Environment International 188 (2024) 108733

Contents lists available at ScienceDirect

Environment International

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

Full length article

Bioassay predictive values for chemical health risks in drinking water

Tessa E. Pronk [a,*], Renske P.J. Hoondert [a], Stefan A.E. Kools [a], Vikas Kumar [b,c,d], Milo L. de Baat [e]

[a] KWR Water Research Institute, Groningenhaven 7, 3433 PE Nieuwegein, the Netherlands
[b] Environmental Engineering Laboratory, Departament d' Enginyeria Quimica, Universitat Rovira i Virgili, Av. Països Catalans 26 43007, Tarragona, Catalonia, Spain
[c] IISPV, Hospital Universitari Sant Joan de Reus, Universitat Rovira i Virgili, Reus, Spain
[d] German Federal Institute for Risk Assessment (BfR), Max-Dohrn-Str. 8-10 10589, Berlin, Germany
[e] Dept. of Freshwater and Marine Ecology, Inst. for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Science Park 904 1098XH, Amsterdam, the Netherlands

ARTICLE INFO

Handling Editor: Adrian Covaci

Keywords:
Effect-based methods
Water quality
Risk assessment
In vitro
Predictive value

ABSTRACT

Bioanalytical tools can be used for assessment of the chemical quality of drinking water and its sources. For water managers it is important to know the probability that a bioassay response above an established health-based 'effect-based trigger value' (EBT) indeed implies a harmful chemical (mixture) concentration. This study presents and applies a framework, based on Bayes' theorem, to derive such risk probabilities for bioassay responses. These were evaluated under varying (*in silico*) chemical mixture concentrations relevant to drinking water (sources), with toxicity data for six *in vitro* assays from the ToxCast database. For single chemicals and *in silico* mixtures, the negative predictive value (NPV) was 100 % for all assays. For water managers, this means that when a bioassay response is below the EBT, a chemical risk is reliably absent, and no further action is required. The positive predictive value (PPV) increased with increasing chemical concentrations (2 μg/L) up to 40–80 %, depending on the assay. For *in silico* mixtures of increasing numbers of chemicals, the PPV did not increase until higher sum concentrations (>2–10 μg/L). Hence, the ability to accurately signal a harmful chemical (mixture) using bioassays will be lowest for highly diverse, low-concentration chemical mixtures. For water managers, this means in practice that further investigations after an EBT exceedance will, in many cases, not reveal chemicals at harmful concentrations. A solution offered is to increase the trigger value for positive responses to achieve a higher PPV and maintain the EBT for negative responses to ensure an optimal NPV.

1. Introduction

The monitoring and assessment of the chemical quality of drinking water and its sources can, in addition to (non–) targeted chemical analysis (Brunner et al., 2020; Renaud et al., 2022) be performed with effect-based methods (EBMs) using bioassays (Neale et al., 2021). EBMs can quantify effects caused by (unknown) chemical mixtures (Escher et al., 2015; Escher et al., 2023; Oskarsson et al., 2021). More potent or higher concentrations of chemicals that elicit a particular effect will cause a higher observed response in the specific corresponding bioassay, resulting in a risk-scaled assessment of mixture toxicity.

In recent decades, the applicability of EBMs to water quality monitoring has been demonstrated and EBMs are becoming increasingly available and accessible to water quality managers, regulators, and policymakers (Neale et al., 2023a). Due to their ability to characterize the hazards of complex chemical mixtures, EBMs are now recommended for the diagnosis and monitoring of water quality (Brack et al., 2019). However, many *in vitro* bioassays are highly sensitive and exhibit responses to contaminants at levels well below their guideline values or drinking water quality standards. Therefore, the use of EBMs for chemical water quality assessment requires threshold values that distinguish between acceptable and hazardous levels of pollution (Neale et al., 2023b). An effect-based 'trigger value' (EBT) can be applied to assess whether an observed effect exceeds a threshold above which risks are not negligible (Brand et al., 2013; Escher et al., 2015; Béen et al., 2021). Above this value, (mixtures of) chemicals may be present at concentrations that are harmful to human health. As such, the EBT can distinguish between poor and acceptable water quality and ensure the safety of its consumption. Hence, ideally, if the bioassay response is lower than the EBT, there is no risk of adverse effects on consumer health, while an exceedance of the EBT signifies a health risk.

At a high rate of correctly signalled health risks, further research into

* Corresponding author.
E-mail address: Tessa.Pronk@kwrwater.nl (T.E. Pronk).

https://doi.org/10.1016/j.envint.2024.108733
Received 3 March 2024; Received in revised form 17 April 2024; Accepted 8 May 2024
Available online 10 May 2024

the identity of the causative chemicals is more easily justified. A performance evaluation of bioassays can give more insight into this aspect. In addition, a performance evaluation can help to select the best-performing bioassay from several analogous alternative bioassays. An additional measure to evaluate bioassays on their ability to distinguish between background- and genuine chemical signal (not necessarily related to harmfulness) is the z-factor (Zhang et al., 1999).

Previous studies have assessed the accuracy of bioassays to signal whether a sample containing a chemical (mixture) is harmful by applying the concept of true or false positives and negatives (Swanson and Chirmule, 2009) to quantify *specificity* and *sensitivity* in bioassay responses to environmental water samples (Brion et al., 2019; CIS working group EBM, 2021). Sensitivity is a measure of the portion of samples containing chemicals at harmful concentrations that are signalled correctly by the bioassay as 'positive', i.e. a response above the EBT. The specificity is the portion of samples containing chemicals at non-harmful concentrations signalled correctly by the bioassay as 'negative', i.e. a response below the EBT. Hence, sensitivity and specificity are appropriate performance evaluation parameters to test how accurate the test (i.e. bioassay) result is when the outcome is already known: chemicals in the samples are harmful or not harmful at a certain concentration and the test is correct for a certain percentage of samples from either category. Although this gives a quantification of the reliability of the test, it does not equate to the predictive power of the bioassay for chemical risks in practice.

Instead, a more relevant question in risk quantification is what can be predicted about the harmfulness of a sample containing a (mixture of) chemical(s) when only the test result is known (Johnson, 2017). To answer that question, predictive values of bioassay responses are required. This statistical evaluation to assess test performance is very common in the medical domain (Webb and Sidebotham, 2020) but is currently not yet applied in effect-based water quality assessment. Bayes' theorem (or rule) is the mathematical relationship that relates sensitivity and specificity to predictive values (Johnson, 2017; Webb and Sidebotham, 2020). In Bayes' theorem, the prevalence of the risk is used to calculate a positive predictive value (PPV) and a negative predictive value (NPV). If an event (or in the case of this analysis, a chemical (mixture) at harmful concentrations) is very rare, false negatives or true positives will seldom occur at all, compared to the number of true negatives and false positives. By considering this, Bayes' theorem can provide the predictive values for bioassays to express risks as probabilities, which more accurately approaches the definition of risk in toxicology, i.e., the likelihood that harm from a specific hazard will occur. For any given test it applies that, as the prevalence of chemicals at harmful concentrations decreases, the PPV decreases because there will be more false positives for every true positive. Consequently, the probability that a response above the EBT is induced by a harmful chemical present in the sample, decreases.

In summary, a high PPV implies that, if the bioassay gives a response > EBT, the chemical (mixture) is in many cases present at a truly harmful concentration. A low PPV implies that on many of those occasions, this is not the case. However, another important performance measure for water managers is the number of times a bioassay gives a response > EBT altogether when water is tested. The more frequently a bioassay indicates a risk, the more often a water manager must follow up with additional analyses to identify potentially hazardous chemicals and formulate mitigation measures. Hence, an ideal scenario for a water manager is a bioassay with a low number of 'hits' (responses > EBT) combined with a high PPV. Although the number of hits is also highly dependent on the water quality, it is preferable to decrease the likelihood of having to further evaluate false positives.

The present study provides a framework to analyse the performance of bioassays regarding their predictive power for chemical health risks in drinking water (sources). This framework is subsequently demonstrated using the ToxCast database (US EPA, 2023), which contains a large amount of data on individually tested chemicals for many different *in vitro* toxicity assays that can potentially be used as bioassays in water quality assessment. To evaluate bioassay performance, twenty of these assays with human-relevant endpoints and a high number of tested chemicals were selected. EBTs were derived and bioassay predictive power to signal harmful chemicals and their mixtures was evaluated *in silico* under varying chemical concentrations relevant to the practice of drinking water quality assessment.

