Trends in Analytical Chemistry 102 (2018) 343–358

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

Screening and risk management solutions for steroidal estrogens in surface and wastewater

Robert Kase [a, *], Barbora Javurkova [b], Eszter Simon [a], Kees Swart [c], Sebastian Buchinger [d], Sarah Könemann [a, e], Beate I. Escher [f, g], Mario Carere [h], Valeria Dulio [i], Selim Ait-Aissa [i], Henner Hollert [e], Sara Valsecchi [j], Stefano Polesello [j], Peter Behnisch [c], Carolina di Paolo [e], Daniel Olbrich [a], Eliska Sychrova [b], Michael Gundlach [e], Rita Schlichting [f, g], Lomig Leborgne [k], Manfred Clara [l], Christoph Scheffknecht [m], Yves Marneffe [n], Carole Chalon [n], Petr Tusil [o], Premysl Soldan [o], Brigitte von Danwitz [p], Julia Schwaiger [q], Antonio Moran Palao [r], Francesca Bersani [s], Olivier Perceval [t], Cornelia Kienle [a], Etienne Vermeirssen [a], Klara Hilscherova [b], Georg Reifferscheid [d], Inge Werner [a]

[a] Swiss Centre for Applied Ecotoxicology, Eawag-EPFL, Ueberlandstrasse 131, 8600 Dübendorf, Switzerland
[b] Research Centre for Toxic Compounds in the Environment (RECETOX), Kamenice 753/5, Pavillion A29, 62500 Brno, Czech Republic
[c] BioDetection Systems b.v., Science Park 406, 1098 XH Amsterdam, Netherlands
[d] Bundesanstalt für Gewässerkunde, Am Mainzer Tor 1, 56068 Koblenz, Germany
[e] Institute for Environmental Research, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany
[f] Helmholtz Centre for Environmental Research (UFZ), Permoserstrasse 15, 04318 Leipzig, Germany
[g] Eberhard Karls University Tuebingen, Hoelderlinstr. 12, 72074 Tuebingen, Germany
[h] Italian Institute of Health, Department Environment and Health, Roma, Italy
[i] INERIS, Rue Jaques Taffanel, Parc Technologique ALATA, 60550 Verneuil-en-Halatte, France
[j] Water Research Institute - National Research Council of Italy, Istituto di Ricerca sulle Acque, Via del Mulino 19, 20861 Brugherio (MB), Italy
[k] Agence de l'eau Adour-Garonne, 90 rue de Fétéra, CS 87801, 31078 Toulouse Cedex 4, France
[l] Umweltbundesamt, Spittelauer Lände 5, 1090 Wien, Austria
[m] Umweltinstitut, Institut für Umwelt und Lebensmittelsicherheit des Landes Vorarlberg, Montfortstraße 4, 6901 Bregenz, Austria
[n] Institut Scientifique de Service Public (ISSeP), Rue Chéra 200, 4000 Liège, Belgium
[o] T.G. Masaryk Water Research Institute, Podbabská 2582/30, Praha 6, 16000, Czech Republic
[p] Landesamt für Natur, Umwelt und Verbraucherschutz NRW (LANUV), Auf dem Draap 25, 40221 Düsseldorf, Germany
[q] Bayerisches Landesamt für Umwelt, Demollstr. 32, 82407 Wielenbach, Germany
[r] Instituto de Recursos Naturales, Universidad de León, Avenida de Portugal 42, 24071 León, Spain
[s] Centro Ricerche SMAT, Società Metropolitana Acque Torino S.p.A., C.so Unità d'Italia 235/3, 10127 Torino, Italy
[t] The French Agency for Biodiversity, 94300 Vincennes, Paris, France

ARTICLE INFO

Article history:
Available online 8 March 2018

Keywords:
Effect-based methods
Science-policy interface
Oestrogen screening
Endocrine disruption
Surface and wastewater assessment
Risk-based trigger values
EU watch-list
Water framework directive
Chemical monitoring of Emerging pollutants

ABSTRACT

Background: The European Commission Implementing Decision EU 2015/495 included three steroidal estrogens, namely 17α-ethinyl estradiol, 17β-estradiol, and estrone, in the so-called "watch list" of the EU Water Framework Directive (WFD). The monitoring of these compounds is difficult because the detection limits of the majority of the available analytical methods cannot achieve the very low target concentrations required to meet proposed environmental quality criteria. In 2014, a combined Science-Policy Interface/Chemical Monitoring of Emerging Pollutants project was launched to meet this monitoring challenge. The project involved 24 research organizations and environmental agencies from 12 different countries.

Methods: Sixteen surface water (SW) and 17 wastewater (WW) samples were collected across Europe and analysed using five *in vitro* effect-based and three chemical analytical methods. A general description of the project and data evaluation is provided by Könemann and colleagues in the companion publication 2018. In our study, we compared bioanalytical and chemical analytical results with regard to their application for aquatic status assessment. Therefore we considered the potential to predict population-relevant risks for aquatic organisms and the specificity and sensitivity of the various methods used in

* Corresponding author. Swiss Centre for Applied Ecotoxicology Eawag-EPFL, Überlandstrasse 133, 8600 Dübendorf, Switzerland.
E-mail addresses: Robert.Kase@oekotoxzentrum.ch, kase.robert@googlemail.com (R. Kase).

https://doi.org/10.1016/j.trac.2018.02.013

both approaches. Finally, we tested and discussed the applicability and relevance of previously suggested effect-based trigger values (EBT).

Results and discussion: Results of the risk assessment based on chemical analytical data correlated highly with estrogenic activities (expressed as 17β-estradiol equivalents (EEQ) determined using effect-based methods), demonstrating the ability of the bioassays to predict the mixture risk caused by steroidal estrogens. For about 15% of SW and 40% of WW samples detection limits of chemical-analytical methods were too high to allow a status assessment, while detection limits of all effect-based methods were below proposed EBT. This demonstrates that effect-based methods are suitable for status assessment of surface waters. The *in vitro* effect-based methods were quite specific for steroidal estrogens and highly sensitive, meaning they have a low probability to detect false positive or negative results. After testing of three alternative EBT proposals, we concluded to use preliminary 400 pg/L EEQ as screening EBT corresponding to the AA-EQS of E2. Further test specific refinements in the application of this value are not excluded.

Conclusions: We conclude that water quality assessment can progress from a purely analytical approach to effect-based monitoring, from single substance to known and unknown mixture assessment and from *in vitro* screening to population-relevant risk assessment. Despite a few limitations, effect-based *in vitro* methods are recommendable for monitoring steroidal estrogens under the WFD because they, a) are able to sensitively quantify the activity of steroidal estrogens in all surface and wastewater samples, b) are able to detect the combined effect of estrogen mixtures including unknown chemicals with estrogen receptor activating properties, c) allow an ecotoxicological status assessment using EBT to calculate risk quotients. This approach is similar to the risk ratio used in regulatory environmental risk assessments, but allows for an integrated mixture assessment.

Outlook: The results of this study support the introduction of a holistic approach for the regulation of chemicals in the aquatic environment under the EU WFD, as proposed recently by EU Water Directors. An ecotoxicological status assessment for one of the most relevant modes of action of endocrine disruption will allow authorities responsible for water quality assessment to focus available monitoring resources and to improve the ecological status of EU waterbodies.

1. Introduction

1.1. State of the art and current regulation

Steroidal estrogens, which are commonly present in wastewaters (WW) and surface waters (SW) (e.g. Ref. [1]), can cause reproductive toxicity to aquatic biota, especially to fish [2–6]. The most potent steroidal estrogens: the synthetic hormone 17α-ethinyl estradiol (EE2), the natural hormones 17β-estradiol (E2), and estrone (E1) the main transformation product of both, were recently included in the European Union's watch list [7–9] of substances for monitoring in SW. This watch-list mechanism is designed to allow target EU-wide monitoring of substances of possible concern to support the prioritization process in future reviews of the priority substances list under the Water Framework Directive (WFD). According to the WFD, the good chemical status of a waterbody is reached when all substances included in the priority list are detected below their defined Environmental Quality Standards (EQS) that are based on annual average and maximum allowable concentrations. The previously proposed and not regulatory binding annual average Environmental Quality Standards (AA-EQS) for E1, E2 and EE2 are as low as 3600 pg/L, 400 pg/L and 35 pg/L, respectively [10,11]. So far, only a very limited number of institutes from both the public and the private sector, and major environmental agencies in Europe have developed capacities to quantify the steroidal estrogens at their EQS levels. Such low EQS, together with complexity of matrices such as SW and WW, and the instability of some of the analytes, make the monitoring of these compounds under the watch-list mechanism of the WFD and national surveillance schemes difficult.

Estrogenic adverse effects in environmental watersare generally caused by mixtures of different estrogenic chemicals including metabolites, which increases their risk to wildlife [12]. A recent study demonstrated that mixtures of priority pollutants present at and below their individual EQS concentrations can cause relevant biological effects and may pose significant risks to wild species and ecosystems in spite of the fact that individual chemicals were at concentrations in compliance with regulations [13]. Mixture toxicity has also been highlighted in the context of the European strategy on endocrine disrupting chemicals [14]. The European Commission recently acknowledged the need to develop and implement methodologies for the identification of chemical mixtures of potential concern and for the assessment of their impacts on both environmental and human health [15]. Such methodologies should help to link the knowledge of chemical contamination and the observation of adverse effects on and via the aquatic environment. For all described reasons (risks at low concentrations, difficulty and costs of high-end analytics, mixture toxicity, and linking chemical contamination to ecological status) alternative monitoring and risk assessment methods are urgently needed.

Effect-based methods detect cumulative effects and are useful to bridge the gap between chemical contamination and ecological status [16,17]. Complementary, mechanism-specific bioassays can provide information on modes of action (MoA) that are intrinsically of concern for ecosystems and human health [18]. Focusing on estrogenic effects, *in vitro* bioassays, can detect an activation of estrogenic receptor(s) (ER) by mixtures of estrogens and xenoestrogens: In parallel, they can detect single analytes at sufficiently low concentration levels [19]. The response of the assays is expressed as E2-equivalent (EEQ) values. The applicability of this approach has been demonstrated in different projects during the past decade e.g. Refs. [20–23]. Also the combination of effect-based and chemical analytical monitoring to identify and assess the risks from steroidal estrogens has been discussed and proposed as a potential tool for WFD monitoring [24–29].

1.2. Current trends

The WFD also involves evaluation of the ecological status of waterbodies, e.g. monitoring of biological diversity. Consequently, severe adverse effects at the population level of aquatic organisms should be captured, but there are currently no tools under WFD implemented to monitor Endocrine Disruption efficiently. The current proposal for the WFD [26–29a,b] is to screen water samples by *in vitro* assays for estrogenic activity and subsequently target the more demanding chemical monitoring on a reduced number of samples that show up as positive in the bioanalytical screening, as was previously elaborated by two international workshops in 2013 and 2017 [29a,b]. The water status evaluated with *in vitro* bioassays would be newly called an ecotoxicological status for risk from ER-mediated effects and is intended to improve chemical and ecological status assessment for one of the most relevant modes of action of endocrine disruption and fish reproduction toxicity [4]. As described, numerous studies recommend the use of these effect-based methods. However, clear recommendations for water managers regarding the use of *in vitro* methods, especially as regards harmonized data evaluation and effect-based trigger values (EBT) distinguishing “acceptable” and “not acceptable” water quality are still missing.

Moreover, these methods with their clearly defined EBT are also necessary for wastewater management. Wastewaters or Wastewater Treatment Plants (WWTPs) effluents are the dominant sources of steroidal estrogens in waterbodies, e.g. Refs. [30,31]. Therefore, although levels of estrogens in WW and WWTPs effluents are not directly regulated by the WFD, their monitoring is necessary to quantify the loads emitted into waterbodies and to show reduction of risks after having applied proper mitigation actions. Various options for the removal of pharmaceuticals and hormones by WWTPs were recently reviewed [24,32]. Although elimination rates of estrogenicity (caused mainly by steroidal estrogens) at conventional WWTPs with tertiary treatment are high (usually from about 80% to more than 90%, [33,34]) the residual activity in discharges can still represent a risk for aquatic biota and remains, besides untreated WW, their major known source [35]. Advanced treatment steps such as e.g. ozonation, UV treatment etc. can further eliminate well above 90% of steroidal estrogens [32]. Under such circumstances, concentrations in surface waters cannot be measured by most currently available chemical analytical methods due to matrix effects [36].

