Effect-based and chemical analytical methods to monitor estrogens under the European Water Framework Directive

Effect-based and chemical analytical methods to monitor estrogens under the European Water Framework Directive

Sarah Könemann‡ [a,b,*], Robert Kase‡ [b], Eszter Simon [b], Kees Swart [c], Sebastian Buchinger [d], Michael Schlüsener [d], Henner Hollert [a], Beate I. Escher [e,f], Inge Werner [b], Selim Aït-Aïssa [g], Etienne Vermeirssen [b], Valeria Dulio [g], Sara Valsecchi [h], Stefano Polesello [h], Peter Behnisch [c], Barbora Javurkova [i], Olivier Perceval [k], Carolina Di Paolo [a], Daniel Olbrich [b], Eliska Sychrova [i], Rita Schlichting[e], Lomig Leborgne [l], Manfred Clara [m], Christoph Scheffknecht [n], Yves Marneffe [o], Carole Chalon [o], Petr Tušil [p], Přemysl Soldàn [p], Brigitte von Danwitz [q], Julia Schwaiger [r], Maria Isabel San Martín Becares [s], Francesca Bersani [t], Klara Hilscherová [i], Georg Reifferscheid [d], Thomas Ternes [d], Mario Carere [u]

[a] *Institute for Environmental Research, RWTH Aachen University, Worringerweg 1, 52074 Aachen, DE;*

[b] *Swiss Centre for Applied Ecotoxicology Eawag-EPFL, Überlandstrasse 131, 8600 Dübendorf, CH;*

[c] *BioDetection Systems b.v., Science Park 406, 1098 XH Amsterdam, NL;*

[d] *Bundesanstalt für Gewässerkunde, Am Mainzer Tor 1, 56068 Koblenz, DE;*

[e] *Helmholtz Centre for Environmental Research - UFZ, Permoserstrasse 15, 04318 Leipzig, DE;*

[f] *Eberhard Karls University Tübingen, Environmental Toxicology, Center for Applied Geosciences, 72074 Tübingen, DE;*

[g] *INERIS, Rue Jaques Taffanel, Parc Technologique ALATA, 60550 Verneuil-en-Halatte, FR;*

[h] *Istituto di Ricerca sulle Acque, Via del Mulino 19, 20861 Brugherio (MB), IT;*

[i] *Masaryk University, Research Centre for Toxic Compounds in the Environment (RECETOX), Kamenice 753/5, 625 00 Brno, CZ;*

[k] *French National Agency for Water and Aquatic Environments, 5 Square Felix Nadar, 94300 Vincennes, FR;*

[l] *Agence de l'eau Adour-Garonne, 90 rue de Fétéra, CS 87801, 31078 Toulouse Cedex 4, FR;*

[m] *Umweltbundesamt, Spittelauer Lände 5, 1090 Wien, AT;*

[n] *Umweltinstitut, Institut für Umwelt und Lebensmittelsicherheit des Landes Vorarlberg, Montfortstraße 4,6901 Bregenz, AT;*

[o] *Institut Scientifique de Service Public (ISSeP), Rue Chéra 200, 4000 Liège, BE;*

[p] *T.G. Masaryk Water Research Institute, Podbabská 2582/30, Praha 6, 16000, CZ;*

[q] *Landesamt für Natur, Umwelt und Verbraucherschutz NRW (LANUV), Auf dem Draap 25, 40221 Düsseldorf, DE;*

[r] *Bayerisches Landesamt für Umwelt, Demollstrasse 32, 82407 Wielenbach, DE;*

[s] *Instituto de Recursos Naturales, Universidad de León, Avenida de Portugal 42, 24071 León, ES;*

[t] *Centro Ricerche (SMAT), Società Metropolitana Acque Torino S.p.A.C. so Unità d'Italia 235/3, 10127 Torino, IT;*

[u,*] *National Institute of Health, Department Environment and Health, Roma, IT;*

‡*Authors contributed equally to this manuscript.*

**Corresponding author:*
Sarah Könemann
Institute for Environmental Research, RWTH Aachen University
Worringerweg 1, 52074 Aachen, Germany
Tel.:+41587655254
E-mail: sarah.koenemann@rwth-aachen.de

Detailed description of the authors contributions can be found in the Supplementary Information (SI, Table S1).

Abstract

The European Decision EU 2015/495 included three steroidal estrogens, estrone, 17β-estradiol and 17α-ethinyl estradiol, in the "watch-list" of the Water Framework Directive (WFD). As consequence, these substances have to be chemically monitored at the level of their environmental quality standards, which can be challenging. This project aimed to identify reliable effect-based methods (EBMs) for screening of endocrine disrupting compounds, to harmonise monitoring and data interpretation methods, and to contribute to the current WFD review process. Water and wastewater samples were collected across Europe and analysed using chemical analyses and EBMs. The results showed that 17β-estradiol equivalents were comparable among methods, while results can vary between methods based on the relative potencies for individual substances. Further, derived 17β-estradiol equivalents were highly correlated with LC-MS/MS analyses. This study shows that the inclusion of effect-based screening methods into monitoring programmes for estrogens in surface waterbodies would be a valuable complement to chemical analysis.

Keywords

Science-policy interface	Estrogen screening	Endocrine disruption
Surface and waste water assessment	Emerging pollutants	EU watch-list
Steroid analyses	*In vitro* bioassays	Integrated effects of mixtures

1 State of the Art

Over the past two decades, numerous scientific studies have demonstrated that endocrine disrupting chemicals (EDCs) elicit adverse effects on sensitive aquatic species, such as fish [1-7]. Steroidal estrogens, like the natural hormones estrone (E1) and 17β-estradiol (E2), as well as the synthetic hormone 17α-ethinyl estradiol (EE2), are of particular environmental concern [8-11]. Due to their steady release via waste water effluents into surface waters [12, 13] and their high biological activity, even very low concentrations of E2 and EE2 have been shown to cause reproductive toxicity with negative effects at the population level [14-16]. As a consequence, E1, E2, and EE2 were included in a European Union (EU) Water Framework Directive (WFD) "watch-list" [17-20]. The WFD watch-list mechanism aims to collect high-quality monitoring data on concentrations of emerging pollutants and potentially hazardous substances, whose currently available monitoring information shows either quantitative or qualitative deficiencies [21]. To collect more high-quality data, listed substances have to be monitored at representative EU sampling sites for a period of at least 12 and up to 48 months. The watch-list mechanism is expected to support future substance prioritisation processes, enable the implementation of measures, and facilitate environmental risk assessment across the EU.

Chemical monitoring of estrogens for the watch-list mechanism is challenging, because the European Commission set maximum acceptable method detection limits (MDLs) at EQS levels of 400 pg/L for E1 and E2, and 35 pg/L for EE2 [18, 22]. Most routine analytical methods used by the Member States cannot meet these requirements, especially for EE2, based on [23, 24]. Hence, the quality assessment of water bodies based on current methods is a challenge for the detection/quantification limits that are too high to detect if EQS are being exceeded or not. Effect-based methods are able to detect estrogenic substances at sub-ng or even pg levels and have the potential to be used as a complementary screening tool [12, 25-27]. In addition, they do not require *a priori* knowledge of the substances to be monitored, as they are able to determine the biological response caused by complex mixtures of unknown compounds. Thus, effect-based methods may be suitable to serve as a valuable link between chemical analytical and ecological quality assessments, since the effects can rarely be linked to individual compounds.

As described in an EU technical report, which was elaborated in the context of the Chemical Monitoring and Emerging Pollutants (CMEP) expert group under the Common Implementation Strategy (CIS) of the WFD, effect-based tools can be categorised into three main groups: Bioassays (*in vitro*, *in vivo*), biomarkers, and ecological methods [28]. With regard to steroidal estrogens and other EDCs, *in vitro* reporter gene assays have been used predominantly to determine the total estrogen receptor (ER) mediated estrogenicity of an environmental sample [29]. Among the most commonly applied assays are *in vitro* methods such as estrogen receptor transactivation assays (ER-TAs), which use various cell types including yeast, human and other mammalian cell lines that were transfected with a human estrogen receptor coupled to a reporter gene [30]. Activation of the ER leads to the expression of the reporter

105 gene product, usually an enzyme that modifies another chemical, causing a quantifiable response. The
106 resulting estrogenic potential of a sample is expressed as an E2 equivalent concentration (EEQ),
107 indicating the estrogenic activity of the sample or sample dilution in terms of equivalency to the
108 estrogenic activity of the corresponding E2 reference concentration [31].

109 Although ER-TAs are highly advantageous methods for the detection of ER activation and
110 quantification of very low estrogen concentrations in surface waters [23], these methods are not included
111 within current WFD monitoring programmes [20]. One reason for this is the lack of data that
112 demonstrate their applicability as a monitoring and screening tool in combination with chemical
113 analytical methods (see e.g. [14]). Such information would greatly increase their regulatory acceptance.
114 As a response to this need, an EU-wide project involving 24 research organisations and environmental
115 agencies from 12 countries was carried out to evaluate the usefulness of specific *in vitro* methods for
116 identifying the presence of the watch-list substances, E1, E2, and EE2, in surface and waste waters. The
117 project aimed to compare the chemical and effect-based data resulting from the analysis of 16 surface
118 and 17 waste water treatment plant effluent samples. Analyses were conducted in seven participating
119 laboratories using different LC/MS- (three laboratories) and effect-based methods (five laboratories).
120 The objectives of the study were (i) the demonstration of reliable effect-based screening methods for the
121 monitoring of estrogenic EDCs in waste water and surface water, (ii) the harmonisation of data
122 interpretation methods, and (iii) providing recommendations for the implementation of cost-effective
123 and reliable effect-based methods in WFD monitoring programmes.

124 # 2 The Project

125 ## 2.1 Sampling

126 A total number of 16 surface water (SW) and 17 waste water (WW) samples were collected according to
127 a protocol developed by the participants (SI, Part A). Selected sampling sites were located in seven
128 European countries in Central and Southern Europe (Figure 1): Austria (1 SW/ 3 WW), Belgium (2/2),
129 Czech Republic (2/2), France (1/1), Germany (4/4), Italy (5/3), and Spain (1/2). Sample collection was
130 carried out from September to November 2015 by ten participating institutions. The samples were taken
131 based on prior knowledge on their contamination with estrogens and represented a gradient of
132 contamination from high to moderate.

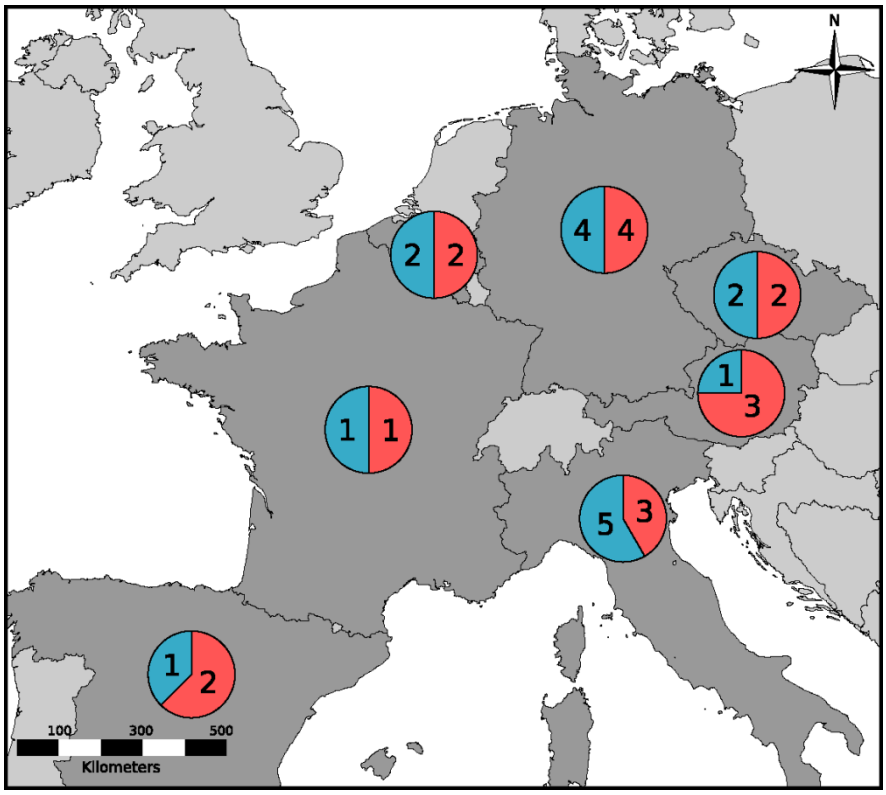


Figure 1: Samples taken in various European States (dark grey). The circles indicate the number of surface water (blue) and waste water samples (red) taken in each country.

2.2 Sample preparation

The sample preparation included the filtering of a part of the SW (see SI, Part A) and all WW samples over glass fibre filters (Millipore, type 4, retention 2.7 µm, circle size 4.7 cm). Since a filtration step can have an impact on the composition of a sample and its estrogenic activity [32], the filtration step was investigated during a feasibility study prior to the main study presented here. The results of the pre study did neither show a significant reduction in estrogenicity in the control nor in tested environmental samples (data not shown). Subsequently, all samples were enriched by means of solid-phase extraction (SPE; 11 L sample to 11 mL extract) and extracts were passed over silica gel (SiOH) columns (methods focusing on E1, E2 and EE2). While for surface water each extract was split into eleven 1 mL aliquots that were each passed over a single SiOH column, for waste water a single column was inadvertently used to treat the whole extract (11 mL). For LC-MS/MS analysis this means that matrix was less efficiently removed from WW extracts (relative to SW extracts) and higher matrix loads would have impeded low LOQs in WW LC-MS/MS analysis. For bioassay analysis this means that, should additional ER-agonists (i.e. other than E1, E2 and EE2) have been present in the extracts, a reduced clean-up efficiency would have reduced ER-agonist removal which in turn would have caused enhanced effects in bioassays. Full details of sample preparation are provided in SI, Part A.

2.3 Chemical and effect-based analyses

Participating laboratories received spiked reference samples, blanks and encoded water extracts. The chemical analyses were conducted in three different labs, which applied an LC-MS/MS with negative electron spray ionisation (detailed information in SI, Part D Table S2). The effect-based methods were conducted in five different labs: Estrogen Receptor Chemical Activated LUciferase gene eXpression (ER-CALUX) at Biodetection Systems (BDS), luciferase-transfected human breast cancer cell line (MELN) gene-reporter assay at INERIS [33], ER-GeneBLAzer assay at the Helmholtz Centre for Environmental Research (UFZ) [34], the stably transfected human estrogen receptor-alpha transcriptional activation Assay using hERα-HeLa-9903 cells (HeLa-9903 assay) at RECETOX [35], and planar Yeast Estrogen Screen (pYES) at the German Federal Institute of Hydrology (BfG) [36, 37]. The pYES is a method, which combines a chromatographic separation of the sample by thin layer chromatography (TLC) with a subsequent performance of the YES on the planar surface of the TLC-plate [38-40]. Like the common assays which are performed in micro-well-plates, this approach allows the quantification of the overall estrogenic activity present in the sample by means of E2-equivalence concentrations. Furthermore, like methods based on LC/MS, it also allows the estimation of concentrations of individual estrogenic compounds, e.g. E1, E2 and EE2, due to the chromatographic separation of the sample. For this purpose the respective standard compounds are used for a calibration on the same TLC plate – in the present study E1, E2, EE2, and estriol (E3) were applied in a mixture at three different levels. Due to the limited separation power of the thin layer chromatography compared to HPLC and GC in particular, a co-migration of estrogenic compounds cannot be excluded. Therefore, under the assumption of effect addition, the estimated individual concentrations represent the possible maximal concentration of the respective compound. This approach can be used to identify and quantify substance groups causing ER-activation.

2.4 Blanks and positive controls

Ultrapure water (11 L) was used as extraction blank. An extraction blank was included with each extraction run of 10 samples, subjected to clean-up and distributed the same as the sample extracts. Further, each analysis using effect-based methods included a negative control. To avoid solvent effects on cell viability, its concentrations did not exceed a defined value (see SI, Part D Table S3). As positive controls for ensuring the validity and enabling a comparison of the methods, surface water samples (11 L each) from the Netherlands were spiked with E2 and EE2 at two concentrations by the central lab (BDS). The “low spike” (600 pg/L) represented a concentration slightly above the proposed EQS for E2 (400 pg/L). The “high spike” (6000 pg/L) represented a concentration that is quantifiable with high certainty by both effect-based and chemical methods.

2.5 Data evaluation – effect-based methods

Raw data and information on relative enrichment factors (REF) of the extracts were collected from participating laboratories. The REF expresses the combination of: 1) sample enrichment using SPE and 2) extract dilution steps in each of the applied effect-based methods. Estrogenic activity of the extracts was expressed as E2-equivalence concentration (pg EEQ/L water) (described in detail in SI, Part B). Briefly, dose-response curves of the reference compound, E2, and the dilution series of the water extracts and blanks were fitted using a five-parametric non-linear regression with normalised data. The concentration of the positive control (E2) needed to induce 10 % effect of the maximum E2-induction (PC_{10}), was calculated. Subsequently, the relative REF of the sample, that stimulates the assay at PC_{10} level was determined by interpolation. The PC_{10} reference concentration was divided by the corresponding sample dilution (REF) to obtain the EEQ of the sample. EEQs derived by the PC_{10} method are presented in the results section.

2.6 Data evaluation – chemical analysis

Internal standard calibration and interpolation using a linear regression model were performed to determine concentrations (pg/L) of the individual steroidal estrogens in sample extracts. Identification of selected analytes was performed based on two to three Multiple Reaction Monitoring (MRM) transitions between the precursor ion and two or three most abundant product ions, depending on the laboratory where analyses were done. The first transition was used for quantification purposes whereas the second and third transitions were used to confirm the presence of the target compound in the sample. Quantified analytes were identified by comparing the retention time (RT) of the corresponding standard and the ratio between two ion transitions recorded (±20 %) in the standard and water samples.

2.7 Calculation of sample-dependent LOD and LOQ

The Limits of quantification (LOQ) for effect-based methods the LOQs were calculated as 3-fold the standard deviation (SD) of the averaged response of the negative control on each assay plate. The effect level of 3-fold the SD was interpolated from the E2 reference curve and divided by the REF of the sample to derive the LOQ. The actual reporting for effect-based methods occurred at the 10% effect level which was always above LOQ (typically at 2-5 % effect levels).

