Global Water
Research Coalition

Effect Based Monitoring in Water Safety Planning

PROJECT REPORT

Global Water Research Coalition

Effect Based Monitoring in Water Safety Planning

5.3 Development of protocols and user guides
5.4 Development of a decision-making tool for evaluation, selection and harmonization of candidate *in vitro* bioassays and implementation in water-related policies

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

Prepared by: Peta Neale, Beate Escher, Frederic Leusch

14 June 2022

Global Water Research Coalition

Alliance House
12 Caxton Street
London SW1H 0QS
United Kingdom

Phone: +31 345 683820
www.globalwaterresearchcoalition.net

Publisher UFZ
ISBN 978-3-944280-29-5

Global Water Research Coalition

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

Publisher UFZ
ISBN 978-3-944280-29-5

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 Network (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

Effect-based monitoring using *in vitro* bioassays and well plate-based *in vivo* assays has been recommended for water quality monitoring as they can capture the mixture effects of groups of chemicals that elicit the same mode of action. Applied as a complementary tool to targeted chemical analysis, effect-based monitoring can provide valuable input for risk analysis and risk management through Water Safety Plans (WSP). This report aims to help support the integration of effect-based monitoring into WSP frameworks by providing protocols and developing decision-making tools to assist both laboratory staff and WSP teams.

Decision-making tools for bioassay selection and sample collection and processing are provided, with a focus on applying bioassays indicative of activation of the aryl hydrocarbon receptor (AhR), activation of the estrogen receptor (ER) and oxidative stress response for wastewater and water reuse for non-potable use. An assay indicative of genotoxicity or mutagenicity is recommended in addition to the above endpoints for drinking water treatment or water reuse for potable use. Further information can be found in the WP3.2 and WP3.3 reports.

To assist laboratory staff, generic guidance for bioassays is provided, with a focus on quality assurance and quality control. Further, available International Organization for Standardization (ISO) and Organisation for Economic Co-operation and Development (OECD) guidelines for *in vitro* bioassays and well plate-based *in vivo* assays are listed. Technical guidance on bioassay data evaluation is also provided, including how to derive effect concentration (EC) values and how to calculate bioanalytical equivalent concentrations (BEQ).

To assist WSP teams, advice is provided on how effect-based monitoring can be used within WSPs (e.g., for system assessment monitoring, validation monitoring, operational monitoring and verification monitoring), including where in the catchment to customer process to apply bioassays, at what test frequencies and how to set alert level triggers. Guidance is also provided on what to do if the effect in a sample exceeds its effect-based trigger value (EBT) for both Category 1 and Category 2 assays. Finally, case studies that demonstrate how effect-based monitoring can be used to describe water quality (e.g., system assessment monitoring), verify treatment efficacy (e.g., verification monitoring) and validate control measures (e.g., validation monitoring) are provided

Abbreviations: AhR: aryl hydrocarbon receptor; AR: androgen receptor; BEQ: bioanalytical equivalent concentration; CAS: Chemical Abstract Service; CV: coefficient of variation; DBP: disinfection by-product; DF: dilution factor; DMSO: dimethyl sulfoxide; DPD: N,N-diethyl-p-phenylenediamine; DWTP: drinking water treatment plant; EBT: effect-based trigger value; EC: effect concentration; EDA: Effect-directed analysis; EEQ: 17β-estradiol equivalent concentration; EF: enrichment factor; ER: estrogen receptor; EQ: equivalent concentration; FSMS: Food Safety Management Systems; GR: glucocorticoid receptor; GV: guideline value HPLC: High-performance liquid chromatography; IC: inhibitory concentration; IR: induction ratio; ISO: International Organization for Standardization; LOD: limit of detection; LOEC: lowest observed effect concentration; LOQ: limit of quantification; LVSPE: large volume solid-phase extraction; MR: mineralocorticoid receptor; NTU: nephelometric turbidity units; OECD: Organisation for Economic Co-operation and Development; OSR: oxidative stress response; PPAR: peroxisome-proliferator activated receptor; PR: progesterone receptor; PXR: pregnane X receptor; QA: quality assurance; QC: quality control; RAR: retinoid acid receptor; REF: relative enrichment factor; RFU: relative fluorescence units; RLU: relative light units; RXR: retinoid X receptor; SOP: standard operating procedures; SPE: solid-phase extraction; SPR: suppression ratio; tBHQ: tert-butylhydroquinone; TCDD: 2,3,7,8-Tetrachlorodibenzodioxin; TR: thyroid receptor; US EPA: United States Environmental Protection Agency; WSP: Water Safety Plan; WWTP: wastewater treatment plant; XETA: Xenopus eleutheroembryonic thyroid assay; YES: Yeast Estrogen Screen

Table of Contents

List of Tables

List of Figures

Introduction

Chemical pollution in water is considered a threat to both human and ecosystem health (Malaj et al., 2014; Landrigan et al., 2018). As of February 2022, the Chemical Abstract Service (CAS) Registry contains over 193 million chemicals (https://www.cas.org/about/cas-content), with 350,000 of these currently registered for production and use on the global market (Wang et al., 2020). These registries do not include chemicals formed in the environment by biotic and abiotic transformation processes or during water treatment processes (e.g., disinfection by-products (DBPs)). Consequently, the actual number of chemicals is much larger and there are likely to be diverse mixtures of numerous chemicals present in drinking water, wastewater, and surface waters at any given time. Monitoring programs aimed at protecting human health and the environment traditionally rely on targeted chemical analysis as a means of assessing the chemical burden in water, with detected concentrations of individual identified chemicals compared against an available guideline value. However, this approach only provides information about a pre-defined list of 'targeted' chemicals, and any chemicals missing from the list are omitted, even if they are present in a water sample. Further, chemicals may be present in a water sample below the limit of detection (LOD) of the analytical instrument, but it is possible for such concentrations to still contribute to a biological effect. This is particularly relevant for estrogenic chemicals, such as 17β-estradiol and 17α-ethinylestradiol, where the analytical LOD is often higher than proposed environmental quality standards in the European Union Water Framework Directive (Kase et al., 2018). Another challenge with targeted chemical analysis is that this approach ignores the mixture effects that can occur between the many chemicals present in a sample, despite previous studies showing that mixtures of chemicals present at low concentrations can result in significant effects (Silva et al., 2002; Walter et al., 2002).

Due to the various limitations and challenges associated with chemical analysis, effect-based monitoring using *in vitro* bioassays and well plate-based *in vivo* assays has been recommended for water quality monitoring (Brack et al., 2019). *In vitro* bioassays typically use mammalian cell lines or bacterial strains and are often run in 96-well or 384-well plate format, making them high-throughput, which is essential for routine water quality monitoring. *In vitro* bioassays indicative of different stages of cellular toxicity pathways, including induction of xenobiotic metabolism, receptor-mediated effects, adaptive stress responses and cytotoxicity, have been applied to a range of water samples and are reviewed in WP3.2 *Medium-to-high throughput bioanalytical tools and decision-making tool for selection of bioassays* (Neale *et al.*, 2020c). Well plate-based *in vivo* assays are typically indicative of apical effects, such as growth, immobilization, and mortality in whole organisms, though some *in vivo* assays provide information about specific effects (e.g., photosystem II inhibition in algae). Well plate-based *in vivo* assays can capture effects from multiple toxicity pathways that result in the same apical effect (Wernersson et al., 2015). While *in vitro* bioassays and well plate-based *in vivo* assays cannot identify the individual chemicals in a water sample, they can capture the mixture effects of groups of chemicals that elicit the same mode of action. Further, effect-based monitoring provides information that is scaled to risk, with more potent chemicals having a greater effect in the assays than less potent chemicals. The strengths of effect-based monitoring strike an elegant balance with chemical analysis, since one limitation of the former includes the inability to identify specific chemicals responsible for causing a biological response. As such, it should be noted that we do not propose replacing chemical analysis with effect-based monitoring, but rather that effect-based monitoring can be applied complementary to chemical analysis as both provide different information that together can improve our understanding of the chemical burden in water.

Applied as a complementary tool to targeted chemical analysis, effect-based monitoring can provide valuable input for risk analysis and risk management through Water Safety Plans (WSP) and Food Safety Management Systems (FSMS). This is discussed in detail in WP5.1 *Defining requirements to fit effects-based monitoring with WSP and FSMS frameworks*. However, as identified in WP5.2 *Identification of needs for protocols and user guides*, the implementation of effect-based monitoring into WSP and FSMS frameworks requires the development of protocols and guidance documents for laboratory staff and frameworks for managers. This includes decision-making tools to assist with the selection of bioassays, sample collection and processing, and standard operating procedures (SOPs).

One perceived limitation of effect-based monitoring relates to the interpretation of bioassay results, specifically whether the chemical water quality is acceptable or not if an effect is detected in a bioassay. For example, some highly sensitive reporter gene *in vitro* bioassays can detect an effect in sufficiently enriched drinking water samples, despite them being

considered clean based on targeted chemical analysis. To better understand if an observed effect is acceptable or unacceptable, effect-based trigger values (EBTs) have been developed for a range of assays for both surface water and drinking water (Brand *et al.*, 2013; Escher *et al.*, 2015; van der Oost *et al.*, 2017; Escher *et al.*, 2018a). Different approaches have been used to derive EBTs and this has been discussed in detail in WP3.4 *Effect-based trigger values for different water quality classes considering hazards for human and environment health* (Neale *et al.*, 2020e). EBTs are required for the wider acceptance of effect-based monitoring by regulators and the water industry, and before bioassays can be implemented into WSPs.

This report focuses on two of the deliverables in WP5 *Water Safety Plan Protocol Development*, specifically WP5.3 *Develop protocols and user guides* and WP5.4 *Develop decision-making tool for evaluation, selection and harmonization of candidate* in vitro *bioassays and implementation in water related policies*. This report aims to help support the integration of effect-based monitoring into WSP frameworks and focuses on establishing protocols and developing decision-making tools to assist both laboratory staff and WSP teams, including managers and treatment plant operators. Specifically, this report includes three sections:

- Section 1 gives a high-level synthesis for decision-makers
- Section 2 provides technical guidance to laboratory staff including:
    - Decision-making tools for bioassay selection and sample collection and processing
    - Generic guidance for bioassays, with a focus on quality assurance (QA) and quality control (QC)
    - Available guidelines
    - Technical guidance on bioassay data evaluation

  SOPs for sample collection and processing are also provided in the Appendix.
- Section 3 provides guidance for WSP teams and includes:
    - A framework to help integrate effect-based monitoring into Water Safety Plans
    - Case studies to show how bioassays have been applied to describe water quality, assess treatment efficacy and understand treatment processes as examples of how effect-based monitoring fits into Water Safety Plan monitoring categories

Section 1 High-level synthesis for decision-makers

Our way of life requires and produces a large number of chemicals – a recent assessment estimates that there are more than 350,000 chemicals and mixtures of chemicals registered for production and use. As a result, there is increasing concern about the presence of chemicals in the aquatic environment, with chemicals detected in both source and treated drinking water. Further, treatment processes such as disinfection can result in the formation of disinfection by-products (DBPs) and other transformation products. In addition, environmental processes such as photodegradation and bacterial activity can also transform chemicals in the natural environment. This can result in a complex mixture of countless chemicals, DBPs and transformation products in source and product waters, which together elicit the total chemical burden. Current water quality guidelines are chemical-specific and are typically limited to a few hundred chemicals at the most. However, mixtures of chemicals present at concentrations below their guideline values can still have significant mixture effects (e.g., the "something from nothing" effect) (Silva et al., 2002).

The complex mixture of chemicals in water means that targeted chemical analysis alone cannot assess the total chemical burden. Effect-based monitoring uses *in vitro* bioassays and well plate-based *in vivo* assays, which are indicative of specific endpoints relevant for human and/or ecological health. Effect-based monitoring can be applied complementarily to overcome the limitations of targeted chemical analysis as effect-based monitoring can detect all chemicals in a water sample that are active in a bioassay, including both known and unknown chemicals. Additionally, bioassays can account for mixture effects that occur between the many chemicals present and can group chemicals that elicit the same mode of action. This is a major advantage of effect-based monitoring compared to the approach of individually detecting known chemicals using targeted chemical analysis.

A single bioassay cannot capture all effects in water, so a battery of at least three to four assays representative of effects commonly detected in water samples is recommended. This includes an assay indicative of activation of the estrogen receptor (ER) to detect estrogenic activity, an assay indicative of activation of the aryl hydrocarbon receptor (AhR) as an indicator of xenobiotic metabolism and an assay indicative of the oxidative stress response, which is an adaptive stress response pathway that is activated after damage by chemical stressors. The decision to focus on these endpoints is based on more than a decade of research on applying bioassays for water quality assessment, with the three endpoints responsive to a range of water types, including wastewater, surface water and, in some cases, drinking water. Further, an assay indicative of mutagenicity or genotoxicity, such as the Ames test or umuC assay, is recommended for drinking water or recycled water for potable use due to the potential formation of mutagenic and/or genotoxic DBPs.

The measured bioassay results can then be compared with available effect-based trigger values (EBT) to determine if the chemical water quality is acceptable or not. For example, the reported EBTs for estrogenic activity using the ERα CALUX assay range from 0.2 to 3.8 ng/L 17β-estradiol equivalents for drinking and recycled water and 0.1 to 0.5 ng/L 17β-estradiol equivalents for surface water. A flow chart was developed to provide advice to WSP teams on the steps to take should the measured effect exceed the EBT. This includes conducting targeted analysis of chemicals known to be active in a particular assay and checking whether their concentration exceeds available chemical guideline values.

While effect-based monitoring has mostly been applied as an investigative monitoring tool to date, it is also applied for routine water quality monitoring. For example, bioassays are applied to monitor surface water and drinking water quality in the Netherlands as part of the Key Factor Toxicity (SFT2) in the KIWK TOX project (https://www.sleutelfactortoxiciteit.nl/nl/).

As this is a newer field of science, there may be a lack of bioanalysis capacity at commercial laboratories currently. However, as the application of effect-based monitoring becomes more widespread, commercial laboratory capacity will increase to meet the demand.

Section 2

2.1 Decision-making tools for bioassay selection, sample collection and processing

Decision-making tools have been developed as part of the GWRC Effect-Based Monitoring in Water Safety Planning project to assist users in selecting appropriate bioassays for water quality monitoring and in choosing suitable sample collection and processing strategies. Further information about the decision-making tools can be found in WP3.2 and WP3.3 *Sampling strategies and sample pre-treatment options and decision-making tool for selection of sampling methods* (Neale et al., 2020d). This section repeats some of the key information already provided in WP3.2 and WP3.3 (Neale et al., 2020c; Neale et al., 2020d).

2.1.1 Decision-making tools for bioassay selection, sample collection and processing

Due to the common occurrence of diverse assortments of chemicals in environmental water extracts, a single bioassay cannot capture the full range of effects that can be induced by complex mixtures. While numerous assays are available, a practical test battery consisting of at least three or four bioassays representative of effects commonly detected in water samples and aligned with relevant steps of adverse outcome pathways is recommended. Although it remains possible that other relevant effects may be missed with only three to four bioassays, aligning the selected bioassays with results from multiplex assays that target many endpoints (e.g., Attagene trans-FACTORIAL™ and cis-FACTORIAL™ assays) can help prevent common effects from being overlooked. Further, assay selection will depend both on the context (e.g., water type, treatment type) and the purpose of the sampling campaign (e.g., to assess product quality or treatment process efficacy), which is discussed further in Section 2.1.2 In choosing an appropriate assay battery, practical guidance can be offered based on considerable prior experience evaluating different water types and sampling campaigns with a range of bioassays.

