Global Water Research Coalition

Tools to Detect Estrogenic Activity in Environmental Waters

Global Water
Research Coalition

IWA affiliate

TOOLS TO DETECT ESTROGENIC ACTIVITY IN ENVIRONMENTAL WATERS

by:

Frederic DL Leusch, Ph.D.
CRC Water Quality and Treatment

2008

Global Water Research Coalition: Global cooperation for the generation of water knowledge

GWRC is a non-profit organization that serves as the collaborative mechanism for water research. The product the GWRC offers its members is water research information and knowledge. The Coalition will focus in water supply and wastewater issues and renewable water resources: the urban water cycle.

The founding members of the GWRC are: the Awwa Research Foundation (US), CRC Water Quality and Treatment (Australia), EAWAG (Switzerland), Kiwa (Netherlands), Suez Environment – CIRSEE (France), Stowa – Foundation for Applied Water Research (Netherlands), PUB—Singapore, DVGW – TZW Water Technology Centre (Germany), UK Water Industry Research (UK), Veolia – Anjou Recherche (France), Water Environment Research Foundation (US), Water Research Commission (South Africa), Water Reuse Foundation and the Water Services Association of Australia.

These organizations are all in charge of a national research program addressing the different parts of the water cycle. They have provided the impetus, credibility, and initial funding for the GWRC. Each brings a unique set of skills and knowledge to the Coalition. Through its member organizations GWRC represents the interests and needs of 500 million consumers.

The Global Water Research Coalition is affiliated with the International Water Association (IWA). The GWRC was officially formed in April 2002 with the signing of the partnership agreement at the International Water Association 3rd World Water Congress in Melbourne. With the U.S. Environmental Protection Agency a partnership agreement was signed in July 2003.

Endocrine Disrupting Compounds (EDCs) comprise an important research area for GWRC members and a number of joint efforts have been undertaken. The project *Tools to Detect Estrogenic Activity in Enviromental Waters* (the EDC Toolbox) was jointly funded by the GWRC members AwwaRF, UKWIR, WERF, and WSAA, and carried out by a project consortium of CRC WQT, Kiwa, TZW, and WRC. WERF was the lead agent of this joint effort and CRC WQT was the coordinator of the consortium.

ACKNOWLEDGMENTS

Project Team
Heather Chapman, Ph.D.
Frederic Leusch, Ph.D.
CRC Water Quality and Treatment (Australia)

Lead Agent
Margaret Stewart
Water Environment Research Foundation (WERF) (U.S.)

Project Steering Group
Seth Kullman, Ph.D.
Duke University (U.S.)

Deb Lester
King County (U.S.)

Elaine Francis, Ph.D.
U.S. Environmental Protection Agency (U.S.)

Djanette Khiari, Ph.D.
Awwa Research Foundation (AwwaRF) (U.S.)

Gordon Wheale
UK Water Industry Research (U.K.)

Issy Cafoor, Ph.D.
Yorkshire Water (U.K.)

Peter Donlon, Ph.D.
Water Services Association of Australia (WSAA) (Australia)

Project Participants
Tiaan de Jager, Ph.D.
Natalie Aneck-Hahn, Ph.D.
Catherina Van Zijl
Water Research Commission/University of Pretoria (South Africa)

Richard Lim, Ph.D.
Anne Colville, Ph.D.
University of Technology Sydney (Australia)

Louis Tremblay, Ph.D.
Katherine Trought

Landcare Research (New Zealand)

Leo Puijker, Ph.D.
Kiwa Water Research (The Netherlands)

Frank Sacher, Ph.D.
DVGW - Technologiezentrum Wasser (Germany)

Benjamin Tan, Ph.D.
Griffith University (Australia)

Yves Levi, Ph.D.
Viviane Huteau
Université Paris Sud 11 (France)

Nadine Dumoutier, Ph.D.
Jean-Michel Laine, Ph.D.
Suez Environment (France)

Vickie Wilson, Ph.D.
Earl Gray Jr, Ph.D.
Chad Blystone
Kathy Bobseine
U.S. Environmental Protection Agency (U.S.)

ABSTRACT

The occurrence of estrogenic endocrine disruptors in water is of international concern because of potential adverse health effects on wildlife and humans. Chemical analysis and quantification of estrogenic compounds in water is problematic due to the great range of compounds with endocrine activity. Also, the ultra-low concentrations that can cause biological effects make it clear that additional methods are needed for this type of analysis. Bioanalytical methods have become increasingly popular and are seen as a possible screening tool for measuring estrogenic activity in water. Bioassays generally have significantly lower detection limits than chemical methods, provide an integration of potency and dose and, most importantly, require no prior knowledge of the specific chemical nature of a sample. Several *in vitro* bioassays have emerged over the past decade to test the estrogenicity of environmental samples. There are, however, concerns about their reproducibility, robustness, interlaboratory variability and their ability to integrate into a regulatory framework based on individual chemicals.

This report describes an international effort to evaluate the performance of five *in vitro* bioassays to assess estrogenic activity in a variety of water matrices (http://www.edctoolbox.info). The project was jointly funded by members of the Global Water Research Coalition (GWRC), and evaluated a selected set of bioassays, including yeast estrogen screen (YES), ER-CALUX, MELN, T47D-KBluc and E-Screen assays. Spiked artificial (tap water spiked with known estrogenic chemicals such as hormones, alkylphenols, phthalates, pesticides and phytosterols) and real samples from sewage, river, groundwater and drinking water were tested.

The results indicate that the ER-CALUX and E-Screen assays in this study successfully detected estrogenicity in environmental water samples even at very low levels of estrogenicity (from 0.1 to 320 ng/L EEq). The estrogenic activity measured in these bioassays could be correlated to the predicted estrogenic activity based on comprehensive chemical analysis (GC/MS, GC/ECD, and HPLC/MS/MS), suggesting that either of these two bioassays could be used as initial screening tools to detect estrogenicity in environmental water samples. The KBluc assay was very similar to the ER-CALUX, but these conclusions are based on a more limited dataset, and should be considered critically. The YES performed well with highly polluted environmental samples (such as sewage samples) but its relatively high detection and quantification limits meant that it was unable to measure low-level estrogenicity (eg ground and river water). With artificial samples, the performance of the YES assay was also significantly affected by octylphenol. The MELN assay tested in this study provided good qualitative data, clearly identifying low and high estrogenic activity in the samples. However, accurate quantification was more problematic, possibly due to matrix interference from complex matrices (such as sewage) in this assay.

This study shows that some bioassay techniques are now sufficiently advanced that they can be used either as a cost-effective first-pass detection system or in combination with standard analytical methods to measure estrogenic pollutants in environmental waters. Standardization of bioassay data analysis was identified as a crucial step forward towards accurate bioassay-derived estrogenicity measurements.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PICTURES

LIST OF ABBREVIATIONS

BA	Bisphenol A (CASRN: 80-05-7)
CASRN	Chemical abstracts service registry number
CRCWQT	Cooperative Research Centre for Water Quality and Treatment
E-Screen	MCF-7 cell proliferation assay
E1	Estrone (CASRN: 53-16-7)
E2	17β-Estradiol (CASRN: 50-28-2)
E3	Estriol (CASRN: 50-27-1)
EDC	Endocrine disrupting compound
EE2	Ethynylestradiol (CASRN: 57-63-6)
EEq	Estradiol equivalents
ER	Estrogen receptor
ERE	Estrogen responsive element
GC/MS	Gas chromatography – mass spectrometry
Luc	Luciferase
MCF-7	Breast cancer cell line
MELN	MCF-7-ERE-βGlob-Luc-Neo
NP	4-Nonylphenol (CASRN: 84852-15-3)
OP	4-*t*-Octylphenol (CASRN: 140-66-9)
P/N	Product number
QSAR	Quantitive structure-activity relationship
SPE	Solid phase extraction
T47D	Breast cancer cell line
TBBPA	Tetrabromobisphenol A (CASRN: 79-94-7)
TMX	Tamoxifen (CASRN: 10540-29-1)
TZW	TechnologieZentrum Wasser
WRC	Water Research Commission
YES	Yeast estrogen screen

EXECUTIVE SUMMARY

The presence of estrogenic compounds in drinking waters, source waters and wastewater is of international concern because of potential adverse effects on wildlife and humans. Chemical analysis in environmental matrices is problematic due to both the large numbers of compounds with endocrine activity that may be present in the environment and the ultra-low concentrations that have been reported in the literature to cause estrogenic effects. Biological methods are becoming increasingly popular as screening tools because the specific chemical nature of an environmental sample is often unknown. As the effects of chemical mixtures cannot always be inferred from their concentrations, bioassays are an important component of examining the presence of and integrating the effects of complex mixtures of endocrine disrupting chemicals. However, no single assay can accurately predict the total estrogenic activity of a complex sample to all organisms. Therefore, there is a need to evaluate the strength and weaknesses of the available methods. This report describes a Global Water Research Coalition (GWRC) project to evaluate *in vitro* bioassays for screening environmental waters for estrogenic activity.

The first stage of this research (reported separately) was a biological methods review that comprehensively examined an initial set of 24 bioassays for their suitability to measure the estrogenic activity of environmental water samples (1). From this initial set, five *in vitro* assays were selected for validation of the methods based on criteria such as global applicability, reliability, robustness, maturity and potential for high-throughput screening. The bioassays tested during this project were the yeast estrogen screen (YES), the ER-CALUX, the MELN, the T47D-KBluc and the MCF7 cell proliferation (E-Screen) assays.

Artificial samples were prepared by mixing known estrogenic compounds in tapwater. Eight environmental samples were also collected from different sites around Brisbane (Qld, Australia), including two sewage treatment plants, two river sites, and two groundwater sites. All samples were extracted using solid-phase extraction, eluted and split into 12 aliquots, which were then sent to all collaborating laboratories for analysis. The five bioassays (YES, ER-CALUX, MELN, KBluc and E-Screen assays) were then applied to determine the estrogenic activity in all artificial and environmental samples. All samples were concurrently analyzed using standard analytical chemistry methods, with gas chromatography and high pressure liquid chromatography (in combination with mass spectrometry) used to measure a range of estrogenic compounds, including natural and synthetic hormones, drugs, industrial estrogen mimics, pesticides and a phytosterol.

The results show that:

- The ER-CALUX and E-Screen assays are robust, in good agreement with chemical analysis and have a very low method quantification limit suggesting that they may be ideal as preliminary screening tools for environmental monitoring in combination with analytical chemistry.
- The KBluc assay likewise appears well-suited to that purpose and may be a useful tool in the future, although conclusions for this assays are based on a limited dataset.

- The YES assay was useful in determining estrogenicity of model compounds but its comparatively high method quantification limit meant that many environmental samples were nondetects, thus limiting YES assay's usefulness in environmental monitoring of very low-level contamination.
- The results from the MELN assay were significantly different from those obtained with the other assays and generally lower than what would have been predicted from the chemical analysis alone, suggesting possible matrix interference or a higher sensitivity to anti-estrogenic chemicals in that assay. Further work is needed to determine the reliability of that assay for environmental monitoring.
- With the exception of the MELN assay, the results from all bioassays in the environmental samples showed very similar trends: very high estrogenic activity in raw sewage, a markedly lower activity in treated sewage, and barely detectable activity in ground and river water.

In conclusion, this study shows that bioassay techniques are now sufficiently advanced that they can be used either as a cost-effective first-pass detection system or in combination with standard analytical methods to measure estrogenic pollutants in environmental waters. Each assay has its advantages and limitations, and the notion of "fit-for-purpose" is critical in determining what bioassay to use in a particular project. For example, the YES assay may be suitable for testing sewage where its poor sensitivity may not be a liability but its low cost is clearly an advantage; when testing estrogenicity in drinking water however, a more sensitive (but also more expensive) bioassay such as the ER-CALUX or the E-Screen bioassays may be more appropriate.

Standardization of bioassay data analysis was identified as a crucial step forward towards accurate bioassay-derived estrogenicity measurements.

The concepts and approach used in this study offer a robust method to test the usefulness of bioassays for other endocrine effects (such as androgenicity) or even other endpoints relevant to human health, such as carcinogenicity, immunotoxicity, neurotoxicity, etc. A combined research approach could eventually develop a battery of bioassays to detect multiple endpoints relevant to human health to screen large numbers of samples for biologically-active contaminants.

1.0 Introduction

Phase 1 of this research (reported separately) was a biological methods review that comprehensively examined an initial set of 24 bioassays for their suitability to measure the estrogenic activity of environmental water samples [1]. From this initial set of 24 bioassays, five in vitro assays were selected for validation of the methods based on criteria such as global applicability, reliability, robustness, maturity and potential for high-throughput screening. The bioassays tested during this project were the yeast estrogen screen (YES), the ER-CALUX, the MELN, the T47D-KBluc and the MCF7 cell proliferation (E-Screen) assays. All samples were concurrently analyzed using standard analytical chemistry methods.

The presence of estrogenic compounds in drinking waters, source waters and wastewater is of international concern because of potential adverse effects on wildlife and humans. Chemical analysis in environmental matrices is problematic due to both the large numbers of compounds with endocrine activity that may be present in the environment and the ultra-low concentrations that have been reported in the literature to cause estrogenic effects. Biological methods are becoming increasingly popular as screening tools because the specific chemical nature of an environmental sample is often unknown. As the effects of chemical mixtures cannot always be inferred from their concentrations, bioassays are an important component of examining the presence of and integrating the effects of complex mixtures of endocrine disrupting chemicals. However, no single assay can accurately predict the total estrogenic activity of a complex sample to all organisms. Therefore, there is a need to evaluate the strength and weaknesses of the available methods.

