Global Water
Research Coalition

Effect Based Monitoring in Water Safety Planning

PROJECT REPORT

Effect Based Monitoring in Water Safety Planning

WP3.4: Effect-based trigger values for different water quality classes considering hazards for human and environment health

Lead institute: Helmholtz Centre for Environmental Research – UFZ and Griffith University

Prepared by: Peta Neale, Frederic Leusch, Beate Escher

11 November 2020

Global Water Research Coalition

Alliance House
12 Caxton Street
London SW1H 0QS
United Kingdom

Phone: +31 345 683820
www.globalwaterresearchcoalition.net

Copyright © 2020
by
Global Water Research Coalition
Publisher UFZ
ISBN 978-3-944280-14-1

Disclaimer

This study was jointly funded by PUB – Public Utilities Board, Stowa- Foundation for Applied Water Research, Water Research Australia, Water Research Commission, Water Services Association of Australia, KWR – Water B.V., SUEZ, Veolia Research and Innovation (VERI), Griffith University, and the Helmholtz-Zentrum für Umweltforschung – UFZ (the "Collaborators"). The Collaborators and the Global Water Research Coalition (GWRC) assume no responsibility for the content of the research study reported in this publication or for the opinions or statements of fact expressed in the report. The mention of trade names for commercial products does not represent or imply the approval or endorsement thereof by the Collaborators or GWRC. This report is presented solely for informational purposes

Copyright © 2020
by
Global Water Research Coalition
Publisher UFZ
ISBN 978-3-944280-14-1

GWRC in brief

In 2002, twelve leading research organisations established an international water research alliance: the Global Water Research Coalition (GWRC). GWRC is a non-profit organisation that serves as a focal point for the global collaboration for research planning and execution on water and wastewater related issues.

The Coalition focuses on water supply and wastewater issues and renewable water resources: the urban water cycle. The function of the GWRC is to leverage funding and expertise among the participating research organisations, coordinate research strategies, secure additional funding not available to single country research foundations, and actively manage a centralised approach to global issues. GWRC offers its members the opportunity to leverage resources through cooperative planning and implementation of research.

The GWRC Members are: Canadian Water Association (Canada), KWR – Water B.V. (Netherlands), PUB – Public Utilities Board (Singapore), Stowa- Foundation for Applied Water Research (Netherlands), SUEZ - CIRSEE (France), TZW - Water Technology Center (Germany), UK Water Industry Research (UK), Veolia Research and Innovation (VERI) (France), Water Research Australia (Australia), Water Research Commission (South Africa), The Water Research Foundation (USA), and the Water Services Association of Australia.

The US Environmental Protection Agency has been a formal partner of the GWRC since 2003. The Global Water Research Coalition is affiliated with the International Water Association (IWA).

GWRC members represents the interests and needs of 500 million consumers and have access to research programs with a cumulative annual budget of more than €150 million. The research portfolio of the GWRC members spans the entire urban water cycle and covers all aspects of resource management.

Executive summary

In vitro bioassays are highly sensitive and some can detect effects in clean water samples. Effect-based trigger values (EBTs) can be applied to distinguish an acceptable response in a water sample from an unacceptable response. This report reviews different approaches used to derive EBTs for the protection of both human health (e.g., drinking water and recycled water for indirect potable use) and ecosystem health (e.g., surface water and wastewater). The approaches applied range from simple translation from existing guideline values or acceptable daily intake values to approaches that determine the *in vitro* effect at the guideline value concentration and considers mixture effects. Other approaches apply multiple lines of evidence or compare experimental i*n vitro* and *in vivo* responses to derive EBTs. Some approaches are only suitable for drinking water or assays where few potent chemicals dominate the effect, while other approaches can be applied to any water type with guideline values and any assay.

The majority of EBTs have been derived for assays indicative of estrogenic activity, with surface water EBTs for six different mammalian reporter gene assays. It is recommended to use assay-specific EBTs rather than a generic EBT for an endpoint as differences in assay sensitivity and chemical potency can result in different EBTs. There are fewer EBTs available for assays indicative of xenobiotic metabolism or adaptive stress responses, with no EBTs expressed in bioanalytical equivalent concentrations for genotoxicity or mutagenicity assays. Further, there are more EBTs for surface water compared to drinking water. However, EBTs derived from guideline values for surface water are expected to also be protective for drinking water.

EBTs have been applied in several monitoring studies to determine if the bioassay response reported in drinking water or surface water poses a potential risk to human health. There is also guidance available on what steps to take if the effect in a sample exceeds its EBT. Work on EBT derivation is continuously evolving and future work should focus on deriving EBTs for assays recommended for routine water quality monitoring. This includes activation of the aryl hydrocarbon receptor, oxidative stress response and mutagenicity/genotoxicity, with EBTs for drinking water are particularly needed.

Abbreviations: AA-EQS: average environmental quality standard; ADI: acceptable daily intake; ADWG: Australian Drinking Water Guidelines; AGWR: Australian Guidelines for Water Recycling; AhR: aryl hydrocarbon receptor; B[a]P: benzo[a]pyrene; BEQ: bioanalytical equivalent concentration; CALUX: Chemical Activated LUciferase gene eXpression; DEHP: di(2-ethylhexyl)-phthalate; DHT: dihydrotestosterone; DWGV: drinking water guideline value; E1: estrone; E2: 17β-estradiol; E3: estriol; EE2: 17α-ethinylestradiol; EBT: effect-based trigger values; EC: effect concentration; EDA: effect-directed analysis; EEQ: 17β-estradiol equivalent concentration; EEQ-SSE: EEQ safe regarding steroid estrogens; EQ: equivalent concentration; EQS: environmental quality standards; EU: European Union; FSMS: Food Safety Management Systems; GR: glucocorticoid receptor; GV: guideline value; HC: hazard concentration; IR: induction ratio; LOEC: lowest observable effect concentrations; NOEC: no observable effect concentrations; NOAEL: no observed adverse effect levels; PNEC: predicted no-effect concentration; PPAR: Peroxisome proliferator-activated receptor; PSII: photosystem II; PXR: pregnane X receptor; QIVIVE: quantitative *in vitro* to *in vivo*; REF: relative enrichment factor; REP: relative effect potency; RfD: reference dose; SIMONI: Smart Integrated Monitoring; SPE: solid-phase extraction; SSD: species sensitivity distribution; TCDD: 2,3,7,8-tetrachlorodibenzodioxin; TDI: tolerable daily intake; TTC: Threshold of Toxicological Concern; TU: toxic units, WFD: Water Framework Directive; WHO: World Health Organisation; WSP: Water Safety Plan; YES: yeast estrogen screen

Table of Contents

List of Tables

List of Figures

1 Introduction

Many *in vitro* bioassays, particularly mammalian reporter gene assays, are highly sensitive by design and can detect effects in relatively clean waters, such as drinking water and recycled water, after sufficient enrichment (e.g., Jia *et al.*, 2015; Conley *et al.*, 2017; Neale *et al.*, 2020b). However, just because an effect is detected it does not necessarily mean that the chemical water quality is unacceptable. To help bioassay users differentiate between an acceptable and unacceptable response, effect-based trigger values (EBTs) have been proposed. What is acceptable or not depends on the water type and its usage and should be related to safe concentrations of regulated chemicals and be protective for the target organisms e.g., aquatic life in surface water, humans for drinking water.

The main application for bioassays is the assessment of treatment efficacy of a certain technical or natural process, to evaluate trends in effect over time and to benchmark the quality of water from different origins. Hence the effects are typically compared within a process, along a time axis or across different locations. Comparing measured effects in water samples against EBTs is a way to compare to an effect that is considered safe.

The classical approach used for chemical water quality monitoring is to compare detected chemical concentrations measured by targeted chemical analysis to chemical guideline values (GV). For example, the European Union (EU) Water Framework Directive (WFD) contains environmental quality standards (EQS) for (groups of) 45 prioritized compounds (European Parliament and European Council, 2013), while the Australian Drinking Water Guidelines (ADWG) contains GVs for over 180 organic chemicals (NHMRC/NRMMC, 2011). Chemical guidelines cannot possibly capture all chemicals potentially present in water, including contaminants of emerging concern. Consequently, the recent revision of the EU Drinking Water Directive amendments allow risk-based monitoring approaches, provided that they ensure full protection of public health (European Commission, 2018). This revision allows monitoring programs to focus on chemicals that are relevant for a specific water system and potential risks. In the Netherlands, a low concentration (0.1-1 µg/L) is used as a so called signalling value for anthropogenic substances in drinking water and drinking water sources, which is considered safe for most contaminants of emerging concern (RIVM, 2017). This is a similar threshold as the traditional 0.1 µg/L accepted level of pesticides in different water types (Hamilton *et al.*, 2003) that does not account for differences in effect potencies between pesticides but basically stems from previous analytical detection limits. As detection limits have decreased over the last decades, awareness has increased that detection alone is per se not sufficient, but that potency of pesticides has to be considered. Further, focusing on individual chemicals does not account for the mixture effects that can occur between the many chemicals present, emphasising the need to also include an integrative method of water quality analysis such as effect-based monitoring and the development of relevant EBTs.

EBTs are usually derived per bioassay and are most commonly expressed as bioanalytical equivalent concentrations (BEQ) but can also be expressed as effect concentration (EC) values or toxic units (TU). The BEQ relates the effect of a water sample to the effect of the assay reference compound, while EC values and TU, which is the inverse of the EC value, relates the effect to the sample enrichment in the assay. EBTs have been derived using a number of different approaches for both human and ecosystem health (Dingemans *et al.*, 2019). This includes EBTs for drinking water, surface water and wastewater effluent. Further, EBTs have been derived for a wide range of endpoints and cover all stages of cellular toxicity pathways (e.g., induction of xenobiotic metabolism, receptor-mediated effects, adaptive stress responses and cytotoxicity), as well as apical (whole-organism) effects in well plate-based *in vivo* assays.

EBTs allow the evaluation of the significance of bioassay results and thus are critical for the wider acceptance of *in vitro* bioassays and well plate-based *in vivo* assays by regulators and the water industry. EBTs are needed before effect-based monitoring can be implemented into Water Safety Plans (WSP) and Food Safety Management Systems (FSMS) (see WP5.1 and 5.2).

