Joint Research Programme
BTO 2020.053 | September 2020

Development of a framework to derive effect-based trigger values to interpret CALUX data for drinking water quality

Joint Research Programme

KWR

Bridging Science to Practice

Report

Development of a Framework to Derive Effect-Based Trigger Values to Interpret CALUX Data for Drinking Water Quality

BTO 2020.053 | September 2020

This research is part of the Joint Research Programme of KWR, the water utilities and Vewin.

Project number
402045-048

Project manager
Patrick Bauerlein

Client
BTO - Thematical research - Chemical safety

Author(s)
Frederic Been, Tessa Pronk, Jochem Louisse, Corine Houtman, Tineke van der Velden-Slootweg, Ron van der Oost, Milou M.L. Dingemans

Quality Assurance
Thomas Ter Laak

Sent to
This report is distributed to BTO-participants and is public.

Keywords

Effect-based trigger values, CALUX, bioassay, drinking water, monitoring, uncertainty

Year of publishing
2020

More information
Dr. M.L. Dingemans
T
E Milou.Dingemans@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

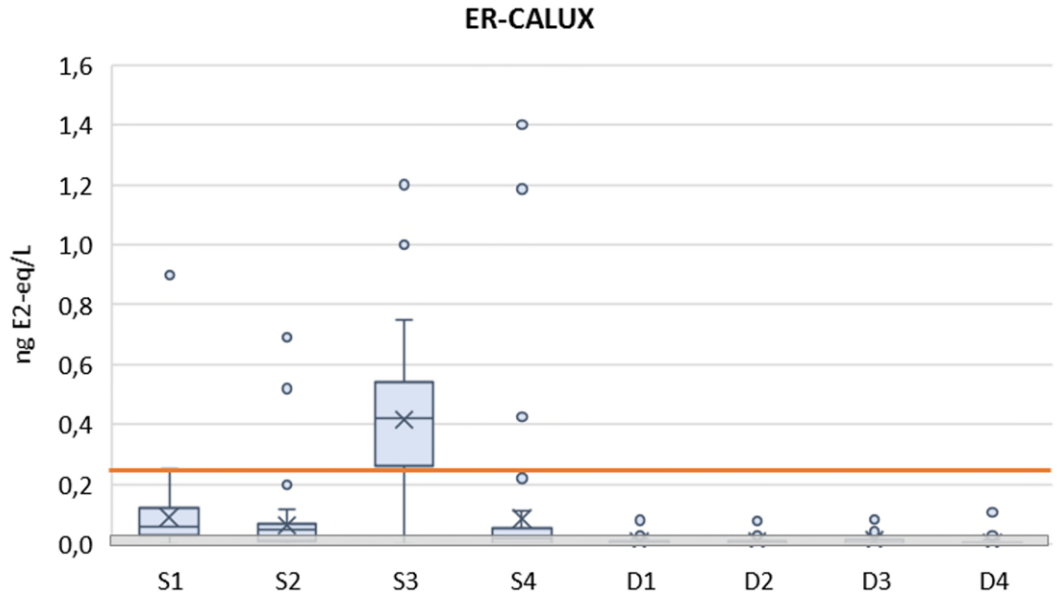
T +31 (0)30 60 69 511
F +31 (0)30 60 61 165
E info@kwrwater.nl
I www.kwrwater.nl

BTO *Managementsamenvatting*

Vier verschillende in-vitro CALUX bioassays geschikt bevonden voor het afleiden van effect-signaalwaarden

Auteur(s) Frederic Been PhD, dr.ir. Tessa Pronk, dr. Milou Dingemans

Van vier beschikbare Chemical Activated LUciferase gene eXpression (CALUX) *in vitro* bioassays is aangetoond dat deze geschikt zijn om bruikbare en betrouwbare effect-signaalwaarden (effect-based trigger value ofwel EBT) af te leiden. In dit onderzoek zijn EBTs afgeleid die zijn afgestemd op humane gezondheidsrisico's. CALUX-bioassays worden momenteel in Nederland toegepast als bioanalytische tool voor waterkwaliteitsbeoordeling. Overschrijding van de EBT geeft aan dat een mogelijke gezondheidsrisico bij drinkwaterconsumptie niet kan worden uitgesloten. Vervolgonderzoek naar de samenstelling van de verontreiniging is dan nodig om aan te tonen of daadwerkelijk sprake is van zo'n risico. Bioassays worden in toenemende mate ingezet als screeningmethode voor waterkwaliteit, omdat ze een integraal effect laten zien van complexe mengsels werkzame stoffen in water.



Vergelijking van de vastgestelde gezondheidskundige EBT (oranje lijn) met ER-CALUX waterkwaliteitsmonitoringsdata. S1-4 zijn drinkwaterbronnen, D1-4 is het bijbehorende drinkwater.

Belang: indicatie voor een mogelijk gezondheidsrisico in waterkwaliteit

Het afleiden van betrouwbare en bruikbare effectsignaalwaarden (EBTs) voor humane gezondheid is belangrijk omdat op basis van het al dan niet overschrijden hiervan in een bioassay-respons, bepaald kan worden of actie nodig is. Bij een te hoge EBT bestaat het gevaar dat gezondheidsrisico's onopgemerkt blijven. Is een EBT te laag, dan worden mogelijk onnodige verdere analyses gedaan om de oorzaak en het werkelijke risico te achterhalen. Dit is inefficiënt en niet kosteneffectief.

Aanpak: EBTs uit literatuurstudie vergeleken en onzekerheid vastgesteld met bioassay-responsen

In deze studie zijn verschillende methoden uit de literatuur voor het afleiden van EBTs met elkaar

vergeleken. Resultaten hiervan zijn gelegd naast bioassay-responsen in routinematige kwaliteitsmonitoringsdata. Vervolgens is de meest geschikte methode toegepast om daarmee in de praktijk bruikbare EBTs voor humane gezondheidsrisico's af te leiden. Daarnaast is voor deze EBTs een betrouwbaarheidsanalyse uitgevoerd, gebaseerd op het risico dat potentiele schadelijke stoffen niet worden opgemerkt. Dit is gedaan met behulp van extra meetgegevens van stoffen, getest in analoge ToxCast bioassays, met gezondheidskundige richtwaarden uit de literatuur en met een voor deze analyse ontwikkeld algoritme, gebaseerd op Monte Carlo simulatie.

Resultaten: afleiding van betrouwbare en bruikbare EBT voor waterkwaliteitsmonitoring

Voor de CALUX bioassays ER-, anti-AR-, AR- en GR-CALUX, zijn voldoende data beschikbaar om een betrouwbare en bruikbare EBT-waarde af te leiden. Deze EBTs zijn bruikbaar voor duiding van bioassay-metingen in de praktijk tijdens het monitoren van de waterkwaliteit van drinkwater(bronnen).

Implementatie: ontbreken van meetgegevens en gezondheidskundige richtwaarden van stoffen

De methodiek om de bruikbaarheid en betrouwbaarheid van EBTs af te leiden, is inzetbaar voor alle CALUX bioassays. Voor sommige bioassays (in dit project: PR-, PAH, NRF2- en p53-CALUX) is zo'n afleiding niet mogelijk omdat onvoldoende stof-specifieke gegevens beschikbaar zijn. Deze gegevens kunnen worden verkregen door in de bioassays meer water-relevante stoffen te testen. Daarnaast kunnen voor stoffen die actief zijn in CALUX bioassays voorlopige gezondheidskundige richtwaarden worden afgeleid.

Rapport

Dit onderzoek is beschreven in het rapport 'Development of a framework to derive effect-based trigger values to interpret CALUX data for drinking water quality' (BTO 2020.053).

Contents

1 Introduction

It is widely acknowledged that with targeted chemical monitoring, unknown chemicals, such as new emerging chemicals or transformation products, may go unnoticed (Altenburger et al., 2015; Brand et al., 2013; van der Oost et al., 2017b). Bioassays can be used as bioanalytical tools to detect such unknown (mixtures of) chemicals, as they give an integrated response to all known and unknown chemicals that induce a response (Escher and Leusch, 2011). An increased bioassay response while acceptable concentrations of target chemicals are measured with chemical analysis may indicate that other bioactive chemicals causing this effect are present. Despite this clear added value of using bioassays in water quality screening, the interpretation of bioassay results remains a challenge. Bioassays can be very sensitive, i.e. already give a positive response at low concentrations. Therefore, a positive bioassay response does not necessarily mean an increased risk. For several bioassays used in water quality monitoring, response levels that provide a trigger to further assess potential adverse effects for humans or the environment are proposed in the scientific literature. For reporter gene assays, such thresholds, also known as effect-based trigger (EBT) values (Escher et al., 2015), are expressed in equivalents of a reference chemical in the bioassay. At responses below the EBT value, adverse effects for humans or the environment are not expected, while possible adverse effects cannot be excluded at higher responses, so a further in-depth assessment is needed.

In the practice of water quality management, EBT values need to be sufficiently conservative to be useful as indicators of hazardous bioactive micropollutants in water but also not overly conservative, preventing that assets are spent on unnecessary follow-up actions (Brand et al., 2013). EBT values of *in vitro* bioassays are based on the biological equivalent (BEQ) concept (Escher et al., 2015; van den Berg et al., 2006). This concept assumes that if chemicals in a mixture that elicits a response in a specific bioassay share the same mode of action, their combined effects can be described by the concentration addition concept. A BEQ concentration of an unknown mixture of chemicals can be measured in a bioassay, expressed as reference compound equivalents by comparing the response in the bioassay to the concentration-response curve of the reference chemical in that same bioassay.

Endpoint- and bioassay specific EBT values for responses of water samples are commonly established based on relative effect potencies (REP) compared to a reference compound, of a range of compounds known or expected to have the same mode of action as captured in the bioassay. In the scientific literature, different approaches have been proposed to derive effect-based trigger (EBT) values, for different purposes (Escher et al., 2013, 2015, 2018; Tang et al., 2013; Van der Oost et al., 2017; Brand et al., 2013). EBT values for the protection of ecosystems can be based upon a species-sensitivity distribution of BEQ that cause adverse effects to aquatic organisms (van der Oost et al., 2017a, 2017b). In other studies, human health-based EBT values have been derived based on safe intake levels for humans based on toxicity data (Brand et al., 2013). In other studies EBT values are based on environmental or drinking water guideline values (Escher et al., 2018, 2015, 2013; Tang et al., 2013).

In the present study we aim to derive EBT values to detect potential risks to drinking water safety for eight different Chemical Activated LUciferase gene eXpression (CALUX) *in vitro* bioassays that are currently being used in The Netherlands as bioanalytical tools for water quality assessment (van der Oost et al., 2017b) (Table 1). Bioassays selected for this purpose are a series of nuclear receptor activation and stress-response bioassays. To derive EBT, an approach similar to the one suggested by Escher et al. (2015) (Escher et al., 2015) was used, which considers provisional health-based guidance values in drinking water (pGLV) as a point of departure for EBT derivation, by evaluating the effect concentration (EC50 or lowest observed effect concentrations (LOEC) in the bioassay) of selected chemicals. We first (1) carried out a sensitivity analysis of the impact of different selection criteria (i.e., used to select relevant compounds for EBT derivation) and calculation approaches (i.e., used to mathematically derive the EBT) on calculated EBT values. Using the most appropriate combination of selection criteria and

derivation approach, (2) we calculated EBT values to detect potential risks to drinking water safety for bioassays listed in Table 1. Finally, (3) we introduced a method to quantify the uncertainty surrounding the derived EBT values, given that only for a fraction of compounds that can reasonably be expected to trigger a response in the assay (based on known effects in analogue *in vitro* tests) REP values are available to derive EBT values.

Table 1: Overview of CALUX bioassays considered in this work.