To test if *in vitro* assays may be suited for regulatory monitoring and risk assessment of low levels of estrogens in both WW and SW, 24 research organizations and environmental agencies from 12 different European countries joined forces in a project, which also supports the activities of the Working Group “Chemicals” under the Common Implementation Strategy (CIS) for the WFD and as a follow-up to their “Science-Policy Interface” activity [17,37]. The project results are summarized in two publications: the general description of the project, experimental details and recommendations on harmonized *in vitro* data evaluation, are described in a companion paper by Könemann et al. [36]. The current publication focuses on testing the most useful EEQ EBT for screening and discussion about environmental risk for SW and WW. A demonstration of applicability of the proposed EBT value to 16 SW and 17 WW samples collected across Europe is also presented here. The samples represented a gradient from low polluted to highly polluted samples and were analysed by sensitive HPLC-MS/MS methods at three different institutes and by five different *in vitro* assays at five different institutes in order to:

- Evaluate the applicability and relevance of the *in vitro* methods for the monitoring of steroidal estrogens with reference to the classification of the chemical status of waterbodies
- Discuss and recommend the extent to which the bioassays can be used for screening and prioritizing environmental samples, while considering risks for aquatic organisms
- Propose suitable effect-based trigger values (EBT) for screening and discriminating between unpolluted and polluted samples with the aim of classifying waterbodies.
- Contribute towards the review process of WFD and to integrated effect-based-methods (EBM) into regulation

Within this project, we aimed to bridge the gap between conventional analytical and effect-based monitoring and risk assessment for steroidal estrogens.

2. Methods

Sampling, sample preparation, positive and negative controls, and chemical and biological analyses methods, are described in detail in Ref. [36] and are briefly summarized here.

2.1. Samples, sample preparation, in vitro and chemical analyses

A total of 16 SW and 17 WW samples with 11 L sampling volume were collected by 10 participating institutes at sites expected to be polluted from Austria, Belgium, Czech Republic, Germany, France, Italy and Spain. The samples were frozen within 48 h and sent to Bio Detection Systems (BDS), Amsterdam, The Netherlands. Samples were filtered (see Könemann et al., 2018) and subsequently extracted by using solid phase extraction with C18 cartridges (Phenomenex Strata C18-E, 55 μm, 70 Å, 500 mg/6 mL). Additional silica gel clean-up was applied to the extracts to reduce matrix effects and reach detection limits in the sub ng/L range in the chemical analysis. For some wastewater samples, a single silica gel (SiOH) column was inadvertently used to treat the entire sample extract (11 mL), while for each surface water sample, extract was split into eleven 1 mL aliquots for clean-up. Extracts were then homogenized, divided into 1 mL aliquots, and sent to Federal Institute of Hydrology (BfG), Institut National de l'Environnement Industriel et des Risques (INERIS), Research Centre for Toxic Compounds in the Environment (RECETOX), and Helmholtz Centre for Environmental Research (UFZ), where analyses by five *in vitro* effect-based methods were performed: ER-Calux (at BDS), pYES, MELN, HeLa 9903, and ER-GeneBlazer and to Joint Research Centre (JRC), BfG, and Swiss Centre for Applied Ecotoxicology (Ecotox Centre, EC) performing HPLC MS/MS analysis. EC used an additional silica gel clean-up for 3 of 17 WW samples prior to the chemical analysis. More detailed information on methods is available in the companion publication Könemann et al., 2018 [36]. All data were analysed centrally in a harmonized way.

2.2. Chemical analytical data evaluation and compliance assessment

Measured concentrations for E1, E2, and EE2 were expressed in pg/L. Measurements below LOQ, but above LOD were indicated as < LOQ. Measurements below LOD were indicated as < LOD (SI Tables 1–3).

The measurements were compared with EQS proposals (AA-EQS EE2 = 35 pg/L, AA-EQS E2 = 400 pg/L, AA-EQS E1 = 3600 pg/L) in order to assess potential compliance. Because the European Commission did not propose an EQS for E1 [10] the Swiss EQS proposal for E1 of 3600 pg/L [11] was used. The compliance of samples was set to 0 if the measurement result exceeded an EQS proposal. Compliance was designated as “not assessable” if the results were below LOD, with an LOD (LOQ/3) above the EQS proposal. If EQS proposals were not exceeded and LODs were below the proposed

EQS, the compliance assessment was set to 1 (compliant). For comparison reasons, the results for WW samples were treated the same way as the SW samples, even though WW discharges are usually diluted by the receiving waters. Only if concentrations of estrogens in the flow of the receiving waterbody and the effluent are known, dilution factors can be applied to calculate a final estrogenicity (Equation (8) in chapter 3.4.7.2) and the receiving water can be assessed as compliant even though the WW could be "non-compliant". On the other hand in our project the concentrations of estrogens in the receiving waters were not measured and the precautionary principle (to set the same quality requirement for WW and SW) was applied in order to avoid the risk of a false negative assessment (extrapolation of compliance in case of non-compliance) for the receiving waterbody. Our approach intends to stimulate the consideration and measurements of background concentrations without only focusing on the estrogenicity of the WW discharges. The overall compliance corresponds to the rounded average compliance to 0 or 1 of all three analytical assessments.

2.3. Effect-based trigger values (EBT)

In recently published studies, a narrow concentration range of published EBT values was proposed, although different approaches were applied to derive them. The following shortly described EBT were tested and discussed in our study in section 3.4.4:

Jarošová et al. [33] derived "safe environmental concentrations" of EEQ in municipal wastewater effluents, based on a simplified assumption that mainly (>90%) steroidal estrogens are causing ER-mediated estrogenicity. These potentially safe concentrations were derived using the estrogenic relative potencies in bioassays, the *in vivo* predicted no-effect concentrations of the compounds, and their relative contributions to the measured EEQ of WW effluents. The predicted safe concentrations ranged from 100 to 400 pg/L EEQ with a median EBT of 300 pg/L EEQ. We used in our study the median EBT because we worked with five different bioassays which we intended to characterize with alternative EBT proposals.

Van der Oost at al [38]. used bioanalytical equivalents (BEQ) of selected substances that trigger the bioassay and a background BEQ to derive an EBT of 500 pg/L EEQ. The background BEQ was calculated with 60 pg EEQ/L, and a safe BEQ (based on lowest NOEC of triggering substances) was 7 pg/L EEQ. The finally proposed EBT (500 pg/L EEQ) was mainly based on a BEQ-converted species sensitivity distribution (SSD) that provided the concentration that is a potential hazard for 5% of aquatic species (HC_5 BEQ = 500 pg/L EEQ).

Kase et al. and Kunz et al. [25,26] proposed to use the proposed AA-EQS of E2 as an EBT for estrogen receptor mediated estrogenic activity, thus proposed 400 pg/L EEQ as EBT. This was done primarily for different reasons: a) The EBT is compatible with the EU AA-EQS proposal for E2 which is based on fish toxicity SSD for population relevant effects 400 pg/L EEQ, b) E2 is a natural steroid hormone and has an *in vitro* and *in vivo* potency between E1 and EE2, therefore it is likely better suited than E1 or EE2 for assessing mixture effects, c) E2-equivalents are commonly used in bioanalysis and biomonitoring, thus data are easily comparable with previous studies, d) EE2 has a slightly higher potency *in vitro* than E2, but *in vivo* it is 10–20 times more potent. If EE2 equivalents were to be used, there is a high probability for risk overestimation and obtaining false positive results, due to the possibility of E2 and E1 playing a more prominent role. e) Steroidal estrogens normally occur as a mixture in WW and in receiving waterbodies. Jarosova and colleagues [33] compiled data of 353 wastewater measurements from three studies with a median concentration of 7–12 ng/L E1, 1.3–1.7 ng/L E2 and 0.47–0.6 ng/L EE2. Based on relative potencies of used bioassays this means that the *in vitro* ER mediated mixture effect is likely dominated by E2. Therefore as a simplified approach the use of 400 pg/L EEQ as EBT seem to be arbitrary, however this can be justified and was successfully tested in this study.

2.4. Effect-based compliance assessment and validation status of methods

The measured EEQs were compared with EBTs to assess the potential compliance of samples. The compliance of samples was set to 0 ("non-compliant") if the measured EEQ exceeded an EBT. The compliance assessment was set to "not assessable" if LOQ or LOD were above the EBT, but this never occurred during the measurements. If EBT were not exceeded and LOD were below EBT the compliance was set to 1 ("compliant"). The overall compliance corresponds to the rounded average compliance to 0 or 1 of all five effect-based assessments. After EBT discussion in section 3.4.4 we used the 400 pg/L EEQ as preliminary EBT for compliance assessment.

Validation activities: Two of our five *in vitro* assays used in this study are currently being OECD validated (HeLa 9903 and ER-Calux [39]) and the ER-Calux, A-YES and L-YES are DIN/EN/ISO standardized in 2018. The ER-ER-GeneBlazer is used in the US within the Tox21 program of the National Institute of Health and US Environmental Protection Agency [40].

2.5. Sensitivity and specificity analysis for methods in compliance assessments

Sensitivity and specificity were determined for both chemical and effect-based techniques. This was done to evaluate their suitability for European WFD monitoring programs. High sensitivity means that the method is less prone to detect false negatives, in other words the method has a low risk of erroneous compliant assessments. While high specificity means that the method is less prone to detect false positives, in other words the method has a low risk of erroneous non-compliant assessments.

In order to assess the sensitivity and specificity of three chemical analytical and five *in vitro* effect-based methods, each method was compared with the overall compliance derived from the chemical or effect-based analyses (see section 2.2 and 2.4)

Four options regarding conformity in the results of the comparison of the specific method with the overall compliance were possible:

- Overall compliance is 1 and the compliance of the specific method is 1, this means "conformity in compliance ":= CC
- Overall compliance is 1 and the compliance of the specific method is 0, this means "non-conformity in compliance":= NCC
- Overall compliance is 0 and the compliance of the specific method are 0, this means "conformity in non-compliance":=CNC
- Overall compliance is 0 and the compliance of the specific method is 1, this means "non-conformity in non-compliance":=NCNC

Based on the conformity rating of each method sensitivity and specificity was calculated according to Equations (1) and (2).

$$\text{Sensitivity } [\%] = \frac{\sum \text{CNC}}{(\sum \text{CNC} + \sum \text{NCC})} * 100 \quad (1)$$

$$\text{Specificity } [\%] = \frac{\sum CC}{(\sum CC + \sum NCNC)} * 100 \quad (2)$$

2.6. Risk quotient and -scenario calculations for chemical and ecotoxicological status assessments

The chemical analytical Risk Quotient (RQ_{chem}) was calculated by dividing the Measured Environmental Concentration (MEC) by the proposed EQS (Equation (3)). The biological RQ_{bio} was calculated by dividing the measured Bioanalytical Equivalent Concentration (BEQ), in our case Estradiol Equivalent Concentration (EEQ) by the effect-based trigger value (EBT) (Equation (4)). RQ > 1 signifies an unacceptable risk for aquatic organisms.

$$RQ_{chem} = \frac{\text{Measured Environmental Concentration (MEC)}}{\text{Proposed EQS}} \quad (3)$$

$$RQ_{effect-based} = \frac{EEQ}{\text{Proposed EBT}} \quad (4)$$

The chemical analytical RQ is used for chemical status assessment and the effect-based RQ for an eco-toxicological status assessment. After EBT discussion in section 3.4.4 we used the 0.4 ng/L EEQ as preliminary EBT. Equation (3) for single analytes, can be adapted to mixture effects of multiple measured substances with the same MoA, via calculation of cumulative RQs, according to Kortenkamp 2007 [41]. Kortenkamp proposed the concentration addition concept as an accurate approach for regulatory use if EDC have the same MoA (Equation (5)).

$$\sum RQ_{E1,\ E2,\ EE2} = \frac{\text{MEC E1}}{3600\ \text{pg/L}} + \frac{\text{MEC E2}}{400\ \text{pg/L}} + \frac{\text{MEC EE2}}{35\ \text{pg/L}} \quad (5)$$

The application of this mixture concept is well supported by additional evidence for endocrine disruptors and other relevant mixtures [42,43]. Moreover, the equation can be further improved to consider unknown unquantified risks by taking into account the specific LODs and LOQs of chemical analytical methods, i.e. by setting concentrations of samples with non-detectable analytes either to 0, LOD/2 of LOD. Three cumulative risk scenarios were calculated in this way to derive the minimum known, the likely, and the maximal risks of steroidal estrogens in the samples:

Mixture risk scenarios for chemical analytical methods:

1) Minimal cumulative risk scenario: $\sum RQ_{EE2,\ E2,\ E1} = \sum (MEC_{EE2,\ E2,\ E1}/AA\text{-}EQS_{EE2,\ E2,\ E1})$
2) Medium cumulative risk scenario: $\sum RQ_{EE2,\ E2,\ E1} = \sum (MEC_{EE2,\ E2,\ E1}$ or $LOD/2_{EE2,\ E2,\ E1}/AA\text{-}EQS_{EE2,\ E2,\ E1})$
3) Maximal cumulative risk scenario: $\sum RQ_{EE2,\ E2,\ E1} = \sum (MEC_{EE2,\ E2,\ E1}$ or $LOD_{EE2,\ E2,\ E1}/AA\text{-}EQS_{EE2,\ E2,\ E1})$

The cumulative RQs for these three risk scenarios are based on the chemical measurements (SI Tables 1–3).