This report describes a Global Water Research Coalition (GWRC) project to evaluate *in vitro* bioassays for screening environmental waters for estrogenic activity. A similar GWRC study has been completed regarding analytical chemical methods [2].

1.1 Aims and Objectives of this Research

The main objective of this study was to determine the usefulness of different bioassays to detect (and if possible quantify) estrogenicity in environmental waters. The project was undertaken to provide a sound scientific basis on which community and government concerns regarding environmental and human health can be addressed and managed. In particular, this study aimed to compare and critically evaluate the suitability and performance of the selected bioassays to detect and measure estrogenicity in a variety of environmental water matrices. As such, this project is akin to a performance evaluation of the tests, and not a formal validation exercise. The conclusions drawn in this report are only relevant to the assays tested. It should be noted that this study was not designed to address the issue of inter-laboratory variability, which would require running the same assay in multiple laboratories.

An additional objective was to compare bioassay results with standard analytical chemistry methods to determine how well the two approaches could be correlated. This is a critical step towards integrating bioassays into a risk assessment framework currently designed around single chemicals.

1.2 Estrogenic Chemicals in Environmental Waters

The issue of endocrine disruption and the presence of estrogenic compounds in the aquatic environment has already been reviewed in a previous publication from this project [1]. In brief, some compounds in the environment have the ability to mimic or interfere with the function of hormones. Hormones regulate a variety of biological functions including growth, metabolism, cell growth and proliferation, cell function and differentiation, sexual development and behaviour, and development of the immune system. In particular, some chemicals can mimic or interfere with the function of estrogens, the primary sex hormone in females responsible for sexual development and maintenance of the reproductive cycle. Estrogens are also present in males, albeit at lower levels, where they are also involved in maintenance of the reproductive system. Estrogenic chemicals in treated sewage have been implicated with sexual abnormalities in fish in the United Kingdom [3], and while there is no evidence that exposure to environmental levels of endocrine disruptors (EDCs) can affect humans [4] there is a need to monitor levels of estrogenicity in the aquatic environment.

1.3 Issues with Standard Chemical Methods

Environmental monitoring for estrogenicity currently relies on chemical analysis to measure individual chemicals in the environment. For estrogenic chemicals there are several limitations to this approach:

- Some estrogenic chemicals such as the natural and synthetic hormones can be extremely potent and cause adverse effects at concentrations below the analytical detection limit.
- Because analytical chemistry methods rely on chemical structure for analysis, only those chemicals that are tested for will be detected. Other chemicals that may be biologically active will not be detected if they are not included in the chemical screening.
- For that same reason, chemical methods have to be continually updated to measure emerging contaminants, which may have very different chemical structures as currently- monitored chemicals.
- The chemical analysis results provide limited biological information (such as bioavailability when combined with QSAR models) to inform a risk assessment.
- They provide no measure of mixture toxicity or possible interactions between the different components of complex mixtures.
- Finally, with hundreds of potential endocrine disruptors and thousands of registered chemicals, the list of potential analytes is gargantuan.

As such, there is an identified need to develop alternative methods to screen samples and help direct chemical analysis towards particular contaminants.

1.4 Bioassays

In vitro bioassays have several significant features that make them ideal candidates to serve as screening tools:

- Bioassays are generally more sensitive than chemical methods, and can detect biologically-active chemicals at very low concentrations.

- Bioassays detect pollutants measuring the physiological effects *in vitro* and not their chemical structure. This means that bioassays can detect estrogenic chemicals irrespective of their chemical structure and do not require any *a priori* knowledge of the chemical composition of a sample.
- *In vitro* bioassay techniques can integrate the effects of many chemicals in a complex mixture with similar modes of action.
- Bioassays are therefore ideal to identify emerging pollutants and do not need to be updated to detect unexpected biologically-active contaminants.
- Bioassays can inform the exposure assessment.
- Finally bioassays provide targeted biological information that may inform the risk assessment process and guide epidemiology studies.

While bioassays have undoubtedly proved their worth in pharmaceutical and laboratory conditions to test the biological activity of individual compounds, there are still many questions about their usefulness with environmental samples. Bioassay results may be affected by other unrelated toxic effects from complex mixtures, and there are questions about the reliability and robustness of bioassays in such conditions. There appear to be significant issues with interlaboratory variability due to poor standardization of bioassays [5], and there is a perception that bioassays are too different from chemical techniques to allow comparison between the two methods.

There is therefore a need to examine the usefulness of *in vitro* bioassays to detect (and possibly quantify) estrogenic activity in environmental waters.

1.5 Significance of this Research

This research is significant for the following reasons:

- To assess the complex and multiple modes of action of EDCs appropriate bioassays need to be identified so that these compounds can be effectively managed and or regulated. It is impractical and extremely costly to solely rely on chemical analysis of effluent due to the number of analytes present and because many of these substances are known to be biologically-active below the analytical detection limits. Bioassays also provide an assessment of integrated exposure (e.g., additive, synergistic and/antagonistic effects) which also cannot be predicted by chemical methods alone.
- The use of mechanism-based bioassays will enable rapid screening of waters as part of exposure assessment and/or effects assessment of whole waters. This may in some cases negate or justify the need for further investigation using *in vivo* methods and/or chemical identification of active substances. In other words, a significant response in an *in vitro* bioassay could be used as a trigger for further investigation using *in vivo* test systems, while a negative response would suggest such financially and ethically-expensive investigation is not warranted.
- This project also aimed to provide a framework for future evaluation of biological tests for different modes of action (e.g., neurotoxins) responding to bioactive substances in environmental waters.

2.0 Materials and Methods

Artificial samples were prepared by mixing known estrogenic compounds (such as estradiol, 17α-ethinylestradiol, nonylphenol, bisphenol A etc) in tapwater. Eight environmental samples were collected from different sites around Brisbane (Qld, Australia), including two sewage treatment plants, two river sites, and two groundwater sites. A duplicate of each environmental sample was also spiked with a standard mixture of known estrogenic compounds as a positive control. All samples were extracted using solid-phase extraction, eluted and split into 12 aliquots, which were then sent to all collaborating laboratories for analysis.

The five bioassays (YES, ER-CALUX, MELN, KBluc and E-Screen assays) were then applied to determine the estrogenic activity in all artificial and environmental samples. All samples were concurrently analyzed using standard analytical chemistry methods with gas chromatography and high pressure liquid chromatography (in combination with mass spectrometry) used to measure a range of estrogenic compounds, including natural and synthetic hormones, drugs, industrial estrogen mimics, pesticides and a phytosterol.

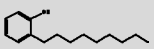
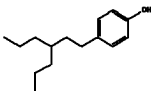
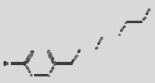
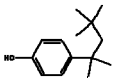
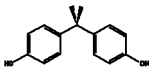
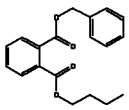
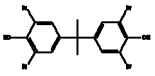
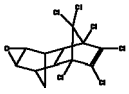
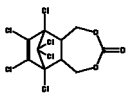
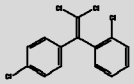
2.1 Samples

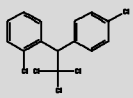
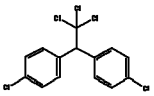
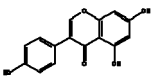
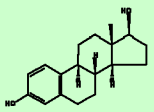
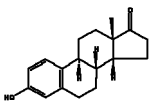
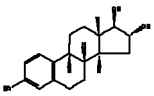
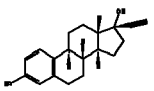
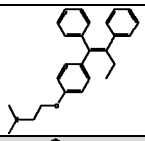
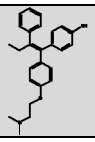
All samples were prepared or collected in Brisbane (Qld, Australia). They were extracted in the laboratory at EnTox (Brisbane, Qld, Australia), and shipped to all laboratories for analysis.

2.1.1 Artificial Samples

Artificial (synthetic) samples were created in the laboratory by adding a suite of selected estrogenic chemicals to tap water (Brisbane municipal water, Brisbane, Australia). Fourteen estrogenic chemicals were selected for this study, including the natural hormones 17β-estradiol (E2; Sigma P/N E8875), estrone (E1; Sigma P/N E3201) and estriol (E3; Aldrich P/N E1149), the pharmaceuticals ethinylestradiol (EE2; Aldrich P/N 28586-2) and tamoxifen (TMX; Aldrich P/N 330396), the industrial compounds nonylphenol (NP technical grade; Aldrich P/N 29085-8), 4-*t*-octylphenol (OP; Aldrich P/N 290823), bisphenol A (BA; Aldrich P/N 239658), benzyl butyl phthalate (BBP; Aldrich P/N 308501) and tetrabromobisphenol A (TBBPA; Aldrich P/N 330396), the pesticides dieldrin (Supelco ChemService P/N PS76), endosulfan (α and β; Supelco ChemService P/N PS81-1 and PS81-2, respectively) and p,p'-DDT (Supelco ChemService P/N PS699), and the natural phytoestrogen genistein (Sigma P/N G6649). The physico-chemical properties of these chemicals and their potency relative to 17β-estradiol in the assays evaluated in this study are presented in Table 1 below.

Table 1. Physico-Chemical Properties of Selected Chemicals and Potency (Log) Relative to 17β-Estradiol in the Different Bioassays Used in this Project.

Chemical	Structure	YES[(a)]	MELN[(a)]	KBluc[(a)]	ER-CALUX[(a)]	E-Screen[(a)]
Nonylphenol (mixture) CASRN: 25154-52-3 Formula: $C_{15}H_{24}O$ MW: 220.35		-4.66 [6]	-4.29 [7]			-4.54 [6]
4-Nonylphenol (This information is for 4-n-Nonylphenol, an unbranched isomer) CASRN: 104-40-5 Formula: $C_{15}H_{24}O$ MW: 220.35		-3.24 [8] -4.00 [9] -4.60 [10] -6.14 [11]	MVLN -4.52 [9] -5.80 [7]		-4.64 [8] -4.64 [12]	-4.89 [13] -4.11 [14] -6.63 [5] -4.24 [15] -4.12 [16]
4-Nonylphenol (NP) CASRN: 84852-15-3 Formula: $C_{15}C_{24}O$ MW: 220.35		**-3.57 (*)** **-2.96 (*)**	**-5.02 (*)**	**-4.43 (*)**	**-3.92 (*)**	**-4.41 (*)** **-4.16 (*)**
4-Octylphenol (This information is for 4-n-Octylphenol, one of the 4-Octylphenol isomers) CASRN: 1806-26-4 Formula: $C_{14}H_{22}O$ MW: 206.33		-5.00 [8] -5.11 [10] -4.52 [6]	-3.49 [7]		-5.85 [8]	-4.00 [13] -6.80 [5] -5.14 [15] -4.46 [6]
4-tert-Octylphenol (OP) CASRN: 140-66-9 Formula: $C_{14}H_{22}O$ MW: 206.33		-3.32 [17] -5.44 [6] **-2.67 (*)** **-2.77 (*)**	**-5.32 (*)**	**-4.72 (*)**	**cytotoxic (*)**	-4.19 [14] -4.01 [15] -4.12 [16] **-3.22 (*)** **-3.38 (*)**
Bisphenol A (BA) CASRN: 80-05-7 Formula: $C_{15}H_{16}O_2$ MW: 228.29		-5.00 [8] -1.99 [18] -3.96 [10] -4.25 [5] -4.07 [17] -4.30 [6]	-3.74 [7]	**-5.70 (*)**	-4.48 [19] -5.11 [8] -5.11 [12]	-4.60 [13] -4.53 [14] -5.60 [5] -4.10 [15] -4.28 [16] -4.78 [6] **-4.43 (*)**
Benzyl butyl phthalate (BBP) CASRN: 85-68-7 Formula: $C_{19}H_{20}O_4$ MW: 312.36		-6.00 [8] -5.40 [6] **-4.84 (*)** **ND (-5.90) (*)**	**-5.88 (*)**	**-5.51 (*)**	-5.85 [8] **-5.43 (*)**	-6.50 [5] -5.62 [16] -5.60 [6] **-5.80 (*)** **-5.54 (*)**
Tetrabromobisphenol (TBBPA) CASRN: 79-94-7 Formula: $C_{15}H_{12}Br_4O_2$ MW: 543.87		**cytotoxic (*)** **ND (-5.47) (*)**	**-6.50 (*)**	**-6.37 (*)**	ND [19] **-6.88 (*)**	-6.00 [20] **-5.61 (*)** **ND (6.87) (*)**
Dieldrin CASRN: 60-57-1 Formula: $C_{12}H_8Cl_6O$ MW: 380.92					-6.62 [12]	-6.70 [21]
Endosulfan CASRN: 115-29-7 Formula: $C_9H_6Cl_6O_3S$ MW: 406.93					-6.00 [12]	-6.00 [21] -6.04 [5] -5.32 [16] -5.84 [6]
o,p'-DDE CASRN: 3424-82-6 Formula: $C_{14}H_8Cl_4$ MW: 318.03		-6.40 [6]	-5.20 [7]		-5.64 [22]	-7.56 [5]