In case of the chemical analysis the limits of detection (LOD) were determined for each compound in each sample based on the signal intensity of the internal standards or the analyte peak by a signal-to-noise (S/N) ratio of 3:1 and LOQ by a S/N ratio of 10:1.

When comparing LOQs of effect-based methods with those of chemical analyses the various key differences between the two approaches need to be taken into account (for further background see SI, Part C).

2.8 Comparison of chemical and biological analysis

The EEQ_{bio} is the ratio of the effect concentration of the reference compound estradiol EC_{50}(E2) (pg/L) and the sample EC_{50}(sample) (Equation 1) and was derived in this study using the PC_{10} approach (see above). The EEQ_{chem} was calculated from the sum of the relative effect potencies REP_i times the detected concentration of estrogenic chemical i, c_i [41]. The REP, in turn, is the ratio of the effect concentration of the reference compound estradiol EC_{50}(E2) and the chemical i's EC_{50}(i) (Equation 2).

$$\mathrm{EEQbio} = \frac{\mathrm{EC50(E2)}}{\mathrm{EC50(sample)}} \quad (1)$$

$$\mathrm{EEQchem} = \sum_{i=1}^{n} \mathrm{REPi} \cdot \mathrm{ci} = \sum_{i=1}^{n} \frac{\mathrm{EC50(E2)}}{\mathrm{EC50(i)}} \cdot \mathrm{ci} \quad (2)$$

Due to the analytical method detection limits of E2 and EE2, we evaluated the potential contribution of non-detected estrogens to the overall $EEQ_{chem,LOD/2}$ using Equation 3, where values below the LOD ("non-detects") were included as LOD/2. If the analytical lab reported data as <LOQ, we used LOQ/2 in Equation 3 instead of LOD/2. In Equation 3, n refers to the total number of chemicals included in the analysis, m refers to the number of chemicals below LOD. Ci is the average value of three analytical measurements,

$$\mathrm{EEQchem,\,LOD/2} = \sum_{i=1}^{n-m} \mathrm{REPi} \cdot \mathrm{ci} + \sum_{j=1}^{m} \mathrm{REPj} \cdot \mathrm{LODj/2} \quad (3)$$

2.9 Correlation analysis

The correlation analysis among effect-based methods ($EEQ_{bio)}$ was performed with GraphPad Prism, using the Pearson correlation (r). [42].

3 Results and discussion

3.1 Reference chemicals and validation

All essential criteria for method performance were fulfilled in this study (described in more detail in the SI, Part E). As shown in Table S4 (SI, Part E), the chemical analytical as well as effect-based methods showed good recovery in the spiked samples. No estrogenic activity or quantifiable concentrations of E1, E2, and EE2 were measured in the blank samples (i.e. procedure-, extraction- and solvent blanks). As the derived effect concentrations in the effect-based methods and chemically measured EE2 concentrations matched with the nominal concentrations of the spiked samples, the observed effects can be ascribed to the samples themselves.

3.2 Results of chemical analysis

Measured concentrations of the three estrogens E1, E2 and EE2 differed widely between sampling sites as well as between surface and waste water samples. Differences among SW samples can be explained by varying river characteristics, e.g. flow (dilution factor), or temperature, as well as differences in estrogenicity of treated WW, that are released into the SW. The results of the analyses, which are summarised in Figure 2, show a 3.2 to 3.6 times higher mean concentration for E1 and E2 in WW (Figure 2B) compared to SW (Figure 2A). Due to the highly contaminated WW sample M(23), possibly influenced by an industrial discharge of EE2, the mean concentration of EE2 across all WW samples was approximately 20 times higher compared to SW (Figure 2). Estrone (E1) was quantified in all samples. For E1 maximum concentrations of 5.6 ng/L (sample P(7)) and 20.5 ng/L (sample Q(20)) in SW and WW were measured, respectively. E2 was the second most frequently quantified estrogen and measured above LOQ in nine of 16 SW and six of 17 WW samples. Measured concentrations ranged from 0.4 ng/L (sample N(33)) to 1.1 ng/L (sample Q(20)) in WW, and from 0.06 ng/L (sample J(10)) to 0.5 ng/L (sample N(15)) in SW. The synthetic EE2 was least frequently quantified and measured above LOQ in four of 16 SW and four of 17 WW samples with a maximum concentration of 0.3 ng/L in SW sample O(3) and 7.5 ng/L in WW sample M(23). These concentration ranges and patterns are in accordance with recent review studies [43, 44].

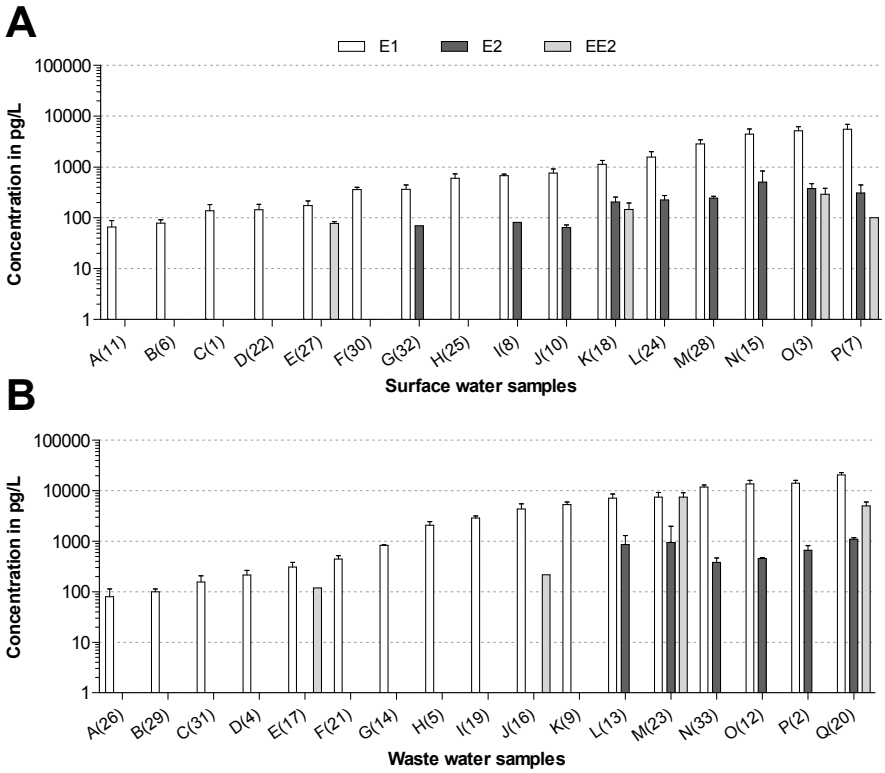


Figure 2: Chemical analytically measured concentrations for SW (A) and WW extracts (B) above LOQ for E1, E2 and EE2. The bars show the mean concentration of all three applied methods for each analyte showing results > LOQ, the standard deviation is shown when two or three methods reported results. The sample-dependent LOQs are listed in the supplementary information together with the measurement data of analytical methods (SI, Part F, Table S6 and S7).

Our results underline the analytical difficulties that have recently been highlighted for E2 and EE2 by several studies and workshops [16, 45], stressing the challenges that emerge for routine methods used in national monitoring programmes. Despite the use of quite advanced chemical analytical techniques (status 2015), the detection and quantification of E2 and EE2 in SW and WW samples was problematic in some cases. While it was possible to quantify E1 in almost all samples, the percentage of quantifications was significantly reduced for E2 and even more for EE2 (Figure 3). This was partially due to the fact that insufficient silica gel was used to reduce the matrix effects in WW. WW is considered as worst-case regarding matrix effects [46, 47].

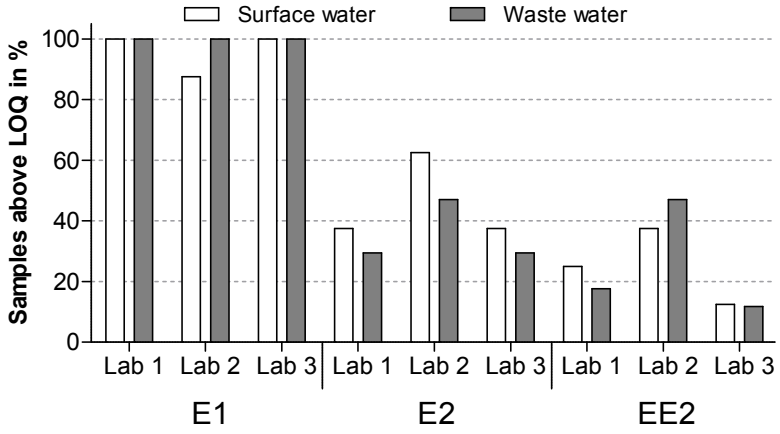


Figure 3: Mean percentage of quantified (>LOQ) samples for each substance in SW and WW. The sample-dependent LOQs are listed in the supplementary information together with the measurement data of the analytical methods (SI Part F, Table S7).

However, the quantification of substances itself is not the only challenge faced by those routinely applying analytical methods for watch-list monitoring. According to the EU Commission Decision 2015/495, which established the first watch-list, the indicative methods applied by Member States have to meet the minimum requirement for method detection limits (MDL) equal to the proposed EQSs of E1 at 3.6 ng/L, E2 at 0.4 ng/L and EE2 at 0.035 ng/L [18]. To take into consideration the matrix effects of different waters, LODs and LOQs had to be calculated for each sample (SI Part F, Table S7). The three techniques used in the current study were able to meet MDL requirements for E1 in all SW and WW samples. Also for E2, in 96 % of surface water samples and 94 % of waste water samples detection was possible at the level of the proposed EQS. In the case of EE2, the minimum criteria were not met, since only 56 % and 16 % of SW and WW samples, respectively, could be monitored at the EQS level. These findings are in accordance with a recent report from 2015, which showed that the lowest LOQ found in literature at that time was sufficient for compliance monitoring of E1 and E2 in inland surface waters, while the criteria were not met for EE2 by several Member States [24]. It has to be pointed out that, in this project, the silica clean-up step for the sample extracts differed between WW and SW samples (see methods section) favouring the presence of polar compounds in extracts of WW samples. This difference likely reduced the sensitivity of the analytical method for the target compounds in WW samples. Furthermore, sample extraction was performed at pH 3 possibly increasing concentrations of

296 humic acids and thus lowering sensitivity of LC/MS-based methods applied. Under ideal conditions, we
297 estimate that analytical methods can achieve LODs and LOQs of a factor 2 to 3 lower in WW samples.
298 It has to be recognised that the LODs of chemical analytical methods used exclusively for steroidal
299 estrogens already significantly decreased from 2013 (LOD E2 and EE2 of 100 pg/L) to 2015 (E2: 60
300 pg/L, EE2: 85 pg/L) and will certainly decrease further [16, 23].

301 Nevertheless, if steroidal estrogens were to be included in the EU priority list for monitoring, very strict
302 minimum performance criteria would apply. As stated in the Commission Directive 2009/90/EC, an
303 analytical method used for monitoring of priority substances needs a LOQ equal or below a value of
304 30 % of the EQS [48]. These requirements can presently be met only for E1, but not for E2 or EE2 in all
305 SW. Regarding the quantification of E2, and EE2, existent routine analytical techniques still lag behind
306 the requirements. This result is supported by two recent reviews on the performance of current analytical
307 methods that have shown that 35 % of reviewed methods complied with the EQS for E2, while only one
308 method complied with the EQS for EE2 [49, 50]. In order to not only detect but also quantify at such
309 low concentrations as required for regulatory monitoring application, a further decrease of LOQs is
310 necessary, which is difficult to achieve for routinely used non-tailored analytical methods in the short-
311 term.

312 ### 3.3 Quantification limits of chemical-analytical and *in vitro* effect-based methods

313 The LOQs for all methods applied in this study are summarised in Figure 4. Since E2 is used as the
314 reference compound for all effect-based methods, the LOQ of E2 is shown for the chemical-analytical
315 methods as an example. When comparing LOQs across the different methods it has to be taken into
316 account that LOQs were derived along different approaches (see method section and SI, Part C for
317 further details). The effect-based *in vitro* methods were generally able to quantify effects at one to two
318 orders of magnitude lower concentrations than the analytical methods used. For effect-based methods,
319 LOQs ranged between 0.002 ng/L and 0.2 ng/L for SW as well as WW, while for chemical-analytical
320 methods LOQs for E2 were 0.04 ng/L to 1.5 ng/L in SW and 0.05 ng/L to 3 ng/L in WW. This increase
321 in LOQs for chemical-analytical methods in WW samples (Figure 4B) compared to surface water
322 (Figure 4A) can be ascribed to the higher complexity of the waste water matrix [46, 47] as well as the
323 less efficient clean-up used for WW samples.

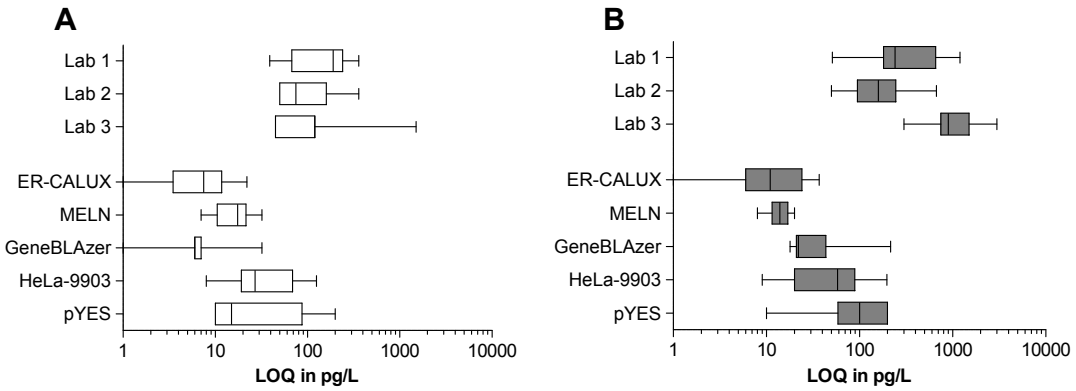
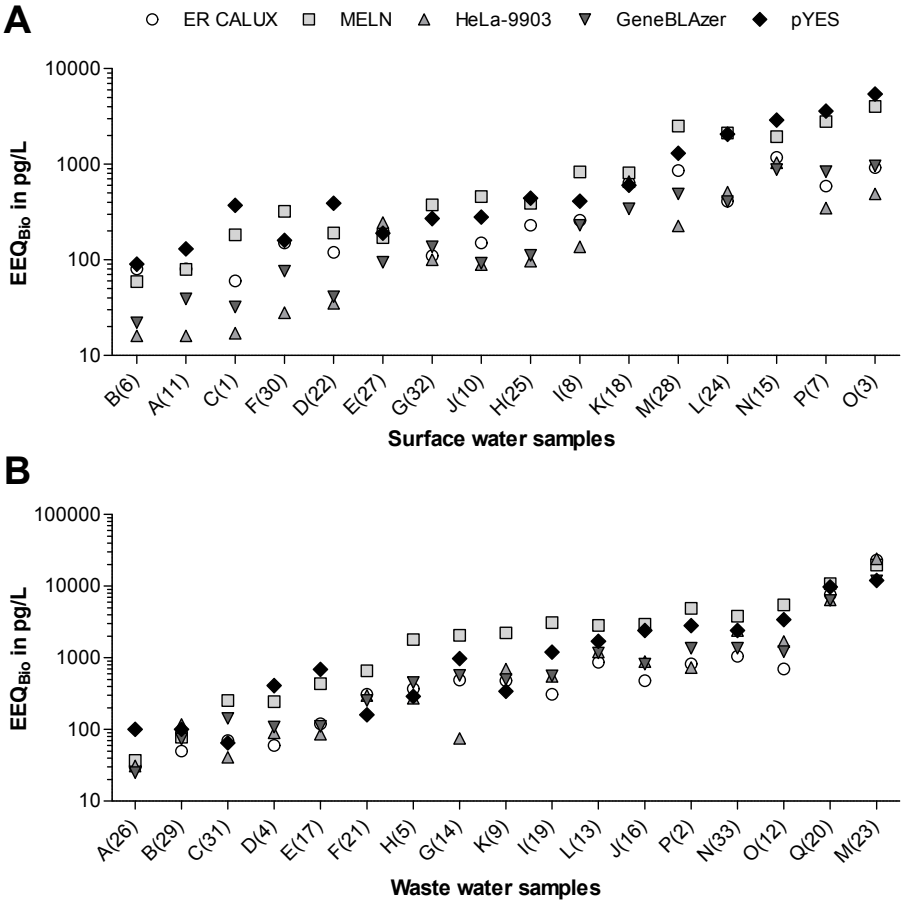


Figure 4: Sample-dependent LOQs in surface water (A) and waste water (B) extracts. For the chemical analytical method the LOQ of E2 is shown as an example and for the effect-based methods the LOQ of the integrated effects is represented. Plots indicate the distribution of data, thereby the bottom and the top of the box are the first and third quartiles, while the line inside the box is the median. The whiskers show the minimum and maximum of all data.

3.4 Measured estrogenic effects

As a result of these low effect-based quantification limits, estrogenic activities were detected in all tested samples. As expected, highest EEQs were measured in WW samples (Figure 5A and B) . In SW, EEQ_{bio} ranged from 0.16 ng/L measured with HeLa-9903 in sample B(6) to up to 5.4 ng/L measured with pYES in sample O(3). In WW, the lowest EEQ_{bio} of 0.03 ng/L was measured in sample A(26) with ER-GeneBLAzer, while the highest EEQ_{bio} of 24 ng/L was measured in sample M(23) with HeLa-9903. Further, it is evident that EEQ_{bio} for SW samples determined with the MELN, as well as the pYES, were higher (> 50 %) than the EEQ_{bio} measured with the other effect-based methods. A possible reason for this pattern, which was less pronounced in WW, could be a higher sensitivity of the MELN and pYES towards E1 (see SI Part F, Table S8), combined with a larger proportion of E1 in surface water. Additionally, alterations in the method's performance occur due to differences between the test systems, which was already mentioned in previous studies [23, 44, 51] and is further discussed for this project in an associated publication [52].