This report reviews the different approaches applied to derive EBTs and collates currently available EBTs. Further, this document outlines potential steps to take in cases where a water sample induces a bioassay response that exceeds an EBT.

2 Introduction

2.1 Overview

EBTs for the protection of both human and ecosystem health have been derived based on safe concentrations *in vivo* and translating those to an *in vitro* equivalent concentration. Safe concentrations can be taken from established GVs for single chemicals or derived from *in vivo* toxicity data. Drinking water GVs are typically derived from No Observed Adverse Effect Levels (NOAEL) or acceptable daily intake (ADI) (WHO, 2017; Baken *et al.*, 2018) and likewise surface water GVs are derived from No Observed Effect Concentrations (NOECs) or Predicted No Effect Concentrations (PNEC) (European Commission, 2011; ANZG, 2018) (Figure 1).

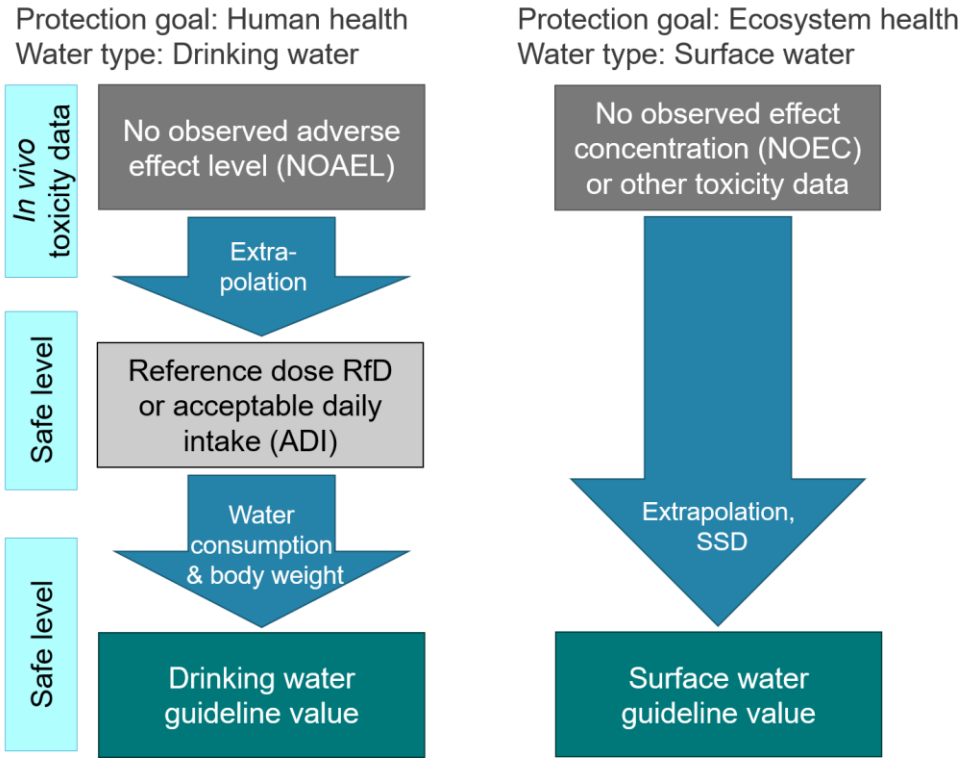

Figure 1: Guideline values for single chemicals for drinking water and surface water and how they are derived from in vivo toxicity data. SSD = species sensitivity distribution.

Several different approaches have commonly been applied to derive EBTs for drinking water and surface water (Figure 2). These will be outlined in more detail in this report but are briefly introduced here. The simplest approach is to translate the reference dose (RfD) or ADI of a prominent chemical acting through the mechanism of the bioassays via the amount of consumed water (typically 2L/d) to a concentration that is equivalent to the GV for drinking water described in Figure 1. The bioanalytical equivalent of this concentration is then the EBT (Approach 1 in Figure 2). An additional approach has been proposed to convert the concentrations of reference compounds considered safe *in vivo* to concentrations that can be detected using *in vitro* assays using differences in the toxicokinetics of different compounds to correct the EBT (Brand *et al.*, 2013) (Approach 2 in Figure 2). While *in vitro* effects can be extrapolated to *in vivo* for risk assessment by so-called (quantitative) *in vitro* to *in vivo* extrapolations (QIVIVE) (Wetmore, 2015; Yoon *et al.*, 2015), the derivation of EBTs takes the inverse route using safe concentration *in vivo* and extrapolating them to the *in vitro* situation, typically for reference compounds assuming that this is representative for all chemicals with this effect in the bioassay. Both Approaches 1 and 2 are applicable for drinking water only.

Many other studies translated existing chemical GVs directly into *in vitro* BEQ or *in vitro* bioassay effect thresholds (Approaches 3 and 4 in Figure 2). These approaches are suitable for either drinking water or surface water GVs. In the simplest way the GV

is directly translated into the BEQ for a given bioassay related to the bioassay's reference compound (Approach 3 in Figure 2). This implies that this reference compound is representative for all chemicals causing the specific effect of the bioassay. This is not necessarily the case and therefore a number of studies have determined the *in vitro* effect at the GV concentration using the different potencies of the bioactive chemicals, with some studies also accounting for mixture effects (by applying the concept of concentration addition for effects of chemicals with the same mode of action) (Approach 4 in Figure 2).

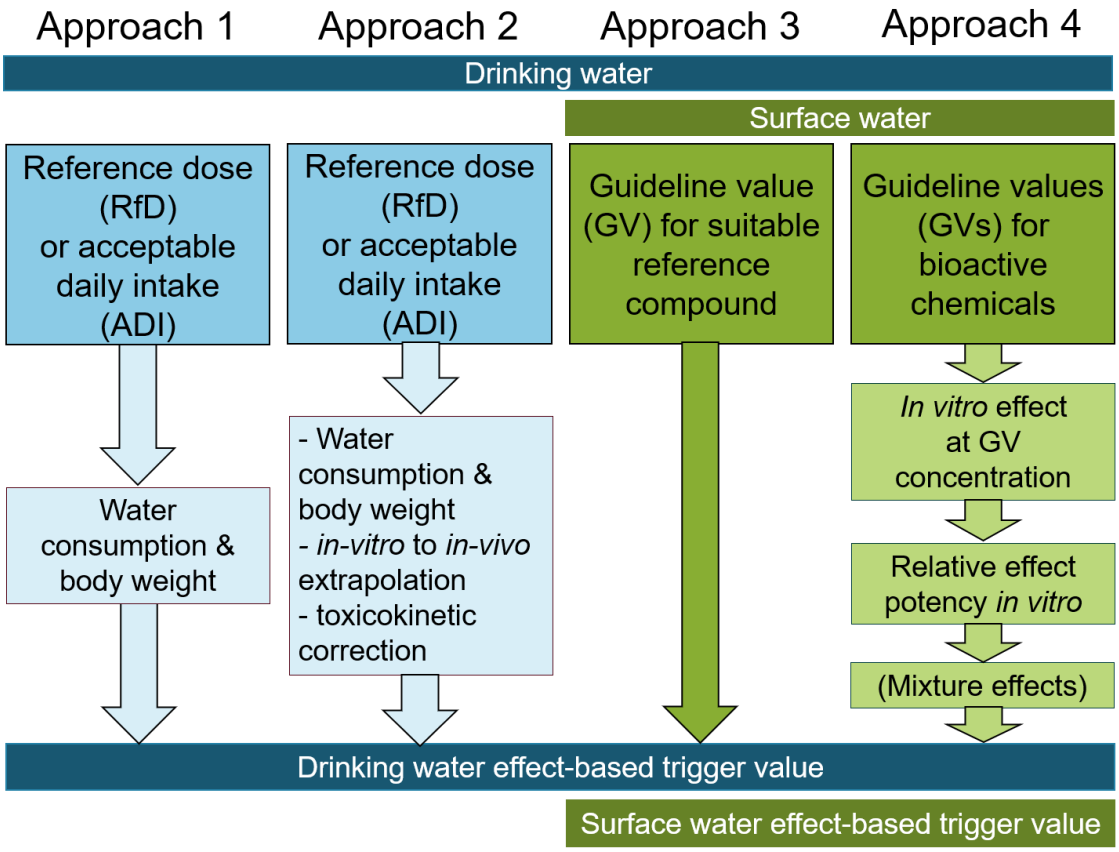

Figure 2: Derivation of effect-based trigger values (EBT) for drinking water and surface water by different approaches: (1) from an acceptable daily intake (ADI); (2) from an ADI incorporating toxicokinetic parameters; (3) from an established guideline value (GV); and (4) from an established GV incorporating relative potencies and mixture factors.
NB: Figure 2 does not cover all approaches used to derive EBTs.

It should be noted that not all studies that have derived EBTs fall into the four approaches outlined in Figure 2. For example, van der Oost *et al.* (2017) used a combination of approaches, including converting from *in vivo* toxicity data and field investigations, to derive EBTs, often following multiple lines of evidence in a more qualitative manner.

The following sections review the different studies, with a summary of the currently available EBTs provided in Table 1. All EBTs in Table 1 are expressed as BEQ, though the reference compound used for a particular endpoint can vary between studies. For example, the EBT for oxidative stress response assays is in units of dichlorvos equivalent concentrations (dichlorvos EQ) in Escher *et al.* (2018) but in curcumin equivalent concentrations (curcumin EQ) in van der Oost *et al.* (2017). The EBTs in Table 1 are sorted according to endpoint and bioassay, with different columns for drinking and recycled water (i.e., human-health relevant EBTs) and surface waters (i.e., ecological-health relevant EBTs). Bioassay specific EBTs, rather than generic EBTs for an endpoint, are provided in Table 1 as differences in assay sensitivity and chemical potency can result in different EBTs for assays indicative of the same endpoint. For example, natural estrogen estrone (E1) has a relative effect potency (REP) value of 0.02 in ERα CALUX but has a REP of 0.10 in ERα GeneBLAzer (Escher *et al.*, 2018) (Figure 3). REP relates the effect of an individual chemical to the effect of the assay reference compound (17β-estradiol (E2) for ERα CALUX and Erα GeneBLAzer), with REP values closer to 1 indicating the individual chemical has a similar potency as the reference chemical and small REP

values indicating the individual chemical is much less potent than the reference chemical. Therefore, E1 is more potent in ERα GeneBLAzer than ERα CALUX. Approaches 2 and 4 and many other EBT derivation approaches consider chemical potency, which is assay specific, so it is important to use EBTs derived for a particular assay.