Chemical	Structure	YES[(a)]	MELN[(a)]	KBluc[(a)]	ER-CALUX[(a)]	E-Screen[(a)]
o,p'-DDT CASRN: 789-02-6 Formula: $C_{14}H_9Cl_5$ MW: 354.49		-5.96 [6]			-5.04 [12] -5.04 [22]	-6.17 [5] -4.78 [6]
p,p'-DDT CASRN: 50-29-3 Formula: $C_{14}H_9Cl_5$ MW: 354.49		-3.86 [18] **ND (-5.76) (*)** **ND (-5.60) (*)**	**-5.87 (*)**	**-6.17 (*)**	**-6.02 (*)**	-5.40 [6] **-6.41 (*)** **-4.74 (*)**
Genistein CASRN: 446-72-0 Formula: $C_{15}H_{10}O_5$ MW: 270.24		-3.61 [18] -3.31 [6]	-3.19 [7]	-4.52 [23]	-4.22 [12]	-4.89 [13] -3.55 [16] -3.85 [6] **-4.05 (*)**
17β-Estradiol (E2)[(b)] CASRN: 50-28-2 Formula: $C_{18}H_{24}O_2$ MW: 272.39		-9.60 [18] -9.80 [9] -9.68 [10] -9.40 [5] -9.09 [17] -9.68 [11] **-9.64 (*)** **-9.04 (*)**	MVLN -10.82 [9] -10.75 [7] -10.70 [24] **-11.25 (*)**	-11.52 [23] **-11.52 (*)**	-11.00 [19] -11.30 [8] -11.22 [12] -11.68 [25] **-11.70 (*)**	-11.00 [21] -12.12 [5] -11.05 [15] -10.49 [11] -11.21 [16] **-11.31 (*)** **-11.27 (*)**
Estrone (E1) CASRN: 53-16-7 Formula: $C_{18}H_{22}O_2$ MW: 270.37		-1.00 [8] -0.40 [9] -0.42 [26] -1.02 [6] **-0.61 (*)**	MVLN -0.70 [9] -1.60 [7] -0.60 [24]	**-1.67 (*)**	-1.25 [8]	-2.00 [13] -1.91 [14] -0.87 [15] -1.02 [16] -1.36 [6] **-1.95 (*)**
Estriol (E3) CASRN: 50-27-1 Formula: $C_{18}H_{24}O_3$ MW: 288.39		-2.62 [26] -2.20 [6]	-0.75 [7] -1.09 [24]		0.00 [27]	-1.15 [13] -0.53 [15] -0.60 [6] **-1.07 (*)**
Ethinylestradiol (EE2) CASRN: 57-63-6 Formula: $C_{20}H_{24}O_2$ MW: 296.41		0.08 [8] -0.05 [9] 0.08 [10] 0.36 [5] -0.02 [17] -0.15 [11] -0.05 [6] **0.09 (*)**	MVLN 0.20 [9] 0.39 [7] 0.06 [24]	-0.45 [23]	0.08 [8]	0.10 [13] 0.03 [8] -0.78 [5] 0.13 [15] 0.28 [11] -0.04 [16] 0.05 [6] **-0.17 (*)**
Tamoxifen (TMX) CASRN: 10540-29-1 Formula: $C_{26}H_{29}NO$ MW: 371.52		-3.15 [5] -4.33 [6] **-3.07 (*)** **-3.40 (*)**	**-3.92 (*)**	-5.52 [23] **-3.08 (*)**	**-3.83 (*)**	-4.39 [13] -3.30 [6] **-3.25 (*)** **-3.22 (*)**
4-Hydroxytamoxifen CASRN: Formula: $C_{26}H_{29}NO_2$ MW: 387.52		-4.14 [6]				-1.52 [5] -1.42 [6]

Notes: Gray rows (such as o,p'-DDT) highlight chemicals that were not used in this project but are given for comparison. ND = no detectable effect (although when available, the highest possible potency is given in brackets based on the highest concentration tested). Note that Fang et al 2000 [6] is a review and reports potencies from several different papers.

(a) The assays and assay number in the GWRC 2006 literature review [1] are: YES = yeast estrogen screen (assay #04); MELN = MELN reporter gene assay (assay #10); KBluc = T47D-KBluc reporter gene assay (assay #12); ER-CALUX = chemical-activated luciferase gene expression assay (assay #11); E-Screen = MCF7 cell proliferation assay (assay #23).

(b) Note that for 17β-estradiol, the EC_{50} (in M) is reported instead of the log RP (the log RP of 17β-estradiol is 0). The EC_{50}s obtained in this study are reported by the participating laboratories from E2 standard curves.

(*) and bold: Potency obtained in this study, by dividing the measured EEq in samples A01-A07 (Table 19) by the actual chemical concentration in the sample as measured by chemical methods (Table 9 and Table 10). Some potencies were also provided for individual chemicals by the participating laboratories (such as EE2 and E1 for the YES assay, and BA, genistein, E1, E3 and EE2 for the E-Screen).

Twelve different artificial samples were prepared, seven from individual compounds, four from compound mixtures designed to simulate different types of wastewater, and one blank (tap water only). The concentration of compound(s) in each mix (Table 2) were chosen based on literature-based potency figures to achieve a total estrogenicity of roughly 10-50 ng/L estradiol equivalents (EEq). The concentrations were also chosen so that the individual chemical would be detectable by chemical analysis and still represent an environmentally-relevant figure.

The compounds were pre-dissolved in ethanol and added to 2.5L tap water (maximum 1125 µL, ie. 0.05% ethanol). After addition of the compounds, the sample was acidified to pH 2 and extracted as described below in Section 2.1.3.

Table 2. Artificial Effluent Mixes and Constituent Chemical Concentrations.

Mix ID	Mix name	Sample no.	Chemical(s)	Nominal Concentration (µg/L)
A01	Estradiol	11, 51, 159	17β-Estradiol (E2)	0.015
A02	Nonylphenol	55, 109, 173	Nonylphenol (NP)	400
A03	Octylphenol	81, 84, 144	4-*t*-Octylphenol (OP)	5 000
A04	Benzyl butyl phthalate	87, 113, 179	Benzyl butyl phthalate (BBP)	5 000
A05	Tetrabromobisphenol A	21, 120, 151	Tetrabromobisphenol A (TBBPA)	5 000
A06	p,p'-DDT	43, 68, 139	p,p'-DDT	5 000
A07	Tamoxifen	25, 56, 122	Tamoxifen (TMX)	50
A08	Hormone mix	97, 99, 184	17β-Estradiol (E2)	0.01
			Estrone (E1)	0.05
			Estriol (E3)	0.05
			Ethinylestradiol (EE2)	0.01
A09	Industrial mimics mix	37, 41, 192	Nonylphenol (NP)	100
			4-*t*-Octylphenol (OP)	1 000
			Bisphenol A (BA)	50
			Tetrabromobisphenol A (TBBPA)	1 000
			Benzyl butyl phthalate (BBP)	1 000
A10	Agro-industrial mix	33, 63, 75	Dieldrin	1 000
			Endosulfan (45% α and 55% β)	1 000
			p,p'-DDT	1 000
			Genistein	20
A11	Combo mix	2, 86, 104	Benzyl butyl phthalate (BBP)	1 000
			Bisphenol A (BA)	50
			Dieldrin	1 000
			Ethinylestradiol (EE2)	0.01
			Endosulfan (45% α and 55% β)	1 000
			17β-Estradiol (E2)	0.01
			Estrone (E1)	0.05
			Estriol (E3)	0.05
			Genistein	20
			Nonylphenol (NP)	100
			4-*t*-Octylphenol (OP)	1 000
			p,p'-DDT	1 000
			Tetrabromobisphenol A (TBBPA)	1 000
			Tamoxifen (TMX)	50
A12	Blank (tap water only)	165, 167, 188	None	0

2.1.2 Environmental Samples

Water from four different matrices was tested in this project. All samples were collected in Brisbane (Qld, Australia) in methanol-rinsed 1 – 2.5 L amber glass bottles and immediately acidified to pH 2 to prevent further microbial degradation. Different sites were sampled (Pictures 1 to 6), including groundwater, raw and treated sewage, and river water to provide an environmentally-relevant range of water matrices and concentrations to test for estrogenic activity. Six replicate samples (2.5 L samples for groundwater, treated sewage and river water; 1 L samples for raw sewage) were taken at each site, half of which (3) were spiked immediately after pH adjustment to pH 2 with a standard mix of estrogenic chemicals (125 ng E1, 37.5 ng E2, 125 ng E3, 25 ng EE2, 125 µg BA, 250 µg NP and 2.5 mg BBP; pre-dissolved and pre-mixed in 875 µL ethanol, ie. maximum 0.09% ethanol) as positive controls. The samples were then brought back to the laboratory and extracted as described below (Section 2.1.3).

Groundwater (GW): Both groundwater samples were taken at bores in South Brisbane, Australia. The first groundwater sample (GW1) was taken at a recently installed shallow bore (16 m deep) in a new residential development in the South of Brisbane, while the second groundwater sample (GW2, Picture 4) was from a much deeper aquifer in a less urbanized area, also in the South of Brisbane.

Sewage, raw (SR): Raw sewage was collected at two large municipal sewage treatment plants in the Greater Brisbane area. The first raw sewage samples (SR1, Picture 5) came from a large wastewater treatment plant receiving approximately 58,000 m^3/d of mixed domestic and industrial waste, and the second set of samples (SR2) came from a medium-size plant receiving approximately 26,000 m^3/d of mixed domestic and industrial waste.

Sewage, treated (ST): The treated sewage samples were taken at the same plants and at the same time as the raw sewage samples (ST1 is from the same plant where SR1 was taken, while ST2 is from the plant where SR2 was collected). The treatment train at the first treatment plant consists of screens, grit channels, primary sedimentation, aerobic/anaerobic bioreactors, clarifiers, and chlorine contact tanks. The treatment train at the second plant is very similar and starts with manual screens followed by passage through grit channels, primary sedimentation, aeration tanks, clarifiers, and finally chlorine contact tanks. Treated sewage samples (ST1 and ST2) were taken post-chlorination.

River water (RW): River water samples were also collected in the Brisbane area. The first set of samples (RW1, Picture 2) was taken from the Brisbane River at Orleigh Park pier (S27.487477°, E152.996539°). At this level, the Brisbane River is a large tidal river with small boat traffic. The second set of river water samples (RW2, Picture 3) was taken from Oxley Creek at the Kendall St boat ramp (S27.558608°, E152.98243°). Oxley Creek is a small tributary to the Brisbane River that flows through industrial areas and is typically high in suspended matter.

Table 3. Description of Environmental Samples and Spiking Levels.

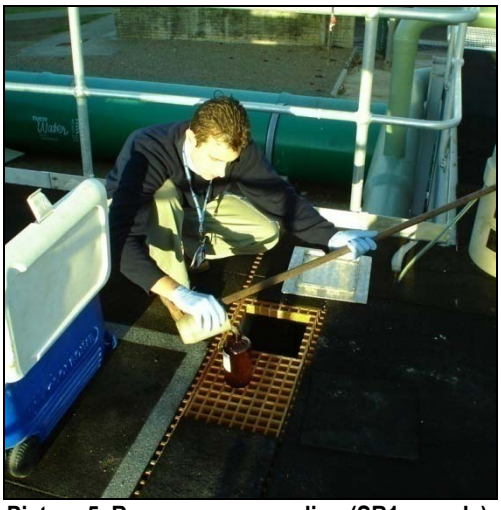
Sample type	ID	Source	Volume	Spike [(a)]	Sample ID
Groundwater 1	GW1-	Shallow aquifer, 16m deep	2.5 L	-	121, 141, 155
	GW1+			+++	5, 119, 176
Groundwater 2	GW2-	Deep aquifer, 90m deep	2.5 L	-	71, 72, 174
	GW2+			+++	8, 62, 157
Sewage, raw 1	SR1-	Large WWTP 58,000 m^3/d	1 L [(b)]	-	58, 78, 94
	SR1+			+++	24, 77, 146
Sewage, raw 2	SR2-	Medium WWTP 26,000 m^3/d	1 L [(b)]	-	32, 138, 170
	SR2+			+++	12, 49, 52
Sewage, treated 1	ST1-	Large WWTP 58,000 m^3/d	2.5 L	-	53, 65, 185
	ST1+			+++	20, 34, 160
Sewage, treated 2	ST2-	Medium WWTP 26,000 m^3/d	2.5 L	-	117, 129, 145
	ST2+			+++	80, 95, 154
River water 1	RW1-	Brisbane River	2.5 L	-	22, 106, 110
	RW1+			+++	142, 171, 181
River water 2	RW2-	Oxley Creek	2.5 L	-	54, 92, 178
	RW2+			+++	48, 132, 150

Notes: In the sewage samples, EP = equivalent people.
(a) "-" = non spiked; "+++" = spiked. The same spike amount was added irrespective of sample volume, so spike concentration in 2.5 L samples (groundwater, treated sewage and river water) were 50 ng/L E1, 15 ng/L E2, 50 ng/L E3, 10 ng/L EE2, 50 µg/L BA, 100 µg/L NP and 1 mg/L BBP. In 1 L samples (raw sewage), the concentrations were 125 ng/L E1, 37.5 ng/L E2, 125 ng/L E3, 25 ng/L EE2, 125 µg/L BA, 250 µg/L NP and 2.5 mg/L BBP.
(b) With raw sewage, 1 L was sampled instead of 2.5 L. The sample was first centrifuged to remove large solids and then filtered through AP20 filters prior to SPE.

Picture 1. Tap water (used to create artificial samples).

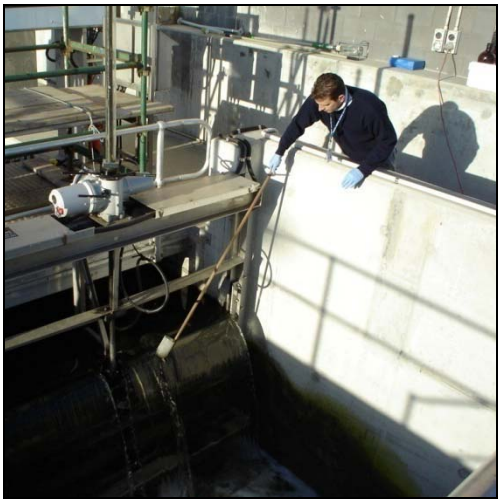
Picture 4. Groundwater pump (GW2 sample).