343

344 **Figure 5: Measured E2-equivalents for all SW (A) and WW (B) extracts.** The symbols show the EEQs for each bioassay,
345 which were calculated according to the method described in section 2.5. The sample-dependent LOQs are mentioned in the
346 supplementary information, together with the measurement data of effect-based methods (SI Part F, Table S8 and S9).

347 3.5 Comparison of chemical analysis and *in vitro* effect-based methods

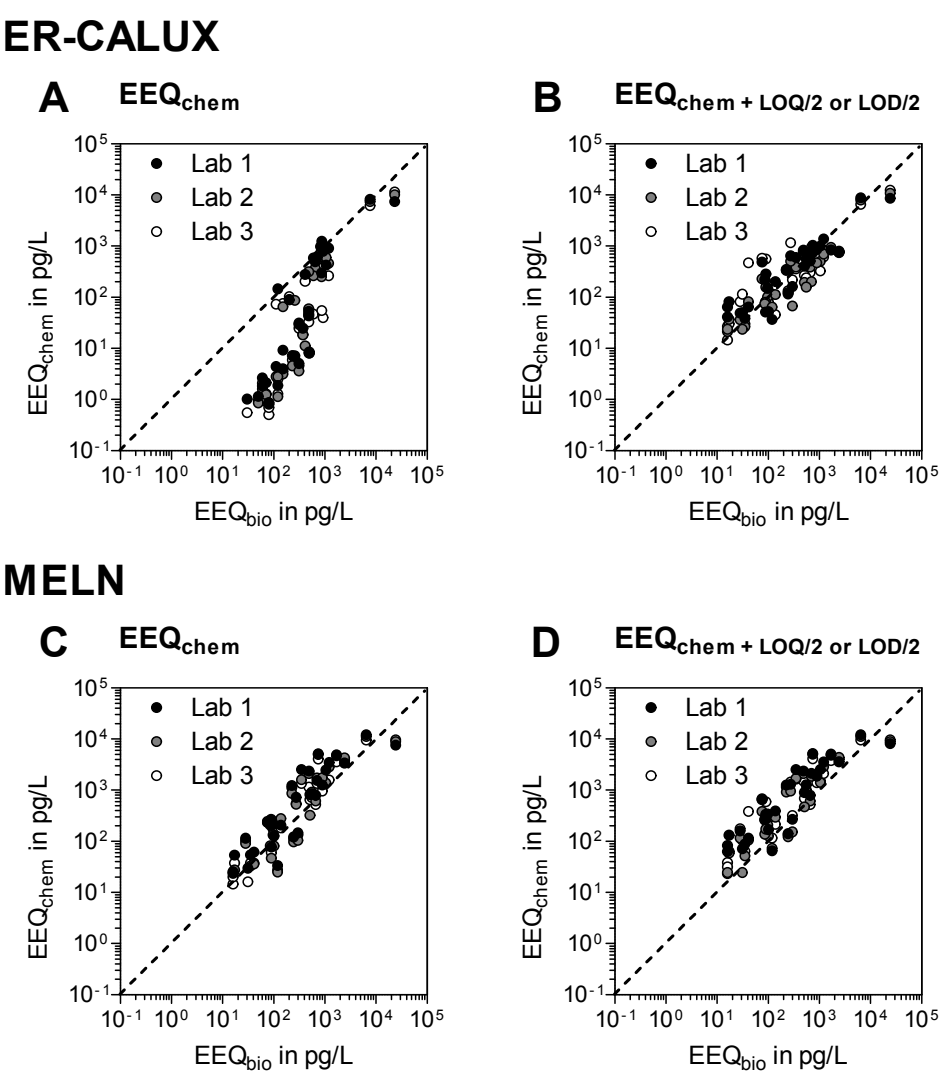
348 We cannot a priori expect consistency between EEQ_{chem} calculated from E1, E2, and EE2 concentrations
349 and EEQ_{bio}. Although the extraction and clean-up method focused on E1, E2, and EE2, other natural
350 estrogens and xenoestrogens (both agonists and antagonists) might still be present in the extracts and
351 contribute to the mixture effects detected by effect-based methods. Thus, there can be situations where
352 EEQ_{chem} is lower than EEQ_{bio} because: 1) agonists other than E1, E2, and EE2 were present in the
353 sample but not quantified by LC-MS/MS analyses or 2) some target compounds were present but below
354 LOQ or LOD, thus they were not included in EEQ_{chem} but still contributed to EEQ_{bio}. Alternatively,
355 EEQ_{chem} can be higher than EEQ_{bio} when antagonists supress the response of the assay.

356 For ER-CALUX, the comparison of EEQ_{bio} with EEQ_{chem} (Figure 6A) indicated an underestimation of
357 EEQ_{bio} by EEQ_{chem} at low concentrations of steroidal estrogens. When E1 concentrations are low,
358 typically E2 and EE2 concentrations are below LOQ (Figure 2). However, as stated above, also below
359 their LOD/LOQ, these chemicals may be present and contribute to the biological mixture effect (i.e.
360 EEQ_{bio}). We therefore also calculated the $EEQ_{chem,LOD/2}$ that uses the LOD/2 or LOQ/2 for those E2 and
361 EE2 concentrations below the LOD or LOQ. The increase in EEQ_{chem}, due to the inclusion of LOQ/2

and LOD/2 data (SI, Part F, Table S10-14), shifts the EEQ_{chem} - EEQ_{bio} data cluster towards the one-to-one line (Figure 6B). In fact, there is now a slight overestimation of the biological effect in the range where EEQ concentrations are low (up to ca.100 pg/L). The fact that the agreement between EEQ_{chem} and EEQ_{bio} has become much better (going from Figure 6A to 6B) is a good indication that E2 and EE2 are indeed present and were captured by effect-based methods.

The situation for MELN is markedly different from that of ER-CALUX. For MELN the direct comparison between EEQ_{chem} and EEQ_{bio} is already very good (Figure 6C). In fact, EEQ_{chem} tends to be above EEQ_{bio} already before adding the additional EEQ_{chem} component using LOD/2 or LOQ/2 for E2 and EE2. The inclusion of LOD/2 or LOQ/2 in the EEQ_{chem} calculation caused a notable overestimation of EEQ_{chem} for almost all samples (>90 % of data above the 1 to 1 line in Figure 6C). The other three bioassays show results that are intermediate between ER-CALUX and MELN, with a general trend towards a slight underestimation of EEQ_{chem} for samples with low EEQ_{bio} and an overestimation after adding LOD/2 or LOQ/2 (see Figure S1).

The marked differences between ER-CALUX and MELN are not unexpected. MELN has the highest relative E1 effect potency of all tested bioassays (0.29 compared to 0.01 for ER-CALUX; Table S5). Thus, EEQ_{chem} results for MELN are strongly based on E1 concentrations – a compound that was always measured (except for a few samples by Lab 2, Figure 3). Consequently, for MELN the relative contribution of E2 and EE2 at LOD/2 or LOQ/2 on top of measured E1 concentrations is relatively small though still noticeable for samples with low EEQ concentrations (compare Figure 6C to 6D).



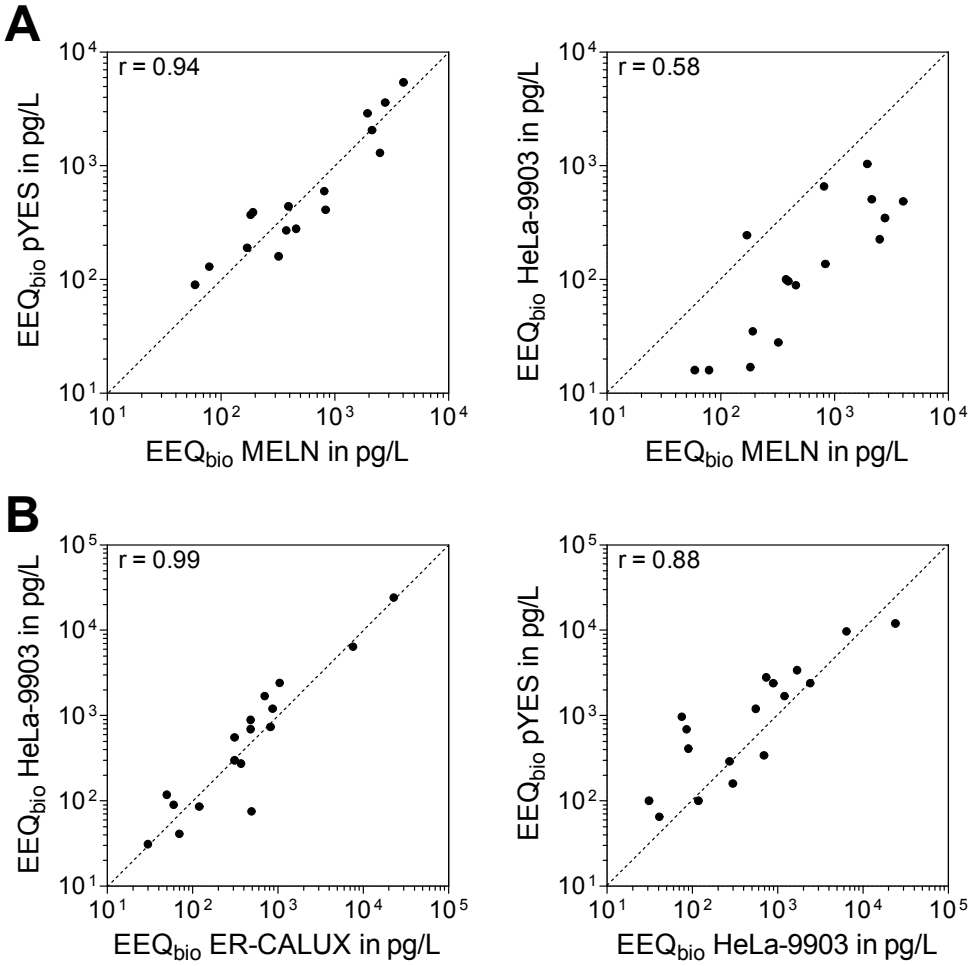
381

382 **Figure 6: Comparison of EEQ_{chem} with EEQ_{bio}.** Exemplary graphs are shown for the ER-CALUX (A, B) and MELN assay
383 (C, D) (further figures in the SI, Part G). Graphs on the left show the EEQ_{chem} derived from values >LOQ, while the graphs on
384 the right show the $EEQ_{chem + LOD/2 \text{ or } LOQ/2}$ calculated by including LOD/2 or LOQ/2. The dashed line indicates perfect agreement
385 of EEQ_{chem} with EEQ_{bio}.

386 ## 3.6 Comparison of effect-based methods

387 To compare the five effect-based methods amongst each other, a correlation analysis was conducted by
388 plotting the EEQs of one method against the EEQs of all other methods for SW samples and WW
389 samples, respectively (Figure 7).

390



391

392 **Figure 7: Exemplary graphs of correlation analysis of effect-based methods for SW (A) and WW (B) showing the**
393 **strongest and weakest correlations.** The correlation analysis was based on the method described in section 2.9. The dashed
394 line indicates perfect agreement of the compared effect-based methods. All correlations were significant with a p value <0.0001
395 except for MELN and HeLa-9903 (top right panel) which had a p value ≈ 0.01. Further graphs are shown in SI, Part H, Figures
396 S2 and S3.

397 The results of this analysis are summarised in Table 1 and Table 2 and show a strong correlation and
398 thus good comparability of pYES, MELN and ER-CALUX. For SW samples, the strongest correlations
399 were seen for pYES/MELN ($r^{\circ}= 0.94$) and pYES/ER-GeneBLAzer ($r^{\circ}= 0.94$), while the weakest
400 correlation was determined for MELN/HeLa-9903 ($r^{\circ}= 0.58$). For WW samples, test results correlated
401 strongly among all methods (Table), and the strongest correlation ($r^{\circ}= 0.99$) was observed for ER-
402 CALUX/HeLa-9903. It is known that effect-based methods differ in their REPs for individual ER-
403 agonists [53-55] which can explain that results obtained by the HeLa-9903 assay correlated less strongly
404 with other test results . Based on these differences effect-based methods can be split into two groups:
405 pYES and MELN with high E1 REP and ER-CALUX, HeLa-9903 and ER-GeneBLAzer with lower E1
406 REP.

Table 1: Pearson correlation coefficients of all bioassays for SW. The values were calculated according to the method mentioned in section 2.9. All correlations were significant with a p value <0.0001 (***) and a p value ≈ 0.01 (*).

	MELN	ER-GeneBLAzer	HeLa-9903	pYES
ER-CALUX	0.81 ***	0.91 ***	0.86 ***	0.76 ***
MELN		0.93 ***	0.58 *	0.94 ***
ER-GeneBLAzer			0.77 ***	0.94 ***
HeLa-9903				0.61 *

Table 2: Pearson correlation coefficients of all bioassays for WW. The values were calculated according to the method mentioned in section 2.9. All correlations were significant with a p value <0.0001 (***).

	MELN	ER-GeneBLAzer	HeLA-9903	pYES
ER-CALUX	0.94 ***	0.98 ***	0.99 ***	0.89 ***
MELN		0.98 ***	0.94 ***	0.97 ***
ER-GeneBLAzer			0.97 ***	0.96 ***
HeLa-9903				0.88 ***

4 Conclusions and trends

By including E1, E2, and EE2 in the watch-list of the WFD, the European Commission recognised the need to assess environmental occurrence and impact of these endocrine disrupting substances. However, the current WFD monitoring approach, which is based on chemical analytical measurements and compliance with specific EQSs, has been shown to be limited with regard to the ability to detect these substances at required concentrations [18, 51]. As demonstrated in this study, chemical analytical methods (status 2015) were unable to quantify the steroidal estrogens E2 and EE2 at EQS concentrations in all samples although E1 was measured effectively. Using effect-based methods, EEQ concentrations could be determined in all samples. As these EEQ concentrations are the responses to mixtures of known as well as unknown substances, effect-based methods have the potential to be highly valuable tools complementing routine monitoring and water quality assessment for estrogenic compounds. Effect-based methods are of particular regulatory interest as tools to screen and prioritise samples for further analysis by chemical analytical methods. Furthermore, DIN/EN/ISO standards to determine the estrogenic potential of water samples – covering human cell lines (e.g. ER-CALUX) and yeast based assays – will be available in early 2018 under ISO/DIS19040. The availability of such standards will facilitate the integration of effect-based methods into regulatory schemes.

Our study showed that EEQ results obtained from all effect-based methods applied were comparable – especially at higher concentrations found in WW – but results can vary between methods based on the

relative effect potencies for individual substances. This has to be considered for the interpretation of data and determination of threshold values. As stated above: 1) *in vitro* effect-based methods cannot deliver single substance based measurements, but are suitable to assess overall estrogenicity in water samples and 2) results of these methods need to be confirmed by advanced chemical analysis. Along these lines, the inclusion of effect-based methods into monitoring programmes as a screening tool (detailed description in Kase et al., [52]) for estrogenic substances in surface water bodies would be a valuable complement to chemical analysis currently foreseen by the Directive 2013/39/EU and WFD [28, 56, 57].

438 **Acknowledgements**

439 The major part of the project was funded by in-kind contribution from numerous project partners showing high
440 interest in this project. Essential parts of the project, e.g. extraction and reporting were externally co-funded by the
441 Swiss Centre of Applied Ecotoxicology EAWAG-EPFL, NORMAN Network, INERIS and Pharmaceutical
442 Associations (Pfizer, Johnson & Johnson, TEVA which joined the project later and only supported the extraction at
443 central Lab BDS). In total, 70 colleagues associated to 24 institutes and agencies in 12 nations contributed to this
444 project. The authors would like to thank: Joint Research Centre (EC), The French Agency for Biodiversity
445 (previously ONEMA) (FR), INERIS (FR), Bio Detection Systems (NL), Swiss Centre for Applied Ecotoxicology
446 (CH), Federal Institute of Hydrology (DE), Federal Environment Agency (DE), RWTH Aachen (DE), RECETOX
447 (CZ), NORMAN-Network, Helmholtz Centre for Environmental Research-UFZ (DE), IRSA-CNR (IT), Italian
448 Institute of Health (IT), University of Leon (ES), Water Research Institute T.G.Masaryk (CZ), Bavarian State
449 Office for Environment (DE), LANUV (DE), Environment Agency Austria (AT), ISSeP (Scientific Institute of
450 Public Service) Wallonia (BE), SMAT (IT), BrianzAcque (IT), Vettabbia scarl, MilanoDepur (IT), Agence de
451 l'Eau Adour-Garonne (FR), Ontario Ministry of the Environment and Climate Change (CAN), McGill University
452 (CAN), Environmental Institute (SK) and DG Environment of the European Commission (EC). The project
453 partners from UFZ, RWTH Aachen, INERIS and RECETOX were supported by the SOLUTIONS project, that is
454 funded by the European Union Seventh Framework Programme (FP7-ENV-2013-two-stage Collaborative project)
455 under grant agreement number 603437. Additionally, we would like to thank Christin Kuehnert (UFZ), Maria
456 König (UFZ) and Michael Gundlach (RWTH) for experimental assistance. Special thanks for stewardship of this
457 study we would like to address to Helen Clayton and Stéphanie Schaan from European Commission, Environment
458 DG.

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

461

462 **Conflict of interests**

463 The Federal Institute of Hydrology did not receive any kind of financial support from the Pharmaceutical
464 Associations. Other authors declare no conflict of interest.