2. Methods

2.1. Selection of in vitro assays and tested chemicals

ToxCast data were downloaded (US EPA, 2023) and processed to include the concentration at which 50 % of the maximum activity is shown in an assay (AC_{50} values) for chemicals with a well-modelled dose response, i.e. a typical S-curve (US EPA, 2014; Ryan & Becker, 2017; Jonker & Van der Heijden, 2007; Groothuis et al., 2015; Feshuk et al., 2023; Filer et al., 2014). A description of the inclusion criteria used in curating the dataset is given in Supplementary Information (SI) I. To demonstrate the proposed framework for deriving bioassay predictive values for chemical health risks, 20 *in vitro* assays with human-relevant endpoints and a high number of tested chemicals were selected (SI I Table A). The selected assay endpoints represent cellular toxicity pathways including hormone receptor activation, regulation of metabolism, oxidative stress, and genotoxicity. Chemicals can be tested more than once in a ToxCast assay, sometimes leading to divergent dose–response results. To represent the full recorded range of responses, these duplicated – but varying – chemical activities in the assays were maintained in the dataset.

To interpret assay responses that are indicative of human health risks, a link must be made between a health risk and an assay response (in this case the AC_{50}). To facilitate this, available provisional guideline values (pGLVs) for active chemicals in the assays were collected. A pGLV for an individual chemical is the provisional, non-binding maximum advised daily dose in drinking water (Brand et al., 2013; Baken et al., 2018). The pGLV for a chemical is established under the assumption of a lifelong consumption of 2 L of water a day by a person of average body weight (70 kg). The pGLVs used in the present study were assembled from Béen et al. (2021), Baken et al. (2018), the Australian drinking water guidelines (2023), and an unpublished project report that used EFSA and WHO-advised daily intake values, amongst others (van den Berg, 2021). This resulted in a dataset of 415 chemicals with a pGLV. For some chemicals, the used sources provided different pGLVs. This can occur if there is a difference in the data sources used to derive the pGLV. Also here, all duplicate values were maintained in the dataset to represent the spread in pGLVs in the available literature.

The collected pGLVs were subsequently used to derive effect-based trigger values (EBT) for the assays according to the method described in Béen et al. (2021). A caveat with the pGLV in general is that this concentration is based on the most sensitive reported endpoint for the chemical in literature. This is not necessarily the target endpoint of the assay for which a threshold value is to be derived based on the pGLV. A sign that the pGLV is not relevant for the assay is if the AC_{50} deviates strongly from the pGLV. If the pGLV is very low compared to the AC_{50} of the bioassay, the pGLV is likely not valid for the activity that the bioassay represents. Therefore, a condition for the inclusion of the pGLV was applied as proposed by Béen et al. (2021):

Select if:

$$\frac{pGLV_i}{AC50_i} > 0.1 \quad (1)$$

Here the $pGLV_i$ is the provisional guideline value of chemical i and $AC50_i$ is the concentration at which half of the maximum activity is reached in the assay for chemical i. In SI II, the result of this selection is visualised in plots per assay, with AC_{50} values for chemicals plotted against their pGLV for selected and excluded chemicals. The resulting

numbers of selected chemicals per assay are shown in SI I Table A.

2.2. Bioassay selection criteria

To select assays for performance analyses, their ability to give a response up to EBT values and the number of tested chemicals with a pGLV were taken into account. For bioassays to be useful in chemical water quality assessment they should detect chemicals at environmentally relevant concentrations with activity above the EBT at harmful concentrations. In the practice of effect-based water quality assessment, a water sample is concentrated or diluted to obtain a concentration range (above the limit of detection and below the threshold at which any cytotoxic effects occur in the bioassay) for the derivation of the AC_{50} value, which is expressed in bioanalytical equivalents of a reference chemical (Neale et al., 2021). It was calculated for how many of the chemicals, at a concentration of 1 μg/L, the bioanalytical equivalent concentration (*BCA*, Equation (6)) exceeds the EBT, as this value represents the signalling value for anthropogenic chemicals defined in the Dutch drinking water directive (2023). If chemical concentrations are detected above this value, further assessments of risks for human consumption and/or mitigation actions are required. Since chemicals in drinking water sources often occur below 1 μg/L – at an average of 10 ng/L (RIWA-Rijn, 2022) – some bioassays would expectedly rarely show an effect above the EBT unless highly potent chemicals are present, or a few chemicals are present at high concentrations. If a bioassay would also rarely elicit an effect at chemical concentrations of 10 μg/L, the bioassay may be considered too insensitive for use in the assessment of (sources of) drinking water. A threshold of > 6 % of tested chemicals at 10 ug/L causing an exceedance of the EBT was used to identify assays as 'sensitive'. This is an arbitrary threshold and simply a means to focus on the topmost sensitive bioassays. Simultaneously, the selected bioassays should have a sufficient number of tested chemicals with a pGLV, otherwise the EBT and performance indicators are less reliably derived. If fewer than 16 chemicals (SI I Table A) with a relevant (Equation (1)) available pGLV were tested in the assay, it was excluded from the data analyses.

2.3. Calculation of health-based trigger values

Health-based EBTs for the selected bioassays were established using the method outlined in Béen et al. (2021) which is briefly explained below. Firstly, the effect potencies of chemicals are calculated relative to a (potent) reference chemical.

Table 1
Health-based EBTs for 20 ToxCast assays derived in the present study vs previously derived EBTs in analogous bioassays with the same reference chemical.

	Bioassay name		EBT (μg/L)	Reference chemical CAS number	Reference chemical name	Earlier EBT (μg/L)
1	ATG_ERa_TRANS_up	Estrogen receptor alpha agonism	8.4	50–28-2	17β-estradiol	0.0002 (1) 0.0018 (1) 0.0012 (1) 0.00025 (2) 0.0038 (3)
2	TOX21_ERa_LUC_VM7_Agonist	Estrogen receptor alpha agonism	0.006	50–28-2	17β-estradiol	0.0002 (1) 0.0018 (1) 0.0012 (1) 0.00025 (2) 0.0038 (3)
3	TOX21_AR_BLA_Antagonist_ratio	Androgen receptor antagonism	533	13311–84-7	flutamide	4.8 (2)
4	TOX21_ARE_BLA_agonist_ratio	ARE receptor agonism	908*	62–73-7	dichlorvos	284 (4)
5	ATG_NRF2_ARE_CIS_up	Nrf2 receptor agonism	354	458–37-7	curcumin	No health-based EBT
6	ATG_PXR_TRANS_up	PX receptor agonism	314	51218–45-2	metolachlor	59 (1)
7	TOX21_PXR_Agonist	PX receptor agonism	163	51218–45-2	metolachlor	59 (1)
8	TOX21_ERa_LUC_VM7_Antagonist_0.1nM_E2	Estrogen receptor alpha antagonism	0.25*	84449–90-1	raloxifene	No health-based EBT
9	TOX21_AR_LUC_MDAKB2_Antagonist_0.5nM_R1881	Androgen receptor antagonism	56.4*	13311–84-7	flutamide	4.8 (2)
10	OT_AR_ARSRC1_0960	Androgen receptor agonism	1.13*	521–18-6	androstanolone (DHT)	0.0045 (2) 0.011 (4)
11	OT_AR_ARSRC1_0480	Androgen receptor agonism	1.60*	521–18-6	androstanolone (DHT)	0.0045 (2) 0.011 (4)
12	ATG_PPARg_TRANS_up	PPARg receptor agonism	5.8*	1461–22-9	tributyltin chloride	No health-based EBT
13	TOX21_AhR_LUC_Agonist	Ah receptor agonism	2045*	63–25-2	carbaryl	18 (1)
14	TOX21_PR_BLA_Antagonist_ratio	Progesterone receptor antagonism	0.038	84371–65-3	mifepristone	No health-based EBT
15	TOX21_GR_BLA_Antagonist_ratio	Glucocorticoid receptor antagonism	0.102*	84371–65-3	mifepristone	No health-based EBT
16	TOX21_TR_LUC_GH3_Antagonist	Thyroid hormone receptor antagonism	118*	133–07-3	folpet	No health-based EBT
17	TOX21_p53_BLA_p2_ratio	Cellular tumor antigen p53 activation	0.123*	50–76-0	actinomycin D	No health-based EBT
18	TOX21_ERa_BLA_Antagonist_ratio	Estrogen receptor alpha antagonism	3269*	84449–90-1	raloxifene	No health-based EBT
19	ATG_ERE_CIS_up	Estrogen receptor alpha agonism	1.97	50–28-2	17β-estradiol	No health-based EBT
20	ACEA_ER_80hr	Estrogen receptor agonism	0.118	50–28-2	17β-estradiol	0.0002 (1) 0.0018 (1) 0.0012 (1) 0.00025 (2) 0.0038 (3)

* Not reliable because the EBT is based on few chemicals (<16). References: (1) Escher et al., 2015 (2) Béen et al., 2021 (3) Brand et al., 2015 (4) Escher et al., 2013.

$$REP_i = \frac{AC50_y}{AC50_i} \quad (2)$$

Here REP_i represents the relative effect potency of chemical *i*, $AC50_i$ is the concentration at which half of the maximum activity is reached for chemical *i*, and $AC50_y$ represents this value for the reference chemical (see Table 1 for an overview of reference chemicals per bioassay).