Picture 2. Brisbane river at Orleigh Park pier (RW1 sample).



Picture 5. Raw sewage sampling (SR1 sample).

Picture 3. Oxley Creek at Kendall St boat ramp (RW2 sample).



Picture 6. Sewage effluent (ST1 sample).

2.1.3 Sample Preparation

Samples were processed immediately upon return to the laboratory (less than 2h after collection). Raw sewage samples were first centrifuged at 4°C and 2,000×*g* for 12 min to remove large particulate matter. The supernatant was transferred to a new methanol-rinsed bottle and extracted using the standard protocol described below.

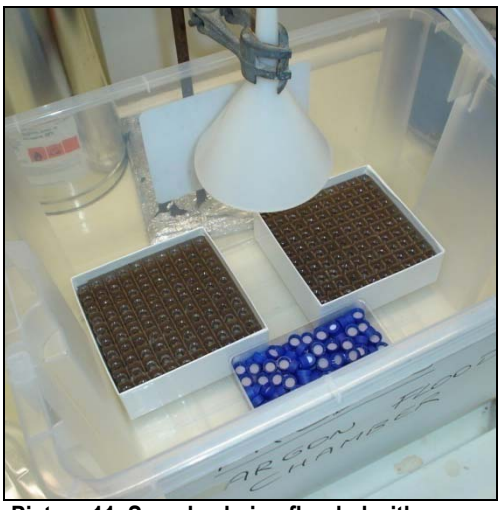
All samples (environmental and artificial) were extracted using a standard extraction protocol (cf protocol GWRC-TDE-02A in Appendix I; Pictures 7 to 12). In brief, samples were pre-filtered with Millipore AP20 filters under gentle vacuum. Pre-filtered samples were passed by vacuum through Oasis HLB reversed phase solid-phase extraction cartridges (Waters Corp., Milford, MA, USA; P/N 18600115) on a 12-port Visiprep SPE vacuum manifold (Supelco, Bellefonte, PA, USA; P/N 57030-U). Once the entire sample had passed, the cartridges were dried on the manifold until dry (usually 1-2 h). Cartridges were then wrapped in aluminum foil and stored at room temperature until elution (maximum 10 days). The cartridges were eluted with 5 mL of methanol followed by 5 mL of 1:1 acetone:hexane. The solvent was then evaporated to dryness under gentle nitrogen stream, and the extract reconstituted in 500 µL ethanol. Samples were split into twelve 40 µL aliquots in MaxRecovery vials (Waters Corp.; P/N 186000326c). The ethanol was evaporated to dryness, and the vials flooded with argon. The samples were then shipped to all participating laboratories for analysis.

Preliminary tests showed that samples preserved, extracted and evaporated using protocols similar to those described above were stable when kept in an opaque cardboard box at room temperature for at least 24 days and at -20°C for at least 12 months (data not shown).

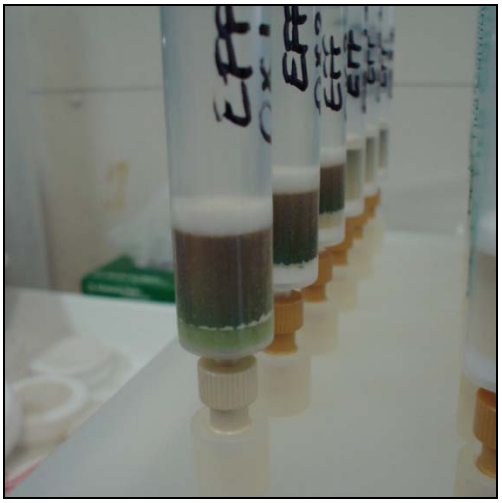
Picture 7. Samples are passed by vacuum through solid-phase extraction cartridges.

Picture 10. Eluted SPE cartridges. Note extracted sample in tube underneath.

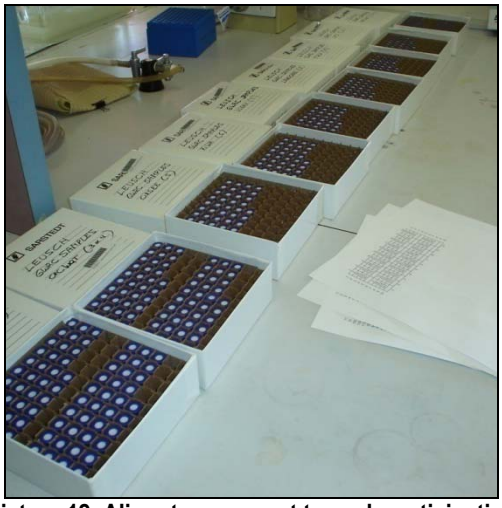
Picture 8. Pre-conditioned Waters Oasis HLB cartridges.



Picture 11. Samples being flooded with argon.



Picture 9. SPE cartridges after sample extraction.



Picture 12. Aliquots were sent to each participating laboratory by courier delivery.

2.2 Bioassays

Five different bioassays selected after an extensive review of available bioassays in the scientific community [1] were evaluated in this project: the yeast estrogen screen (YES), the ER-CALUX, the MELN, the T47D-KBluc, and the E-Screen assays. The following sections provide a brief overview of these bioassays.

2.2.1 Yeast Estrogen Screen (YES)

In the Yeast Estrogen Screen (YES), yeast cells *Saccharomyces cerevisiae* have been stably transfected with the gene for human ERα and a plasmid containing an ERE-linked *lac-Z* gene. Activation of the receptor by binding of an agonistic ligand causes expression of the *lac-Z* gene, which produces β-galactosidase [28]. To perform this assay, yeast cells are distributed in a 96-well plate and exposed to the sample in culture medium for 3 d. The yellow chromogenic substrate CPRG (chlorophenol red-β-D-galactopyranoside) is then added and its transformation into a red product by β-galactosidase is measured in a spectrophotometer at 540 nm [18]. Galactosidase activity, a measure of the ability of the sample to induce ER-mediated gene expression, is then compared against an E2 standard curve, and the estrogenic potency of the sample expressed as estradiol equivalents (EEq). The YES assay is by far the most widely used yeast-based reporter gene assay, and a significant amount of data for individual compounds are available for this assay. It is suitably sensitive and relatively robust, but cell toxicity appears to be an issue with highly concentrated environmental aqueous samples [1]. Furthermore, the thick yeast cell wall may impede active and passive transport of chemicals to the intracellular space, thus increasing the risk of false negatives.

In this study, the YES assay was performed in two laboratories: the University of Pretoria on behalf of the Water Research Commission (WRC) in South Africa, and the University of Technology in Sydney on behalf of the CRC for Water Quality and Treatment (CRCWQT) in Australia. A current standard operating protocol (SOP) for the YES assay is available upon request.

2.2.2 ER-CALUX

The ER-mediated chemical-activated luciferase gene expression assay (ER-CALUX) is based on T47D breast cancer cells stably transfected with an ERE-Luc plasmid [12]. T47D cells endogenously express both ERα and ERβ. To perform this assay, cells are seeded into 96-well plates two days prior to induction. A day later, the medium is changed to steroid-free medium. On the day of induction, the medium is changed again and replaced by steroid-free medium with the sample. After 24 h of exposure, cells are lysed, luciferin is added to the incubation medium and luciferase activity measured by luminescence plate reader. Estrogenicity is expressed relative to that of an estradiol standard curve, as EEq. This assay has been used quite extensively by research groups in the Netherlands, and a review of the literature suggests it is quite robust and appropriate for environmental monitoring [1]. It can however be slightly more expensive than the other bioassays, as a per-sample fee has to be paid to the patent holder for every use.

In this study, this assay was performed by BioDetection Systems on behalf of Kiwa Water Research in the Netherlands. The SOP is copyrighted and can be obtained by contacting BioDetection Systems in the Netherlands.

2.2.3 MELN

The MELN cell line is derived from MCF-7 breast cancer cells that have been stably transfected with a plasmid containing the luciferase gene (*Luc*) driven by an ERE in front of the β-globin promoter (MELN stands for MCF-7-ERE-βGlob-Luc-Neo, where neomycin is the antibiotic used for selection of transfected cells) [29]. MCF-7 cells have endogenous ERα and ERβ, and do not need to be transfected with an external ER. Exposure to estradiol leads to induction of the *Luc* gene. Luciferase production is then measured by addition of the substrate luciferin and quantification of luminescence in intact cells or in cell lysate with a luminometer [29]. To perform this assay, cells are seeded into 96-well plates two days prior to induction. A day later, the medium is changed to steroid-free medium. On the day of induction, the medium is changed again and replaced by steroid-free medium with the sample. After 16-24 h of exposure, cells are (generally) lysed, luciferin is added to the incubation medium and luciferase activity measured by luminescence plate reader. Estrogenicity is expressed relative to that of an estradiol standard curve, as EEq. Luciferase activity is generally measured in cell lysate to achieve greater signal amplification, but the luminescent signal can also be detected from whole cells without loss of sensitivity [30].

In this study, the MELN assay was performed by Université Paris Sud 11 on behalf of the International Research Center on Water and Environment (CIRSEE) in France. A current SOP for the MELN assay is available upon request.

2.2.4 T47D-KBluc

This relatively new assay uses T47D cells stably transfected with a triplet ERE-promoter-luciferase reporter gene construct (available as CRL-2865 from American Type Culture Collection – ATCC) [23]. The protocol for the assay is very similar to the ER-CALUX assay. Cells are withdrawn into steroid-free medium for 1 week prior to being seeded into 96-well plates for 24 h, after which they are exposed to the samples for a further 24 h. At the end of the incubation period, cells are lysed and luciferase activity measured by luminescence. Estrogenicity is expressed relative to that of an estradiol standard curve, as EEq. Although more recently established, the assay appears slightly more sensitive than the ER-CALUX assay and the cell line is freely available [23].

In this study, a limited set of samples were analysed by the Office of Research and Development of the U.S. Environmental Protection Agency (U.S. EPA). A current SOP for the KBluc assay is freely available upon request.

2.2.5 E-Screen

The E-Screen uses human breast cancer cells, which are estrogen-dependent for growth. In this assay, the number of cells present after 5 d of exposure to a sample is compared with the number of cells present in an estradiol standard curve [31]. To perform the assay, breast cancer cells are seeded in 96-well plates [32, 33] in steroid-free medium. After 24 h, the medium is exchanged for fresh steroid-free medium with the sample. After 5 d of exposure, the number of live cells in each well is determined using a standard MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay. Generally, the MCF-7 cell line is used, although the T47-D cell line has been shown to be equally sensitive [34]. Although this assay provides a measure of estrogenic activity at the cellular level incorporating both genomic and non-genomic effects, there is considerable variation between different MCF-7 cell lines, with the MCF-7 BOS stock

(used in this study) showing the highest proliferative effect under estradiol stimulation [35]. The E-Screen assay could lead to false positives, as cell growth can be induced by a range of mitogens, cytokines, growth factors, nutrients and hormones other than estrogens [1]. The E-Screen assay is also generally more time-consuming (and thus expensive) than other assays, limiting its application for large-scale screening [36].

In this study, the E-Screen assay was performed by the National Research Centre for Environmental Toxicology (EnTox; Pictures 13 to 15) and Landcare Research NZ Ltd on behalf of the CRCWQT in Australia. A current SOP for the E-Screen is freely available upon request.

2.3 Chemical Analysis

Two chemical analyses were performed simultaneously to the bioassay analysis in two different laboratories: a rapid wide-spectrum screening and a more sensitive and targeted analysis.

2.3.1 Wide Spectrum Screening

Gas chromatography / mass spectrometry (GC/MS) wide spectrum screening was used to measure 17β-estradiol, estrone, estriol, tamoxifen, α-endosulfan, β-endosulfan, dieldrin, p,p'-DDT, 4-*t*-octylphenol, nonylphenol, bisphenol A and benzyl butyl phthalate in most samples.

The analysis was performed using an Agilent 6890 gas chromatograph coupled to an Agilent 5973 mass-selective detector (Agilent Technologies, Palo Alto, CA, USA). Separation was accomplished on a DB-5MS fused silica column (30 m × 0.25 mm i.d.; 0.5 µm film thickness, Agilent Technologies). The oven temperature program was 4 min at 50°C, 8°C/min to 150°C, 7°C/min to 250°C, 8°C/min to 300°C and then held at 300°C for 4 min. Column pressure was set at 70 kPa. Helium was used as the gas carrier at a constant flow of 1.2 mL/min. The transfer line was heated to 300°C and the source at 250°C. Sample injection (2 µL) was in splitless mode.

Table 4. GC/MS Parameters for Wide Spectrum Screening.

GC/MS system:	GC 6890 (Agilent Technologies)
Injection system:	Split/splitless injector
Injection volume:	2 µL (splitless)
Injection temperature:	250°C
Separation column:	DB-5MS (30 m x 0.25 mm x 0.5 µm)
Carrier gas:	Helium (1.2 mL/min)
Temperature program:	50°C (4 min), 8°C/min to 160°C, 7°C/min to 250°C, 8ºC/min to 300ºC, 300ºC (4 min)
Detector:	MSD 5973 (Agilent Technologies)
Detector temperature:	300°C
Transfer-line temperature:	250°C
Scan Mode:	Single ion monitoring (SIM)

This analysis was done at Griffith University on behalf of the CRC for Water Quality and Treatment in Australia. A standard operating protocol (SOP) is freely available upon request.