465

Bibliography

1. John P Sumpter and Andrew C Johnson, *10th anniversary perspective: Reflections on endocrine disruption in the aquatic environment: From known knowns to unknown unknowns (and many things in between).* Journal of Environmental Monitoring, 2008. **10**(12): p. 1476-1485.
2. Georg Streck, *Chemical and biological analysis of estrogenic, progestagenic and androgenic steroids in the environment.* TrAC Trends in Analytical Chemistry, 2009. **28**(6): p. 635-652.
3. Daniel J Caldwell, Frank Mastrocco, Paul D Anderson, Reinhard Länge, and John P Sumpter, *Predicted-no-effect concentrations for the steroid estrogens estrone, 17β-estradiol, estriol, and 17α-ethinylestradiol.* Environmental Toxicology and Chemistry, 2012. **31**(6): p. 1396-1406.
4. Helmut Segner, Ayako Casanova-Nakayama, Robert Kase, and Charles R Tyler, *Impact of environmental estrogens on yfish considering the diversity of estrogen signaling.* General and comparative endocrinology, 2013. **191**: p. 190-201.
5. Jinmiao Zha, Liwei Sun, Yiqi Zhou, Philip A Spear, Mei Ma, and Zijian Wang, *Assessment of 17α-ethinylestradiol effects and underlying mechanisms in a continuous, multigeneration exposure of the chinese rare minnow (gobiocypris rarus).* Toxicology and applied pharmacology, 2008. **226**(3): p. 298-308.
6. Karen A Kidd, Paul J Blanchfield, Kenneth H Mills, Vince P Palace, Robert E Evans, James M Lazorchak, and Robert W Flick, *Collapse of a fish population after exposure to a synthetic estrogen*, in *Proceedings of the National Academy of Sciences*. 2007. p. 8897-8901.
7. Stefanie Grund, Eric Higley, René Schönenberger, Marc JF Suter, John P Giesy, Thomas Braunbeck, Markus Hecker, and Henner Hollert, *The endocrine disrupting potential of sediments from the upper danube river (germany) as revealed by in vitro bioassays and chemical analysis.* Environmental Science and Pollution Research, 2011. **18**(3): p. 446-460.
8. Barbara V Rutishauser, Maija Pesonen, Beate I Escher, Gabriele E Ackermann, Hans-Rudolf Aerni, Marc J-F Suter, and Rik IL Eggen, *Comparative analysis of estrogenic activity in sewage treatment plant effluents involving three in vitro assays and chemical analysis of steroids.* Environmental Toxicology and Chemistry, 2004. **23**(4): p. 857-864.
9. Andrew C Johnson, Egon Dumont, Richard J Williams, Rik Oldenkamp, Iwona Cisowska, and John P Sumpter, *Do concentrations of ethinylestradiol, estradiol, and diclofenac in european rivers exceed proposed eu environmental quality standards?* Environmental science & technology, 2013. **47**(21): p. 12297-12304.
10. Adam R Schwindt, Dana L Winkelman, Kristen Keteles, Mark Murphy, and Alan M Vajda, *An environmental oestrogen disrupts fish population dynamics through direct and transgenerational effects on survival and fecundity.* Journal of applied ecology, 2014. **51**(3): p. 582-591.
11. Eric Higley, Stefanie Grund, Paul D Jones, Tobias Schulze, Thomas-B Seiler, Urte Lübcke-von Varel, Werner Brack, Jan Wölz, Hanno Zielke, and John P Giesy, *Endocrine disrupting, mutagenic, and teratogenic effects of upper danube river sediments using effect-directed analysis.* Environmental Toxicology and Chemistry, 2012. **31**(5): p. 1053-1062.
12. CEJR Desbrow, EJ Routledge, GC Brighty, JP Sumpter, and M Waldock, *Identification of estrogenic chemicals in stw effluent. 1. Chemical fractionation and in vitro biological screening.* Environmental science & technology, 1998. **32**(11): p. 1549-1558.
13. Patricia Burkhardt-Holm, Walter Giger, Herbert GUttinger, Ueli Ochsenbein, Armin Peter, Karin Scheurer, Helmut Segner, Erich Staub, and Marc J-F Suter, *Where have all the fish gone?* Environmental science & technology, 2005. **39**(21): p. 441A-447A.
14. Frederic DL Leusch, Christiaan De Jager, Yves Levi, Richard Lim, Leo Puijker, Frank Sacher, Louis A Tremblay, Vickie S Wilson, and Heather F Chapman, *Comparison of five in vitro bioassays to measure estrogenic activity in environmental waters.* Environmental science & technology, 2010. **44**(10): p. 3853-3860.
15. Jon P Nash, David E Kime, Leo TM Van der Ven, Piet W Wester, François Brion, Gerd Maack, Petra Stahlschmidt-Allner, and Charles R Tyler, *Long-term exposure to environmental concentrations of the pharmaceutical ethynylestradiol causes reproductive failure in fish.* Environmental health perspectives, 2004: p. 1725-1733.
16. Christiane Heiss. *Recommendation for a monitoring strategy for estrogens in coastal and continental surface waters*. 2013 [cited 2018 February 5th]; Available from: http://www.bafg.de/DE/05_Wissen/02_Veranst/2013/2013_02_27_votum_en.pdf?__blob=publicationFile.
17. European Union, *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 substances in the field of water policy.* Official Journal of the European Union, 2013.
18. EU Decision, *495/2015, 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.* Off. J. Eur. Union L, 2015. **78**: p. 40-42.
19. Mario Carere, Stefano Polesello, Robert Kase, and Bernd Manfred Gawlik, *The emerging contaminants in the context of the eu water framework directive*, in *Emerging contaminants in river ecosystems*. 2015, Springer. p. 197-215.

20. EU WFD, *Directive 2000/60/ec of the european parliament and of the council establishing a framework for the community action in the field of water policy.* The European Parliament and the Council of the European Union: Brussels, Belgium, 2000.
21. RN Carvalho, L Ceriani, and A Ippolito, *Development of the first watch list under the environmental quality standards directive water policy*. 2015.
22. AS Wernersson, C Maggi, and M Carere, *Technical report on aquatic effect-based monitoring tools.* Office for Official Publications of the European Communities, 2014.
23. Petra Y Kunz, Cornelia Kienle, Mario Carere, Nadzeya Homazava, and Robert Kase, *In vitro bioassays to screen for endocrine active pharmaceuticals in surface and waste waters.* Journal of pharmaceutical and biomedical analysis, 2015. **106**: p. 107-115.
24. Robert Loos. *Analytical methods for possible wfd 1st watch list substances*. 2015 [cited 2018 February 5th]; Report EUR 27046 EN:[Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.829.1423&rep=rep1&type=pdf.
25. Mario Carere, Stefano Polesello, Caterina Sollazzo, and Bernd Manfred Gawlik, *Chemical monitoring and emerging pollutants in the common implementation strategy of the water framework directive.* TrAC Trends in Analytical Chemistry, 2012. **36**: p. 12-14.
26. Markus Hecker and Henner Hollert, *Endocrine disruptor screening: Regulatory perspectives and needs.* Environmental Sciences Europe, 2011. **23**(1): p. 1.
27. AD Vethaak and AC Belfroid, *Estrogens and xeno-estrogens in the aquatic environment of the netherlands: Occurence, potency and biological effects*. 2002: Rijkswaterstaat, RIKZ.
28. Ann-Sofie Wernersson, Mario Carere, Chiara Maggi, Petr Tusil, Premysl Soldan, Alice James, Wilfried Sanchez, Valeria Dulio, Katja Broeg, and Georg Reifferscheid, *The european technical report on aquatic effect-based monitoring tools under the water framework directive.* Environmental Sciences Europe, 2015. **27**(1): p. 1-11.
29. Karin Kinnberg, *Evaluation of in vitro assays for determination of estrogenic activity in the environment*. 2003, Danish Ministry of the Environment, Danish Environmental Protection Agency.
30. Diane M Klotz, Barbara S Beckman, Steven M Hill, John A McLachlan, Marian R Walters, and Steven F Arnold, *Identification of environmental chemicals with estrogenic activity using a combination of in vitro assays.* Environmental health perspectives, 1996. **104**(10): p. 1084.
31. Beate I Escher, Nadine Bramaz, Jochen F Mueller, Pamela Quayle, Sibylle Rutishauser, and Etiënne LM Vermeirssen, *Toxic equivalent concentrations (teqs) for baseline toxicity and specific modes of action as a tool to improve interpretation of ecotoxicity testing of environmental samples.* Journal of Environmental Monitoring, 2008. **10**(5): p. 612-621.
32. Charles W Walker and John E Watson, *Adsorption of estrogens on laboratory materials and filters during sample preparation all rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher.* Journal of environmental quality, 2010. **39**(2): p. 744-748.
33. Patrick Balaguer, Fabienne François, Franck Comunale, Hélène Fenet, Anne-Marie Boussioux, Michel Pons, Jean-Claude Nicolas, and Claude Casellas, *Reporter cell lines to study the estrogenic effects of xenoestrogens.* Science of the Total Environment, 1999. **233**(1): p. 47-56.
34. Beate I Escher, Mayumi Allinson, Rolf Altenburger, Peter A Bain, Patrick Balaguer, Wibke Busch, Jordan Crago, Nancy D Denslow, Elke Dopp, and Klara Hilscherova, *Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro bioassays.* Environmental science & technology, 2013. **48**(3): p. 1940-1956.
35. OECD, *Test no. 455: Performance-based test guideline for stably transfected transactivation in vitro assays to detect estrogen receptor agonists*. OECD Publishing.
36. Denise Spira, Georg Reifferscheid, and Sebastian Buchinger, *Combination of high-performance thin-layer chromatography with a specific bioassay-a tool for effect-directed analysis.* JPC-Journal of Planar Chromatography-Modern TLC, 2013. **26**(5): p. 395-401.
37. M. B. Müller, C. Dausend, C. Weins, and F. H. Frimmel, *A new bioautographic screening method for the detection of estrogenic compounds.* Chromatographia, 2004. **60**(3): p. 207-211.
38. Sebastian Buchinger, Denise Spira, Kathrin Bröder, Michael Schlüsener, Thomas Ternes, and Georg Reifferscheid, *Direct coupling of thin-layer chromatography with a bioassay for the detection of estrogenic compounds: Applications for effect-directed analysis.* Analytical chemistry, 2013. **85**(15): p. 7248-7256.
39. Ines Klingelhöfer and Gertrud E Morlock, *Bioprofiling of surface/wastewater and bioquantitation of discovered endocrine-active compounds by streamlined direct bioautography.* Analytical chemistry, 2015. **87**(21): p. 11098-11104.
40. Andreas Schönborn and Andrea Grimmer, *Coupling sample preparation with effect-directed analysis of estrogenic activity-proposal for a new rapid screening concept for water samples.* JPC-Journal of Planar Chromatography-Modern TLC, 2013. **26**(5): p. 402-408.

41. Maria König, Beate I Escher, Peta A Neale, Martin Krauss, Klára Hilscherová, Jiří Novák, Ivana Teodorović, Tobias Schulze, Sven Seidensticker, and Muhammad Arslan Kamal Hashmi, *Impact of untreated wastewater on a major european river evaluated with a combination of in vitro bioassays and chemical analysis.* Environmental Pollution, 2017. **220**: p. 1220-1230.
42. Lothar Sachs, *Angewandte statistik: Anwendung statistischer methoden*. 2013: Springer-Verlag.
43. Ze-hua Liu, Yoshinori Kanjo, and Satoshi Mizutani, *Removal mechanisms for endocrine disrupting compounds (edcs) in wastewater treatment—physical means, biodegradation, and chemical advanced oxidation: A review.* Science of the Total Environment, 2009. **407**(2): p. 731-748.
44. Barbora Jarošová, Luděk Bláha, John P Giesy, and Klára Hilscherová, *What level of estrogenic activity determined by in vitro assays in municipal waste waters can be considered as safe?* Environment international, 2014. **64**: p. 98-109.
45. R Loos, *Analytical methods relevant to the european commission's 2012 proposal on priority substances under the water framework directive.* JRC scientific and policy report, 2012.
46. Tom Benijts, Riet Dams, Willy Lambert, and André De Leenheer, *Countering matrix effects in environmental liquid chromatography–electrospray ionization tandem mass spectrometry water analysis for endocrine disrupting chemicals.* Journal of chromatography A, 2004. **1029**(1): p. 153-159.
47. YKK Koh, TY Chiu, A Boobis, E Cartmell, JN Lester, and MD Scrimshaw, *Determination of steroid estrogens in wastewater by high performance liquid chromatography–tandem mass spectrometry.* Journal of chromatography A, 2007. **1173**(1): p. 81-87.
48. EC Directive, *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.* Official Journal of the European Union L, 2009. **201**: p. 36.
49. Sándor Görög, *Advances in the analysis of steroid hormone drugs in pharmaceuticals and environmental samples (2004–2010).* Journal of pharmaceutical and biomedical analysis, 2011. **55**(4): p. 728-743.
50. Helena Tomšíková, Jana Aufartová, Petr Solich, Lucie Nováková, Zoraida Sosa-Ferrera, and José Juan Santana-Rodríguez, *High-sensitivity analysis of female-steroid hormones in environmental samples.* TrAC Trends in Analytical Chemistry, 2012. **34**: p. 35-58.
51. Carolina Di Paolo, Richard Ottermanns, Steffen Keiter, Selim Ait-Aissa, Kerstin Bluhm, Werner Brack, Magnus Breitholtz, Sebastian Buchinger, Mario Carere, and Carole Chalon, *Bioassay battery interlaboratory investigation of emerging contaminants in spiked water extracts–towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring.* Water research, 2016. **104**: p. 473-484.
52. Robert Kase, Barbora Javurkova, Eszter Simon, Kees Swart, Sebastian Buchinger, Sarah Könemann, Beate Escher, Mario Carere, Valeria Dulio, Selim Ait-Aissa, Henner Hollert, Sara Valsecchi, Stefano Polesello, Peter Behnisch, Carolina Di Paolo, Daniel Olbrich, Eliska Sychrova, Michael Gundlach, Rita Schlichting, Lomig Leborgne, Manfred Clara, Christoph Scheffknecht, Yves Marneffe, Carole Chalon, Petr Tusil, Premysl Soldan, Brigitte von Danwitz, Julia Schwaiger, Antonio Moran, Francesca Bersani, Olivier Perceval, Cornelia Kienle, Etienne Vermeirssen, Klara Hilscherova, Georg Reifferscheid, and Inge Werner, *Screening and risk management solutions for steroidal oestrogens in surface and wastewater.* Trends in Analytical Chemistry, 2018. Under review.
53. Juliette Legler, Martine Dennekamp, A Dick Vethaak, Abraham Brouwer, Jan H Koeman, Bart van der Burg, and Albertinka J Murk, *Detection of estrogenic activity in sediment-associated compounds using in vitro reporter gene assays.* Science of the Total Environment, 2002. **293**(1): p. 69-83.
54. Albertinka J Murk, Juliette Legler, Marola MH Van Lipzig, John HN Meerman, Angelique C Belfroid, Albertus Spenkelink, Bart Van Der Burg, Gerard BJ Rijs, and Dick Vethaak, *Detection of estrogenic potency in wastewater and surface water with three in vitro bioassays.* Environmental Toxicology and Chemistry, 2002. **21**(1): p. 16-23.
55. Frederic DL Leusch. *Tools to detect estrogenic activity in environmental waters*. 2008 [cited 2018 February 5th]; Available from: http://www.waterrf.org/resources/Lists/SpecialReports/Attachments/2/GWRC_EDC_ToolsToDetect.pdf.
56. Werner Brack, Valeria Dulio, Marlene Ågerstrand, Ian Allan, Rolf Altenburger, Markus Brinkmann, Dirk Bunke, Robert M Burgess, Ian Cousins, and Beate I Escher, *Towards the review of the european union water framework directive: Recommendations for more efficient assessment and management of chemical contamination in european surface water resources.* Science of the Total Environment, 2017. **576**: p. 720-737.
57. Ian J Allan, Branislav Vrana, Richard Greenwood, Graham A Mills, Benoit Roig, and Catherine Gonzalez, *A "toolbox" for biological and chemical monitoring requirements for the european union's water framework directive.* Talanta, 2006. **69**(2): p. 302-322.

Supplementary Information

Effect-based and chemical analytical methods to monitor estrogens under the European Water Framework Directive

Sarah Könemann, Robert Kase, Eszter Simon, Kees Swart, Sebastian Buchinger, Michael Schlüsener, Henner Hollert, Beate I. Escher, Inge Werner, Selim Aït-Aïssa, Etienne Vermeirssen, Valeria Dulio, Sara Valsecchi, Stefano Polesello, Peter Behnisch, Barbora Javurkova, Olivier Perceval, Carolina Di Paolo, Daniel Olbrich, Eliska Sychrova, Rita Schlichting, Lomig Leborgne, Manfred Clara, Christoph Scheffknecht, Yves Marneffe, Carole Chalon, Petr Tušil, Premysl Soldán, Brigitte von Danwitz, Julia Schwaiger, Maria Isabel San Martín Becares, Francesca Bersani, Klara Hilscherova, Georg Reifferscheid, Thomas Ternes, Mario Carere

Table S1: Author contributions calculated according to Clement Prabhakar, T. [1].