In the second step, the health-related bioanalytical equivalent pGLV of each chemical is calculated. This involves determining the expected pGLV-related bioassay activity expressed as reference chemical equivalents (based on the *REP*).

$$BpGLV_i = pGLV_i \bullet REP_i \quad (3)$$

Here $BpGLV_i$ is the bioanalytical pGLV equivalent of chemical *i*, and REP_i is the relative effect potency of chemical *i*. In Béen et al., *BpGLV* is referred to as pGLV BEQ.

A minor adjustment was made in the EBT derivation for a bioassay compared to the study of Béen et al., in which a normal distribution was fitted to the available activity data for all chemicals per assay, and the lower 5 % of the *BpGLV* distribution was taken as the EBT. This approach was followed because few data points were available in Béen et al., and the caveat was acknowledged that the lowest available *BpGLV* may not be within the lowest 5 % of values. Therefore, this 5 % point was explicitly derived from the distribution. However, because the presently used ToxCast dataset has a more extensive set of data points, the lowest value expectedly falls within the lower 5 % of the distribution. Therefore, the lowest *BpGLV* can now be used to derive the EBT instead of the lower 5 % limit of a distribution.

$$EBT = \text{minimum}(BpGLV_{i-n}) \quad (4)$$

The derived EBTs were compared to EBTs of analogous bioassays with the same endpoint and derived based on the same reference chemical (Table 1).

2.4. Calculation of chemical mixture characteristics

For the simulations of *in silico* mixtures, each mixture was assigned a fixed sum concentration and a fixed number of chemicals. For the selected chemicals in the *in vitro* assays, an additive effect is assumed (Cedergreen et al., 2008; Hadrup et al., 2013). Assuming concentration addition is reasonable because the chemicals were selected for their relevance to the bioassay based on their pGLV (Equation (1)) and are expected to act through the same mode of action within the bioassay. Chemicals in the *in silico* mixtures were selected at random from the tested chemicals with appropriate pGLVs. The chemicals were subsequently appointed a random portion of the sum concentration. All analyses were automated in the statistical language 'R' (v. 3.6.3).

The total equivalent concentration of the chemicals in the mixture is expressed as equivalents of the chemical with the lowest (i.e. most potent) pGLV in the mixture.

$$BCH_{mix} = \sum_n^i \frac{pGLV_x}{pGLV_i} \bullet C_i \quad (5)$$

Here BCH_{mix} is the health-based bioanalytical equivalent concentration of the mixture, $pGLV_x$ is the pGLV of chemical *x*. $pGLV_i$ is the pGLV of chemical *i*. C_i is the concentration of chemical *i*.

As an example, a BCH_{mix} is calculated for two mixtures of each two chemicals A and B at a sum concentration of 1 μg/L.

A: pGLV 5 μg/L, concentration 0.9 μg/L OR pGLV 5 μg/L, concentration 0.1 μg/L.

B: pGLV 100 μg/L, concentration 0.1 μg/L OR pGLV 100 μg/L, concentration 0.9 μg/L.

$$BCH_{mix} = \left(\tfrac{5}{5} \bullet 0.9 + \tfrac{5}{100} \bullet 0.1\right) = 0.905\ \mu g/L \quad \text{OR} \quad BCH_{mix} = \left(\tfrac{5}{5} \bullet 0.1 + \tfrac{5}{100} \bullet 0.9\right) = 0.145\ \mu g/L.$$

The predicted activity of the mixture in bioanalytical equivalents is calculated as the bioanalytical concentration activity (*BCA*):

$$BCA_{mix} = \sum_n^i REP_i \bullet C_i \quad (6)$$

Here REP_i is the potency of chemical *i* (Equation (2) and C_i is the concentration of chemical *i*.

2.5. Statistics for bioassay performance

The bioassay performance is evaluated by calculating the statistics for the assay itself (specificity, sensitivity) and the predictive values considering samples containing *in silico* mixtures of chemicals (one or more).

$$True\ Positive(TP)\ if:\ pGLV_x \le BCH_{mix}\ \&\ BCA_{mix} > EBT \quad (7)$$

$$False\ Negative(FN)\ if:\ pGLV_x \le BCH_{mix}\ \&\ BCA_{mix} < EBT \quad (8)$$

$$False\ Positive(FP)\ if:\ pGLV_x \geqslant BCH_{mix}\ \&\ BCA_{mix} > EBT \quad (9)$$

$$True\ Negative(TN)\ if:\ pGLV_x \geqslant BCH_{mix}\ \&\ BCA_{mix} < EBT \quad (10)$$

From these values, the sensitivity (*Sens*) and specificity (*Spec*) of the assay are calculated:

$$Sens = \frac{TP}{TP + FN} \quad (11)$$

$$Spec = \frac{TN}{TN + FP} \quad (12)$$

Prevalence (*Prev*) or occurrence is the number of samples with $pGLV_x \le BCH_{mix}$ divided by the total number of samples.

$$\text{Prev} = \frac{samples\ pGLVx \le BCH_{mix}}{samples} \quad (13)$$

Then, the probabilities of a positive test ($BCA_{mix} > EBT$) to signal a sample with a harmful chemical mixture concentration exceeding the pGLV ($BCH_{mix} > pGLV_x$) (positive predictive value; PPV) and the probabilities of a negative test to signal a safe sample (negative predictive value; NPV) are calculated by:

$$PPV = \frac{Prev \bullet Sens}{Prev \bullet Sens + (Prev \bullet Sens + ((1 - Prev) \bullet (1 - Spec)))} \text{ OR } PPV = \frac{TP}{TP + FP} \quad (14)$$

$$NPV = \frac{(1 - Prev) \bullet Spec}{(1 - Prev) \bullet Spec + Prev \bullet (1 - Sens)} \text{ OR } NPV = \frac{TN}{TN + FN} \quad (15)$$

The added value of PPV and NPV to sensitivity and specificity can be illustrated with an example. Assume the sensitivity of a bioassay is 0.9, meaning that nine out of ten chemicals at harmful concentrations elicit a response above EBT in the bioassay. Similarly, assume the specificity is 0.9 as well, indicating that nine out of ten times chemicals below the harmful concentration do not elicit a response above the EBT. Now, consider a scenario with 1100 tested samples with chemicals of which 100 are harmful and 1000 are not, at the tested concentration. If a test result is positive (> EBT), this corresponds to 1000 x 0.1 (=100) false positives and 0.9 x 100 (=90) true positives. The positive predictive value is 90 out of 190 (100 + 90) resulting in a value of 0.47. This illustrates that the probability that a chemical (mixture) that elicits a positive response is indeed harmful, is a little less than 50 % for this evaluated bioassay, even with a high sensitivity and specificity.

Lastly, the proportion of samples for which a bioassay responds above the EBT ('hit') is calculated as

$$Hit_B = \frac{FP + TP}{FN + TN + FP + TP} \quad (16)$$

Here Hit_B is the proportion of samples with responses above the EBT relative to all possible samples for assay *B*. The term 'hit' as it is used in this study should not be confused with the term 'hit call' in ToxCast, where it refers to a label for (in)active tested chemicals in the bioassay.