2.7. Correlation analysis of risk quotients and EEQ measurements

We compared the chemical analytical mixture risk scenarios with effect-based biological responses. Cumulative RQs of the minimum and maximum risk scenarios were plotted on a logarithmic scale against the biological EEQ (BEQ) responses and a log-linear regression line ($y = ax^b$) was calculated. Data were tested for log-linearity (scatter-plot), constant variance (TA-plot) and normality (Q-Q plot). Moreover, a double-sided correlation analysis significance test was performed with the Pearson correlation coefficient at $p \leq 0.0001$; $p \leq 0.001$, $p \leq 0.01$ and $p \leq 0.05$ [44]. More specific p-values and confidence intervals were calculated with graph pad 5 using a two-tailed column Pearson normality test after normality check.

2.8. Calculation of a risk indication score (RIS) and screening score for effect-based methods

To measure how precisely a biological response indicates a population relevant mixture risk, we calculated a Risk Indication Score (RIS). A chemical analytical cumulative $RQ_{mix} > 1$ (Equation (5) above) for estrogens indicates an "unacceptable" risk for aquatic organisms and their populations. This is mainly the case for fish species as, based on current knowledge, they include the most sensitive species for estrogenic effects and were used to derive the EQS. RQmix was compared to exceedances of EEQ measured by effect based methods of different EBT (see chapter 2.3) ranging from 300 to 500 pg/L.

There were two possible outcomes:

- If the cumulative chemical $RQ_{E1E2EE2}$ was >1 and the respective EBT was exceeded by the biological response, it was counted as successful risk indication.
- If the cumulative chemical $RQ_{E1E2EE2}$ was >1 and the respective EBT was not exceeded by the biological response, it was counted as failed risk indication.

The number of successful risk indications was scored and normalized to the maximal number of possible risk indications by calculation of the RIS (Equation (6)).

$$RIS\ [\%] = \frac{(\sum \#RQ_{E1,\ E2,\ EE2} > 1 \text{ and BEQ} > \text{EBT}) x100}{\sum \#RQ_{E1,\ E2,\ EE2} > 1} \quad (6)$$

In a few cases, an EBT exceedance was observed where cumulative chemical $RQ_{E1E2EE2}$ was <1. This was scored as "ncr (no chemical risk indication, but positive biological response) and an "ncr* screening score" was calculated (Equation (7)).

$$ncr^*\ [\%] = \frac{(\sum \#RQ_{E1,\ E2,\ EE2} < 1 \text{ and BEQ} > \text{EBT}) x100}{\sum \#RQ_{E1,\ E2,\ EE2} > 1} \quad (7)$$

RIS and ncr* score were calculated using three proposed EBTs. The two parameters identify the specificity of effect-based methods to predict risks caused by steroidal estrogens and their potential as screening methods. This screening allows to detect additional risks caused by estrogen receptor activating substances other than steroidal estrogens or where chemical analysis was not able to quantify steroidal estrogens due to high LOQs.

3. Results and discussion

3.1. Compliance assessment with chemical analytical methods

Here we describe the results of our compliance assessment based on chemical analytical data. Results of the chemical analysis of surface water and effluent samples tested in this study including LOD and LOQ values are provided in SI Tables 1–3 and by Könemann et al. [36].

On average, SW water samples had a much higher percentage of compliant samples compared to WW samples (54% SW vs. 12% WW, Table 1). Due to matrix effects and associated higher LOD/

Table 1
Chemical analytical compliance assessments of 16 surface water (SW) samples and 17 wastewater (WW) samples, which were analysed by HPLC MS/MS methods. Compliance frequency is shown as percentage, arithmetic mean and coefficient of variation (CV). Analytical data is provided in SI Tables 1–3.

	Lab1 [%]	Lab2 [%]	Lab3 [%]	Mean ± CV [%]
SW compliant samples	44	63	56	54.2 ± 9.5
SW non-compliant samples	31	25	38	31.3 ± 6.3
SW not assessable samples	25	12	6	14.6 ± 9.5
WW compliant samples	18	18	0	11.8 ± 10.2
WW non-compliant samples	53	47	41	47.1 ± 5.9
WW not assessable samples	29	35	59	41.2 ± 15.6

LOQs, a much higher number of WW samples (41%) could not be assessed compared to SW samples (15%). For SW and WW sample assessments, the CV is in the range of 6–10%. Due to insufficient silica gel clean-up of WW samples, matrix effects were not reduced in an optimal way. Under ideal conditions, analytical methods can achieve LODs and LOQs of a factor 2 to 3 lower in WW samples. It has to be recognized that LODs of chemical analytical methods for steroidal estrogens have been lowered significantly since 2013 and are likely to decrease further [25,29]. The differences in sample assessment ("compliant" vs. "non-compliant" vs. "not assessable") among the three analytical methods were far more frequently caused by differences in LOQs rather than by differences in detected concentrations (SI Tables 1–3). The percentage of "not assessable" samples (up to 25% in SW and 59% in WW) confirmed the existence of challenges for the chemical monitoring of estrogens E2 and especially EE2. Moreover, if the strict requirements of compliance assessment currently applied to monitoring of priority substances under the WFD (i.e. the LOQ should be 1/3 of the EQS [45]), none of the results obtained in this study could be considered acceptable for compliance assessment of the chemical status of surface waters.

3.2. Ecotoxicological status assessment with in vitro effect-based methods

Bioanalytical equivalent concentrations (BEQ), or EEQ for estrogenicity, are a measure of the effect caused by mixtures of unknown and potentially unidentified chemicals expressed as the equivalent concentration of a known reference compound that would elicit the same effect as the sample [46]. Effect-based trigger values (EBT) for *in vitro* bioassays can be derived by combining calculated or measured BEQs of selected substances that trigger a specific effect, with a benchmark approach using known chemical, toxicological and biological data [38]. Similar to conventional risk assessment using chemical concentration and EQS, exceedance of effect-specific EBTs indicates an elevated, unacceptable risk (hazard & exposure) for the aquatic ecosystem due to chemicals with a particular MoA such as estrogenicity. Measured BEQs below an EBT indicate a low and acceptable ecological risk.

Similar to the results of chemical methods, a greater percentage of SW samples (61%) were "compliant" compared to WW samples (39%) (Table 2). The matrix effects did not lead to any "not assessable" categorization and ER mediated estrogenicity was quantified in all samples. Results obtained using different effect-based methods showed good agreement in average status assessments with CVs of 10–11% in SW and WW samples.

3.3. Comparison of status assessment by chemical analytical and in vitro effect-based methods

The *in vitro* effect-based methods were less matrix-dependent and provided generally lower LOQs than chemical analytical methods (Table 3). As it was mentioned before not all matrix effects were optimally removed during WW samples extraction. All results obtained with *in vitro* effect-based methods allowed a risk assessment because the LOQ was below the target EBT values for all samples. Status assessment with chemical analytical and *in vitro* effect-based methods showed overlapping values and means. The main difference in chemical analytical detection methods is the single substance based approach which can be matrix dependent. *In vitro* effect-based methods allow an integrative activity measurement of all ER activating substances and are not matrix independent as long no cytotoxic and ant-estrogenic effects occur, which was not the case in our samples. Therefore the different responses and LOQs of both method can be compared, a detailed discussion is available in Könemann et al. [36].

3.4. Comparison of chemical analytical and in vitro effect-based methods

3.4.1. Mixture risk scenarios of chemical analytical measurements

Since the main MoA of E1, E2, and EE2 is activation of ER, their interactions in environmental mixtures are most probably concentration additive [36]. The risk quotient addition model, which is derived from the mixture model of concentration addition, was shown to be sufficiently accurate for regulatory use [41]. Therefore, the sum of individual RQs of E1, E2, and EE2, represents the combined risk for the mixture of E1, E2, and EE2 and is called cumulative RQ. To show full ranges of potential RQs, the minimal, the medium, and maximal cumulative RQs were thus determined for each sample.

Depending on the mixture risk scenario, the sum of RQ and mean percentage of samples at unacceptable risk increased (Table 4). The minimal cumulative mixture risk scenario showed in 7 SW and 9 WW samples RQs above 1, whereas 9 SW and all WW samples presented an unacceptable risk in the medium mixture risk scenario. The maximal possible mixture risk scenario showed 10 out of 16 SW samples and all WW samples RQs above 1. A dilution factor was not taken into account for the WW samples (see chapter 3.5.2).

Table 2
In vitro effect-based (ecotoxicological) compliance assessments of 16 surface water (SW) and 17 wastewater (WW) samples using an EBT of 400 pg/L EEQ. Results of compliance assessments are shown as percentage, arithmetic mean, and coefficient of variation (CV). EEQ concentrations are provided in SI Table 4,5.

	ER-Calux [%]	p-YES [%]	MELN [%]	HeLa 9903 [%]	ER-Gene-Blazer [%]	Mean ± CV [%]
SW compliant samples	62	50	50	75	69	61.3 ± 11.2
SW non-compliant samples	38	50	50	25	31	38.8 ± 11.2
SW not assessable samples	0	0	0	0	0	0.0
WW compliant samples	47	35	24	47	35	37.6 ± 9.8
WW non-compliant samples	53	65	77	53	65	62.4 ± 9.8
WW not assessable samples	0	0	0	0	0	0.0

Table 3
LOQ comparison of chemical analytical (LOQ for E2) and *in vitro* effect-based (EEQ corresponding to E2 equivalents) measurements of 16 surface water (SW) and 17 wastewater (WW) samples. Data are shown as arithmetic mean, range, and coefficient of variation (CV). The measurements are provided in SI Tables 1–5.

Methods	LOQ_{SW}			LOQ_{WW}		
	Mean ± SD [pg/L]	Min-max range [pg/L]	CV%	Mean ± SD [pg/L]	Min-max range [pg/L]	CV%
3 chemical analytical methods [E2]	181 ± 291	39–1500	161	627 ± 726	50–3000	116
5 *in vitro* effect-based methods [EEQ]	28 ± 33	2–200	116	60 ± 62	1–216	104

Table 4
Average cumulative risk quotients (RQ) for EE2, E2, and E1 based on mean concentrations of three chemical analytical measurements for 16 surface water (SW) and 17 wastewater (WW) samples calculated for minimal, medium and maximal risk scenarios.

Cumulative RQ for SW samples				Cumulative RQ for WW samples			
Sample code	Minimal	Medium	Maximal	Sample code	Minimal	Medium	Maximal
A (11)	0.02	0.41	0.77	A (26)	0.02	1.78	3.53
B (6)	0.02	0.38	0.72	B (29)	0.03	2.01	3.78
C (1)	0.04	0.45	0.8	C (31)	0.04	7.78	15.48
D (22)	0.04	0.41	0.76	D (4)	0.06	3.45	6.29
E (27)	2.27	2.34	2.38	E (17)	3.49	3.25	5.06
F (30)	0.1	1.77	3.37	F (21)	0.12	2.74	5.32
G (32)	0.28	0.58	0.97	G (14)	0.23	11.2	21.21
H (25)	0.17	1.6	2.62	H (5)	0.58	18.41	34.99
I (8)	0.39	1.21	2.06	I (19)	0.8	5.5	9.75
J (10)	0.37	0.66	0.98	J (16)	7.45	14.54	25.5
K (18)	4.99	4.99	4.99	K (9)	1.49	18.3	34.52
L (24)	1	4.18	7.06	L (13)	4.15	6.93	10.07
M(28)	1.41	4.11	6.36	M(23)	219.18	219.1	219.4
N (15)	2.5	4.99	7.26	N (33)	4.26	15.37	25.69
O (3)	10.71	12.11	16.6	O (12)	4.94	17.01	27.51
P (7)	5.17	5.92	8.19	P (2)	5.6	7.85	9.88
				Q (20)	151.6	151.6	151.6
Mean cumulative RQ	1.84	2.88	4.12	Mean cumulative RQ	23.76	29.81	35.85
Percentage of samples presenting an unacceptable risk	44%	63%	63%		53%	100%	100%

3.4.2. Mixture risk scenarios of chemical analytical measurements compared with in vitro effect based methods

We compared the risk derived for the minimum and the maximum mixture risk scenarios based on chemical analytical data with risk indicated by the effect based methods. Figs. 1 and 2 show that the sum of the risk quotients ("mixture risk of steroidal estrogens") derived from chemical measurements were highly correlated with the measured EEQs in the respective samples. Depending on the choice of LOD, LOD/2 or 0 to replace non-detects, the respective cumulative mixture risk estimation (minimal to maximal) can change considerably (by several orders of magnitude) if non-detects occur.

Overall, EEQ measured with the five *in vitro* effect-based methods correctly assigned the "chemical status" of wastewater samples as determined by the sum of the risk quotients for E1, E2, and EE2, with highest EEQ signals detected at sites where EE2 was present at concentrations above the LOQ and thus quantified. In most cases, the fit for maximal RQ was worse than for the minimal scenario indicating that undetected compounds were likely not present in the mixtures and non-detected compounds did not play a significant role. The significant correlations indicate that for our selection of samples (33 samples from seven countries with variable pollution levels) estrogenicity was mainly caused by the steroidal estrogens E1, E2, and EE2 and anti-estrogenicity or other xenoestrogens played a minor role. This is supported by the iceberg modelling presented by Könemann et al. [36]. However, this result can be partially influenced by the choice of more or less specific extraction methods.