	Ideas	Work			Writing	Stewardship	Ideas	Work			Writing	Stewardship	Sum	Total contribution
		Sampling	Labwork	Data eval.				Sampling	Labwork	Data eval.				
weighting	0.1	0.2	0.2	0.2	0.3	0.2	0.1	0.2	0.2	0.2	0.3	0.2	**1.0**	
Sarah Könemann				15.0	30.0		0.0	0.0	0.0	2.3	9.0	0.0	11.3	**11.3**
Robert Kase	44.4			15.0	13.0	10.5	2.2	0.0	0.0	2.3	3.9	2.1	10.5	**10.5**
Mario Carere	44.4				13.0	10.5	2.2	0.0	0.0	0.0	3.9	2.1	8.2	**8.2**
Eszter Simon				25.0	10.0		0.0	0.0	0.0	3.8	3.0	0.0	6.8	**6.8**
Kees Swart			39.1				0.0	0.0	5.9	0.0	0.0	0.0	5.9	**5.9**
Sebastian Buchinger				20.0	5.0	5.3	0.0	0.0	0.0	3.0	1.5	1.1	5.6	**5.6**
Michael Schlüsener			17.4	10.0			0.0	0.0	2.6	1.5	0.0	0.0	4.1	**4.1**
Henner Hollert					5.0	10.5	0.0	0.0	0.0	0.0	1.5	2.1	3.6	**3.6**
Beate Escher					8.0	5.6	0.0	0.0	0.0	0.0	2.4	1.1	3.5	**3.5**
Inge Werner						15.3	0.0	0.0	0.0	0.0	0.0	3.1	3.1	**3.1**
Selim Ait-Aissa			8.7	10.0			0.0	0.0	1.3	1.5	0.0	0.0	2.8	**2.8**
Etienne Vermeirssen					5.0	5.3	0.0	0.0	0.0	0.0	1.5	1.1	2.6	**2.6**
Valeria Dulio						10.5	0.0	0.0	0.0	0.0	0.0	2.1	2.1	**2.1**
Sara Valsecchi		8.3			2.5		0.0	1.2	0.0	0.0	0.8	0.0	2.0	**2.0**
Stefano Polesello		8.3			2.5		0.0	1.2	0.0	0.0	0.8	0.0	2.0	**2.0**
Peter Behnisch				5.0		5.3	0.0	0.0	0.0	0.8	0.0	1.1	1.8	**1.8**
Olivier Perceval	11.1					5.3	0.6	0.0	0.0	0.0	0.0	1.1	1.6	**1.6**
Daniel Olbrich			8.7				0.0	0.0	1.3	0.0	0.0	0.0	1.3	**1.3**
Eliska Sychrova			8.7				0.0	0.0	1.3	0.0	0.0	0.0	1.3	**1.3**
Rita Schlichting			8.7				0.0	0.0	1.3	0.0	0.0	0.0	1.3	**1.3**
Isabel San Martin Becares			8.7				0.0	0.0	1.3	0.0	0.0	0.0	1.3	**1.3**
Lomig Leborgne		8.3					0.0	1.2	0.0	0.0	0.0	0.0	1.2	**1.2**
Manfred Clara		8.3					0.0	1.2	0.0	0.0	0.0	0.0	1.2	**1.2**
Christoph Scheffknecht		8.3					0.0	1.2	0.0	0.0	0.0	0.0	1.2	**1.2**
Yves Marneffe		8.3					0.0	1.2	0.0	0.0	0.0	0.0	1.2	**1.2**
Carole Chalon		8.3					0.0	1.2	0.0	0.0	0.0	0.0	1.2	**1.2**
Petr Tusil		8.3					0.0	1.2	0.0	0.0	0.0	0.0	1.2	**1.2**
Premysl Soldan		8.3					0.0	1.2	0.0	0.0	0.0	0.0	1.2	**1.2**
Brigitte von Danwitz		8.3					0.0	1.2	0.0	0.0	0.0	0.0	1.2	**1.2**
Julia Schwaiger		8.3					0.0	1.2	0.0	0.0	0.0	0.0	1.2	**1.2**
Francesca Bersani		8.3					0.0	1.2	0.0	0.0	0.0	0.0	1.2	**1.2**
Klara Hilscherova						5.3	0.0	0.0	0.0	0.0	0.0	1.1	1.1	**1.1**
Thomas Ternes						5.3	0.0	0.0	0.0	0.0	0.0	1.1	1.1	**1.1**
Georg Reifferscheid						5.3	0.0	0.0	0.0	0.0	0.0	1.1	1.1	**1.1**
Barbora Javurkova					3.0		0.0	0.0	0.0	0.0	0.9	0.0	0.9	**0.9**
Carolina di Paolo					3.0		0.0	0.0	0.0	0.0	0.9	0.0	0.9	**0.9**
Sum	99.9	99.6	100.0	100.0	100.0	100.0	5.0	14.9	15.0	15.0	30.0	20.0	99.9	100.0

Calculation of the minimal total contribution (MTC): $MTC = \frac{0.3 \cdot 100}{n}$ with n = 36

Part A: Sampling and sample preparation

Given that chemicals from a broad range of compound classes can interact with the estrogen receptor – both in an agonistic and in an antagonistic way – there is a wide scope of possible sources of contamination of the samples. To allow for a robust analysis of the samples, any contamination needs to be avoided.

Samples are taken by means of a scoop. The sample container on the scoop was cleaned before use by rinsing its in and outside **three times** with acetone. Rinsing is performed by spraying acetone from a PTFE wash bottle onto the scoop. Subsequently, the container was rinsed using ambient water (i.e. river water or effluent that will be sampled) by dipping it in ambient water and filling and emptying the container **five times**. After this preparation, bottles were filled one after the other. One person – **using clean nitrile gloves** – removed the lid from the first bottle and places this lid on a clean surface. The other person used the scoop to collect a sample and poured the sample into the bottle. Sample scooping and addition of sample to the bottle was repeated until the bottle was 2/3 full (i.e. 3.75 L of sample). Next, the lid was removed from bottle 2 and used to seal bottle 1; now bottle 2 was be filled. This process was continued until all bottles were filled. The last bottle in the sampling series received the lid that came off the first bottle. In order to prevent degeneration of substances, 3.75 mL of 3 M H_2SO_4 were added to adjust the pH to approximately 3. As soon as possible, the samples were frozen and sent to the central lab on dry ice, where they were stored at -20 °C until extraction.

The further preparation of samples generally involved the extraction of samples according to the SPE procedure to a 1000-fold concentration. For the solid-phase extraction (SPE) 11 L of each sample were thawed overnight at room temperature. Then the samples were homogenised and divided into 1 L aliquots. By addition of 3 M H_2SO_4 the pH was adjusted to 3. Approximately 40 % of the surface water samples (A(11), B(6), D(22), E(27), I(8), M(28), O(3)) and all waste water samples were filtered over glass fiber filters (Millipore, type 4, retention 2.7 µm, circle size 4.7 cm) before extraction. A 1 L procedure blank with ultrapure water adjusted to pH 3 was processed in parallel. In case a sample needed to be filtered, the procedure blank was filtered as well. Prior to the extraction, the cartridges (Phenomenex Strata C_{18}-E, 55 µm, 70 Å, 500 mg/6 mL) were conditioned with 6 mL of n-hexane, 2 mL of acetone, 6 mL of methanol and finally 10 mL of ultrapure water adjusted to pH 3. Subsequently, each cartridge was loaded with 1 L of sample using a continuous flow setup with 5 mL/min. To determine the exact volume of the sample that was extracted, the glass bottles were weighed before and after the extraction process. After 30 min of drying with full vacuum, the cartridges were eluted with 4 mL acetone. Pertinent extracts were combined and the collection tubes were rinsed with 1 mL acetone. Then the extracts were evaporated to a final volume of 200 µL acetone. Procedural blank extracts were subjected to the same protocol.

Subsequently, all samples underwent a silica clean-up. For the clean-up step, all extracts were reconstituted in 1 mL n-hexane / acetone (65:35, v/v) and applied to a silica column (CHROMABOND SiOH, Macherey-Nagel, 6 mL, bed size 1000 mg), which was previously activated (3 h, 85 °C) and conditioned with 10 mL n-hexane / acetone (65:35, v/v). For surface water extracts one column was used for one mL of sample, whereas for waste water extracts one column was used for all 11 mL of the sample. After sample application the column was rinsed with 2 mL of n-hexane / acetone (65:35, v/v) and eluted with 4 mL of the same mixture. The extracts were then dried with N_2 at 45 °C and reconstituted in 11 mL ethanol resulting in a 1000-fold concentration of the initial water sample (enrichment factor 1000).

The final 1 mL of all extracted samples including the blanks and the spiked water reference samples were then sent to all project partners to be analysed by LC MS/MS as well as effect-based techniques.

Part B: Data evaluation and EEQ calculation

Data collection

The bioassays of this study are based on the same principle (i.e. binding of a ligand to an estrogen receptor with a subsequent activation of a reporter gene), however they vary in the measured endpoint (i.e. luciferase activity by bioluminescence or the photometric measurements after the induction of the β-galactosidase enzyme) and in certain experimental parameters (e.g. cell types and density, % and type of organic solvent in the exposure medium, cell culture and exposure mediums, and extracts dilution applied in the assay).

Each participating laboratory received the same number of E1, E2 and EE2 stocks, ethanolic extracts of extraction blanks, positive controls and encoded surface and wastewaters (1 mL) sent by the extraction lab together with the list of weights of the extracts. The participating labs were asked to analyze all blanks, references, controls and sample extracts in dilution series following their own protocol and report any further handling the extracts underwent during the analysis (e.g. solvent exchange, dilution etc).

All raw data together with the relative enrichment factors (REFs - the multiplication of the enrichment factor of the sample extraction and the dilution factor of the extract in each of the bioassays) and the test concentrations (dilutions) of the samples were provided by the participating laboratories and evaluated centrally.

Data evaluation

Bioassay data were normalised based on the measured response for the negative control (i.e. negative control) in the respective bioassay (0%) and the highest induction measured for the reference substance (100%). The PC_{10} approach was used as EEQ derivation procedure for data evaluation, [2]
First the measured bioassay responses were fitted using the five-parameter logistic regressions (Equation 1). This model - unlike the four-parameter model - is suitable to handle asymmetric curves, that are often observed when analysing environmental samples [2].

$$y = \frac{a - d}{\left(1 + \left(\frac{x}{C}\right)^{b}\right)^{f}} + d \qquad (1)$$

y =	the calculated effect measure (e.g. corrected absorbance) at concentration x
x =	the compound concentration which activates the test system to effect measure y
a =	minimum response (mean value fixed to the measurement of the solvent blank; bottom curve point)
d =	maximum response (mean value of y with the maximal activation of the test; curve plateau)
C =	the curve point of inflection, in case of a symmetric concentration-response curve C equals the mean effect concentration at which the estrogenic effect reaches half of its maximum (EC50 – 50 % effect concentration)
b =	Slope refers to the steepness of the curve
f =	assymetry factor, which reflects the asymmetry of the curve. In case of a symmetric curve f is equal to one.

The REF of the sample at a certain effect level (e.g. 10 %-effect) of the reference compound E2 – was calculated by inserting the associated effect measure into the inverse of the function describing the concentration response relationship of the sample. The respective REF of the sample equals the PCx of the reference compound. The appropriate effect level of the reference compound was defined as 10 % for each bioassay. This effect level in the sample extracts was above the assays sample-dependent LOQs and we assumed that this level can be reached and quantified by low contaminated surface water samples as well. Further, the 10 % threshold is statistically different from the control. The estrogenic activity of the sample was determined by dividing the PC_{10} by the derived REF (REF_{10}) and expressed as pg EEQ/L water.

Calculation of 17β-estradiol equivalents (EEQs) for test samples

First, the effect measure y_{PCx} at the x%-effect of the reference compound was calculated by the following equation (Equation 2):

$$y^{PCx} = \left(\frac{d-a}{100} \cdot e\right) + a \quad (2)$$

y_{PCx} = the effect measure at the x%-effect level of the reference compound

e = the x%-effect level of the reference (e.g. 10 for the estimation of the 10%-effect level)

d = the mean value of y with the maximal activation of the test (curve plateau derived from the concentration response relationship of the reference by the curve fitting)

a = the mean value of y without estrogenic effects (bottom curve point derived from the concentration response relationship of the reference by the curve fitting)

Then, y_{PCx} was inserted into the inverse of the function describing the concentration-response relationship of the sample for calculating the REF of the sample that induce the bioassay to the same extend as the PC_x of the reference compound (Equation 3).

$$REF = \left(\left(\frac{a-d}{yPCx-d}\right)^{\left(\frac{1}{f}\right)} - 1\right)^{\left(\frac{1}{b}\right)} \cdot C \quad (3)$$

REF = the REF of the sample that induce the same effect as the PCx of the reference compound

y_{PCx} = the effect measure at the x%-effect level of the reference compound

d = the mean value of y with the maximal activation of the test (curve plateau derived from the concentration response relationship of the sample by the curve fitting)

a = the mean value of y without estrogenic effects (bottom curve point derived from the concentration response relationship of the sampleby the curve fitting)

C = the curve point of inflection derived from the concentration response relationship of the sample by the curve fitting

b = Slope refers to the steepness of the curve and is proportional to the slope of the function at C (derived from the concentration response relationship of the sample by the curve fitting)

f = assymetry factor, which reflects the asymmetry of the curveand derived from the concentration-response relationship of the sample by the curve fitting. In case of a symmetric curve f is equal to one.

The estrogenic potential of the sample as was estimated by the quotient of PCx of the reference compound, E2 and the calculated REF_s of the sample.

Part C: LOQ determination for bioassays and LC-MS/MS analyses

LOQs for bioassays

The sample-specific LOQs in the bioassays were determined based on the variability of the response of the negative controls tested along the sample(s) and the highest sample-specific REF (relative enrichment factor) tested. The average response of the negative control (SC) replicates plus three times the standard deviation (SD) was interpolated from the E2 dose-response curve. This resulting E2-equivalent concentration (ng EEQ/L) was then divided by the highest tested REF of the sample to obtain a sample-specific LOQ. The figure below illustrates the procedure using an example from ER-CALUX.

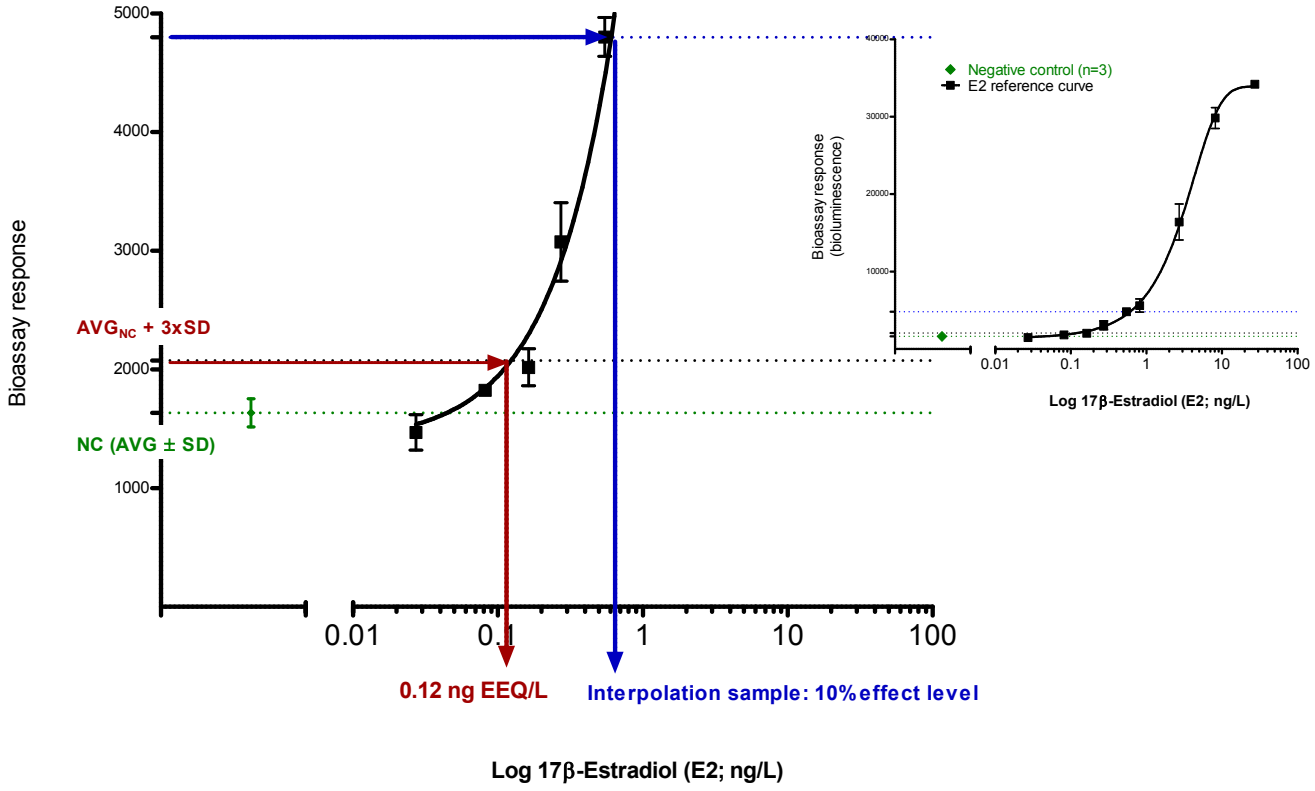


Figure S1. Graphical representation of the sample-specific LOQ determination in the ERα-CALUX bioassay. The small figure shows the full E2 dose-response curve, the large figure shows the lower part of the curve for better clarity. The response measured for the negative control was 1634 ± 118 RLU (relative light units; bioluminescence). Based on this measurement the LOQ threshold was determined: $AVG_{negative\ control}$ + 3xSD = 2075 RLU. This value was then interpolated from the E2 curve and an EEQ of 0.12 ng/L was obtained. Dividing this EEQ by the highest REF of the sample (in this particular case: 23.3) leads to a sample-specific LOQ of 0.005 ng EEQ/L (5.3 pg EEQ/L).

It is important to note two aspects: 1) sample EEQ concentrations were derived using the 10 % effect level of the E2 curve (PC_{10}); 2) sample curve fitting was only performed when "at least two effect measurements above 10 % were obtained" (see Section E of the SI).

Considering the first point: a "PC_{10} reporting level LOQ" for the above sample can be interpolated from the E2 curve along the blue arrows. The 10 % effect level equates to 4720 RLU and intercepts the E2 curve at 0.70 ng/L. For the sample above, with an REF of 23.3, the sample-specific PC_{10} reporting level LOQ is 29 pg/L.

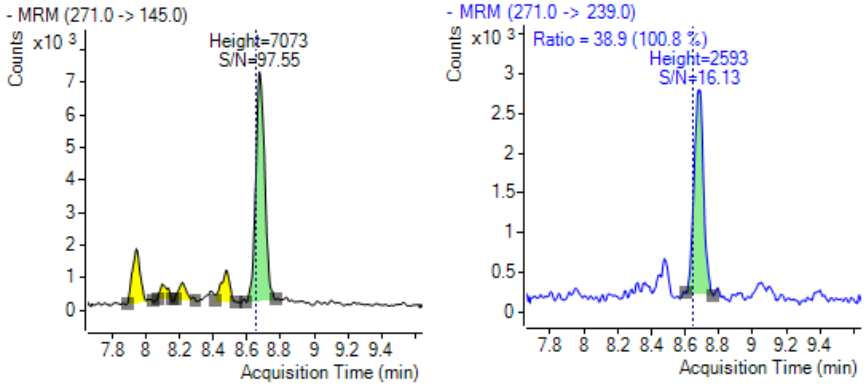
Considering the second point: for a robust fitting of the PC_{10}, a higher sample EEQ concentration than 29 pg/L is required and the effective LOQ is thus higher than the PC_{10} derived LOQ.