2.6. Experiments with bioassay performance

Three *in silico* experiments were executed to evaluate the performance of the selected bioassays. First, performance was assessed for single tested chemicals at concentrations ranging from 0.01-10 μg/L. Chemical concentrations influence the outcome of equations (5)–(16). The higher the BCA_{mix}, the likelier an EBT exceedance, which will increase the number of both false and true positives. Simultaneously, the chemical mixture concentration BCH_{mix} (in pGLV equivalents) will surpass its human health-relevant pGLV leading to a higher number of true positives. The maximum sum concentration of 10 μg/L is relatively high in a drinking water context, which was intentionally chosen to cover the complete range of possible PPVs and NPVs for the investigated bioassays.

Second, performance based on single chemicals tested in ToxCast was evaluated for a range of 'hit-thresholds' of 0.1-10x the EBT. The hit-threshold is an alternative trigger value (also see Leusch & Snyder, 2015). Whereas the EBT was derived based on health-based values (Equation (4)) the hit-threshold represents an assignable level for follow-up action (like analysing mixture composition) to bioassay responses, which can be used to tailor assay performance. If the hit-threshold is lower than the EBT, more hits (Equation (16)) are expected, and vice-versa. The range of 0.1-10x EBT for hit thresholds was based on a Dutch study, in which the risk of chemicals at harmful concentrations was categorized into classes ranging from a response < EBT (low-risk class) to a response > 10x EBT (high-risk class) (de Baat et al., 2021). For this second experiment, the concentration of chemicals was fixed at 10 μg/L. At lower (more realistic) concentrations, the number of bioanalytical responses > EBT was in some cases too low to calculate performance.

Third, performance was evaluated for *in silico* mixtures randomly generated from all chemicals with a relevant pGLV for each bioassay. Each time, 50 *in silico* mixtures were randomly generated for a range of sum concentrations (0.01–10 μg/L) and a fixed number of chemicals in the mixture (1, 2, 5, 10). The concentrations of the individual chemicals were randomly assigned within the mixture, adding up to the fixed sum concentration. Chemicals could be selected more than once per mixture (i.e. sampling was done with replacement).

3. Results

3.1. EBT derivation

Human-relevant health-based EBTs were derived for the 20 assays selected from the ToxCast database (See Methods section). If the EBTs were derived based on fewer than 16 chemicals, the EBT was flagged (*) as less reliable (Table 1).

3.2. Selection of assays for performance analysis

Relatively few of the tested chemicals (regardless of their pGLV) showed an expected bioanalytical response (*BCA*, Equation (6)) above the EBT at a concentration of 1 μg/L. This percentage ranges from 0 to 19 % depending on the assay (Table 2). At 10 μg/L, this percentage increased for all assays and ranged from 1 to 42 % (Table 2).

Six of the evaluated assays were identified as the most sensitive for detecting harmful chemicals in drinking water (sources) (Table 2) *and* are supported by enough tested chemicals with appropriate pGLVs (SI I Table A). To illustrate the findings and applicability of the derived EBTs, the remainder of the current study is focused on only two of the investigated assays. To cover the range from worst to best case scenario, one is a relatively poorly performing assay (*TOX21_PR_BLA_Antagonist_ratio)* and the other is a well-performing assay (*TOX21_ERa_LUC_VM7_Agonist*). The figures for the other four selected assays can be found in SI II-V.

3.3. Bioassay performance with data for single tested chemicals

At any assumed chemical (mixture) concentration a response in the bioassay is either TP, TN, FP, or FN (Equations (7)–(10)). This is depicted for a concentration of 1 μg/L of the tested chemicals for both example assays (Fig. 1) as an illustration. Chemicals that were included in the derivation of the pGLV (dark blue), as well as chemicals that were not (light blue) (Equation (1)), are shown. Most of the selected (dark blue) chemicals that do not elicit an effect in the bioassay are indeed not harmful at that concentration (i.e., do not exceed their pGLV) and are

Table 2
Selected ToxCast assays with derived EBTs and sensitivity to signal chemical health risks (>EBT) in water.

	Assay name	% chemicals with effect > EBT at 1 μg/L	% chemicals with effect > EBT at 10 μg/L	Sensitive for drinking water (sources)?	Selection for performance analysis
1	ATG_ERa_TRANS_up	5	11	Yes	Selected
2	TOX21_ERa_LUC_VM7_Agonist	10	15	Yes	Selected
3	TOX21_AR_BLA_Antagonist_ratio	4	11	Yes	Selected
4	TOX21_ARE_BLA_agonist_ratio	1	3	No [#] [*]	No
5	ATG_NRF2_ARE_CIS_up	0	3	No [#]	No
6	ATG_PXR_TRANS_up	0	1	No [#]	No
7	TOX21_PXR_Agonist	0	1	No [#]	No
8	TOX21_ERa_LUC_VM7_Antagonist_0.1nM_E2	1	5	No [#] [*]	No
9	TOX21_AR_LUC_MDAKB2_Antagonist_0.5nM_R1881	1	5	No [#] [*]	No
10	OT_AR_ARSRC1_0960	16	25	Yes [*]	No
11	OT_AR_ARSRC1_0480	16	31	Yes [*]	No
12	ATG_PPARg_TRANS_up	1	2	No [#] [*]	No
13	TOX21_AhR_LUC_Agonist	0	2	No [#] [*]	No
14	TOX21_PR_BLA_Antagonist_ratio	2	7	Yes	Selected
15	TOX21_GR_BLA_Antagonist_ratio	3	7	Yes [*]	No
16	TOX21_TR_LUC_GH3_Antagonist	2	6	No [#] [*]	No
17	TOX21_p53_BLA_p2_ratio	5	12	Yes [*]	No
18	TOX21_ERa_BLA_Antagonist_ratio	3	5	No [#] [*]	No
19	ATG_ERE_CIS_up	6	12	Yes	Selected
20	ACEA_ER_80hr	19	42	Yes	Selected

[*] The EBT is not reliable because it is based on too few chemicals (<16) (see SI I Table A).
[#] At 10 μg/L, 6 % or less of the tested chemicals induced a bioanalytical response above the EBT.

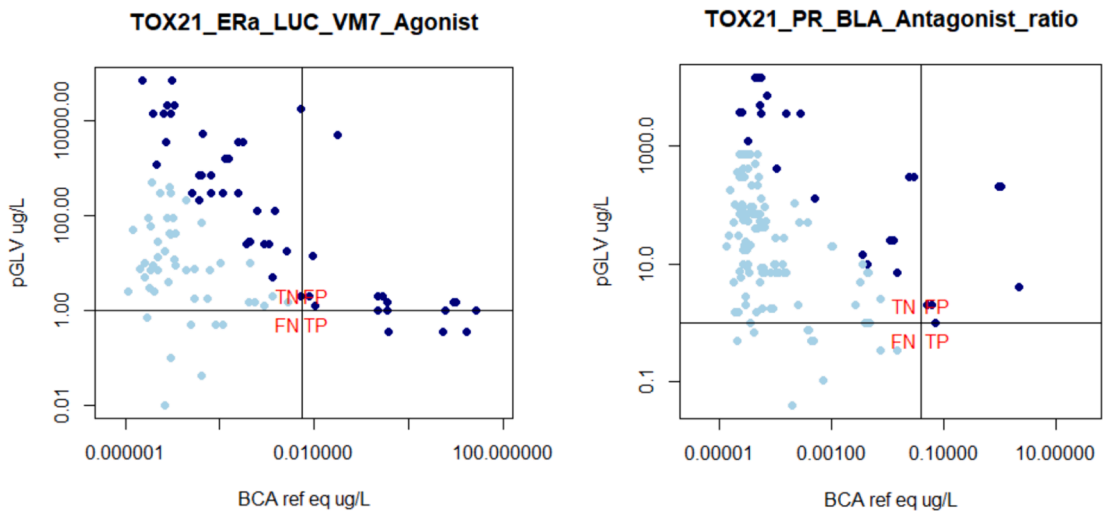

Fig. 1. The bioanalytical concentration activity (BCA) of chemicals at 1 μg/L (Equation (6)) versus pGLV and their status as true/false positive/negative in two example bioassays for selected (dark dots) and non-selected (light dots) chemicals (Equation (1)). Axes are on a natural log-scale. The horizontal line depicts 1 μg/L and the vertical line depicts the derived EBT.

thus true negatives (TN) (Fig. 1). None of the selected chemicals are FN at 1 μg/L. In the remaining analyses, only selected chemicals (Equation (1)) were considered.