2.3.2 Detailed Analysis

A combination of techniques was used at TechnologieZentrum Wasser (TZW) to perform a more comprehensive analysis of selected chemical contaminants in the samples. First, the dried sample aliquot was reconstituted with 1 mL of a 1:1 mixture (v:v) acetone/n-hexane. To better dissolve the residue, the solution was treated for 5 minutes in an ultrasonic bath. The solution was split then into four parts and each aliquot was analysed by a different analytical technique for different analytes. For each method, quantification was done using a calibration recorded in the direct injection mode, i.e. by adding different known amounts of target analytes to an acetone/n-hexane mixture and by treating these solutions the same way that the samples were treated.

Analysis of steroidal hormones and alkylphenols

Analysis of the steroidal hormones 17α-ethinylestradiol, 17β-estradiol, estrone and estriol as well as of the xeno-estrogens 4-nonylphenol (technical mixture of isomers), 4-*t*-octylphenol, and bisphenol A was done by GC/MS after silylation of the analytes. First, 500 ng of chrysene-d12 (100 ng/µL solution in acetone) were added as internal standard to an aliquot of the original extract. Then the organic solvent was evaporated to dryness in a drying oven at 80°C. The dry residue was reconstituted with 100 µL of a silylating reagent mixture (N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA)/2 % trimethyliodo silane). After a reaction time of 20 minutes at 80°C in a drying oven, determination of the derivatives was done by GC/MS. Details of the GC/MS method are summarised in the following table:

Table 5. GC/MS Parameters for the Determination of Steroidal Hormones and Alkylphenols.

GC/MS system:	GC 6890 (Agilent Technologies)
Injection system:	Split/splitless injector
Injection volume:	2 µL (splitless, 1.5 min)
Injection temperature:	250°C
Separation column:	HP 5-ms (30 m x 0.25 mm x 0.25 µm)
Carrier gas:	Helium (1.2 mL/min)
Temperature program:	50°C (2 min), 16°C/min to 180°C, 5°C/min to 290°C (10 min)
Detector:	MSD 5973N (Agilent Technologies)
Detector temperature:	280°C
Transfer-line temperature:	280°C
Scan Mode:	Single ion monitoring (SIM)

Analysis of benzylbutyl phthalate (BBP), tetrabromobisphenol A (TBBPA) and tamoxifen

Analysis of BBP, TBBPA and tamoxifen was done by GC/MS. If necessary (e.g. for some of the spiked samples), the aliquot was further diluted with a 1:1 mixture (v:v) acetone/n-hexane. Then 1 µL of a 45 µg/µL solution of BDE-7 in acetone was added as internal standard and an aliquot was injected into the GC/MS system. Major parameters and conditions of the GC/MS analysis are summarised in the following table:

Table 6. GC/MS Parameters for the Determination of Benzylbutyl Phthalate (BBP), Tetrabromobisphenol A (TBBPA) and Tamoxifen.

GC/MS system:	GC 6890 (Agilent Technologies)
Injection system:	Cold-injection system (Gerstel)
Injection volume:	2 µL (splitless, 1 min)
Injection temperature:	60°C (0.5 min), 12°C/s to 320°C (10 min)
Separation column:	Zebron ZB-1 (15 m x 0.25 mm x 0.1 µm)
Carrier gas:	Helium (1.2 mL/min)
Temperature program:	60°C (1 min), 20°C/min to 180°C (2 min), 5°C/min to 260°C (1 min), 25°C/min to 310°C
Detector:	MSD 5973N (Agilent Technologies)
Detector temperature:	280°C
Transfer-line temperature:	280°C
Scan Mode:	Single ion monitoring (SIM)

Analysis of pesticides

Analysis of the pesticides p,p'-DDT, dieldrin and endosulfan was done by gas chromatography / electron capture detection (GC/ECD). For confirmation of results double-column technique was used. For endosulfan, both isomers (α- and β-form) were determined. If necessary (e.g. for some of the spiked samples), the aliquot was further diluted with a 1:1 mixture (v:v) acetone/n-hexane. Then 10 µL of 1 ng/µL solution of hexachloro-p-xylene in acetone were added as internal standard and an aliquot was injected into the GC system. Details of the GC analysis are summarised in the following table:

Table 7. GC Parameters for the Determination of P,P'-DDT, Dieldrin and Endosulfan.

GC system:	Autosystem XL (PerkinElmer)
Injection system:	Programmable split/splitless injector (PSSI)
Injection volume:	2 µL (splitless, 0.3 min)
Injection temperature:	50°C (0.3 min), 999°C/min to 270°C (56 min)
Separation column A:	Supelco MDN 5 (30 m x 0.32 mm x 0.25 µm)
Separation column B:	Restek CLP (30 m x 0.32 mm x 0.5 µm)
Carrier gas:	Helium (3.8 mL/min for each column)
Temperature program:	60°C (1 min), 20°C/min to 180°C, 3°C/min to 207°C, 1.5°C/min to 260°C (5 min)
Detector:	Electron-capture detector (ECD)
Make-up gas:	Argon/methane

Analysis of genistein

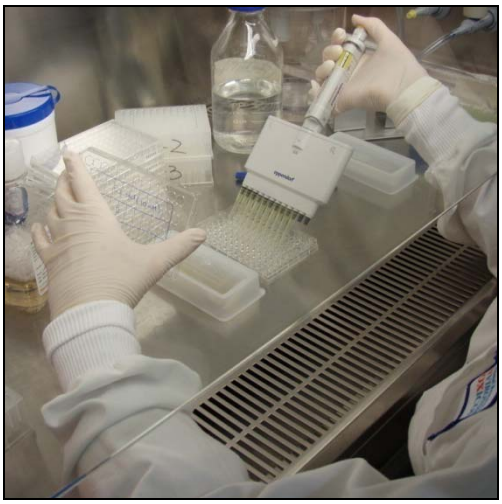
Analysis of genistein was done by liquid chromatography coupled to tandem mass spectrometry by an electrospray interface (HPLC-ESI-MS-MS). If necessary (e.g. for some of the spiked samples), the aliquot was further diluted with MilliQ water. Then an aliquot was injected into the HPLC-ESI-MS-MS system. Details of the HPLC-ESI-MS-MS method are summarised in the following table:

Table 8. HPLC-MS-MS Parameters for the Determination of Genistein.

HPLC system:	HPLC 1090 (Agilent Technologies)
Injection volume:	50 µL
Injection temperature:	50°C (0.3 min), 999°C/min to 270°C (56 min)
Separation column:	Phenomenex Luna C-18 (250 mm x 2 mm, 5 µm)
Eluent A:	Water + 20 mM ammonium formiat
Eluent B:	2/3 Acetonitrile, 1/3 methanol + 20 mM ammonium formiat
Gradient:	0 min: 95% A 5% B
	2 min: 95% A 5% B
	12 min: 0% A 100% B
	19 min: 0% A 100% B
	20 min: 95% A 5% B
	21 min: 95% A 5% B
Flow rate:	0.2 mL/min
Detector:	PE Sciex API 2000 triple quadrupole MS
Interface:	Electrospray (ESI)
Ionisation mode:	negative
Ionisation voltage:	-4500 V

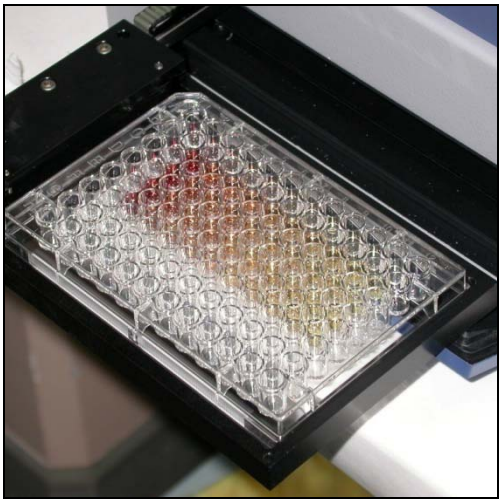
Picture 13. Performing the E-Screen at the EnTox cell laboratory.

Picture 16. GCMS analysis at Griffith University.



Picture 14. Seeding MCF7 cells in 96-well plates.

Picture 17. GCMS at TZW.



Picture 15. The number of viable cells is measured by addition of MTT.

Picture 18. HPLC at TZW.

3.0 Results and Discussion

With the exception of the MELN assay, the results from all bioassays (YES, ER-CALUX, KBluc and E-Screen assays) with both artificial and environmental samples showed very similar trends. Some inter-assay differences were observed, such as a higher estrogenic activity with p,p'-DDT in the MELN assay or octylphenol in the YES and E-Screen assays. The results from the MELN assay were significantly different from those obtained with the other assays, although the trends were again similar, suggesting perhaps that the induction in the MELN assay was not as high.

When combined with chemical analysis, there was good agreement between the measured activity in the ER-CALUX and E-Screen assays and the predicted estrogenic activity based on chemical measurement, suggesting that both of those assays can be reliably combined with chemical analysis to provide a comprehensive screening of complex samples. Based on a more limited dataset, the KBluc assay appears very similar to the ER-CALUX, and may be a useful tool in the future. There was also fair agreement with the YES assay, but its comparatively high detection and quantification limits meant many samples were nondetects, thus limiting the YES assay's usefulness in environmental monitoring of very low-level contamination. The MELN assay results were generally lower than what would have been predicted from the chemical analysis alone, suggesting possible matrix interference or a higher sensitivity to anti-estrogenic chemicals in that assay.

3.1 Chemistry

With a few minor exceptions, most of the chemical data from the two laboratories were in good agreement (Table 20, Table 21, and Table 22). The data from both laboratories were combined using weighted averages to give more influence to the more precise data (i.e. with smaller standard deviation; see notes on Table 9 for equations). The combined data are presented in Table 9 and Table 10.

Table 9. Combined Chemical Data (1 Of 2) - Weighted Averages ± SD for Industrial Chemicals.

Sample ID	Description	4t-OP (µg/L)	NP (µg/L)	Bisphenol A (µg/L)	BBP (µg/L)	TBBPA (µg/L)
A01	Artificial - Estradiol	< 5	< 5	< 5	< 50	
A02	Artificial - Nonylphenol	< 5	238 ± 20.1	< 5	< 50	
A03	Artificial - Octylphenol	1,609 ± 113	< 5	< 5	< 50	
A04	Artificial - Benzy Butyl Phthalate	< 5	< 5	< 5	3,994 ± 535	
A05	Artificial - TBBPA	< 5	< 5	< 5	< 50	1,483 ± 839
A06	Artificial - p,p'-DDT	5.97 ± 6.00	< 5	< 5	< 50	
A07	Artificial - Tamoxifen	< 5	< 5	< 5	< 50	
A08	Artificial - Hormone mix	< 5	< 5	< 5	< 50	
A09	Artificial - Industrial mix	524 ± 132	78.1 ± 2.87	39.7 ± 2.97	606 ± 86.4	450 ± 180
A10	Artificial - Agro-chemical mix	< 5	< 5	< 5	< 50	
A11	Artificial - Combo mix	527 ± 62.4	92.8 ± 2.64	45.4 ± 2.17	729 ± 204	192 ± 14.4
A12	Artificial - Blank	< 0.025	< 0.125	< 0.025	< 5	< 5
GW1-	Shallow aquifer (non spiked)	0.583 ± 0.454	0.271 ± 0.219	< 0.025	< 5	
GW1+	Shallow aquifer (spiked)	5.80 ± 5.10	94.1 ± 9.08	57.4 ± 5.50	720 ± 51.5	
GW2-	Deep aquifer (non spiked)	< 0.025	< 0.125	< 0.025	< 5	
GW2+	Deep aquifer (spiked)	0.300 ± 0.100	90.7 ± 11.6	53.2 ± 4.30	775 ± 103	
SR1-	Raw sewage 1 (non spiked)	2.62 ± 2.53	6.33 ± 1.15	0.433 ± 0.029	< 12.5	
SR1+	Raw sewage 1 (spiked)	7.28 ± 7.35	20.7 ± 2.03	47.2 ± 2.24	88.7 ± 52.1	
SR2-	Raw sewage 2 (non spiked)	0.400 ± 0.087	3.64 ± 5.52	0.633 ± 0.126	< 12.5	
SR2+	Raw sewage 2 (spiked)	0.450 ± 0.100	14.8 ± 0.763	46.1 ± 2.77	71.5 ± 10.1	
ST1-	Treated sewage 1 (non spiked)	0.500 ± 0.278	1.12 ± 0.076	0.050 ± 0.000	< 5	
ST1+	Treated sewage 1 (spiked)	4.20 ± 5.47	27.8 ± 4.48	46.4 ± 2.87	332 ± 74.5	
ST2-	Treated sewage 2 (non spiked)	0.233 ± 0.104	0.338 ± 0.239	< 0.025	< 5	
ST2+	Treated sewage 2 (spiked)	0.817 ± 0.679	12.2 ± 1.02	18.5 ± 1.75	373 ± 97.3	
RW1-	River 1 (non spiked)	0.483 ± 0.029	0.221 ± 0.146	< 0.025	< 5	
RW1+	River 1 (spiked)	1.22 ± 0.321	38.7 ± 1.93	49.6 ± 3.90	341 ± 45.8	
RW2-	River 2 (non spiked)	0.367 ± 0.153	5.27 ± 4.97	< 0.025	< 5	
RW2+	River 2 (spiked)	0.383 ± 0.126	51.5 ± 2.52	48.2 ± 3.88	551 ± 58.4	

Notes: n = 3 independent samples. Weighted averages (x) calculated as $x = \frac{\sum_{i=1}^{n} x_i / \sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}$ and the weighted standard deviation (σ) is calculated as $\sigma^2 = \frac{1}{\sum_{i=1}^{n} 1/\sigma_i^2}$.