Bioassay	Endpoint	Reference compound
ERα-CALUX	Activation of the estrogen receptor	17β-estradiol (E2)
AR-CALUX	Activation of the androgen receptor	Dihydrotestosterone (DHT)
anti-AR-CALUX	Inhibition of the androgen receptor	Flutamide (Flut)
PR-CALUX	Activation of the progesterone receptor	Progesterone (P4)
GR-CALUX	Glucocorticoid receptor-mediated signalling	Dexamethasone (DEX)
PAH-CALUX	Activation of the aryl hydrocarbon receptor	Benzo[a]pyrene (BaP)
Nrf2-CALUX	Activation of the Nfr2-mediated stress response related to electrophilic and oxidative stress	Curcumin (Cur)
p53-CALUX	Activation of the p53-mediated stress response indicative of genotoxicity	2-acetylaminofluorene (2-AAF)

2 Materials and methods

2.1 Calculation of pGLV-BEQ

In vitro potencies of individual chemicals in the bioassays were collected from literature, allowing the determination of REP values (see Supporting Information). These were based on concentrations eliciting a response equal to 50% of the maximum response of the reference compounds (EC_{50}). For stress-response assays (i.e., NRF2- and p53-CALUX), no EC_{50} are available, and lowest-observed effect concentrations (LOECs) were used. Reference compounds used are reported in Table 1. The REP for each compound *i* was calculated as

$$REP_i = EC_{50ref}/EC_{50i} \qquad \text{Equation 1}$$

In the second step, safe chemical concentrations in drinking water, represented by provisional health-based guidance values in drinking water (pGLV), were calculated based on reported safe intake levels (see Supporting Information). Preferably, safe intake levels as derived by competent authorities were used, but if these were not available, data from (sub-)chronic animal toxicity studies were used to derive safe intake levels (Baken et al., 2018). In that case, safe intake levels (ADI or TDI) were obtained by dividing no-observed adverse effect levels (NOAELs) by 100 (to account for intra- and interspecies differences in toxicokinetics and toxicodynamics) (Baars et al., 2001). If only data from subacute studies were available, an extra uncertainty factor of 3 was applied. For pharmaceuticals, safe intake levels were derived by dividing therapeutic doses [defined daily dosis (DDD)] by an uncertainty factor of 100 (Houtman et al., 2014; Schriks et al., 2010). From these safe intake levels, pGLV are calculated assuming a bodyweight of 70 kg, an allocation factor to water of 20% and a drinking water consumption of 2 L/day (Equation 2; Baken et al. 2018):

$$pGLV\ (\mu g/L) = \frac{\text{ADI}\ (\mu g/kg\ bw/day) * 70\ (\text{kg}) * 20\%}{2\ (L/day)} \qquad \text{Equation 2}$$

In this equation, *pGLV* is the provisional guideline value, *ADI* is the Acceptable Daily Intake (or Tolerable Daily Intake for unavoidable chemicals). For genotoxic chemicals, it is assumed that no safety thresholds exist and that each exposure increases the risk of adverse health effects. Therefore, an *acceptable* intake level is calculated which is the dose that is associated with an increase of an extra cancer incidence of 1 in a million individuals upon a lifetime exposure (Fewtrell and Bartram, 2001). This intake level associated with an acceptable risk can be determined using data on dose-dependent cancer formation in laboratory animals. In the present study, LTD_{10} values reported in the Carcinogenic Potency Database (CPDB 2019) were used, which are the lower confidences of TD_{10} values, daily doses that cause an increase in cancer incidences of 10% in laboratory animals. Using linear extrapolation, doses that cause a 1 in a million cancer incidences (0.0001%) were calculated, which were used to derive pGLV for genotoxic carcinogens, assuming a bodyweight of 70 kg and a drinking water consumption of 2 L/day (Equation 3; (van der Aa et al., 2017)):

$$pGLV\ (\mu g/L) = \frac{\text{dose with acceptable risk}\ (\mu g/kg\ bw/day) * 70\ (\text{kg})}{2\ (L/day)} \qquad \text{Equation 3}$$

In a third step, by multiplying the obtained pGLV by their REP values, pGLV expressed in reference compound equivalents (µg/L) were obtained (pGLV-BEQ):

$$pGLV - BEQ = pGLV * REP \qquad \text{Equation 4}$$

2.2 Calculation of EBT values

Prior to calculation of EBT values, six criteria were considered to select the compounds from which EBT values were derived. These criteria were defined because not all the compounds for which CALUX-specific data are available are relevant to establish EBT values that can be used in routine water monitoring. These criteria take into account whether the bioassay contributes to the detection of a substance (criterium 2), whether bioactivity can be reasonably expected at concentrations found in drinking water sources (criterium 3), whether internal (blood) concentration are expected to increase by oral intake (criterium 4), and discrepancies between potencies *in vitro* and in humans (criterium 5 and 6). Alternatively, no selection was made and all available data was included (criterium 1). Furthermore, five mathematical approaches (A to E) to derive EBT values were evaluated: A) 5^{th} percentile of normal distribution of pGLV-BEQs; B) 5^{th} percentile of pGLV-BEQs; C: linear regression of pGLV-BEQs; D) quantile regression of pGLV-BEQs; E) minimal pGLV-BEQ). These were applied for each of the six selection criteria (Supplemental Figure S1). The criteria and mathematical approaches tested in this study and the findings are described in detail in the Supporting Information. Based on the results, the EBT presented here are computed by taking the 5th percentile of the log-normal cumulative distribution function of the available pGLV values (approach A) of substances for which 0.1 < pGLVi/EC50i (or LOECi) were selected (criterion 6).

2.3 Simulations and uncertainty analysis

Due to the limited number of pGLV-BEQ values available for CALUX bioassays, an algorithm based on Monte Carlo simulations was developed to obtain a profile of the variability in derived EBT values and their protective power as a function of the number of compounds available. In other terms, the goal of these simulations is to estimate the probability that EBT values derived from a limited number of chemicals will be protective (i.e., sufficiently low to detect the presence of potentially harmful (mixtures of) chemical(s)), also when considering a larger range of chemicals. The procedure used is detailed below.

Step 1: Obtaining EC50 values for compounds in analogous bioassays

EC50 values for all chemicals (> 9000) and biological endpoints (> 1000) listed in the U.S. EPA's ToxCast (United States Environmental Protection Agency, 2016) were downloaded from the ToxCast data repository (United States Environmental Protection Agency, 2018). Subsequently, only EC50 values measured in bioassays considered as analogues of the CALUX bioassays (i.e., same biological endpoint, see Supplemental Table S1) in this study were used, also including the EC50 of the reference compound. Inactive compounds (i.e., listed in the ToxCast data repository as having an EC50 of 10^6 µM) were removed. The output of this selection consisted in a large matrix of EC50 values that have been measured in bioassays analogues to the CALUX bioassays. It should be noted that analogue bioassays (and EC50 values of tested compounds) in the ToxCast data repository were retrieved only for ERα-, AR-, anti-AR-, GR-, NRF2-, PAH- and p53-CALUX bioassays.

Step 2: Obtaining pGLV for compounds in analogue bioassays

pGLVs of a broad range of water relevant chemicals (not limited to compounds that have been tested with CALUX bioassays) were collected from the literature (Baken et al., 2018). The obtained dataset was collated to the available data collected for active chemicals in CALUX bioassays, yielding a dataset of more than 160 pGLV for water relevant compounds.

Step 3: REP and pGLV-BEQ calculation

REP were then calculated from the EC50 values reported for CALUX-analogue assays (Equation 1) and then used to calculate pGLV-BEQ (Equation 4).

Step 4: Merge and select

Compounds within the same group of analogue assays (e.g., all compounds tested in assays analogues to ERα-CALUX) were merged into a single list, yielding a dataset of compounds with pGLV-BEQ for each of the CALUX assays for which analogue assays were available. Compounds from the CALUX assay itself were also included in this list. Subsequently, the selection criterion defined above was applied, i.e. only compounds with a pGLV/EC50 ratio above 0.1 were retained. The number of compounds available in each dataset before and after applying the selection criterion are shown in Table S2.

Step 5: Simulations

Per dataset of CALUX analogue bioassays, it was verified that the obtained pGLV-BEQ values could be approximated using a log-normal distribution. Subsequently, the following steps were done:

(i) for each dataset, a certain number of pGLV-BEQ were drawn randomly. The number of pGLV-BEQ to be drawn depends on the number of compounds in the corresponding CALUX assay (see Table 2) which remain available after applying the selection criterion. For instance, if for compounds available for ERα-CALUX 10 compounds are relevant to derive an EBT value after applying the selection criterion (i.e., pGLV/EC50 > 0.1), then 10 compounds are randomly sampled from the compounds in the dataset of ERα-CALUX analogue assays (see Step 4).
(ii) The selected pGLV-BEQ are then log-transformed and the corresponding mean and variance are computed.
(iii) A normal distribution and its 5^{th} percentile, used as a theoretical EBT value, are then computed.
(iv) The probability of obtaining a value above the theoretical EBT value for any of the compounds from the analogues dataset is then computed using the derived 5^{th} percentile and the log-transformed cumulative distribution function of all *n* compounds in the dataset. This yields the proportion of compounds in the analogous bioassays which have a pGLV-BEQ greater or equal to the derived theoretical EBT value and are covered in terms of risk with this EBT.
(v) Points (i) to (iv) are repeated iteratively 10000 times and the outcomes are plotted as histograms and fitted to a Beta distribution with shape parameters α and β. The 95% confidence intervals were computed from the Beta distribution. All calculations and simulations were carried out using RStudio (RStudio Team, 2016).

2.4 Empirical data from water quality monitoring in The Netherlands

Water quality monitoring data was collected by The Water Laboratory (HWL) in Haarlem, The Netherlands, where a number of *in vitro* CALUX bioassays are implemented and presently used for water quality monitoring of sources of and produced drinking waters of three drinking water production plants in the Netherlands. EBT values derived in the context of this study were compared with responses found in real water samples. In particular, responses found for ERα-, GR, and Anti-AR-CALUX in the period 2013-2017 were used for comparison. Surface water grab samples and corresponding drinking water samples (D1-D4) were collected at four abstraction points used by Dutch drinking water companies at the rivers Meuse (Enclosed Meuse) (source 1; S1) and Rhine (Lek Channel) (S2), a reclaimed land area Bethune Polder (S3), Lake IJssel (S4). Details about the collected surface and drinking water samples and extraction protocols are provided in the Supplementary Information.

3 Results and discussion

3.1 pGLV-REP values of chemicals in bioassays

REP values were determined as described in the methods section. For each of the bioassays, Figure 1 shows the relationships between REP, log pGLV and log pGLV-BEQ values for individual substances. For chemicals tested in the bioassays, pGLV values were collected or calculated based on reported safe intake levels or from safe intake levels determined in the present study, as described in materials and methods.

The pGLV values were multiplied with their REP values to obtain pGLV-BEQ (expressed in reference chemical equivalents). The data underlying the derivation of the pGLV and pGLV-BEQ values are presented in the Supplementary Information. Most data points (substances) were available for the anti-AR-CALUX assay, followed by ERα-CALUX, AR-CALUX and p53-CALUX assays. For the PR-CALUX and PAH-CALUX assays only a few data points (substances) are available. The figure also shows that for each bioassay, the majority of the REP values are lower than 1 (i.e., $\log_{10}$ REP < 0), especially for ERα-CALUX. This means that generally the reference compound has a high potency in the bioassay, relative to other compounds. However, this is not necessarily the case for the toxicity of these chemicals in humans, as the pGLV of the reference compound is not the lowest in all assays.

The symbol colour per substance indicates the value of the pGLV-REP, with low values shown in blue and high values shown in red. To minimize the risk of not detecting potentially harmful chemicals the EBT value needs to be sufficiently low to detect all these substances. On the other hand, an EBT value should not be too low to avoid unnecessary follow-up analyses. To prevent this, some of the substances were excluded from the derivation of the EBT if they did not fulfil the selection criterion, as explained previously. These compounds are indicated in Figure 1 by a black circle (these are also reported as cumulative distributions in Figure S2). A line is included in the plots to represent a level of 100 times the signalling value of 0.1 µg/L to identify which substances may pose a risk at water concentrations not detected using conventional chemical analyses (if below this line).