Even though our results confirmed that the steroidal estrogens were dominantly responsible for triggering the estrogenic activities measured by the bioassays, an important question is if the assumptions are applicable to all surface and wastewaters. Although most studies comparing steroidal estrogen concentrations with biological activity support the applicability of effect-based methods [13,18,36,38,46,47], there can be some exceptions. In our dataset, the risk for one SW sample was evaluated as unacceptable in all three chemical analyses, whereas not by any effect-based method (Table 5, sample E (27)). In this sample, EE2 was detected in the range of 73–85 pg/L, which indicated an elevated risk (RQ = 2.27), while concentrations of E2 and E1 were very low (Table SI 1–3, sample E (27)). We can explain this by comparing the potential of E1, E2 and EE2 to induce responses *in vitro* and *in vivo*. While E1 and E2 usually trigger *in vitro* responses at similar concentrations as *in vivo* responses, EE2 is 10–20 times more potent *in vivo* than *in vitro* [33,48]. This is reflected in very low EQS of EE2, but in contrast to the relatively high EBT, which integrates the risk of chemical mixtures. The precautionary principle (to work with EE2-equivalents and an EBT of 35 pg/L) cannot be applied here, because then all samples including background samples [26,38,49] would be assessed as presenting an elevated risk. Since the main source of EE2 is its excretion after use of contraceptive pills, however, EE2 is usually present together with other natural steroidal estrogens, and risk is therefore correctly indicated by results of effect based methods and the proposed EBT. Sample E (27) appears to be an exception as other WW samples (Q (20), M(23)) contained relatively high concentrations of EE2 (which contributed to more than 93% of cumulative RQ) and high EEQs were measured in all five *in vitro* effect-based methods due to the mixture effects in the samples. However, in the aquatic environment EE2 degrades less rapidly than the natural hormones [50] and may still be present

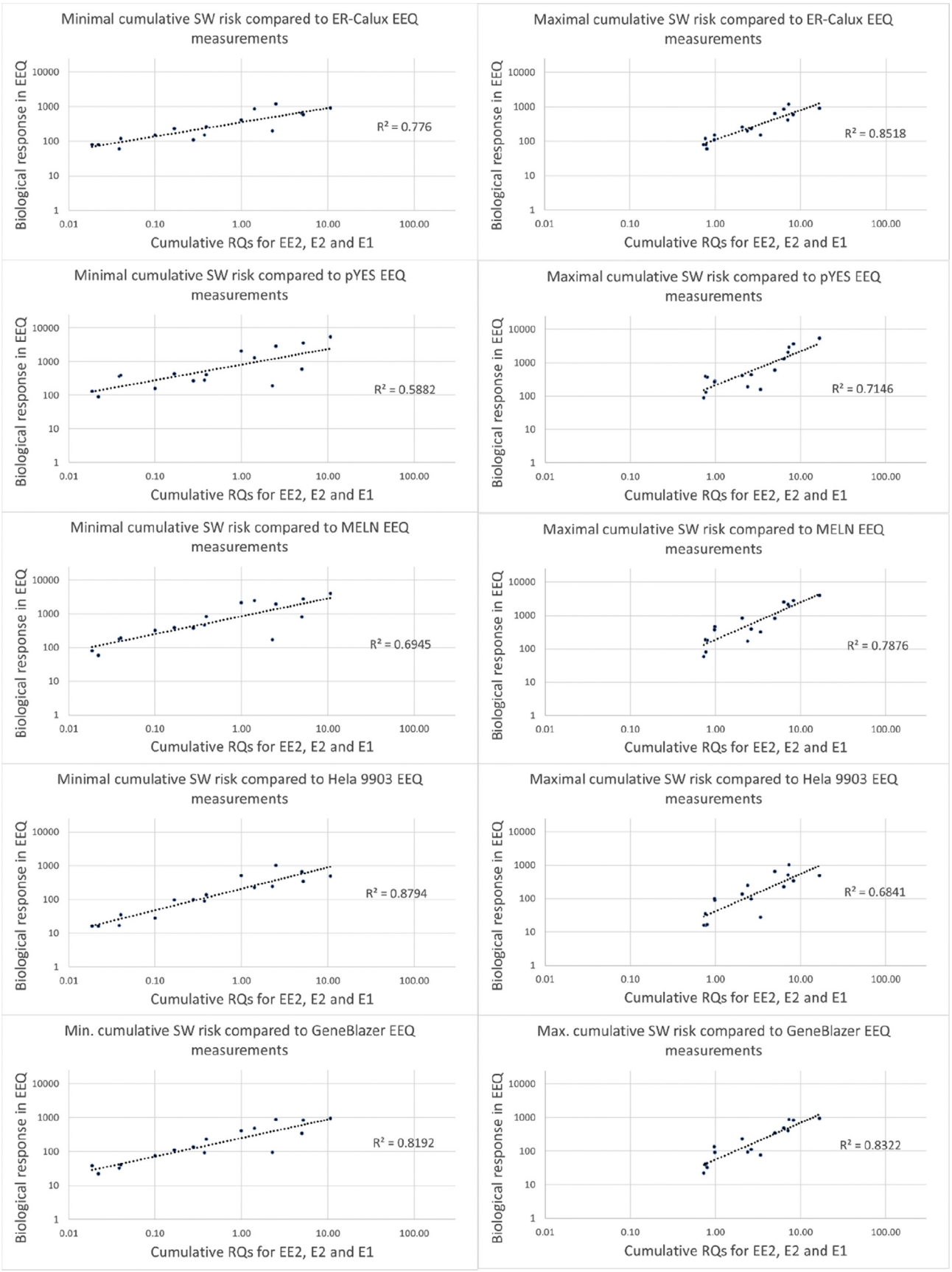

Fig. 1. Minimal and maximal cumulative risk quotients (RQ) compared with measured ER-Calux, pYES, MELN, Hela 9903, ER-GeneBlazer biological responses of ER activation [EEQ in pg/L] in 16 surface water (SW) samples. The *in vitro* effect-based methods are shown vertically from row 1 to 5. All correlations shown in the figures are highly significant with $p < 0.0001$ (SI Table 13).

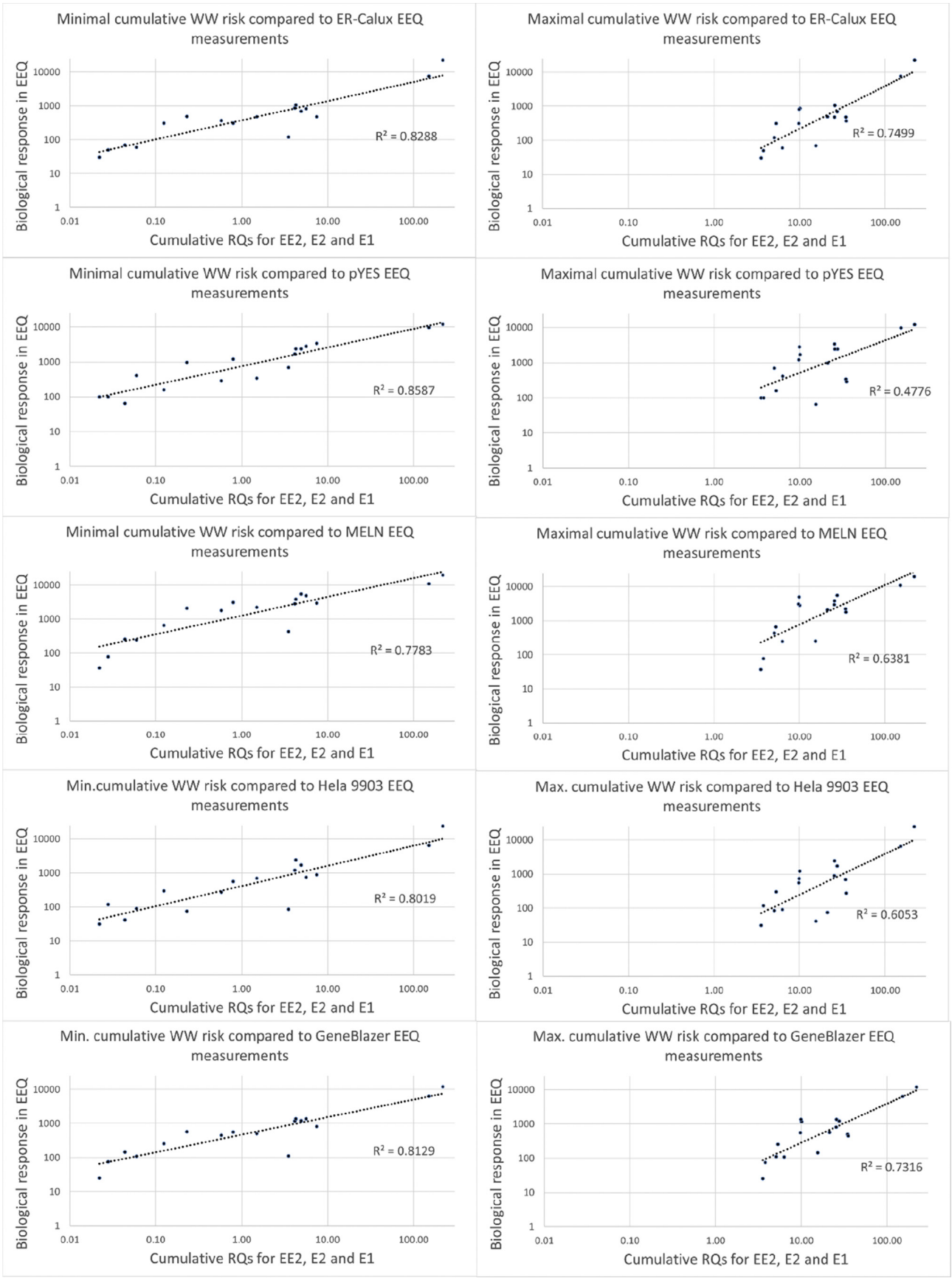

Fig. 2. Minimal and maximal cumulative risk quotients (RQ) compared with measured ER-Calux, pYES, MELN, Hela 9903, ER-GeneBlazer biological responses of ER activation [EEQ in pg/L] in 17 wastewater (WW) samples. The *in vitro* effect-based methods are shown vertically from row 1 to 5. All correlations shown in the figures are highly significant with $p < 0.0001$., only pYES graph at maximal risk had $p < 0.01$ (SI Table 14).

further downstream from the source. Where this is the case, EE2 could be missed both by chemical analysis and effect-based methods, but for effect-based methods the test specific relative potency for EE2 can be considered and further refinements in EBT are possible. EE2 alone can also be discharged by industrial sources, but such sources can be expected to discharge concentrations far above EBT which could be detected by *in vitro* effect-based methods.

In theory, anti-estrogenic compounds could modulate ER activity and, therefore, interfere with EEQ measurements. While most literature focuses on agonistic (estrogenic) modes of action, substantially less information is available on anti-estrogenic MoA in environmental waters [51]. Overall, anti-estrogens are not considered to be a major issue in common municipal WW [33]. So far, anti-estrogenic activity has not been commonly detected in environmental waters by effect-based methods and many limitations exist in measuring anti-estrogenic effects (lack of standardization, and potential artefact problems due to DOC) [48]. The highly significant correlation between measured estrogens and bioassay results found in this study supports the hypothesis that anti-estrogenic substances or other ER-receptor activating substances play a minor role in WW and SW samples containing WW (Figs. 1 and 2).

Given that the EQS values of the steroidal estrogens are based on population-relevant long-term effect data (the EQS were derived from Species Sensitivity Distribution based on data from 9 to 11 fish species), the mixture risk can be considered as directly indicative for population-relevant effects in fish species. As a consequence, the (receptor activation-based) biological response measured with the bioassays, which was highly correlated with cumulative RQs (Figs. 1 and 2), can also be considered to estimate the risk for aquatic species. Therefore, we investigated whether the biological EEQ response exceeded the EBT of 400 pg/L in those cases where a cumulative population relevant mixture-risk was identified.

3.4.3. Risk indication of in vitro effect-based methods for cumulative population relevant effects

The risk indication for all three EBT scenarios was calculated in the SI Tables 10–12 and showed a good agreement of chemical and biological risk indicators. For the moderate EBT of 400 pg/L the relative RIS (score of biological responses which indicates quantifiable chemical mixture risk) of all five *in vitro* assays was 77% ± 13% CV for SW samples and 91% ± 5% CV for WW samples. There are two reasons for the higher percentage of RIS in WW. First, there is naturally higher variability in the evaluation of samples with lower activities, such as SW compared to WW (EEQs close to the EBT can result in different category). Second, the composition of municipal wastewaters in Europe have been shown to typically contain steroidal estrogens with EE2 contributing ≤40% of total estrogenic activity (EEQ) [33]. As discussed above, the bioassays indicate the risks most precisely when EE2 is not the predominant estrogenicity driver but occurs in combination with other steroids (chapter 3.4.2).