LOQs for LC-MS/MS

The figure below illustrates the procedure followed by Lab 2 for two examples. In the first example, the analyte was detected above a S/N of 10, in the second example the analyte was not detected and standard addition was used to determine the LOQ. Both the quantifier and the qualifier should have a S/N ≥ 10.

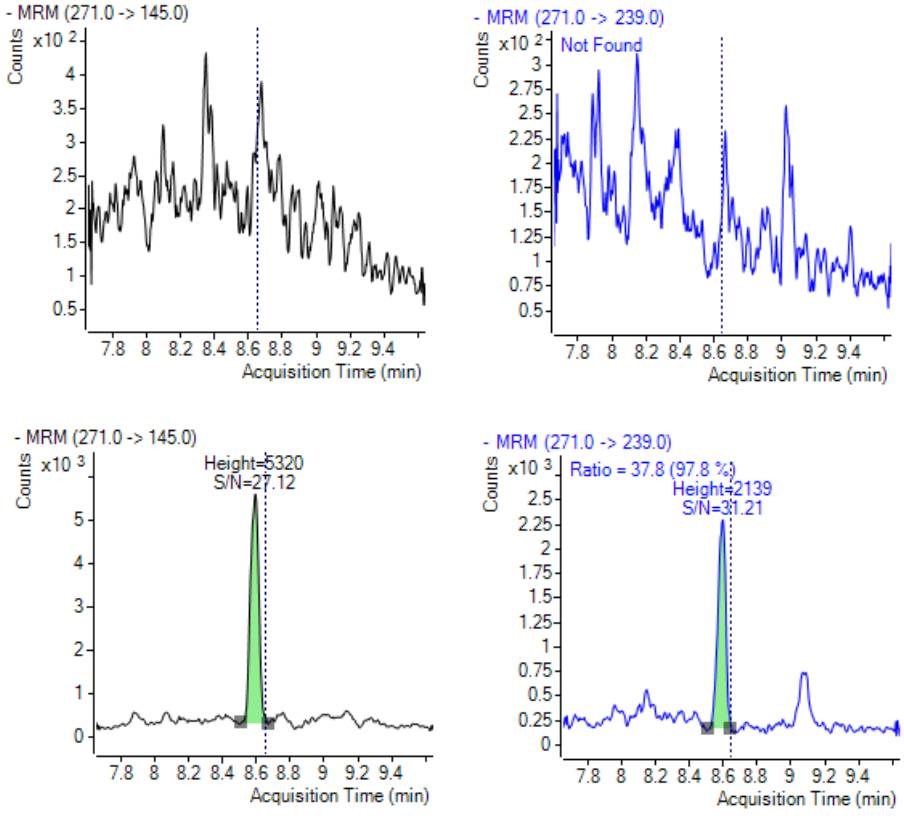
Example 1

The first panel shows the E2 peak in a wastewater sample. With a concentration of 0.55 ng/L and an S/N of 97 (graph on the left; quantifier) and S/N 16 (graph on the right; qualifier) the data were normalized to a S/N of 10 as follows: 0.55 ng/L /(16/10)= 0.34 ng/L. This theoretical concentration should have a S/N of 10:1 and is defined as LOQ for this analyte in this sample.



Example 2

The top panel shows the absence of a peak (quantifier on the left and qualifier o the right). The bottom panel shows the same sample with the addition of 0.5 ng/L with a S/N of 27 and S/N 31. After normalization to a S/N of 10 the LOQ was calculated 0.19 ng/L (i.e. 0.5 ng/L/(27/10)= 0.19 ng/L).



Part D: Method summary

Table S2: Detailed information on analytical methods used for the analysis of steroidal estrogens.

	Laboratory 1	Laboratory 2	Laboratory 3
Sample preparation	Extracts were evaporated to dryness under gentle N2 stream and then reconstituted to 0.2 ml reconstituting solution (NH4OH 0.1%: AcN, 9:1, % v/v)	No additional clean-up was performed on the samples, except for sample 20 and sample 23. For these samples the silica SPE clean-up was repeated as carried out by BDS (but this time using 1 SPE column per extract) Generally, 500 µL extract were taken, IS added, evaporated till dryness, and reconstituted in 50 µL MeOH and 50 µL Water; concentration factor 1:5000	Internal Standards were added. Extracts were evaporated to dryness under gentle N2 stream and then reconstituted in 0.5 ml methanol followed by 0.5 ml water.
System	LC-MS/MS	LC-MS/MS	LC-MS/MS
Pumps	Binary Solvent Manager, Model UPB, Waters (Milford, MA, USA).	BinaryPump: Agilent G7120A Quaternary Pump for post-column addition: G1311C	Agilent G1312B binary LC pump, a G1310B isocratic LC pump
Autosampler	Sample Manager, Model UPA, Waters (Milford, MA, USA).	Multisampler: Agilent G7167B	Agilent G1367E
Ionization method	ESI Negative	ESI Negative	ESI Negative
Detector	QTRAP 5500 MS	Agilent 6495 QQQ	Sciex 6500 Qtrap One Quantifier MRM transition Two Qualifier MRM transition
Flow rate	400 µl/min	500 µL/min	300 µL/min
Injection volume	30 µl	20µL	10, 30, 100 µL
Column	Hypersil GOLD, 1.9 µm, 50 x 2.1 mm, Thermo Scientific	ACQUITY UPLC BEH Shield RP18 Column, 130Å, 1.7 µm, 2.1 mm X 100 mm, [Waters 186002854] at 40°C	Poroshell C18-EC column (3 x 50 mm, 2.7 µm, Agilent) at 25 °C
Gradient	A: NH4OH 0.1 % B: Acetonitrile 0 to 0.5 min 10% B, from 1 to 5 min 40% B, from 5 to 6 min 90% B, from 6.5 to 12 min 10% B	A: 5mM NH3 in water B: Methanol, postcolumn addition of 0.05ml/min 600µM NH4F 0 to 0.19 min, 10% B; from 0.19 to 0.2 min, 10 to 30% B; from 0.2 to 9.0 min, 30 to74% B; from 9.0 to 10.3 min 74 to 100% B; from 10.3 to 14 min, 100% B; equilibration time 3min before analysis	A: Water B: Acetonitrile 0 to 0.5 min 10% B; from 0.5 to 1.0 min 10 to 45% B; from 1.0 to 9.0 min 45 to 60% B; from 9.0 to 9.1 min 60 to 98% B; from 9.1 to 12 min 98% B; from 12.0 to 12.1 min 98% to 10% B, from 12.1 to 15.0 min 10% B
Internal Standards	E1 13C3, E2 d4, EE2 d4	Estrone-2,4,16,16-d4 (D-3650, cdnisotopes), Estradiole-2,4,16,16-d4 (DLM-2487-0, CIL), 17-α-EE2-2,4,16,16-D4 (DLM-4691, CIL)	Estrone 2, 4, 16, 16 - d4 (E2-d4), 17beta-estradiol 2,4,16,16-d4,and 17a-ethinyl estradiol 2, 4, 16, 16-d4 (EE2-d4) were obtained from Toronto Research Chemicals (North York, ON, Canada)

158 **Table S3: Detailed information on effect-based methods used for determination of EEQs.**

	ERα CALUX	MELN	ER-GeneBLAzer	HeLa-9903	pYES
Cell-line	Human U2OS osteo-sarcoma cells stably transfected with 3xHRE-TATA-Luc and pSG5-neo-hERα constructs using calcium phosphate co-precipitation method	Human MCF-7 breast cancer cells stably transfected with an ERE-βGlobine-luciferase construct. This cell line expresses endogenously hERα and hERβ while only hERα is functional and activates the reporter gene	GeneBLAzer® ER alpha -UAS-bla GripTite™ cells (genetically engineered from HEK293) using β-lactamase reporter gene under control of a UAS response element	Human HeLa-9903 cervical tumor cells stably transfected with two constructs: (i) the hERα expression construct (encoding the full-length human receptor), and (ii) a firefly luciferase reporter construct bearing five tandem repeats of a vitellogenin Estrogen-Responsive Element (ERE) driven by a mouse metallothionein promoter TATA element.	*Saccharomyces cerevisiae* BJ3505 transfected with the plasmids YEPE10 and YRPEG3. YEPE10 contains a CUP1::hER fusion encoding the human estrogen receptor alpha cloned from the MCF-7 human cell lineage under the control of the metallothionein promoter CUP1. YRPEG3 contains the fusion gene 2ERE-CyC1::lacZ. This fusion gene express β-galactosidase under the control of the CyC1 promoter from *S. cerevisiae* fused to two copies of the vitellogenin A2-gene from *Xenopus laevis*.
Endpoint for estrogenicity	Luciferase activity	Luciferase activity	β-Lactamase via fluorescence (FRET) reagent with combined fluorescence cytotoxicity measure	Luciferase activity	β-Galactosidase activity
Assay medium	Phenol red free D-MEM/F 12) containing 5% DCC-FCS	Phenol red free DMEM containing 3% DCC-FBS (v/v)	Phenol red free DMEM (Gibco 10569-010) containing 2% DCC-FBS (Gibco 12676-011), 1% Penicillin-Streptomycin (Gibco 15140-122), 1mM Sodium Pyruvate (Gibco 11360-070), 0.1 mM NEAA (Gibco 11140-050)	Phenol red free DMEM - F12 (Sigma Aldrich, USA) + fetal calf serum (10% v/v) for maintanance. For experiments: Medium DMEM - F12 (Sigma Aldrich, USA) + dialyzed DCC-FBS (10% v/v)	28 mg/ml Yeast nitrogen base w/o amino acids, 130 mg/ml Glucose, 150 µg/ml L-Lysine-HCL, 100 µg/ml L-Histidine-HCl, 25 µg/ml $CuSO_4 \bullet 7H_2O$, supplemented with Ampicillin and Streptomycin
Assay format	96-well microtiter plates	96-well microtiter plates	384-well microtiter plates	96-well microtiter plates	High performance thinlayer plate (silica)
Cells per well seeded	1 x 10^4	0.5-1 x 10^4	2 x 10^4	1 x 10^4	n.a.
Dosing of cells	After 24 h seeding, the growth medium is replaced by the assay medium containing the reference compound or the sample extract in respective dilutions (100 µl/well).	After 24h seeding in medium (100 µl/well) and another 50 µl/well of assay medium containing the test chemical or sample (3x concentrated) is added to the cells. Only internal wells on the assay plate	8 µl/well of dosing media containing the test chemical or sample	Cells are seeded for 24h in medium (100 µl/well) and another 100 µl/well of dosing media containing the test chemical or sample is added for cells exposure. Total volume of media is therefore 200 µl/well.	Overnight-culture of yeast cells is adjusted to 1000 FNU. 5 ml of the cell suspension is applied evenly on a TLC-plate (10x20 cm) after chromatographic separation of the extract. Applied volumes of the

		are used.			extracts vary between 5 µl and 100 µl depending on the expected contamination level.
Incubation period for exposure	22-24h	16-24h	16h	24h	3h
Number of replicate wells	3	3 (manual method) or 4 (automated method)	2 replicates + minimum 2 independent repeats of the assay	3	n.a. Assay is repeated three times independently
Negative controls	0.1% DMSO in the assay medium in triplicate pro assay plate (negative control)	0.1-0.5% DMSO in the assay medium (depending on the expected low activities) and assay medium without DMSO	32 replicates of assay medium per assay plate (No negative controls[a])	0.1% (v/v) MeOH on every plate in triplicates and medium in triplicates	5 µl of Ethanol for negative control, additionally 100 µl of the extracted field blank sample applied on a separate lane on the TLC-plate
Concentration range tested of the reference compound, 17β-estradiol (E2)	0.03-27 ng/L The full concentration range was tested on each assay plate	0.03 – 27 ng/L The full concentration range was tested in each experimental series and a fixed concentration of E2 (10 nM) E2 on each assay plate (in in sextuplicate)	0.3 – 545 ng/L The full concentration range (6 - 12 concentrations in duplicate) was tested on each plate	0.001 - 10 ng/L The full concentration range (at least 5 concentrations in triplicate) was tested on each plate	1 pg/L to 10 pg/L E2 and EE2, 10 pg/L to 100 pg/L E1, 100 pg/L to 1000 pg/L E3
Cell harvesting at the end of the exposure and response detection	Exposure medium was removed and cells were lysed with 30 µL Triton-lysis buffer to open up the cell membrane. Then the luciferin substrate mix was added to the cells and the luciferase activity was measured at 0.1 min/well	Cells were first washed with PBS buffer (optional) and 50 µL of medium containing 30 mM of D-luciferin substrate was added. Cells were not lysed. After 5 min plates were read on a luminometer, 1s per well.	8µl per well and incubated for 2h at room temperature Fluorescence was measured immediately after adding the substrate buffer (time 0h for correction of autofluorescence) and then after 2h-incubation using the same gain for both measurements	50 µL of cell lysis solution enriched by luciferin solution was added to each well. Plates were read with a luminometer after 10 min of shaking (150 rotations/min) at room temperature	After exposure the TLC-plate is developed by the application of the following buffer: 10 mg/ml $Na_2HPO_4 \bullet 2 H_2O$, 0.75 mg/ml KCl 0.25 mg/ml $MgSO_4 \bullet 7 H_2O$, 1 mg/ml SDS, 0.5 mg/ml MUG, pH=7.0 After application the plate is incubated for 15 min at 37°C. The fluorescence-signal of the developed Methyl-umbelliferon is detected at 254 nm.

Abbreviations:**DCC-FCS**: Dextran-charcoal treated Fetal Calf Serum; **DCC-FBS**: Dextran-charcoal treated Fetal Bovine Serum; **D-MEM**: Dulbecco´s modified Eagle´s medium; **MUG**: 4-Methylumbelliferyl β-D-galactopyranoside; **SDS**: Sodium dodecyl sulfate

[a]The extracts were dried and resuspended in assay medium. DMSO stock solution of reference compound is diluted $5x10^5$-times in the assay medium. Therefore, the amount of solvent in the positive control is negligible and negative control was not tested.

[b] Phosphate buffered saline

Part E: General and specific validity criteria of the bioanalytical and chemical analysis methods

- No estrogenic activity or quantifiable concentration of E1, E2 and EE2 were measured in the blank samples (i.e. procedure-, extraction- and solvent blanks)
- Derived effect concentrations in the bioassays and chemically measured EE2 concentrations matched with the nominal concentrations of the spiked samples
- Participating laboratories reported their raw data in a standardized format and all data were then analysed centrally in a harmonized way.

Table S4: Measured chemical concentrations and 17β-estradiol (E2) equivalent concentrations (EEQ) of the positive control samples with high or low concentrations of E2 and EE2 are compared with the nominal chemical concentrations and calculated nominal EEQ concentrations.

Spiked water	Nominal spike conc.	Chemically determined spike conc.*	Measured conc. as % of nominal spike conc.	EEQ determined in the *in vitro* assays**	$EEQ_{nominal}$***	EEQ as % of $EEQ_{nominal}$
	pg/L	pg/L		pg/L	pg/L	
$E2_{low}$	600	617	103%	772	678	114%
$E2_{high}$	6000	5651	94%	4914	5717	86%
$EE2_{low}$	600	770	128%	914	975	94%
$EE2_{high}$	6000	6341	106%	6134	7548	81%
AVG ± SD			108 ± 15%			94 ± 14%

*Averaged spiked concentration (pg/L) of the three measured values provided by the analytical labs (JRC, BfG, OZ)

**Averaged EEQ (pg/L) obtained in the 5 *in vitro* bioassay

***$EEQ_{nominal}$ (pg/L) was calculated by multiplying the chemically measured and averaged E2 or EE2 concentrations by their assay-specific relative potency (REP). In all four spiked water samples E1 concentration was measured. Similarly to E2 and EE2, the chemically measured E1 concentrations were also translated into EEQ concentrations and added to the pertinent $EEQ_{nominal}$ concentration.

Effect-based methods

- Each participating laboratory used the same batch of certified E2 standard as reference compound in their bioassay and the same batch of E1 and EE2 standard solutions to determine their assay-specific relative potencies (REPs). REPs are summarised in Table S4.
- Sample dilutions with observed cytotoxicity were not taken for data evaluation.
- Additionally to the assessment criteria of the respective laboratory (e.g. acceptable CV of replicate measurements), the reported data fulfilled sufficient criteria elements that allowed for a robust data evaluation. The evaluation was performed collectively by a group of three experts. Bioassay data should have included a full reference dose-response curve (below 10% and above 90 %) and had sufficient values at the lower effect levels (e.g. <10 % effect) for both reference and samples. These latter criteria were especially important for the PC_{10} EEQ derivation method. In the case of incomplete curves, at least two effect measurements above 10 % were obtained that allowed curve fitting.

Table S5: Assay-specific mass-based 17β-estradiol relative potency factors (REPs). REP shows the relative potency of the compound in the certain bioassay compared to the reference compound (E2 in this case). REPs were determined by dividing the 50 % effect concentration (EC_{50}; g/L) of the reference compound by the EC_{50} of the test compound (g/L).

REP (g_{E2}/g_i)	E1	E2	EE2
ERα-CALUX	0.01	1.0	1.2
MELN	0.29	1.0	0.79
ER-GeneBLAzer	0.08	1.0	1.67
HeLa-9903	0.02	1.0	1.18
pYES	0.11	1.0	1.0

LC-MS/MS

- Only peaks with a signal-to-noise (S/N) ratio > 10:1 were evaluated and quantified.
- The acceptable retention time deviation relative to the pertinent internal standard was <0.2 min within the same sequence and the uncertainty of the data qualifier <30 %.

Part F: Results of chemical analytical and effect-based methods

Table S6: Results of the three chemical analytical methods. Measured concentrations are shown for surface and waste water in pg/L.