Table 10. Combined Chemical Data (2 Of 2) - Weighted Averages ± SD for Natural Hormones, Drugs, Pesticides, and Phytoestrogens.

Sample ID	Description	Estrone (µg/L)	β-Estradiol (µg/L)	Estriol (µg/L)	Tamoxifen (µg/L)	EE2 (µg/L)	Endosulfan (µg/L)	Dieldrin (µg/L)	p,p'-DDT (µg/L)	Genistein (µg/L)
A01	Artificial - Estradiol	< 0.01	< 0.005	< 0.01	< 0.1		< 10	< 10	< 50	
A02	Artificial - Nonylphenol	< 0.01	< 0.005	< 0.01	< 0.1		< 10	< 10	< 50	
A03	Artificial - Octylphenol	< 0.01	< 0.005	< 0.01	< 0.1		< 10	< 10	< 50	
A04	Artificial - Benzy Butyl Phthalate	< 0.01	< 0.005	< 0.01	< 0.1		< 10	< 10	< 50	
A05	Artificial - TBBPA	< 0.01	< 0.005	0.028 ± 0.040	< 0.1		< 10	< 10	< 50	
A06	Artificial - p,p'-DDT	< 0.01	< 0.005	< 0.01	< 0.1		11.8 ± 8.80	15.0 ± 17.4	2,003 ± 218	
A07	Artificial - Tamoxifen	< 0.01	< 0.005	< 0.01	17.4 ± 1.53		< 10	< 10	< 50	
A08	Artificial - Hormone mix	0.014 ± 0.003	< 0.005	0.006 ± 0.003	< 0.1	0.006 ± 0.002	< 10	< 10	< 50	
A09	Artificial - Industrial mix	< 0.01	< 0.005	< 0.01	< 0.1		< 10	< 10	< 50	
A10	Artificial - Agro-chemical mix	< 0.01	< 0.005	< 0.01	< 0.1		568 ± 28.9	443 ± 40.4	424 ± 48.6	4.01 ± 3.49
A11	Artificial - Combo mix	0.039 ± 0.007	0.009 ± 0.001	0.048 ± 0.006	21.6 ± 2.35	0.010 ± 0.002	488 ± 18.8	456 ± 22.9	475 ± 47.5	8.50 ± 0.500
A12	Artificial - Blank	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.05	< 0.05	< 0.05	< 0.05
GW1-	Shallow aquifer (non spiked)	< 0.005	< 0.005	< 0.005	< 0.1	< 0.005	< 10	< 10	< 50	
GW1+	Shallow aquifer (spiked)	0.075 ± 0.022	0.025 ± 0.005	0.018 ± 0.013	< 0.1	< 0.005	< 10	< 10	< 50	
GW2-	Deep aquifer (non spiked)	< 0.005	< 0.005	< 0.005	< 0.1	< 0.005	< 10	< 10	< 50	
GW2+	Deep aquifer (spiked)	0.061 ± 0.005	0.014 ± 0.004	0.042 ± 0.003	< 0.1	0.008 ± 0.004	< 10	< 10	< 50	
SR1-	Raw sewage 1 (non spiked)	0.017 ± 0.012	0.016 ± 0.011	0.181 ± 0.009	< 0.25	< 0.013	< 25	< 25	< 125	
SR1+	Raw sewage 1 (spiked)	0.277 ± 0.014	0.033 ± 0.008	0.197 ± 0.010	< 0.25	< 0.013	< 25	< 25	< 125	
SR2-	Raw sewage 2 (non spiked)	0.083 ± 0.021	0.024 ± 0.011	0.159 ± 0.057	< 0.25	< 0.013	< 25	< 25	< 125	
SR2+	Raw sewage 2 (spiked)	0.134 ± 0.001	0.043 ± 0.010	0.279 ± 0.065	< 0.25	< 0.013	< 25	< 25	< 125	
ST1-	Treated sewage 1 (non spiked)	0.023 ± 0.021	< 0.005	< 0.005	< 0.1	< 0.005	< 10	< 10	< 50	
ST1+	Treated sewage 1 (spiked)	0.138 ± 0.003	0.027 ± 0.005	0.055 ± 0.005	< 0.1	< 0.005	< 10	< 10	< 50	
ST2-	Treated sewage 2 (non spiked)	< 0.005	< 0.005	< 0.005	< 0.1	< 0.005	< 10	< 10	< 50	
ST2+	Treated sewage 2 (spiked)	< 0.005	0.007 ± 0.005	< 0.005	< 0.1	< 0.005	< 10	< 10	< 50	
RW1-	River 1 (non spiked)	< 0.005	< 0.005	< 0.005	< 0.1	< 0.005	< 10	< 10	< 50	
RW1+	River 1 (spiked)	0.070 ± 0.007	0.010 ± 0.001	0.040 ± 0.005	< 0.1	0.006 ± 0.004	< 10	< 10	< 50	
RW2-	River 2 (non spiked)	0.007 ± 0.007	< 0.005	< 0.005	< 0.1	< 0.005	< 10	< 10	< 50	
RW2+	River 2 (spiked)	0.088 ± 0.005	0.008 ± 0.003	0.045 ± 0.005	< 0.1	0.008 ± 0.003	< 10	< 10	< 50	

Notes: n = 3 independent samples. Weighted averages (x) calculated as $x = \frac{\sum_{i=1}^{n} x_i / \sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}$ and the weighted standard deviation (σ) is calculated as $\sigma^2 = \frac{1}{\sum_{i=1}^{n} 1/\sigma_i^2}$.

To estimate recovery of estrogenic activity from unprocessed environmental samples, all samples were spiked *before* pre-treatment (i.e. before centrifugation and/or filtration). Recovery is thus a combination of loss due to partitioning into the solid phase as well as extraction efficiency. It is therefore not surprising to find relatively poor recovery of the spiked compounds in raw sewage where a significant portion of chemicals would likely have partitioned into the suspended phase. Based on the combined data, recovery efficiencies were calculated for both the artificial and the environmental samples. Spike recoveries in the artificial samples varied between 12 to 100% (average 54.7%) for the different compounds (Table 11). Although these recoveries are lower than would be expected when compared to previous studies [14], they were relatively consistent between sample replicates with an average coefficient of variation of 17 %. Possible explanations for the lower-than-expected recoveries are the larger volume of original sample used (2.5 L instead of 1 L) that could lead to decreased performance of the SPE cartridges, as well as the more extensive sample preparation and worldwide shipping involved in this project. As expected, net recovery of the spiked compounds (nonylphenol, bisphenol A, benzyl butyl phthalate, estrone, estradiol, estriol and ethynylestradiol) in the environmental samples was strongly dependent upon the sample matrix (Figure 1). It was most reliable in groundwater, with an average net recovery of 96 ± 31 % for the seven compounds combined. In raw sewage on the other hand, net recovery was much more variable with an average of 46 ± 40 %. The industrial compounds in particular (nonylphenol, bisphenol A and benzyl butyl phthalate) yielded poor recoveries in raw sewage with an average recovery of 15 ± 18 %. Treated sewage and river water yielded 66 ± 38 and 78 ± 38 % average net recoveries, respectively (Figure 1).

It is important to highlight that all comparisons between chemical and bioanalytical data in this document are based on actual concentrations of each selected chemical as determined by chemical analysis (reported in Table 9 and Table 10) and not the nominal concentrations (given in Table 2). Thus the recovery statistics are given purely to illustrate the difficulty of extracting organic compounds from complex matrices such as raw sewage, as recovery does not affect any of the results or their interpretation.

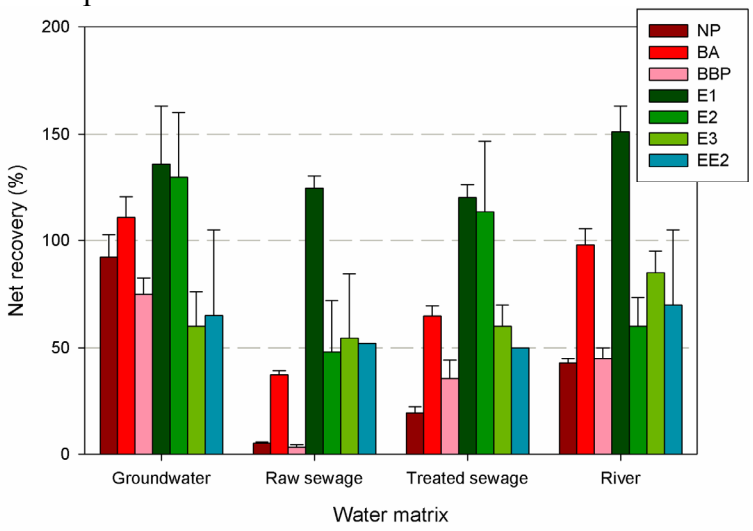


Figure 1. Net Recovery of Spiked Compounds in the Environmental Samples Based on Combined Chemical Data.
Notes: n = 3 independent samples.

Net recovery R was calculated as $R = \frac{\text{(Amount in spiked sample) - (Amount in unspiked sample)}}{\text{Spiked amount}} \times 100$.

Table 11. Average Extraction Recoveries of Artificial Samples in Tap Water Based on Combined Chemical Data.

Mix ID	Mix name	Chemical(s)	Average recovery (% ± SD)
A01	Estradiol	17β-Estradiol (E2)	< 33.3
A02	Nonylphenol	Nonylphenol (NP)	59.5 ± 5.0
A03	Octylphenol	4-*t*-Octylphenol (OP)	32.2 ± 2.3
A04	Benzyl butyl phthalate	Benzyl butyl phthalate (BBP)	79.9 ± 10.7
A05	Tetrabromobisphenol A	Tetrabromobisphenol A (TBBPA)	29.7 ± 16.8
A06	p,p'-DDT	p,p'-DDT	40.1 ± 4.4
A07	Tamoxifen	Tamoxifen (TMX)	34.8 ± 3.1
A08	Hormone mix	17β-Estradiol (E2)	< 50.0
		Estrone (E1)	28.0 ± 6.0
		Estriol (E3)	12.0 ± 6.0
		Ethinylestradiol (EE2)	60.0 ± 20.0
A09	Industrial mimics mix	Nonylphenol (NP)	78.1 ± 2.9
		4-*t*-Octylphenol (OP)	52.4 ± 13.2
		Bisphenol A (BA)	79.4 ± 5.9
		Tetrabromobisphenol A (TBBPA)	45.0 ± 18.0
		Benzyl butyl phthalate (BBP)	60.6 ± 8.6
A10	Agro-industrial mix	Dieldrin	44.3 ± 4.0
		Endosulfan (α and β)	56.8 ± 2.9
		p,p'-DDT	42.4 ± 4.9
		Genistein	20.1 ± 17.5
A11	Combo mix	Benzyl butyl phthalate (BBP)	72.9 ± 20.4
		Bisphenol A (BA)	90.8 ± 4.3
		Dieldrin	45.6 ± 2.3
		Ethinylestradiol (EE2)	100 ± 20.0
		Endosulfan (α and β)	48.8 ± 1.9
		17β-Estradiol (E2)	90.0 ± 10.0
		Estrone (E1)	78.0 ± 14.0
		Estriol (E3)	96.0 ± 12.0
		Genistein	42.5 ± 2.5
		Nonylphenol (NP)	92.8 ± 2.6
		4-*t*-Octylphenol (OP)	52.7 ± 6.2
		p,p'-DDT	47.5 ± 4.8
		Tetrabromobisphenol A (TBBPA)	19.2 ± 1.4
		Tamoxifen (TMX)	43.2 ± 4.7
A12	Blank (tap water only)	None	N/A

Notes: n = 3 independent samples. For initial spike concentration, see Table 2.

Recovery R in artificial samples are calculated as $R = \frac{\text{Amount in extracted sample}}{\text{Spiked amount}} \times 100$.

3.2 Bioassays

All measurements in the bioassays are expressed as estradiol equivalents (EEq). This is equivalent to the amount of estradiol that would have to be present in the sample to induce the specific response measured in the bioassay. The results for the artificial samples and environmental samples are presented in Figure 2 and Figure 4, respectively. All data were log-transformed to be compatible with statistical assumptions of normality. Samples below the quantification limit (nondetects) were dealt with separately depending on purpose (see Section 3.5.4 on page 69 for further discussion into this issue):

- For calculations (e.g., averages, standard deviations) and graphical representations, nondetects were assigned a value of half the quantification limit.
- For statistical analysis (correlation and statistical differences between the datasets), nondetects were excluded from the dataset. Nondetects are discussed semi-quantitatively in Figure 3 and Figure 5.
- The full dataset (including nondetects) was used for survival (Kaplan-Meier) analysis when comparing the performance of the assays irrespective of the sample source. This analysis is presented in Section 3.2.3 on page 52.

3.2.1 Artificial Samples

Bioassays conducted on the different artificial samples displayed a very similar trend (Figure 2), with the highest reported estrogenicity for octylphenol (except in the ER-CALUX assay, where OP was cytotoxic and its estrogenicity could not be determined); there was no detectable estrogenic activity in the blank sample.