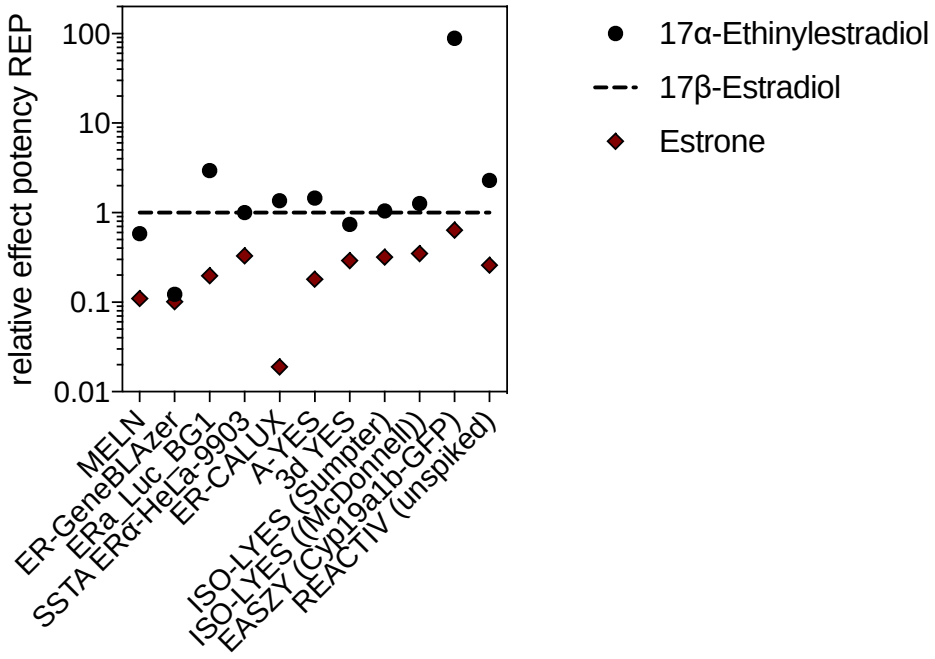

Figure 3: Relative effect potency (REP) of 17α ethinylestradiol and estrone in different assays indicative of estrogenic activity. All data available in Appendix A of Escher et al. (2018).

In the sections below we differentiate between EBTs for drinking water and environmental waters because the difference in the health target (humans consuming drinking water versus wildlife living in and consuming surface water) might cause substantial differences in safe concentrations even if the very same assays can be applied to evaluate both drinking and surface water quality. To date, only one study has derived EBTs for wastewater effluent (Jarošová *et al.*, 2014) and this study is discussed with the EBTs for surface water in Section 2.3. As an example, available EBTs for estrogenic activity are shown in Figure 4, with surface water EBTs typically lower than drinking water EBTs. This is not unexpected as EBTs derived for surface water to protect ecosystem health are often lower than EBTs derived for drinking water as GVs for surface water tend to be more protective for specifically vulnerable species than the drinking water GVs (Figure 5). For example, estrogenic chemicals cause adverse effects at very low concentrations in aquatic organisms, whereas terrestrial animals exposed primarily through dietary intake are less adversely affected (Escher *et al.*, 2018). However, differences between EBTs indicative of the same assay and the same water type still exist (e.g., EBTs for ERα CALUX for drinking water) due to differences in the derivation approaches applied.

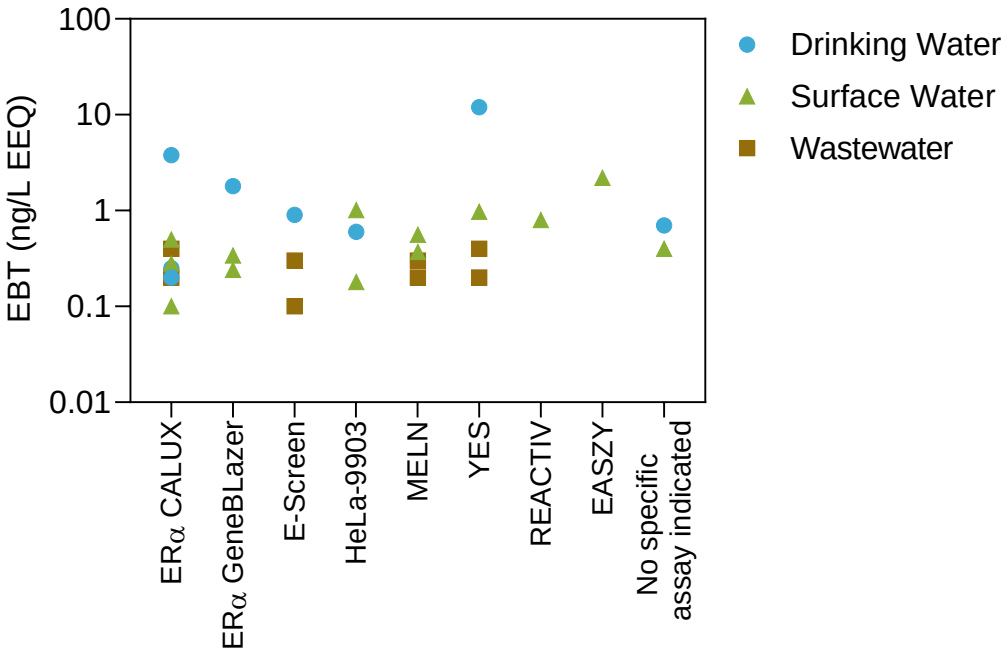

Figure 4: Overview of published effect-based trigger values (EBT) for estrogenic activity in units of 17β-estradiol equivalent concentration (EEQ) for drinking water, surface water and wastewater effluent. See Table 1 for further details.

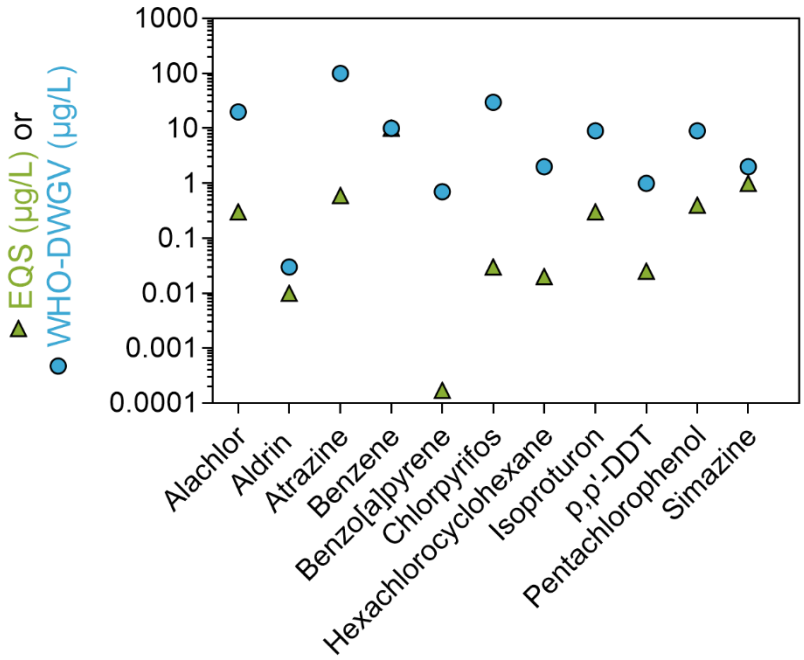

Figure 5: Comparison of common chemicals with both European Union Water Framework Directive and Swiss environmental quality standards (EQS) and WHO Drinking Water Guideline values (DWGV). Data compiled in the Supplementary Information of Escher et al. (2018)

2.2 EBT for drinking water and recycled water for indirect potable use

The simplest approach is to translate the ADI of a chemical that also serves as a potent reference compound to water concentrations in a similar fashion as World Health Organisation (WHO) drinking water guidelines are derived (WHO, 2017). Effectively, the ADI of a reference compound is converted to a GV by considering body weight (60-70 kg), water consumption (2 L) and a certain percentage allocation for drinking water (10-20%) (Approach 1 in Figure 2). Genthe *et al.* (2009) derived an EBT for estrogenic activity from the ADI with this method.

Brand *et al.* (2013) also focused on drinking water and derived EBTs using the ADI of a potent reference compound and considered oral bioavailability and the fraction unbound to plasma as indicators of adsorption and distribution to estimate the safe internal concentration (Approach 2 in Figure 2). Toxicokinetic data (i.e., bioavailability and fraction unbound to proteins) were also considered for other potent chemicals that act by the same mode of action (e.g., dexamethasone was the reference compound for the glucocorticoid receptor (GR), but other GR agonists, such as cortisol and prednisolone, may also be present in water). This approach was applied to a battery of CALUX assay and used REP data specific to these assays. The estrogenic activity EBT derived by Genthe *et al.* (2009) (0.7 ng/L), which did not account for adsorption or distribution, was around 5 times lower than the ERα CALUX EBT in Brand *et al.* (2013). Another difference was that Genthe *et al.* (2009) and Brand *et al.* (2013) started with the same ADI but assumed a different body weight (65 kg vs. 60 kg) and drinking water allocation (10% vs. 20%). The EBT of Genthe *et al.* (2009) was not assay specific but Brand *et al.* (2013) reported EBTs for four CALUX assays.

Approaches 1 and 2 applied by Genthe *et al.* (2009) and Brand *et al.* (2013), respectively, are only applicable to assays where a few chemicals dominate the effect, such as hormone receptor-mediated effects and where ADI or tolerable daily intake (TDI) values are available. Therefore, Approach 4 or an additional derivation approach is required for assays where many chemicals can contribute to the observed effect.

Béen *et al.* (submitted) evaluated and compared various established approaches used to determine EBT values with variations in the mathematical approach and the selection criteria to select the compound dataset and concluded that the most influential factor determining the value of the derived EBT is the criterion used to select data to be included. Chemical potency data were collected from the literature and provisional health-based GVs were calculated to determine a provisional health-based GV equivalent concentration. The EBT was derived from the 5th percentile of a log-normal distribution of all provisional health-based GV equivalent concentrations after removing low potency chemicals. Sufficient single chemical data was obtained from the literature to derive EBTs for the ERα CALUX, anti-AR CALUX and the AR CALUX. It was found that EBTs for PR CALUX, GR CALUX and PAH CALUX can be considered, at most, preliminary, and more single chemical effect data was also needed to derive even a preliminary EBT for Nrf2 CALUX and p53 CALUX. They developed an uncertainty analysis framework using ToxCast data from analogue bioassays (with similar target as the bioassay evaluated) to determine the protective power of the derived EBTs and the chance that potentially harmful substances might not be detected (Béen *et al.*, submitted).