		E1 in pg/L			E2 in pg/L			EE2 in pg/L		
		Lab 1	**Lab 2**	**Lab 3**	**Lab 1**	**Lab 2**	**Lab 3**	**Lab 1**	**Lab 2**	**Lab 3**
Spikes and Blanks	Blank	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD
	Blank-Spike	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD	<LOD
	E2 600	1020	760	800	741	571	540	<LOD	<LOD	<LOD
	E2 6000	773	531	640	6022	5531	5400	<LOD	<LOD	<LOD
	EE2 600	703	524	560	<LOD	<LOD	<LOD	685	874	750
	EE2 6000	204	108	120	<LOD	<LOD	<LOD	6170	6654	6200
Surface water samples	**A(11)**	82	<LOQ	51	<LOQ	<LOD	<LOD	<LOD	<LOD	<LOD
	B(6)	88	<LOQ	70	<LOQ	<LOD	<LOD	<LOD	<LOD	<LOD
	C(1)	186	96	130	<LOQ	<LOQ	<LOQ	<LOD	<LOD	<LOD
	D(22)	189	114	130	<LOQ	<LOD	<LOQ	<LOD	<LOD	<LOD
	E(27)	221	136	160	<LOQ	<LOQ	<LOQ	73	75	85
	F(30)	400	318	360	<LOQ	<LOD	<LOQ	<LOD	<LOD	<LOD
	G(32)	444	282	360	<LOQ	<LOQ	70	<LOD	<LOD	<LOD
	H(25)	732	461	600	<LOQ	<LOQ	<LOQ	<LOD	<LOQ	<LOD
	I(8)	717	691	630	<LOQ	81	<LOD	<LOD	<LOD	<LOD
	J(10)	933	702	640	<LOQ	59	70	<LOD	<LOD	<LOD
	K(18)	1348	1106	960	261	171	180	186	161	90
	L(24)	1971	1124	1600	259	<LOQ	190	<LOD	<LOQ	<LOD
	M(28)	3353	2212	3000	267	230	240	<LOD	<LOQ	<LOD
	N(15)	5626	3307	4400	862	431	220	<LOD	<LOQ	<LOD
	O(3)	6186	5220	4000	311	441	<LOD	354	230	<LOD
	P(7)	7136	4846	4700	403	215	<LOD	100	<LOQ	<LOD
Waste water samples	**A(26)**	103	<LOQ	56	<LOQ	<LOD	<LOD	<LOD	<LOD	<LOD
	B(29)	116	87	96	<LOQ	<LOQ	<LOD	<LOD	<LOQ	<LOD
	C(31)	215	125	130	<LOD	<LOQ	<LOD	<LOD	<LOD	<LOD
	D(4)	268	162	210	<LOQ	<LOQ	<LOD	<LOD	<LOQ	<LOD
	E(17)	390	284	250	<LOQ	<LOD	<LOD	119	<LOQ	<LOD
	F(21)	509	364	460	<LOQ	<LOD	<LOD	<LOD	<LOD	<LOD
	G(14)	852	816	810	<LOQ	<LOD	<LOD	<LOD	<LOQ	<LOD
	H(5)	2502	1848	1900	<LOQ	<LOQ	<LOD	<LOD	<LOQ	<LOD
	I(19)	2955	3150	2500	<LOQ	<LOQ	<LOD	<LOD	<LOQ	<LOD
	J(16)	4351	5490	3300	<LOQ	<LOQ	<LOD	<LOD	218	<LOD
	K(9)	5350	6020	4700	<LOQ	<LOQ	<LOD	<LOD	<LOQ	<LOD
	L(13)	8022	7971	5500	1177	551	<LOD	<LOD	<LOQ	<LOD
	M(23)	9485	5940	7200	<LOQ	1670	220	6124	7022	9400
	N(33)	10606	13080	12000	323	478	350	<LOD	<LOQ	<LOD
	O(12)	14746	15300	11000	461	471	430	<LOD	<LOQ	<LOD
	P(2)	14482	16000	12000	839	540	620	<LOD	<LOQ	<LOD
	Q(20)	22648	21000	18000	1031	1074	1200	5864	5160	4000

Table S7: Limits of quantification (LOQs) of the three chemical analytical methods. Determined concentrations are shown for surface and waste water in pg/L.

		LOQ for E1 in pg/L			LOQ for E2 in pg/L			LOQ for EE2 in pg/L		
		Lab 1	Lab 2	Lab 3	Lab 1	Lab 2	Lab 3	Lab 1	Lab 2	Lab 3
Spikes and Blanks	Blank	42	50	90	240		360			300
	Blank-Spike	42	50	90	312		360			300
	E2 600		50	90		60	360			300
	E2 6000		50	90		80	360			300
	EE2 600		50	90			360		50	300
	EE2 6000		50	90			360		80	300
Surface water samples	**A(11)**	9	50	15	108	50	120	45	60	30
	B(6)	3	50	15	66	50	45	36	70	30
	C(1)	6	50	15	147	50	45	36	50	30
	D(22)	6	50	15	63	50	45	24	80	30
	E(27)	9	50	15	39	50	45	60	70	30
	F(30)	18	50	15	90	70	120	600	100	90
	G(32)	12	50	15	63	80	45	78	50	30
	H(25)	30	50	15	234	80	120	120	90	90
	I(8)	24	50	45	360	50	120	72	120	90
	J(10)	9	50	15	72	50	45	30	70	30
	K(18)	21	50	15	240	80	120	135	150	90
	L(24)	192	50	45	240	190	120	600	140	300
	M(28)	18	50	45	240	130	120	240	140	300
	N(15)	168	50	45	240	170	120	390	70	300
	O(3)	327	170	45	240	360	1500	60	230	900
	P(7)	45	100	45	240	270	1500	72	200	300
Waste water samples	**A(26)**	21	50	90	60	50	300	90	70	300
	B(29)	3	50	90	51	50	300	48	60	300
	C(31)	24	50	90	228	110	900	114	120	1500
	D(4)	12	100	90	450	160	3000	279	120	300
	E(17)	75	50	90	126	200	900	120	60	300
	F(21)	15	50	90	135	140	900	420	190	300
	G(14)	42	90	90	579	210	1500	900	290	1500
	H(5)	90	90	90	1080	670	3000	399	240	3000
	I(19)	27	50	90	480	140	900	1410	90	300
	J(16)	90	100	90	738	270	1500	2100	140	1500
	K(9)	123	50	90	1020	180	1500	720	80	3000
	L(13)	33	50	90	240	80	3000	600	110	300
	M(23)	144	1000	90	1200	290	600	810	1200	3000
	N(33)	33	70	90	240	220	900	1500	250	1500
	O(12)	306	120	90	240	290	900	1110	500	1500
	P(2)	24	50	90	240	80	360	240	70	300
	Q(20)	156	90	90	240	150	900	1737	140	300

Table S8: EEQ determined by effect-based methods using the PC_{10} approach.

		EEQ_{Bio} in pg/L				
		ER CALUX	MELN	HeLa	ER-GeneBlazer	pYES
Surface water samples	A(11)	80	79	16	39	130
	B(6)	80	59	16	22	90
	C(1)	60	182	17	32	370
	D(22)	120	191	35	41	390
	E(27)	200	170	245	94	190
	F(30)	150	321	28	76	160
	G(32)	110	375	100	137	270
	H(25)	230	391	97	111	440
	I(8)	260	831	137	230	410
	J(10)	150	458	89	92	280
	K(18)	640	809	659	342	600
	L(24)	410	2124	509	407	2060
	M(28)	860	2498	226	487	1300
	N(15)	1180	1942	1039	876	2900
	O(3)	920	4019	487	959	5430
	P(7)	590	2787	347	835	3600
Waste water samples	A(26)	30	37	31	25	100
	B(29)	50	78	118	74	100
	C(31)	70	254	41	143	65
	D(4)	60	244	90	108	410
	E(17)	120	434	86	111	690
	F(21)	310	656	298	254	160
	G(14)	490	2066	75	567	970
	H(5)	370	1793	273	452	290
	I(19)	310	3104	553	560	1200
	J(16)	480	2936	888	813	2400
	K(9)	480	2233	692	501	340
	L(13)	870	2824	1205	1166	1700
	M(23)	22930	19716	24144	11892	12000
	N(33)	1050	3807	2407	1369	2400
	O(12)	700	5493	1695	1212	3400
	P(2)	820	4928	734	1360	2800
	Q(20)	7590	10851	6442	6317	9700

205 **Table S9: Limits of detection (LODs) and quantification (LOQs) of effect-based methods.**

		ER-CALUX		MELN		HeLa-993		ER-GeneBLAzer		pYES	
		LOD in pg/L	LOQ in pg/L	LOD in pg/L	LOQ in pg/L	LOD in pg/L	LOQ in pg/L	LOD in pg/L	LOQ in pg/L	LOD in pg/L	LOQ in pg/L
Surface water samples	A(11)	0	12	2	16	8	37	11	43	3	10
	B(6)	4	19	0	9	5	19	7	26	3	10
	C(1)	6	11	3	15	0	57	0	48	3	10
	D(22)	0	5	3	8	18	73	5	20	3	10
	E(27)	4	7	8	22	2	8	6	22	3	10
	F(30)	7	22	9	32	6	33	6	23	3	10
	G(32)	2	3	8	22	2	9	6	22	3	10
	H(25)	5	8	0	7	8	29	6	22	7	20
	I(8)	4	0	5	16		24	7	24	7	20
	J(10)	0	13	8	25	44	125	7	24	3	10
	K(18)		5	3	8		24	8	27	7	20
	L(24)	3	9	5	15		79	7	24	33	100
	M(28)	4	10	3	19		79	32	113	17	50
	N(15)		2	8	20	2	10	6	21	33	100
	O(3)	5	6	6	19		20	6	23	67	200
	P(7)	1	3	7	20	5	25	6	22	33	100
Waste water samples	A(26)	6	11	3	10	31	85	11	43	3	10
	B(29)	4	12	5	17	9	58	6	23	3	10
	C(31)	9	27	6	17	12	43	11	42	17	50
	D(4)	4	6	0	14	8	12	55	216	33	100
	E(17)	16	37	6	17	5	43	6	22	22	67
	F(21)	7	22	3	17	31	85	6	22	22	67
	G(14)	10	26	0	12	0	14	5	21	17	50
	H(5)	0	0	5	12	1	17	5	21	67	200
	I(19)	3	9	8	18	6	23	5	21	33	100
	J(16)	0	2	0	11	24	93	7	27	67	200
	K(9)	3	6	2	12	14	70	55	213	67	200
	L(13)	4	16	5	14	0	70	5	21	67	200
	M(23)	6	15	0	13	8	46	11	44	67	200
	N(33)	5	8	6	20	29	132	5	21	67	200
	O(12)	2	6	4	8	66	197	13	53	33	100
	P(2)	15	36	3	17	2	9	5	18	33	100
	Q(20)	4	7	0	9	27	131	6	22	67	200

Part G: Iceberg modelling

Table S10: Sums of EEQ_{chem} based on the REPs obtained with ER-CALUX. For the left part of the table (>LOQ) only analytical data above the LOQ was included, while for the right part data below LOD or LOQ was substituted by LOD/2 and LOQ/2.

			>LOQ			<LOD = LOD/2 <LOQ = LOQ/2		
		ER-CALUX EEQ in pg/L	Sum EEQ_{chem} in pg/L			Sum EEQ_{chem} in pg/L		
			Lab 1	Lab 2	Lab 3	Lab 1	Lab 2	Lab 3
Surface water samples	N(15)	1180	918	464	264	996	506	324
	K(18)	640	498	375	298	498	375	298
	M(28)	860	301	252	270	349	336	330
	O(3)	920	798	769	40	798	769	470
	F(30)	150	4	3	4	49	35	82
	P(7)	590	594	263	47	594	383	357
	L(24)	410	279	11	206	399	190	266
	H(25)	230	7	5	6	148	99	84
	J(10)	150	9	66	76	51	80	82
	E(27)	200	90	91	104	109	116	126
	G(32)	110	4	3	74	52	53	80
	I(8)	260	7	88	6	202	112	44
	A(11)	80	1	0	1	64	21	27
	B(6)	80	1	0	1	41	23	14
	C(1)	60	2	1	1	83	36	30
	D(22)	120	2	1	1	38	25	30
Waste water samples	E(17)	120	147	3	3	210	72	213
	A(26)	30	1	0	1	49	23	111
	L(13)	870	1257	631	55	1377	697	615
	K(9)	480	54	60	47	708	198	897
	F(21)	310	5	4	5	157	65	215
	H(5)	370	25	18	19	645	497	1119
	D(4)	60	3	2	2	283	154	562
	O(12)	700	608	624	540	830	924	840
	J(16)	480	44	317	33	833	452	583
	P(2)	820	984	700	740	1032	742	800
	Q(20)	7590	8294	7476	6180	8294	7476	6180
	M(23)	22930	7444	10156	11572	8044	10156	11572
	G(14)	490	9	8	8	478	217	558
	I(19)	310	30	32	25	552	156	235
	B(29)	50	1	1	1	36	62	111
	N(33)	1050	429	609	470	729	759	770
	C(31)	70	2	1	1	63	80	451

Table S11: Sums of EEQ_{chem} based on the REPs obtained with MELN. For the left part of the table (>LOQ) only analytical data above the LOQ was included, while for the right part data below LOD or LOQ was substituted by LOD/2 and LOQ/2.

			>LOQ			<LOD = LOD/2 <LOQ = LOQ/2		
		MELN EEQ in pg/L	Sum EEQ_{chem} in pg/L			Sum EEQ_{chem} in pg/L		
			Lab 1	Lab 2	Lab 3	Lab 1	Lab 2	Lab 3
Surface water samples	N(15)	1942	2494	1390	1496	2545	1418	1536
	K(18)	809	799	619	530	799	619	530
	M(28)	2498	1239	871	1110	1271	927	1150
	O(3)	4019	2385	2137	1160	2385	2137	1529
	F(30)	321	116	92	104	161	117	176
	P(7)	2787	2551	1620	1363	2551	1699	1653
	L(24)	2124	831	326	654	910	476	694
	H(25)	391	212	134	174	345	209	246
	J(10)	458	271	263	256	311	272	260
	E(27)	170	122	99	114	141	124	136
	G(32)	375	129	82	174	171	128	178
	I(8)	831	208	281	183	397	297	215
	A(11)	79	24	0	15	84	23	39
	B(6)	59	26	0	20	63	25	32
	C(1)	182	54	28	38	132	59	64
	D(22)	191	55	33	38	89	52	64
Waste water samples	E(17)	434	207	82	73	270	139	262
	A(26)	37	30	0	16	72	25	106
	L(13)	2824	3503	2863	1595	3582	2906	2135
	K(9)	2233	1552	1746	1363	2156	1867	2008
	F(21)	656	148	106	133	270	154	323
	H(5)	1793	726	536	551	1318	966	1446
	D(4)	244	78	47	61	339	174	600
	O(12)	5493	4737	4908	3620	4883	5106	3818
	J(16)	2936	1262	1764	957	1907	1899	1405
	P(2)	4928	5039	5180	4100	5070	5208	4140
	Q(20)	10851	12231	11240	9580	12231	11240	9580
	M(23)	19716	7589	8940	9734	8189	8940	9734
	G(14)	2066	247	237	235	655	386	682
	I(19)	3104	857	914	725	1283	1019	915
	B(29)	78	34	25	28	65	74	117
	N(33)	3807	3399	4271	3830	3596	4370	4028
	C(31)	254	62	36	38	115	107	385

Table S12: Sums of EEQ_{chem} based on the REPs obtained with HeLa-9903. For the left part of the table (>LOQ) only analytical data above the LOQ was included, while for the right part data below LOD or LOQ was substituted by LOD/2 and LOQ/2.

			>LOQ			<LOD = LOD/2 <LOQ = LOQ/2		
		HeLa-9903 EEQ in pg/L	Sum EEQ_{chem} in pg/L			Sum EEQ_{chem} in pg/L		
			Lab 1	Lab 2	Lab 3	Lab 1	Lab 2	Lab 3
Surface water samples	N(15)	1039	975	497	308	1051	538	367
	K(18)	659	507	383	305	507	383	305
	M(28)	226	334	274	300	381	357	359
	O(3)	487	852	817	80	852	817	507
	F(30)	28	8	6	7	171	38	85
	P(7)	347	664	312	94	664	430	403
	L(24)	509	298	22	222	416	200	281
	H(25)	97	15	9	12	155	102	90
	J(10)	89	19	73	83	61	87	89
	E(27)	245	91	91	104	110	116	126
	G(32)	100	9	6	77	56	55	83
	I(8)	137	14	95	13	209	118	50
	A(11)	16	2	0	1	64	21	27
	B(6)	16	2	0	1	42	23	15
	C(1)	17	4	2	3	84	37	31
	D(22)	35	4	2	3	40	26	31
Waste water samples	E(17)	86	148	6	5	211	74	214
	A(26)	31	2	0	1	50	23	110
	L(13)	1205	1337	710	110	1455	775	669
	K(9)	692	107	120	94	759	258	934
	F(21)	298	10	7	9	160	68	218
	H(5)	273	50	37	38	669	514	1128
	D(4)	90	5	3	4	285	154	563
	O(12)	1695	756	777	650	974	1072	945
	J(16)	888	87	367	66	869	502	611
	P(2)	734	1129	860	860	1176	901	919
	Q(20)	6442	8403	7583	6280	8403	7583	6280
	M(23)	24144	7416	10075	11456	8016	10075	11456
	G(14)	75	17	16	16	484	222	561
	I(19)	553	59	63	50	576	186	259
	B(29)	118	2	2	2	37	62	111
	N(33)	2407	535	740	590	830	887	885
	C(31)	41	4	3	3	65	81	448

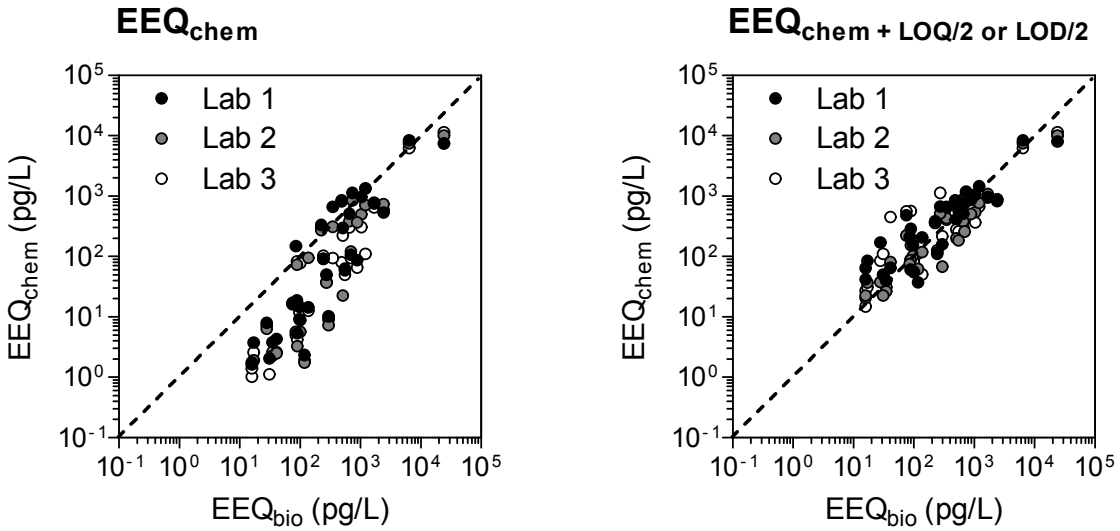
Table S13: Sums of EEQ_{chem} based on the REPs obtained with ER-GeneBLAzer. For the left part of the table (>LOQ) only analytical data above the LOQ was included, while for the right part data below LOD or LOQ was substituted by LOD/2 and LOQ/2.

