively new A-YES assay (Table B9, Gehrmann et al., 2016; Itzel et al., 2017), while the number of studies was overwhelming for the popular and longer established 3d YES (Table B10), with experience going back as far as 2004 (Pawlowski et al., 2004; Tyler et al., 2005; Hashimoto et al., 2007; Williams et al., 2007; Coleman et al., 2008; Escher et al., 2008a; Escher et al., 2009; Stalter et al., 2011; Fang et al., 2012; Alvarez et al., 2013; Margot et al., 2013; Escher et al., 2014; French et al., 2015; Gehrmann et al., 2016). The case studies with the EASZY assay (Table B11) had the issue that the tested extracts were not enriched high enough to test for compliance (Neale et al., 2017a) and the REACTIV assay was so far only applied on WWTPs (Valitalo et al., 2017), so we cannot judge if the EBT is in a practical range (Table B12).

When comparing EEQ for the same samples between four different estrogenicity assays from the same study (Figure 2) it becomes evident why it is necessary to set specific EBT-EEQ for each bioassay. Both EBTs and samples have different EEQ levels but the EBT-EEQ differentiated clearly in

all cases between contaminated water (untreated wastewater and stormwater) and treated water/surface water.

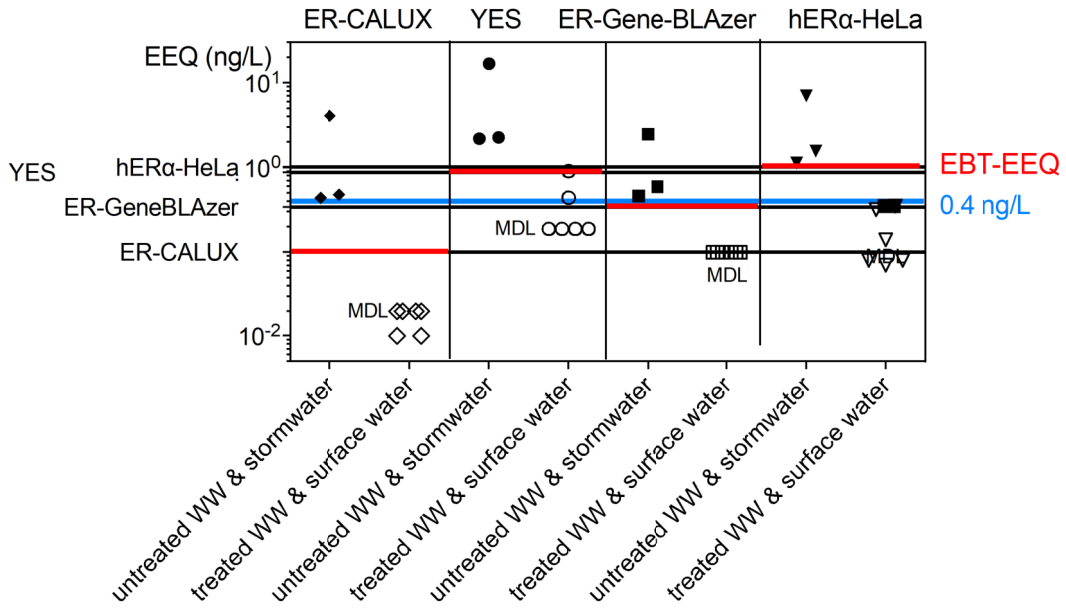


Figure 2 Comparison of EBT for different estrogenicity assays (red dashed lines) applied to one set of diverse water samples. The filled symbols refer to untreated wastewater (WW) and stormwater, the empty symbols to treated WW and surface water. Data from Escher et al. (2014).

As discussed above, the antagonistic effects on hormone receptors for AR and PR are fairly non-specific and need to be treated with some caution. At least it appears that they are in the right order of magnitude. The effects in the anti AR-GeneBLAzer were compliant up and downstream of a raw wastewater discharge in Novi Sad but not compliant at the point of discharge (Table B13, (König et al., 2017)). A similar picture was obtained for anti MDA-kb2 (Table B14) and anti AR-CALUX (Table B15). Scott et al. (2014) observed that only 16 % of Australian surface water samples were active in anti PR-CALUX with the highest concentrations being non-compliant and the lower concentrations being compliant (Table B16).

Few case studies with water samples were available for the TTR binding assays and the anti-GR-Luc-GH3 and anti GR-CALUX assays (Leusch et al., 2018). Most samples in WWTP effluent and surface water were below the detection limit in the XETA assay (Leusch et al., 2018; Tousova et al., 2017;

Valitalo et al., 2017) but those active were typically just above the EBT-T3EQ (Table B17).

There is an abundance of case studies available for the oxidative stress response. Their discriminatory power was rather mixed, though. For AREc32 (Escher et al., 2012; Escher et al., 2014; Nivala et al., 2018), the general trend was that wastewater was not compliant but surface water was (Table B18). However, some of the highly treated water exceeded the EBT. This is likely due to the formation of disinfection by-products, which activate oxidative stress response typically very strongly (Neale et al., 2012; Hebert et al., 2018). A consistent picture (Table B19) was seen for the ARE-GeneBLAzer (Neale et al., 2015; König et al., 2017; Neale et al., 2017a). We have only one case study for Nrf2-CALUX (Escher et al., 2014) and here none of the samples would have been compliant, so more experience should be gained with this assay.

Validation of EBTs with case studies is especially important for bioassays with apical endpoint because the choice of the mixture factor also depends if specifically acting or non-specifically acting chemicals dominate the mixture effects. The Microtox assay with the proposed effect threshold of 10 was generally able to differentiate between wastewater and highly treated and surface water in five monitoring studies (Table B21, Escher et al., 2008b; Escher et al., 2009; Macova et al., 2010; Macova et al., 2011; Margot et al., 2013; Escher et al., 2014). Of the algal assays, only the combined algae assay was applied in diverse monitoring studies (Escher et al., 2008b; Escher et al., 2014; Neale et al., 2017a). The EBT-DEQ for the growth endpoint could not differentiate very well between untreated and treated water (Table B22) but the EBT for photosynthesis inhibition (Table B23) achieved this differentiation. Unfortunately, only one surface monitoring study using SPE extracts and *Daphnia magna* was located in the literature (Bettinetti et al., 2014), and the EC were right around the EBT-chlorpyrifos-EQ (Table B24). The 48h FET was applied in two studies (Escher et al., 2014; Neale et al., 2015) and all but untreated wastewater was compliant (Table B25), while for the 96h FET (Tousova et al., 2017) the comparison was less conclusive (Table B26).

4 Conclusions

This analysis provides a first proof of principle for a read across approach to derive EBTs from existing EQS values and existing effect data for single chemicals. Bioassays with EBTs clearly have the potential to be used to support classification of the surface water status according to the WFD. The numerical EBT values derived here are preliminary due to a lack of complete data sets but this can be overcome in the future by targeted bioassay experiments. An improved quality control of bioassays is also required to assure the accuracy, precision, robustness, selectivity, sensitivity and specificity of each bioassay and each test performed (Escher and Leusch, 2012). However, additional measurements of EC for chemicals with accepted EQS cannot surmount the problem of lack of appropriate chemicals included in the list of priority chemicals for some bioassays. The list of existing EQS from which we have drawn does not include potent chemicals in some of the bioassays that still cover biological effects of environmental concern, i.e., effects that are frequently observed in surface water.