Escher et al. developed an approach to address mixture effects in drinking water and recycled water for indirect potable use (Escher *et al.*, 2013; Tang *et al.*, 2013; Escher *et al.*, 2015). They tried to find as many bioactive chemicals as possible for existing sets of drinking water and recycled water guidelines. Their approach (Approach 4 in Figure 2) can be applied to any water type including surface water, provided chemical GVs are available, and is discussed in more detail in Section 2.4 below. The State Water Resources Control Board of the State of California (2019) proposed monitoring trigger values of 3.5 ng/L EEQ in ERα assays, and 0.5 ng/L TCCD-EQ for AhR assays for recycled water.

2.3 EBT for surface water and wastewater

EBTs for surface water and wastewater need to be protective for ecosystem health and are therefore derived from safe concentrations for different wildlife species. After translation to an *in vitro* bioassay the cell line applied does not necessarily need to be of the origin of an aquatic species because the *in vitro* effects are used as bioanalytical measures and not as ecological effect measures.

The simplest approach is to directly translate from a GV to the EBT (Approach 3 in Figure 2). For example, Kunz *et al.* (2015) proposed to use the E2 annual average environmental quality standard (AA-EQS) of 0.4 ng/L as the EBT to assess whether the risk was tolerable or intolerable for *in vitro* assays indicative of estrogenic activity. The E2 AA-EQS was selected rather than the

E1 or 17α-ethinylestradiol (EE2) AA-EQS values because E2 is commonly used to express the results of bioassay studies and because *in vitro* and *in vivo* REP values for E1 and EE2 are expressed relative to E2. Using a similar approach, Leusch *et al.* (2014) proposed a threshold of 0.1 ng/L EEQ for the E-screen assay and 0.2 µg/L diuron EQ for photosynthesis inhibition in the green algae *Chlorella vulgaris* based on Australian and New Zealand Guidelines for Fresh and Marine Water Quality.

Jarošová *et al.* (2014) derived "17β-estradiol equivalent concentration (EEQ) Safe regarding Steroid Estrogens" (EEQ-SSE) values based on the assumption that four potent estrogens, E1, E2, estriol (E3) and EE2, explained most estrogenic activity in water. EEQ-SSE was defined as the EEQ at which no adverse effects should be observed in municipal effluent based on the *in vivo* PNEC. A literature review was conducted to identify general concentrations of the four estrogens in municipal effluent, along with their REP in common *in vitro* estrogenic activity assays. The long term EEQ-SSE ranged from 0.1 to 0.4 ng/L EEQ. While based on wastewater effluent, the EEQ-SSE can be applied to surface water receiving effluent after appropriate dilution factors are considered.

In an alternative approach, Brion *et al.* (2019) derived EBTs for five *in vitro* estrogenic activity assays using experimental *in vitro* and *in vivo* results for 16 surface water and 17 wastewater extracts. The response in the four mammalian reporter gene assays and one yeast reporter gene assay were compared with the whole organism EASZY assay, with true negatives (i.e., samples with *in vitro* activity below the EBT and no response *in vivo*) and true positives (i.e., samples with *in vitro* activity above the EBT and an *in vivo* response) used to determine specificity and sensitivity, respectively. Logistic regression models were applied to determine the maximum sensitivity and specificity cut-off for each assay, and this was used to estimate the EBT. While this is based on a limited number of samples, the EBTs are within a similar range to other estrogenic *in vitro* assay (Table 1). However, this approach is only valid for endpoints with both *in vitro* and *in vivo* assays available.

In an approach that combined *in vivo* effect data with experimental data, van der Oost *et al.* (2017) applied three different methods to determine BEQ, which were used to derive EBTs for a battery of CALUX assays indicative of xenobiotic metabolism (PAH CALUX, DR CALUX, PPARγ CALUX, PXR CALUX), hormone receptor-mediated effects (ERα CALUX, Anti-AR CALUX, GR CALUX) and adaptive stress responses (Nrf2 CALUX). A list of chemicals was selected based on their available toxicity data and reported concentrations in water, with chemicals with low REP values removed by filtering to prevent them from biasing the EBT. *In vivo* toxicity data, including NOECs, lowest observable effect concentrations (LOECs), PNECs and effect concentration causing 50% effect (EC_{50}), were converted to BEQ using the REP, with an acute to chronic ratio of 10 applied to acute data. Safe BEQ, which indicated no risk to the ecosystem, were derived using the lowest BEQ and dividing by an assessment factor based on the endpoint. The 5% hazard concentration (HC5) BEQ, which indicates low risk, was derived using a species sensitivity distribution (SSD) approach with the BEQ values, with the 5th percentile HC determined. Finally, bioanalysis of surface water with good ecological status was used to determine background BEQ values. An EBT for each assay was derived based on the evaluation of the safe BEQ, HC5 BEQ and background BEQ values, with different multiplication factors applied based on expert judgement. In the case of non-specific effects (e.g., non-specific toxicity to bacteria and zooplankton), an average acute-to-chronic ratio of 10 and a safety factor of 2 for assumed 50% recovery by solid-phase extraction (SPE) or passive sampling was used to determine an EC at a relative enrichment factor (REF) of 20 or a TU of 0.05. REF takes into consideration both sample enrichment and dilution in the bioassay, so samples that have a non-specific effect after 20 times or more enrichment would be considered acceptable. A similar approach was taken for genotoxicity (Ames, umuC and p53 CALUX), with EBTs derived for non-specific endpoints reduced by an assessment factor of 10. This gave an EC REF of 200 or a TU of 0.005. To date, no studies have derived EBTs for genotoxicity in units of BEQ due to the lack of experimental single chemical data, but research is ongoing making use of newly developed databases with genotoxicity data (KWR, personal communication). One limitation of the EBT approach described in van der Oost *et al.* (2017) is that it relies on expert judgement for the selection of assessment factors and multiplication factors, which may make it difficult for others to apply this approach to other assays.

Kase *et al.* (2018) compared the proposed EBTs for *in vitro* estrogenic activity assays from Kunz *et al.* (2015), Jarošová *et al.* (2014) and van der Oost *et al.* (2017) against estrogenic activity reported in surface water and wastewater samples from Europe. The EBT proposed by Kunz *et al.* (2015) of 0.4 ng/L EEQ was found to be the most suitable based on the risk indicator score and the need for additional chemical analysis. However, this one EBT was applied to five different *in vitro* assays, with the fraction of compliant samples varying for the different assays (e.g., between 50 to 75% compliance in surface water). As discussed above, differences in assay sensitivity means that a single EBT should not be applied for all assays indicative of the same endpoint, with most studies deriving EBTs for individual assays (e.g., Escher *et al.*, 2015; Escher *et al.*, 2018; Brion *et al.*,

2019). For example, the human-health relevant EBT for the yeast estrogen screen (YES) (12 ng/L EEQ) was much higher than the other mammalian reporter gene assay human-health relevant EBTs for estrogenic activity (0.2 to 3.8 ng/L EEQ) (Table 1).

2.4 EBT considering mixtures for different water types

Directly translating from a GV (Approach 3) is likely only suitable for endpoints where very few high potency chemicals dominate the effect, such as specific hormone receptor-mediated responses (e.g., Konemann *et al.*, 2018) and photosynthesis inhibition (e.g., Bengtson Nash *et al.*, 2006). Therefore, a different approach is required for endpoints where many chemicals can contribute to the observed effect, such as adaptive stress responses and apical effects (e.g., Escher *et al.*, 2013; Neale *et al.*, 2015). Consequently, this section will focus on EBT derivation approaches that have addressed mixture effects using a read across approach (Approach 4) (Escher *et al.*, 2013; Tang *et al.*, 2013; Escher *et al.*, 2015; Escher *et al.*, 2018). While these studies either derived EBTs for drinking water and recycled water for indirect potable use or surface water, the approaches presented here can be applied to any water type, provided chemical GVs are available.

Tang *et al.* (2013) derived EBTs for bacterial toxicity based on bioluminescence inhibition in *Aliivibrio fischeri* (Microtox assay) by reading across from the Australian Drinking Water Guidelines (ADWG) and the Australian Guidelines for Water Recycling (AGWR). The EBT was derived based on the predicted mixture effect of all chemicals in the guidelines using the model of concentration addition divided by the sum molar concentration of all chemicals in the guidelines and included an extrapolation factor to account for the number of chemicals included in the derivation, model uncertainties and the acceptable fraction of chemicals present at their guideline value. This yielded an EBT-EC_{50} of 3 in units of REF for drinking water and REF 2.8 for recycled water for indirect potable reuse. This means that a drinking water extract would exceed the EBT if it induced 50% bacterial toxicity after less than 3 times enrichment. Converted to TU, the EBT would be 0.33 for drinking water and 0.36 for recycled water. The EC_{50} values measured in actual recycled water samples were much higher (i.e., less toxic) than the EBT-EC_{50} REF 2.8, and the measured EC_{50} value for river water at the inlet of a drinking water treatment plant was also higher than EBT-EC_{50} REF 3, suggesting no further action was required. A similar approach was applied by Escher *et al.* (2013) for the oxidative stress response assay. A tentative EBT of an effect concentration causing an induction ratio of 1.5 ($EC_{IR1.5}$) at REF 6 was proposed for both drinking water and recycled water. Based on the $EC_{IR1.5}$ value of potent pesticide dichlorvos, this translates to a dichlorvos equivalent concentration (dichlorvos EQ) of 284 µg/L. Both endpoints are indicative of integrative effects and thus reflect the presence of many chemicals in a sample.