In the case of wastewater and non-potable water reuse, we recommend assays indicative of activation of the estrogen receptor (ER), oxidative stress response (OSR) and activation of the aryl hydrocarbon receptor (AhR) as a minimum test battery based on the literature review in WP3.2. These three endpoints are responsive to a range of water types, as demonstrated by both individual and multiplexed assays, and represent different stages of the cellular toxicity pathway (i.e., receptor-mediated effects, adaptive stress responses, xenobiotic metabolism). For example, Escher et al. (2014) found that AhR, ER, peroxisome proliferator-activated receptor gamma (PPARγ), pregnane X receptor (PXR) and OSR (Nrf2/ARE) were the most responsive endpoints in a range of water extracts, including wastewater effluent, river water and drinking water (Figure 1). Assays indicative of binding to PPARγ and activation of PXR were not included in the practical test battery as they are indicative of induction of xenobiotic metabolism, like activation of AhR. The recommended assays are also commonly applied assays, with 77%, 21% and 27% of reviewed studies in WP3.2 applying assays indicative of activation of ER, OSR and activation of AhR, respectively. Proposed EBTs are also available for these endpoints (Brand et al., 2013; Escher et al., 2013; Escher et al., 2015; van der Oost et al., 2017; Escher et al., 2018a). This recommendation aligns with recommendations for testing surface water quality (Brack et al., 2019), and such harmonisation is important given that rivers are receiving effluent input and may at the same time be source water for drinking water treatment plants (DWTP). Further, consistently applying the same test battery allows for the comparison of effects over time.

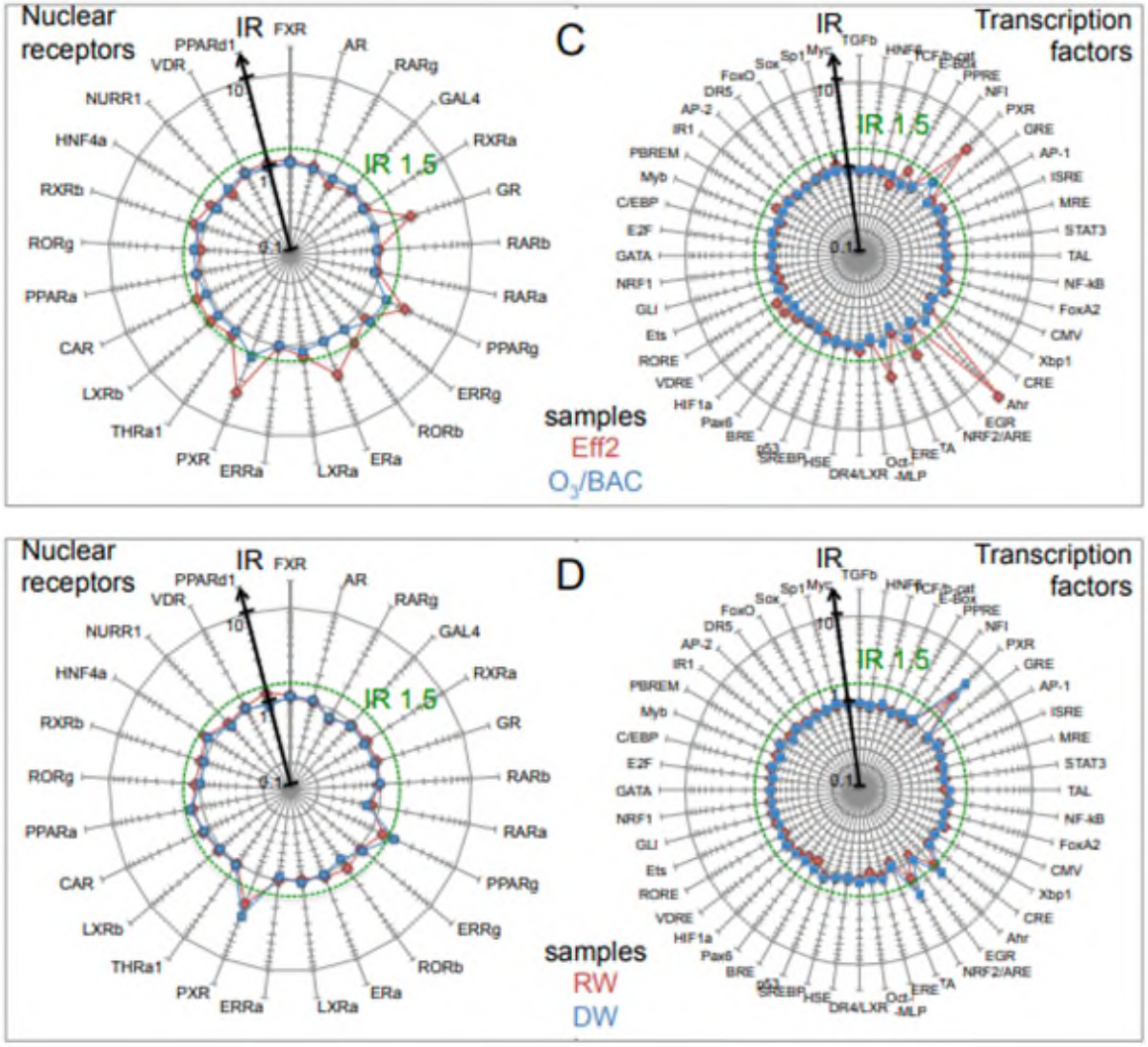

Figure 1: Screening of 25 nuclear receptors and 48 transcription factors in wastewater effluent (Eff2), after ozone/biologically activated carbon (O3/BAC), river water (RW) and drinking water (DW) in Attagene multiplex assays. Reprinted with permission from Escher et al., (2014). Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro *bioassays. Environmental Science & Technology, 48,1940-1956. DOI: 10.1021/es403899t. Copyright 2014 American Chemical Society.*

In the context of drinking water treatment or water reuse for potable use, an assay indicative of either genotoxicity or mutagenicity plus an assay on OSR are recommended to detect any DBPs formed during disinfection. It is worth noting that OSR assays can detect increased effects after drinking water disinfection, as can be seen in Figure 1 where the Nrf2/ARE response is higher in disinfected drinking water compared to river water. However, OSR cannot replace mutagenicity or genotoxicity testing (with traditional bacterial assays such as the Ames test or umuC assay) but will often also be triggered by genotoxic chemicals, not only those with direct reactive toxicity. Unfortunately, EBTs are not currently available for genotoxicity or mutagenicity assays in the scientific literature and the EBT for OSR were derived for micropollutants and would need to be adapted to DBPs. To assure that micropollutants from source water have been adequately removed, assays on activation of ER and AhR may complement the assays that target DBPs.

There are often multiple assays available for each endpoint, so guidance is required to assist users in selecting appropriate assays. Consequently, we developed a decision-making tool that groups assays into three test batteries based on assay sensitivity (Figure 2), with test battery selection depending on the sampling campaign context (water type) and purpose

(Figure 3). For example, if the purpose of a sampling campaign is to assess wastewater treatment plant (WWTP) product quality alone, a battery of low sensitivity assays, such as yeast reporter gene assays, could be applied since these are typically sufficiently sensitive to detect effects in treated effluent (test battery 1). However, a battery of high sensitivity assays, namely mammalian reporter gene assays, are recommended for understanding critical processes in WWTPs or for any purpose in a water reuse context (test battery 2). This is because yeast reporter gene assays are unlikely to be sufficiently sensitive to detect effects after advanced treatment processes. Most mammalian reporter gene assays are of similar sensitivity, so any of the reviewed assays in WP3.2 can be applied.

In contrast to wastewater, a battery of high sensitivity assays is required for all sampling purposes in a drinking water context (test battery 3). Test battery 3 adds a genotoxicity or mutagenicity assay, such as Ames or umuC, which will be particularly important if disinfected (e.g., chlorinated) water is being evaluated. It should be noted that the assay detection limit depends on the volume of sample enriched, so larger sample volumes are recommended for cleaner samples, such as drinking water or recycled water.

In all test batteries, the specific effects measurements must always be accompanied by cytotoxicity assessment. This is because cytotoxicity may cause false-negative results (e.g., masking the effect) or false-positive results (e.g., "cytotoxicity burst" phenomena (Judson et al., 2016)). Consequently, without cytotoxicity assessment it is not possible to confirm whether the detected effect is real or an artefact. Ideally, the reporter gene assay used can be duplexed with a quantitative cytotoxicity assay in the same cell line that reports effect concentrations for cytotoxicity (e.g., inhibitory concentration IC_{10}). If not, then it is imperative to include an assay with an apical endpoint, such as a bacterial bioluminescence inhibition assay.

Test Battery	Bioassays			
Battery 1	Low sensitivity ER	Oxidative Stress	AhR	
Battery 2	High sensitivity ER	Oxidative Stress	AhR	
Battery 3	High sensitivity ER	Oxidative Stress	AhR	Mutagen/ genotoxicity

Figure 2: Recommended endpoints in the different test batteries to apply for water quality monitoring (Neale et al., 2020c). For battery selection depending on the context and purpose of the sampling campaign, see Figure 3 below.

Water context	Purpose		
	Assess product quality	Assess treatment efficacy	Understand treatment processes
Wastewater treatment	Battery 1	Battery 1	Battery 2
Water reuse (non-potable)	Battery 2	Battery 2	Battery 2
Drinking water (including potable use)	Battery 3	Battery 3	Battery 3

Figure 3: Battery selection depending on sampling campaign context and purpose (Neale et al., 2020c). See Figure 2 above for a description of the different test batteries.

To assist with the selection of a suitable bioassay for each endpoint, Tables 1 to 3 summarise some key features of commonly used assays indicative of estrogenic activity, OSR and AhR activation, including the availability of an EBT, whether the assay is commonly used for water quality monitoring, and assay sensitivity. The reference compound for the OSR assay is tert-butylhydroquinone (tBHQ), but the results can also be expressed in units of dichlorvos equivalents (Neale et al., 2020a; Neale et al., 2020f) because tBHQ is not a stable chemical and is not found in water samples. Therefore, the effect concentration (EC) for dichlorvos is provided in Table 2. Tables 1 to 3 will need to be updated when new EBTs are developed for assays currently without EBTs, and when new bioassays become available. We cannot provide such a table for mutagenicity/genotoxicity assessment because most of the available assays only give yes/no responses and future work will need to focus on suggesting quantitative measures and EBTs for mutagenicity/genotoxicity assays. Either the umuC assay for genotoxicity or the Ames assay for mutagenicity could be used as both are commonly applied to water samples. In addition to the umuC and Ames assays, there are other reactive toxicity assays available, such as the γH2AX/p-H3 assay, micronucleus assay and CHO cell acute genotoxicity assay. The majority have not been widely applied for water quality assessment, with a focus instead on chemical assessment. However, the CHO cell acute genotoxicity assay has been used to both fingerprint the effect of many DBPs (Wagner and Plewa, 2017) and assess genotoxicity in drinking water and wastewater (e.g., Jeong et al., 2012; Dong et al., 2019), with the results expressed as a genotoxicity index.

Table 1: Overview of key parameters of common assays applied to evaluate estrogenic activity.

Assay	Cell line/test system	17β-estradiol EC_{10} (ng/L)	Sensitivity	Availability of EBT	Commonly applied in case studies	Experience with water quality testing
Yeast reporter gene						
YES	Yeast	10.2	Low	+	+	Applied in wastewater influent, wastewater effluent, recycled water, surface water, drinking water
Mammalian reporter gene						
ERα CALUX	U2OS	0.19	High	+	+	Applied in wastewater influent, wastewater effluent, recycled water, surface water, ground water, drinking water
ERα GeneBLAzer	HEK 293	2.7	High	+	+	Applied in wastewater influent, wastewater effluent, recycled water, surface water, ground water, riverbank filtrate drinking water
HeLa-9903	HeLa	2.1	High	+	+	Applied in wastewater effluent, recycled water, surface water, drinking water
MELN	MCF-7	0.66	High	+	+	Applied in wastewater influent, wastewater effluent, surface water, drinking water
MVLN	MCF-7	0.86	High	+	+	Applied in wastewater influent, wastewater effluent, surface water
T47D-Kbluc	T47D	0.13	High	-	+	Applied in wastewater effluent, surface water, drinking water
Cell proliferation						
E-Screen	MCF7	0.22	High	+	-*	Applied in wastewater influent, wastewater effluent, recycled water, surface water, drinking water
Whole organism						

EASZY	Embryonic zebrafish	EC_{50} 168	Low	+	-	Applied to wastewater effluent, surface water

"+" indicates availability of previously published EBT or commonly applied in case studies; "-" indicates that an EBT is not currently available or the assay is infrequently applied in case studies.

**E-Screen has been widely used in the past, but it is a seven-day test, so is much more time consuming than other assays. Consequently, it is less commonly applied now.*

Table 2: Overview of key parameters of common assays applied to evaluate the oxidative stress response (OSR).

Assay	**Cell line**	**tBHQ $EC_{IR1.5}$ (µg/L)**	**Dichlorvos $EC_{IR1.5}$ (µg/L)**	**Sensitivity**	**Availability of EBT**	**Commonly applied in case studies**	**Experience with water quality testing**
AREc32	MCF-7	219	1702	High	+	+	Applied in wastewater influent, wastewater effluent, recycled water, surface water, ground water, riverbank filtrate, drinking water
ARE GeneBLAzer	HepG2	406	3867	High	+	+	Applied in wastewater effluent and surface water
Nrf2 CALUX	U2OS	166	880*	High	+	+	Applied in wastewater effluent, recycled water, surface water, drinking water
Nrf2 reporter gene assay	HepG2	332	N/A	High	-	-	Applied in wastewater influent, wastewater effluent, surface water and drinking water
Nrf2-MDA-MB	MDA-MB-231-745	5490	N/A	Moderate	-	-	Applied in wastewater effluent, recycled water, surface water, ground water, drinking water

"+" indicates availability of previously published EBT or commonly applied in case studies; "-" indicates that an EBT is not currently available or the assay is infrequently applied in case studies.

**PC_{10} value reported*

tBHQ: tert-butylhydroquinone

Table 3: Overview of key parameters of common cell-based reporter gene assays applied to evaluate the aryl hydrocarbon receptor (AhR) activity.

Assay	Cell line	TCDD EC_{10} (ng/L)	Benzo(a)pyrene EC_{10} (ng/L)	Sensitivity	Availability of EBT	Commonly applied in case studies	Experience with water quality testing
AhR CAFLUX	H1.G1.1c3, H4.G1.1c2	0.21-0.22		High	-	+	Applied in wastewater influent, wastewater effluent, recycled water, surface water, groundwater, drinking water
AhR CALUX	H4L1.1c4	0.19	211	High	+	+	Applied in wastewater effluent, wastewater effluent, surface water, riverbank filtrate, drinking water
AhR reporter gene assay	HepG2	20		Low	-	-	Applied in wastewater influent, wastewater effluent, surface water, drinking water
H4IIE-luc	H4IIE	0.05		High	-	+	Applied in wastewater influent, wastewater effluent, recycled water, surface water, groundwater, drinking water
PAH CALUX	H4IIE		50	High	+	-	Applied to wastewater effluent*

"+" indicates availability of previously published EBT or commonly applied in case studies; "-" indicates that an EBT is not currently available or the assay was infrequently applied in case studies.

**Also applied to surface water passive sampler extracts.*

TCDD: 2,3,7,8-Tetrachlorodibenzodioxin

It should be noted that while test batteries including a minimum of three or four assays are recommended in most situations, water utilities could also apply more comprehensive test batteries. This could include any assay previously found to have a response in water extracts from a relevant context and could also include whole organism assays indicative of apical effects (e.g., algal growth inhibition assay or fish embryo toxicity assay). The selection of additional assays may be related to specific water quality concerns of the water utility. For example, a phytotoxicity assay could be included if raw drinking water is collected from a catchment impacted by agriculture.

Water utilities in developing countries may not have access to laboratories that can run mammalian reporter gene assays but will most likely have access to a microbiology laboratory. In these cases, a simple bacterial toxicity assay, such as Microtox or BLT-Screen, could be applied. Both assays are similarly sensitive, have been applied to wastewater, surface water and drinking water (though Microtox is more widely used), and EBTs are available for Microtox (Tang *et al.*, 2013; Escher *et al.*, 2018a). It should be noted that these assays only provide information about non-specific effects and should be complemented with assays indicative of specific effects, when possible, but they can be powerful as indicators of the relative chemical water quality. As such, they are particularly useful for measuring changes over time, or for comparing different water sources.