The proposed method for EBT derivation is simple and straightforward and is provided in the form of an excel sheet as Appendix A of this manuscript. It was possible to derive preliminary EBT for 32 bioassays out of the 48 bioassays included in the analysis (tabs A4 to A51 in Appendix A). There is even a blank sheet included at the end (tab “A52. General template” in Appendix A) to encourage readers to derive EBT values. Therefore, as EQS are evolving and new EQS are being implemented or revised, the database for the derivation of EBTs can be appended. Moreover, as new bioassay data are becoming available the number of input data will increase and make the approach more robust and less sensitive to outliers. There are many data available for single chemicals in the included *in vitro* bioassays in the dashboard of the US EPA (Tox21 and ToxCAST; (U.S. EPA, 2015)) and in diverse publications (van der Linden et al., 2014; Di Paolo et al., 2016; Neale et al., 2017b) but as discussed above, complete datasets for all chemicals with EQS would greatly improve the robustness and quality of the derived EBTs.

At present, lack of data is the largest impediment for the definition of EBTs with the proposed read across method. As this exercise has demonstrated, a lack of effect data for chemicals that have EQS is one of the most urgent gaps to close to advance the derivation of EBTs. Proposals for bioassay test batteries are sometimes very comprehensive and cover as many chemicals as possible (Escher et al., 2014; Wernersson et al., 2015), while others only include assays that are likely to light up with water samples (Neale et al., 2017b). Here we have mainly included bioassays for which there exist monitoring data with WWTP effluent or surface water samples. Thus, we are certain about their relevance for water quality assessment.

However, sufficient chemicals were not available for all endpoints in our POD list of EQS values or experimental data for the chemicals with EQS were lacking. Sometimes there are data, but the EC values are pointing to a rather non-specific effect. This includes all assays for anti-estrogenicity, as well as activation of PR and GR. Also many of the compounds active in some assays, especially in AhR and TTR, are very hydrophobic. They hardly dissolve in the water phase of the aquatic environment, but rather adsorb to sediment and suspended particulate matter. Consequently, no EQS values in water are available for many of the active compounds in these bioassays, hampering the derivation of EQS-based EBT values but also posing the question if EBTs need to be expanded to sediment and soils. In principle, there is no limitation, provided that there are EQS available for these compartments and single chemical data in the bioassays is sufficiently abundant.

We have clearly demonstrated that Option B, i.e. the mean of BEQ_i, performed best for bioassays with only high-potency compounds in the list of chemicals with EQS. If the list contains high- and low-potency compounds a filtering step was necessary to exclude those compounds with too low potency because they would have decreased the EBT to unrealistically low levels (Option F). The estrogenicity assays and whole organism assay specifically sensitive to certain chemicals, e.g., algal toxicity dominated by herbicides or daphnia toxicity dominated by insecticides, yielded robust EBT-BEQ after filtering.

Those bioassays, where the filtering excluded most or all chemicals, were all category 2 bioassays with many bioactive but low potency chemicals, for which read across is more difficult and only possible if mixture factors are included in the algorithm. The category 2 bioassays include the bioassays indicative of activation of metabolism (PXR, PPAR, AhR) and assays for adaptive stress responses, with the oxidative stress response activated via the keap-Nrf2-ARE pathway particularly relevant for water quality. For these assays, previous iceberg modeling has shown that even if hundreds of chemicals are analyzed and the effects are known, their predicted mixture toxicity explains far less than 1% of the observed biological effect of the sample (Escher et al., 2013; Neale et al., 2017b; Tang et al., 2014; Tang et al., 2013). Many of the chemicals have very low potencies and therefore one or few BEQ cannot be representative. In the present study, we applied mixture factors of 100 for receptor-mediated effects with low specificity (PXR, AhR) and 1000 for oxidative stress response. In the case of the Microtox assay, every chemical can contribute to mixture toxicity and the antibiotics do not have a highly specific effect in the standard 30 min incubation, therefore the mixture factor was increased to 10000. For *Daphnia magna* the mixture factor was reduced to 10 because the mixture effect is driven by insecticides and non-specifically acting chemicals acting together. If only insecticides were included then the EBT came also to 15 $ng_{chlorpyrifos}/L$ with option F but only based on three chemicals, therefore less robust than the mixture method. The derivation of the mixture factor is the Achilles' heel of the approach. The analysis of existing single-chemical effect data as well as the iceberg modeling in the case studies clearly indicates the need for the mixture factor approach but the derivation of the mixture factor is not mature yet and the proposed values have to be considered preliminary until further information on mixture effects becomes available.

It must be noted that the list of priority compounds in the WFD was not defined with any consideration of covering relevant MOAs. One could now argue that EBTs should be derived from other PODs than the EQS. Alternatively, one could argue that the WFD list of priority pollutants should be expanded to include chemicals representative for those MOAs. The latter

option is preferred because chemical assessment is the current gold standard of water quality monitoring. Most proposals to date are suggesting EBM as a screening tool and not as a replacement of chemical analytical monitoring and this team of authors supports this view. Chemical analyses will always be necessary in risk assessment, but it is most relevant at sites where bioanalytical screening indicates that micropollutants' levels may pose a risk. This strategy is also applied in the food industry (Hoogenboom et al., 2010).

Despite the limited effect data availability and limitations of the existing lists of EQS, the proposed generic methods to derive EBTs are a first step to harmonize existing approaches and explore various different options for a large diversity of bioassays commonly applied for water quality assessment. Research groups active in bioassay research are encouraged to fill gaps in availability of effect concentrations for chemicals that are relevant in surface water and have a defined EQS. Excel spreadsheets are provided that allow inclusion of more chemicals and more effect data to derive more and more robust EBTs.

Acknowledgements

The project SOLUTIONS is supported by the European Union Seventh Framework Programme (FP7-ENV-2013-two-stage Collaborative project) under grant agreement number 603437. Escher and Neale were supported by the National Health and Medical Research Council (NHMRC) – European Union Collaborative Research Grant (APP1074775). We also acknowledge support by the NORMAN network and want to thank all NORMAN members and especially Valeria Dulio for support of this activity. We thank the RWTH Aachen, the Swiss Centre for Applied Ecotoxicology and the Helmholtz Centre for Environmental Research UFZ for hosting workshops that discussed the EBT derivation. We thank Barbora Jarosova (Recetox), for helpful discussions, Rolf Altenburger (UFZ) and Wibke Busch (UFZ) for reviewing the manuscript and Nils Klüver for sharing literature effect data.

Appendices

Appendix A: Excel sheet with POD (Table A1), extracts from the ETOX database (Table A2), detailed summary of all EBT options (Table A3) and all data on EQS and single chemical effect concentrations used for the derivation of the EBTs (Tables A4 to A51). Table A52 template of the EBT derivations proposed here including all equations for use by reader. Appendix B: Excel sheets with diverse case studies from literature for application testing of the proposed EBT (Tables B1-B26).

References

Altenburger R, Bodeker W, Faust M, Grimme LH. Evaluation of the isobologram method for the assessment of mixtures of chemicals - Combination effect studies with pesticides in algal biotests. Ecotoxicology and Environmental Safety 1990; 20: 98-114.

Alvarez DA, Shappell NW, Billey LO, Bermudez DS, Wilson VS, Kolpin DW, et al. Bioassay of estrogenicity and chemical analyses of estrogens in streams across the United States associated with livestock operations. Water Research 2013; 47: 3347-3363.

Backhaus T, Faust M. Predictive Environmental Risk Assessment of Chemical Mixtures: A Conceptual Framework. Environmental Science & Technology 2012; 46: 2564-2573.

Bain PA, Williams M, Kumar A. Assessment of multiple hormonal activities in wastewwater at different stages of treatment. Environmental Toxicology and Chemistry 2014; 33: 2297-2307.