The sensitivity and specificity of the bioassays were calculated (considering only selected chemicals) for a range of concentrations between 0.01–10 μg/L. Specificity and sensitivity are comparable in the two assays (Fig. 2a,c). The sensitivity in both bioassays is 1 (i.e., 'perfect') at all concentrations. In other words, when a chemical is present at a harmful concentration, the bioassay elicits a response above the EBT (i.e. no 'false negative'). With regards to the specificity, a portion of the chemicals at a concentration below their respective pGLV elicit a false positive bioanalytical response (above EBT).

The PPV is variable along the assumed concentrations (Fig. 2 c,d). For the *TOX21_ERa_LUC_VM7_Agonist*, PPV has a steep increase at lower concentrations, starting from 0.2 and increasing to 0.4–0.9 around 2 μg/L, implying an increase in the number of true positives compared to the

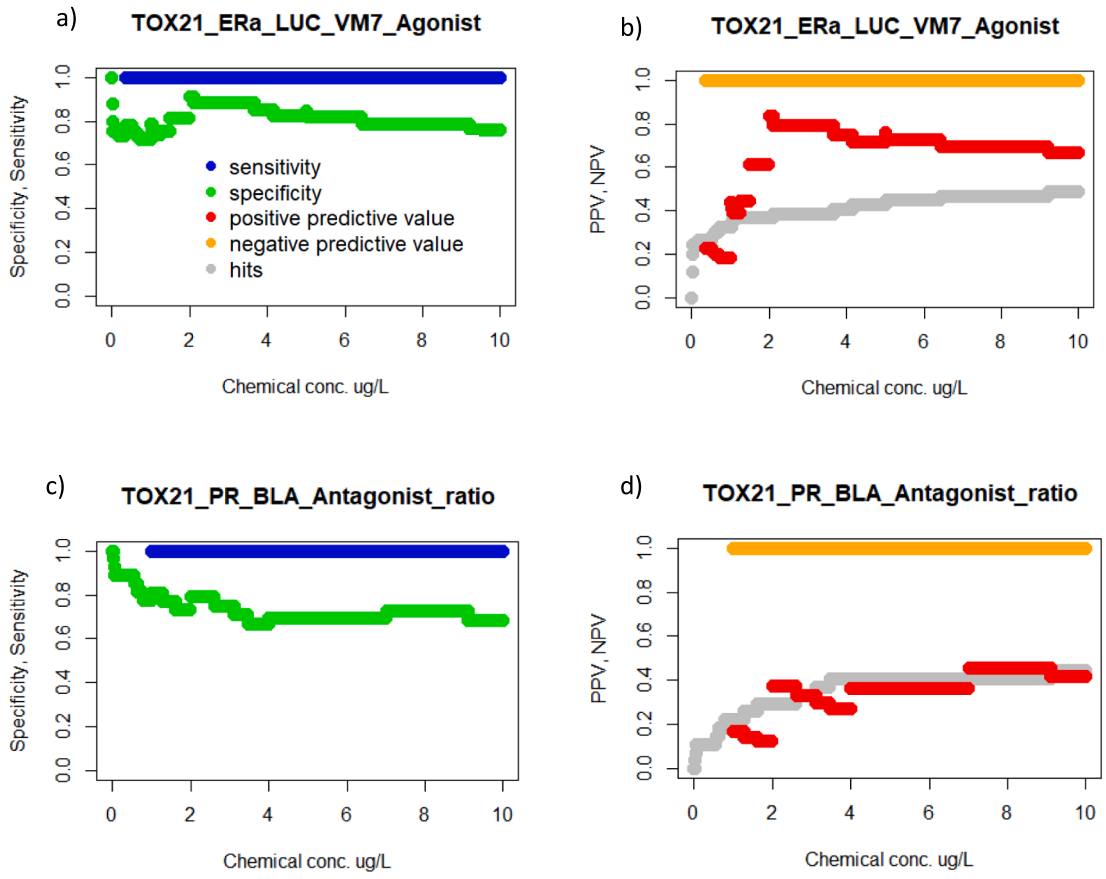

Fig. 2. Sensitivity, Specificity (a,c), Positive, Negative Predictive Value (b,d) for two example bioassays with individually tested chemicals. See Fig. 1 for an explicit visualization of FN, FP, TN, TP for the selected (Equation (1)) chemicals in these assays at 1 ug/L.

number of false positives, followed by a gradual decline at increasing concentrations. At the lowest concentrations, there are no true positives (not shown). Consequently, the sensitivity and the PPV cannot be calculated for those low concentrations. The PPV for the *TOX21_ERa_LUC_VM7_Agonist* is generally higher than for the *TOX21_PR_BLA_Antagonist_ratio*, and the PPV and NPV broadly follow a similar pattern for both bioassays. Similar results were obtained for the other selected bioassays (SI III).

The change in PPV with increasing chemical concentration can be explained as follows. At higher concentrations, chemicals approach, and in some instances exceed, their pGLV. Concurrently, the bioassay response increases with concentration and more frequently exceeds the EBT. The sharp increase in PPV is therefore caused by an increase in true positive (TP) chemicals (exceedance of the EBT and pGLV) compared to false positive (FP) chemicals (exceedance of the EBT but not pGLV). A comparison of the results with random assigned pGLV to chemicals can be found in SI IV.

The proportion of 'hits' (Equation (16)) increases with increasing chemical concentration. This is caused by the combined higher number of TP and FP. The proportion of hits is similar in both bioassays. Because the PPV is higher in the *TOX21_ERa_LUC_VM7_Agonist*, the bioassay 'hits' are more likely to appropriately signal a sample containing a harmful chemical.

3.4. Bioassay performance at different hit-thresholds

The bioassay performance was also assessed for variable 'hit-threshold' values of 0.1 up to 10 times the EBT (x-axis, Fig. 3). The difference between the EBT and the hit-threshold values is that the EBT is derived based on health-based values (Equation (4)), while the hit-thresholds are other values of the bioassay response used for calculation of the bioassay performance.

At higher hit-thresholds (>1 x EBT), the sensitivity drops whereas the specificity increases (Fig. 3a,c). This is reflected in a drop in NPV, whereas the PPV increases (Fig. 3b,d). This increase in PPV can be explained by the decrease in the number of false positives at higher hit-thresholds. False positives are chemicals for which concentrations exceed the hit-threshold but not their respective pGLV. At higher hit-thresholds, this phenomenon is negated. True positives also decrease slightly when higher hit-thresholds are applied but less strongly than the decrease in false positives.

The NPV decreases with increasing hit-thresholds because false negatives start to appear, caused by chemicals for which concentrations exceed their pGLV but do not trigger an effect > hit-threshold. The NPV is only slightly impacted because there are many more true negatives compared to false negatives. The true negatives increase with a higher hit-threshold, mostly attributable to a decrease in false positives (Fig. 3b,d).

In addition, higher hit-thresholds decrease the ratio of positives (above threshold responses or 'hits') in the assays (Fig. 3b,d). This means a water manager will be required to follow up less frequently based on the bioanalytical response at higher hit-thresholds. And, especially for the *TOX21_ERa_LUC_VM7_Agonist*, these sparse hits are increasingly worthwhile to follow up (high PPV) (Fig. 3b). Similar results were obtained for the other selected bioassays (SI V).