			>LOQ			<LOD = LOD/2 <LOQ = LOQ/2		
		ER-GeneBLAzer EEQ in pg/L	Sum EEQ_{chem} in pg/L			Sum EEQ_{chem} in pg/L		
			Lab 1	Lab 2	Lab 3	Lab 1	Lab 2	Lab 3
Surface water samples	N(15)	876	1312	696	572	1421	754	656
	K(18)	342	679	528	407	679	528	407
	M(28)	487	535	407	480	602	524	564
	O(3)	959	1397	1243	320	1397	1243	821
	F(30)	76	32	25	29	244	65	114
	P(7)	835	1141	603	376	1141	770	710
	L(24)	407	417	90	318	584	302	402
	H(25)	111	59	37	48	209	152	133
	J(10)	92	75	115	121	119	135	130
	E(27)	94	140	136	155	159	161	177
	G(32)	137	36	23	99	89	76	107
	I(8)	230	57	136	50	257	170	95
	A(11)	39	7	0	4	73	27	32
	B(6)	22	7	0	6	50	30	21
	C(1)	32	15	8	10	98	47	41
	D(22)	41	15	9	10	53	40	41
Waste water samples	E(17)	111	230	23	20	293	106	254
	A(26)	25	8	0	4	63	30	138
	L(13)	1166	1819	1189	440	1986	1281	1024
	K(9)	501	428	482	376	1138	638	1461
	F(21)	254	41	29	37	225	105	270
	H(5)	452	200	148	152	851	683	1487
	D(4)	108	21	13	17	324	193	600
	O(12)	1212	1641	1695	1310	1950	2113	1728
	J(16)	813	348	803	264	1302	938	932
	P(2)	1360	1998	1820	1580	2064	1878	1664
	Q(20)	6317	12636	11371	9320	12636	11371	9320
	M(23)	11892	10986	13872	16494	11586	13872	16494
	G(14)	567	68	65	65	608	342	732
	I(19)	560	236	252	200	869	397	434
	B(29)	74	9	7	8	48	82	141
	N(33)	1369	1171	1524	1310	1589	1733	1728
	C(31)	143	17	10	10	87	98	578

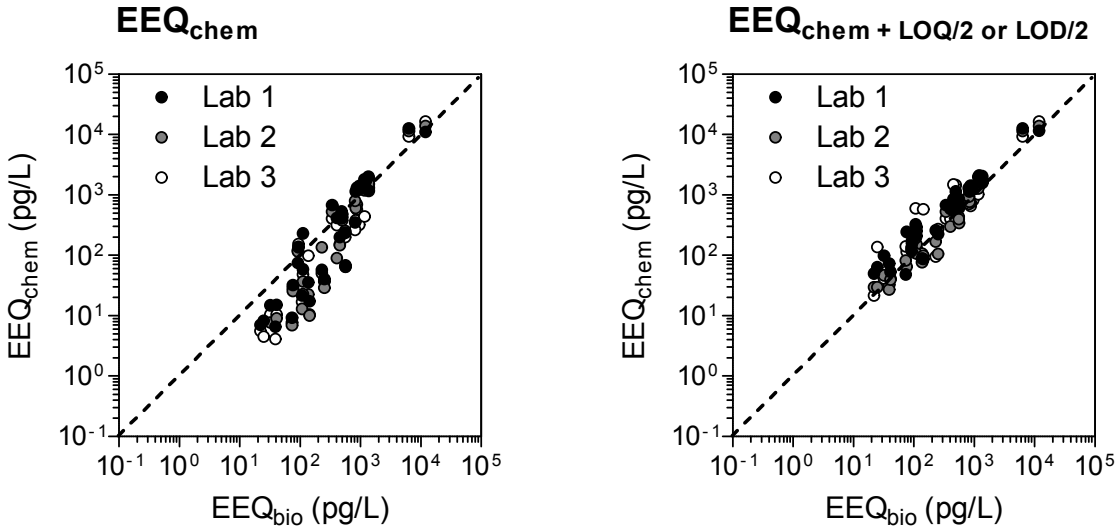
Table S14: Sums of EEQ_{chem} based on the REPs obtained with pYES. For the left part of the table (>LOQ) only analytical data above the LOQ was included, while for the right part data below LOD or LOQ was substituted by LOD/2 and LOQ/2.

			>LOQ			<LOD = LOD/2 <LOQ = LOQ/2		
		pYES EEQ in pg/L	Sum EEQ_{chem} in pg/L			Sum EEQ_{chem} in pg/L		
			Lab 1	Lab 2	Lab 3	Lab 1	Lab 2	Lab 3
Surface water samples	N(15)	2900	1481	795	704	1546	830	754
	K(18)	600	595	454	376	595	454	376
	M(28)	1300	636	473	570	676	543	620
	O(3)	5430	1345	1245	440	1345	1245	840
	F(30)	160	44	35	40	189	63	115
	P(7)	3600	1288	748	517	1288	848	817
	L(24)	2060	476	124	366	576	289	416
	H(25)	440	81	51	66	218	136	141
	J(10)	280	103	136	140	144	148	145
	E(27)	190	97	90	103	117	115	125
	G(32)	270	49	31	110	93	79	115
	I(8)	410	79	157	69	271	177	104
	A(11)	130	9	0	6	71	21	31
	B(6)	90	10	0	8	49	23	20
	C(1)	370	20	11	14	100	44	42
	D(22)	390	21	13	14	56	34	42
Waste water samples	E(17)	690	162	31	28	225	95	228
	A(26)	100	11	0	6	56	23	106
	L(13)	1700	2059	1428	605	2159	1483	1155
	K(9)	340	589	662	517	1219	792	1267
	F(21)	160	56	40	51	193	95	251
	H(5)	290	275	203	209	882	658	1209
	D(4)	410	29	18	23	301	158	573
	O(12)	2400	2083	2154	1640	2268	2404	1890
	J(16)	3400	479	822	363	1198	957	863
	P(2)	2800	2432	2300	1940	2472	2335	1990
	Q(20)	9700	9386	8544	7180	9386	8544	7180
	M(23)	12000	7167	9345	10412	7767	9345	10412
	G(14)	970	94	90	89	533	270	589
	I(19)	1200	325	347	275	800	462	475
	B(29)	100	13	10	11	46	65	111
	N(33)	2400	1490	1917	1670	1740	2042	1920
	C(31)	65	24	14	14	81	89	414

HeLa



ER-GeneBLAzer



pYES

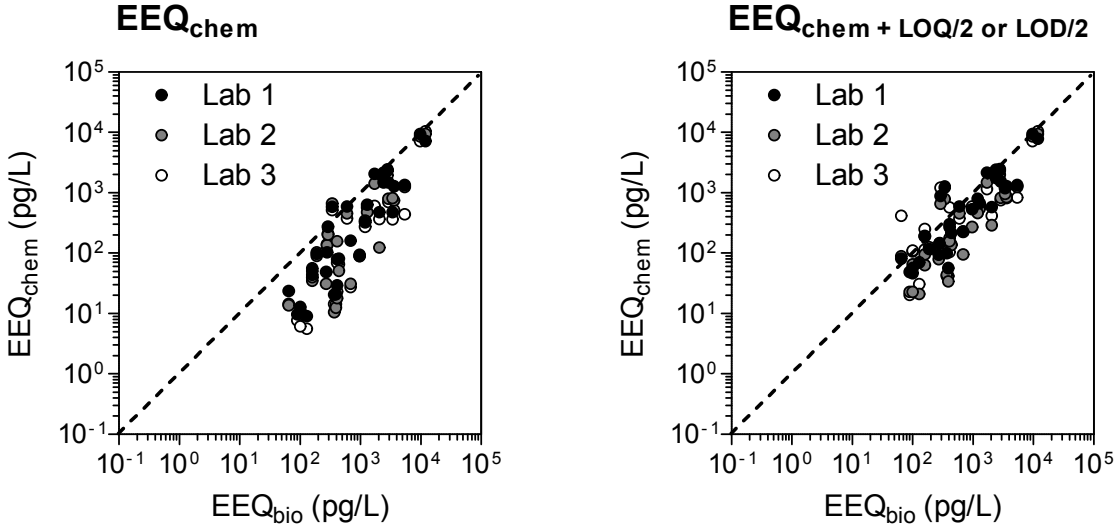


Figure S1: Comparison of EEQ_{chem} and EEQ_{bio}. Graphs on the left show the EEQ_{chem} derived from values >LOQ, while the graphs on the right show the $EEQ_{chem+LOD/2\ or\ LOQ/2}$ calculated by including LOD/2 and LOQ/2. The dashed line indicates perfect agreement of EEQ_{chem} with EEQ_{bio}.

Part H: Correlation analysis

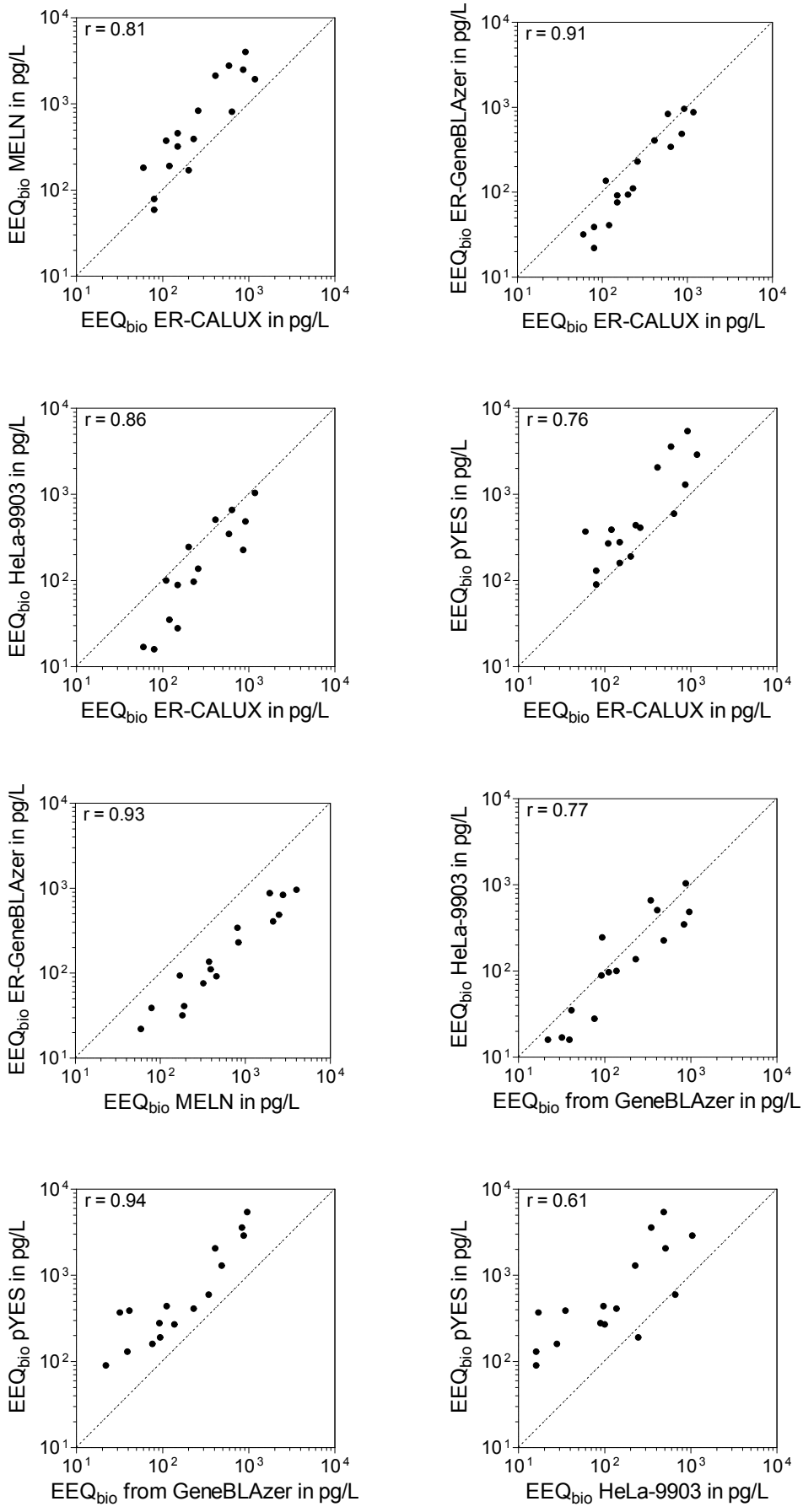


Figure S2: Comparison of effect-based methods by correlating the measured EEQs for surface water samples. The correlation analysis was based on the method described in section 2.9 with a fixed slope of 1. The dashed line indicates perfect agreement of the compared effect-based methods.

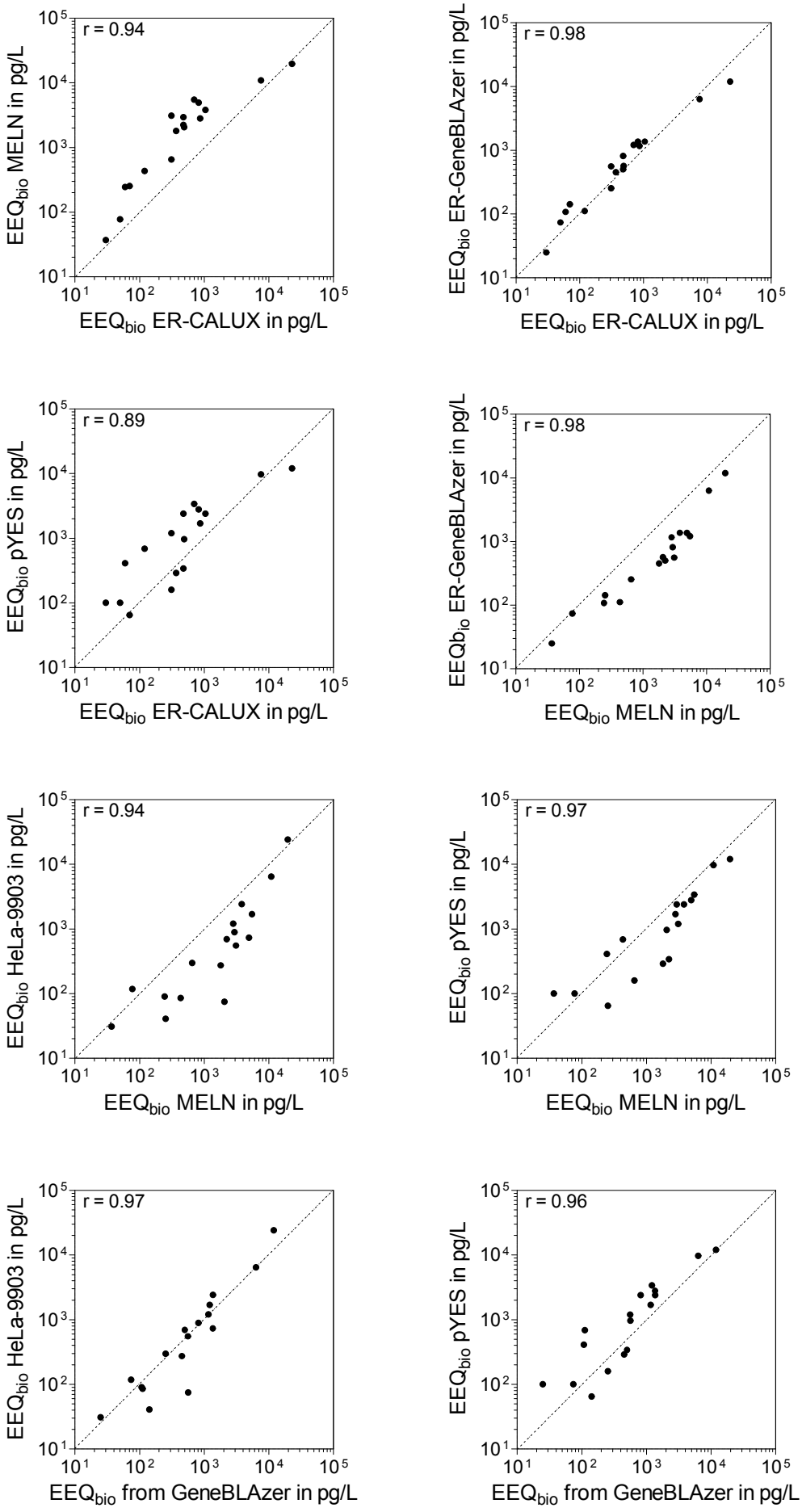


Figure S3: Comparison of effect-based methods by correlating the measured EEQs for waste water samples. The correlation analysis was based on the method described in section 2.9. The dashed line indicates perfect agreement of the compared effect-based methods.

Bibliography

1. Clement, T.P., *Authorship matrix: A rational approach to quantify individual contributions and responsibilities in multi-author scientific articles.* Science and Engineering Ethics, 2014. **20**(2): p. 345-361.
2. Kunz, P.Y., et al., *Effect-based tools for monitoring estrogenic mixtures: Evaluation of five in vitro bioassays.* Water Research, 2017. **110**: p. 378-388.