Balaguer P, Francois F, Comunale F, Fenet H, Boussioux AM, Pons M, et al. Reporter cell lines to study the estrogenic effects of xenoestrogens. Science of the Total Environment 1999; 233: 47-56.

Bettinetti R, Ponti B, Quadroni S. An Ecotoxicological Approach to Assess the Environmental Quality of Freshwater Basins: A Possible Implementation of the EU Water Framework Directive? Environments 2014: 92-106.

Brack W, Ait-Aissa S, Burgess RM, Busch W, Creusot N, Di Paolo C, et al. Effect-directed analysis supporting monitoring of aquatic environments - An in-depth overview. Science of the Total Environment 2016; 544: 1073-1118.

Brack W, Dulio V, Agerstrand M, Allan I, Altenburger R, Brinkmann M, et al. Towards the review of the European Union Water Framework Directive: Recommendations for more efficient assessment and management of chemical contamination in European surface water resources. Science of the Total Environment 2017; 576: 720-737.

Brand W, De Jongh CM, van Linden SC, Mennes W, Puijker LM, van Leeuwen C, et al. Trigger values for investigation of hormonal activity in

drinking water and its sources using CALUX bioassays. Environment International 2013; 55: 109-118.

Brennan JC, He G, Tsutsumi T, Zhao J, Wirth E, Fulton MH, et al. Development of Species-Specific Ah Receptor-Responsive Third Generation CALUX Cell Lines with Enhanced Responsiveness and Improved Detection Limits. Environmental Science & Technology 2015; 49: 11903-11912.

Brion F, Le Page Y, Piccini B, Cardoso O, Tong SK, Chung BC, et al. Screening estrogenic activities of chemicals or mixtures *in vivo* using transgenic (cyp19a1b-GFP) zebrafish embryos. Plos One 2012; 7.

Castillo L, Seriki K, Mateos S, Loire N, Guedon N, Lemkine GF, et al. In vivo endocrine disruption assessment of wastewater treatment plant effluents with small organisms. Water Science and Technology 2013; 68: 261-268.

Coleman HM, Khan SJ, Watkins G, Stuetz RM. Fate and analysis of endocrine disrupting chemicals in some sewage treatment plants in Australia. Water Science and Technology 2008; 58: 2187-2194.

Creusot N, Kinani S, Balaguer P, Tapie N, LeMenach K, Maillot-Marechal E, et al. Evaluation of an hPXR reporter gene assay for the detection of aquatic emerging pollutants: screening of chemicals and application to water samples. Analytical and Bioanalytical Chemistry 2010; 396: 569-583.

den Besten PJ, Leverett D, Tonkes M. Chapter 2- Bioassays and tiered approaches for monitoring surface water quality and effluents. In: den Besten, editor. Ecotoxicological Testing of Marine and Freshwater Ecosystems. Emerging Techniques, Trends and Strategies, CRC Press 2005.

Di Paolo C, Ottermanns R, Keiter S, Ait-Aissa S, Bluhm K, Brack W, et al. Bioassay battery interlaboratory investigation of emerging contaminants in spiked water extracts - Towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring. Water Research 2016; 104: 473-484.

Ecotox Centre. EQS proposal of the Ecotox Centre for diuron:http://www.ecotoxcentre.ch/expert-service/quality-standards/. 2016.

Escher BI, Baumer A, Bittermann K, Henneberger L, König M, Kühnert C, et al. General baseline toxicity QSAR for non-polar, polar and ionisable chemicals and their mixtures in the bioluminescence inhibition assay with *Alliivibrio fischeri*. Environmental Science: Processes & Impacts 2017; 19: 414-428.

Escher B, Leusch F. Bioanalytical tools in water quality assessment. London, UK: IWA Publishing, 2012.

Escher BI, Allinson M, Altenburger R, Bain P, Balaguer P, Busch W, et al. Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro bioassays. Environmental Science & Technology 2014; 48: 1940-1956.

Escher BI, Bramaz N, Mueller JF, Quayle P, Rutishauser S, Vermeirssen ELM. Toxic equivalent concentrations (TEQs) for baseline toxicity and specific modes of action as a tool to improve interpretation of ecotoxicity testing of environmental samples. Journal of Environmental Monitoring 2008a; 10: 612-621.

Escher BI, Bramaz N, Ort C. JEM Spotlight: Monitoring the treatment efficiency of a full scale ozonation on a sewage treatment plant with a mode-of-action based test battery. Journal of Environmental Monitoring 2009; 11: 1836-1846.

Escher BI, Bramaz N, Quayle P, Rutishauser S, Vermeirssen ELM. Monitoring of the ecotoxicological hazard potential by polar organic micropollutants in sewage treatment plants and surface waters using a mode-of-action based test battery. Journal of Environmental Monitoring 2008b; 10: 622-631.

Escher BI, Dutt M, Maylin E, Tang JYM, Toze S, Wolf CR, et al. Water quality assessment using the AREc32 reporter gene assay indicative of the oxidative stress response pathway. J. Environ. Monitor. 2012; 14: 2877-85.

Escher BI, Neale PA, Leusch FDL. Effect-based trigger values for in vitro bioassays: Reading across from existing water quality guideline values. Water Research 2015; 81: 137-148.

Escher BI, van Daele C, Dutt M, Tang JYM, Altenburger R. Most oxidative stress response in water samples comes from unknown chemicals: the need for effect-based water quality trigger values. Environmental Science & Technology 2013; 47: 7002-7011.

ETOX. Information System Ecotoxicology and Environmental Quality Targets, Umweltbundesamt, https://webetox.uba.de/webETOX. 2017.

European Parliament and European Council. Directive 2000/60/EC of the European Parliament and the Council of 23 October 2000 establishing a framework for Community action in the field of water policy (short: Water Framework Directive) Official Journal of the European Communities 2000; 327/1.

EuropeanCommission. Common implementation strategy for the water framework directive (2000/60/EC). Guidance document No. 27. Technical guidance for deriving environmental quality standards. European Communities, 2011.

Fang Y-X, Ying G-G, Zhao J-L, Chen F, Liu S, Zhang L-J, et al. Assessment of hormonal activities and genotoxicity of industrial effluents using *in vitro* bioassays combined with chemical analysis. Environmental Toxicology and Chemistry 2012; 31: 1273-1282.

Federal Ministry for the Environment; Nature Conservation & Nuclear Safety; Ordinance on Requirements for the Discharge of Waste Water into Waters (Waste Water Ordinance - AbwV). Federal Law Gazette (BGBl) I 2004: 1108.

Fini J-B, Mughal BB, Le Mevel S, Leemans M, Lettmann M, Spirhanzlova P, et al. Human amniotic fluid contaminants alter thyroid hormone signalling and early brain development in Xenopus embryos. Scientific Reports 2017; 7.

Fini JB, Le Mevel S, Turque N, Palmier K, Zalko D, Cravedi JP, et al. An *in vivo* multiwell-based fluorescent screen for monitoring vertebrate thyroid hormone disruption. Environmental Science & Technology 2007; 41: 5908-5914.

Freitas J, Cano P, Craig-Veit C, Goodson ML, David Furlow J, Murk AJ. Detection of thyroid hormone receptor disruptors by a novel stable *in vitro* reporter gene assay. Toxicology in Vitro 2011; 25: 257-266.

French VA, King SC, Kumar A, Northcott G, McGuinness K, Parry D. Characterisation of microcontaminants in Darwin Harbour, a tropical estuary of northern Australia undergoing rapid development. Science of the Total Environment 2015; 536: 639-647.

Gartiser S, Hafner C, Oeking S, Paschke A. Results of a "Whole Effluent Assessment" study from different industrial sectors in Germany according to OSPAR's WEA strategy. Journal of Environmental Monitoring 2009; 11: 359-369.