3.5. Bioassay performance for chemical mixtures

For the subsequent analysis, 50 mixture compositions were generated *in silico* (Equations (5) and (6)) per total mixture concentration and

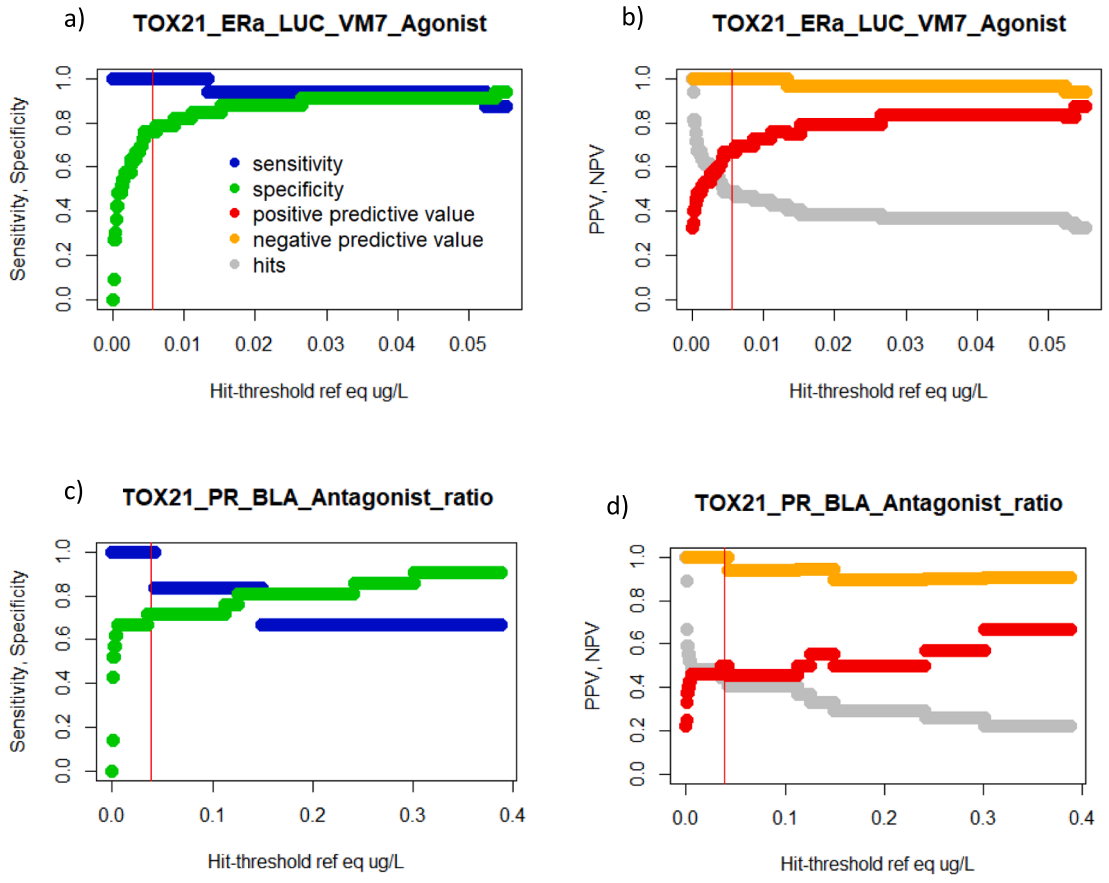

Fig. 3. Performance for two example assays at chemical concentrations of 10 μg/L, at varying hit-thresholds from 0.1 up to 10 times the health-based EBT (Table 1) (vertical red line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the NPVs, PPVs, and hits were calculated for assays exposed to each mixture concentration in steps of 0.01 μg/L (Fig. 4). Because the chemical composition of the 50 mixtures is different for each mixture concentration and the concentration steps are small, results appear like a cloud of values. The leftmost subfigures in Fig. 4 are comparable with the results in Fig. 2b,d with the only difference between the figures being the calculations: Fig. 2b,d is based on concentrations of all individual chemicals with relevant pGLVs whereas in the leftmost panels in Fig. 4, 50 simulations are with compositions of randomly selected individual chemicals with relevant (Equation (1)) pGLVs.

The NPV is 1 ('perfect') for all mixtures at all concentrations while the PPV generally starts low and rises with increasing mixture concentrations (Fig. 4). The higher the number of chemicals in the mixture, the higher the total mixture concentration at which PPVs start to increase. The variability in NPV and PPV becomes lower with a higher number of mixture constituents. This is, however, a consequence of the mixtures becoming more alike with more chemicals selected from the same group of selected, tested chemicals in the bioassay.

Compared to single chemicals, the number of hits is highest for the most diverse (i.e. including the most chemicals) mixtures and the PPV is lower. This is caused by the combined high number of TP and (more) FP (see SI VI). For the most diverse mixtures at the higher concentrations, almost all bioanalytical responses are a 'hit'. For the *TOX21_PR_BLA_Antagonist_ratio* this means that, at low chemical mixture concentrations (<5 μg/L), the required follow-ups by water managers (i.e., chemical analysis of target chemicals and comparison of the detected concentrations to pGLVs) in most cases would not yield pGLV exceedances and thus human health risks. For the other assays, the same applies, but to a lesser extent (not shown).

4. Discussion

Determining positive and negative predictive values (PPV and NPV) (Johnson, 2017) is common practice in assessing test performance in the medical domain (Webb and Sidebotham, 2020). However, it has not been applied to assess the performance of bioassays until now. The calculation of PPV and NPV provides important additional information on bioassay interpretation and performance. The PPV statistics quantify the probability that a tested sample containing a chemical (mixture) actually exceeds its safe pGLV if the bioanalytical response exceeds the EBT. Similarly, the NPV quantifies the probability that a sample containing a chemical (mixture) concentration is indeed safe if the bioassay response remains below the EBT.

To perform these analyses, we utilised data on single tested chemicals in six human-relevant *in vitro* assays from the ToxCast database (US EPA, 2023). Additionally, we derived *in silico* chemical mixtures based on the single chemical data. Six assays were selected to simulate their use as bioassays for effect-based drinking water (source) quality assessment. Among the selected assays, the best-performing bioassay at different chemical concentrations was *TOX21_ERa_LUC_VM7_Agonist* while the least-performing bioassay was *TOX21_PR_BLA_Antagonist_ratio*. Some assays were excluded due to a potential lack of sensitivity or an insufficient number of tested chemicals with a pGLV. With more pGLVs and data on individual tested chemicals becoming available, additional bioassays can undergo a similar analysis of their performance in effect-based water quality assessment. The required assay performance for effective use can be decided by water managers, depending on the expected chemical mixtures, the expected levels of pollution, and the intended use (e.g. drinking water, irrigation). The minimum required performance will depend on a preferred balance between the effort to further investigate the identity and concentrations of chemicals in the mixture versus the probability that a bioassay correctly signals a harmful chemical (mixture) concentration. For example, if the follow-up investigation requires little effort, and the proportion of 'hits' (responses > EBT) is low, the PPV can be allowed to be low. The framework presented here allows for the derivation of the values that are required for water managers to make these informed decisions.

The analyses required the derivation of health-based EBTs (Brand et al., 2013; Escher et al., 2015; Béen et al., 2021) for the assays. To recap, the EBT is established based on a health-based safety threshold. Above this threshold, health risks cannot be excluded (Brand et al., 2013; Escher et al., 2015; Béen et al., 2021). This way, the EBT allows the evaluation of assay performance in signalling harmful compound concentrations. In some cases, the derived EBTs deviated from

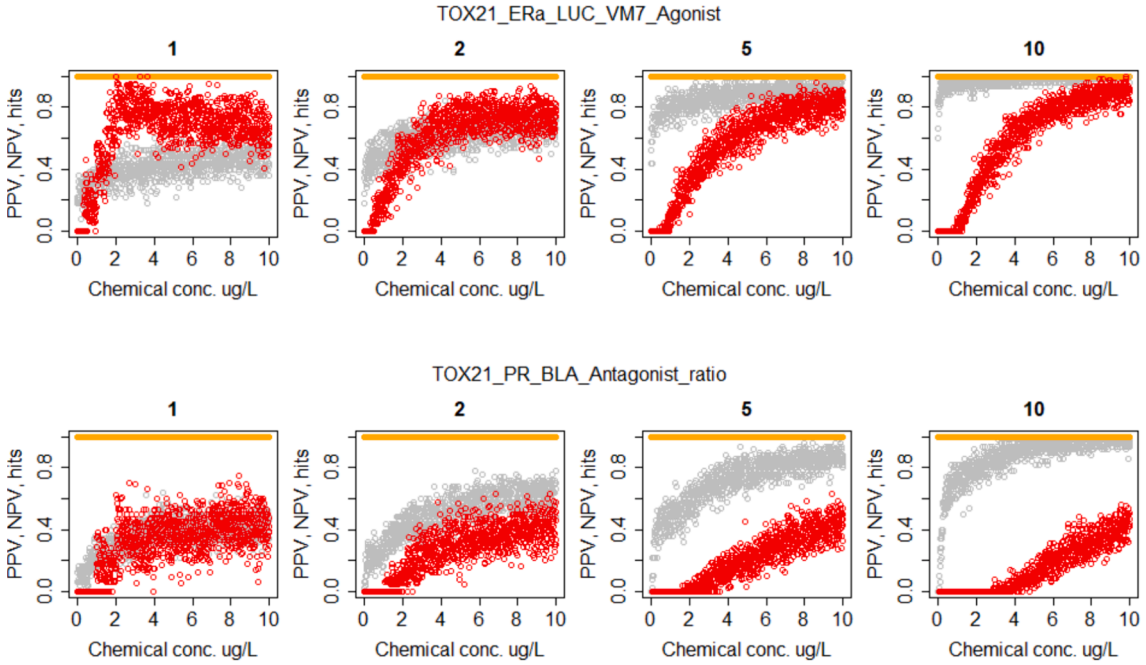

Fig. 4. The positive predictive value (PPV) (red circles), negative predictive value (NPV) (orange circles), and 'hits' (grey circles) at increasing concentrations for each time 50 *in silico* samples containing random mixtures of 1, 2, 5, or 10 chemicals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