2.1.2 Sample collection and processing

To ensure that bioassay results are meaningful, it is important to select an appropriate sampling strategy and to use suitable sample pre-treatment and processing methods. The sampling strategy will depend on the purpose, objectives and sample context. If the purpose of a sampling campaign is to assess the product quality, with the objective of comparing the effect in the final water to an EBT, then only the product water needs to be collected. In contrast, both source water and product water are required if the purpose is to evaluate treatment process efficiency. Samples can also be collected after intermediate steps throughout the treatment train, such as after advanced oxidation or disinfection, if the purpose of the sampling campaign is to understand critical processes. This is discussed further in WP3.3.

The required volume of water to be collected will depend on the expected level of chemical contamination (e.g., less volume needed for wastewater influent, more volume needed for drinking water). Other decisions to be made include whether to adjust the pH after collection, how to store samples prior to extraction, and whether to filter prior to extraction. Consequently, a decision-making flow chart has been developed to guide users through some of the key decisions (Figure 4), with further information about each step provided in WP3.3. Once the final sample pre-treatment and processing methods have been selected, it is important to use the same approach for all samples that will be compared. It is not possible to truly compare changes over time or differences between sites if different sample pre-treatment and processing methods are used as this can affect the chemical mixture in the final extract. The same sample pre-treatment and processing methods should also be used for both bioanalysis and chemical analysis, when possible, to facilitate comparison of the results.

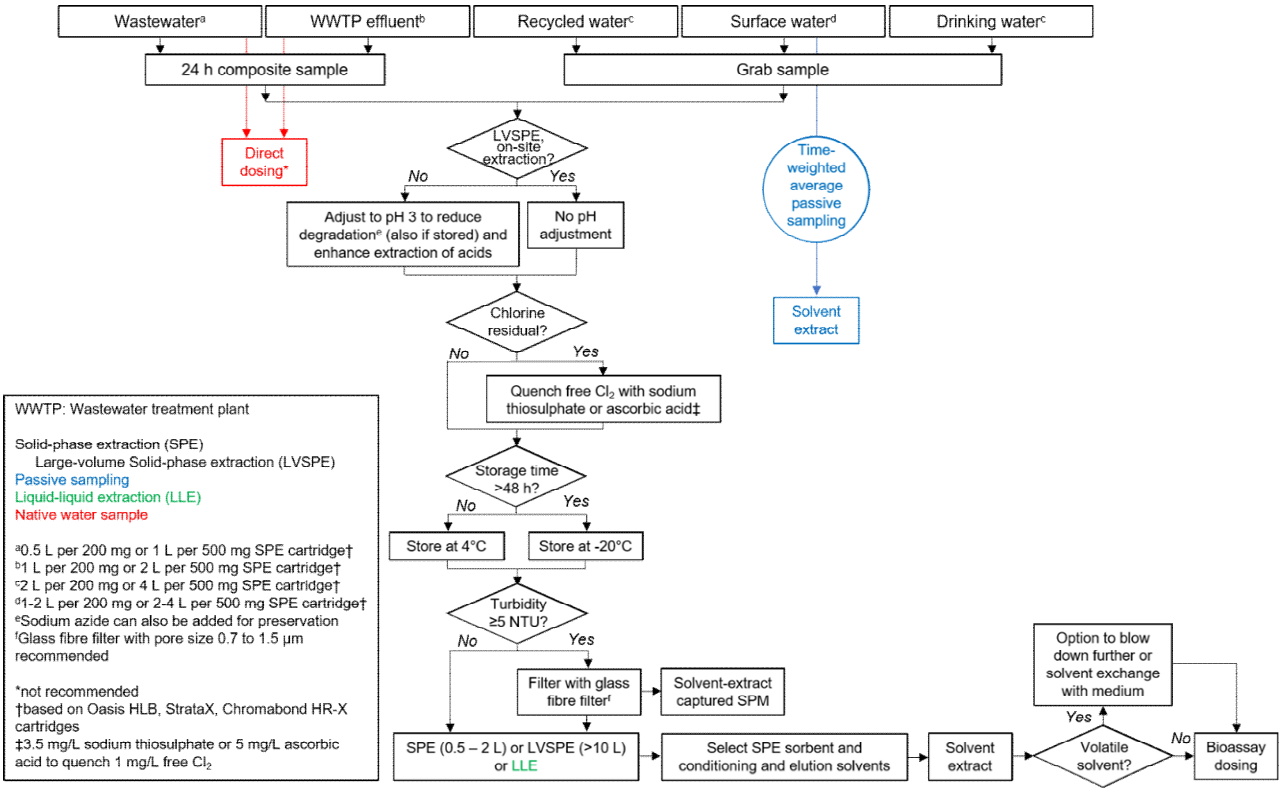

Figure 4: Sample pre-treatment and processing decision-making flow chart (Neale et al., 2020d).

2.2 Generic guidance for bioassays

This section outlines some important considerations when running *in vitro* bioassays and well plate-based *in vivo* assays, particularly around QA and QC. The focus of QA is managing quality by reducing the chance of issues arising, which is

often achieved through good record-keeping and using appropriate SOPs and can be thought of as a proactive process (OECD, 2018). QC describes the procedures in place that are used to verify the quality of the data, such as QC samples and replicates, and is a reactive process. Both will be discussed further below, with more information available in Chapter 11 of Escher et al. (2021). Due to the large number of *in vitro* bioassays and well plate-based *in vivo* assays used in water quality monitoring it is not possible to provide specific information for each, but rather this section aims to provide generic widely applicable guidance. All bioassays used for water quality monitoring should be validated to ensure they are accurate, precise, robust, and sensitive. Other important considerations for running bioassays include wearing appropriate personal protective equipment and, particularly in the case of cell-based assays, working in a sterile environment (e.g., working in a biological safety cabinet and spraying any items coming into the cabinet with 70% v/v ethanol). It is also important to run extracts at multiple concentrations using dilution series, rather than just at a single extract concentration, as this allows concentration-effect curves to be generated. This is discussed further in Section 2.4.

2.2.1 Quality assurance

An SOP that includes consumable and equipment requirements, detailed assay procedures and data analysis information is essential to ensure reproducible assay results. SOPs are often available from assay suppliers (e.g., Bio Detection Systems, Thermo Fisher Scientific) or published in the peer-reviewed literature (e.g., Neale et al. 2017). Further, Organisation for Economic Co-operation and Development (OECD) and International Organization for Standardization (ISO) guidelines are available for *in vitro* bioassays and well plate-based *in vivo* assays. This is discussed further in Section 2.3.

A well-maintained laboratory book (either hardcopy or electronic) with information including cell passage number, batch/lot numbers and any deviations from the SOP is important for good record-keeping and to ensure that all necessary information is readily available. Further, raw and analysed bioassay data should be stored safely.

Finally, instruments, such as plate readers, and equipment, such as pipettes, should be calibrated on an annual basis to ensure they are working correctly and are not a source of error.

2.2.2 Quality control

2.2.2.1 Replication

Replication is essential for any form of bioanalysis and different types of replicates are necessary for appropriate bioassay QC. Firstly, samples should be analysed in replicate, typically duplicate or triplicate, together on the same plate. This is called intra-plate replication and is used to determine the variability between wells. Variability may be due to the operator (e.g., pipetting errors) or environmental factors (e.g., variable loss of media due to evaporation during incubation). The intra-plate replication coefficient of variation (CV), which is calculated using Equation 1 where σ is the standard deviation and μ is the mean of the replicate samples, should not exceed the predetermined CV threshold. The CV threshold can vary from assay to assay but is often set at 10-15%, and samples should be re-run if the CV threshold is exceeded.

$$CV = \frac{\sigma}{\mu} \quad (1)$$

The same sample can be run on different plates if running multiple plates on the same day or on different parts of the same plate to ensure there is no temporal drift, which could be due to environmental factors such as temperature or issues with the measurement instrument. This is called intra-assay replication and the CV threshold is often set to 10-15% but can be assay specific. Intra-assay replication is not required for all samples, but at least one randomly selected sample should be used.

Each sample should always be run at least twice in independent assay runs performed on different days. This is called inter-assay replication and helps ensure there is no assay drift or bias over time. The inter-assay CV threshold can vary between assays but should be no more than 15-20%, with greater variability warranting a third run.

Finally, true replicate samples, which are water samples collected in duplicate or triplicate and independently extracted, can also be included in monitoring programs. The average and standard deviation of the independent samples can be reported, though this value will include the variability associated with sample collection and processing, as well as bioanalysis.

2.2.2.2 Quality control samples

QC samples, including a reference compound, negative control, solvent control and any blank samples, should be included in every bioassay run to ensure the assay results are within an acceptable range and the assay is sufficiently sensitive. This is because variations in cell culture can potentially cause altered behaviour of the test system, resulting in inaccurate results. Relevant QC samples are summarised in Table 4.

The reference compound is typically a potent chemical in the assay, such as 17β-estradiol in the activation of ER assays, and ideally should also be an environmentally relevant chemical (i.e., a chemical likely to be present in water samples). Common reference compounds for *in vitro* bioassays and well plate-based *in vivo* assays used for water quality monitoring are provided in WP3.2. A standard curve, which could be either a log-sigmoidal concentration-effect curve or a linear concentration-effect curve, of the reference compound should be included in every assay run. Concentration-effect curves will be discussed further in Section 2.4, which focuses on bioassay data evaluation. The standard curve should have a minimum of 8 points and include concentrations that range from 0 to 100% effect for log-sigmoidal concentration-effect curves or 0 to 30% effect or an induction ratio (IR) of 1 to 4 for linear concentration-effect curves. The standard curve is used to generate an EC value, such as the concentration causing 50% effect (EC_{50}) for log-sigmoidal concentration-effect curves or the concentration causing 10% effect (EC_{10}) or an IR of 1.5 ($EC_{IR1.5}$) for linear concentration-effect curves. The reference compound EC value can be compared between runs and over time. In addition to the reference compound, a positive control, a bioactive chemical with different physico-chemical properties to the reference compound, can also be run in the assay.

The negative control, where the test medium is added to the cells, indicates the minimum response of the test system. As water extracts in solvent are typically tested in bioassays, it is also important to confirm that the solvent itself does not induce a response in the assay. This is tested using a solvent control, where the same volume of solvent as the water extracts is added to the test medium. Ideally, the negative control and solvent control should yield the same results. If the solvent control has a greater effect than the negative control, the amount of solvent added to the assay should be reduced. Using the example of common solvents, the maximum concentration of dimethyl sulfoxide (DMSO) is typically 0.1% in cell-based reporter gene assays, while up to 1% methanol can be used in some reporter gene assays (Leusch et al., 2017). Some assays also include cell-free controls, where the medium is added to a well in the absence of cells. This gives a background reading and can be subtracted from the readings to determine background-corrected values. It is also important to test blank samples, such as field blanks and laboratory blanks, to ensure that the sampling or sample processing procedures are not contributing to the observed effect.

In addition, an inter-assay sample can be included as a QC sample. Intra-assay samples are previously investigated water extracts that have a response in the assay. The effect in the intra-assay sample can be compared with the previous results to confirm that the analysis is reliable and repeatable.

Table 4: Overview of common quality control samples used in bioanalysis.

Quality Control Sample	Definition
Reference compound	A potent chemical used to generate a standard curve. The standard curve should be included in every assay run and be benchmarked against previous assay runs
Positive control	A known active chemical that has different physico-chemical properties to the reference compound
Negative control	Used to determine the minimal response in assay media without any other influences
Solvent control	Used to verify that the solvent itself is not having an effect in the assay

Field blank	Ultrapure water sample taken into the field and exposed to the same conditions as the collected water samples (e.g., temperature). It is also processed using the same sample processing procedures as the actual water samples
Laboratory blank	Ultrapure water sample that is processed in the laboratory using the same sample processing procedures as the actual water samples
Inter-assay sample	Previously measured water extract that is known to have a response in the assay

2.2.2.3 Control charts

Control charts help to benchmark key bioassay parameters, such as the EC value over time, which can help to identify any abnormal results. Examples of control charts for ERα GeneBLAzer and AREc32 are provided in Figure 5. Control charts include the mean value of all previous runs, with upper and lower warning and control limits. The warning limits are typically the mean plus or minus two standard deviations and the control limits are typically the mean plus or minus three standard deviations. If the bioassay parameter (e.g., EC value) from a run is outside the control limit the data should not be used, with re-testing required. Bioassay parameters that fall between the warning and control limit should be further investigated. Control charts can also reveal gradual shifts in assay performance, which could be due to factors such as cell passage number or degrading reagents.

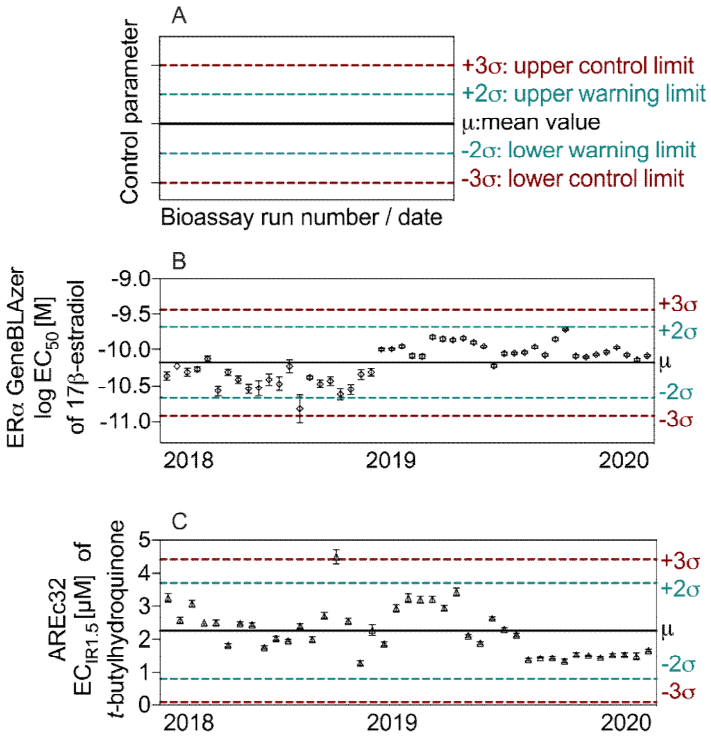

Figure 5: Makeup of a control chart showing the mean value, upper and lower warning limits (mean ± 2σ) and upper and lower control limits (mean ± 3σ); B. example control chart for the log EC_{50} of 17β-estradiol in ERα GeneBLAzer; C. example control chart for the $EC_{IR1.5}$ of t-butylhydroquinone in AREc32 (Reproduced from Escher, B.I., Neale, P.A., Leusch, F.D.L. (2021) Bioanalytical Tools in Water Quality Assessment - Second Edition, IWA Publishing, London)

2.2.2.4 Assay sensitivity

The sensitivity of an assay determines its LOD and limit of quantification (LOQ). The LOD is typically calculated as 3× the standard deviation of the effect of negative control and LOQ is calculated as 10× the standard deviation of the effect of the negative control. At low effect cut-offs, such as EC_{10}, it is important to ensure that the bioassays are sufficiently sensitive, that is, LOD should be lower than 10% effect. This should not be an issue for most reporter gene assays, which are typically

very robust and there is often little variability in the controls but may be a problem for assays with larger variability. For example, Escher et al. (2012) selected an IR of 1.5 for AREc32 assay for oxidative stress as it was above the LOD, but was still close to the threshold of effect and similar to the lowest observed effect concentration (LOEC).