Our results also demonstrate the potential of the effect-based methods to screen samples for other estrogens than the three target compounds. For example, two WW samples were evaluated as “compliant” or “not assessable” when their RQ were calculated based on chemical data (SI Table 11, samples G (14) and I (19)), however, most bioassays indicated elevated risk. The most probable reason is that the screening function of bioassays is not limited to steroidal estrogens and confirms findings of recently published approaches for screening endocrine active pharmaceuticals and other receptor activating substances [20,21,25]. In this study, the screening for other receptor activating compounds was measured by an ncr* score (the relative positive risk indication without chemical analytical verification ratio normalized to the number of chemical positive findings) and resulted in 11.4% “biological positives” for SW samples and 26.7% for WW samples. In other words, with the selected EBT, the effect-based methods were able to screen 11–27% more positive samples for SW and WW.

Finally, one of the *in vitro* effect-based methods (HeLa 9903) occasionally showed EEQs below the EBT where all other bioassays and chemical methods showed a risk (SI Table 11, SW samples P (7), M(28)). Variability can generally account for some negative risk indications, which occur when the detected EEQs are close to the EBT. These samples contained very high concentrations of E1. E1 is typically a less potent ER ligand than E2, but it is particularly less potent in HeLa 9903 with an estrogenic potency relative to E2 of 0.018 (the relative potencies of other used *in vitro* effect-based methods are listed in the SI Table 5 of the companion publication [36]), and thus the contribution of E1 to EEQ was lower than for most other bioassays with exception ER-Calux. Low potencies and higher variabilities, which are indicated by LOQs can lead to reduced detectability by some bioassays, and test specific refinements should be considered. This is also a matter of identifying criteria for benchmarking of bioassays suitable for these application purpose and to add test specific “sensitivity factors” which can be multiplied with EEQs to meet a screening EBT. This was identified as a future need and included as one of the aims of a subsequent project (see Conclusions and Outlook).

3.4.4. Comparison of trigger value (EBT) scenarios to assess the risk indication of in vitro effect-based methods

Different trigger values were applied to assess risk indication and screening function of used methods. Results provided in chapter 3.4.3 show that an EBT of 400 pg/L can distinguish with high precision (77% ± 13%–91% ± 5%) between more and less polluted SW and WW sites, indicated by a quantifiable population relevant mixture risk. To investigate the impact of the choice of EBT on results all three proposed EBT, 300 pg/L EEQ [33], 400 pg/L EEQ [25,26] and 500 pg/L EEQ [38] were compared.

Application of the lowest EBT of 300 pg/L resulted in the highest RIS of 83 ± 6% for SW samples, and 93 ± 6% for WW samples, as well as in the highest ncr* score with 26% for SW and 33% for WW. Use of the moderate EBT of 400 pg/L led to slightly (2–6%) lower RIS for WW and SW samples, compared to the strictest EBT scenario of 300 pg/L. On the other hand, the moderate EBT scenario reduced the ncr*to half for SW (11%) and to two thirds for WW (27%), compared to the strictest EBT scenario. The least stringent EBT of 500 pg/L lowered the RIS for WW and SW samples to 8% and 16%, respectively, compared to the strictest EBT scenario. The ncr* score decreased to 2% in SW and 20% in WW. The ncr* score can vary

Table 5
Mean positive risk indication scores (RIS) and coefficients of variations (CV) of the 5 *in vitro* effect-based methods for the identification of population relevant risks (RQs>1) applying different trigger values (EBT) 300, 400 and 500 pg/L. Additionally the percentage of positive biological responses without chemical verification (ncr*) was calculated. Data were used from SI Tables 10–12.

EBT approach	Risk indication of steroidalestrogens RIS [%]				Screening of other xenestrogens and unquantifiable steroidal oestrogens Mean percentage of ncr* related to chemical positives [%]			
	SW	CV	WW	CV	SW	CV	WW	CV
EBT = 300 pg/L	82.9	6.4	93.3	6.1	25.7	35.6	33.3	13.6
EBT = 400 pg/L	77.1	12.8	91.1	5.0	11.4	15.7	26.7	14.9
EBT = 500 pg/L	65.7	21.7	84.4	9.9	2.2	4.9	20.0	16.4

depending on the chemical composition and activity of estrogen mixtures.

Generally the ncr* should allow screening for non-target ER receptor-activating substances. A high ncr* as shown in our strictest EBT scenario would mean that there was a need for additional analyses by costly high end chemical analytical methods (26–33%). Application of a higher EBT of 500 pg/L results in a low ncr*. (2% in SW) indicating that samples containing unknown estrogens will not be selected. The moderate EBT scenario means that fewer (half to one third) samples have to be analysed further by chemical analysis. This moderate EBT scenario is still quite protective and specific with 77–91% of positive RIS based on quantified chemical analytical mixture risks. The moderate EBT scenario of 400 pg/L EEQ has an additional screening function for other non-target ER activating substances, combined with a high specificity for the risks of steroidal estrogens.

3.4.5. Specificity and sensitivity of chemical analytical and in vitro effect-based methods in compliance assessments

A suitable method should be specific and sensitive. A specificity and sensitivity analysis was therefore performed with each methods applied in this study in order to characterize and compare their suitability for monitoring (Fig. 3AB).

For SW samples the three chemical analytical methods performed with moderate specificity >73% and moderate sensitivity >69%. The Lab3 method achieved higher sensitivity of 75% and the Lab2 method achieved the highest specificity level of close to 85%. The five *in vitro* effect-based methods had, in most cases, high specificity and high sensitivity >90% in SW. Only the sensitivity of one method HeLa 9903 was lower (66.7%). ER-Calux, p-YES and MELN were the most sensitive assays (100% sensitivity), and ER-Calux, HeLa 9903 and ER-ER-GeneBlazer had the highest specificity (100%).

For WW samples, the chemical analytical methods performed with low specificity (in a range of 52–56%) and with low to moderate sensitivity (59–67%), likely due to matrix effects which were not removed efficiently by the silica gel cleaning step (see methods section 2.1). Most *in vitro* effect-based methods performed well in WW and showed both high specificity and high sensitivity >85%. Only the p-YES and MELN were less specific with 71% and 57% specificity, respectively. This can be explained by the higher sensitivity for E1 of both methods. For WW, MELN and ER-ER-GeneBlazer were the most sensitive assays (100% sensitivity), and ER-Calux and HeLa 9903 were the most specific. Finally, most (exception of MELN and pYES in SW which performed similarly) (Fig. 3 AB). This can be explained by the quantification problems of HPLC MS/MS that often occurred in WW and in some SW samples (SI Tables 1–3 and 6,7).

3.4.6. Comparability of chemical analytical and in vitro effect-based methods

A recent literature review [52] highlighted the need for sufficiently sensitive analytical methods for E2 and EE2 in order to be able to comply with the WFD reporting requirements [8,9]. Our study applied advanced analytical methods and confirmed this finding. The main advantages of chemical analytical methods are the quantification of single analytes, however, for both E2 and EE2 LOQs were often >EQS. Chemical analytical methods were able to detect steroidal estrogens above their EQS in 56% of SW samples and only in 16% of WW samples [36], demonstrating that the chemical analytical detection of E2 and EE2 is currently at the limit of feasibility with advanced methods. First monitoring results of the EU watch-list substances in 2017 confirm these results. In 21 EU member states the LOQs for EE2 E2 were above their EQS in >95% and approx. 50% of unquantified water samples analysed. In our study, if steroidal estrogens were detected, the average coefficients of variation CV % of quantifiable concentration measurements in SW (for E1 = 22.2%, E2 = 28.3%, EE2 = 15.8%) and WW (for E1 = 18.9%, E2 = 36.2%, EE2 = 14.6%) showed good agreement between the three chemical analytical analyses (SI Tables 1–3), but the methods showed also significant absolute variability in LOQs (Table 3) making a comparable risk-assessment difficult.

Overall, i*n vitro* effect-based methods were highly sensitive (90–94%) and specific (83–92%) in both SW and WW assessments (Fig. 3AB). The main advantage of the *in vitro* effect-based methods is their ability to account for the mixture toxicity and integrate the effects of unknown chemicals with the same MoA (e.g. metabolites) as well as synergistic or antagonistic mixture effects. In our study the CV of SW and WW sample assessments was in the range of 10–11%, showing good comparability of all five *in vitro* methods regarding the status assessment (Table 2) without any not-assessable samples.

Variability of *in vitro* effect based methods is similar to that of chemical analytical methods. To quantify intra- and inter-test variability, five *in vitro* effect-based methods were recently compared [47]. In this comparison the CV of EEQ concentrations measured in the five in *vitro* assays and for all samples was around 32% for comparing artificial mixtures. CV was lower for intra-day experiments (30%) compared to inter-day experiments (37%). ER-Calux had the best precision and repeatability with an overall CV of 13%. Further validation, inter-laboratory comparison studies and standardization of these effect-based methods may still improve their suitability for monitoring (Mehinto et al. [53]). In our study the five used *in vitro* effect-based methods correlated well among each other as it was shown in the companion publication by Könemann et al. [36]. In line with the results of our study, other studies have also confirmed that *in vitro* effect-based methods are

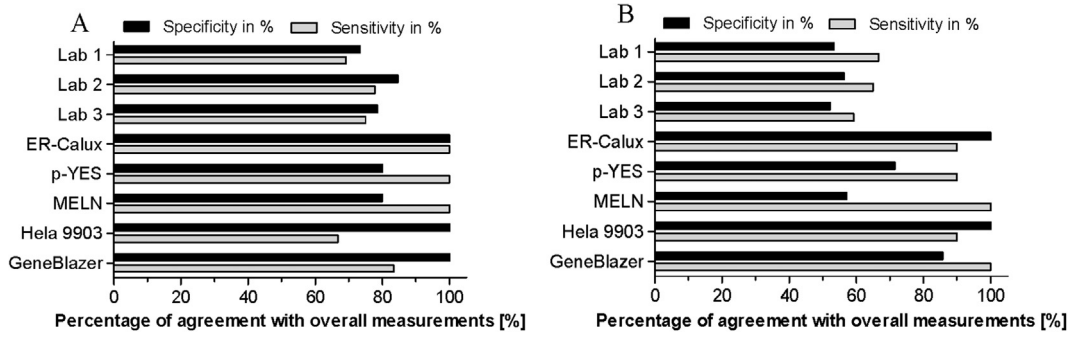

Fig. 3. AB: Specificity and sensitivity assessment of three analytical LC MS/MS methods (Lab 1, Lab 2, Lab 3) and five *in vitro* effect-based methods for specificity and sensitivity applied in this study. Results are given as relative values in % for 16 surface water samples (A) and 17 wastewater samples (B) and were compared to the overall compliance assessment. A high specificity indicates if a method is less prone to false positive assessments. A high sensitivity indicates if a method is less prone to false negative assessments.

able to benchmark contamination by estrogenic compounds correctly [23,54].

The results of our study support the application of *in vitro* effect-based methods for surface and especially for WW monitoring. Results of individual bioassays and chemical analyses correlated highly (Figs. 1 and 2) demonstrating the suitability of all effect-based methods to indicate steroidal estrogen pollution with population relevant mixture risks. Our proposed approach for an ecotoxicological status assessment is in line with the latest results of the EU project SOLUTIONS from König et al. [21], which measured pollution patterns in the river Danube using a large set of effect-based as well as chemical analytical methods. In this study an integrated analytical-bioanalytical approach was well suited to detect the impact of untreated wastewater on Danube River water quality. Both chemical and biological effect patterns were affected in a consistent way.

Limitations and recommendations: To prevent overestimation of these *in vitro* effect-based methods to protect the environment, their limitations are summarized as follows:

a) Cell culture based assays cannot account for organism-level toxicokinetic changes e.g. metabolism. Differences in toxicokinetics are likely in less frequently investigated species and modes of actions might in specific cases result in over- or underestimations of *in vivo* effects when evaluated purely by *in vitro* effect-based methods with one trigger value. Despite this our results showed that also the highly potent EE2 was correctly identified in mixtures by *in vitro* methods, see chapter 3.4.3. But generally the chemical analytical approach also does not consider the toxicokinetics and bioavailability, and EQS are always limited to available effect-data sets.
b) Although not confirmed in any study including this one, the interaction of the generally much less potent (compared to the steroidal estrogens) anti-estrogenic compounds could lead to underestimation of risks of estrogenicity cannot be excluded for very specific sample compositions. In case if high concentrations of anti-estrogenic substances would bind to ER receptor the detection of ER-agonists can be lowered.