analogous bioassays with the same reference compound. It can be expected that analogous bioassays may not have exactly the same EBT as bioassays may differ in their sensitivity (e.g. observed effect size in response to a given chemical) to the specific reference compound or other compounds (Neale et al., 2023b). E.g. if an assay is less sensitive to the reference compound, relative effect potencies of other compounds are higher, leading to a higher EBT value. Moreover, EBTs of different analogous assays may be based on a different set of tested compounds and the lowest combination of pGLV and REP could have been absent in any analogous assay.

In all six selected bioassays, the PPV increased sharply with increasing concentrations of the tested single chemicals, reaching a gradually declining plateau from approximately 2 μg/L onwards. This means that an EBT exceedance at low concentrations does not necessarily signify a chemical risk, but that this becomes a more reliable indicator of pGLV exceedances at concentrations at or above 2 μg/L, as true positives start to increase. Contrastingly, the NPV was perfect at all concentrations (evaluated up to 10 μg/L), which is perhaps even more important from a risk assessment perspective, as chemical concentrations labelled as harmless truly are. With a perfect NPV, it can be assumed that there is no chemical present at a harmful concentration in the sample if the bioanalytical response remains below the EBT. In other words, a 'safe' bioanalytical response reliably signifies the absence of a chemical risk. This underlines the added value of EBM in addition to chemical analyses for chemical risk screening of (drinking) water.

Replacing the EBT with a higher (non-health-based) hit-threshold decreased the ratio of positive (above hit-threshold, 'hits') to negative (below hit-threshold) responses in the assay. This results in fewer false positives and thus fewer instances in which water managers are required to do follow-up analyses based on the assay outcomes. Alongside this decrease in 'hits', there is a higher PPV. However, the NPV is no longer perfect, which means that some chemical concentrations lead to false negative bioanalytical responses and a harmful chemical (mixture) could go unnoticed, which is undesirable. Hence, a threshold higher than the EBT can be used to achieve a higher PPV, making investigating harmful chemical concentrations after a hit more worthwhile. Yet, this also renders the bioanalysis unable to definitively rule out risk, as false negatives start to occur.

Follow-up actions to bioanalytical EBT exceedances constitute multiple options. Further investigations to identify chemicals at harmful concentrations include chemical analytical target, suspect, and non-target analyses (Brunner et al., 2020; Hollender et al., 2023; Renaud et al., 2022). Furthermore, more elaborate effect-based approaches can provide insight into chemical mixture hazards such as the application of a more expansive bioanalytical test battery or the identification of causative compounds using effect-directed analysis (EDA) (Tian et al., 2023; Brack et al., 2016; Zwart et al., 2018). Once concentrations of potentially harmful chemicals have been detected, these can be compared to health-based guideline values or used to calculate the contribution to the observed effects. Ultimately, chemical mixture risks detected by EBT exceedances may result in a requirement for optimization of the water treatment process or mitigation of pollution at the source. To guide water managers in these follow-up actions, Neale et al., (2023a) developed an interpretation framework in which the magnitude of the response is moderated by the magnitude of the exceedance.

The bioassay PPV decreases with increasingly diverse mixtures of chemicals and the number of hits increases. This may seem counterintuitive but results from the disproportionate influence of potent chemicals in a mixture. A collection of mixtures with even a small portion of a highly potent chemical has inherently higher induced bioanalytical responses than a collection of individual chemicals at the same concentration, leading to a higher proportion of false positives. Simultaneously, mixtures are inherently more toxic due to the disproportionally high contribution of potent hazardous (low pGLV) chemicals, increasing the number of true positives. Together, the false and true positives result in more hits. This can be illustrated with a thought experiment: suppose there are five chemicals that, when individually tested, yield four negative (< EBT) and one positive (> EBT) bioanalytical response. The positively tested chemical is 100 times more potent than the other chemicals. If the chemicals form a mixture, this mixture is around 20 times more potent than four of the five tested chemicals individually and will cause a positive (> EBT) response. The positive rate is thus replaced from 20 % to 100 % for the same five compounds when present in a mixture rather than individually.

Combined with the finding that the PPV is low at lower chemical concentrations, this leads to the conclusion that the ability of the bioassays to accurately signal a harmful chemical (mixture) at responses > EBT is lowest at highly diverse low mixture concentrations. This provides a less optimal scenario for water managers, considering that most samples of drinking water (sources) will contain low-level mixtures with diverse chemical constituents. For instance, the median concentration of individual chemicals in the river Rhine is about 0.01 μg/L (10 ng/L) when considering chemicals measured above the limit of quantification (RIWA-Rijn, 2022), and many chemicals are detected at this low concentration. The implications for risk assessment of drinking water (sources) with bioassays is that it is to be expected that in many cases a positive response (> EBT) in a bioassay will not yield chemicals at harmful concentrations after further investigation. Nonetheless, most bioassays hold some positive predictive value even at these low concentration-diverse mixtures and the negative predictive power is 100 %, highlighting the value of EBM for chemical risk screening purposes. Evaluating the mixtures only at higher exceedances of the EBT can be expected to yield harmful concentrations more often in further investigations because the PPV increases, and a pragmatic solution may lie in the application of increased hit-thresholds for follow-up investigations as was explored in the present study.

A caveat in the presented results is the assumption that the chemicals with relevant pGLVs tested in the ToxCast database are a representative sample of all relevant chemicals, with no selection for water-relevant chemicals. Additionally, results are based on *in silico* mixtures at fixed sum concentrations. In practice, it is unknown exactly which and how many chemicals induce the effect and at which concentrations they are present. Therefore, water managers additionally need knowledge of the nature and concentrations of chemicals to estimate the expected bioassay performance in their water systems.

5. Conclusions

The positive predictive value of bioassays for chemical health-based guideline value exceedances may be low, especially when exposed to highly diverse low-concentration mixtures. Nevertheless, some predictive power is maintained, and the negative predictive power is perfect. This means that the bioassays can still identify potential risks even if the probability is small, and the absence of a bioanalytical response reliably confirms that no individual- or mixture-pGLV is exceeded. Replacing the EBT with a higher hit-threshold will increase the PPV as well as lower the proportion of positive bioanalytical responses that must be followed up. Unfortunately, the NPV is no longer perfect in that case and chemicals at harmful concentrations may be missed, which poses a trade-off between certainty and feasibility that may be best made at a regulatory level. It can be considered to use the EBT as a threshold to ensure bioanalytical responses < EBT are indeed of low risk and using a higher hit-threshold to optimize the PPV. The presented framework provides insight into what to expect from bioassays in water quality assessment and serves as a means to select bioassays that have a relatively large probability of signaling harmful chemical concentrations when their response exceeds the EBT. This study is the first to explore the concept of predictive values of bioassays for chemical health risks in drinking water and provides the opportunity for the results to be further validated with case studies.

CRediT authorship contribution statement

Tessa E. Pronk: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Renske P.J. Hoondert:** Writing – review & editing, Data curation. **Stefan A.E. Kools:** Writing – review & editing, Methodology. **Vikas Kumar:** Writing – review & editing. **Milo L. de Baat:** Writing – review & editing, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Open data was used and other data will be shared as supplementary data

Acknowledgments

This work was carried out in the framework of the European Partnership for the Assessment of Risks from Chemicals (PARC) and has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101057014. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. The authors thank Sebastian Buchinger (BAFG) for helpful suggestions and considerations regarding the analyses and reporting. Sebastiaan Wesseling (WUR) is acknowledged for helpful initial discussions.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2024.108733.