Escher *et al.* (2015) used the ADWG to derive EBTs for assays indicative of receptor-mediated effects. EC values for individual chemicals with GVs were collected from the literature for each bioassay, then filtered, with EC values over an order of magnitude smaller or larger than the GV excluded. This was to prevent extremely more potent or less potent chemicals from skewing or dominating the distribution. The REP of each filtered chemical was calculated, which normalised the potency of the chemical to the assay reference chemicals, then the GV was converted to a BEQ by multiplying the REP and the guideline concentration. The EBT was derived from the 5th percentile of a cumulative distribution of the BEQ values. The 5th percentile was selected to be protective for the majority of chemicals, while still accounting for mixture effects and also aligns with the HC5 used by van der Oost *et al.* (2017). This process was repeated for each bioassay, with 11 of the 18 assays having sufficient data to derive a preliminary EBT. The preliminary EBTs were able to differentiate between recycled water for indirect potable reuse and secondary treated effluent.

Following on from this approach, Escher *et al.* (2018) used current and proposed AA-EQS values to derive EBTs for 48 *in vitro* and *in vivo* assays, with sufficient data available to derive 32 preliminary EBT. Similar to Escher *et al.* (2015), single chemical data was collected from the literature to translate the EQS into a BEQ value. However, rather than using one algorithm for all assays, different approaches were used for different classes of bioassays. Bioassays were grouped into two categories: category 1 assays where few potent chemicals explain most of the effect (e.g., receptor-mediated effects) and category 2 assays where many low potency chemicals have a response (e.g., adaptive stress response, non-specific effects). For category 1 assays, the EBT was derived based on the average BEQ of all chemicals, with a filtering step applied if low potency chemicals were included to prevent these chemicals reducing the EBT to unrealistically low levels. In the case of estrogenic activity, where the mixture composition of potent estrogens often had a similar pattern in environmental samples, an exposure-corrected approach was applied. BEQ was multiplied by the fraction of potent estrogens commonly found in wastewater and surface water (e.g., 11% E2, 9% EE2, 80% E1 (Kase *et al.*, 2018)). In the case of category 2 assays, it was necessary to include a mixture

factor to account for the many chemicals present with a mixture factor of 100 set for assays indicative of xenobiotic metabolism and 1,000 for adaptive stress responses. The mixture factor was multiplied by the average BEQ. The mixture factor values were based on experience with the fraction of effect explained in a particular assay by known chemicals using iceberg modelling. For example, often less than 0.1% of the effect was explained in the oxidative stress response assay (e.g., Escher *et al.*, 2013; Neale *et al.*, 2017), hence a mixture factor of 1,000. The proposed EBTs were compared with available literature data, with most EBTs able to differentiate between surface water and wastewater effluent. The EBTs for estrogenic activity are the most robust, but for many of the other endpoints more work would be required, including fingerprinting the effect of more single chemicals.

It is not possible to compare EBTs expressed as BEQ values for different endpoints as the reference compound is different. Therefore, the EBTs in Escher *et al.* (2018) were converted to an effect threshold using the EBT-BEQ value and the EC value of the corresponding assay reference compound. This way the EBT-BEQ were expressed as an effect threshold EC value in units of REF. For the apical effects in well plate-based *in vivo* assays the effect threshold EC_{10} ranged from REF 10 to 37 for bacterial toxicity, daphnia immobilization and fish embryo toxicity (Escher *et al.*, 2018). This is consistent with the effect threshold EC_{50} REF 20 proposed by van der Oost *et al.* (2017).

Ma *et al.* (2019) applied the surface water EBTs from Escher *et al.* (2018) to evaluate the compliance of advanced treated wastewater. Rather than using the published EBTs, they recalculated the effect threshold using their own reference compound EC value and the published EBTs in units of BEQ for inhibition of bioluminescence, photosynthesis inhibition with *Chlorella vulgaris* and YES. As explained above, an EBT from Escher *et al.* (2018) was calculated by converting EQS values to BEQ using the REP values based on the assay reference compound. Therefore, it does not make sense to re-calculate the effect threshold using different reference compound data. To do this correctly, the BEQ values would need to be updated using the new assay reference compound data. Ma *et al.* (2019) also estimated an EBT for the umuC assay for genotoxicity of $EC_{IR1.5}$ REF 28. This was determined using the approach discussed in van der Oost *et al.* (2017) using the EBT for non-specific toxicity and an assessment factor of 10.

A limitation of the read across approach is that the EBT will be dictated by the availability of single chemical data for each assay, as well as the quality and availability of the chemical GVs. For example, the fit of the cumulative distribution in Escher *et al.* (2015) was affected by the number of BEQ values for each assay. However, more fingerprinting of individual chemicals with GVs can improve the EBT. For example, the EBT for the AhR CALUX assay in Escher *et al.* (2018) was derived using four experimental EC values, but new single chemical effect data for nine additional chemicals was used to refine the EBT (Neale *et al.*, 2020a). The revised EBTs for AhR CALUX, PPARγ GeneBLAzer and AREc32 from Neale *et al.* (2020a), which used the same derivation approach as Escher *et al.* (2018), are provided in Table 1.

Note that several of these EBTs are at or close to detection limits of the bioassays. This is often the concern for end users as it implies that any bioactivity is of concern. This is not the case and it must be assured in all cases that the detection limits of bioassays is sufficiently low to capture the EBTs.

Further, chemicals included in current guidelines may not always match with commonly detected effects. For example, an EBT could not be derived for glucocorticoid activity in Escher *et al.* (2018) as all chemicals with an EQS were of low potency. However, experience with water quality monitoring has shown that glucocorticoid activity is often detected in wastewater effluent and surface water (e.g., Daniels *et al.*, 2018; Houtman *et al.*, 2018; Alygizakis *et al.*, 2019). An option could be to derive provisional GVs for emerging or potentially relevant chemicals based on available toxicity data (using the methods in Figure 1 to derive safe concentration) or to use generic target values using the Threshold of Toxicological Concern (TTC) approach if no toxicity data is available (Baken *et al.*, 2018).

Table 1: Summary of proposed effect-based trigger values (EBT) for both human health and ecological health expressed as bioanalytical equivalent concentrations (BEQ) that are currently available in the literature. Note: EBTs in units of BEQ are not currently available for assays indicative of reactive toxicity, so these assays are not included in Table 1.

Endpoint	Assay name	Human EBT (Drinking and recycled water for indirect potable reuse)	Ecological EBT (Surface water)
Xenobiotic metabolism			
AhR activity	AhR-cisFACTORIAL	18 µg/L Carbaryl EQ (1)	

Endpoint	Assay name	Human EBT (Drinking and recycled water for indirect potable reuse)	Ecological EBT (Surface water)
	PAH CALUX		150 ng/L B[a]P EQ [2] 6.2 ng/L B[a]P EQ [3] 62.1 ng/L B[a]P EQ [4]
	DR CALUX		0.05 ng/L TCDD EQ [2]
	H4L1.1c4 AhR assay		6.4 ng/L B[a]P EQ [3] 4.3 ng/L B[a]P EQ [5]
PPARγ activity	PPARγ CALUX		10 ng/L Rosiglitazone EQ [2]
	PPARγ-GeneBLAzer		36 ng/L Rosiglitazone EQ [3] 19 ng/L Rosiglitazone EQ [5]
PXR activity	PXR-cisFACTORIAL	59 µg/L Metolachlor EQ [1]	
	PXR CALUX		3.0 µg/L Nicardipine EQ [2] 272 µg/L DEHP EQ [3] corresponding to 54 µg/L Nicardipine EQ 5.4 µg/L Nicardipine EQ [4]
	HG5LN-hPXR		16 µg/L DEHP EQ [3]
Receptor-mediated effects			
Estrogenic activity	-#	0.7 ng/L EEQ [6]	0.4 ng/L EEQ [9]
	ERα CALUX	0.2 ng/L EEQ [1] 3.8 ng/L EEQ [7] 0.25 ng/L EEQ [8]	0.5 ng/L EEQ [2] 0.10 ng/L EEQ [3] 0.28 ng/L EEQ [10] 0.2 – 0.4 ng/L EEQ# [11]
	ERα GeneBLAzer	1.8 ng/L EEQ [1]	0.34 ng/L EEQ [3] 0.24 ng/L EEQ [10]
	E-SCREEN	0.9 ng/L EEQ [1]	0.1 – 0.3 ng/L EEQ# [11]
	YES	12 ng/L EEQ [1]	0.2 – 0.4 ng/L EEQ# [11]
	HeLa-9903	0.6 ng/L EEQ [1]	1.0 ng/L EEQ [3] 0.18 ng/L EEQ [10]
	MELN		0.37 ng/L EEQ [3] 0.56 ng/L EEQ [10] 0.2 – 0.3 ng/L EEQ# [11]
	MVLN		0.1 – 0.3 ng/L EEQ# [11]
	ERα-Luc-BG1		0.62 ng/L EEQ [3]
	A-YES		0.56 ng/L EEQ [3]
	3d YES		0.88 ng/L EEQ [3]
	ISO-LYES (Sumpter)		0.97 ng/L EEQ [3]
	ISO-LYES (McDonnell)		1.1 ng/L EEQ [3]
	pYES		0.5 ng/L EEQ [10]
	EASZY (Cyp19a1b-GFP)		2.2 ng/L EEQ [3]
	REACTIV (unspiked)		0.80 ng/L EEQ [3]
Androgenic activity	AR CALUX	11 ng/L DHT EQ [7] 4.5 ng/L DHT EQ [8]	
	AR GeneBLAzer	14 ng/L Testosterone EQ [1]	
Anti-androgenic activity	Anti-AR CALUX	4.8 µg/L Flutamide EQ [8]	25 µg/L Flutamide EQ [2] 14 µg/L Flutamide EQ [3]
	Anti-AR GeneBLAzer		3.3 µg/L Flutamide EQ [3]
	Anti-MDA-kb2		3.5 µg/L Flutamide EQ [3]