YES assay:

Many of the artificial samples had an estrogenic activity below the method quantification limit (MQL) of the YES assay (Figure 2A), which was relatively high (3.5 and 5 ng/L for the first and second YES assays, respectively, compared with 0.2 ng/L for the E-Screen assay for example). In fact, only 16 out of 36 samples displayed estrogenic activity higher than the quantification limit of the YES assays (i.e., 56% of nondetects). When detectable however, the responses were generally higher than in other bioassays. The activities reported from the two laboratories conducting the assay were not significantly different (paired t-test, $p = 0.80$; Table 12) and the two datasets were well correlated ($R^2 = 0.71$; Table 12). There were a few disagreements however with the benzyl butyl phthalate samples (BBP) inducing an average of 58.3 ng/L EEq in the first assay and < 5 ng/L EEq in the second, and tetrabromobisphenol A (TBBPA) was cytotoxic in the first but not the second YES assay (although it did not induce measurable estrogenicity in either assays). Of all the bioassays tested, the YES assay was the only one that was not able to quantify the estrogenic activity in the estradiol samples (low-level activity was detected, but insufficient for accurate quantification). Octylphenol was more potent than expected from literature reports (Table 1) and the samples containing octylphenol had the highest estrogenic activity. On the other hand the p,p'-DDT samples did not produce any estrogenic activity in either of the YES assays and neither did the agro-chemical mix (which contained p,p'-DDT, endosulfan, dieldrin and genistein). The other chemicals were within the range of reported potencies (Table 1).

When combined (by averaging the results from the two assays for each sample), the estrogenic activity demonstrated by the YES assay was significantly higher than that observed

for most of the other bioassays (paired t-test, $p \leq 0.01$ for all assays except ER-CALUX where $p = 0.17$; Table 12) but well correlated with all other tests with the exception of the MELN assay (average R^2 excluding MELN = 0.61; Table 12). This indicates that while YES assay results for the artificial samples were generally higher (paired t-test), the data trends were very similar with what is reported for other assays (correlation). This is also clear in the pairwise comparison including nondetects (Figure 3) where there is generally a clear linear trend between the YES assay results and the other assays, except the MELN assay. The YES assays had the highest number of samples below quantification limit, with 64% of nondetects in the combined YES assay dataset. This is also evident in the YES assay pairwise comparisons (Figure 3) with a large proportion of points below the quantification limit (dashed line).

ER-CALUX assay:

The ER-CALUX (Figure 2B) had the lowest MQL (0.1 ng/L) of all assays tested in this study. It detected activity in all samples except the octylphenol (OP) samples (which could not be measured because of cytotoxicity) and the blank samples (< 0.1 ng/L EEq) (19% of nondetects). The estrogenic activity of the BBP sample was similar to that previously reported in the literature (Table 1); TBBPA showed only limited estrogenic activity. Although OP was cytotoxic on its own, the cytotoxic effect was not detectable in the combo and industrial mixes. This may be because OP is present at a lower concentration in the mixtures (1000 µg/L; A09 and A11, Table 2) than on its own (5000 µg/L; A03, Table 2). In fact those mixes displayed the highest estrogenic activity.

The ER-CALUX assay results with the artificial samples were very similar to and well correlated with those obtained in the other bioassays with the exception of the MELN assay (paired t-test, $p > 0.10$ except for MELN where $p = 0.03$; average R^2 excluding MELN = 0.66; Table 12). Its excellent quantification limit is evidenced by the limited number of samples below quantification limit in the ER-CALUX assay pairwise comparisons including nondetects (Figure 3).

MELN assay:

The MELN assay (Figure 2C) was the most different of all bioassays, as indicated by its much lower correlation with the other assays (average $R^2 = 0.26$; Table 12). In particular, the p,p'-DDT samples and the pesticide mix showed relatively high estrogenic activity, a trend that is different from all the other assays and may indicate a higher sensitivity of the MELN assay to pesticides (Figure 2). There is only a limited set of potencies for pesticides available from the literature, so it is unclear how the MELN assay results from this study compared with previous MELN assays (Table 1). Although some estrogenic activity was detected with the hormone mix, this was not sufficient to be accurately quantifiable (below quantification limit) in this assay.

Overall, the MELN assay results were significantly lower than those obtained in the (combined) YES assays, the (combined) E-Screen and the ER-CALUX assays (paired t-test, $p < 0.05$; Table 12). For example, the highest activity in the octylphenol sample was 7.77 ng/L EEq compared to 3457 ng/L EEq in the first YES assay or 972 ng/L EEq in the first E-Screen assay. It was poorly correlated with most other assays in this study (Figure 3) except the KBluc assay limited dataset, and had the second highest proportion of nondetects (39%). The results with individual compounds (samples A01 – A07, Table 2) was however very similar to those obtained with the KBluc assay limited dataset (paired t-test, $p = 0.95$; R2 = 0.43; Table 12).

KBluc assay:
The KBluc dataset is more limited than that of the other assays, with only the individual compounds analyzed (samples A01 – A07; Figure 2D). The limited results indicate that the KBluc assay appears well correlated with the other assays (average $R^2 = 0.63$; Table 12) and comparable in values to the ER-CALUX, the E-Screen and the MELN assays, while being generally lower than the (combined) YES results ($p > 0.25$ for all assays except the combined YES assay where $p = 0.01$; Table 12). This conclusion is supported by pairwise comparison that includes nondetects (Figure 3). The octylphenol samples again displayed the highest estrogenic activity; TBBPA was only barely quantifiable. The KBluc assay is a relatively recent assay, and few data are available in the literature for comparison (Table 1).

E-Screen assay:
Finally, the E-Screen assay data (Figure 2E) were similar and well correlated with data from most other assays (average R^2 excluding MELN = 0.74; Table 12). The data from the two E-Screen assays were not significantly different (paired t-test, $p = 0.08$; Table 12) and well-correlated ($R^2 = 0.67$; Table 12) although the second E-Screen assay appeared more sensitive to p,p'-DDT (Figure 2E). Estrogenic activity was detected in all samples except the TBBPA sample in the second assay and the blank samples in both E-Screen assays (Figure 2E). The potency of individual chemicals tested in this study was within the range of those reported previously (Table 1). Pairwise comparisons including nondetect) highlight the good correlation between the E-Screen assay and other assays (except the MELN assay) and its small proportion of nondetects (22% and 33% for the first and second E-Screen assays, respectively).

Table 12. P Values of Paired T-Tests, Coefficient of Determination (R^2) and Sample Size (N) on Log-Transformed Data Between the Different Bioassay for the Artificial Samples (Excluding Nondetects).

Assay	Statistic	YES [a]	ER-CALUX	MELN	KBluc	E-Screen [a]
YES [a]	p	0.80 [b]	0.17	**< 0.01 ****	**0.01 ***	**< 0.01 ****
	R^2	**0.71 ** [b]**	**0.47 ***	0.27	0.65	**0.71 ****
	n	13 [b]	10	9	5	12
ER-CALUX	p	0.17	--	**0.03 ***	0.50	0.11
	R^2	**0.47 ***	--	< 0.01	**0.69 ***	**0.81 ****
	n	10	--	18	11	18
MELN	p	**< 0.01 ****	**0.03 ***	--	0.95	**< 0.01 ****
	R^2	0.27	< 0.01	--	**0.43 ***	0.08
	n	9	18	--	11	14
KBluc	p	**0.01 ***	0.50	0.95	--	0.25
	R^2	0.65	**0.69 ***	**0.43 ***	--	**0.74 ****
	n	5	11	11	--	9
E-Screen [a]	p	**< 0.01 ****	0.11	**< 0.01 ****	0.25	0.08 [b]
	R^2	**0.71 ****	**0.81 ****	0.08	**0.74 ****	**0.67 ** [b]**
	n	12	18	14	9	21 [b]

Notes: Numbers in bold are statistically significant. The * indicates $p < 0.05$, ** indicates $p < 0.01$. Analysis does not include nondetects. The statistical significance of the coefficient of determination (R^2) was calculated from its t-value using a two-tailed t-distribution.

(a) Data from multiple assays combined into one larger dataset.

(b) Value is for comparison of the same assay but performed in two different laboratories (eg YES 1 vs YES 2).

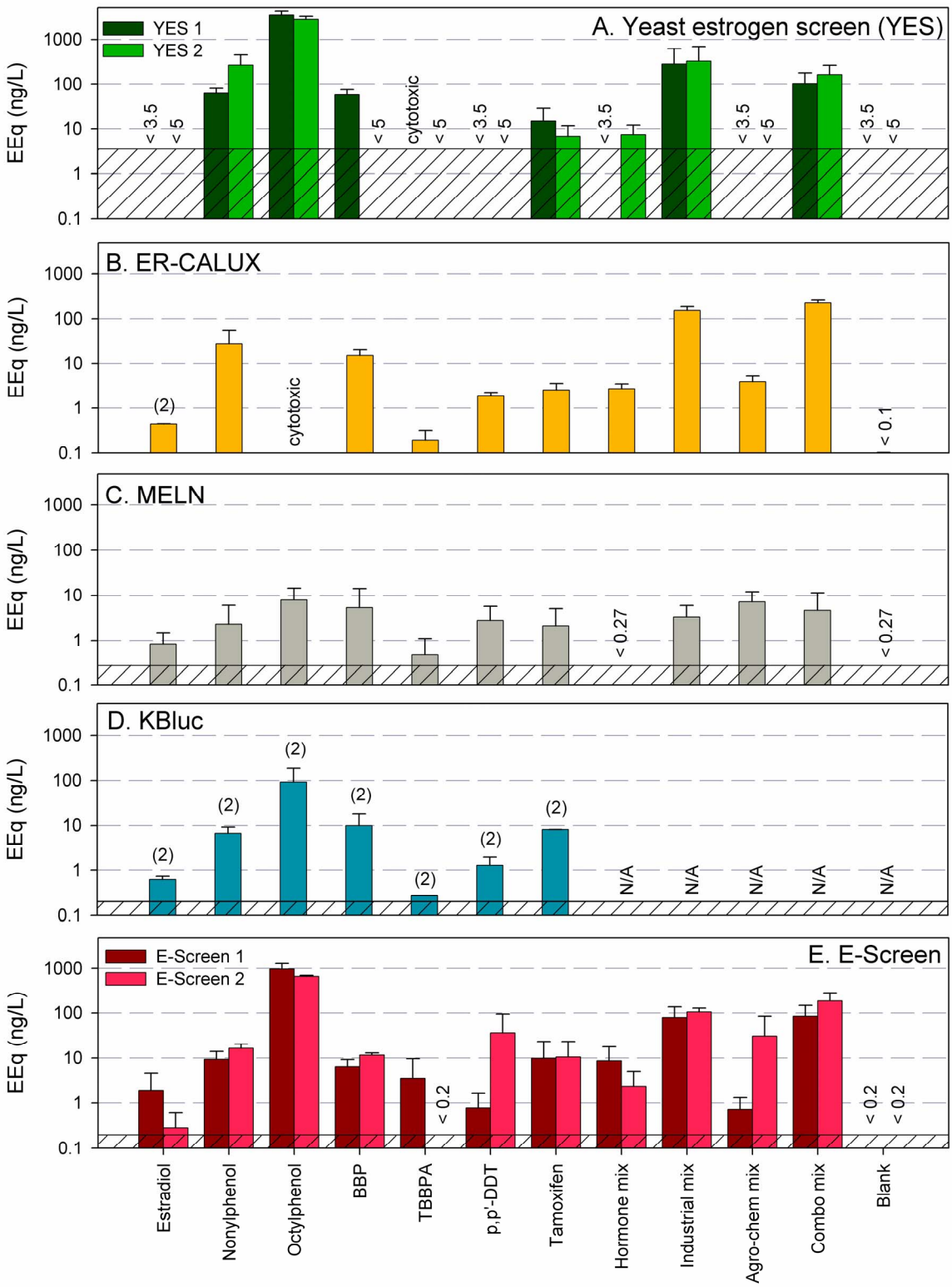


Figure 2. Estrogenicity (Expressed as Estradiol Equivalents, Eeq) in the Artificial Samples as Determined by Bioassay.
Notes: n = 3 independent samples unless otherwise indicated by a number in brackets above the bars. The cross-haired area indicates data below the method quantification limit. The YES and E-Screen assays were run in two separate laboratories. Error bars represent standard deviation. Data includes nondetects at half the quantification limit.