Gehrmann L, Bielak H, Behr M, Itzel F, Lyko S, Simon A, et al. (Anti-) estrogenic and (anti-)androgenic effects in wastewater during advanced treatment: comparison of three *in vitro* bioassays. Environmental Science and Pollution Research 2016: DOI 10.1007/s11356-016-7165-4.

Gerlach T, Knaust J, Kaiser C, Korner M, Hettwer K, Uhlig S, et al. Development and assessment of a novel Arxula adeninivorans androgen screen (A-YAS) assay and its application in analysis of cattle urine. Science of the Total Environment 2014; 490: 1073-1081.

Ghorbanzadeh M, van Ede KI, Larsson M, van Duursen MBM, Poellinger L, Lucke-Johansson S, et al. In Vitro and in Silico Derived Relative Effect Potencies of Ah-Receptor-Mediated Effects by PCDD/Fs and PCBs in Rat, Mouse, and Guinea Pig CALUX Cell Lines. Chemical Research in Toxicology 2014; 27: 1120-1132.

Gijsbers L, Man HY, Kloet SK, de Haan LH, Keijer J, Rietjens IM, et al. Stable reporter cell lines for peroxisome proliferator-activated receptor gamma (PPARgamma)-mediated modulation of gene expression. Analytical Biochemistry 2011; 414: 77-83.

Hamers T, Kamstra JH, Sonneveld E, Murk AJ, Kester MHA, Andersson PL, et al. *In vitro* profiling of the endocrine-disrupting potency of brominated flame retardants. Toxicological Sciences 2006; 92: 157-173.

Hamers T, Legler J, Blaha L, Hylland K, Marigomez I, Schipper CA, et al. Expert opinion on toxicity profiling - Report from a NORMAN expert group meeting. Integrated Environmental Assessment and Management 2013; 9: 185-191.

Hashimoto T, Onda K, Nakamura Y, Tada K, Miya A, Murakami T. Comparison of natural estrogen removal efficiency in the conventional activated sludge process and the oxidation ditch process. Water Research 2007; 41: 2117-2126.

Hashmi MAK, Escher BI, Krauss M, Teodorovic I, Brack W. Effect-directed analysis (EDA) of Danube River water sample receiving untreated municipal wastewater from Novi Sad, Serbia. Science of The Total Environment 2018; 624: 1072 - 1081.

He GC, Tsutsumi T, Zhao B, Baston DS, Zhao J, Heath-Pagliuso S, et al. Third-generation Ah receptor-responsive luciferase reporter plasmids: Amplification of dioxin-responsive elements dramatically increases CALUX bioassay sensitivity and responsiveness. Toxicological Sciences 2011; 123: 511-522.

Hebert A, Felier C, Lecarpentier C, Neale P, Schichting R, Thibert S, et al. Bioanalytical assessment of adaptive stress responses in drinking water as a tool to differentiate between micropollutants and disinfection by-products. Water Research 2018; 132: 340-349.

Hettwer K, Jähne M, Frost K, Giersberg M, Kunze G, Trimborn M, et al. Validation of Arxula Yeast Estrogen Screen assay for detection of estrogenic activity in water samples: Results of an international interlaboratory study. Science of The Total Environment 2018; 621: 612- 625.

Hoogenboom L, Haedrich J, Eppe G, Goeyens L, Elskens M, Malagocki P, et al. Revised EU-criteria for applying bioanalytical methods for screening of feed and food for dioxins and dioxin-like PCBs. Organohalogen Compounds 2010; 72.

Huang R, Sakamuru S, Martin MT, Reif DM, Judson RS, Houck KA, et al. Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway. Scientific reports 2014; 4: 5664-5664.

Huang RL, Xia MH, Cho MH, Sakamuru S, Shinn P, Houck KA, et al. Chemical genomics profiling of environmental chemical modulation of human nuclear receptors. Environmental Health Perspectives 2011; 119: 1142-1148.

ISO/DIS19040-1. Water quality -- Determination of the estrogenic potential of water and waste water -- Part 1: Yeast estrogen screen (*Saccharomyces cerevisiae*). International Organization for Standardization (ISO), Geneva, Switzerland, 2017.

ISO/DIS19040-2. Water quality — Determination of the estrogenic potential of water and waste water —Part 2: Yeast estrogen screen (A-YES, Arxula adeninivorans). International Organization for Standardization (ISO), Geneva, Switzerland, 2017.

Itzel F, Gehrmann L, Bielak H, Ebersbach P, Boergers A, Herbst H, et al. Investigation of full-scale ozonation at a municipal wastewater treatment plant using a toxicity-based evaluation concept. Journal of Toxicology and Environmental Health, Part A 2017; 80: 1242-1258.

Jarosova B, Blaha L, Giesy JP, Hilscherova K. What level of estrogenic activity determined by *in vitro* assays in municipal wastewaters can be considered as safe? Environment International 2014; 64: 98-109.

Jia A, Escher BI, Leusch FDL, Tang JYM, Prochazka E, Dong B, et al. In vitro bioassays to evaluate complex chemical mixtures in recycled water. Water Research 2015; 80: 1-11.

Judson R, Houck K, Martin M, Richard AM, Knudsen TB, Shah I, et al. Analysis of the Effects of Cell Stress and Cytotoxicity on In Vitro Assay Activity Across a Diverse Chemical and Assay Space. Toxicological Sciences 2016; 152: 323-339.

Jugan ML, Oziol L, Bimbot M, Huteau V, Tamisier-Karolak S, Blondeau JP, et al. *In vitro* assessment of thyroid and estrogenic endocrine disruptors in wastewater treatment plants, rivers and drinking water supplies in the greater Paris area (France). Science of the Total Environment 2009; 407: 3579-3587.

Kase R, Javurkova B, Simon E, Swart K, Carere M, Könemann S, et al. Screening and risk management solutions for steroidal estrogens in surface and wastewater. Trends Analyt. Chem. 2017: submitted.

Kienle C, Vermeirssen E, Kunz P, Werner I. Grobbeurteilung der Wasserqualität von abwasserbelasteten Gewässern anhand von ökotoxikologischen Biotests. Studie im Auftrag des BAFU (aktualisiert 2017). Évaluation sommaire de la qualité de l'eau dans les cours d'eau pollués par des effluents d'épuration à l'aide de bioessais écotoxicologiques. Étude réalisée sur mandat de l'OFEV (actualisé 2017). Centre for Applied Ecotoxicology, Dübendorf, Switzerland. 2015.

König M, Escher BI, Neale PA, Krauss M, Hilscherová K, Novák J, et al. Impact of untreated wastewater on a major European river evaluated with a combination of *in vitro* bioassays and chemical analysis. Environmental Pollution 2017; 220: 1220-1230.

Kunz PY, Kienle C, Carere M, Homazava N, Kase R. *In vitro* bioassays to screen for endocrine active pharmaceuticals in surface and waste waters. Journal of Pharmaceutical and Biomedical Analysis 2015; 106: 107-115.

Kunz PY, Simon E, Creusot N, Jayasinghe BS, Kienle C, Maletz S, et al. Effect-based tools for monitoring estrogenic mixtures: Evaluation of five *in vitro* bioassays. Water Research 2017; 110: 378-388

Lemaire G, Mnif W, Pascussi JM, Pillon A, Rabenoelina F, Fenet H, et al. Identification of new human pregnane X receptor ligands among pesticides using a stable reporter cell system. Toxicological Sciences 2006; 91: 501-509.