Endpoint	Assay name	Human EBT (Drinking and recycled water for indirect potable reuse)	Ecological EBT (Surface water)
	Anti-AR RADAR (spiked)		3.6 µg/L Flutamide EQ [3]
Glucocorticoid activity	GR CALUX	150 ng/L Dexamethasone EQ [1] 21 ng/L Dexamethasone EQ [7]	100 ng/L Dexamethasone EQ [2]
Progestagenic activity	PR CALUX	724 ng/L Levonorgestrel EQ* [7]	
Anti-progestagenic activity	Anti-PR CALUX		1967 ng/L Endosulfan EQ [3]
Thyroid activity	TTR RLBA		0.06 µg/L Thyroxine EQ [3]
Thyroid activity	TTR FITC-T4		0.49 µg/L Thyroxine EQ [3]
Thyroid activity	XETA (unspiked)		0.62 ng/L Triiodothyronine EQ [3]
Anti-thyroid activity	Anti-TR-LUC-GH3		0.60 µg/L Bisphenol A EQ [3]
Photosynthesis inhibition	Combined algae assay (2 h–PSII)	0.6 µg/L Diuron EQ [1]	0.07 µg/L Diuron EQ [3]
Acetylcholinesterase inhibition	AChE assay	26 µg/L Parathion EQ [1]	
Adaptive stress response			
Oxidative stress response	AREc32	284 µg/L Dichlorvos EQ† [12]	156 µg/L Dichlorvos EQ [3] 140 µg/L Dichlorvos EQ [5]
Oxidative stress response	Nrf2 CALUX		10 µg/L Curcumin EQ [2] 26 µg/L Dichlorvos EQ [3]
Oxidative stress response	ARE GeneBLAzer		392 µg/L Dichlorvos EQ [3]
Apical effects in well plate-based in vivo assays			
Bacterial toxicity	Microtox	4100-4392 µg/L Baseline TEQ‡ [13]	1264 µg/L Baseline TEQ [3]
Algal growth	72 h algal growth inhibition		0.12 µg/L Diuron EQ [3]
Algal growth	24 h synchronous algae reproduction		0.11 µg/L Diuron EQ [3]
Algal growth	Combined algae assay (24 h–growth)		0.13 µg/L Diuron EQ [3]
Immobilization	48 h daphnia immobilization test		15 ng/L Chlorpyrifos EQ [3]
Mortality	Fish embryo toxicity (48 h)		276 µg/L Bisphenol A EQ [3]
Mortality	Fish embryo toxicity (96 - 120 h)		183 µg/L Bisphenol A EQ [3]

*#No specific assay indicated; *Converted from Org2058 equivalent concentration to levonorgestrel equivalent concentration using REP in Brand et al. (2013); †Converted to dichlorvos equivalent concentration using dichlorvos EC value in Escher et al. (2018); ‡Converted to baseline toxic equivalent concentration (TEQ) using the virtual baseline toxicant EC value in Escher et al. (2018); #EBT calculated specifically for wastewater effluent.*

AhR: aryl hydrocarbon receptor; B[a]P: benzo[a]pyrene; DEHP: di(2-ethylhexyl)-phthalate; DHT: dihydrotestosterone; EEQ: 17β-estradiol equivalent concentration; PPARγ; peroxisome proliferator-activated receptor gamma; PSII: photosystem II; PXR: pregnane X receptor; TCDD: 2,3,7,8-tetrachlorodibenzodioxin

References for EBT: [1] (Escher et al., 2015); [2] (van der Oost et al., 2017); [3] (Escher et al., 2018); [4] (De Baat et al., 2020); [5] (Neale et al., 2020a); [6] (Genthe et al., 2009); [7] (Brand et al., 2013); [8] (Béen et al., submitted); [9] (Kunz et al., 2015); [10] (Brion et al., 2019); [11] (Jarošová et al., 2014); [12] (Escher et al., 2013); [13] (Tang et al., 2013)

3 Application of EBTs in monitoring studies

A number of studies have compared experimental bioassay results to the available EBTs to assess whether drinking water or surface water poses a potential risk to human or ecological health (e.g., Hamers *et al.*, 2018; Muller *et al.*, 2018; Alygizakis *et al.*, 2019; Kienle *et al.*, 2019; Neale *et al.*, 2020b). De Baat *et al.* (2019) compared the response in surface water passive sampler extracts from The Netherlands to the EBTs derived by van der Oost *et al.* (2017). The effect in the extracts exceeded the EBT for nine of the studied assays, with the effect at all sites exceeding the EBT for PAH CALUX. Further, Leusch *et al.* (2018) found that some treated wastewater and surface water samples from France and Spain exceeded the proposed EBTs for estrogenic activity, but all drinking water samples were below the drinking water EBTs.

As noted earlier, when comparing bioassay results with existing EBTs it is important to choose EBTs either for human or ecosystem health and use EBTs derived for the specific assay where available (e.g., ERα CALUX or ERα GeneBLAzer) rather than an endpoint (e.g., estrogenic activity) as differences in sensitivity and chemical potency between assays can result in different EBTs (Table 1). That said, the EBTs for mammalian reporter gene assays for estrogenic activity are all generally within an order of magnitude, despite the different methods used to derive them. From Table 1, the majority of EBTs are derived for receptor-mediated effects, with estrogenic activity the most common endpoint. There are fewer EBTs available for induction of xenobiotic metabolism, adaptive stress responses and apical effects, with the majority of these EBTs derived for surface water rather than drinking water. As discussed in Section 2.1, EBTs based on surface water GVs are typically lower than drinking water GVs, so EBTs for surface water should be protective for drinking water.

The need to derive bioassay-specific EBTs was also stressed by Escher *et al.* (2018), who compared the generic EBT-EEQ of 0.4 ng/L and bioassay-specific EBT-EEQs with data from field studies (Figure 6). Here untreated wastewater (WW) and stormwater was considered to be of poor water quality, which was compared with treated WW and surface water. Note that the experimental data stemmed from many different studies, hence they are not directly comparable. The generic EBT-EEQ does not appear to differentiate the different bioassays well enough but a clearer picture was obtained with bioassay-specific EBT-EEQs (Figure 6).

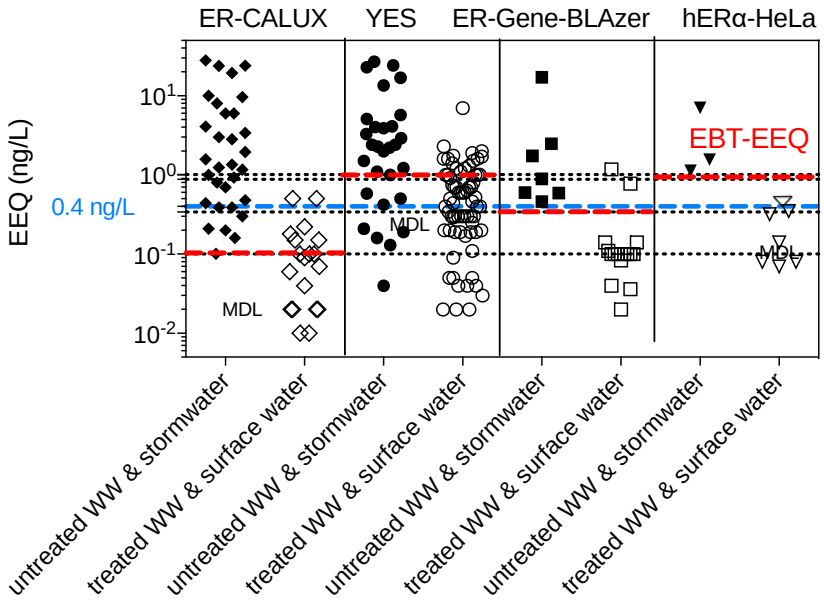

Figure 6: EBT-EEQ (or EBT-BEQ for ER) compared with measured EEQ from all data available in Appendix B of Escher et al. (2018).

4 What to do if a water extract exceeds the EBT?

To facilitate the implementation of EBTs for regulatory purposes, there is a need for a framework to determine what steps to take if the measured effect in a water sample exceeds the proposed EBT. Leusch and Snyder (2015) have proposed such a framework with three tiers of screening, targeted analysis and exploration recommended. If the measured effect exceeds the EBT during the screening tier, re-testing is recommended to 1) confirm the results and 2) determine if this is an on-going issue. A second tier of targeted analysis of known potent chemicals with GVs is then recommended if the re-tested sample still exceeds the EBT. If a detected chemical concentration exceeds the GV then operators would need to follow the GV exceedance response procedure. If the measured effect is over 10 times higher than the EBT, a number of options are suggested in a third tier including full chemical analysis of all chemicals in the relevant guidelines, effect-directed analysis (EDA) to identify the causative chemicals or bench-scale experiments to identify effective treatment methods. Non-targeted screening approaches can be used to investigate the potential presence of unknown chemicals and transformation products (Brunner *et al.*, 2020). Consultation with the local or national regulator to determine if further action is required is also recommended for any exceedances of the EBT. van der Oost *et al.* (2017) have also suggested a two-tier system consisting of hazard identification and risk analysis as part of the Smart Integrated Monitoring (SIMONI) approach. If the effect of a water sample exceeds the EBTs in the hazard identification tier, chemical analysis and EDA are among the steps recommended in the risk analysis tier. It should be noted that the approach proposed by Leusch and Snyder (2015) focused on category 1 bioassays, where detected chemicals will explain most of the effect. Therefore, chemical analysis of known potent chemicals and EDA are not likely to be useful for category 2 assay, where many low potency chemicals can contribute to the effect. Experiments to identify effective treatment methods could still be applied for category 2 assays.

Further, the NEREUS-COST Action project (nereus-cost.eu) has also proposed actions should a water sample exceed an EBT. Focusing on wastewater effluent discharged into surface water, different actions, including data quality checking, more frequent monitoring and source identification, are proposed if the measured effect exceeds the EBT by <3, <10, <100 and >100 times (Joint NORMAN and Water Europe Position Paper, 2019). For example, data quality checking, immediate re-sampling and targeted chemical analysis is recommended if the measured effect is between 3 to 10 times higher than the proposed EBT. The approaches proposed by Leusch and Snyder (2015) and the Joint NORMAN and Water Europe Position Paper (2019) are compared in Table 2.

A similared tiered approach was also proposed for ERα and AhR bioassays in the "Water quality control policy for recycled water" of the State of California, US (State Water Resources Control Board, 2019). If the ratios of BEQ to MTL (monitoring trigger levels), which are effectively equivalent to EBTs, were ≤ 0.15 in ERα assays, and <1 for AhR assays, then these assays were not further applied in monitoring. The ratio BEQ/MTL ≤ 0.15 in in ERα assays means that the response must be above the detection limit, which is typically in the range of an EEQ of 0.5 ng_{E2}/L. Hence, the MTL thresholds imposed by the State of California assure that no false negative results are implicated. If ratio BEQ/MTL was 100 < BEQ/MTL ≤ 1000, contacting the regional water board to discuss additional action and testing is required and if BEQ/MTL > 1000, then further increased testing is called for.