Considering these limitations and options, we recommend direct use of *in vitro* effect-based methods with a preliminary screening EBT of 400 pg/L EEQ under the WFD for the following pragmatic reasons: a) there are currently no better available tools to monitor this type of endocrine disruptor pollution providing important link between MoA and adverse effects, b) they will circumvent current monitoring problems of steroidal estrogens, c) the methods are cost-efficient and can decrease the financial burden of monitoring, d) they are readily available, e) they address mixture effects.

3.4.7. Risk management

3.4.7.1. Surface water risk management options. The development of analytical techniques to detect EDCs in environmental matrices still remains one of the main challenges for environmental chemists [54,55]. Due to analytical difficulties in the last decades no representative EU-wide monitoring dataset is available and risk characterizations are mainly known from modelling. For example Johnson and colleagues [1] estimated at median flow conditions an average an EQS exceedance of EE2 in 12% by length of Europe's rivers, which can in some countries also be higher than 30%. This single substance related population relevant risk is certainly increased by other ER activating substances.

One of the main recommendations (in view of the future review of the WFD) is to integrate effect-based methods into monitoring of water quality and to adopt them as a key approach for addressing chemical mixtures interactions with aquatic organisms [16]. Our study supports this recommendation: specific effect-based methods proved to be suitable tools to indicate the risks associated to the quantified and unquantified fractions of EE2, E2, and E1 for various water samples and should therefore be applied as screening tools to identify polluted waterbodies. Especially because of the low LOQs and low absolute variability in LOQs they are reliable and suitable regarding risk assessments and prioritizations.

Besides, effect-based methods are the only currently available tools to address unknown mixture risks and circumvent the monitoring limitations of current chemical analytical methods, as mentioned above. But another question needs to be solved for risk management: How to proceed when an EBT is exceeded?

First of all, an EBT exceedance can identify waterbodies at risk for receptor-mediated estrogenicity. This can allow focusing of monitoring resources on priority sites. For example, if 100 waterbodies are screened and only 10% are at risk, for 90% of remaining waterbodies no costly high-end chemical analysis needs to be performed. Taking into account our chemical analytical findings for SW assessment: 54% ± 10% of the samples were rated as “compliant”, 31% ± 6% “non-compliant” and 15% ± 10% “not-assessable” due to too high LOD/LOQ (Table 1). With effect-based methods, the results overlapped with those obtained with analytical methods, (61% ± 11% of the samples assessed as “compliant” and 39% ± 11 “non-compliant” at an EBT of 400 pg/L EEQ) (Table 2), but thanks to the shift from chemical analytical assessments to an ecotoxicological effect-based assessment, the percentage of not assessable samples can be reduced to zero with an obvious benefit in terms of assessment feasibility and costs.

Secondly, the choice is given to use directly the EBT RQ (Equation (4) in chapter 2.6) for an ecotoxicological status assessment or if additionally an identification of substances for an investigative purpose is needed, e.g. via application of mini Effect-Directed Analysis (EDA), but this would definitely increase the costs. A cost estimation for a mini EDA for ER activation is in the range of 5 k Euro per sample (pers. communication Timo Hamers VU University of Amsterdam). Based on our study, if a ecotoxicological risk is identified, we suggest to mitigate if possible the risk or to identify the cause substance. Considering the current costs of mini-EDAs for substance identification and the high probability that most of the effects of concern are caused by mixtures this favours a direct risk reduction. An ecotoxicological effect-based assessment can be established if in the EU context the EBT is harmonized and highly validated and comparable effect-based methods are used for screening. In this study, we characterized and discussed the screening value of *in vitro* effect-based methods. Most of the *in vitro* effect-based methods are less expensive compared to high-end chemical analytical methods considering installation costs and analysis costs per sample. A short cost discussion subchapter is provided in the SI.

3.4.7.2. Wastewater risk management options. Although no legal discharge limits for micropollutants exist at the EU level [31], WW are often monitored as the main known sources of these compounds to waterbodies. Jarosova and colleagues [33] compiled data of 353 wastewater measurements from three studies with a median concentration of 7–12 ng/L E1, 1.3–1.7 ng/L E2 and 0.47–0.6 ng/L EE2, so it will depend on the dilution factor and the background concentration of the receiving water if the EQS can be met and population relevant risks can be excluded. Also, several activities aim at limiting unnecessary risks of pharmaceuticals, such as the EU Strategy on pharmaceuticals [8] which aims at reducing discharges, emissions, and losses, or the Eco-Pharmaco-Stewardship Initiative for industrial wastewaters [48]. Our study identified a high RIS (mean 91%) of 5 *in vitro* effect-based methods for WW and

notable chemical analytical limitations for the detection of steroidal estrogens in these samples because of their complex matrix composition. This offers a direct use of effect-based methods to WW risk regulation at local, national, and EU-wide level. This is of special importance, as the main entrance pathway of the synthetic EE2 as well as the overall ER-mediated estrogenicity into our waterbodies is municipal WW. ER-mediated estrogenicity in WW can be reduced by around a factor of 10 with additional wastewater treatment techniques and the *in vitro* effect-based methods can be also used to monitor these technical options to reduce the pharmaceutical and anthropogenic mixture risks before entering the aquatic environment [56–58].

Without knowledge of the ER mediated estrogenicity risk in the receiving waterbody, we suggest using the same ecotoxicological status assessment for WW as for SW to ensure an assessment compliance, because a compliant WW cannot lead to a change in SW compliance assessment to a non-compliant assessment. With further knowledge about the ER mediated estrogenicity in the receiving waterbody the EEQ of WW can be combined with a dilution factor to estimate a more appropriate overall ER mediated estrogenicity RQ in SW (Equation (8)). Similarly, for chemical analytical cumulative risk assessment, the combined risk can be calculated with a dilution factor and Measured Environmental Concentrations MEC (Equation (9)).

$$\text{RQ effect}-\text{based} = \frac{\text{WW EEQ}}{\text{EBT}} * \frac{1}{\text{dilution factor}} + \frac{\text{SW EEQ}}{\text{EBT}} * (1 - 1/\text{dilution factor}) \quad (8)$$

$$\text{RQ chem} = \left(\frac{\text{WW MEC EE2}}{\text{EQS EE2}} + \frac{\text{WW MEC E2}}{\text{EQS E2}} + \frac{\text{WW MEC E1}}{\text{EQS E1}}\right) * \frac{1}{\text{dilution factor}} + \left(\frac{\text{SW MEC EE2}}{\text{EQS EE2}} + \frac{\text{SW MEC E2}}{\text{EQS E2}} + \frac{\text{SW MEC E1}}{\text{EQS E1}}\right) * (1 - 1/\text{dilution factor}) \quad (9)$$

Because we intended a 1:1 comparability in SW vs. WW assessments, our sampling locations had varying dilution factors, and we had, in most cases, too limited knowledge about the EEQ in the receiving water, we used the simplified Equation (4) (Methods 2.6) for our calculations. For further studies, another practical solution would be to measure EEQ directly in the mixing zone of the receiving waterbody, meaning one EEQ for one ecotoxicological assessment.

4. Conclusions and Outlook

Considering their relevance, applicability for screening as well as limitations, we propose the application of effect-based methods under the WFD, in particular the use of *in vitro* effect based methods for identifying ER-mediated risk in WFD monitoring programs. The methods are: I) capable of addressing relevant combined mixture effects, II) able to overcome detection problems encountered with analytical techniques for the EU watch list substances (EE2 and E2), III) suitable as screening tools for the identification and prioritization of waterbodies requiring further examination, and IV) suitable for measuring ecotoxicological status in relation to receptor-mediated estrogenicity, one of the most relevant MoA of EDCs.

A recent study [16] emphasized the need for the harmonization and standardization of EBT. The derivation of EBT is one of the tasks identified for the activity on effect-based methods started by the EU Working Group Chemicals under the Common Implementation Strategy for the WFD [59]. The results of our study confirm that a preliminary screening EBT of 400 pg/L EEQ is suitable for the identification of population relevant analytical (and mixture) risks from steroidal estrogens, at the same time as achieving the screening of other ER-activating substances. This EBT is recommended as a suitable threshold or cut-off value to discriminate samples of greater level of estrogenic pollution with the aim of classifying waterbodies. The application of this value can be further refined, taking into account differences in sensitivity of the used methods via sensitivity factors and risk classifications.

Our study demonstrated SW and WW risk management options by using risk indication scores (RIS) based on tested EBT covering population relevant effects for aquatic organisms. The tested concepts proved to be applicable for WW and for most SW (91% of RIS vs. 77% of RIS). Effect-based methods were highly sensitive (90–94%) and specific (83–92%) in both SW and WW assessments. In special situations where EE2 occurs at low, yet EQS-exceeding, concentrations and alone mainly contributes to ER-mediated estrogenicity, a false negative assessment might occur. This issue was only discovered in one of 16 SW or in total 33 samples. Based on the highly significant correlations between all measured estrogens risks and bioassay results found in this study (Fig. 1 + 2) it was possible to identify that anti-estrogenicity and matrix effects played a minor role in most of our samples. This presented approach allows us to screen, prioritize, and manage environmental samples using the ER-EBT concept very similar and compatible to the current chemical status assessment of the WFD. Furthermore, ER-Calux, A-YES, and L-YES will be standardized at DIN/EN/ISO level by early 2018 supporting their availability for regulatory use. Our study showed the use of very specific *in vitro* effect-based methods with tested EBT is able to bridge the gap between conventional analytical and effect-based monitoring and risk assessment for steroidal estrogens.

The combination of the results of this study demonstrates that water quality assessment can progress from a purely analytical approach to effect-based monitoring, from single substance to known and unknown mixture assessment, and from *in vitro* screening to population-relevant risk assessment. This approach can support the introduction of the proposed new holistic approach to the regulation of chemicals in the aquatic environment under the EU Water Framework Directive, an objective which EU water directors agreed in November 2016 to investigate [60] and which has also been recommended by international platforms such as, the NORMAN network and the EU-funded SOLUTIONS project [16]. A follow-up study regarding the use of different effect-based methods under the EU watch list mechanism in 2017 and 2018 is intended. This follow-up study aims to characterize the screening function for ER-mediated effects with regulatory relevant EU watch list samples. Moreover, it intends to apply an integrative effect-based approach for other relevant pharmaceutical MoA such as COX inhibition [61,62].

Acknowledgements

The major part of the project was funded by in-kind contribution from numerous project partners showing high interest in this project. Essential parts of the project, e.g. extraction and reporting were externally co-funded by the Swiss Centre of Applied Ecotoxicology EAWAG-EPFL, NORMAN Network, INERIS and Pharmaceutical Associations (Pfizer, Johnson & Johnson-Johnson & Johnson, TEVA which joined the project later and only supported the extraction at central Lab BDS). In total, 70

colleagues associated to 24 institutes and agencies in 12 nations contributed to this project. The authors would like to thank: Joint Research Centre (EC), The French Agency for Biodiversity (previously ONEMA) (FR), INERIS (FR), Bio Detection Systems (NL), Swiss Centre for Applied Ecotoxicology (CH), Federal Institute of Hydrology (DE), Federal Environment Agency (DE), RWTH Aachen (DE), RECETOX (CZ), NORMAN-Network, Helmholtz Centre for Environmental Research-UFZ (DE), IRSA-CNR (IT), Italian Institute of Health (IT), University of Leon (ES), Water Research Institute T.G.Masaryk (CZ), Bavarian State Office for Environment (DE), LANUV (DE), Environment Agency Austria (AT), ISSeP (Scientific Institute of Public Service) Wallonia (BE), SMAT (IT), BrianzAcque (IT), Vettabbia scarl, MilanoDepur (IT), Agence de l'Eau Adour-Garonne (FR), Ontario Ministry of the Environment and Climate Change (CAN), McGill University (CAN), Environmental Institute (SK) and DG Environment of the European Commission (EC). The project partners from UFZ, RWTH Aachen, INERIS and RECETOX were supported by the SOLUTIONS project, that is funded by the European Union Seventh Framework Programme (FP7-ENV-2013-two-stage Collaborative project) under grant agreement number 603437. Additionally, we would like to thank Christin Kuehnert (UFZ) and Maria König (UFZ) for experimental assistance. Special thanks for stewardship of this study we would like to address to Helen Clayton and Stéphanie Schaan from European Commission, Environment DG and Joint Research Centre (EC) for chemical analysis of samples.

The Federal Institute of Hydrology got financial support by the Federal Ministry for the Environment, Nature Conservation, and Building (BMUB).

List of abbreviations

AA-EQS	Annual-Average Environmental Quality Standard
EDA	Effect directed Analysis
EDC	Endocrine Disrupting Compounds
EQS	Environmental Quality Standard
BEQ	Bioanalytical equivalents
COX	cyclooxygenase
CV	Coefficient of variation
EE2	17-alpha-ethinylestradiol
E2	17-beta-estradiol
E1	Estrone
EBT	Effect-based trigger values
EEQ	17-beta-estradiol-equivalents
ER	Estrogen Receptor
EU	European Union
HPLC MS	High Pressure liquid chromatography–mass spectrometry
LOD	Limit of Detection
LOQ	Limit of Quantification
MoA	Mode of Action
MEC	Measured Environmental Concentration
ncr	No Chemcical Risk indication but positive biological response
RIS	Risk Indication Score
RQ	Risk Quotient
SW	Surface waters
WFD	Water Framework Directive
WW	Wastewaters

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.trac.2018.02.013.