Leusch FDL, Aneck-Hahn NH, Cavanagh J-AE, Du Pasquier D, Hamers T, Hebert A, et al. Comparison of in vitro and in vivo bioassays to measure thyroid hormone disrupting activity in water extracts. Chemosphere 2018; 191: 868- 875.

Leusch FDL, Chapman HF, Korner W, Gooneratne SR, Tremblay LA. Efficacy of an advanced sewage treatment plant in southeast Queensland, Australia, to remove estrogenic chemicals. Environmental Science & Technology 2005; 39: 5781-5786.

Leusch FDL, De Jager C, Levi Y, Lim R, Puijker L, Sacher F, et al. Comparison of five *in vitro* bioassays to measure estrogenic activity in environmental waters. Environmental Science & Technology 2010; 44: 3853-3860.

Leusch FDL, Khan SJ, Laingam S, Prochazka E, Trinh T, Froscio S, et al. Assessment of the application of bioanalytical tools as surrogate measure of chemical contaminants in recycled water. Water Research 2014; 49: 300-315.

Leusch FDL, Neale PA, Hebert A, Scheurer M, Schriks MCM. Analysis of the sensitivity of *in vitro* bioassays for androgenic, progestagenic,

glucocorticoid, thyroid and estrogenic activity: Suitability for drinking and environmental waters. Environment International 2017; 99: 120-130.

Leusch FDL, Snyder SA. Bioanalytical tools: half a century of application for potable reuse. Environmental Science: Water Research Technolology 2015: 606-621.

Macova M, Escher BI, Reungoat J, Carswell S, Chue KL, Keller J, et al. Monitoring the biological activity of micropollutants during advanced wastewater treatment with ozonation and activated carbon filtration. Water Research 2010; 44: 477-492.

Macova M, Toze S, Hodgers L, Mueller JF, Bartkow ME, Escher BI. Bioanalytical tools for the evaluation of organic micropollutants during sewage treatment, water recycling and drinking water generation. Water Research 2011; 45: 4238-4247.

Maletz S, Floehr T, Beier S, Klumper C, Brouwer A, Behnisch P, et al. In vitro characterization of the effectiveness of enhanced sewage treatment processes to eliminate endocrine activity of hospital effluents. Water Research 2013; 47: 1545-1557.

Margot J, Kienle C, Magnet A, Weil M, Rossi L, de Alencastro LF, et al. Treatment of micropollutants in municipal wastewater: Ozone or powdered activated carbon? Science of the Total Environment 2013; 461: 480-498.

Miege C, Karolak S, Gabet V, Jugan ML, Oziol L, Chevreuil M, et al. Evaluation of estrogenic disrupting potency in aquatic environments and urban wastewaters by combining chemical and biological analysis. TRAC-Trends in Analytical Chemistry 2009; 28: 186-195.

Mnif W, Zidi I, Hassine AIH, Gomez E, Bartegi A, Roig B, et al. Monitoring endocrine disrupter compounds in the Tunisian Hamdoun River using *in vitro* bioassays. Soil & Sediment Contamination 2012; 21: 815-830.

Muschket M, di Paolo C, Tindall AJ, Touak.G, Phan A, Krauss M, et al. Identification of Unknown Antiandrogenic Compounds in Surface Waters by Effect-Directed Analysis (EDA) Using a Parallel Fractionation Approach. Environmental Science & Technology 2018; 52: 288–297.

Neale PA, Ait-Aissa S, Brack W, Creusot N, Denison MS, Deutschmann B, et al. Linking in Vitro Effects and Detected Organic Micropollutants in Surface Water Using Mixture-Toxicity Modeling. Environmental Science & Technology 2015; 49: 14614-14624.

Neale PA, Antony A, Bartkow M, Farre M, Heitz A, Kristiana I, et al. Bioanalytical assessment of the formation of disinfection by-products in a drinking water treatment plant. Environmental Science & Technology 2012; 46: 10317–10325.

Neale PA, Escher BI. Does co-extracted dissolved organic carbon cause artefacts in cell-based bioassays? Chemosphere 2014; 108: 281-288.

Neale PA, Munz NA, Ait-Aissa S, Altenburger R, Brion F, Busch W, et al. Integrating Chemical Analysis and Bioanalysis to Evaluate the

Contribution of Wastewater Effluent on the Micropollutant Burden in Small Streams. Science of the Total Environment 2017a; 576: 785-795.

Neale PN, Altenburger R, Ait-Aissa S, Brion F, Busch W, de Aragão Umbuzeiro G, et al. Development of a bioanalytical test battery for water quality monitoring: Fingerprinting identified micropollutants and their contribution to effects in surface water. Water Research 2017b; 123: 734-750.

Neale PN, Brack W, Ait-Aissa S, Busch W, Hollender J, Krauss M, et al. Solid-phase extraction as sample preparation of water samples for cell-based and other *in vitro* bioassays. Environ. Sci. Process. Impacts 2018: submitted.

Nivala J, Neale PN, Haasis T, Kahl S, König M, Müller R, et al. Application of bioanalytical tools to evaluate treatment efficacy of conventional and intensified treatment wetlands. Environmental Science: Water Research & Technology 2018: in press, DOI: 10.1039/C7EW00341B.

NRMMC & EPHC & NHMRC. Australian guidelines for water recycling: managing health and environmental risks (phase 2). Augmentation of drinking water supplies. National Water Quality Management Strategy (NWQMS), Natural Resource Management Ministerial Council (NRMMC), Environment Protection and Heritage Council (EPHC) and National Health and Medical Research Council (NHRMC), Canberra, Australia. http://www.environment.gov.au/resource/national-water-quality-management-strategy-australian-guidelines-water-recycling-managing 2008.

OECD. Test guideline no. 201. Alga, growth inhibition test. OECD guidelines for testing of chemicals. Environmental Directorate, Organisation for Economic Co-operation and Development, Paris, France, 1984.

OECD. Test guideline no. 202. *Daphnia sp.* acute immobilisation test. OECD guidelines for testing of chemicals. Environmental Directorate, Organisation for Economic Co-operation and Development, Paris, France, 2004.

OECD. Test guideline no. 236. Fish Embryo Acute Toxicity Test. OECD guidelines for testing of chemicals. Environmental Directorate, Organisation for Economic Co-operation and Development, Paris, France, 2013.

OECD. Test guideline no. 455. Performance-Based Test Guideline for Stably Transfected Transactivation In Vitro Assays to Detect Estrogen Receptor Agonists and Antagonist. OECD guidelines for testing of chemicals. Environmental Directorate, Organisation for Economic Co-operation and Development, Paris, France, 2015.

Pawlowski S, Ternes TA, Bonerz M, Rastall AC, Erdinger L, Braunbeck T. Estrogenicity of solid phase-extracted water samples from two municipal sewage treatment plant effluents and river Rhine water using the yeast estrogen screen. Toxicology in Vitro 2004; 18: 129-138.

Pieterse B, Felzel E, Winter R, van der Burg B, Brouwer A. PAH-CALUX, an Optimized Bioassay for AhR-Mediated Hazard Identification of Polycyclic Aromatic Hydrocarbons (PAHs) as Individual Compounds

and in Complex Mixtures. Environmental Science & Technology 2013; 47: 11651-11659.

Prasse C, Stalter D, Schulte-Oehlmann U, Oehlmann J. Spoilt for choice: A critical review on the chemical and biological assessment of current wastewater treatment technologies. Water Research 2015; 87: 237-270.

Reifferscheid G, Maes HM, Allner B, Badurova J, Belkin S, Bluhm K, et al. International round-robin study on the Ames fluctuation test. Environmental and Molecular Mutagenesis 2012; 53: 185-197.