2.2.3 Cytotoxicity

For endpoints other than cell viability, it is essential to ensure that cytotoxicity is measured in parallel to activation for *in vitro* bioassays. As discussed above, cytotoxicity can cause false-positive or false-negative results, meaning that laboratory staff cannot confirm that an observed effect is real or an artefact without cytotoxicity assessment. This can be assessed using dyes, such as neutral red or MTT, which assesses cell viability based on the ability of cells to uptake dye or mitochondrial metabolic activity, or by live-cell imaging. For assays run in agonist mode, any concentrations causing more than 10% cytotoxicity should be excluded from further data evaluation. Further, a cut-off of 1% cytotoxicity has been proposed for assays run in antagonist mode (Nivala et al., 2018) as cytotoxicity and antagonism cannot be distinguished. Data evaluation for cytotoxicity is discussed further in Section 2.4.

2.3 Available guidelines

There are several test guidelines available for *in vitro* bioassays and well plate-based *in vivo* assays. ISO guidelines for water quality assessment include mutagenicity testing with the Ames test (ISO 11350, 2012), genotoxicity testing using the umu-test (ISO 13829, 2000), bioluminescence inhibition of *Aliivibrio fischeri* (ISO11348-3, 1998), algal toxicity (ISO8692, 2004) and the fish embryo toxicity test (ISO15088, 2007). These guidelines test water directly, rather than after extraction, with sample filtration recommended prior to running the assays. ISO guidelines are also available for reporter gene assays for estrogenicity using two types of the Yeast Estrogen Screen (YES) (ISO/DIS19040-1, 2017; ISO/DIS19040-2, 2017) and a human cell-based reporter gene assay (ISO/DIS19040-3, 2017), with the option of testing water samples directly or SPE extracts.

OECD guidelines for chemical testing are available, with guidelines for well plate-based *in vivo* assays including fish embryo acute toxicity (Test No. 236, 2013), algae growth inhibition (Test No. 201, 2011), the EASZY assay to detect endocrine active substances using transgenic zebrafish embryos (Test No. 250, 2021) and the Xenopus eleutheroembryonic thyroid assay (XETA) (Test No. 248, 2019). There are also guidelines for reporter gene assays indicative of estrogenic activity (Test No. 455, 2021) and androgenic activity (Test No. 458, 2020), as well as acute toxicity in the fish cell line RTgill-W1 (Test No. 249, 2021). The OECD guidelines are developed for chemical testing, but the information provided can also be adapted for water extracts.

2.4 Technical guidance on bioassay data evaluation

Running single chemicals or water extracts in a bioassay will typically generate data in units of absorbance, relative fluorescence units (RFU) or relative light units (RLU), depending on the detection method. Therefore, data evaluation steps are required to derive an EC value. EC values are derived from concentration-effect curves, with log-sigmoidal concentration-effect curves traditionally used for bioassay data evaluation. These curves can range from 0 to 100% effect, with a four-parameter logistic fit commonly applied to determine the minimum, maximum, slope and EC_{50} value (Escher et al., 2018b). When plotted on a linear concentration scale, log-sigmoidal concentration-effect curves are linear up to about 30% effect. Therefore, linear concentration-effect curves have been proposed for evaluating *in vitro* bioassay data at low effect levels (Escher et al., 2018b). A standardised data analysis approach is critical for comparing results as different analysis pipelines will produce different results (Leusch et al., 2010). In this section a standardised approach for deriving EC values and equivalent concentrations from both log-sigmoidal and linear concentration-effect curves is discussed.

2.4.1 Inhibitory concentration

For *in vitro* bioassays, only concentrations causing less than 10% cytotoxicity in agonist mode or less than 1% cytotoxicity in antagonist mode should be used for further data evaluation to prevent cytotoxicity from masking the effect. Cell viability

can be calculated using Equation 2, based on the % confluency of the sample extract and % confluency of the solvent controls if cell viability is assessed using cell imaging, or the signal of the sample extract ($signal_{extract}$) and the signal of the solvent control ($signal_{control}$) if cell viability is measured using spectroscopic methods. The signal can be expressed as absorbance, RFU or RLU. Cytotoxicity can then be calculated using Equation 3.

$$\text{Cell viability (\%)} = \frac{\text{\% Confluency (extract)}}{\text{\% Confluency (control)}} \text{ or } \frac{\text{Signal}_{extract}}{\text{Signal}_{control}} \quad (2)$$

$$\text{Cytotoxicity (\%)} = 1 - \text{cell viability (\%)} \quad (3)$$

IC_{10} can be calculated from linear concentration-effect curves (Equation 4) up to 50% cytotoxicity using Equation 5, with the concentration causing 1% inhibition (IC_{01}) calculated using Equation 6.

$$\text{\% inhibition} = \text{concentration} \cdot \text{slope} \quad (4)$$

$$IC_{10} = \frac{10\%}{\text{Slope}} \quad (5)$$

$$IC_{01} = \frac{1\%}{\text{Slope}} \quad (6)$$

2.4.2 Effect concentration

Assays indicative of induction of xenobiotic metabolism, receptor-mediated effects and apical effects in whole organisms can be expressed as % effect as a maximum effect can be reached in these assays. The % effect is calculated using Equation 7 with $signal_{extract}$, $signal_{control}$ and the signal of the maximum response of the positive reference control ($signal_{max}$) (Figure 6 A).

$$\text{\% effect} = \frac{\text{Signal}_{extract} - \text{Signal}_{control}}{\text{Signal}_{max} - \text{Signal}_{control}} \quad (7)$$

EC_{50} can be determine using log-sigmoidal concentration-effect curves using Equation 8, where min is minimum effect (0%) and max is the maximum effect (100%). The slope and the EC_{50} are adjustable parameters.

$$\text{\% effect} = \text{min} + \frac{\text{max} - \text{min}}{1 + 10^{\text{slope} \cdot (\log EC_{50} - \log \text{concentration})}} \quad (8)$$

While EC_{50} is commonly reported for well plate-based *in vivo* assays, there can be some problems with deriving EC_{50} values for environmental extracts in *in vitro* bioassays. This is because some environmental samples, such as surface water and drinking water, may only induce low effects in an assay and the effect would need to be extrapolated to reach 50% effect, which is not ideal. Secondly, cytotoxicity can occur at high concentration, which may mask the effect. This is illustrated in Figure 6 B, where reduced cell viability occurs at the highest tested concentrations. Linear concentration-effect curves can be applied for data evaluation to avoid these issues. For assays that reach a maximum effect in the

assay, linear concentration-effect curves up to 30% (Equation 9) can be used to determine the EC_{10} value (Equation 10) (Figure 6 C).

$$\% \text{ effect} = \text{concentration} \cdot \text{slope} \tag{9}$$

$$EC_{10} = \frac{10\%}{\text{slope}} \tag{10}$$

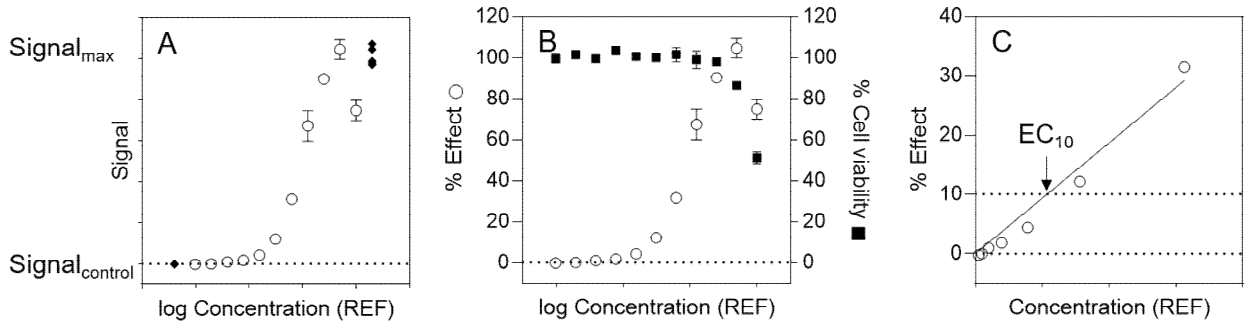

Figure 6: Data evaluation overview for assays that reach a maximum effect in agonist mode.

In contrast to assays where a maximum effect can be reached, there is no defined maximum effect in some *in vitro* bioassays, such as those indicative of adaptive stress responses and some indicative of reactive toxicity (e.g., umuC assay). Therefore, an IR can be calculated based on $signal_{extract}$ and $signal_{control}$ (Equation 11). An IR of 1 means the extract has the same effect as the control. Linear concentration-effect curves up to a maximum IR of 4 can be applied to determine the $EC_{IR1.5}$ (Equations 12 and 13). An IR of 1.5 means that the extract has a 50% greater effect than the control.

$$IR = \frac{signal_{extract}}{signal_{control}} \tag{11}$$

$$IR = 1 + \text{concentration} \cdot \text{slope} \tag{12}$$

$$EC_{IR1.5} = \frac{0.5}{\text{slope}} \tag{13}$$

All these examples focus on assays run in agonist mode. In the case of antagonist mode, a suppression ratio (SPR) can be calculated using Equation 14 with $signal_{extract}$, $signal_{control}$ and the signal of the agonist ($signal_{agonist}$), which is typically the maximum response in the assay (Figure 7 A and B). Assays run in antagonist mode contain a background concentration of agonist, with agonist concentrations causing between 50 to 100% effect used in the literature, though the agonist concentration causing 80% effect has previously been shown to be optimal for both robustness and sensitivity (Neale and Leusch, 2015). A linear concentration-effect curve similar to Equation 9 can be applied to determine the effect concentration causing an SPR of 0.2 ($EC_{SR0.2}$) (Equation 15, Figure 7 C), which is equivalent to 20% suppression of the agonist response.

$$SPR = 1 - \frac{Signal_{extract} - Signal_{control}}{Signal_{agonist} - Signal_{control}} \quad (14)$$

$$EC_{SPR0.2} = \frac{0.2}{slope} \quad (15)$$

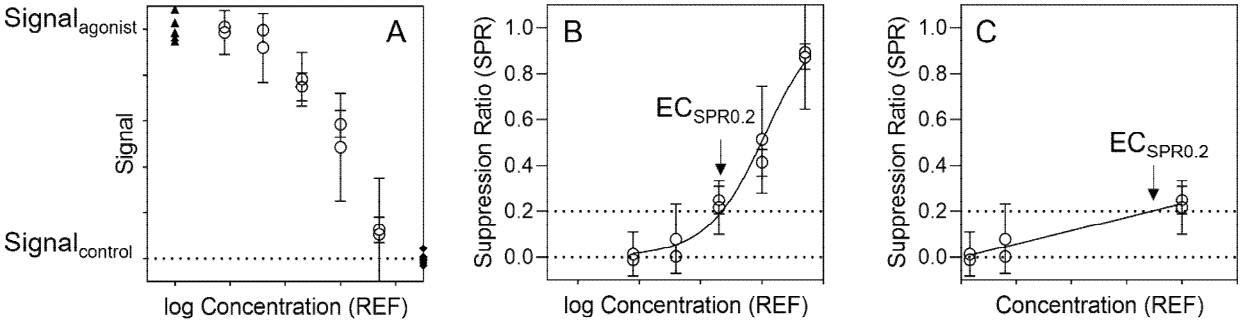

Figure 7: Data evaluation overview for assays in antagonist mode.

As the signal may vary slightly from plate to plate, data should be analysed per plate. Further information about bioassay data evaluation can be found in Escher et al. (2014) and Neale et al. (2017), while more information specifically about linear concentration-effect curves is available in Escher et al. (2018b).

2.4.3 Concentration-effect curves of water samples

Unlike chemicals, the concentration of a water extract cannot be expressed in mass (e.g., µg/L or ng/L) or molar units, but the EC values are expressed as a relative enrichment factor (REF) in units of $L_{water}/L_{bioassay}$. REF is calculated based on the sample EF and the dilution factor (DF) in the assay (Equation 16). The EF is calculated using the volume of water extracted and the volume of the final sample extract (Equation 17) and the DF is calculated based on the volume of extract dosed and the final volume in the bioassay (Equation 18). The observed effect is plotted against the REF using linear or log-sigmoidal concentration-effect curves to determine the EC value. Lower EC values indicate a greater effect than high EC values as less enrichment is required to detect an effect in the assay.

$$\text{Relative enrichment factor } REF = EF \cdot DF \quad (16)$$

$$\text{Enrichment factor } EF = \frac{\text{Volume extracted } (L_{water})}{\text{Final volume of extract } (L_{extract})} \quad (17)$$

$$\text{Dilution factor } DF = \frac{\text{Volume of extract dosed } (L_{extract})}{\text{Final volume in bioassay } (L_{bioassay})} \quad (18)$$

2.4.4 Bioanalytical equivalent concentrations

The EC value can be expressed as a bioanalytical equivalent concentration from bioanalysis (BEQ_{bio}) value, such as a 17β-estradiol equivalent concentration (EEQ). BEQ_{bio} converts the effect of a sample into the effect of a known mass or molar concentration of a reference compound and has the advantage that a high BEQ_{bio} indicates a high effect, whereas a high EC value indicates a low effect. Consequently, translating to BEQ_{bio} can help people unfamiliar with bioanalysis better understand bioassay results. Further, EBTs are commonly given as a BEQ value (EBT-BEQ), so expressing bioassay results as BEQ_{bio} can assist with integrating effect-based monitoring into WSPs. BEQ_{bio} can be calculated using Equation 19 using the EC_{50}, EC_{10} or $EC_{IR1.5}$ value of the assay reference compound in mass or molar units and the corresponding EC_{50}, EC_{10} or $EC_{IR1.5}$ value of the water extract expressed as REF.

$$BEQ_{bio} = \frac{EC_{50}\ (reference)}{EC_{50}\ (extract)} \text{ or } \frac{EC_{10}\ (reference)}{EC_{10}\ (extract)} \text{ or } \frac{EC_{IR.15}\ (reference)}{EC_{IR.15}\ (extract)} \quad (19)$$

The BEQ_{bio} calculation is only valid for EC values derived from log-sigmoidal concentration-effect curves when the slopes are parallel (Villeneuve et al., 2000). However, this is not a requirement for EC values derived from linear concentration-effect curves with a common intercept at the effect axis as BEQ_{bio} is the inverse ratio between the linear concentration-effect curve slopes (Neale et al., 2015). This is another advantage of using linear concentration-effect curves for bioassay data evaluation. Further, the uncertainty associated with BEQ_{bio} values derived from log-sigmoidal concentration-effect curves can be higher than BEQ_{bio} values derived from linear concentration-effect curves because of non-parallel log-sigmoidal concentration-effect curves and extrapolation to untested REFs.

Section 3

3.1 A framework to help integrate effect-based monitoring into Water Safety Plans

The following section provides advice to WSP teams regarding how effect-based monitoring can be used within WSPs, including where in the catchment to customer process, at what test frequencies and how to set alert level triggers. This section also provides guidance on what to do if the effect in a sample exceeds its EBT.

3.1.1 How can effect-based monitoring fit within WSPs

Effect-based monitoring can potentially be applied in the four different monitoring categories in the WSP framework, namely system assessment monitoring, operational monitoring, verification monitoring and validation monitoring (see WP5.1 for more information). An overview of how effect-based monitoring can be applied within the different monitoring categories and WSP modules, as well as where and how frequently to apply effect-based monitoring is provided in Figure 8.

System assessment monitoring is used to characterise water quality and help inform risk assessments and define treatment requirements, with effect-based monitoring suitable in Modules 2 (Describe the water supply system), 3 (Identify hazards and hazardous events and assess the risks) and 10 (Plan and carry out a periodic review of the WSP). In a system assessment monitoring context, effect-based monitoring can be applied anywhere in the system from catchment to customer. The monitoring frequency will depend on the module and on the characteristics of the catchment. For example, effect-based monitoring can be used to describe the water supply system every 3 to 5 years for large catchments (Module 2), though smaller catchments or catchments with more variable water quality may require more frequent monitoring. In contrast, effect-based monitoring should be applied after a hazardous event in the catchment, such as a chemical spill or bushfire (Module 3), or after changes in the catchment or the addition of new treatment processes (Module 10).