References

[1] A.C. Johnson, E. Dumont, R.J. Williams, R. Oldenkamp, I. Cisowska, J.P. Sumpter, Do concentrations of ethinylestradiol, estradiol, and diclofenac in European rivers exceed proposed EU environmental quality Standards? Environ. Sci. Technol. 47 (2013) 12297–12304. https://doi.org/10.1021/es4030035.

[2] J. Zha, L. Sun, Y. Zhou, P.A. Spear, M. Ma, Z. Wang, Assessment of 17α-ethinylestradiol effects and underlying mechanisms in a continuous, multi-generation exposure of the Chinese rare minnow (Gobiocypris rarus), Toxicol. Appl. Pharmacol. 226 (2008) 298–308. https://doi.org/10.1016/j.taap.2007.10.006.

[3] A.R. Schwindt, D.L. Winkelman, K. Keteles, M. Murphy, A.M. Vajda, An environmental oestrogen disrupts fish population dynamics through direct and transgenerational effects on survival and fecundity, J. Appl. Ecol. 51 (2014) 582–591. https://doi.org/10.1111/1365-2664.12237.

[4] H. Segner, A. Casanova-Nakayama, R. Kase, C.R. Tyler, Impact of environmental estrogens on Yfish considering the diversity of estrogen signaling, Gen. Comp. Endocrinol. 191 (2013) 190–201. https://doi.org/10.1016/j.ygcen.2013.05.015.

[5] L. Viganò, J.L. Loizeau, A. Mandich, G. Mascolo, Medium- and Long-Term Effects Of estrogenic contaminants on the middle river Po fish community as reconstructed from a sediment core, arch, Environ. Contam. Toxicol. 71 (2016) 454–472. https://doi.org/10.1007/s00244-016-0315-3.

[6] K.A. Kidd, P.J. Blanchfield, K.H. Mills, V.P. Palace, R.E. Evans, J.M. Lazorchak, R.W. Flick, Collapse of a fish population after exposure to a synthetic estrogen, Proc. Natl. Acad. Sci. 104 (2007) 8897–8901. https://doi.org/10.1073/pnas.0609568104.

[7] M. Carere, S. Polesello, R. Kase, B.M. Gawlik, The emerging contaminants in the context of the EU water framework directive, in: M. Petrovic, S. Sabater, A. Elosegi, D. Barceló (Editors), Emerg. Contam. River Ecosyst. Occur. Eff. Under Mult. Stress Cond, Springer International Publishing, Cham, 2016, pp. 197–215. https://doi.org/10.1007/698_2015_5011.

[8] Off.J. Eur. Union L 226 56, Directive 2013/39/EU of the European Parliament and of the Council of 12 August 2013 amending Directives 2000/60/EC and 2008/105/EC as Regards Priority Sin the Field of Water Policy, 2013, pp. 1–17. n.d.

[9] Off. J. Eur. Union L 78, Commission Implementing Decision (EU) 2015/495 of 20 March 2015 Establishing a Watch list of Substances for Union-wide Monitoring in the Field of Water Policy Pursuant to Directive 2008/105/EC of the European Parliament and of the Council, 2015, p. 40. n.d.

[10] European Commission, Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL amending Directives 2000/60/EC and 2008/105/EC as Regards Priority Substances in the Field of Water Policy, COM, 2011, p. 876 final 2011/0429 (COD), Brussels, 2012.

[11] R. Kase, R.I.L. Eggen, M. Junghans, C. Götz, J. Hollender, Assessment of micropollutants from municipal wastewater- combination of exposure and ecotoxicological effect data for Switzerland, ISBN: 978-953-307-233-3, in: Wastewater - Evaluation and Management,Fernando Sebastián García Einschlag, InTech, 2011, pp. 32–55 (chapter 2), Available at: http://cdn.intechopen.com/pdfs/14570.pdf. (Accessed 26 May 2017).

[12] WHO/UNEP, State of the science of endocrine disrupting chemicals-2012, An Assessment of the State of the Science of Endocrine Disruptors Prepared by a Group of Experts for the United Nations Environment Programme (UNEP) and World Health Organization (WHO), 2013. http://www.who.int/ceh/publications/endocrine/en/. (Accessed 13 April 2017).

[13] R.N. Carvalho, A. Arukwe, S. Ait-Aissa, A. Bado-Nilles, S. Balzamo, A. Baun, S. Belkin, L. Blaha, F. Brion, D. Conti, N. Creusot, Y. Essig, V.E. V Ferrero, V.V. Flander-Putrle, M. Fürhacker, R. Grillari-Voglauer, C. Hogstrand, A. Jonáš, J.B. Kharlyngdoh, R. Loos, A.K. Lundebye, C. Modig, P.E. Olsson, S. Pillai, N. Polak, M. Potalivo, W. Sanchez, A. Schifferli, K. Schirmer, S. Sforzini, S.R. Stürzenbaum, L. Søfteland, V. Turk, A. Viarengo, I. Werner, S.S. Yagur-Kroll, R. Zounkova, T. Lettieri, Mixtures of chemical pollutants at European legislation safety concentrations: how safe are they? Toxicol. Sci. 141 (2014) 218–233. https://doi.org/10.1093/toxsci/kfu118.

[14] European Commission, 4th Report on the Implementation of the "Community Strategy for Endocrine Disrupters" a Range of Substances Suspected of Interfering with the Hormone Systems of Humans and Wildlife (COM, 1999, p. 706. SEC(2011) 1001 final, Brussels, 2011, http://ec.europa.eu/environment/chemicals/endocrine/pdf/sec_2011_1001.pdf. (Accessed 13 April 2017).

[15] The combination effects of chemicals. Chemical mixtures. COM, European CommissAssessment of Micropollutants from Municipal Wastewater- Combination of Exposure and Ecotoxicological Effect Data for Switzerlandion, Communication from the Commission to the Council, 2012, p. 252. Brussels, 2012, http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52012DC0252. (Accessed 13 April 2017).

[16] W. Brack, V. Dulio, M. Ågerstrand, I. Allan, R. Altenburger, M. Brinkmann, D. Bunke, R.M. Burgess, I. Cousins, B.I. Escher, F.J. Hernández, L.M. Hewitt, K. Hilscherová, J. Hollender, H. Hollert, R. Kase, B. Klauer, C. Lindim, D.L. Herráez, C. Miège, J. Munthe, S. O'Toole, L. Posthuma, H. Rüdel, R.B. Schäfer, M. Sengl, F. Smedes, D. van de Meent, P.J. van den Brink, J. van Gils, A.P. van Wezel, A.D. Vethaak, E. Vermeirssen, P.C. von der Ohe, B. Vrana, Towards the review of the European Union Water Framework management of chemical contamination in European surface water resources, Sci. Total Environ. 576 (2017) 720–737. https://doi.org/10.1016/j.scitotenv.2016.10.104.