Ren XM, Guo LH. Assessment of the Binding of Hydroxylated Polybrominated Diphenyl Ethers to Thyroid Hormone Transport Proteins Using a Site-Specific Fluorescence Probe. Environmental Science & Technology 2012; 46: 4633-4640.

Roberts J, Bain PA, Kumar A, Hepplewhite C, Ellis DJ, Christy AG, et al. Tracking multiple modes of endocrine activity in Australia's largest inland sewage treatment plant and effluent- receiving environment using a panel of *in vitro* bioassays. Environmental Toxicology and Chemistry 2015; 34: 2271-2281.

Rotroff DM, Martin MT, Dix DJ, Filer DL, Houck KA, Knudsen TB, et al. Predictive Endocrine Testing in the 21st Century Using in Vitro Assays of Estrogen Receptor Signaling Responses. Environmental Science & Technology 2014; 48: 8706-8716.

Routledge EJ, Sumpter JP. Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen. Environmental Toxicology and Chemistry 1996; 15: 241-248.

Schiliro T, Porfido A, Spina F, Varese GC, Gilli G. Oestrogenic activity of a textile industrial wastewater treatment plant effluent evaluated by the E-screen test and MELN gene-reporter luciferase assay. Science of the Total Environment 2012; 432: 389-395.

Schmitt S, Reifferscheid G, Claus E, Schlusener M, Buchinger S. Effect directed analysis and mixture effects of estrogenic compounds in a sediment of the river Elbe. Environmental Science and Pollution Research 2012; 19: 3350-3361.

Schulze T, Ahel M, Ahlheim J, Ait-Aissa S, Brion F, Di Paolo C, et al. Assessment of a novel device for onsite integrative large-volume solid phase extraction of water samples to enable a comprehensive chemical and effect-based analysis. Science of the Total Environment 2017; 581: 350-358.

Scott PD, Bartkow M, Blockwell SJ, Coleman HM, Khan SJ, Lim R, et al. An assessment of endocrine activity in Australian rivers using chemical and in vitro analyses. Environmental Science and Pollution Research 2014; 21: 12951-12967.

Sebillot A, Damdimopoulou P, Ogino Y, Spirhanzlova P, Miyagawa S, Du Pasquier D, et al. Rapid Fluorescent Detection of (Anti)androgens with spiggin-gfp Medaka. Environmental Science & Technology 2014; 48: 10919-10928.

Silva E, Rajapakse N, Kortenkamp A. Something from "nothing"– eight weak estrogenic chemicals combined at concentrations below NOECs produce significant mixture effects. Environmental Science & Technology 2002; 36: 1751-1756.

Simmons SO, Fan C-Y, Ramabhadran R. Cellular Stress Response Pathway System as a Sentinel Ensemble in Toxicological Screening. Toxicol. Sci. 2009; 111: 202-225.

Sonneveld E, Jansen HJ, Riteco JAC, Brouwer A, van der Burg B. Development of androgen- and estrogen-responsive bioassays, members of a panel of human cell line-based highly selective steroid-responsive bioassays. Toxicological Sciences 2005; 83: 136-148.

Sonneveld E, Pieterse B, Schoonen WG, van der Burg B. Validation of *in vitro* screening models for progestagenic activities: Inter-assay comparison and correlation with *in vivo* activity in rabbits. Toxicology in Vitro 2011; 25: 545-554.

Spirhanzlova P, De Groef B, Nicholson FE, Grommen SVH, Marras G, Sebillot A, et al. Using short-term bioassays to evaluate the endocrine disrupting capacity of the pesticides linuron and fenoxycarb. Comparative Biochemistry and Physiology C-Toxicology & Pharmacology 2017; 200: 52-58.

Spirhanzlova P, Leleu M, Sebillot A, Lemkine GF, Iguchi T, Demeneix BA, et al. Oestrogen reporter transgenic medaka for non-invasive evaluation of aromatase activity. Comparative Biochemistry and Physiology C-Toxicology & Pharmacology 2016; 179: 64-71.

Stalter D, Magdeburg A, Wagner M, Oehlmann J. Ozonation and activated carbon treatment of sewage effluents: Removal of endocrine activity and cytotoxicity. Water Research 2011; 45: 1015-1024.

Tang JYM, Busetti F, Charrois JWA, Escher BI. Which chemicals drive biological effects in wastewater and recycled water? Water Research 2014; 60: 289-299.

Tang JYM, Escher BI. Realistic environmental mixtures of micropollutants in wastewater, recycled water and surface water: herbicides dominate the mixture toxicity towards algae. Environmental Toxicology and Chemistry 2014; 33: 1427-1436.

Tang JYM, McCarty S, Glenn E, Neale PA, Warne MS, Escher BI. Mixture effects of organic micropollutants present in water: towards the development of effect-based water quality trigger values for baseline toxicity. Water Research 2013; 47: 3300-3314.

Tousova Z, Oswald P, Slobodnik J, Blaha L, Muz M, Hu M, et al. European demonstration program on the effect-based and chemical identification and monitoring of organic pollutants in European surface waters. Science of the Total Environment 2017; 601: 1849-1868.

Tyler CR, Spary C, Gibson R, Santos EM, Shears J, Hill EM. Accounting for differences in estrogenic responses in rainbow trout (*Oncorhynchus mykiss : Salmonidae*) and roach (*Rutilus rutilus : Cyprinidae*) exposed to effluents from wastewater treatment works. Environmental Science & Technology 2005; 39: 2599-2607.

U.S. EPA. US EPA Interactive Chemical Safety for Sustainability (iCSS) Dashboard. http://actor.epa.gov/dashboard/. US Environmental Protection Agency, 2015.

Valitalo P, Massei R, Heiskanen I, Behnisch P, Brack W, Tindall AJ, et al. Effect-based assessment of toxicity removal during wastewater treatment. Water research 2017; 126: 153-163.

Välitalo P, Perkola N, Seiler TB, Sillanpää M, Kuckelkorn J, Mikola A, et al. Estrogenic activity in Finnish municipal wastewater effluents. Water Research 2016; 88: 740-749.

van der Burg B, van der Linden S, Man H-y, Winter R, Jonker L, van Vugt-Lussenburg B, et al. A Panel of Quantitative Calux® Reporter Gene Assays for Reliable High-Throughput Toxicity Screening of Chemicals and Complex Mixtures. In: Steinberg P, editor. High-Throughput Screening Methods in Toxicity Testing. John Wiley & Sons, 2013.

van der Burg B, Winter R, Weimer M, Berckmans P, Suzuki G, Gijsbers L, et al. Optimization and prevalidation of the *in vitro* ER alpha CALUX method to test estrogenic and antiestrogenic activity of compounds. Reproductive Toxicology 2010; 30: 73-80.

Van der Linden SC, Heringa MB, Man HY, Sonneveld E, Puijker LM, Brouwer A, et al. Detection of multiple hormonal activities in wastewater effluents and surface water, using a panel of steroid receptor CALUX bioassays. Environmental Science & Technology 2008; 42: 5814-5820.

van der Linden SC, von Bergh ARM, van Vught-Lussenburg BMA, Jonker LRA, Teunis M, Krul CAM, et al. Development of a panel of high-throughput reporter-gene assays to detect genotoxicity and oxidative stress. Mutation Research-Genetic Toxicology and Environmental Mutagenesis 2014; 760: 23-32.

van der Oost R, Sileno G, Suarez-Munoz M, Nguyen MT, Besselink H, Brouwer A. SIMONI (smart integrated monitoring) as a novel bioanalytical strategy for water quality assessment: Part I–model design and effect-based trigger values. Environmental Toxicology and Chemistry 2017; 36: 2385-2399.