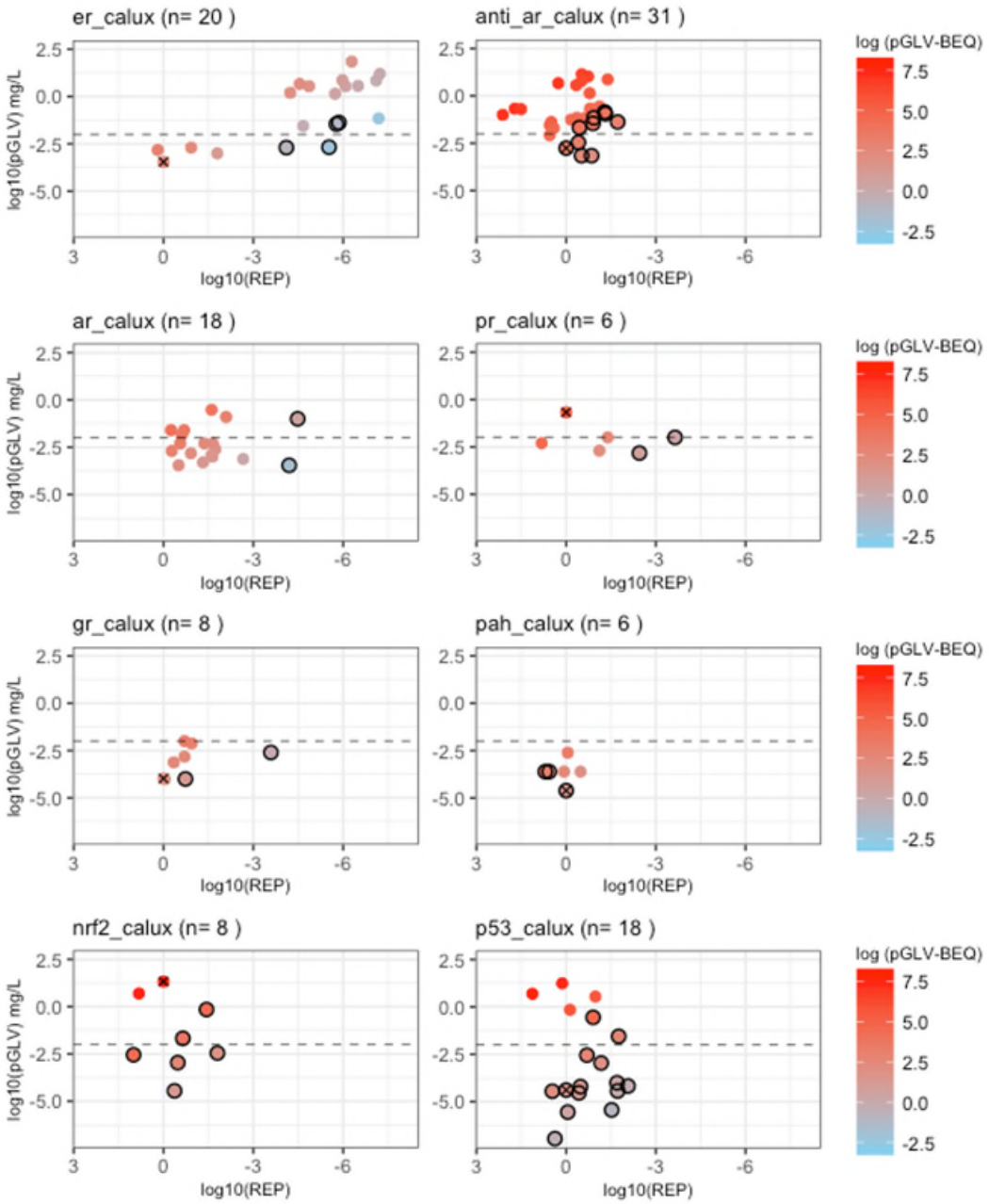


Figure 1: Plots of REP, pGLV and pGLV-BEQ values obtained for substances detected by different bioassays. The data were log10 transformed prior to plotting. The x-axis (i.e., Log10(REP)) was inverted so that high-risk and high-potency chemicals cluster close to the origin, while low-risk and low-potency chemicals cluster in the upper right corner. The dotted line corresponds to a pGLV value of 0.01 mg/L (or -2 in log10 scale), which is equal to 100 times the signal value (0.1 µg/L) (International Association of Waterworks in the Rhine Basin, 2020; van der Aa et al., 2017). n=number of substances. Dots marked with a cross indicate the reference chemical. Dots marked with a black circle did not fulfil the selection criterion (6) as defined previously and were excluded from the EBT calculation. Note: A pGLV value was not available for the reference compound used in the AR-CALUX, namely dihydrotestosterone, which is thus not included in this figure.

3.2 Derivation of EBT values from pGLV-BEQ

EBT values were derived from pGLV-BEQ based on the full dataset or a selection based on criterion (6) and using estimation method (A) detailed previously (Table 2).

Table 2: Obtained EBT values for the considered CALUX bioassays using the entire dataset or using only compounds which fulfil the selection criterion.

Bioassay	Reference compound	EBT value [ng ref-EQ/L] pGLV/EC50 > 0.1	EBT value [ng ref-EQ/L] Full dataset
ER-CALUX	17ß-estradiol (E2)	*0.25 (n = 16)	8.3E-03 (n = 20)
Anti-AR CALUX	Flutamide (Flut)	4800 (n = 21)	2.7E+02 (n = 31)
AR-CALUX	Dihydrotestosterone (DHT)	4.5 (n = 16)	5.1E-01 (n = 19)
PR-CALUX	Progesterone (P4)	15.5** (n = 4)	2.2E-01 (n = 6)
GR-CALUX	Dexamethasone (DEX)	47.9** (n = 6)	1.7E+00 (n = 8)
PAH-CALUX	Benzo[a]pyrene (BaP)	24.4** (n = 3)	1.9E+01 (n = 6)
NRF2-CALUX	Curcumin (Cur)	***NA (n = 2)	NA (n = 8)
P53-CALUX	2-Acetamimofluorene (2-AAF)	***NA (n = 4)	NA (n = 18)

**This value was obtained after removing one compound (i.e., prothioconazole) which had very low pGLV-BEQ value; ** Preliminary value due to lack of underlying data; ***With the current methodology, and in particular the (limited) data available, it was not possible to estimate an EBT value for these bioassays.*

3.2.1 ERα-CALUX

After selection using the previously described criterion, 16 out 20 compounds remained (see Figure 1) to derive the EBT value which amounts to 0.05 ng E2 eq./L. This value is lower compared to other EBT values reported in the literature. For instance, Escher et al. (2018) (Escher et al., 2018) reported an EBT value of 0.1 ng E2 eq./L. Van der Oost et al. (2017) (van der Oost et al., 2017b) obtained an EBT of 0.5 ng E2 eq./L for this bioassay, but it should be mentioned that these authors selected compounds having a REP > 0.001 and focused on ecological risks rather than effects on humans. On the other hand, Brand et al. (2013) (Brand et al., 2013) derived a higher value, namely 3.8 ng E2 eq./L. The difference between the EBT values obtained in the present study and those reported in the literature is likely due to the inclusion of one compound (i.e., prothioconazole) at the lower end of the distribution (as shown in Figure S2), which has a low pGLV-BEQ value. Although it did fulfill the selection criterion, it was decided to remove this low potency compound from the selection. Without this compound, the obtained EBT value amounts to 0.25 ng E2 eq./L, which is in the same range as those from the literature.

3.2.2 Anti-AR-CALUX

Out of the 31 compounds available for anti-AR-CALUX, 21 remained after applying the selection criterion (see Figure 1) and the obtained EBT value was equal to 4.8 µg Flut eq./L. This is in line with the EBT values suggested by Escher et al. (2018) (Escher et al., 2018) and Van der Oost et al. (2017) (van der Oost et al., 2017b) of 14.4 and 25 µg Flut eq./L for this bioassay, respectively, although both these EBT are higher as compared to the figure obtained here. To allow practical use of the EBT both authors used a correction of the lower values they obtained with their regular approach. Escher et al. (2018) used a mixture factor of 100x to account for the potential presence of unknown substances in surface waters that may have a response in this bioassay. EBT in this study focus on

ecological risks but are often considered to be protective also for human health effects based on the assumption that environmental quality standards (EQS) are often lower than pGLVs. If this assumption holds up also with regard to mixture effects requires further study. Van der Oost et al. (2017) focussed on ecological risks and used the background BEQ, consisting of the average BEQ measured at sites with a good ecological situation, multiplied by a factor ranging from 2 to 5. .

3.2.3 AR-CALUX

In the specific case of the AR-CALUX, after selection of relevant compounds, 16 out of the 18 initial chemicals (see Figure 1) were still retained to derive an EBT value, which amounts to 4.5 ng DHT eq./L. This value is in the same order of magnitude, as the EBT value of 11 ng DHT eq./L derived by Brand et al. (2013). Escher et al. (Escher et al., 2015) derived an EBT value of 14 ng testosterone equivalents/L for the AR-CALUX assay, which corresponds to 3.4 ng DHT eq./L, quite similar to the one derived in the present study.

3.2.4 PR-CALUX

For the PR-CALUX, only 4 of the 6 compounds initially available were left after excluding chemicals based on the defined criterion and large differences in pGLV-BEQ values were found between compounds (as shown in Figure 1). Based on remaining compounds, an EBT value of 15.5 ng P4 eq./L was calculated. On the other hand, using the entire dataset, an EBT value of 0.223 ng P4 eq./L would have been obtained, which is much lower than EBT values reported in the literature. Brand et al. 2013 (Brand et al., 2013) calculated an EBT value which was substantially higher, namely 333 ng Org2058-eq/L, which corresponds to 4.5 µg P4 eq./L. As discussed previously, this might be due to the different EBT approach. Nevertheless, due to the limited number of chemicals available and the lack of other data from the literature, the EBT value derived here should be only considered as a preliminary indication.

3.2.5 GR-CALUX

After selection of relevant compounds, a total of 6 out of 8 chemicals initially available was used to derive the EBT value of 47.9 ng DEX eq./L. Due to the limited number of available compounds, the EBT value obtained here should be considered only as a preliminary indication. When compared to results from the literature, the obtained EBT is in the same order of magnitude as the EBT value of 21 ng DEX eq./L derived by Brand et al. 2013 (Brand et al., 2013). Although focused on water quality from an ecotoxicology perspective, Van der Oost et al. (2017) reported a slightly higher EBT of 100 ng DEX eq./L. Whilst the EBT value derived here does not deviate substantially from results reported in the literature, analysis of a broader range of compounds is advisable to corroborate the reliability of the obtained result.

3.2.6 PAH-CALUX

Only 3 chemicals out of the 6 initially fulfilled the selection criterion. The EBT value was estimated to 24.4 ng BaP eq./L which lies within the range of EBT values reported in the literature for surface water, namely 6.2 ng BaP eq./L (Escher et al., 2018) and 150 ng BaP eq./L (van der Oost et al., 2017b). Inclusion of all available chemicals did not substantially modify the derived EBT value (19.3 ng BaP eq./L). As for PR- and GR-CALUX, the obtained EBT value should only be viewed as a preliminary threshold due to the limited number of compounds available to derive it. Additional chemicals should be measured to obtain a larger dataset to derive a reliable EBT for routine monitoring of drinking water quality.

3.2.7 NRF2-CALUX

In the specific case of the NRF2-CALUX, only two out of 8 compounds fulfilled the requirements of the selection criterion and an EBT of 16 mg Cur eq./L was derived, whilst when using all available chemicals, an EBT value of 1.7 ng Cur eq./L was obtained. The large difference is due to the broad range of pGLV-BEQ values, as shown in Figure 1. Van der Oost et al. (2017) (van der Oost et al., 2017b) derived a trigger value for the Nrf2-CALUX equal to 10 µg Cur eq./L, whereas the trigger value derived by Escher et al (2018) (Escher et al., 2018) derived for this bioassay amounts to 26 µg dichlorvos eq./L, which corresponds to 42 µg Cur eq./L. As discussed previously, Van der Oost et

al. (2017) (van der Oost et al., 2017b) focused on ecological risks and used background BEQs to derive an EBT value for oxidative stress due to the effects which are readily measured at sites with good ecological status. Compared to results from actual measurements, the obtained EBT value is substantially higher compared to median responses obtained in the Netherlands for raw water used for drinking water production, namely 2.8 µg Cur eq./L (personal communication The Water Laboratorium, Haarlem, The Netherlands). The limited amount of data available hinders the derivation of reliable EBT values for the NRF2-CALUX and highlight the need to measure a broader range of compounds before a robust threshold can be derived for routine applications. As many chemicals activate oxidative stress bioassays, their use to detect specific chemicals or to track removal of chemicals more generally needs further study.