Validation monitoring provides evidence that control measures are effective and effect-based monitoring can be applied in Modules 4 (Determine and validate the control measures, reassess and prioritize the risk) and 5 (Develop, implement and maintain an improvement/upgrade plan). Verification monitoring is used to verify routine operations, with effect-based monitoring fitting into Module 7 (Verify the effectiveness of the WSP). Operational monitoring is monitoring of control measures on a continual basis to confirm they are working in a timely manner, with the potential to apply effect-based monitoring in Module 6 (Define monitoring of the control measures). The time required to process and run samples in cell-based bioassays (e.g., at least 3 to 4 days) means that current effect-based monitoring may not be suitable for operational monitoring of rapid changes of conditions and their consequences but may be more suitable for operational monitoring of long-term developments of treatment and surveillance. To meet the need for operational monitoring of changes in operation, further research is required to develop rapid-response tools and online effect-based monitoring options. Any effect data generated during verification and operational monitoring can also be collated to inform system assessment monitoring.

For validation, verification, and operational monitoring, monitoring before and after the control measure is required, but the frequency varies greatly. Validation monitoring is only required to validate the existing control measures (Module 4) or any new or upgraded control measures (Module 5), while verification monitoring should be conducted on a quarterly or biannual basis, similar to chemical analysis. As a first application of effect-based monitoring, treatment plant operators may decide to use bioassays for verification monitoring of treated water. Bioassay test batteries, as described in Section 2.1.1, should be applied for both validation and verification monitoring to gain a more thorough understanding of the chemical burden. In contrast, online effect-based monitoring or rapid response tools, such as Microtox or BLT-Screen, are suitable for operational monitoring, which is conducted on an ongoing basis.

For all types of monitoring, the observed effect should be compared to an EBT. EBTs are available for drinking water and surface water, but not readily available for wastewater effluent. Instead, WSP teams can compare the effect in wastewater effluent with the surface water EBT after dilution. If the dilution factor is not known, a dilution factor of 10 can be used as a conservative estimate of wastewater dilution into a receiving waterbody. Available EBTs are provided in WP3.4.

The uptake of effect-based monitoring into WSPs can be facilitated through Modules 8 (Prepare management procedures) and 9 (Develop supporting programmes) of the WSP framework. This includes developing SOPs and training programs to help develop people's skills and knowledge, as well as research and development to improve understanding of water quality.

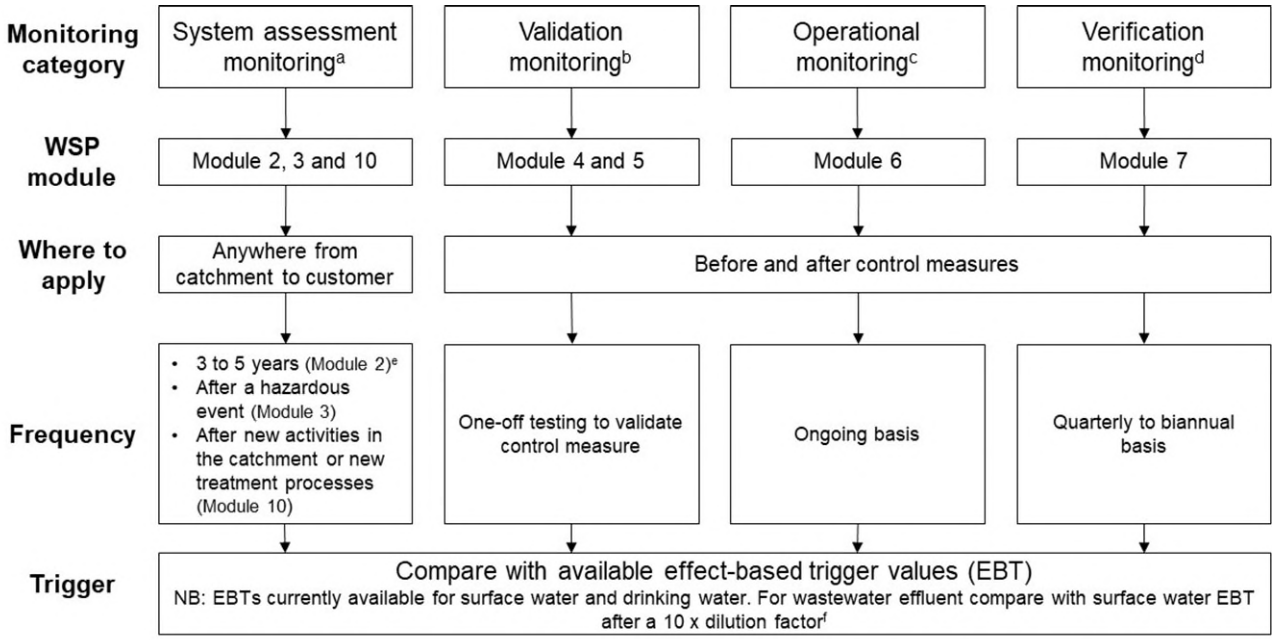

Figure 8: Overview of how effect-based monitoring can be applied within WSPs, with a focus on different monitoring categories, WSP modules, monitoring location and monitoring frequency.

[a]monitoring to characterise water quality and help inform risk assessments and define treatment requirements; [b]monitoring to provide evidence that control measures are effective; [c]monitoring of control measures on a continual basis to confirm they are working in a timely manner; [d]monitoring used to verify routine operations; [e]more frequent system assessment monitoring is required for smaller reservoirs or reservoirs with more variable water quality; [f]a dilution factor of 10 was selected as a conservative estimate of dilution of wastewater effluent into a receiving waterbody. However, if the operator knows that the dilution factor is larger or smaller, the correct dilution factor can be used instead.

3.1.2 What to do if the effect of a sample exceeds its effect-based trigger value

The response in a bioassay, expressed as BEQ_{bio}, can be compared to the EBT-BEQ, with no further action required if BEQ_{bio} is lower than the EBT-BEQ. If the measured BEQ_{bio} value exceeds the EBT-BEQ, the first step is to check the bioassay QC and collect another water sample from the same site and re-test (step ❶ in Figure 9 and Figure 10). This is comparable to what is currently done for chemical analytes. If the BEQ_{bio} of the second sample is below the EBT-BEQ (in other words, if the re-test does not confirm the initial positive result), then no further action is required. If, however, the second test confirms the initial positive result and both samples report a BEQ_{bio} > EBT-BEQ, then further action is needed. What action specifically depends on the category of the bioassay (refer to WP3.4). Bioassays fall broadly into two categories: Category 1 and Category 2 assays.

3.1.2.1 Category 1 bioassay

The framework for interpreting Category 1 assay results is presented in Figure 9. Category 1 assays are highly specific bioassays that are mainly triggered by a limited number of known and generally potent chemicals. For these Category 1 bioassays, almost all the detected effect can typically be explained by known chemicals. A good example of a Category 1 assay is a reporter gene assay for the activation of ER or bioassays for photosynthesis inhibition (such as the Imaging-PAM). For example, >90% of the estrogenic activity in most water samples is caused by natural hormones (such as estradiol, estrone and estriol), synthetic hormones (such as ethinylestradiol) and a few industrial xenoestrogens (such as bisphenol A

or nonylphenol). Thus, in the case of a Category 1 assay response, it is usually possible to target a short list of chemicals for chemical analysis. The concentration of each compound detected (C_i) is then multiplied by the potency of each compound in the bioassay to produce a calculated bioassay response $BEQ_{chem,i}$, which is then summed up over all chemicals to obtain the bioanalytical equivalent concentration from chemical analysis BEQ_{chem} ($BEQ_{chem}=\sum BEQ_{chem,i}$), compared to the actual bioassay response BEQ_{bio} (step ❸ in Figure 9). If the two values agree (i.e., are within 20%), then it is concluded that the identified chemicals are indeed driving most of the bioassay response, and the concentrations of chemicals detected (C_i) can be compared to available conventional chemical guideline values (GV_i). If the concentrations exceed the guideline value (step ❹ in Figure 9), then the usual process is followed for exceedance of regulatory standards. Otherwise, if the chemical concentrations do not exceed individual guideline values, then the water is technically compliant with regulatory expectations. While no further immediate action is necessary, the bioassay response may indicate a potential risk associated with unregulated chemicals, or that the EBT-BEQ is set too low and may need to be revisited.

However, if at step ❸, the predicted bioassay response BEQ_{chem} is less than 80% of the measured bioassay response BEQ_{bio} ($BEQ_{chem}<0.8\times BEQ_{bio}$), then this indicates that other unidentified chemicals are contributing to the bioassay response. In this case, an effort should be made to identify those unknown mixture risk drivers. Effect-directed analysis (EDA) has been applied successfully to identify unknown contributors to the mixture effects in water samples for Category 1 bioassays. If additional causative chemicals are identified, then they are included in the predicted bioassay response BEQ_{chem}, and this is again compared to the measured bioassay response BEQ_{bio} (step ❺ in Figure 9). If the additional chemicals have now improved the agreement between BEQ_{chem} and BEQ_{bio}, then the conventional approach can be used, as described above.

If the measured bioassay response BEQ_{bio} remains significantly different from the predicted bioassay response based on identified chemicals BEQ_{chem} at step ❺ (e.g., $BEQ_{chem}<0.8\times BEQ_{bio}$), then even EDA has not been able to identify all significant bioactive chemicals. At this stage, the response depends on the context of the bioassay testing and the amplitude of the exceedance. In consultation with the relevant regulatory body, additional steps may be needed especially if the exceedance is more than 10 times the EBT. For example, if testing was done in a water treatment plant (step ❻ in Figure 9), then it may be possible to optimise the treatment process to remove the bioassay response. This could be first tested at the bench-scale to fine-tune the treatment process. If testing was done in another context (e.g., monitoring of surface water), then the efforts should focus on assessing the quality of the surface water (e.g., ecosystem assessment) and identifying the source of the pollution. The magnitude of the response should depend on the magnitude of the exceedance and advice from the regulatory body. This is in recognition of the fact that for some assays there is several EBTs available in the literature.

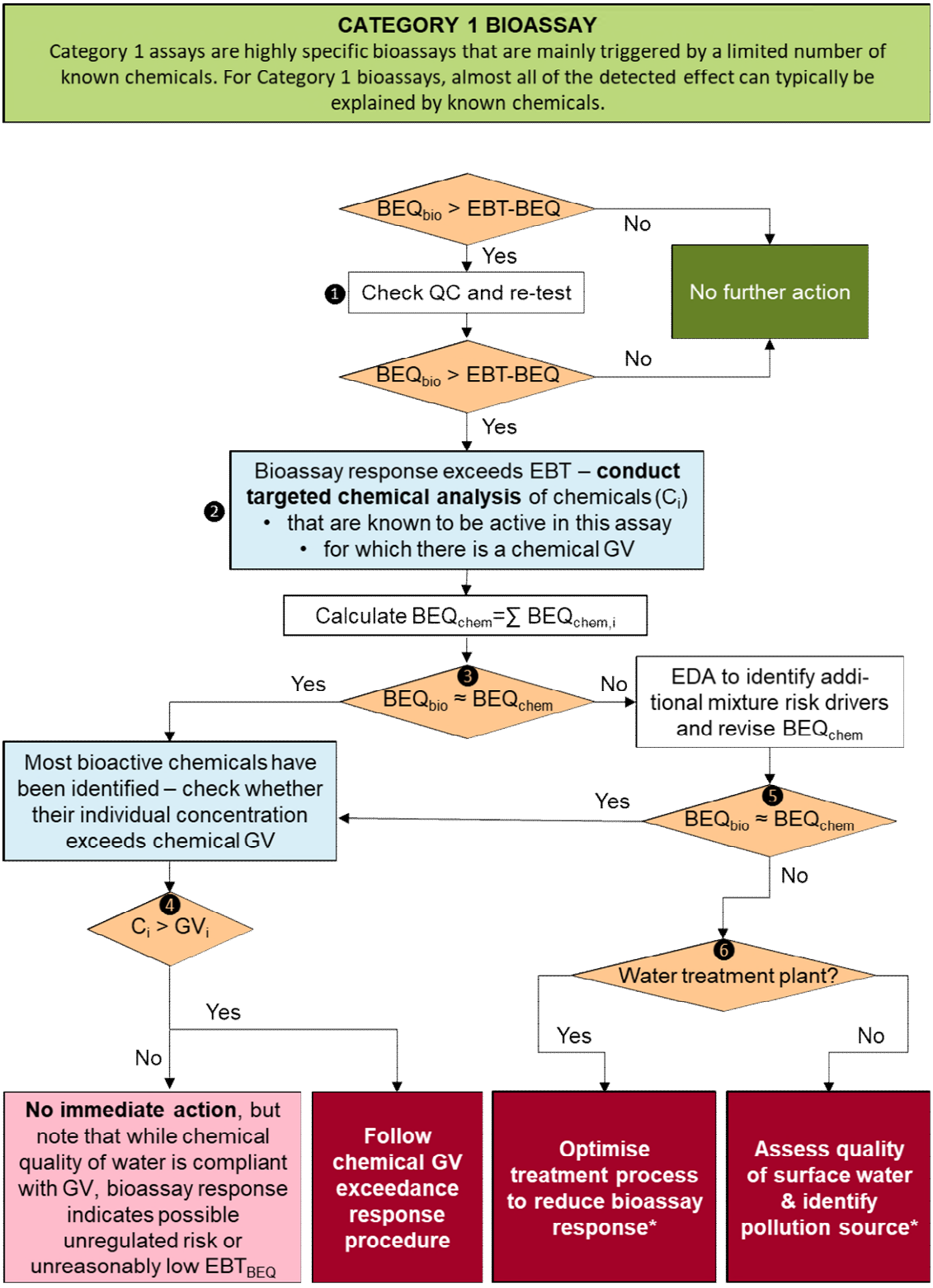

Figure 9: Interpretation framework for Category 1 bioassays.

**The magnitude of the response should depend on the magnitude of the exceedance and regulatory advice.*

3.1.2.2 Category 2 bioassay

In Category 2 bioassays only a small fraction of the measured effect can typically be explained by known and detected chemicals, which has important implications for the interpretation of the bioassay results. Typical Category 2 bioassays are those indicative of adaptive stress responses or apical endpoints (refer to WP3.4). For these types of assays, BEQ_{chem} is usually much lower than BEQ_{bio}, and thus it usually makes little sense to try to identify causative chemicals (step ❷ in Figure 10). In this case, the cytotoxicity response (expressed here as IC10) is compared to the cytotoxicity EBT (referred here as EBT_{IC10}) (step ❸ in Figure 10). If the IC_{10} in the assay is less than the EBT_{IC10}, the sample is additionally cytotoxic and risk mitigation is required. As above, the response here depends on the context of the screening (step ❹ in Figure 10): if within a water treatment plant, then optimisation of the treatment process should be investigated to reduce the bioassay response; otherwise, the quality of the surface water should be investigated (using conventional ecosystem assessment methods) and the source of the pollution identified, if possible. The magnitude of the response should depend on the magnitude of exceedance and regulatory advice.

However, if at step ❸ it is determined that the sample was not significantly toxic (IC10 > EBT_{IC10}), then this suggests that while there are bioactive compounds present (which explain the BEQ_{bio} being greater than the EBT-BEQ), those compounds may not pose an acute risk. At this stage, it is suggested that the water quality be monitored with other bioassays that cover other endpoints, including Category 1 bioassays. The urgency and magnitude of the response may be moderated by the magnitude of the exceedance. In the case of a significant exceedance (e.g., 10 times higher than the EBT), a broad chemical characterisation of the water quality should also be conducted.