Table 2: Comparison of currently available guidance on steps to take if the effect in a sample, expressed as a bioanalytical equivalent concentration (BEQ), exceeds the effect-based trigger value (EBT).

	Leusch and Snyder (2015)	**NEREUS-COST Action project**
BEQ < EBT	No further action required.	No further action required.
BEQ > EBT	Re-test to confirm results and determine if exceedance is an on-going issue. If BEQ > EBT conduct targeted chemical analysis of causative chemicals. If causative chemicals exceed guideline values, then follow guideline value exceedance response procedure.	BEQ < 3 × EBT: Quality check data and monitor quarterly for 1 year and until BEQ < EBT. BEQ < 10 × EBT: Quality check data, re-sampling and conduct chemical analysis of causative chemicals. Monitor quarterly for 1 year and until BEQ < EBT.
BEQ > 10 × EBT	Options include: 1) full chemical analysis of chemicals in relevant guidelines. 2) effect-directed analysis to identify causative chemicals.	All of the above with addition source identification steps.

	3) bench-scale experiments to identify effective treatment methods. Consult with regulator to determine if further action is needed.	
BEQ > 100 × EBT		All of the above and consult with regulator to determine if further action is needed.

A comprehensive framework on what to do if the effect in a sample exceeds its EBT will be developed as part of WP5.3 and 5.4.

5 Conclusions

EBTs are essential to understand the significance of bioassay results and for the wider acceptance of effect-based monitoring by regulators and the water industry as they can distinguish between acceptable and unacceptable chemical water quality. A number of different approaches have been applied to develop EBTs, including simple translation from ADIs and GVs (Approaches 1 and 3), incorporation of chemical potency and mixtures (Approach 4), using multiple lines of evidence (van der Oost *et al.*, 2017) and comparison of *in vitro and in vivo* responses to determine maximum sensitivity and specificity cut-offs (Brion *et al.*, 2019). Some of these approaches are only valid for category 1 assays where few chemicals dominate, while other approaches also suitable for category 2 assays where many chemicals contribute to the effect. Further some approaches are only suitable for drinking water, while other approaches can be applied to different water types. Commonly, more than one bioassay is applied in a test battery, and the preference is to use EBTs derived using a similar method, if possible, to ensure consistency. From a precautionary point of view, one could also choose to include the lowest available EBT for a specific assay in the data analysis.

For many bioassays no EBT is yet available (Table 1), but this can be remediated by collecting bioassays response data and (provisional) GV for individual chemicals found in water. Despite a plethora of different approaches with differing requirements for expert opinion being applied, it is remarkable that most of the current EBTs generally end up within a log unit of each other. EBTs are increasingly applied in the literature to benchmark water quality and to evaluate treatment efficacy, giving input for practical frameworks proposed on steps to take should the effect in a sample exceed the EBT.

Bioassays indicative of activation of AhR, activation of ER and oxidative stress response were recommended for routine monitoring of surface water and wastewater in WP3.2, with surface water EBTs available for a number of assays indicative of these endpoints (Table 1). An assays indicative of mutagenicity or genotoxicity was also recommended for drinking water, in addition to the three mentioned endpoints. Besides the surface water EC REF of 200 from van der Oost *et al.* (2017), there are no EBTs for genotoxicity or mutagenicity, though work is currently underway to derive a suitable EBT. Further, drinking water EBTs are only available for one assay indicative of activation of AhR and oxidative stress response, respectively (Table 1). Further work should prioritise deriving EBTs for bioassays recommended for routine water quality monitoring.

The State of California, US (State Water Resources Control Board, 2019) is the first regulatory body that has implemented bioassays in the “Water quality control policy for recycled water” and this was only possible because clear advice was given as to acceptable effect levels, called monitoring trigger levels. Despite recommendations (Brack *et al.*, 2019), effect-based methods have not been implemented in the European Water Framework Directive as of now, even if EBTs read across from EQS of the WFD are available in research papers.

Acknowledgements

This project was funded by PUB – Public Utilities Board, Stowa- Foundation for Applied Water Research, Water Research Australia, the Water Research Commission and the Water Services Association of Australia. In-kind support was kindly provided by Veolia Research and Innovation (VERI), SUEZ, and KWR.

The authors would like to acknowledge the project team: Geertje Pronk, Stefan Kools, Milou Dingemans, Patrick Smeets (KWR), Jerome Enault, Jean-Francois Loret (SUEZ), Gaelle Meheut, Magali Dechesne, Charlotte Arnal (Veolia), Beate Escher (UFZ), Peta Neale, Fred Leusch (Griffith University), and Stehanie Rink-Pfeiffer (GWRC), as well as the project technical advisory committee: David Cunliffe, Dan Deere, Henner Hollert, Yves Levy, Alvina Mehinto, Leo Posthuman, Shane Snyder, Suzanne van Drunick, Etienne Vermerissen and Viviane Yargeau.

We thank Milou Dingemans, Jerome Enault, Jean-Francois Loret, Charlotte Arnal, Dan Deere, Magali Dechesne, Henner Hollert, Christopher Impellitteri, Yves Levi, Alvina Mehinto and Gaelle Meheut for their helpful feedback on the draft report.

References

Alygizakis, N.A., Besselink, H., Paulus, G.K., Oswald, P., Hornstra, L.M., Oswaldova, M., Medema, G., Thomaidis, N.S., Behnisch, P.A. and Slobodnik, J. (2019). Characterization of wastewater effluents in the Danube River Basin with chemical screening, *in vitro* bioassays and antibiotic resistant genes analysis. *Environment International*, **127**: 420-429.

ANZG (2018). Australian and New Zealand Guidelines for Fresh and Marine Water Quality. Australian and New Zealand Governments and Australian state and territory governments. Canberra ACT, Australia. Available at www.waterquality.gov.au/anz-guidelines.

Baken, K.A., Sjerps, R.M.A., Schriks, M. and van Wezel, A.P. (2018). Toxicological risk assessment and prioritization of drinking water relevant contaminants of emerging concern. *Environment International*, **118**: 293-303.

Béen, F., Pronk, T., Louisse, J., Houtman, C., Van der Velden-Slootweg, T., Van der Oost, R. and Dingemans, M.M.L. (submitted). Development of a framework to derive effect-based trigger values to interpret CALUX data for drinking water quality.

Bengtson Nash, S.M., Goddard, J. and Muller, J.F. (2006). Phytotoxicity of surface waters of the Thames and Brisbane River Estuaries: A combined chemical analysis and bioassay approach for the comparison of two systems. *Biosensors & Bioelectronics*, **21**(11): 2086-2093.

Brack, W., Ait Aissa, S., Backhaus, T., Dulio, V., Escher, B.I., Faust, M., Hilscherova, K., Hollender, J., Hollert, H., Muller, C., Munthe, J., Posthuma, L., Seiler, T.B., Slobodnik, J., Teodorovic, I., Tindall, A.J., Umbuzeiro, G.D., Zhang, X.W. and Altenburger, R. (2019). Effect-based methods are key. The European Collaborative Project SOLUTIONS recommends integrating effect-based methods for diagnosis and monitoring of water quality. *Environmental Sciences Europe*, **31**.

Brand, W., de Jongh, C.M., van der Linden, S.C., Mennes, W., Puijker, L.M., van Leeuwen, C.J., van Wezel, A.P., Schriks, M. and Heringa, M.B. (2013). Trigger values for investigation of hormonal activity in drinking water and its sources using CALUX bioassays. *Environment International*, **55**: 109-118.

Brion, F., De Gussem, V., Buchinger, S., Hollert, H., Carere, M., Porcher, J.M., Piccini, B., Feray, C., Dulio, V., Konemann, S., Simon, E., Werner, I., Kase, R. and Ait-Aissa, S. (2019). Monitoring estrogenic activities of waste and surface waters using a novel *in vivo* zebrafish embryonic (EASZY) assay: Comparison with *in vitro* cell-based assays and determination of effect-based trigger values. *Environment International*, **130**: 104896.

Brunner, A.M., Bertelkamp, C., Dingemans, M.M.L., Kolkman, A., Wols, B., Harmsen, D., Siegers, W., Martijn, B.J., Oorthuizen, W.A. and ter Laak, T.L. (2020). Integration of target analyses, non-target screening and effect-based monitoring to assess OMP related water quality changes in drinking water treatment. *Science of the Total Environment*, **705**.

Conley, J.M., Evans, N., Mash, H., Rosenblum, L., Schenck, K., Glassmeyer, S., Furlong, E.T., Kolpin, D.W. and Wilson, V.S. (2017). Comparison of *in vitro* estrogenic activity and estrogen concentrations in source and treated waters from 25 US drinking water treatment plants. *Science of the Total Environment*, **579**: 1610-1617.

Daniels, K.D., VanDervort, D., Wu, S.M., Leusch, F.D.L., van de Merwe, J.P., Jia, A. and Snyder, S.A. (2018). Downstream trends of *in vitro* bioassay responses in a wastewater effluent-dominated river. *Chemosphere*, **212**: 182-192.

De Baat, M.L., Kraak, M.H.S., Van der Oost, R., De Voogt, P. and Verdonschot, P.F.M. (2019). Effect-based nationwide surface water quality assessment to identify ecotoxicological risks. *Water Research*, **159**: 434-443.

De Baat, M.L., Van der Oost, R., Van der Lee, G.H., Wieringa, N., Hamers, T., Verdonschot, P.F.M., De Voogt, P. and Kraak, M.H.S. (2020). Advancements in effect-based surface water quality assessment. *Water Research*, **183**: 116017.

Dingemans, M.M., Baken, K.A., van der Oost, R., Schriks, M. and van Wezel, A.P. (2019). Risk-based approach in the revised European Union drinking water legislation: Opportunities for bioanalytical tools. *Integrated Environmental Assessment and Management*, **15**(1): 126-134.

Escher, B.I., Ait-Aissa, S., Behnisch, P.A., Brack, W., Brion, F., Brouwer, A., Buchinger, S., Crawford, S.E., Du Pasquier, D., Hamers, T., Hettwer, K., Hilscherova, K., Hollert, H., Kase, R., Kienle, C., Tindall, A.J., Tuerk, J., van der Oost, R., Vermeirssen, E. and Neale, P.A. (2018). 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. *Science of the Total Environment*, **628-629**: 748-765.