3.2.8 p53-CALUX

Out of the 18 compounds initially available for this bioassay, only 4 fulfilled the selection criterion and an EBT value of 54 µg 2-AAF eq./L was estimated, whilst an EBT value of 8 pg 2-AAF eq./L was obtained using the full dataset. This substantial difference is due to the fact that most of the excluded chemicals actually lie in the lower part of the distribution, as shown in Figure 2. These chemicals were excluded due to their low activity in this particular bioassay. Further, it must be noted that besides genotoxic chemicals, also a large set of non-genotoxic chemicals are active in this bioassay. Current guidelines for genotoxic substances that may be potentially carcinogenic assume that there is no safe level. The chances of adverse effects decrease at lower exposure levels, but a theoretical contribution to a health risk always remains. Since low-risk levels of most genotoxic chemicals are much lower than safe levels of non-genotoxic chemicals, the distribution of pGLV-BEQ spans a too large range to derive useful EBT values for all these chemicals. No other studies have derived EBT values for the p53-CALUX or other related stress-response bioassays so far. Similar to the previous bioassay, the limited number of compounds available hinders the derivation of an EBT value. Additional substances need to be tested in this and other genotoxicity bioassays before a reliable threshold can be derived that can be applied in routine.

3.3 Simulations to estimate the probability that EBT values derived from a limited number of chemicals will be protective

In the previous section, EBT values were determined for a set of CALUX bioassays based on available pGLV-BEQ. In an ideal situation, one would derive EBT values from a very large set of pGLV-BEQ (i.e., from a very large number of substances that i) have been tested in the considered bioassays and ii) for which enough toxicity information is available to derive a pGLV). However, this is often not the case and EBT values are derived using data from a limited number of substances. This introduces uncertainty in EBT values because, depending on the representativeness of the available pGLV-BEQ values (i.e., whether they are a representative sample of the population of compounds, which could be encountered in practice), derived EBT values might not be adequate to assess the quality of water intended for drinking water production.

In an attempt to quantify the uncertainty surrounding EBT values estimated using only a limited number of compounds, an algorithm was developed which uses data derived from the literature and the ToxCast data repository (United States Environmental Protection Agency, 2018). As highlighted previously, the goal of the developed algorithm is not to compute actual EBT values to be used in practice, but rather to assess the risk that EBT values derived from a limited number of pGLV-BEQ might not be able to effectively detect the presence of a potentially harmful compound and, as a consequence, be unreliable. In other terms, the goal of the algorithm is to determine the proportion of relevant chemicals giving a response in the bioassay above the established EBT value.

Results of the simulations are reported in Figure 2. In the case of ERα-, anti-AR and AR-CALUX analogues, the obtained 95%-confidence interval (95%-CI) suggest that, with the available number of chemicals, the derived EBTs would be able to detect >83% or more of potentially harmful chemicals. In the case of GR- and PAH-CALUX

analogues, it was not possible to obtain substantially more data from the ToxCast data repository (i.e., only 8 and 9 compounds, respectively) compared to the number of compounds for which CALUX-specific data were available (i.e., 6, see Table 1). Consequently, we cannot exclude that the estimated proportion of pGLV-BEQ ≥ theoretical EBT is overly optimistic and that the true 95%-CI is actually broader than what was obtained here. For Nrf2- and p53-CALUX analogues, more relevant data could be recovered from the ToxCast data repository, namely 22 and 25 pGLV-BEQ versus only 2 and 4 for the CALUX assays (in both cases after applying the selection criterion), respectively. When observing the 95%-CI, these were much broader, indicating that the derived EBT values will likely not be sufficiently low to detect potentially harmful chemicals.

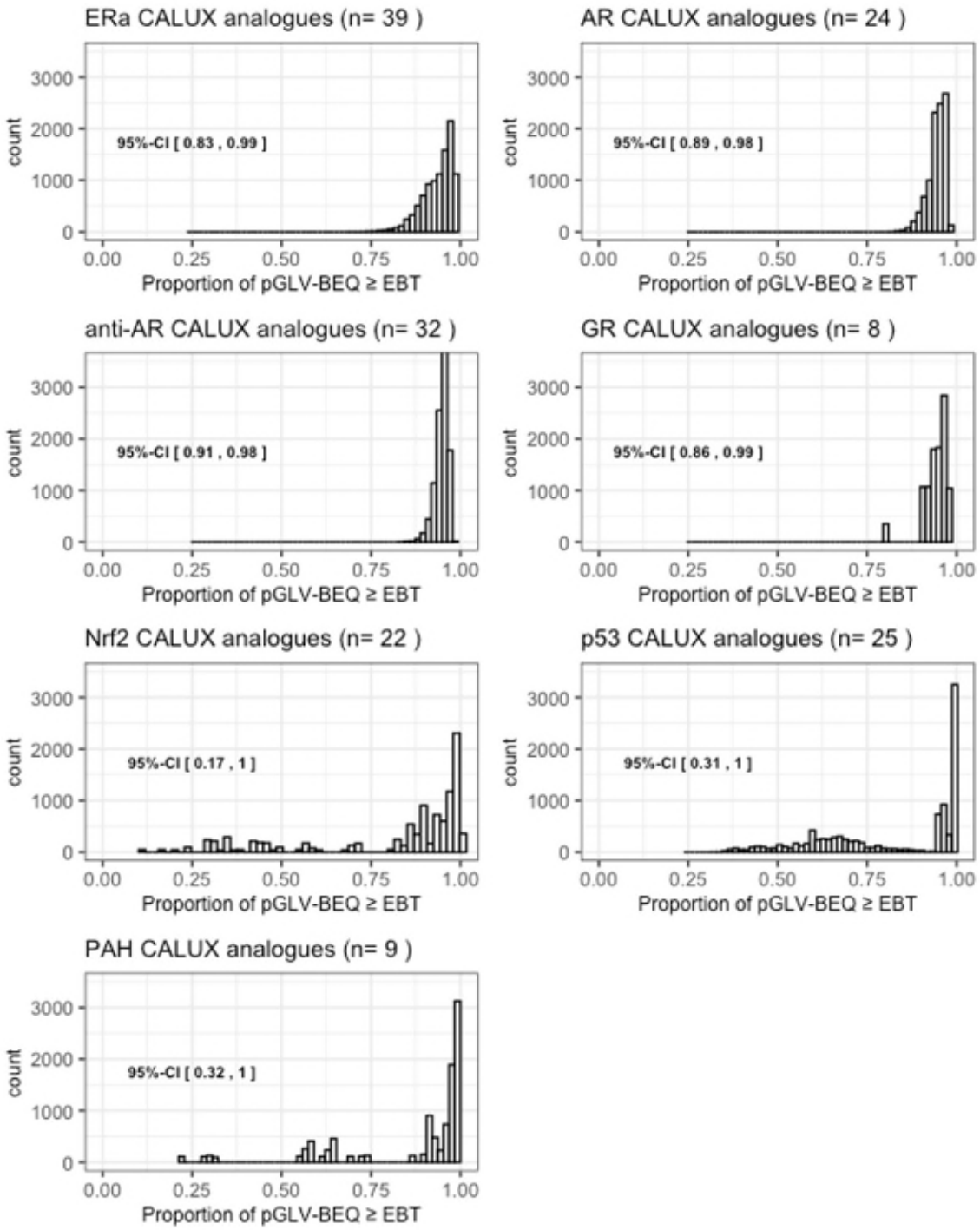


Figure 2: Distribution of the proportion of substances having a pGLV-BEQ ≥ theoretical EBT value in the CALUX analogue assays retrieved from the ToxCast data repository (United States Environmental Protection Agency, 2018). n= number of compounds from the ToxCast data repository (United States Environmental Protection Agency, 2018) which were available for the simulations; Above 95% = Proportion of EBT values which will cover 95% or more of the compounds; 95%-CI = 95% confidence interval computed from a fitted Beta distribution.

Overall, these findings illustrate that if the number of substances relevant for the bioassays available in ToxCast is sufficiently large compared to the total number of potential (unknown) compounds which might be encountered in practice, EBT values that are expected to be protective, can be derived. This was the case for ERα-, anti-AR and AR-CALUX analogues. It should however be emphasized that the difference between the size of the total population is only roughly twice the size of the number of values drawn (e.g., 16 compounds drawn versus a total population of 39 for ERα-CALUX analogues). This population can be expected to be only a small proportion of all substances (with

more or less of this activity) that may be present in water. Consequently, it is possible that the obtained 95%-CI are overoptimistic as the same compounds are more likely to be drawn from smaller populations. On the other hand, if the number of relevant substances available increases, the more likely it is that the derived EBT value will be protective (in particular if it is derived as the 5th percentile of the associated normal distribution).

These results, however, also highlight that it difficult to derive a reliable EBT value when only few compounds are available, as is the case for NRF2- and p53-CALUX. Defining a threshold below which the number of available compounds is deemed too low to derive a reliable EBT value is also difficult, as it depends on the variability of responses given by chemicals in a specific assay (i.e., distribution of REP). For instance, by using 10 or 5 compounds instead of 16 for the ERα-CALUX analogue assays, the 95%-CI interval would broaden to 76-100% or 61-100%, respectively. To have a 95% chance of covering at least 80% of all compounds one would need at least 12 compounds. Likewise, for assays analogue to NRF2-CALUX, going from 2 to 8 available compounds to derive an EBT value increases the 95%-CI to 82-99% (see Figure 3).

Thus, in situations where EBT values are derived using a low number of relevant compounds (i.e., less than 8 or 10 according to the examples shown above), which do not cover a representative range of REP, it would be advisable to introduce a precautionary factor. However, the most appropriate approach to overcome this situation is to collect more experimental data for water relevant substances. For instance, by determining REP for a broader range of compounds for the assay in question or, derive pGLV for a larger set of compounds. In fact, chemicals have often been tested with bioassays (i.e., EC50 have been calculated and REP can hence be derived), however there is a lack of information about guideline values for drinking water (i.e., pGLV have not been determined). Enlarging the dataset provides the empirical basis necessary to derive reliable EBT values which can be implemented in regular drinking water quality monitoring.

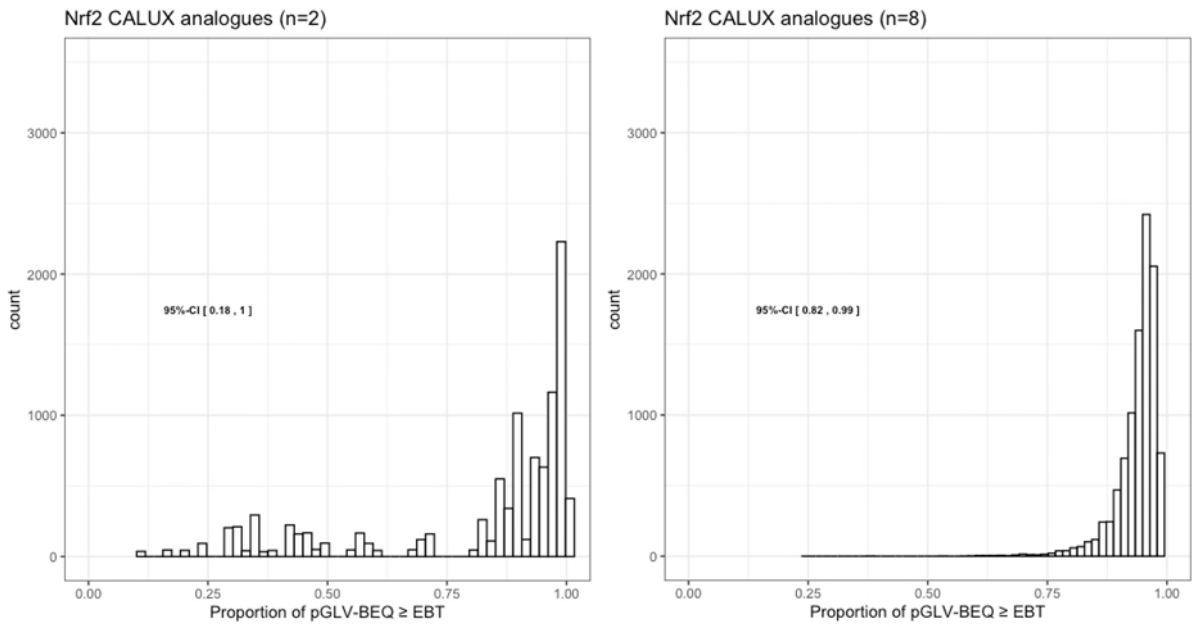


Figure 2: Comparison between EBT values obtained for NRF2-CALUX using 2 (left) or 8 (right) compounds out of the total 22 available.