Walter H, Consolaro F, Gramatica P, Scholze M, Altenburger R. Mixture toxicity of priority pollutants at no observed effect concentrations (NOECs). Ecotoxicology 2002; 11: 299-310.

Warne MSJ, Hawker DW. The number of components in a mixture determines whether synergistic and antagonistic or additive toxicity predominate - the funnel hypothesis. Ecotoxicology and Environmental Safety 1995; 31: 23-28.

Wegner S, Browne P, Dix D. Identifying reference chemicals for thyroid bioactivity screening. Reproductive Toxicology 2016; 65: 402-413.

Wernersson A-S, Carere M, Maggi C, Tusil P, Soldan P, James A, et al. The European technical report on aquatic effect-based monitoring tools under the water framework directive. Environmental Sciences Europe 2015; 27: 1-11.

WHO. Guidelines for drinking-water quality 4th ed. World Health Organisation, Geneva, Switzerland, 2011.

Williams M, Woods M, Kumar A, Ying GG, Shareef A, Karkkainen M. Endocrine disrupting chemicals in the Australian riverine environment: A pilot study on estrogenic compounds. CSIRO and Land & Water Australia, 2007.

Wilson VS, Bobseine K, Gray LE, Jr. Development and characterization of a cell line that stably expresses an estrogen-responsive luciferase reporter for the detection of estrogen receptor agonist and antagonists. Toxicological Sciences 2004; 81: 69-77.

Wilson VS, Bobseine K, Lambright CR, Gray LE. A novel cell line, MDA-kb2, that stably expresses an androgen- and glucocorticoid-responsive reporter for the detection of hormone receptor agonists and antagonists. Toxicological Sciences 2002; 66: 69-81.

1861 Table 1 Overview of bioassays including proposed EBT and effect threshold (eq. 13). Detailed information is given in Appendix A,
1862 Table A3)
1863

Bio-assay cate-gory[a]	Assay name	Measured endpoint or molecular target	Method reference	Expres-sion of effect concen-tration	Reference chemical	EC	Unit of EC and EBT	Abbre-viation for BEQ	Numerical value of EBT	Option	Effect thres-hold
2	H4L1.1c4 AhR assay	Activation of aryl hydrocarbon receptor (AhR)	(Brennan et al., 2015; Neale et al., 2017b)	EC_{10}	Benzo[a] pyrene	211	ng/L	B[a]P-EQ	6.36	H	33
2	PAH-CALUX	Activation of aryl hydrocarbon receptor (AhR)	(Pieterse et al., 2013)	PC_{10}	Benzo[a] pyrene	50	ng/L	B[a]P-EQ	6.21	H	8.1
2	PPARγ-GeneBLAzer	Activation of peroxisome proliferator-activated receptor (PPARγ)	(Neale et al., 2017b)	EC_{10}	Rosiglita-zone	354	µg/L	Rosiglita-zone-EQ	13.5	F	9.8

Bio-assay cate-gory[a]	Assay name	Measured endpoint or molecular target	Method reference	Expres-sion of effect concen-tration	Reference chemical	EC	Unit of EC and EBT	Abbre-viation for BEQ	Numerical value of EBT	Option	Effect thres-hold
2	PPARγ-CALUX	Activation of peroxisome proliferator-activated receptor (PPARγ)	(Gijsbers et al., 2011)	PC_{10}	Rosglita-zone	3574	µg/L	lack of sufficient single-chemical data			
2	HG5LN-hPXR	Activation of pregnane X receptor (PXR)	(Lemaire et al., 2006)	EC_{10}	Di(2-ethylhexyl)-phthalate	108	µg/L	DEHP-EQ	16.3	H	6.7
2	PXR-CALUX	Activation of pregnane X receptor (PXR)	(BDS, unpublished 2017)	PC_{10}	Di(2-ethylhexyl)-phthalate	155	µg/L	DEHP-EQ	272	H	0.6
1	MELN	Activation of estrogen receptor (ER)	(Balaguer et al., 1999)	EC_{10}	17β-Estradiol	0.68	ng/L	EEQ	0.37	G	1.9
1	ER-	Activation	(Rotroff et	EC_{10}	17β-	3.1	ng/L	EEQ	0.34	G	9.1

Bio-assay cate-gory[a]	Assay name	Measured endpoint or molecular target	Method reference	Expres-sion of effect concen-tration	Reference chemical	EC	Unit of EC and EBT	Abbre-viation for BEQ	Numerical value of EBT	Option	Effect thres-hold
	GeneBLAzer	of estrogen receptor (ER)	al., 2014)		Estradiol						
1	ERα-Luc-BG1	Activation of estrogen receptor (ER)	(Wilson et al., 2004)	EC_{10}	17β-Estradiol	2.4	ng/L	EEQ	0.62	G	3.8
1	SSTA ERα-HeLa-9903	Activation of estrogen receptor (ER)	(OECD, 2015)	PC_{10}	17β-Estradiol	2.7	ng/L	EEQ	1.01	G	2.7
1	ER-CALUX	Activation of estrogen receptor (ER)	(Sonneveld et al., 2005; van der Burg et al., 2010)	EC_{10}	17β-Estradiol	0.19	ng/L	EEQ	0.10	G	1.9
1	A-YES	Activation of estrogen receptor (ER)	(Hettwer et al., 2018; ISO/DIS19 040-2, 2017)	EC_{10}	17β-Estradiol	14	ng/L	EEQ	0.56	G	25
1	3d YES	Activation of estrogen receptor	(Routledge and Sumpter,	EC_{10}	17β-Estradiol	82	ng/L	EEQ	0.88	G	93

Bio-assay cate-gory[a]	Assay name	Measured endpoint or molecular target	Method reference	Expres-sion of effect concen-tration	Reference chemical	EC	Unit of EC and EBT	Abbre-viation for BEQ	Numerical value of EBT	Option	Effect thres-hold
		(ER)	1996)								
1	ISO-LYES (Sumpter)	Activation of estrogen receptor (ER)	(ISO/DIS19040-1, 2017)	EC_{10}	17β-Estradiol	14	ng/L	EEQ	0.97	G	14
1	ISO-LYES (McDonnell)	Activation of estrogen receptor (ER)	(ISO/DIS19040-1, 2017)	EC_{10}	17β-Estradiol	79	ng/L	EEQ	1.07	G	74
1	EASZY (Cyp19a1b-GFP)	Activation of estrogen receptor (ER)	(Brion et al., 2012)	EC_{50}	17β-Estradiol	169	ng/L	EEQ	2.15	G	78
1	REACTIV (unspiked)	Estrogenic signaling	(Spirhanzlova et al., 2016)	EC_{10}	17β-Estradiol	123	ng/L	EEQ	0.80	G	142
1	anti ER-GeneBLAzer	Antagonistic activity on the estrogen receptor (ER)	(Huang et al., 2011)	$EC_{SR0.2}$	Tamoxifen	65	not relevant because all regulated chemicals are of low potency -> no read across possible				
1	anti ERa_Luc_BG1	Antagonistic activity on the estrogen receptor (ER)	(Huang et al., 2014)	$EC_{SR0.2}$	Tamoxifen	1035	not relevant because all regulated chemicals are of low potency -> no read across possible				