References

Australian drinking water guidelines (2022) https://www.nhmrc.gov.au/sites/default/files/documents/attachments/publications/Australian_Drinking_Water_Guidelines_ADWG_V3-8_Sep2022.pdf.

Baken, K.A., Sjerps, R.M.A., Schriks, M., van Annemarie, P., 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.

Been, F., Pronk, T., Louisse, J., Houtman, C., van der Velden-Slootweg, T., van der Oost, R., Dingemans, M.M.L., 2021. Development of a framework to derive effect-based trigger values to interpret CALUX data for drinking water quality. Water Res. 193, 116859 https://doi.org/10.1016/j.watres.2021.116859.

Brack, W., Ait-Aissa, S., Burgess, R.M., et al., 2016. Effect-directed analysis supporting monitoring of aquatic environments — An in-depth overview. Sci. Total Environ. 544, 1073–1118. https://doi.org/10.1016/j.scitotenv.2015.11.102.

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.

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

Brunner, A.M., Bertelkamp, C., Dingemans, M.M.L., Kolkman, A., Wols, B., Harmsen, D., Siegers, W., Martijn, B.J., Oorthuizen, W.A., Ter Laak, T.L., 2020. Integration of target analyses, non-target screening and effect-based monitoring to assess OMP related water quality changes in drinking water treatment. Sci. Total Environ. 705, 135779 https://doi.org/10.1016/j.scitotenv.2019.135779.

Cedergreen, N., Christensen, A.M., Kamper, A., Kudsk, P., Mathiassen, S.K., Streibig, J.C., Sørensen, H., 2008. A review of independent action compared to concentration addition as reference models for mixtures of chemicals with different molecular target sites. Environ. Toxicol. Chem. 27 (7), 1621–1632. https://doi.org/10.1897/07-474.1.

CIS working group EBM (2021) Technical Proposal for Effect-Based Monitoring and Assessment under the Water Framework Directive. Report to the Common Implementation Strategy (CIS) Working Group Chemicals on the outcome of the work performed in the subgroup on Effect-Based Methods (EBM).

de Baat, M.L., T.E. Pronk, S.J.P. van den Berg, R. van der Oost (2021) Kalibratie van de risicogrenzen voor interpretatie van bioassays. Versie 1, 21 december 2021. KIWK-Toxiciteit Notitie. Amersfoort, Nederland. Kennis Impuls Water Kwaliteit. https://www.sleutelfactortoxiciteit.nl/sites/default/files/2023-01/SFTox%20document%20achtergrond%20calibraties%20bioassays%20kwaliteitsklassen%20versie%201%281%29.pdf.

Drinking water directive (2023). Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 on the quality of water intended for human consumption (recast) (Text with EEA relevance). https://eur-lex.europa.eu/eli/dir/2020/2184/oj.

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. Tech. 47 (13), 7002–7011.

Escher, B.I., Neale, P.A., Leusch, F.D., 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., Blanco, J., Caixach, J., et al., 2023. In vitro bioassays for monitoring drinking water quality of tap water, domestic filtration and bottled water. J. Expo. Sci. Environ. Epidemiol. https://doi.org/10.1038/s41370-023-00566-6.

Feshuk, M., Kolaczkowski, L., Dunham, K., Davidson-Fritz, S.E., Carstens, K.E., Brown, J., Judson, R.S., Paul Friedman, K., 2023. The ToxCast pipeline: updates to curve-fitting approaches and database structure. Front. Toxicol. 5, 1275980. https://doi.org/10.3389/ftox.2023.1275980.

Filer, D. L., Kothiya, P., Setzer, W. R., Judson, R. S., & Martin, M. T. (2014). The ToxCast analysis pipeline: An R package for processing and modeling chemical screening data. US Environmental Protection Agency: http://www. epa. gov/ncct/toxcast/files/MySQL% 20Database/Pipeline_Overview. pdf.

Groothuis, F.A., Heringa, M.B., Nicol, B., Hermens, J.L.M., Blaauboer, B.J., Kramer, N.I., 2015. Dose metric considerations in in vitro assays to improve quantitative in vitro–in vivo dose extrapolations. Toxicology 332, 30–40. https://doi.org/10.1016/j.tox.2013.08.012.

Hadrup, N., Taxvig, C., Pedersen, M., Nellemann, C., Hass, U., Vinggaard, A.M., 2013. Concentration addition, independent action and generalized concentration addition models for mixture effect prediction of sex hormone synthesis in vitro. PLoS One 8 (8), e70490. https://doi.org/10.1371/journal.pone.0070490.

Hollender, J., Schymanski, E.L., Ahrens, L., et al., 2023. NORMAN guidance on suspect and non-target screening in environmental monitoring. Environ. Sci. Eur. 35, 75. https://doi.org/10.1186/s12302-023-00779-4.

Johnson KM (2017) Using Bayes' rule in diagnostic testing: a graphical explanation. From the journal Diagnosis https://doi.org/10.1515/dx-2017-0011.

Jonker, M.T., Van der Heijden, S.A., 2007. Bioconcentration factor hydrophobicity cutoff: An artificial phenomenon reconstructed. Environ. Sci. Tech. 41 (21), 7363–7369.

Leusch, F.D.L., Snyder, S.A., 2015. Bioanalytical tools: Half a century of application for potable reuse. Environ. Sci.: Water Res. Technol. 1, 606–621.

Neale, P.A., Escher, B.I., de Baat, M.L., Enault, J., Leusch, F.D.L., 2023a. Effect-based trigger values are essential for the uptake of effect-based methods in water safety planning. Environ. Toxicol. Chem. 42 (3), 714–726. https://doi.org/10.1002/etc.5544.

Neale, P.A., Escher, B.I., de Baat, M.L., Dechesne, M., Dingemans, M.M.L., Enault, J., Pronk, G.J., Smeets, P.W.M.H., Leusch, F.D.L., 2023b. Application of effect-based methods to water quality monitoring: Answering frequently asked questions by water quality managers, regulators, and policy makers. Environ. Sci. Technol. 57 (15), 6023–6032. https://doi.org/10.1021/acs.est.2c06365.

Neale PA, Leusch F, Escher B (2021) Bioanalytical Tools in Water Quality Assessment. IWA publishing. https://library.oapen.org/handle/20.500.12657/50078.

Oskarsson, A., Rosenmai, A.K., Mandava, G., Johannisson, A., Holmes, A., Tröger, R., Lundqvist, J., 2021. Assessment of source and treated water quality in seven drinking water treatment plants by in vitro bioassays - Oxidative stress and antiandrogenic effects after artificial infiltration. Sci. Total Environ. 758, 144001 https://doi.org/10.1016/j.scitotenv.2020.144001.

Renaud, J.B., Sabourin, L., Hoogstra, S., Helm, P., Lapen, D.R., Sumarah, M.W., 2022. Monitoring of environmental contaminants in mixed-use watersheds combining targeted and nontargeted analysis with passive sampling. Environ. Toxicol. Chem. 41 (5), 1131–1143. https://doi.org/10.1002/etc.5192.

Ryan, N., & Becker, R. (2017). A user's guide for accessing and interpreting ToxCast data: Bayer.

Swanson, S.J., Chirmule, N., 2009. Assessing specificity for immunogenicity assays. Bioanalysis 1 (3), 611–617. https://doi.org/10.4155/bio.09.41.

Tian Z, McMinn MH, Fang M (2023) Effect-directed analysis and beyond: how to find causal environmental toxicants. Exposome 3:1, osad002, https://doi.org/10.1093/exposome/osad002.

U.S. EPA (2014). US EPA TOXCAST DATA RELEASE ASSAY QUALITY SUMMARY OCTOBER 2014.

U.S. EPA (2023) ToxCast & Tox21 Summary Files from invitrodb_v3. Retrieved from https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data on October 28, 2023. Data released October 2018.

van den Berg, A.E.T. (2021) Refining the Water Quality Index and Defining its Relation to Human Health Risks. KWR water research institute Student report, unpublished.

Webb, M.P.K., Sidebotham, D., 2020. Bayes' formula: A powerful but counterintuitive tool for medical decision-making. BJA Educ. 20 (6), 208–213. https://doi.org/10.1016/j.bjae.2020.03.002.

Zhang, J.H., Chung, T.D., Oldenburg, K.R., 1999. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screen 4 (2), 67–73.

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 (7), 4367–4377. https://doi.org/10.1021/acs.est.7b06604.