3.4 Overview of the factors of influence for deriving the EBT value

With regard to the selection criteria used to include or exclude compounds and the methods used to derive EBT values, results show that the latter have only a secondary influence on the outcomes. Nevertheless, based on the assumption that pGLV-BEQ values follow a normal distribution, we suggested to use the 5th percentile of a normal

distribution as derivation method. The most influential factor in the value of the obtained EBT value is the selection criterion of data inclusion. In this study we bring forward various arguments to support the use of the criterion suggested by Escher et al. (2015), yet without including the upper boundary. This because we consider that highly responsive compounds (i.e., whose pGLV-BEQ to EC50 ratio is above 10) are relevant and there is no obvious reason why these should be excluded. On the other hand, non-responsive compounds should be excluded because they are likely not relevant for the endpoint considered by the specific bioassay.

Using the abovementioned approach, EBT values were derived for CALUX assays. Based on the size of the available datasets for each bioassay and the outcome of the simulations, EBT values which can be implemented in practice could be obtained for ERα-, anti-AR-, AR- and GR-CALUX. This was not the case for PR-, PAH-, NRF2- and p53-CALUX because either the number of compounds available was too small, and the uncertainty obtained from the simulations was too large, or the differences with EBT values from the literature were too large. For these bioassays, additional studies, in particular the collection of a broader dataset (i.e., measure more compounds and increase the number of available pGLV-BEQ) is needed before reliable EBT values can be obtained. For non-specific bioassays such as Nrf2-CALUX, other implementations, e.g. to monitor the removal of bioactive chemicals rather than to detect the presence of potentially risks, are also envisaged.

Finally, it should be stressed that no unique algorithm is able to derive solid trigger values for all bioassays, since unrealistically low EBT values can always be obtained for certain bioassays. Incorporation of precautionary correction factors based on actual measurements in the water cycle can be considered to increase EBTs if these are unrealistically low. For instance, in other studies assessment factors (Tang et al., 2013), background responses (van der Oost et al., 2017a) or mixture toxicity factors (Escher et al., 2018) were used to adjust derived EBT values.

3.5 Implications of the derived EBT-values for drinking water

Useful EBT should indicate potential chemical risks. Yet, at the same time, they should avoid giving (too many) false positives. The EBT values obtained in this study were compared to empirical data from routine monitoring of drinking water sources in The Netherlands to estimate their applicability in practice (see Supporting Information for more details). Figure 4 shows the boxplots with the measured bioassay activities in the drinking water sources S1-S4 and the corresponding drinking waters D1-D4.

For estrogenic activity as measured with the ER-CALUX, the EBT value is 7 times higher compared to the limit of quantification (LOQ) of 0.034 ng E2-eq/L. In the untreated sources estrogenic activity is regularly detected with median concentrations in the range of 0.051-0.35 ng E2-eq/L in the period 2013-2017 (Table S1). S3 has a relative high exceedance rate of the EBT value, which is notable, since it has the lowest anthropogenic influence of the included drinking water sources (as seen with the routine monitoring of chemicals). An effect directed analysis research has been carried out that tentatively identified E2 as the responsible compound causing this activity (in preparation, data not shown). E2 possibly enters S3 via manure from animal farming. In D3, there is no exceedance of the EBT value, indicating that E2 is removed during the different drinking water treatment processes. Also, in the other drinking water treatments the compounds causing the estrogenic activity are mostly removed. The EBT is exceeded only in drinking water D1 in only 3% percent of the drinking water measurements.

Glucorticoid activity as measured with the GR-CALUX is not often detected in untreated drinking water sources and never in drinking water in the period 2013-2017. The EBT value, which is one order of magnitude higher than the LOQ, is only exceeded in S1 and S4 in one of the samples (Figure 4, Table S2).

Anti-androgenic activity as measured with the anti-AR-CALUX has been analysed since 2016 and is regularly detected in the untreated drinking water sources. Anti-androgenic activity is measured above the LOQ in more than

83% of the samples from all sources. Anti-androgenic responses do fluctuate over time in the sources and equivalent concentrations up to 35 µg Flut-eq/L are measured (Figure 4). The anti-androgenic activity decreases during the different water treatment processes and also in D2, D3 and D4, the EBT is sporadically exceeded. Van der Oost et al. (2017) (van der Oost et al., 2017a) found that the background concentrations for anti-androgenic activity in unpolluted sites were almost 100'000 times higher than the BEQ that was considered safe based on ecological toxicity data for the most sensitive organism, but also 35 times higher than the BEQ derived as the 5% hazard concentration in a SSD analysis. The group of compounds that are known to have anti-androgenic potencies is large and heterogeneous and include polychlorinated biphenyls, pesticides and pharmaceuticals (e.g. flutamide) (Aït-Aïssa et al., 2010; Hamers et al., 2011; Houtman et al., 2020), but also natural organic matter and humic acids that are considered non-toxic (Bittner et al., 2012). This makes it more complicated to derive and interpret an EBT value. Since the EBT value is exceeded regularly, it is advisable to get better insight in the compounds that could be responsible for the responses, especially in drinking water. If the compounds can be identified it is possible to perform a more specific risk assessment and fine-tune the EBT to more realistically indicate potential risks.

It must be stressed that an exceedance of the EBT value does not necessary mean that there is a risk, however it signals the presence of a potentially harmful compound (hazard). In cases in which the measured response in drinking water exceeds the EBT value, it is proposed to repeat the bioassay measurement in the specific location. A single exceedance can always be caused by a temporarily contamination caused by for example shipping or industrial discharge. If the measured response continues to exceed the EBT value, it would be advised to initiate an investigation to find out which compounds are responsible. Effect directed analysis (EDA), an approach that combines biological analysis with chemical analysis is a promising tool for the detection and identification of the responsible biologically active compounds (Houtman et al., 2018; Zwart et al., 2018).

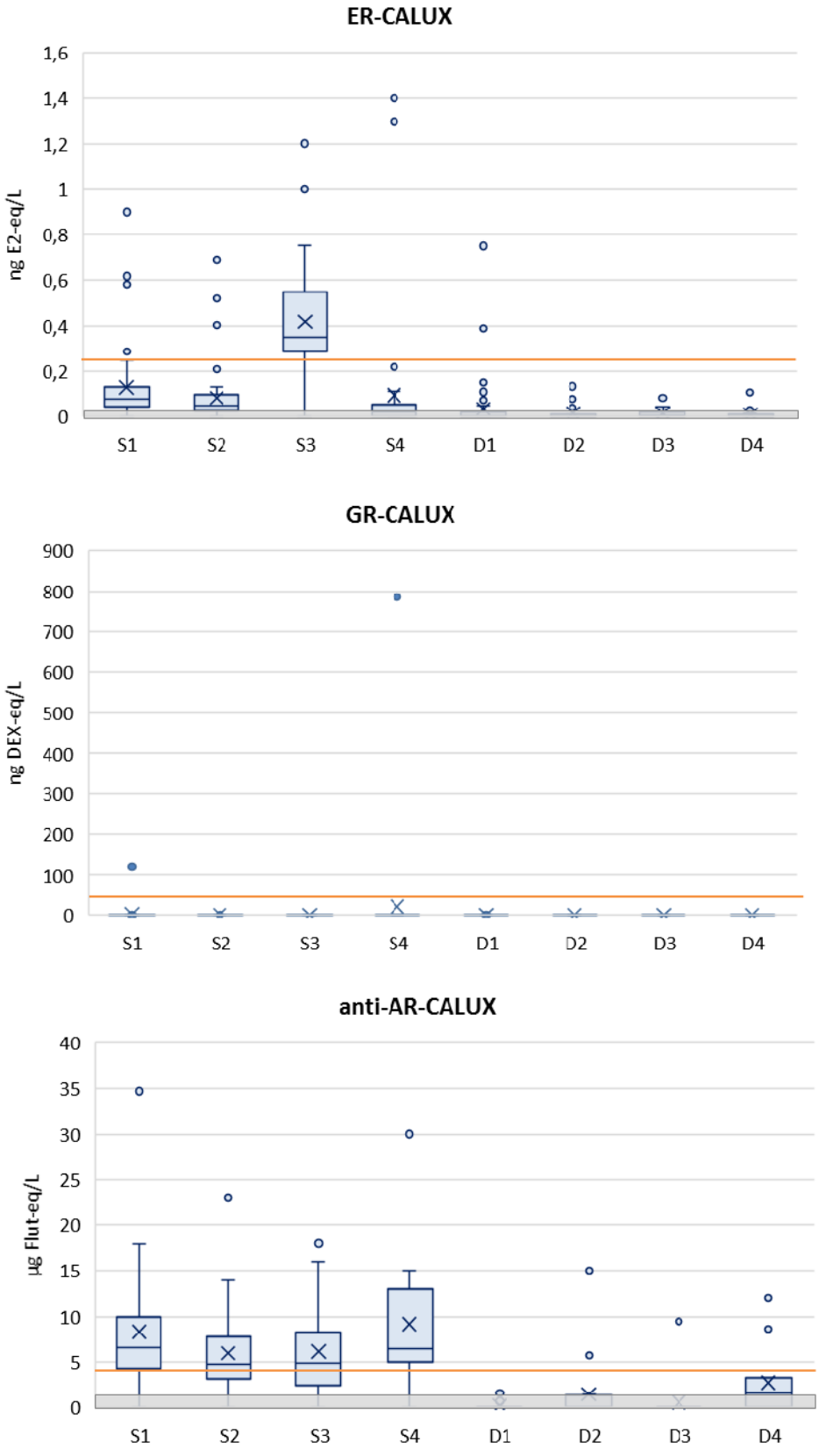


Figure 3: Boxplots of the bioassay activities in the drinking water sources S1-S4 and the corresponding drinking waters D1-D4 as measured with the ER-, GR- and anti-AR-CALUX. The bullet represents the average, whereas the line represents the median value. Outliers are shown as circles. The EBT is indicated by a yellow line in the figures and the area below the LOQ is indicated in grey.

4 Conclusion

Bioassays have become an important tool in assessing water quality, yet the interpretation of results can be challenging as a positive response does not necessarily implicate a risk for humans. Consequently, effect-based trigger values (EBT) that relate response levels to potential human risks are needed. Following established procedures, we attempted to derive EBT values for a set of eight CALUX bioassays which could be implemented in practice. We developed an innovative algorithm which uses available data about bioassays analogues to the CALUX assays, together with Monte Carlo simulations, to derive EBT values and assess their protective power, or in other words, their ability to detect the presence of many potentially hazardous chemicals at concentrations close to their provisional health-based guidance values in drinking water (pGLV). For four of the considered bioassays, namely ERα-, anti-AR-, AR- and GR-CALUX, the available data were deemed sufficient to derive reliable and realistic EBT values which could be used in practice to monitor the quality of drinking water. For the remaining four (i.e., PR-, PAH, NRF2- and p53-CALUX) the limited amount of data available as well as the important differences between EBT values obtained here and EBT values reported in the literature prevented their reliable implementation in practice.

Besides testing a larger number of chemicals, to obtain more EC50 values from which REP can be calculated, there is also a lack of pGLV values. In fact, for quite a large number of substances for which activity was detected in one or more bioassays, health guideline values for drinking water are lacking. Some of the derived EBT values could be compared to actual data from drinking water and sources in the Netherlands. For ERα- and GR-CALUX, the derived EBT did not seem to cause particularly high exceedances and thus appear to be appropriate for use in routine settings. For anti-AR-CALUX, however, higher numbers of exceedances were observed and more insights are required to understand which compounds are causing these. Whilst obtaining a definitive EBT is still a matter of debate, this work provides a framework to derive EBT values and assess their protective power using available data.