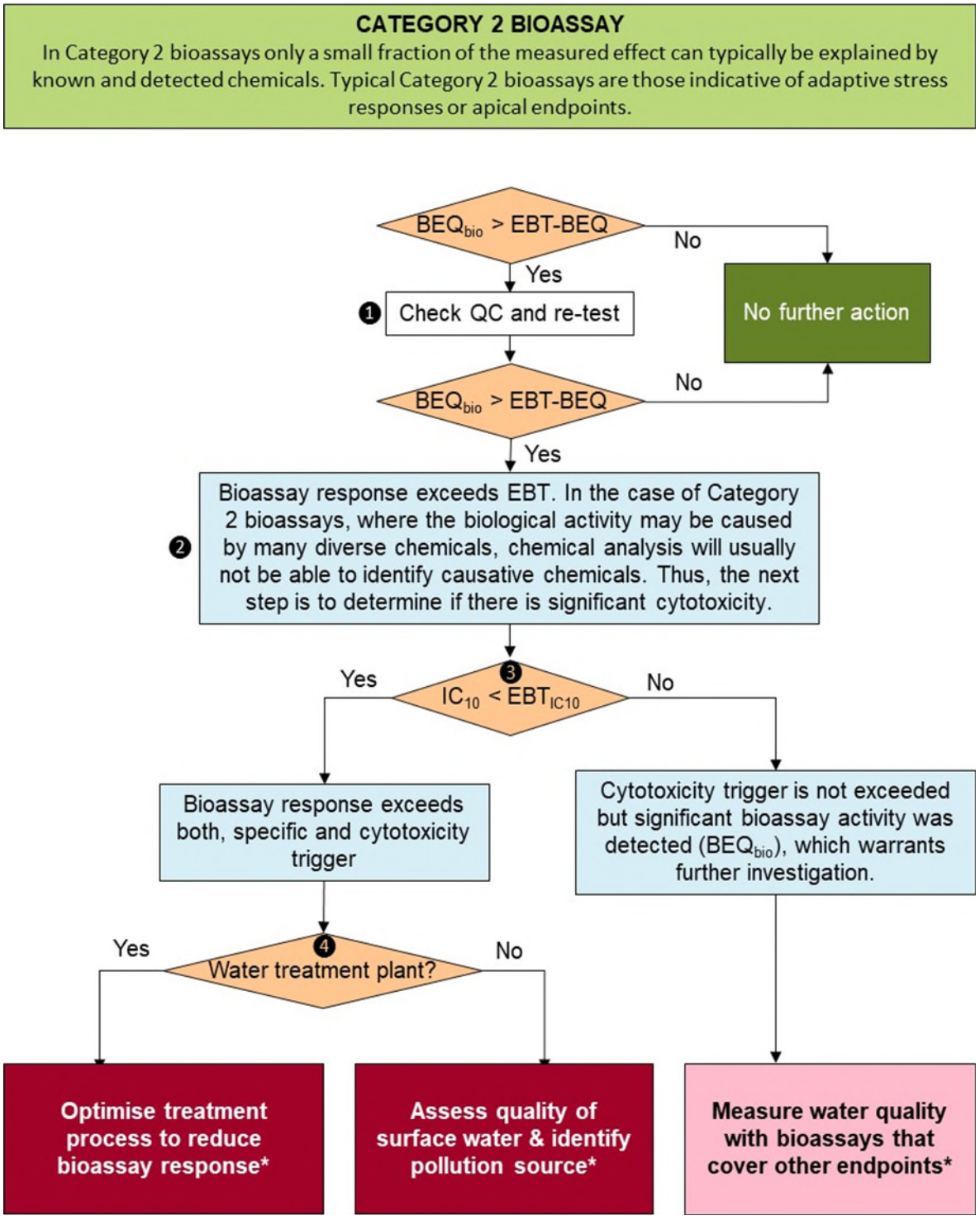

Figure 10: Interpretation framework for Category 2 bioassays.

**The magnitude of the response should depend on the magnitude of the exceedance and regulatory advice.*

3.2 Case studies

To help support the implementation of effect-based monitoring into WSP and FSMS frameworks it is important to include case studies to show how bioassays can be applied to describe water quality (Module 2 of WSP), verify treatment efficacy (Module 7 of WSP) and validate control measures (Module 4 of WSP). This section includes previously published case studies focusing on drinking water, surface water and wastewater where effect-based monitoring has been used for system assessment monitoring, verification monitoring and validation monitoring. Where possible, the observed effect was compared with available EBTs. It should be noted that the case studies below do not exactly follow the recommended bioassay test batteries provided in Figures 2 and 3.

3.2.1 Benchmarking water quality

Effect-based monitoring is commonly used to characterise the quality of drinking water, surface water and wastewater (e.g., Escher et al., 2014; Rosenmai et al., 2018; Medlock Kakaley et al., 2020). This is an example of system assessment monitoring, which can help inform risk assessments and define treatment requirements. Ideally, the bioassay results should be compared with EBTs to have any practical application. In this case study, a battery of bioassays indicative of different hormone receptor-mediated effects was applied to assess endocrine activity in drinking water, surface water and treated wastewater from six countries: Germany, Australia, South Africa, France, the Netherlands and Spain (Leusch et al., 2018). The test battery included commonly studied hormone receptors, including ER, androgen receptor (AR) and glucocorticoid receptor (GR), as well as less frequently applied hormone receptors, including the progesterone receptor (PR), thyroid receptor (TR), retinoid X receptor (RXR), retinoid acid receptor (RARα) and mineralocorticoid receptor (MR) (Table 5). The observed effect, expressed as BEQ_{bio}, was compared with EBTs, where available.

Table 5: Overview of hormone receptor-mediated assays applied in Leusch et al. (2018).

Mode of action	Assay	Positive reference compound
Activation of estrogen receptor (ER) Inhibition of estrogen receptor (ER)	ERα GeneBLAzer ERα GeneBLAzer	17β-Estradiol Hydroxytamoxifen
Activation of androgen receptor (AR) Inhibition of androgen receptor (AR)	AR GeneBLAzer AR GeneBLAzer	Metribolone (R1881) Cyproterone acetate
Activation of glucocorticoid receptor (GR) Inhibition of glucocorticoid receptor (GR)	GR GeneBLAzer GR GeneBLAzer	Dexamethasone Mifepristone (RU486)
Activation of glucocorticoid receptor (GR)	GR CALUX	Dexamethasone
Activation of progesterone receptor (PR) Inhibition of progesterone receptor (PR)	PR GeneBLAzer PR GeneBLAzer	Levonorgestrel Mifepristone (RU486)
Activation of progesterone receptor (PR)	PR CALUX	Org 2058
Activation of thyroid receptor (TR)	GH3.TRELuc	Triiodothyronine
Activation of retinoid X receptor (RXR)	RXR-CALUX	Trans retinoic acid
Activation of retinoid acid receptor (RARα) and RXR Inhibition of retinoid acid receptor (RARα) and RXR	HELN RARα-RXR HELN RARα-RXR	TTNPB BMS493
Activation of mineralocorticoid receptor (MR) Inhibition of mineralocorticoid receptor (MR)	HG5LN MR HG5LN MR	Aldosterone Spironolactone

Endocrine activity was below the LOQ for most water samples. Estrogenic activity (0.78 ng/L EEQ) was detected in treated wastewater from France, while glucocorticoid (130 ng/L dexamethasone equivalent concentration (EQ)) and anti-mineralocorticoid activity (3100 ng/L spironolactone EQ) were detected in treated wastewater from Spain. Surface water from Spain was the most active, with estrogenic (0.31 ng/L EEQ), glucocorticoid (96 ng/L dexamethasone EQ), progestagenic (1.1 ng/L levonorgestrel EQ) and anti-mineralocorticoid activity (910 ng/L spironolactone EQ) detected, while anti-mineralocorticoid activity (660 ng/L spironolactone EQ) was also detected in surface water from France. None of the wastewater or surface water extracts from the other countries had an effect in any of the assays above the LOQ, while none of the tested drinking water extracts had an effect.

The effects in the bioassays were compared with available EBTs (Table 6). EBTs are typically determined for a specific bioassay rather than an endpoint. Therefore, EBTs derived for a specific assay were used where possible, but the lack of

EBTs for all studied assays meant that available EBTs for a particular endpoint were also compared in this case study. Despite anti-mineralocorticoid activity being detected in both treated wastewater and surface water, there is no EBT currently available, so we cannot determine whether the detected activity poses a potential risk to ecological health. The effect in treated wastewater from France was above the treated wastewater EBT for estrogenic activity, while the observed glucocorticoid activity from Spain exceeded the surface water EBT. This suggests that undiluted wastewater discharge may pose a risk to aquatic organisms. However, wastewater effluent is typically diluted in receiving waterbodies and, assuming a conservative dilution factor of 10, the effect would be below the EBT. All surface water extracts were below the surface water EBTs, though it was not possible to compare the progestagenic activity in surface water from Spain as no surface water EBT has been derived for this endpoint to date. Further, all drinking water extracts were below available drinking water EBTs, indicating that the studied drinking water samples do not pose a risk to human health. Effect-based monitoring every 3 to 5 years is recommended to describe the water supply system, though risk managers can also apply effect-based monitoring more frequently, such as after a hazardous event in the catchment.

Table 6: Comparison of BEQ_{bio} values (ng/L) detected in Leusch et al. (2018) with available effect-based trigger values (EBT) for surface waters (ecological health) and drinking waters (human health).

Activity	Units	Treated wastewater‡ and surface water EBT-BEQ	Treated wastewater	Surface water	Drinking water EBT-BEQ	Drinking water
ER	EEQ (ng/L)	0.1 – 0.4[a‡], 0.34[b]	<0.6 – 0.78	<0.1 – 0.31	1.8 – 3.8[c,d]	<0.03
AR	Dihydrotestorone EQ (ng/L)	n/a	<4	<1	11[c]	<1
Anti-AR	Flutamide EQ (ng/L)	3280[b]	<22000	<8900	n/a	<4400
GR	Dexamethasone EQ (ng/L)	100[b]	<120 – 130[†]	<23 – 96	21-150[c,d]	<5.8
PR	Levonorgestrel EQ (ng/L)	n/a	<2.5	<0.5 – 1.1	730[c*]	<0.1

[a]Wastewater effluent EBT-BEQ calculated for YES, MELN, ER-CALUX, E-SCREEN and MVLN (Jarošová et al., 2014); [b]Surface water EBT-BEQ calculated for ERα GeneBLAzer and Anti AR GeneBLAzer (Escher et al., 2018a) and GR-CALUX (van der Oost et al., 2017); [c]Drinking water EBT-BEQ calculated for ER-CALUX, AR-CALUX, GR-CALUX and PR-CALUX (Brand et al., 2013); [d]Drinking water and recycled water EBT-BEQ calculated for ERα GeneBLAzer and GR-CALUX (Escher et al., 2015).

**Calculated from 333 ng/L Org2058 equivalent in the original reference divided by the relative potency of levonorgestrel in the PR-CALUX of 0.46 compared to Org2058 based on the EC_{50} values for levonorgestrel from Leusch et al. (2017) and Org2058 from Van der Linden et al. (2008); ‡treated wastewater; †activity above the surface water EBT, but with a dilution factor of 10 for treated wastewater in the receiving waters assumed BEQ_{bio} would likely be below the surface water EBT.*

Colour scheme: green = detected activity below available EBT-BEQ; yellow = activity detected but no EBT-BEQ available; orange = activity detected <10× above the EBT-BEQ; grey = no activity detected and no EBT-BEQ.

3.2.2 Treatment process efficacy

Bioassays have been applied previously to evaluate the treatment efficacy of WWTPs, advanced water treatment plants and DWTPs (e.g., Bain et al., 2014; Jia et al., 2015; Conley et al., 2017). This is an example of verification monitoring to confirm the quality of the produced water, with the effect in the product water compared with an EBT. Further, treatment efficacy over time can be evaluated by comparing the measured effect with previously measured bioassay data from the same plant. Even if the effect in the treated water does not exceed the EBT, an increasing effect in the treated water over time could indicate the need to replace equipment or upgrade treatment processes.

In this next case study, a battery of bioassays indicative of xenobiotic metabolism, hormone receptor-mediated effects and adaptive stress responses was applied to evaluate effect removal by a conventional WWTP and different constructed wetlands (Nivala et al., 2018). Constructed wetlands are an option for wastewater treatment when centralised treatment is not possible, but few studies have evaluated effect removal by constructed wetlands. Therefore, a test battery covering different stages of cellular toxicity pathways was assembled (Table 7) and applied to influent and effluent samples from pilot-scale conventional and intensified constructed wetlands, as well as a conventional WWTP.

Table 7: Overview of bioassays applied in Nivala et al. (2018). Cytotoxicity was measured in parallel for all assays.

Mode of action	Assay	Positive reference compound
Xenobiotic metabolism		
Activation of aryl hydrocarbon receptor (AhR)	AhR CALUX	2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD)
Binding to peroxisome proliferator-activated receptor gamma (PPARγ)	PPARγ-bla	Rosiglitazone
Hormone receptor-mediated effects		
Activation of estrogen receptor (ER) Inhibition of estrogen receptor (ER)	ERα GeneBLAzer ERα GeneBLAzer	17β-Estradiol Tamoxifen
Activation of androgen receptor (AR) Inhibition of androgen receptor (AR)	AR GeneBLAzer AR GeneBLAzer	Metribolone (R1881) Cyproterone acetate
Activation of glucocorticoid receptor (GR) Inhibition of glucocorticoid receptor (GR)	GR GeneBLAzer GR GeneBLAzer	Dexamethasone Mifepristone (RU486)
Activation of progesterone receptor (PR) Inhibition of progesterone receptor (PR)	PR GeneBLAzer PR GeneBLAzer	Promegestone Mifepristone (RU486)
Adaptive stress response		
Oxidative stress response (OSR)	AREc32	tert-Butylhydroquinone (tBHQ)
NF-κB response	NF-κB-bla	Tumor necrosis factor alpha (TNFα)

The majority of samples were active in assays indicative of activation of AhR, binding to PPARγ, activation of ER, OSR and NF-κB response, with the NF-κB response the most responsive endpoint in the study. In contrast, the effect was often masked by cytotoxicity or there was no effect up to the maximum tested REF for the other endpoints. Consequently, effect removal efficacy could only be calculated for five endpoints.

Effect removal by intensified wetlands was comparable or better than the conventional WWTP (Figure 11), with between 74 to 85% removal of AhR activity and 98 to 99% removal of estrogenic activity observed. Between 46 to 69% of AhR activity was removed after activated carbon filtration and ozonation in a water reclamation plant (Reungoat et al., 2010), while 80 to >99% removal of estrogenic activity by WWTPs has been reported (e.g., Jalova et al., 2013; Houtman et al., 2018). This shows that intensified wetlands can remove biological effects to a greater extent than conventional WWTPs. In contrast to the intensified wetlands, the conventional horizontal flow wetland without aeration had much poorer removal of biological effects, particularly for PPARγ activity, estrogenic activity and NF-κB activity.

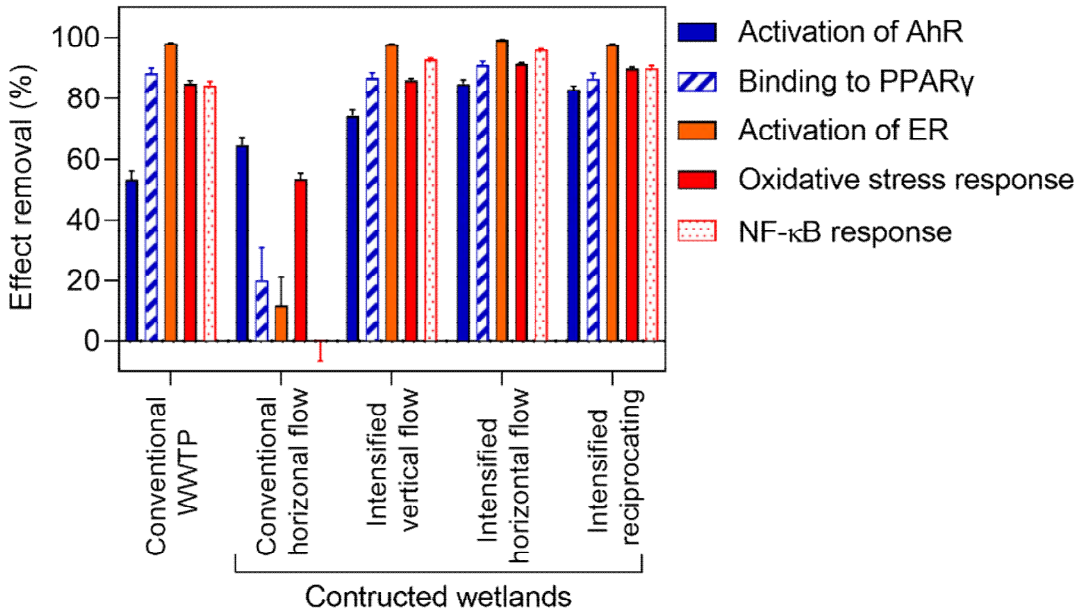

Figure 11: Effect removal efficacy (%) after treatment by a conventional WWTP and different constructed wetlands. Adapted from Nivala et al. (2018)

EBTs have primarily been developed for drinking water and surface water, with no wastewater effluent EBTs specific for the applied bioassays. Surface water EBTs are available for all assays except NF-κB-bla, in Escher et al. (2018a) and Escher and Neale (2021). With the exception of activation of ER, treated wastewater was below the proposed surface water EBT-BEQs. Assuming a dilution factor of 10 for release into surface waters, the effect would be below the surface water EBT-BEQ for activation of ER for all treatment processes, except the conventional horizontal flow wetland which showed poor effect removal. This case study demonstrates how effect-based monitoring can be used for verification monitoring, with routine monitoring (e.g., quarterly to biannual basis) recommended to verify that the treatment processes are working and that the effect in the treated water is at an acceptable level.