Bio-assay cate-gory[a]	Assay name	Measured endpoint or molecular target	Method reference	Expres-sion of effect concen-tration	Reference chemical	EC	Unit of EC and EBT	Abbre-viation for BEQ	Numerical value of EBT	Option	Effect thres-hold
1	anti A-YES	Antagonistic activity on the estrogen receptor (ER)	(Gehrmann et al., 2016)	EC_{50}	Tamoxifen	1259	not relevant because all regulated chemicals are of low potency -> no read across possible				
1	AR-GeneBLAzer	Activation of androgen receptor (AR)	(Huang et al., 2011)	EC_{10}	Methyltrienolone (R1881)	44	not relevant because all regulated chemicals are of low potency -> no read across possible				
1	MDA-kb2	Activation of androgen receptor (AR)	(Wilson et al., 2002)	EC_{10}	5α-Dihydrotestosterone (DHT)	10	not relevant because all regulated chemicals are of low potency -> no read across possible				
1	A-YAS	Activation of androgen receptor (AR)	(Gerlach et al., 2014)	EC_{10}	5α-Dihydro-testosterone (DHT)	217	not relevant because only two chemicals were active, which are also estrogenic at lower concentration				
1	RADAR (unspiked)	Androgenic activity	(Sebillot et al., 2014)	EC_{10}	17α-methyl testosterone (17MT)	1458	not relevant because none of the tested chemicals were active				
2	anti AR-GenBLAzer	Antagonistic activity on the androgen receptor (AR)	(Huang et al., 2011)	$EC_{SR0.2}$	Flutamide	152	µg/L	Flutamide-EQ	3.28	H	46
2	anti MDA-kb2	Antagonistic activity on the	(Wilson et al., 2002)	$EC_{SR0.2}$	Flutamide	57	µg/L	Flutamide-EQ	3.46	H	17

Bio-assay cate-gory[a]	Assay name	Measured endpoint or molecular target	Method reference	Expres-sion of effect concen-tration	Reference chemical	EC	Unit of EC and EBT	Abbre-viation for BEQ	Numerical value of EBT	Option	Effect thres-hold
		androgen receptor (AR)									
2	anti AR-CALUX	Antagonistic activity on the androgen receptor (AR)	(Sonneveld et al., 2005; van der Burg et al., 2010)	$EC_{SR0.2}$	Flutamide	87	µg/L	Flutamide-EQ	14.4	H	6.1
2	anti AR RADAR (spiked)	Anti-androgenic activity	(Sebillot et al., 2014)	EC_{20}	Flutamide	22	µg/L	Flutamide-EQ	3.63	H	6.0
2	anti PR-CALUX	antagonistic activity on the progestogenic receptor (PR)	(Sonneveld et al., 2011)	$EC_{SR0.2}$	Endosulfan	64	ng/L	Endosufan-EQ	1967	H	[b]
1	GR-GeneBLAzer	Activation of glucocorticoid	(Huang et al., 2011)	EC_{10}	Dexamethasone	44	not relevant because all regulated chemicals are of low potency				

Bio-assay cate-gory[a]	Assay name	Measured endpoint or molecular target	Method reference	Expres-sion of effect concen-tration	Reference chemical	EC	Unit of EC and EBT	Abbre-viation for BEQ	Numerical value of EBT	Option	Effect thres-hold
		receptor (GR)									
2	anti GR-GeneBLAzer	Antagonistic activity of glucocorticoid receptor (GR)	(König et al., 2017)	$EC_{SR0.2}$	Mifepristone	29		not relevant because all regulated chemicals are of low potency			
1	TTR RLBA	Competition with T4 for binding to transthyretin (TTR)	(Hamers et al., 2006)	EC_{50}	Thyroxine	43	µg/L	T4-EQ	0.06	B	[b]
1	TTR FITC_T4	Binding to the thyroid hormone transport proteins	(Ren and Guo, 2012)	EC_{50}	Thyroxine	78	µg/L	T4-EQ	0.49	B	[b]
1	XETA (unspiked)	Modulation of thyroid hormone signaling	(Fini et al., 2007)	EC_{20}	Triiodothyronine (T3)	1.3	ng/L	T3-EQ	0.62	B	[b]

Bio-assay cate-gory[a]	Assay name	Measured endpoint or molecular target	Method reference	Expres-sion of effect concen-tration	Reference chemical	EC	Unit of EC and EBT	Abbre-viation for BEQ	Numerical value of EBT	Option	Effect thres-hold
1	Anti-TR-LUC-TRE	Antagonistic activity on the thyroid receptor (TR)	(Freitas et al., 2011)	$EC_{SR0.2}$	Bisphenol A	3173	µg/L	BPA-EQ	0.60	B	[b]
-	Ames Fluctuation Test (TA98)	Mutagenicity (+S9)	(Reifferscheid et al., 2012)	$EC_{IR1.5}$	Benzo[a] pyrene	196	lack of sufficient single chemical data				
-	Ames Fluctuation Test (TA100)	Mutagenicity (+S9)	(Reifferscheid et al., 2012)	$EC_{IR1.5}$	Benzo[a] pyrene	1062	lack of sufficient single chemical data				
2	AREc32	Induction of oxidative stress response	(Escher et al., 2012)	$EC_{IR1.5}$	Dichlorvos	1702	µg/L	Dichlor-vos-EQ	156	H	10.9
2	ARE GeneBLAzer	Induction of oxidative stress response	(König et al., 2017)	$EC_{IR1.5}$	Dichlorvos	3867	µg/L	Dichlor-vos-EQ	392	H	9.9
2	Nrf2-CALUX	Induction of oxidative stress response	(van der Linden et al., 2014)	$EC_{IR1.5}$	Dichlorvos	880	µg/L	Dichlor-vos-EQ	26	H	34

Bio-assay cate-gory[a]	Assay name	Measured endpoint or molecular target	Method reference	Expres-sion of effect concen-tration	Reference chemical	EC	Unit of EC and EBT	Abbre-viation for BEQ	Numerical value of EBT	Option	Effect thres-hold
2	Microtox	Inhibition of biolumines cence	(Escher et al., 2017)	EC_{50}	virtual baseline toxicant	12300	µg/L	Baseline-TEQ	1264	H	9.7
2	72h Algal growth inhibition	Growth inhibition	(OECD, 1984)	EC_{50}	Diuron	29	µg/L	DEQ	0.12	F	247
2	24h Synchrono us algae reproductio n	Growth inhibition	(Altenburg er et al., 1990)	EC_{50}	Diuron	7.7	µg/L	DEQ	0.11	F	70
1	Combined algae assay (24h-growth)	Growth inhibition	(Escher et al., 2008a)	EC_{50}	Diuron	39	µg/L	DEQ	0.13	F	302
1	Combined algae assay (2h-PSII)	Photosynth esis inhibition	(Escher et al., 2008a)	EC_{50}	Diuron	4	µg/L	DEQ	0.07	F	53
1	48h Daphnia immobilizat	Immobilizat ion	(OECD, 2004)	EC_{50}	Chlorpyrifo s	553	ng/L	Chlor-pyrifos-EQ	15	H	37

Bio-assay cate-gory[a]	Assay name	Measured endpoint or molecular target	Method reference	Expres-sion of effect concen-tration	Reference chemical	EC	Unit of EC and EBT	Abbre-viation for BEQ	Numerical value of EBT	Option	Effect thres-hold
	ion test										
2	Fish embryo toxicity	Mortality after 48h	(OECD, 2013)	LC_{50}	Bisphenol A	16368	µg/L	BPA-EQ	276	H	59
2	Fish embryo toxicity	Mortality after 96/120h	(OECD, 2013)	LC_{50}	Bisphenol A	5730	µg/L	BPA-EQ	183	H	31

1864 [a] Bioassay category 1: responsive to defined mixtures of high-potency chemicals; category 2: bioassays responsive to many if not all chemicals (undefined
1865 mixtures); [b] too preliminary to derive final effect threshold due to lack of data.

1866

1867