5 References

Aït-Aïssa, S., Laskowski, S., Laville, N., Porcher, J.-M., Brion, F., 2010. Anti-androgenic activities of environmental pesticides in the MDA-kb2 reporter cell line. Toxicol. In Vitro 24, 1979–1985. https://doi.org/10.1016/j.tiv.2010.08.014

Altenburger, R., Ait-Aissa, S., Antczak, P., Backhaus, T., Barceló, D., Seiler, T.-B., Brion, F., Busch, W., Chipman, K., de Alda, M.L., de Aragão Umbuzeiro, G., Escher, B.I., Falciani, F., Faust, M., Focks, A., Hilscherova, K., Hollender, J., Hollert, H., Jäger, F., Jahnke, A., Kortenkamp, A., Krauss, M., Lemkine, G.F., Munthe, J., Neumann, S., Schymanski, E.L., Scrimshaw, M., Segner, H., Slobodnik, J., Smedes, F., Kughathas, S., Teodorovic, I., Tindall, A.J., Tollefsen, K.E., Walz, K.-H., Williams, T.D., Van den Brink, P.J., van Gils, J., Vrana, B., Zhang, X., Brack, W., 2015. Future water quality monitoring — Adapting tools to deal with mixtures of pollutants in water resource management. Sci. Total Environ. 512–513, 540–551. https://doi.org/10.1016/j.scitotenv.2014.12.057

Baars, A.J., Theelen, R.M.C., Janssen, P.J.C.M., Hesse, J.M., van Apeldoorn, M.E., Meijerink, M.C.M., Verdam, L., Zeilmaker, M.J., 2001. Re-evaluation of human-toxicological maximum permissible risk levels (No. 711701 025). Rijksinstituut voor Volksgezondheid en Milieu (RIVM), Bilthoven.

Baken, K.A., Sjerps, R.M.A., Schriks, M., van Wezel, A.P., 2018. Toxicological risk assessment and prioritization of drinking water relevant contaminants of emerging concern. Environ. Int. 118, 293–303. https://doi.org/10.1016/j.envint.2018.05.006

Brand, W., de Jongh, C.M., van der Linden, S.C., Mennes, W., Puijker, L.M., van Leeuwen, C.J., van Wezel, A.P., Schriks, M., Heringa, M.B., 2013. Trigger values for investigation of hormonal activity in drinking water and its sources using CALUX bioassays. Environ. Int. 55, 109–118. https://doi.org/10.1016/j.envint.2013.02.003

Escher, B.I., Aït-Aïssa, S., Behnisch, P.A., Brack, W., Brion, F., Brouwer, A., Buchinger, S., Crawford, S.E., Du Pasquier, D., Hamers, T., Hettwer, K., Hilscherová, K., Hollert, H., Kase, R., Kienle, C., Tindall, A.J., Tuerk, J., van der Oost, R., Vermeirssen, E., Neale, P.A., 2018. Effect-based trigger values for in vitro and in vivo bioassays performed on surface water extracts supporting the environmental quality standards (EQS) of the European Water Framework Directive. Sci. Total Environ. 628–629, 748–765. https://doi.org/10.1016/j.scitotenv.2018.01.340

Escher, B.I., Leusch, F., 2011. Bioanalytical Tools in Water Quality Assessment - With contributions by Heather Chapman and Anita Poulson. IWA Publishing, London.

Escher, B.I., Neale, P.A., Leusch, F.D.L., 2015. Effect-based trigger values for in vitro bioassays: Reading across from existing water quality guideline values. Water Res. 81, 137–148. https://doi.org/10.1016/j.watres.2015.05.049

Escher, B.I., van Daele, C., Dutt, M., Tang, J.Y.M., Altenburger, R., 2013. Most Oxidative Stress Response In Water Samples Comes From Unknown Chemicals: The Need For Effect-Based Water Quality Trigger Values. Environ. Sci. Technol. 47, 7002–7011. https://doi.org/10.1021/es304793h

Fewtrell, L., Bartram, J., 2001. Water quality: Guidelines, standards and health. Assessment of risk and risk management for water-related infectious disease (No. 924154533X). World Health Organization, Geneva.

Hamers, T., Kamstra, J.H., Cenijn, P.H., Pencikova, K., Palkova, L., Simeckova, P., Vondracek, J., Andersson, P.L., Stenberg, M., Machala, M., 2011. In Vitro Toxicity Profiling of Ultrapure Non–Dioxin-like Polychlorinated Biphenyl Congeners and Their Relative Toxic Contribution to PCB Mixtures in Humans. Toxicol. Sci. 121, 88–100. https://doi.org/10.1093/toxsci/kfr043

Houtman, C.J., Kroesbergen, J., Lekkerkerker-Teunissen, K., van der Hoek, J.P., 2014. Human health risk assessment of the mixture of pharmaceuticals in Dutch drinking water and its sources based on frequent monitoring data. Sci. Total Environ. 496, 54–62. https://doi.org/10.1016/j.scitotenv.2014.07.022

Houtman, C.J., ten Broek, R., Brouwer, A., 2018. Steroid hormonal bioactivities, culprit natural and synthetic hormones and other emerging contaminants in waste water measured using bioassays and UPLC-tQ-MS. Sci. Total Environ. 630, 1492–1501. https://doi.org/10.1016/j.scitotenv.2018.02.273

Houtman, C.J., ten Broek, R., van Oorschot, Y., Kloes, D., van der Oost, R., Rosielle, M., Lamoree, M.H., 2020. High resolution effect-directed analysis of steroid hormone (ant)agonists in surface and wastewater quality monitoring. Environ. Toxicol. Pharmacol. 80, 103460. https://doi.org/10.1016/j.etap.2020.103460

International Association of Waterworks in the Rhine Basin, 2020. European River Memorandum for Quality Assurance of Drinking Water Production.

RStudio Team, 2016. RStudio: Integrated Development for R. RStudio Inc., Boston.

Schriks, M., Heringa, M.B., van der Kooi, M.M.E., de Voogt, P., van Wezel, A.P., 2010. Toxicological relevance of emerging contaminants for drinking water quality. Water Res. 44, 461–476. https://doi.org/10.1016/j.watres.2009.08.023

Tang, J.Y.M., McCarty, S., Glenn, E., Neale, P.A., Warne, M.St.J., Escher, B.I., 2013. Mixture effects of organic micropollutants present in water: Towards the development of effect-based water quality trigger values for baseline toxicity. Water Res. 47, 3300–3314. https://doi.org/10.1016/j.watres.2013.03.011

United States Environmental Protection Agency, 2018. ToxCast and Tox21 Summary Files [WWW Document]. https://doi.org/10.23645/epacomptox.6062479.v1

United States Environmental Protection Agency, 2016. ToxCast Dashboard [WWW Document]. US EPA. URL https://www.epa.gov/chemical-research/toxcast-dashboard (accessed 6.14.19).

van den Berg, M., Birnbaum, L.S., Denison, M., De Vito, M., Farland, W., Feeley, M., Fiedler, H., Hakansson, H., Hanberg, A., Haws, L., Rose, M., Safe, S., Schrenk, D., Tohyama, C., Tritscher, A., Tuomisto, J., Tysklind, M., Walker, N., Peterson, R.E., 2006. The 2005 World Health Organization Re-evaluation of Human and Mammalian Toxic Equivalency Factors for Dioxins and Dioxin-like Compounds. Toxicol. Sci. Off. J. Soc. Toxicol. 93, 223–241. https://doi.org/10.1093/toxsci/kfl055

van der Aa, N.G.F.M., van Leerdam, R.C., van de Ven, B., Janssen, P.J.C.M., Smit, C.E., Versteegh, J.F.M., 2017. Evaluatie signaleringsparameter nieuwe stoffen drinkwaterbeleid (No. RIVM-2017-0091). National Institute for Public Health and the Environment.

van der Oost, R., Sileno, G., Janse, T., Nguyen, M.T., Besselink, H., Brouwer, A., 2017a. SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality assessment: Part II-field feasibility survey: Field feasibility bioanalytical SIMONI hazard identification. Environ. Toxicol. Chem. 36, 2400–2416. https://doi.org/10.1002/etc.3837

van der Oost, R., Sileno, G., Suárez-Muñoz, M., Nguyen, M.T., Besselink, H., Brouwer, A., 2017b. SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality assessment: Part I-model design and effect-based trigger values: SIMONI: Smart bioanalytical monitoring of micropollutants. Environ. Toxicol. Chem. 36, 2385–2399. https://doi.org/10.1002/etc.3836

Zwart, N., Nio, S.L., Houtman, C.J., de Boer, J., Kool, J., Hamers, T., Lamoree, M.H., 2018. High-Throughput Effect-Directed Analysis Using Downscaled in Vitro Reporter Gene Assays To Identify Endocrine Disruptors in Surface Water. Environ. Sci. Technol. 52, 4367–4377. https://doi.org/10.1021/acs.est.7b06604

I Supplementary information: Evaluation of different selection criteria and calculation approaches

I.I Selection criteria

Six criteria were considered to select the compounds from which EBT values were derived. These criteria were defined because not all substances for which CALUX-specific data are available are relevant to establish EBT values that can be used in routine water monitoring.

- Criterium (all data (1)): EBT values were derived using all available data, thus without applying any additional selection criteria.

- Criterium (< 2mg/L (2)): Only compounds whose effective concentrations in the assay medium were below 2 mg/L (LOECs for P53 and NRF2) or 0.2 mg/L (EC_{50} values for the remaining assays) are selected. Based on the sample preparation protocols used in the Netherlands (Houtman et al. 2018), this corresponds to concentrations of
  10-100 µg/L in the original water samples. This upper boundary was chosen for drinking water samples as analytes present at these concentrations would be easily detected using conventional chemical analyses, and therefore bioassays are not necessary to detect these.

- Criterium (REP ≥ 0.001 (3)): Only compounds with a REP ≥ 0.001 are selected to exclude compounds with low potency, as including them would result in unrealistically skewed pGLV-BEQ distributions (Escher et al., 2018; van der Oost et al., 2017). It cannot be excluded however that such low-activity substances in a particular bioassay can represent a health risk via another mode of action. A high pGLV for a substance indicates that a risk occurs only at high doses. If the potency of the reference compound is high (low concentration needed to elicit an effect) then there is a higher chance that other (environmental) substances have a relatively low REP. More compounds would thus be excluded in bioassays that have a highly potent reference compound compared to other substances. In further studies it can be determined if the reference compound is not overly potent (and thus readily detectable at very low concentrations) and precludes the detection of less potent, yet still relevant and potentially active, substances.

- Criterium (Css (4)): Exclude compounds based on modelled internal steady state concentrations (Css) (expressed in reference compound-equivalents) resulting from chronic exposure to 2 L water per day at pGLV (safe) concentrations. If equivalent internal steady state concentrations remain low, it is assumed that that the mode of action as measured in the bioassay is not relevant for the adverse effect based on which the pGLV has been derived. Thus, compounds were excluded if their Css was < 1/1000 compared to the Css of the reference compound. Internal steady-state concentrations (See data file provided as Excel document) were estimated using the high-throughput PBPK model that was developed by Sipes et al. 2017 (Sipes et al., 2017).

- Criterium (0.1 < pGLVi/EC50i (or LOECi) < 10 (5)): exclude compounds if 0.1 < pGLVi/EC50i (or LOECi) < 10, according to Escher et al. (Escher et al., 2015). The lower boundary relates to the case where compound is potent in humans (low pGLV), but not in the bioassay (i.e. the substance also triggers another more

sensitive mode of action in humans, not related to the mode of action captured by the bioassay, to which the pGLV is related). The upper boundary is used to exclude compounds which, for instance, have low EC50, thus the bioassay is responsive, but high pGLVs. This is not per se contradictive as, in humans, the uptake of the compound may be low and/or it may be rapidly metabolized, resulting in low internal concentrations. In such cases the low toxicity in vivo is thus not due to the low intrinsic potency of the chemical, but to the ability of the human body to mitigate internal concentrations. So, compounds that adhere to this criterion have, in most cases, a high pGLV-BEQ value and would thus skew the distribution of pGLV-BEQ for a given bioassay.