Escher, B.I., Neale, P.A. and Leusch, F.D.L. (2015). Effect-based trigger values for *in vitro* bioassays: Reading across from existing water quality guideline values. *Water Research*, **81**: 137-148.

Escher, B.I., van Daele, C., Dutt, M., Tang, J.Y.M. and Altenburger, R. (2013). Most oxidative stress response in water samples comes from unknown chemicals: The need for effect-based water quality trigger values. *Environmental Science & Technology*, **47**(13): 7002-7011.

European Commission (2011). Common Implementation Strategy for the Water Framework Directive (2000/60/EC), Guidance Document No. 27 Technical Guidance For Deriving Environmental Quality Standards.

European Commission (2018). Review of the drinking water directive, https://ec.europa.eu/environment/water/water-drink/review_en.html.

European Parliament and European Council (2013). Directive 2008/105/EC of the European Parliament and of the Council of 16 December 2008 on environmental quality standards in the field of water policy, amending and subsequently repealing

Council Directives 82/176/EEC, 83/513/EEC, 84/156/EEC, 84/491/EEC, 86/280/EEC and amending Directive 2000/60/EC of the European Parliament and of the Council.

Genthe, B., Steyn, M., Aneck-Hahn, N., Van Zijl, C. and De Jager, C. (2009). The feasibility of a health risk assessment framework to derive guidelines for oestrogen activity in treated drinking water. pp. 02.

Hamers, T., Legradi, J., Zwart, N., Smedes, F., de Weert, J., van den Brandhof, E.J., van de Meent, D. and de Zwart, D. (2018). Time-Integrative Passive sampling combined with TOxicity Profiling (TIPTOP): an effect-based strategy for cost-effective chemical water quality assessment. *Environmental Toxicology and Pharmacology*, **64**: 48-59.

Hamilton, D.J., Ambrus, Á., Dieterle, R.M., Felsot, A.S., Harris, C.A., Holland, P.T., Katayama, A., Kurihara, N., Linders, J., Unsworth, J. and Wong, S.-S. (2003). Regulatory limits for pesticide residues in water (IUPAC Technical Report). *Pure and Applied Chemistry*, **75**(8): 1123.

Houtman, C.J., ten Broek, R. and Brouwer, A. (2018). Steroid hormonal bioactivities, culprit natural and synthetic hormones and other emerging contaminants in waste water measured using bioassays and UPLC-tQ-MS. *Science of the Total Environment*, **630**: 1492-1501.

Jarošová, B., Blaha, L., Giesy, J.P. and Hilscherova, K. (2014). What level of estrogenic activity determined by *in vitro* assays in municipal waste waters can be considered as safe? *Environment International*, **64**: 98-109.

Jia, A., Escher, B.I., Leusch, F.D.L., Tang, J.Y.M., Prochazka, E., Dong, B.F., Snyder, E.M. and Snyder, S.A. (2015). *In vitro* bioassays to evaluate complex chemical mixtures in recycled water. *Water Research*, **80**: 1-11.

Joint NORMAN and Water Europe Position Paper (2019). Contaminants of Emerging Concern in Urban Wastewater, https://www.normandata.eu/sites/default/files/files/Publications/Position%20paper_CECs%20UWW_NORMAN_WE_2019_Final_20190910_public.pdf.

Kase, R., Javurkova, B., Simon, E., Swart, K., Buchinger, S., Konemann, S., Escher, B.I., Carere, M., Dulio, V., Ait-Aissa, S., Hollert, H., Valsecchi, S., Polesello, S., Behnisch, P., di Paolo, C., Olbrich, D., Sychrova, E., Gundlach, M., Schlichting, R., Leborgne, L., Clara, M., Scheffknecht, C., Marneffe, Y., Chalon, C., Tusil, P., Soldan, P., von Danwitz, B., Schwaiger, J., Palao, A.M., Bersani, F., Perceval, O., Kienle, C., Vermeirssen, E., Hilscherova, K., Reifferscheid, G. and Werner, I. (2018). Screening and risk management solutions for steroidal estrogens in surface and wastewater. *Trac-Trends in Analytical Chemistry*, **102**: 343-358.

Kienle, C., Vermeirssen, E.L.M., Schifferli, A., Singer, H., Stamm, C. and Werner, I. (2019). Effects of treated wastewater on the ecotoxicity of small streams - Unravelling the contribution of chemicals causing effects. *Plos One*, **14**(12).

Konemann, S., Kase, R., Simon, E., Swart, K., Buchinger, S., Schlusener, M., Hollert, H., Escher, B.I., Werner, I., Ait-Aissa, S., Vermeirssen, E., Dulio, V., Valsecchi, S., Polesello, S., Behnisch, P., Javurkova, B., Perceval, O., Di Paolo, C., Olbrich, D., Sychrova, E., Schlichting, R., Leborgne, L., Clara, M., Scheffknecht, C., Marneffe, Y., Chalon, C., Tusil, P., Soldan, P., von Danwitz, B., Schwaiger, J., Becares, M.I.S., Bersani, F., Hilscherova, K., Reifferscheid, G., Ternes, T. and Carere, M. (2018). Effect-based and chemical analytical methods to monitor estrogens under the European Water Framework Directive. *TRAC-Trends in Analytical Chemistry*, **102**: 225-235.

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

Leusch, F.D.L., Khan, S.J., Gagnon, M.M., Quayle, P., Trinh, T., Coleman, H., Rawson, C., Chapman, H.F., Blair, P., Nice, H. and Reitsema, T. (2014). Assessment of wastewater and recycled water quality: A comparison of lines of evidence from *in vitro*, *in vivo* and chemical analyses. *Water Research*, **50**: 420-431.

Leusch, F.D.L., Neale, P.A., Arnal, C., Aneck-Hahn, N.H., Balaguer, P., Bruchet, A., Escher, B.I., Esperanza, M., Grimaldi, M., Leroy, G., Scheurer, M., Schlichting, R., Schriks, M. and Hebert, A. (2018). Analysis of endocrine activity in drinking water, surface water and treated wastewater from six countries. *Water Research*, **139**: 10-18.

Leusch, F.D.L. and Snyder, S.A. (2015). Bioanalytical tools: half a century of application for potable reuse. *Environmental Science-Water Research & Technology*, **1**(5): 606-621.

Ma, X.Y.Y., Wang, Y.K., Dong, K., Wang, X.C.C., Zheng, K., Hao, L.W. and Ngo, H.H. (2019). The treatability of trace organic pollutants in WWTP effluent and associated biotoxicity reduction by advanced treatment processes for effluent quality improvement. *Water Research*, **159**: 423-433.

Muller, M.E., Escher, B.I., Schwientek, M., Werneburg, M., Zarfl, C. and Zwiener, C. (2018). Combining *in vitro* reporter gene bioassays with chemical analysis to assess changes in the water quality along the Ammer River, Southwestern Germany. *Environmental Sciences Europe*, **30**.

Neale, P.A., Ait-Aissa, S., Brack, W., Creusot, N., Denison, M.S., Deutschmann, B., Hilscherova, K., Hollert, H., Krauss, M., Novak, J., Schulze, T., Seiler, T.B., Serra, H., Shao, Y. and Escher, B.I. (2015). Linking *in vitro* effects and detected organic micropollutants in surface water using mixture-toxicity modeling. *Environmental Science & Technology*, **49**(24): 14614-14624.

Neale, P.A., Braun, G., Brack, W., Carmona, E., Gunold, R., König, M., Krauss, M., Liebmann, L., Liess, M., Link, M., Schäfer, R., Schlichting, R., Schreiner, V.C., Schulze, T., Vormeier, P., Weisner, O. and Escher, B.I. (2020a). Assessing the mixture effects in *in vitro* bioassays of chemicals occurring in small agricultural streams during rain events. *Environmental Science & Technology*, **54**: 8280-8290.

Neale, P.A., Feliers, C., Glauch, L., König, M., Lecarpentier, C., Schlichting, R., Thibert, S. and Escher, B.I. (2020b). Application of *in vitro* bioassays for water quality monitoring in three drinking water treatment plants using different treatment processes including biological treatment, nanofiltration and ozonation coupled with disinfection. *Environmental Science: Water Research & Technology*, **6**: 2444-2453.

Neale, P.A., Munz, N.A., Ait-Aissa, S., Altenburger, R., Brion, F., Busch, W., Escher, B.I., Hilscherova, K., Kienle, C., Novak, J., Seiler, T.B., Shao, Y., Stamm, C. and Hollender, J. (2017). Integrating chemical analysis and bioanalysis to evaluate the contribution of wastewater effluent on the micropollutant burden in small streams. *Science of the Total Environment*, **576**: 785-795.

NHMRC/NRMMC (2011). Australian Drinking Water Guidelines Paper 6 National Water Quality Management Strategy. Canberra, Australia.

RIVM (2017). Evaluation signaling value emerging contaminants drinking water policy, RIVM Rapport 2017-0091, https://www.rivm.nl/bibliotheek/rapporten/2017-0091.pdf.

State Water Resources Control Board (2019). Water quality control policy for recycled water. https://www.waterboards.ca.gov/board_decisions/adopted_orders/resolutions/2018/121118_7_final_amendment_oal.pdf (accessed 16 November 2020).

Tang, J.Y.M., McCarty, S., Glenn, E., Neale, P.A., Warne, M.S.J. and Escher, B.I. (2013). Mixture effects of organic micropollutants present in water: Towards the development of effect-based water quality trigger values for baseline toxicity. *Water Research*, **47**(10): 3300-3314.

van der Oost, R., Sileno, G., Suarez-Munoz, M., Nguyen, M.T., Besselink, H. and Brouwer, A. (2017). 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*, **36**(9): 2385-2399.

Wetmore, B.A. (2015). Quantitative *in vitro*-to-*in vivo* extrapolation in a high-throughput environment. *Toxicology*, **332**: 94-101.

WHO (2017). Guidelines for drinking-water quality, 4th edition. Geneva: World Health Organization.

Yoon, M., Blaauboer, B.J. and Clewell, H.J. (2015). Quantitative *in vitro* to *in vivo* extrapolation (QIVIVE): An essential element for *in vitro*-based risk assessment. *Toxicology*, **332**: 1-3.