3.2.3 Understanding treatment processes

Source waters that feed DWTPs may contain organic micropollutants, while DBPs may form during treatment processes, such as chlorination. Therefore, validation monitoring can be conducted to understand how well treatment processes can reduce or remove effects. This case study aimed to evaluate the efficacy of treatment processes in three DWTPs in the Paris area, France, using a test battery focusing on hormone-receptor mediated effects to evaluate micropollutant removal and reactive toxicity and adaptive stress responses to assess DBP formation (Neale et al., 2020b). Two plants, Choisy-le-Roi and Neuilly-sur-Marne, applied pre-ozonation (Choisy-le-Roi only), clarification, sand filtration, ozonation, granular activated carbon filtration, UV and chlorination. Samples were collected from the source water, after UV treatment and after chlorination over four seasons. A third plant, Méry-sur-Oise, treated 30% of the water using a similar biological treatment process as the other DWTPs, with 70% of the water treated using nanofiltration. Samples were collected from the source water, after biological treatment, after nanofiltration and after chlorination over the same four seasons. The water extracts were analysed with assays indicative of activation and inhibition of the ERα, AR, GR and PR, as well as assays indicative of mutagenicity, OSR, p53 response and NF-κB response.

Of the studied endpoints, the extracts were only active in assays indicative of activation of ER, OSR and NF-κB response (Figure 12). No other hormonal activity was detected in any of the samples, while cytotoxicity often masked the p53 response. Further, none of the source or product samples were mutagenic in the Ames assay using strains S. typhimurium TA98, TA100 and YG7108 (both with and without metabolic activation).

While estrogenic activity was commonly detected in source water and ranged from 0.17 to 3.98 ng/L EEQ, biological treatment and UV were able to remove most of the estrogenic activity in all DWTPs, with the effect in the product water often below the assay LOD. Estrogenic activity was only detected in one product water sample at 0.04 ng/L EEQ, which translated to a treatment efficacy of 95.7%. Similarly high removal efficacy of estrogenic activity in DWTPs has previously been observed In the literature (Lv et al., 2016; Shi et al., 2018).

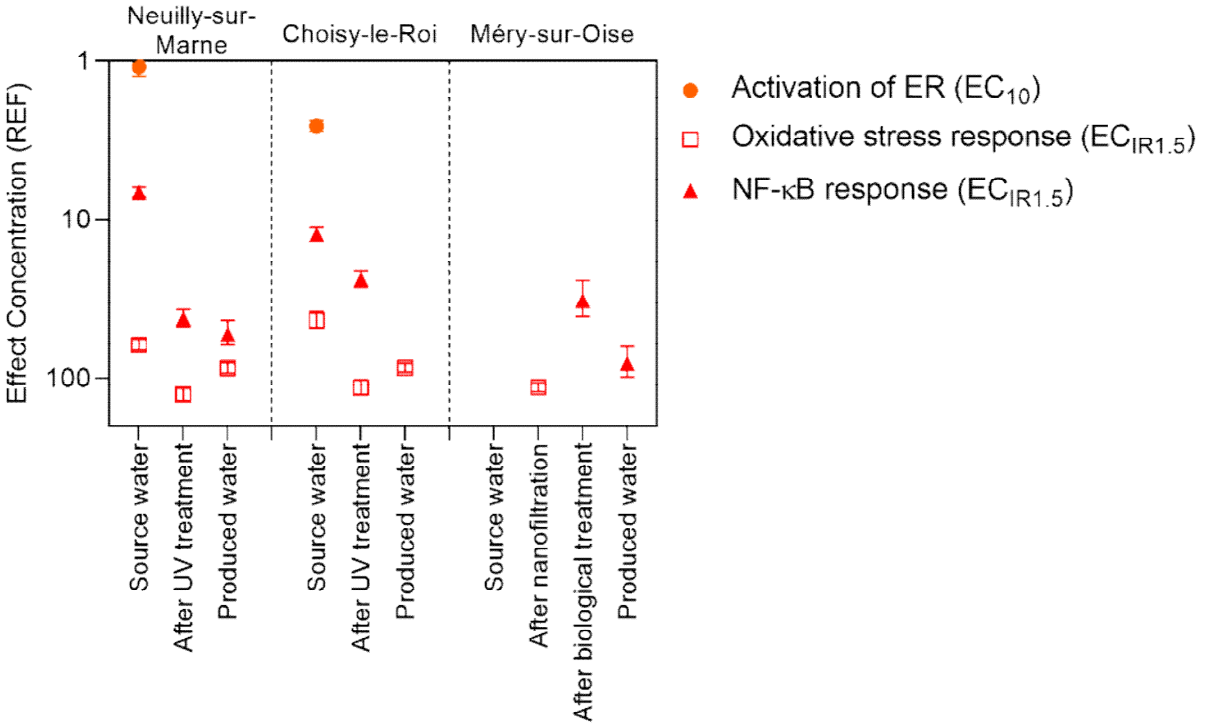

Figure 12: Effect concentrations for activation of ER (EC_{10}, orange circles), oxidative stress response ($EC_{IR1.5}$, red squares) and NF-κB response ($EC_{IR1.5}$, red triangles) in units of relative enrichment factor (REF) for all DWTPs in May 2018. Adapted from Neale et al. (2020b).

The OSR was often highest in the source water, decreased after UV treatment and then increased slightly in the product water after chlorination (Figure 12). However, the extracts still needed to be enriched between 78 to 136 times in the assay before an effect could be detected in the treated and product water. The effect before and after chlorination was assessed to determine the contribution of formed DBPs to the observed OSR based on the approach described in Hebert et al. (2018). Between 25 to 32% of the OSR could be attributed to formed DBPs in May 2018, showing that most of the effect was due to compounds already present in the source water.

To determine whether the effect in the product water was acceptable or unacceptable, the effect was compared with available EBTs (Figure 13). Estrogenic activity in product water was 45 times lower than the proposed drinking water EBT-BEQ of 1.8 ng/L EEQ (Escher et al., 2015). The EBT for the OSR assay was expressed as $EC_{IR1.5}$ of REF 6 (Escher et al., 2013), but was converted to an equivalent concentration using the oxidative stress reference compound tBHQ. The effect in the product water was 13 to 17 times lower than the converted EBT-BEQ of 85.5 µg/L tBHQ-EQ. The large difference between the observed effect and the proposed EBTs shows the high quality of the final treated water.

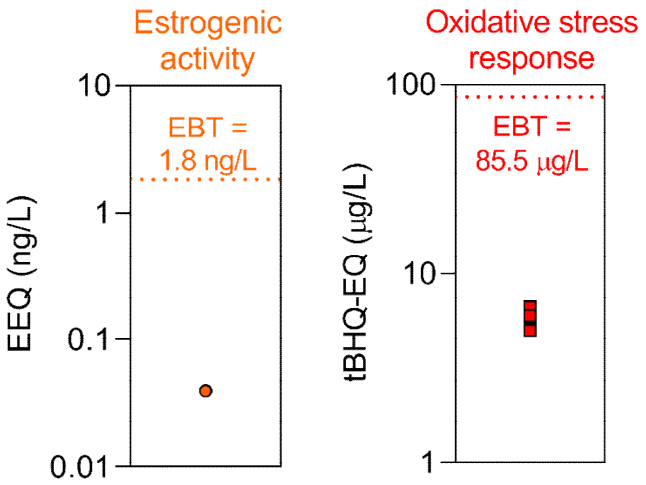

Figure 13: The effect in the product water, expressed as 17β-estradiol equivalent concentration (EEQ) for estrogenic activity and tert-butylhydroquinone (tBHQ) equivalent concentration (tBHQ-EQ), compared to the effect-based trigger value (EBT-BEQ), which was indicated by the dotted line.

Conclusions

To help support the integration of effect-based monitoring in WSP and FSMS frameworks, there is a need for protocols, guidance documents and decision-making tools to assist both laboratory staff and WSP teams. In this report, protocols and guidance documents for sample collection and processing, bioassay QA and QC, and bioassay data evaluation were developed for laboratory staff. Further, decision-making tools to assist users with bioassay selection, sample collection and processing were provided, as well as a framework targeted at WSP teams on how to integrate effect-based monitoring in a WSP framework, including the steps to take if the effect of a sample exceeds its EBT. Finally, three case studies were selected to demonstrate how effect-based monitoring can be applied to benchmark water quality, assess treatment efficacy, and understand critical treatment processes, with the examples covering drinking water, surface water and wastewater.

Acknowledgments

We thank Milo de Baat, Patrick Smeets (both KWR), Jerome Enault (Suez), Charlotte Arnal and Faten Belhadj-Kaabi (both Veolia), Alvina Mehinto (Southern California Coastal Water Research Project (SCCWRP)), David Cunliffe (SA Health), Leo Posthuma (National Institute for Public Health and the Environment, (RIVM)) and Steve Melvin (Griffith University) for their helpful feedback.

References

Bain, P.A., Williams, M. and Kumar, A. (2014). Assessment of multiple hormonal activities in wastewater at different stages of treatment. *Environmental Toxicology and Chemistry*, **33**(10): 2297-2307.

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.

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.

Dong, S.K., Page, M.A., Massalha, N., Hur, A., Hur, K., Bokenkamp, K., Wagner, E.D. and Plewa, M.J. (2019). Toxicological comparison of water, wastewaters, and processed wastewaters. *Environmental Science & Technology*, **53**(15): 9139-9147.

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. (2018a). 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., Allinson, M., Altenburger, R., Bain, P.A., Balaguer, P., Busch, W., Crago, J., Denslow, N.D., Dopp, E., Hilscherova, K., Humpage, A.R., Kumar, A., Grimaldi, M., Jayasinghe, B.S., Jarosova, B., Jia, A., Makarov, S., Maruya, K.A., Medvedev, A., Mehinto, A.C., Mendez, J.E., Poulsen, A., Prochazka, E., Richard, J., Schifferli, A., Schlenk, D., Scholz, S., Shiraish, F., Snyder, S., Su, G.Y., Tang, J.Y.M., van der Burg, B., van der Linden, S.C., Werner, I., Westerheide, S.D., Wong, C.K.C., Yang, M., Yeung, B.H.Y., Zhang, X.W. and Leusch, F.D.L. (2014). Benchmarking organic micropollutants in wastewater, recycled water and drinking water with *in vitro* bioassays. *Environmental Science & Technology*, **48**(3): 1940-1956.

Escher, B.I., Dutt, M., Maylin, E., Tang, J.Y.M., Toze, S., Wolf, C.R. and Lang, M. (2012). Water quality assessment using the AREc32 reporter gene assay indicative of the oxidative stress response pathway. *Journal of Environmental Monitoring*, **14**(11): 2877-2885.

Escher, B.I. and Neale, P.A. (2021). Effect-based trigger values for mixtures of chemicals in surface water detected with *in vitro* bioassays. *Environmental Toxicology and Chemistry*, **40**(2): 487-499.

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., Neale, P.A. and Leusch, F.D.L. (2021). Bioanalytical Tools in Water Quality Assessment - Second Edition. IWA Publishing, London.

Escher, B.I., Neale, P.A. and Villeneuve, D.L. (2018b). The advantages of linear concentration-response curves for *in vitro* bioassays with environmental samples. *Environmental Toxicology and Chemistry*, **37**(9): 2273-2280.

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.

Hebert, A., Feliers, C., Lecarpentier, C., Neale, P.A., Schlichting, R., Thibert, S. and Escher, B.I. (2018). Bioanalytical assessment of adaptive stress responses in drinking water: A predictive tool to differentiate between micropollutants and disinfection by-products. *Water Research*, **132**: 340-349.

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.

Jalova, V., Jarosova, B., Blaha, L., Giesy, J.P., Ocelka, T., Grabic, R., Jurcikova, J., Vrana, B. and Hilscherova, K. (2013). Estrogen-, androgen- and aryl hydrocarbon receptor mediated activities in passive and composite samples from municipal waste and surface waters. *Environment International*, **59**: 372-383.

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.

Jeong, C.H., Wagner, E.D., Siebert, V.R., Anduri, S., Richardson, S.D., Daiber, E.J., McKague, A.B., Kogevinas, M., Villanueva, C.M., Goslan, E.H., Luo, W.T., Isabelle, L.M., Pankow, J.F., Grazuleviciene, R., Cordier, S., Edwards, S.C., Righi, E., Nieuwenhuijsen, M.J. and Plewa, M.J. (2012). Occurrence and toxicity of disinfection byproducts in European drinking waters in relation with the HIWATE epidemiology study. *Environmental Science & Technology*, **46**(21): 12120-12128.

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.

Judson, R., Houck, K., Martin, M., Richard, A.M., Knudsen, T.B., Shah, I., Little, S., Wambaugh, J., Setzer, R.W., Kothya, P., Phuong, J., Filer, D., Smith, D., Reif, D., Rotroff, D., Kleinstreuer, N., Sipes, N., Xia, M.H., Huang, R.L., Crofton, K. and Thomas, R.S. (2016). Analysis of the effects of cell stress and cytotoxicity on *in vitro* assay activity across a diverse chemical and assay space. *Toxicological Sciences*, **152**(2): 323-339.

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.

Landrigan, P.J., Fuller, R., Acosta, N.J.R., Adeyi, O., Arnold, R., Basu, N., Balde, A.B., Bertollini, R., Bose-O'Reilly, S., Boufford, J.I., Breysse, P.N., Chiles, T., Mahidol, C., Coll-Seck, A.M., Cropper, M.L., Fobil, J., Fuster, V., Greenstone, M., Haines, A., Hanrahan, D., Hunter, D., Khare, M., Krupnick, A., Lanphear, B., Lohani, B., Martin, K., Mathiasen, K.V., McTeer, M.A.,

Murray, C.J.L., Ndahimananjara, J.D., Perera, F., Potocnik, J., Preker, A.S., Ramesh, J., Rockstrom, J., Salinas, C., Samson, L.D., Sandilya, K., Sly, P.D., Smith, K.R., Steiner, A., Stewart, R.B., Suk, W.A., van Schayck, O.C.P., Yadama, G.N., Yumkella, K. and Zhong, M. (2018). The Lancet Commission on pollution and health. *Lancet*, **391**(10119): 462-512.

Leusch, F.D.L., De Jager, C., Levi, Y., Lim, R., Puijker, L., Sacher, F., Tremblay, L.A., Wilson, V.S. and Chapman, H.F. (2010). Comparison of five *in vitro* bioassays to measure estrogenic activity in environmental waters. *Environmental Science & Technology*, **44**(10): 3853-3860.