- Criterion (6): exclude compounds if 0.1 < pGLVi/EC50i (or LOECi), as for criterion (5) (i.e. ,0.1 < pGLVi/EC50i (or LOECi) <10) yet without the upper boundary (i.e., excluding compounds where the pGLV is more than ten times higher than the EC50 concentration).

I.II EBT value estimation approaches

Five approaches to derive EBT values were applied for each of the six selection criteria.

- Approach (5th norm. dist. (A)): Available pGLV-BEQs were assumed to follow a normal distribution and the 5th percentile was calculated from the corresponding cumulative distribution function and used as EBT.

- Approach (5th direct (B)): EBT values are derived by taking the 5th percentile, computed directly, of pGLV-BEQs of all compounds available for each bioassay (Escher et al., 2015).

- Approach (gen. linear model (C)): Compute a generalized linear model using the rank of the available compounds as predictor and the corresponding pGLV-BEQs as response variable. The intercept of the obtained regression was used as EBT value, which is thus below the lowest pGLV-BEQ available.

- Approach (quant. reg. (D)): Similar to approach (C), yet in this case a quantile regression estimated at the 5th quantile was used instead of a conventional linear regression. The advantage of this approach is that it is more robust against outliers as it makes no assumptions regarding the distribution of the residuals.

- Approach (lowest (E)): Take the lowest pGLV-BEQ value available for each bioassay as EBT value.

From the available data (Supplemental Tables #-#), EBT values were computed for each CALUX bioassay using the previously described selection criteria and mathematical approaches. Results are summarised in Figure S1.

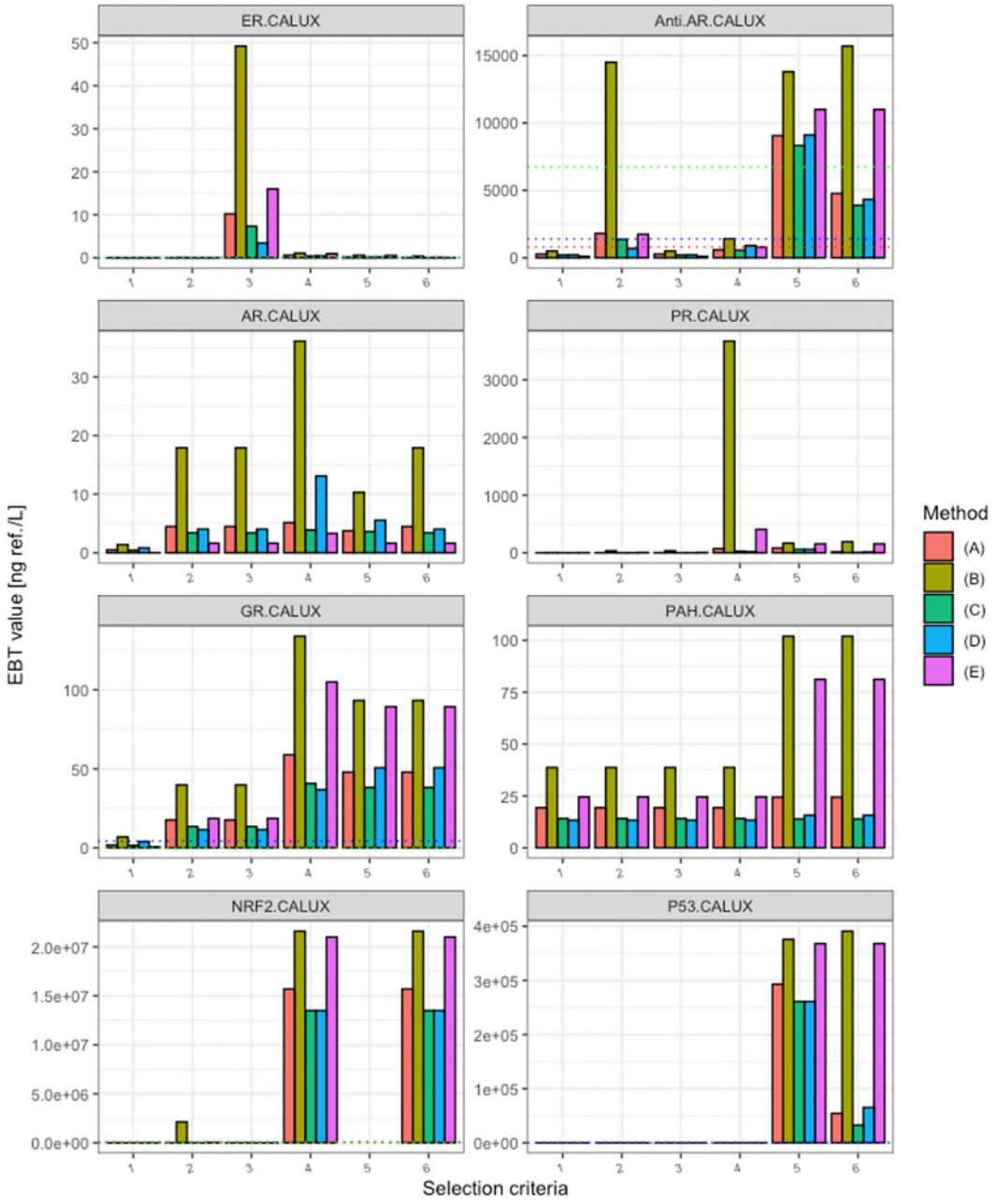


Figure S4: Calculated EBT values for the different CALUX assays using the described selection criteria (1-6) and estimation approaches (A-E). Where available, the horizontal lines indicate the reporting limit (blue), the average response measured in drinking (red) and raw (green) water.

With regard to the approaches to deriving the EBT, a trend can be seen between the different approaches. In terms of conservativeness, the approaches can be ranked 5th direct (B), lowest (E), 5th norm. dist (A) and gen. linear model (C), quant. reg. (D) (high EBT to low EBT).

Approach 5th direct (B) generally results in a high EBT value. This approach takes the value at the 5th percentile of the available values. lowest (E) also gives a relatively high EBT value, likely due to a less steep slope which results in the origin having a higher value. 5th norm. dist. (A) and lowest (E) give about the same result. Approach lowest (E) takes the minimum among the available pGLV-BEQ values. This corresponds in many cases to the 5th percentile of a normal distribution of the available values (A). This has to do with the number of substances available from which the normal distribution is derived. The more compounds available to derive the normal distribution, the more likely

that the 5th percentile will be above the lowest available value, and then 5th norm. dist (A) will give higher EBTs than lowest (E). Gen. linear model (C) and quant. reg. (D) are in most cases the most conservative approaches as they both derive an EBT that is below the lowest available pGLV-BEQ value.

I.III Selection criteria per bioassay

The effect of using a selection criterion differs from one bioassay to another. We therefore discuss the results per bioassay.

In the case of the ERα-CALUX, selection criterion (3) (i.e., REP > 0.001) provides the highest EBT values because only four compounds with very high REPs are retained. The discarded compounds have REPs 4 to 8 orders of magnitude lower. The other selection criteria provided EBT values which were generally 2 to 3 orders of magnitude lower compared to method REP ≥ 0.001 (3). Most derived EBT values were above the reporting limit currently in use at Het Water Laboratorium (HWL), namely 0.034 ng E2 eq./L.

For the anti-AR-CALUX, highest EBT values were obtained with methods (5) (i.e., 0.1 < pGLVi/EC50i (or LOECi) <10) and (6) (i.e., 0.1 < pGLVi/EC50i (or LOECi)) because, in both cases, compounds in these bioassays had pGLV and EC50 values yielding pGLV/EC50 ratios above the 0.1 threshold. Furthermore, the remaining compounds for this assay had high REPs (above 0.05) and high pGLV, resulting in high pGLV-BEQs.

For the AR-CALUX, differences between selection methods were less pronounced. Only when no selection criterion was applied all data (1), low EBT values were derived due to 17ß-estradiol, which was not excluded and this compound has a very low pGLV-BEQ of 0.02 ng DHT eq./L.

For the PR-CALUX, a particularly high EBT value was obtained with selection criterion (4) (Css > 1/1000 of the reference compound), as this coincided with excluding compounds with the lowest pGLV-BEQ. In fact, this method excluded compounds with the lowest pGLV-BEQs, in all bioassays.

In the case of GR-CALUX, methods 1-3 yielded lower EBT values in this bioassay than the other methods, which did not exclude the substances with the lowest pGLV-BEQ. Obtained EBT values were always higher compared to values from routine measurements.

In the case of PAH-CALUX, differences between the obtained EBT values were less pronounced between the considered selection criteria. Only in (5) (0.1 < pGLVi/EC50i (or LOECi) <10) and (6) (i.e., 0.1 < pGLVi/EC50i (or LOECi)) the derivation approaches 5th direct (B) and lowest (E) yielded high EBT values. For this bioassay only a few values are available and methods 5th direct (B) and lowest (E) tend to provide an EBT value above the lowest value.

For Nfr2-CALUX, substantially higher EBT values were obtained with selection criteria Css (4) and (6) (i.e., 0.1 < pGLVi/EC50i (or LOECi)) compared to the rest. It should be noticed that for selection criteria (5) (0.1 < pGLVi/EC50i (or LOECi) <10), no substances were available for calculations as all had pGLV-BEQs values either below 0.1 or above 10. With criteria Css (4) and (6) (i.e., 0.1 < pGLVi/EC50i (or LOECi)), only two compounds were available for derivation of EBT values and both had high pGLV-BEQs.

For p53-CALUX, criteria (5) (i.e., 0.1 < pGLVi/EC50i (or LOECi) <10) and (6) (i.e., 0.1 < pGLVi/EC50i (or LOECi)) yielded the highest EBTs. Only substances with high pGLV-BEQ values were selected after applying criteria (5) (i.e., 0.1 < pGLVi/EC50i (or LOECi) <10) and (6) (i.e., 0.1 < pGLVi/EC50i (or LOECi)).

I.IV Overview

The obtained results show that the criterion used to select (or exclude) substances (i.e., pGLV-BEQ values) is the main factor affecting the obtained EBT values, while the influence of derivation approaches is only secondary.

Nevertheless, some differences exist between derivation approaches and a trend can be observed, from those generally giving lower EBT values to those giving higher values. The suggested method to derive the EBT is to take the 5^{th} percentile of the normal distribution of the pGLV-BEQs. This is a reasonably robust criterion, also between the different selection criteria of compounds, as the true distribution of pGLV-BEQs of relevant substances (including those that have not been measured) can be assumed to be normal. When based on a sufficiently large number of representative substances, any new substance (and its pGLV-BEQ) should fall within this normal distribution.

As mentioned, the selection criteria for relevant substances appears to have a large impact on the derived EBT values. Yet differences exist between selection criteria and bioassays. This is partly because the specific exclusion (or rather selection) of one substance in a selection method can make a large difference in the derived EBT for a bioassay, especially if there are few tested compounds. Consequently, derived EBT values will be less robust (i.e., representative). This also explains why differences exist between EBT values reported in the literature for the same bioassay: not only the substances (and their pGLV-BEQ) used to derive the EBT values are different, also selection criteria used to include or exclude differ.

A modified version of the selection criterion of Escher et al. (2015), e.g. exclude compounds based on pGLVi/EC50i < 0.1, appears to be a good compromise. Compounds excluded based on this criterion cannot reliably be used to derive a trigger value, as they are not responsive enough in the considered bioassay and thus less relevant for the measured endpoint. Omitting the other half of the criterion defined by Escher et al. (2015) (i.e., pGLV-BEQ to EC50 ratio > 10), means that more compounds in the higher ends of the pGLV-BEQ are included. This would in particular influence derivation approaches based on percentile and distributions (i.e., A, B and D) as they will yield slightly higher EBT values compared to the original Escher et al., 2015 criterion (as can be seen for instance in Figure S1 for anti-AR-CALUX, where derived EBT values decrease from criteria 5 to 6).