[17] R. Kase, H. Clayton, F. Martini, Science-Policy Interface (SPI) Activity on Prioritization of Research Needs, Knowledge Availability and Dissemination for the Working Group E (Chemical Aspects) 2010-2012, 2012. Available at: https://circabc.europa.eu/w/browse/5bf63ff3-b24b-4365-8a57-38e4d56b941c. (Accessed 13 April 2017).
[18] C. Di Paolo, R. Ottermanns, S. Keiter, S. Ait-Aissa, K. Bluhm, W. Brack, M. Breitholtz, S. Buchinger, M. Carere, C. Chalon, X. Cousin, V. Dulio, B.I. Escher, T. Hamers, K. Hilscherová, S. Jarque, A. Jonas, E. Maillot-Marechal, Y. Marneffe, M.T. Nguyen, P. Pandard, A. Schifferli, T. Schulze, S. Seidensticker, T.-B. Seiler, J. Tang, R. van der Oost, E. Vermeirssen, R. Zounková, N. Zwart, H. Hollert, Bioassay battery interlaboratory investigation of emerging contaminants in spiked water extracts – towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring, Water Res. 104 (2016) 473–484. https://doi.org/10.1016/j.watres.2016.08.018.
[19] M. Hecker, H. Hollert, Endocrine disruptor screening: regulatory perspectives and needs, Environ. Sci. Eur. 23 (2011) 15. https://doi.org/10.1186/2190-4715-23-15.
[20] RIWA, T. Slootweg, C. J Houtman, Evaluatie van hormonale activiteit gemeten in de Rijn bij Lobith (2010-2011), 2012. http://www.riwa-rijn.org/evaluatie-van-hormonale-activiteit-gemeten-in-de-rijn-bij-lobith-2010-2011/.
[21] M. König, B.I. Escher, P.A. Neale, M. Krauss, K. Hilscherová, J. Novák, I. Teodorović, T. Schulze, S. Seidensticker, M.A. Kamal Hashmi, J. Ahlheim, W. Brack, Impact of untreated wastewater on a major European river evaluated with a combination of in vitro bioassays and chemical analysis, Environ. Pollut. 220 (2017) 1220–1230. https://doi.org/10.1016/j.envpol.2016.11.011.
[22] B. Jarošová, A. Erseková, K. Hilscherová, R. Loos, B.M. Gawlik, J.P. Giesy, L. Bláha, Europe-wide survey of estrogenicity in wastewater treatment plant effluents: the need for the effect-based monitoring, Environ. Sci. Pollut. Res. Int. 21 (2014) 10970–10982. https://doi.org/10.1007/s11356-014-3056-8.
[23] F.D.L. Leusch, C. De Jager, Y. Levi, R. Lim, L. Puijker, F. Sacher, L.A. Tremblay, V.S. Wilson, H.F. Chapman, Comparison of five in vitro bioassays to measure estrogenic activity in environmental waters, Environ. Sci. Technol. 44 (2010) 3853–3860.
[24] P. Schröder, B. Helmreich, B. Škrbić, M. Carballa, M. Papa, C. Pastore, Z. Emre, A. Oehmen, A. Langenhoff, M. Molinos, J. Dvarioniene, C. Huber, K.P. Tsagarakis, E. Martinez-Lopez, S.M. Pagano, C. Vogelsang, G. Mascolo, Status of hormones and painkillers in wastewater effluents across several European states—considerations for the EU watch list concerning estradiols and diclofenac, Environ. Sci. Pollut. Res. 23 (2016) 12835–12866. https://doi.org/10.1007/s11356-016-6503-x.
[25] P.Y. Kunz, C. Kienle, M. Carere, N. Homazava, R. Kase, In vitro bioassays to screen for endocrine active pharmaceuticals in surface and wastewaters, J. Pharm. Biomed. Anal 106 (2015) 107–115. https://doi.org/10.1016/j.jpba.2014.11.018.
[26] R. Kase, P.Y. Kunz, H. Hollert, I. Werner, Contribution on Bioanalytical Assays for Steroidal Estrogens, in EU JRC report by R. Loos "Analytical Methods Relevant to the European Commission's 2012 Proposal on Priority Substances Under the Water Framework Directive, 2012. https://doi.org/10.2788/51497.
[27] A.-S. Wernersson, M. Carere, C. Maggi, P. Tusil, P. Soldan, A. James, W. Sanchez, V. Dulio, K. Broeg, G. Reifferscheid, S. Buchinger, H. Maas, E. Van Der Grinten, S. O'Toole, A. Ausili, L. Manfra, L. Marziali, S. Polesello, I. Lacchetti, L. Mancini, K. Lilja, M. Linderoth, T. Lundeberg, B. Fjällborg, T. Porsbring, D.J. Larsson, J. Bengtsson-Palme, L. Förlin, C. Kienle, P. Kunz, E. Vermeirssen, I. Werner, C.D. Robinson, B. Lyons, I. Katsiadaki, C. Whalley, K. den Haan, M. Messiaen, H. Clayton, T. Lettieri, R.N. Carvalho, B.M. Gawlik, H. Hollert, C. Di Paolo, W. Brack, U. Kammann, R. Kase, The European technical report on aquatic effect-based monitoring tools under the water framework directive,, Environ. Sci. Eur. 27 (2015) 7. https://doi.org/10.1186/s12302-015-0039-4.
[28] A.-S. Wernersson, et al., Technical Report on Aquatic Effect-Based Monitoring Tools, Technical Report 2014–077, 2014. https://doi.org/10.2779/7260.
[29a] C. Heiss, B. Gawlik, B. von Danwitz, D. Wilke, G. Reifferscheid, G. Maack, H. Hollert, L. Portugall, M. Fuerhacker, M. Carere, N. Homazava, O. Perceval, P. Kunz, R. Kase, S. Buchinger, Selim, V. Dulio, Recommendation for a Monitoring Strategy for Estrogens in Coastal and Continental Surface Waters, 2013. http://www.bafg.de/DE/05_Wissen/02_Veranst/2013/2013_02_27_votum_en.pdf?__blob=publicationFile. (Accessed 30 May 2017);
[b] Valeria Dulio, Robert Kase, Recommendations for the use of effect-based methods for monitoring of estrogens in surface waters – Outcomes of the Dübendorf Workshop (21 June 2017), 2017. Available at, http://www.ecotoxcentre.ch/media/178820/estrogen-monitoring-recommendations-2017_final.pdf. (Accessed 23 March 2018).
[30] S. Jobling, R. Williams, A. Johnson, A. Taylor, M. Gross-Sorokin, M. Nolan, C.R. Tyler, R. van Aerle, E. Santos, G. Brighty, Predicted exposures to steroid estrogens in U.K. Rivers correlate with widespread sexual disruption in wild fish populations, Environ. Health Perspect. 114 (2005) 32–39. https://doi.org/10.1289/ehp.8050.
[31] N. Bolong, A.F. Ismail, M.R. Salim, T. Matsuura, A review of the effects of emerging contaminants in wastewater and options for their removal,, Desalination 239 (2009) 229–246.
[32] M.O. Barbosa, N.F.F. Moreira, A.R. Ribeiro, M.F.R. Pereira, A.M.T. Silva, Occurrence and removal of organic micropollutants: an overview of the watch list of EU Decision 2015/495, Water Res. 94 (2016) 257–279. https://doi.org/10.1016/j.watres.2016.02.047.
[33] B. Jarosova, L. Blaha, J.P. Giesy, K. Hilscherova, What level of estrogenic activity determined by in vitro assays in municipal wastewaters can be considered as safe? Environ. Int. 64 (2014) 98–109. https://doi.org/10.1016/j.envint.2013.12.009.
[34] F.D.L. Leusch, H.F. Chapman, M.R. van den Heuvel, B.L.L. Tan, S.R. Gooneratne, L.A. Tremblay, Bioassay-derived androgenic and estrogenic activity in municipal sewage in Australia and New Zealand, Ecotoxicol. Environ. Saf. 65 (2006) 403–411. https://doi.org/10.1016/j.ecoenv.2005.07.020.
[35] A.C. Johnson, R.J. Williams, A model to estimate influent and effluent concentrations of estradiol, estrone, and ethinylestradiol at sewage treatment works, Environ. Sci. Technol. 38 (2004) 3649–3658. http://www.ncbi.nlm.nih.gov/pubmed/15296317.
[36] S. Könemann, R. Kase, E. Simon b, K. Swart, S. Buchinger, M. Schlüsener, H. Hollert, B.I. Escher, I. Werner, S. Aît-Aïssa, V. Dulio, S. Valsecchi, S. Polesello, P. Behnisch, B. Javurkova, O. Perceval, C. Di Paolo, D. Olbrich, S. Tavazzi, E. Sychrova, M. Gundlach, R. Schlichting, L. Leborgne, M. Clara, C. Scheffknecht, Y. Marneffe, C. Chalon, P. Tušil, P. Soldàn, B. von Danwitz, J. Schwaiger, A. Moran Palao, I. San Martín Becares, F. Bersani, E. Vermeirssen, K. Hilscherová, G. Reifferscheid, M. Carere, Effect-Based and Chemical Analytical Methods to Monitor Estrogens Under the European Water Framework Directive, TrAC Trends Anal. Chem 102 (May 2018) 225–235. https://doi.org/10.1016/j.trac.2018.02.008 (publication with shared first authorship).
[37] F. Martini, C. Fragakis, S. Michel, A. Natacha, G. Kinga, R. Kase, G. Monacelli, Y. Reyjol, R. Ward, Science–Policy Interface in Support of the Water Framework Directive , 2013. https://doi.org/10.2777/35629. CIS-SPI activity report 2010-12, Luxembourg.
[38] R. Van der Oost, G. Sileno, M. Suárez-Muñoz, M.T. Nguyen, H. Besselink, B. Bram, SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality assessment; Part I: model design and effect-based trigger values, Environ. Toxicol. Chem. 36 (9) (2017 Sep) 2385–2399. https://doi.org/10.1002/etc.3836.
[39] OECD, Test Guideline 455, Stably Transfected Human Estrogen Receptor-α Transcriptional Activation Assay for Detection of Estrogenic Agonist-Activity of Chemicals, 2009. https://ntp.niehs.nih.gov/iccvam/suppdocs/feddocs/oecd/oecd-tg455.pdf. (Accessed 13 April 2017).
[40] R. Huang, S. Sakamuru, M.T. Martin, D.M. Reif, R.S. Judson, K.A. Houck, W. Casey, J.H. Hsieh, K.R. Shockley, P. Ceger, J. Fostel, K.L. Witt, W. Tong, D.M. Rotroff, T. Zhao, P. Shinn, A. Simeonov, D.J. Dix, C.P. Austin, R.J. Kavlock, R.R. Tice, M. Xia, Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway, Sci. Rep. 4 (2014). https://doi.org/10.1038/srep05664.
[41] A. Kortenkamp, Ten years of mixing cocktails: a review of combination effects of endocrine-disrupting chemicals, Environ. Health Perspect. 115 (2007) 98–105. https://doi.org/10.1289/ehp.9357.
[42] A. Kortenkamp, Low dose mixture effects of endocrine disrupters and their implications for regulatory thresholds in chemical risk assessment, Curr. Opin. Pharmacol. 19 (2014) 105–111. https://doi.org/10.1016/j.coph.2014.08.006.
[43] A. Kortenkamp, T. Backhaus, M. Faust, State of the art report on mixture toxicity. http://ec.europa.eu/environment/chemicals/effects/pdf/report_mixture_toxicity.pdf, 2009.
[44] L. Sachs, Angewandte Statistik: Anwendung Statistischer Methoden, ninth ed., Springer-Verlag, 1999.
[45] Off.J. Eur. Union L 201, COMMISSION DIRECTIVE 2009/90/EC of 31 July 2009 Laying Down, Pursuant to Directive 2000/60/EC of the European Parliament and of the Council, Technical Specifications for Chemical Analysis and Monitoring of Water status, 2009, p. 36. n.d.
[46] B.I. Escher, F.D.L. Leusch, Bioanalytical Tools in Water Quality Assessment, IWA publishing, London, 2012.
[47] P.Y. Kunz, E. Simon, N. Creusot, B.S. Jayasinghe, C. Kienle, S. Maletz, A. Schifferli, C. Schönlau, S. Aït-Aïssa, N.D. Denslow, H. Hollert, I. Werner, E.L.M. Vermeirssen, Effect-based tools for monitoring estrogenic mixtures: evaluation of five in vitro bioassays, Water Res. 110 (2017) 378–388. https://doi.org/10.1016/j.watres.2016.10.062.
[48] D.J. Caldwell, B. Mertens, K. Kappler, T. Senac, R. Journel, P. Wilson, R.D. Meyerhoff, N.J. Parke, F. Mastrocco, B. Mattson, R. Murray-Smith, D.G. Dolan, J.O. Straub, M. Wiedemann, A. Hartmann, D.S. Finan, A risk-based approach to managing active pharmaceutical ingredients in manufacturing effluent, Environ. Toxicol. Chem. 35 (2016) 813–822. https://doi.org/10.1002/etc.3163.
[49] B. Jarosova, L. Blaha, B. Vrana, T. Randak, R. Grabic, J.P. Giesy, K. Hilscherova, Changes in concentrations of hydrophilic organic contaminants and of endocrine-disrupting potential downstream of small communities located adjacent to headwaters, Environ. Int. 45 (2012) 22–31. https://doi.org/10.1016/j.envint.2012.04.001.
[50] M.D. Jurgens, K.I.E. Holthaus, A.C. Johnson, J.L. Smith, M. Hetheridge, R.J. Williams, The potential for estradiol and ethinylestradiol degradation in English rivers, Environ. Toxicol. Chem. 21 (2002) 480–488. http://www.ncbi.nlm.nih.gov/pubmed/11883412. (Accessed 13 April 2017).
[51] F.D.L. Leusch, P.A. Neale, A. Hebert, M. Scheurer, M.C.M. Schriks, Analysis of the sensitivity of in vitro bioassays for androgenic, progestagenic, glucocorticoid, thyroid and estrogenic activity: suitability for drinking and environmental waters, Environ. Int. 99 (2017) 120–130. https://doi.org/10.1016/j.envint.2016.12.014.
[52] E.J. Tiedeken, A. Tahar, B. McHugh, N.J. Rowan, Monitoring, sources, receptors, and control measures for three European Union watch list substances of emerging

concern in receiving waters – a 20year systematic review, Sci. Total Environ. 574 (2017) 1140–1163. https://doi.org/10.1016/j.scitotenv.2016.09.084.
[53] A.C. Mehinto, A. Jia, S.A. Snyder, B.S. Jayasinghe, N.D. Denslow, J. Crago, D. Schlenk, C. Menzie, S.D. Westerheide, F.D.L. Leusch, K.A. Maruya, Inter-laboratory comparison of in vitro bioassays for screening of endocrine active chemicals in recycled water, Water Res. 83 (2015) 303–309. https://doi.org/10.1016/j.watres.2015.06.050.
[54] B.I. Escher, M. Allinson, R. Altenburger, P.A. Bain, P. Balaguer, W. Busch, J. Crago, N.D. Denslow, E. Dopp, K. Hilscherova, A.R. Humpage, A. Kumar, M. Grimaldi, B.S. Jayasinghe, B. Jarosova, A. Jia, S. Makarov, K.A. Maruya, A. Medvedev, A.C. Mehinto, J.E. Mendez, A. Poulsen, E. Prochazka, J. Richard, A. Schifferli, D. Schlenk, S. Scholz, F. Shiraishi, S. Snyder, G. Su, J.Y.M. Tang, B. van der Burg, S.C. van der Linden, I. Werner, S.D. Westerheide, C.K.C. Wong, M. Yang, B.H.Y. Yeung, X. Zhang, F.D.L. Leusch, Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro bioassays, Environ. Sci. Technol. 48 (2014) 1940–1956. https://doi.org/10.1021/es403899t.
[55] T.F.T. Omar, A. Ahmad, A.Z. Aris, F.M. Yusoff, Endocrine disrupting compounds (EDCs) in environmental matrices: review of analytical strategies for pharmaceuticals, estrogenic hormones, and alkylphenol compounds, TrAC Trends Anal. Chem. 85 (2016) 241–259. https://doi.org/10.1016/j.trac.2016.08.004.
[56] C. Kienle, R. Kase, I. Werner, Evaluation of Bioassays and Wastewater Quality: in vitro and in vivo Bioassays for the Performance Review in the Project “Strategy MicroPoll,”, 2011. http://www.oekotoxzentrum.ch/media/2229/2011_kienle_bioassays_micropoll.pdf. (Accessed 13 April 2017).
[57] A.R. Ribeiro, O.C. Nunes, M.F.R. Pereira, A.M.T. Silva, An overview on the advanced oxidation processes applied for the treatment of water pollutants defined in the recently launched Directive 2013/39/EU, Environ. Int. 75 (2015) 33–51. https://doi.org/10.1016/j.envint.2014.10.027.
[58] J. Rivera-Utrilla, M. Sánchez-Polo, M.Á. Ferro-García, G. Prados-Joya, R. Ocampo-Pérez, Pharmaceuticals as emerging contaminants and their removal from water. A review, Chemosphere 93 (2013) 1268–1287. https://doi.org/10.1016/j.chemosphere.2013.07.059.
[59] EU WG Chemicals, ACTIVITY: EFFECT-BASED TOOLS/METHODS for WG Chemicals as Part of the Water Framework Directive CIS Work Programme (2016-2018) Endorsed by the Water Directors, 2016. Terms of References for the Effect-Based Activity - 08-07-2016.docx, https://circabc.europa.eu/sd/a/e39a14f1-ae0a-4ec7-9dec-2a8c6c98645a/Draft. (Accessed 13 April 2017).
[60] EU WG Chemicals, Priority Substances – Outcome of Water Directors Meeting, 2016. Chem 2016–12 (10) WD2016 outcome_Priority Substances.pdf, https://circabc.europa.eu/sd/a/096c7e10-1be3-4ffa-b5d3-e8cab1b86c7f/WG. (Accessed 13 April 2017).
[61] K. Bernhard, C. Stahl, R. Martens, H.-R. Köhler, R. Triebskorn, M. Scheurer, M. Frey, Two novel real time cell-based assays quantify beta-blocker and NSAID specific effects in effluents of municipal wastewater treatment plants, Water Res. 115 (2017) 74–83. https://doi.org/10.1016/j.watres.2017.02.036.
[62] R. Kase, A. Duffek, Applied Effect-Based Monitoring of EU Watch List Samples, 2017. https://circabc.europa.eu/w/browse/209ca179-1ab3-4dc1-81f9-0b3e7fdb8885. (Accessed 13 April 2017).