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., Neale, P.A., Hebert, A., Scheurer, M. and Schriks, M.C.M. (2017). Analysis of the sensitivity of *in vitro* bioassays for androgenic, progestagenic, glucocorticoid, thyroid and estrogenic activity: Suitability for drinking and environmental waters. *Environment International*, **99**: 120-130.

Lv, X.M., Xiao, S.H., Zhang, G., Jiang, P. and Tang, F. (2016). Occurrence and removal of phenolic endocrine disrupting chemicals in the water treatment processes. *Scientific Reports*, **6**: 1-10.

Malaj, E., von der Ohe, P.C., Grote, M., Kuhne, R., Mondy, C.P., Usseglio-Polatera, P., Brack, W. and Schafer, R.B. (2014). Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(26): 9549-9554.

Medlock Kakaley, E.K., Blackwell, B.R., Cardon, M.C., Conley, J.M., Evans, N., Feifarek, D.J., Furlong, E.T., Glassmeyer, S.T., Gray, L.E., Hartig, P.C., Kolpin, D.W., Mills, M.A., Rosenblum, L., Villeneuve, D.L. and Wilson, V.S. (2020). De Facto Water Reuse: Bioassay suite approach delivers depth and breadth in endocrine active compound detection. *Science of the Total Environment*, **699**: 134297.

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., Altenburger, R., Ait-Aissa, S., Brion, F., Busch, W., Umbuzeiro, G.D., Denison, M.S., Du Pasquier, D., Hilscherova, K., Hollert, H., Morales, D.A., Novak, J., Schlichting, R., Seiler, T.B., Serra, H., Shao, Y., Tindall, A.J., Tollefsen, K.E., Williams, T.D. and Escher, B.I. (2017). Development of a bioanalytical test battery for water quality monitoring: Fingerprinting identified micropollutants and their contribution to effects in surface water. *Water Research*, **123**: 734-750.

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. and Leusch, F.D.L. (2015). Considerations when assessing antagonism *in vitro*: Why standardizing the agonist concentration matters. *Chemosphere*, **135**: 20-30.

Neale, P.A., Leusch, F.D.L. and Escher, B.I. (2020c). Effect Based Monitoring in Water Safety Planning - WP3.2 Medium-to-high throughput bioanalytical tools and decision-making tool for selection of bioassays. London, UK.

Neale, P.A., Leusch, F.D.L. and Escher, B.I. (2020d). Effect Based Monitoring in Water Safety Planning - WP3.3 Sampling strategies and sample pre-treatment options and decision-making tool for selection of sampling methods. London, UK.

Neale, P.A., Leusch, F.D.L. and Escher, B.I. (2020e). Effect Based Monitoring in Water Safety Planning - WP3.4 Effect-based trigger values for different water quality classes considering hazards for human and the environment health London, UK.

Neale, P.A., O'Brien, J.W., Glauch, L., Konig, M., Krauss, M., Mueller, J.F., Tscharke, B. and Escher, B.I. (2020f). Wastewater treatment efficacy evaluated with *in vitro* bioassays. *Water Research X*, **9**.

Nivala, J., Neale, P.A., Haasis, T., Kahl, S., Konig, M., Muller, R.A., Reemtsma, T., Schlichtinge, R. and Escher, B.I. (2018). Application of cell-based bioassays to evaluate treatment efficacy of conventional and intensified treatment wetlands. *Environmental Science-Water Research & Technology*, **4**(2): 206-217.

OECD (2018). Guidance Document on Good *In Vitro* Method Practices (GIVIMP), OECD Series on Testing and Assessment, No. 286. OECD Publishing, Paris, https://doi.org/10.1787/9789264304796-en.

Reungoat, J., Macova, M., Escher, B.I., Carswell, S., Mueller, J.F. and Keller, J. (2010). Removal of micropollutants and reduction of biological activity in a full scale reclamation plant using ozonation and activated carbon filtration. *Water Research*, **44**(2): 625-637.

Rosenmai, A.K., Lundqvist, J., le Godec, T., Ohisson, A., Troger, R., Hellman, B. and Oskarsson, A. (2018). *In vitro* bioanalysis of drinking water from source to tap. *Water Research*, **139**: 272-280.

Shi, P., Zhou, S.C., Xiao, H.X., Qiu, J.F., Li, A.M., Zhou, Q., Pan, Y. and Hollert, H. (2018). Toxicological and chemical insights into representative source and drinking water in eastern China. *Environmental Pollution*, **233**: 35-44.

Silva, E., Rajapakse, N. and Kortenkamp, A. (2002). Something from "nothing" - Eight weak estrogenic chemicals combined at concentrations below NOECs produce significant mixture effects. *Environmental Science & Technology*, **36**(8): 1751-1756.

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 Linden, S.C., Heringa, M.B., Man, H.Y., Sonneveld, E., Puijker, L.M., Brouwer, A. and Van der Burg, B. (2008). Detection of multiple hormonal activities in wastewater effluents and surface water, using a panel of steroid receptor CALUX bioassays. *Environmental Science & Technology*, **42**(15): 5814-5820.

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.

Villeneuve, D.L., Blankenship, A.L. and Giesy, J.P. (2000). Derivation and application of relative potency estimates based on *in vitro* bioassay results. *Environmental Toxicology and Chemistry*, **19**(11): 2835-2843.

Wagner, E.D. and Plewa, M.J. (2017). CHO cell cytotoxicity and genotoxicity analyses of disinfection by-products: An updated review. *Journal of Environmental Sciences*, **58**: 64-76.

Walter, H., Consolaro, F., Gramatica, P., Scholze, M. and Altenburger, R. (2002). Mixture toxicity of priority pollutants at no observed effect concentrations (NOECs). *Ecotoxicology*, **11**(5): 299-310.

Wang, Z.Y., Walker, G.W., Muir, D.C.G. and Nagatani-Yoshida, K. (2020). Toward a global understanding of chemical pollution: A first comprehensive analysis of national and regional chemical inventories. *Environmental Science & Technology*, **54**(5): 2575-2584.

Wernersson, A.S., Carere, M., Maggi, C., Tusil, P., Soldan, P., James, A., Sanchez, W., Dulio, V., Broeg, K., Reifferscheid, G., Buchinger, S., Maas, H., Van Der Grinten, E., O'Toole, S., Ausili, A., Manfra, L., Marziali, L., Polesello, S., Lacchetti, I., Mancini, L., Lilja, K., Linderoth, M., Lundeberg, T., Fjallborg, B., Porsbring, T., Larsson, D.G.J., Bengtsson-Palme, J., Forlin, L., Kienle, C., Kunz, P., Vermeirssen, E., Werner, I., Robinson, C.D., Lyons, B., Katsiadaki, I., Whalley, C., den Haan, K., Messiaen, M., Clayton, H., Lettieri, T., Carvalho, R.N., Gawlik, B.M., Hollert, H., Di Paolo, C., Brack, W., Kammann, U. and Kase, R. (2015). The European technical report on aquatic effect-based monitoring tools under the water framework directive. *Environmental Sciences Europe*, **27**: 1-11.

Appendix: Generic SOP for sample collection and sample processing

This SOP for sample collection and processing was developed based on WP3.3 and the decision-making tool shown in Section 2.1.2. The SOP focuses on solid-phase extraction (SPE) as this is the most used sample extraction method based on the literature review conducted in WP3.2. Further, the SOP is intended as a generic protocol, rather than providing specific guidance for a particular SPE sorbent or specific conditioning and elution solvents.

Hazards associated with sample collection and processing: Wastewater samples can contain biological hazards. Staff should have appropriate vaccinations, wear personal protective equipment (e.g., gloves, lab coat, safety glasses) and conform with local health and sanitation recommendations. Hazardous chemicals, including hydrochloric acid and solvents used for conditioning and extraction, are required for sample processing and staff should consult the safety data sheets and any risk assessments before extraction. The maximum recommended vacuum on the SPE manifold should not be exceeded.

Requirements
Equipment
Turbidity meter
Filtration unit
SPE manifold and tubing or automated SPE
Evaporator to reduce solvent extract volume
Glassware (e.g., measuring cylinders, beakers, glass serological pipettes)
pH meter (Note: this is optional, pH indicator paper can also be used)

Consumables
Amber glass bottles for sample collection
SPE cartridge (e.g., Oasis HLB, StrataX, Chromabond HR-X)
Pasteur pipettes
pH indicator paper
Glass fibre filters (0.7 to 1.5 µm pore size)
Glass test tubes

Chemicals and solvents
Hydrochloric acid
Sodium thiosulphate or ascorbic acid
High-performance liquid chromatography (HPLC) grade solvents for conditioning and elution (e.g., methanol, hexane:acetone, ethyl acetate)
Ultrapure water
Note: Clean all equipment (e.g., SPE manifold, tubing, filtration unit) that will come into contact with the samples prior to sample collection and sample processing using the solvents used for SPE conditioning and elution.

Sample collection and pre-treatment
On-site

- Collect water samples in solvent-washed amber glass bottles. The volume to be collected will depend on the water type and the number of assays to be run. A 200 mg SPE cartridge allows the extraction of 0.5 L of wastewater influent, 1 L of wastewater effluent or surface water and 2 L of drinking water, highly treated recycled water or clean surface water. Double the volume can be extracted on 500 mg SPE cartridges.
- Grab sampling is sufficient if sampling surface water, recycled water and drinking water. 24 h composite samples are recommended for wastewater influent and wastewater effluent due to diurnal variations in the concentrations of some micropollutants.

- If transporting the water samples back to the laboratory for extraction, adjust the pH to 3 on-site to reduce degradation. pH adjustment is not necessary if extracting on-site (unless wanting to improve the extraction of weak acids) or using large volume SPE (LVSPE). Adjust the pH by adding drops of concentrated hydrochloric acid using a glass Pasteur pipette and checking the pH using pH indicator paper or a pH meter.
- If the water sample is chlorinated, quench free Cl_2 using sodium thiosulphate or ascorbic acid. 3.5 mg/L of sodium thiosulphate is required to quench 1 mg/L Cl_2, while 5 mg/L of ascorbic acid is required to quench 1 mg/L Cl_2. Free Cl_2 can be measured using the US EPA DPD (N,N-diethyl-p-phenylenediamine) method or online sensors.
- Store the water samples on ice and in the dark and return to the laboratory as soon as possible.

Laboratory

- Once in the laboratory, water samples can be stored at 4˚C if extracting within 48 h of sampling. The samples should be stored at -20˚C if the extraction cannot be performed within 48 h of sampling.
- If a water sample contains visible particles, it should be filtered prior to SPE to prevent clogging of the SPE cartridge. We suggest that water samples with a turbidity of 5 nephelometric turbidity units (NTU) should be filtered using glass fibre filters with a pore size between 0.7 to 1.5 µm. Samples with a turbidity less than 5 NTU can be extracted by SPE without filtration. If filtration is required and the purpose of the study is to capture the effects from the whole-water sample, suspended particulate matter in the water sample can be collected on the filter and extracted separately using solvents.
  Note: The filtration unit should be cleaned with the solvents used for conditioning and elution between each sample that is filtered.
- In addition to the water samples, it is also important to include blank samples, such as a field blank and a laboratory blank. The field blank is the same volume of ultrapure water that has been taken into the field and exposed to the same conditions as the water samples, such as temperature in the field and travel time, and processed with the collected water samples. The laboratory blank is ultrapure water that has been processed using the same sample processing procedures as the water samples.

Sample extraction

- Measure the sample volume or weigh the bottle prior to extraction to determine the volume of sample extracted. If weighing the bottle, also remember to weigh the empty bottle after extraction and subtract this from the weight of the bottle prior to SPE to calculate the sample volume.
- Label SPE cartridges with the sample ID names and load them onto the SPE manifold and open the vacuum valves on the manifold. Automated SPE systems can also be used for sample extraction, though this guidance is focused specially on manual SPE as this is most commonly applied.
- Condition the SPE cartridges under gravity. For a 200 mg or 500 mg SPE cartridge, add 2 × 5 mL of each solvent to the cartridge using glass serological pipettes or glass measuring cylinders. If using more than one solvent (e.g., hexane:acetone and methanol or ethyl acetate and methanol), add the least polar solvent first (e.g., hexane:acetone and ethyl acetate in the examples above) and let it drip through before adding the second solvent. After the solvents have dripped through, add 2 × 5 mL ultrapure water, which should be adjusted to pH 3 if the water samples were also adjusted to pH 3. Just before all the ultrapure water has dripped through, close the vacuum valve, and fill the SPE cartridge reservoir with ultrapure water.
  Note: The same solvents should be used for both conditioning and elution.
- Connect the SPE manifold to a vacuum trap (i.e., waste bottle) and connect the vacuum trap to the vacuum inlet.
- Attach tubing to each SPE cartridge and place the other end of the tubing in the corresponding sample bottle.
- Turn the vacuum on and slowly open the vacuum values on the SPE manifold. The water should drip through the SPE manifold with a flow rate of approximately 10 mL/min. You should be able to see separate drops and not a constant stream. It is important to monitor the flow and ensure that the cartridges do not run dry during extraction.
  Note: Make sure the vacuum trap does not overfill during extraction. It may need to be emptied regularly depending on the number of samples being extracted and the size of the vacuum trap.
- After the entire sample has passed through the cartridge, turn off the vacuum and remove the tubing. Turn the vacuum back on and leave the cartridges to dry for 1 to 2 hours or until dry. The drying time will depend on the cartridge size,

with larger sorbent beds requiring more time. It is very important the cartridges are completely dry prior to elution, which is indicated by a colour change and the sorbent bed appearing powdery.

- Once dried, the SPE cartridges can be eluted straight away or can be wrapped in parafilm and aluminium foil and stored in -20˚C until ready for elution.
- Disconnect the SPE manifold and clean all equipment with the solvents used for conditioning and elution.

Sample elution

- If frozen for storage, allow the SPE cartridges to defrost at room temperature for approximately 30 minutes prior to elution.
- Connect the SPE manifold to a vacuum trap (i.e., waste bottle) and connect the vacuum trap to the vacuum inlet.
- Load SPE cartridges on the SPE manifold. In the case of SPE cartridges stored at -20˚C, turn on the vacuum to ensure that the cartridges are completely dry before elution. Turn off the vacuum once dried.
- Label glass test tubes with the sample ID names and place them in a rack inside the SPE manifold, with the labelled test tube directly below the corresponding SPE cartridge. If eluting with more than one solvent (e.g., hexane:acetone and methanol or ethyl acetate and methanol) two test tubes may be required per sample.
- Under gravity, add solvent to the SPE cartridge. Using the example of a 200 mg or 500 mg SPE cartridge, add 2 × 5 mL of each solvent to the cartridge using glass serological pipettes or glass measuring cylinders. Once the majority of the first solvent has passed through, the vacuum can be gently turned on to extract any remaining solvent from the sorbent bed. If eluting with additional solvents, replace the filled test tubes in the rack with new labelled test tubes, then add 2 × 5 mL of the second solvent to the cartridge as described above.
- Seal the full test tubes with parafilm and discard the SPE cartridges.
- Disconnect the SPE manifold and clean all equipment with solvents used for conditioning and elution.

Sample evaporation

- Evaporate the solvent extracts in the test tubes using either a nitrogen evaporator or centrifugal evaporation.
- Once the solvent volume reaches around 0.5 mL pool the contents of the test tubes from the same water sample into one test tube using a glass Pasteur pipette. Rinse the previous test tubes with 3 × 1 mL methanol and add this to the pooled extract.
- Evaporate the pooled solvent extract down to around 0.5 mL and transfer to a labelled 2 mL glass HPLC vial. Rinse the test tube with 3 × 1 mL methanol and add this to the 2 mL HPLC vial. As all the methanol from rinsing will not fit into the HPLC vial at once, you will need to evaporate some of the extract and then add the remaining methanol.
- Evaporate to dryness, being careful to stop drying as soon as the last of the solvent has evaporated, and resuspend in the final bioassay solvent, such as methanol or DMSO. For a 200 mg or 500 mg SPE cartridge extracting 0.5 to 2 L of water, a final extract volume of 0.5 or 1 mL is common.
- Store labelled HPLC vials in a box at -20˚C.