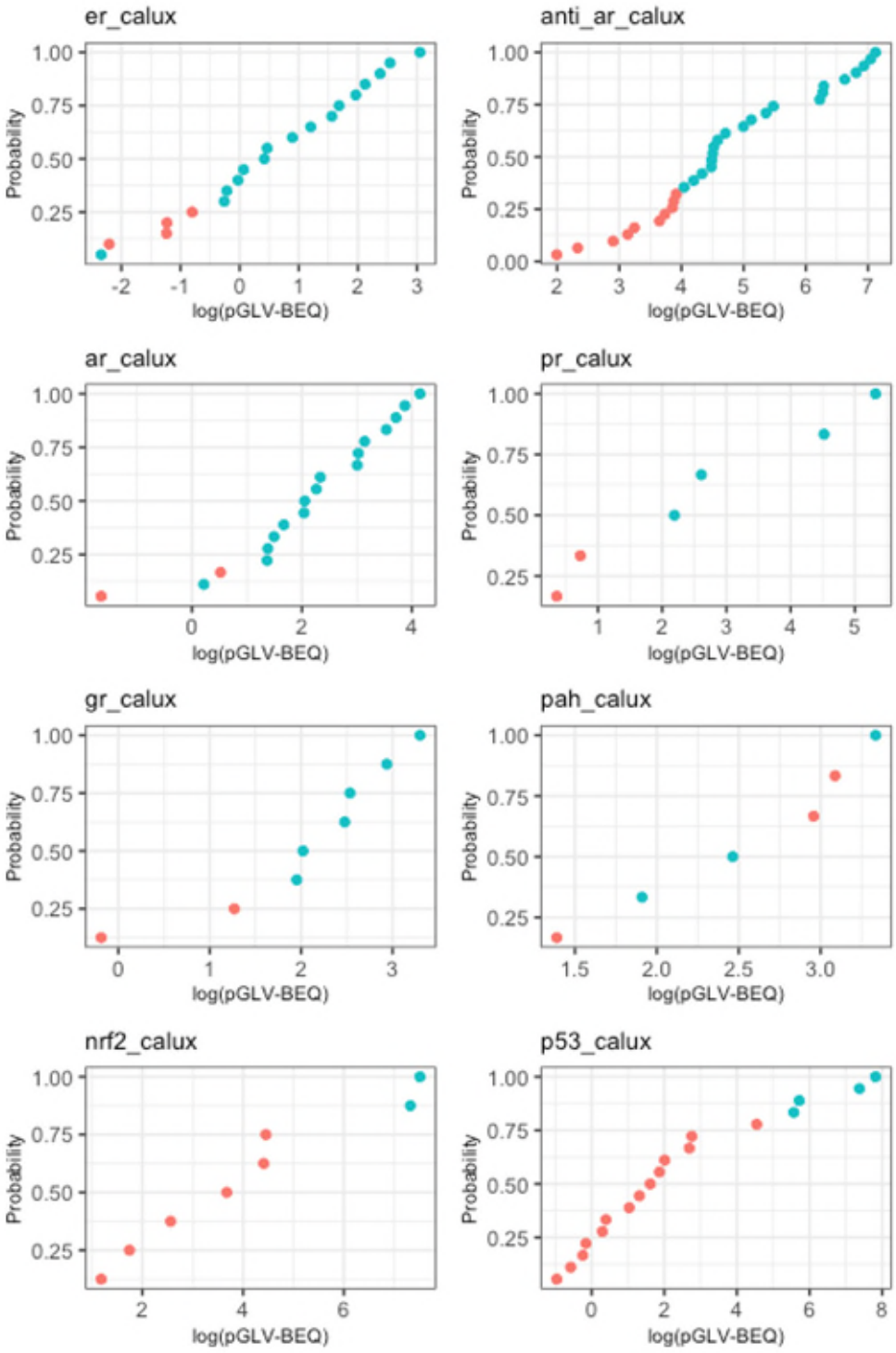


Figure S5: Cumulative distribution function of pGLV-BEQ values calculated for compounds in the various CALUX bioassays. Red dots indicate compounds that were excluded based on the applied selection criterion (6).

Table S3: CALUX-analogue bioassays present in the ToxCast database which were used for the simulations.

CALUX assay	Mechanism	Reference chemical	cas	Corresponding ToxCast
AR CALUX	androgen receptor activation	dihydrotestosterone	58-22-0	ATG_AR_TRANS_up OT_AR_ARELUC_AG_1440 TOX21_AR_BLA_Agonist_ratio TOX21_AR_LUC_MDAKB2_Agonist
anti-AR CALUX	repression of androgen receptor activation	flutamide	13311-84-7	TOX21_AR_BLA_Antagonist_ratio TOX21_AR_LUC_MDAKB2_Antagonist
ERa CALUX	estrogen receptor ?-mediated signaling	17a-estradiol	57-91-0	ATG_ERE_CIS_up ATG_ERa_TRANS_up OT_ERa_EREGFP_0120 OT_ERa_EREGFP_0480 TOX21_ERa_BLA_Agonist_ratio TOX21_ERa_LUC_BG1_Agonist
GR CALUX	glucocorticoid receptor-mediated signaling	dexamethasone	50-02-2	ATG_GRE_CIS_up ATG_GR_TRANS_up TOX21_GR_BLA_Agonist_ratio
anti-PR CALUX	repression of progesterone receptor-mediated signaling	RU486		ATG_PR_TRANS2_up
Nrf2 CALUX	activation of the Nrf2 pathway / oxidative stress response	curcumin	1948-33-0	ATG_NRF2_ARE_CIS_up TOX21_ARE_BLA_agonist_ratio
p53 CALUX (+S9)	p53-dependent pathway activation / genotoxicity response (+/- S9)	actinomycin D	53-96-3	ATG_p53_CIS_up TOX21_p53_BLA_p1_ratio TOX21_p53_BLA_p2_ratio TOX21_p53_BLA_p3_ratio TOX21_p53_BLA_p4_ratio TOX21_p53_BLA_p5_ratio
PAH CALUX	Activation of the aryl hydrocarbon receptor	benzo[a]pyrene (BaP)	50-32-8	TOX21_AhR_LUC_Agonist ATG_Ahr_CIS_up

Table S4: Number of compounds from the analogue bioassays in ToxCast which were available before and after applying the 0.1 < pGLV/EC50 (or LOEC) selection criterion (6).

Analogues to	# of compounds without selection criterium	#compounds with selection criterium
ERα-CALUX	141	23
AR-CALUX	35	8
anti-AR-CALUX	69	11
GR-CALUX	19	2
Nrf2-CALUX	115	20
P53-CALUX	158	21
PAH-CALUX	36	6

I.V Routine monitoring of Dutch surface and drinking water

Sampling

Surface water grab samples and drinking water samples were collected in green glass bottles pre-rinsed with ethyl acetate at four abstraction points used by Dutch drinking water companies at the rivers Meuse (Enclosed Meuse) (source 1; S1) and Rhine (Lek Channel) (S2), a reclaimed land area Bethune Polder (S3), Lake IJssel (S4) and the corresponding drinking waters (D1-D4). S1 and S2 exhibit a stronger anthropogenic influence, and concentrations of compounds (industrial, pharmaceuticals and pesticides) are in general higher in these sources. All surface waters are treated with a multi-barrier water treatment, including amongst others active carbon filtration, advanced oxidation processes and/or dune filtration. For a detailed description of the sites, see Houtman et al. 2019 (Houtman et al., 2019).

Bioassays

CALUX cells were supplied by BioDetection systems B.V. (Amsterdam, the Netherlands). One litre water samples were collected form the influent (feed water) and effluent (treated water of the pilot plants) and stored in green glass bottles at 4°C and processed within one week. Compounds in the samples were extracted on Oasis HLB SPE cartridges as described in Houtman et al. 2018 and extracts were reconstituted in 50 µL dimethylsulfoxide (DMSO). From these extracts, 3-, 10-, 30- and 100- fold dilutions were prepared in DMSO. Extracts and dilutions were tested in the ERα- and GR-CALUX and Anti-AR CALUX at 0.1% extract concentration as described earlier (Houtman et al., 2018; van der Linden et al., 2008). 17β-estradiol (E2), dexamethasone and flutamide were used as reference compounds. Anti-androgenic activity was tested in the presence of DHT at EC50 level.

An overview of the amount of samples per location is given, including the median responses that are detected in sources and drinking water with the CALUX assays and the percentage of samples that exceeds the EBT is shown in Tables S1-S3.

Table S1: Results of the ER-CALUX measurements in sources and drinking waters in the period 2013-2017. LOQ of the ER-CALUX is 3.4E-02 ng E2-eq/L , EBT value is 2.5E-01 ng E2-eq/L.

Sample point (#)	Median [ng E2-eq/L]	>LOQ [%]	Exceedance EBT value [%]
S1 (60)	7.8E-02	83%	8%
S2 (66)	5.1E-02	73%	5%
S3 (27)	3.5E-01	93%	85%
S4 (39)	<LOQ	36%	5%
D1 (95)	<LOQ	19%	3%
D2 (29)	<LOQ	17%	0%
D3 (29)	<LOQ	14%	0%
D4 (28)	<LOQ	7%	0%

Table S2: Results of the GR-CALUX measurements in sources and drinking waters in the period 2013-2017. LOQ of the GR-CALUX is 4.3 ng Dex-eq/L, EBT value is 48 ng Dex-eq/L.

Sample point (#)	Median [ng Dex-eq/L]	>LOQ [%]	Exceedance EBT value [%]
S1 (60)	<LOQ	2%	2%
S2 (66)	<LOQ	5%	0%
S3 (29)	<LOQ	0%	0%
S4 (39)	<LOQ	3%	3%
D1 (91)	<LOQ	0%	0%
D2 (29)	<LOQ	0%	0%
D3 (29)	<LOQ	0%	0%
D4 (28)	<LOQ	0%	0%

Table S3: Results of the Anti-AR-CALUX measurements in sources and drinking waters in the period 2016-2017. LOQ of the anti-AR-CALUX is 1.4E+03 ng Flut-eq/L, EBT value is 4.8E+03 ng Flut-eq/L.

Sample point (#)	Median [ng Flut-eq/L]	>LOQ [%]	Exceedance EBT value [%]
S1 (25)	6.6E+03	96%	64%
S2 (24)	4.8E+03	92%	46%
S3 (24)	4.9E+03	83%	50%
S4 (12)	6.4E+03	92%	83%
D1 (22)	<LOQ	9%	0%
D2 (17)	<LOQ	18%	12%
D3 (17)	<LOQ	6%	6%
D4 (11)	1.6E+03	55%	18%

I.VI References

Escher, B.I., Neale, P.A., Leusch, F.D.L., 2015. Effect-based trigger values for in vitro bioassays: Reading across from existing water quality guideline values. Water Res. 81, 137–148. https://doi.org/10.1016/j.watres.2015.05.049

Houtman, C.J., Kroesbergen, J., Baggelaar, P.K., van Lieverloo, J.H.M., 2019. Statistical analysis of a large set of semi-quantitative GC–MS screening data to evaluate and prioritize organic contaminants in surface and drinking water of the Netherlands. Sci. Total Environ. 697, 133806. https://doi.org/10.1016/j.scitotenv.2019.133806

Houtman, C.J., ten Broek, R., Brouwer, A., 2018. Steroid hormonal bioactivities, culprit natural and synthetic hormones and other emerging contaminants in waste water measured using bioassays and UPLC-tQ-MS. Sci. Total Environ. 630, 1492–1501. https://doi.org/10.1016/j.scitotenv.2018.02.273

Sipes, N.S., Wambaugh, J.F., Pearce, R., Auerbach, S.S., Wetmore, B.A., Hsieh, J.-H., Shapiro, A.J., Svoboda, D., DeVito, M.J., Ferguson, S.S., 2017. An Intuitive Approach for Predicting Potential Human Health Risk with the Tox21 10k Library. Environ. Sci. Technol. 51, 10786–10796. https://doi.org/10.1021/acs.est.7b00650

van der Linden, S.C., Heringa, M.B., Man, H.-Y., Sonneveld, E., Puijker, L.M., Brouwer, A., van der Burg, B., 2008. Detection of Multiple Hormonal Activities in Wastewater Effluents and Surface Water, Using a Panel of Steroid Receptor CALUX Bioassays. Environ. Sci. Technol. 42, 5814–5820. https://doi.org/10.1021/es702897y