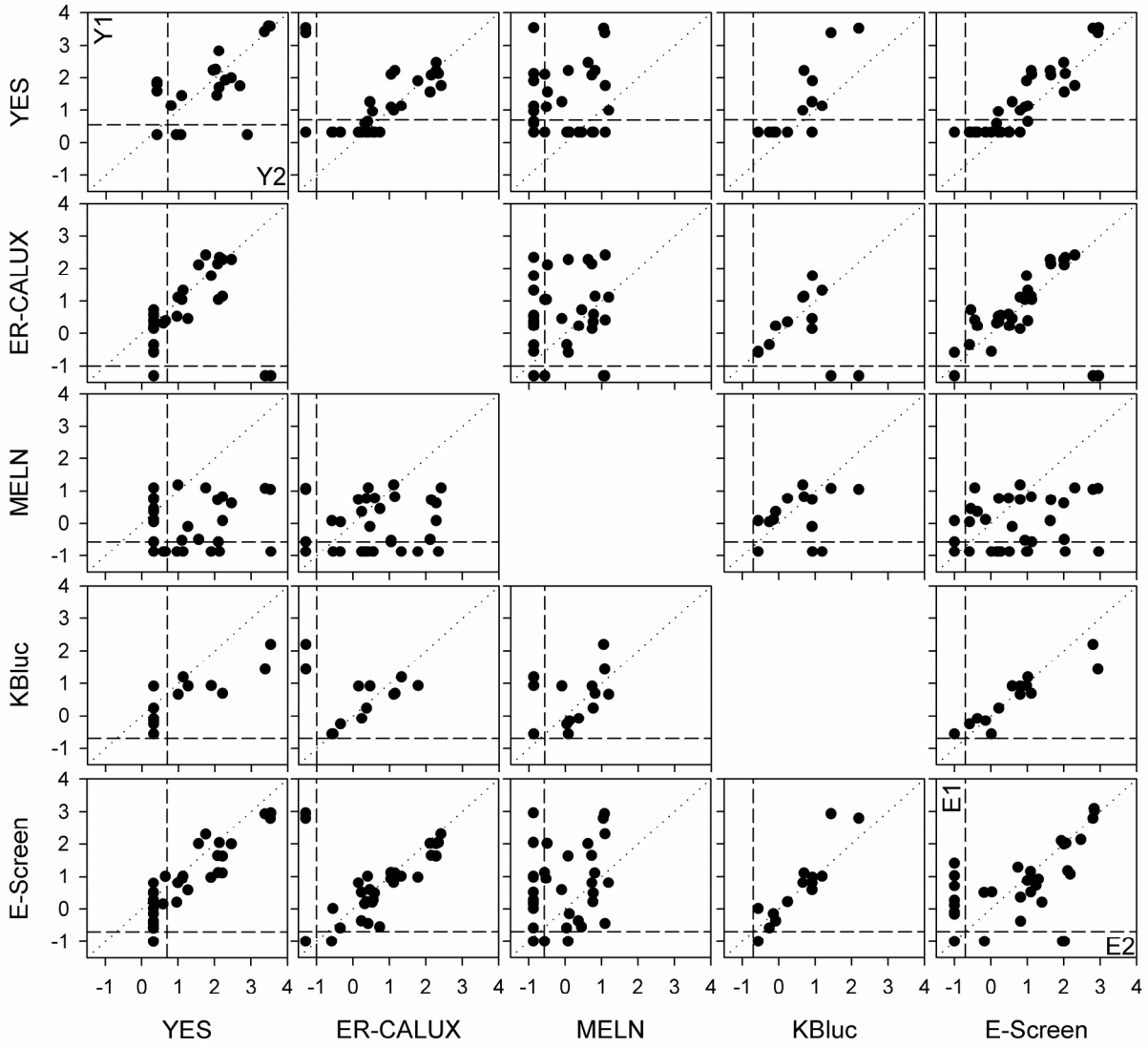


Figure 3. Pairwise Comparisons of Activity (Log Eeq) Between the Different Bioassays with Artificial Samples (Including Nondetects).

Notes: The dotted line is the isometric line. Dashed lines mark the method quantification limit for that assay. Nondetects were assigned a value equal to half the quantification limit. For the YES and E-Screen assays, the dataset was combined for comparisons with other assays but the two assay datasets are compared on their relevant lines (and identified as Y1 and Y2 for the YES datasets 1 and 2, and E1 and E2 for the E-Screen datasets 1 and 2, respectively).

3.2.2 Environmental Samples

Overall, all bioassays displayed similar trends with the environmental samples, albeit with different amplitudes (Figure 4 and Table 13). Statistical correlations between assays were not as pronounced as with the artificial samples, though this was strongly influenced by the removal of the censored (nondetect) data from the analysis. The environmental dataset is strongly binomially-distributed (Figure 5), with high levels of activity in the spiked samples and low levels in the natural samples (Figure 4). When the nondetects were assigned a value of half the quantification limit, the coefficients of correlation were much higher (with R2 > 0.75 in all cases) clearly highlighting the good agreement between the different assays in identifying samples with high vs. low activity (Figure 5).

Of the non spiked environmental samples, only the sewage samples displayed relatively high levels of estrogenic activity (> 10 ng/L EEq) in all assays (Figure 4, left). It may be worth keeping in mind that the median effective concentration of E2 required for a significant induction of the egg yolk precursor protein vitellogenin (a commonly-used biomarker of exposure to estrogenic EDCs) in juvenile rainbow trout after 2 weeks of exposure was estimated to be 10-20 ng/L [37, 38]. The final treated effluent exhibited significantly less estrogenic activity compared to raw sewage, illustrating good removal of estrogenic chemicals during sewage treatment processes. Low-level estrogenic activity was detected in the river water samples with some of the assays (particularly with river site 2), while estrogenicity in groundwater samples was below the quantification limit of most assays except the ER-CALUX and the second E-Screen assays (Figure 4, left). As expected, estrogenic activity was increased following addition of the spike (Figure 4, right). Spiked groundwater and river water samples were increased by a relatively predictable amount (Table 14), while spiked sewage samples were mostly increased but more variable. This again underlines the higher variability involved in extracting this type of sample (as illustrated with spike recoveries in Figure 1).

With the environmental samples, the KBluc assay was only performed on one independent sample (albeit run in triplicate assays). The comparisons with the KBluc dataset are therefore less precise than those among the other datasets (due to the smaller sample size), but the data is included nevertheless as indicative of the performance of the KBluc assay.

<u>YES assay:</u>
With the YES assay, the results of the first assay (Figure 4A) were significantly lower than those of the second (paired t-test, $p < 0.01$; Table 13). The two datasets were significantly correlated (p-correl = 0.03), but this correlation was not particularly meaningful ($R^2 = 0.17$; Table 13). Samples containing high concentrations of octylphenol in particular resulted in very high estrogenicity in the second YES assay. This interference by octylphenol in the YES assay was already identified with the artificial samples, with a very strong tendency for octylphenol to "creep" across the assay plate. When the data from both assays were combined (by averaging the results from the two assays for each sample), the estrogenic activity in the YES assay was significantly higher than what was reported in the ER-CALUX, MELN and KBluc assays but close to that reported in the E-Screen assay (paired t-test, $p < 0.01$, $p = 0.02$, $p < 0.01$ and $p = 0.93$, respectively; Table 13). However, the YES assay data were significantly correlated with those from both the ER-CALUX and the E-Screen assays, but not the MELN or KBluc assay ($R^2 = 0.50$ and 0.50, and $R^2 = 0.02$ and 0.40, respectively; Table 13). This indicates that the data

trends in the YES assay are similar to those reported in the ER-CALUX and E-Screen assays, albeit with different amplitudes. The YES assay again exhibited the highest proportion of nondetects (40%).

In the non spiked samples (Figure 4A, left), no activity was detected in any of the samples except the raw sewage samples (> 8.95 ng/L in the first and > 45.1 ng/L EEq in the second YES assay). Increased estrogenic activity compared to non spiked samples was detected in all spiked samples (Figure 4A, right), and raw sewage again exhibited the highest estrogenic activity of all samples. Based on the relative potency of the chemicals in the spike (Table 1) and assuming 100% recovery (i.e., nominal chemical concentrations in Table 3), it is possible to calculate a predicted increase in estrogenicity from the addition of the spike (Table 14). In the groundwater and river water samples (where spike recovery is usually good, see Figure 1), the measured increase in estrogenic activity (in ng/L EEq) was approximately 31.4% of the predicted increase (combined YES assay data, calculated from Table 14). This lower-than-expected increase could be an indication that the performance of the YES assay with environmental samples is not as commendable as with artificial samples. With the more complex matrix of sewage samples, the measured increase was even lower (Table 14). This could be due to already high levels in the non spiked samples (which would mask the comparatively more subtle effect of spiking) or the poor recoveries of the spikes in this complex matrix (Figure 1).

ER-CALUX assay:

The ER-CALUX assay results (Figure 4B) were significantly different from those of the other assays (paired t-test, $p < 0.05$; Table 13), but the data were significantly correlated with those of all other assays (average $R^2 = 0.67$; Table 13). The ER-CALUX assay detected significant activity in almost all samples, with only 3% of nondetects.

Estrogenic activity was detected in all non spiked samples (Figure 4B, left), including the groundwater samples (albeit at very low levels, < 0.211 ng/L EEq). The raw sewage samples exhibited the highest activity (> 80.1 ng/L), with sewage treatment greatly reducing estrogenic activity - but more so at the second treatment plant. The river water samples also induced low estrogenic activity in this assay, slightly higher in samples from site 2 (Oxley Creek; Picture 3). The river at site 2 is much smaller than at site 1 (Brisbane River; Picture 2) and is located in an industrial area, which may explain the higher estrogenic activity (due to the presence of potential industrial estrogen mimics and a reduced dilution effect). Higher activity was detected in all spiked samples (Figure 4B, right), and as was the case with the YES assay the highest activity was detected in the spiked raw sewage samples. The measured increase in estrogenic activity from addition of the spike in the groundwater and river water samples was 96.6% of the predicted increase (calculated from Table 14). With the sewage samples this ratio decreased slightly (Table 14), a point that has been previously discussed in the YES assay results.

MELN assay:

The results in the MELN assay (Figure 4C) were significantly different from those obtained with the (combined) YES, ER-CALUX and MELN assays but similar to those obtained in the (combined) E-Screen assay (paired t-test, $p = 0.02$, $p < 0.01$, $p < 0.01$ and $p = 0.28$, respectively; Table 13). The data from the MELN assay was fairly correlated with the data generated in the ER-CALUX and E-Screen assays (average $R^2 = 0.36$; Table 13), but not the YES or KBluc

assays ($R^2 < 0.03$; Table 13). Only 29% of the environmental samples were below the quantification limit of this assay.

The results from the non spiked samples (Figure 4C, left) show no activity in any of the groundwater and river water samples, high activity in the raw sewage samples (> 25.6 ng/L EEq), and detectable activity in only the first of the two treated sewage samples (3.70 ng/L EEq). Estrogenic activity was detected in all spiked samples (Figure 4C, right), but the activity in the raw sewage samples did not increase following addition of spike (Table 14). This may suggest interference with the assay in raw sewage, possibly due to cytotoxicity. The increase in activity was only 42.5% of the predicted increase in the groundwater and river water samples and only 24.4% of predicted increase with the treated sewage samples (calculated from Table 14).

KBluc assay:

The KBluc assay results (Figure 4D) were significantly different from those obtained with all other assays (paired t-test, $p < 0.05$; Table 13) but well correlated with the ER-CALUX and E-Screen assays ($R^2 > 0.90$; Table 13). Only 13% of all environmental samples were below quantification limit of this assay. Pairwise comparison including nondetects shows a good binomial agreement between nondetects and positive samples with the ER-CALUX and the E-Screen assays (Figure 5).

In the non spiked samples (Figure 4D, left), the KBluc assay displays a similar pattern to all other assays, with high estrogenic activity in the raw sewage samples (> 65 ng/L), lower activity in the treated sewage samples (undetectable in treated sewage from the second treatment plant), very low estrogenic activity in the river samples (with quantifiable activity in river 2) and even lower estrogenicity in the groundwater samples (with estrogenicity detectable in groundwater 1, albeit below the limit of quantification of 0.2 ng/L). Activity was detected in all spiked samples (Figure 4D, right). The measured increase in activity due to spiking was much higher than predicted (from 129 – 316%, calculated from Table 14), however the predicted value does not include the effect of estriol as no potency was available for that chemical in the KBluc assay. Estriol had a significant effect in other assays (due to its high potency, Table 1), and thus the predicted value is most likely greatly underestimated. Indeed, if one assumes a relative potency of 1 for estriol in the KBluc assay (log RP of 0.00, similar to what it is in the ER-CALUX assay), then the predicted estrogenicity is much closer to the measured value (76.5 ng/L EEq in groundwater, river water and treated sewage, and 191 ng/L EEq in raw sewage).

E-Screen assay:

In the E-Screen assay (Figure 4E), the data from the two assays were significantly different (paired t-test, $p < 0.01$; Table 13) with the data from the second assay generally higher than the first (Figure 5). The two datasets were however well correlated ($R^2 = 0.73$, Table 13). When the data of the two E-Screen assays were combined, they were similar to both the YES and the MELN assay results but different from the ER-CALUX and KBluc assay results (paired t-test, $p = 0.93$ and 0.28, $p < 0.01$ and 0.02; Table 13), but again well correlated with data trends from all assays (average R2 = 0.62; Table 13). There is also good agreement between the E-Screen assay and the other assays in identifying nondetects (pairwise comparison including nondetects, Figure 5).

In the non spiked samples (Figure 4E, left) the raw sewage samples again displayed the highest estrogenic activity, followed by the treated sewage samples. Most groundwater samples displayed no estrogenic activity in the E-Screen assay (except groundwater from site 1 in the second E-Screen assay, which was reported just above the quantification limit at 0.205 ng/L EEq). These data are similar to the ER-CALUX and KBluc assay results, which suggested that groundwater at site 1 contained slightly more estrogenic activity than groundwater at site 2 (albeit both at very low levels). With the river water samples, only samples from site 2 (Oxley Creek) induced detectable estrogenic activity, again comparing favorably with the ER-CALUX and KBluc assay results. Activity was detected in all spiked samples (Figure 4E, right), with the raw sewage samples again displaying the highest activity. The measured increase in activity due to spiking in groundwater and river water was 99.0% of the predicted increase (combined E-Screen data, calculated from Table 14). As with the other assays, this increase in activity was less with the sewage samples (Table 14), possibly due to the presence of anti-estrogenic compounds in sewage or the difficulty of reliably recovering the spike in this complex matrix (Figure 1).

Table 13. P Values of Paired T-Tests, Coefficient of Determination (R^2) and Sample Size (N) on Log-Transformed Data Between the Different Bioassay Datasets for the Environmental Samples (Excluding Nondetects).

Assay	Statistic	YES [a]	ER-CALUX	MELN	KBluc	E-Screen [a]
YES [a]	p	**< 0.01 **** [b]	**< 0.01 ****	**0.02 ****	**< 0.01 ****	0.93
	R^2	**0.17 *** [b]	**0.50 ****	0.02	0.40	**0.50 ****
	n	29 [b]	28	29	9	29
ER-CALUX	p	**< 0.01 ****	--	**< 0.01 ****	**0.03 ***	**< 0.01 ****
	R^2	**0.50 ****	--	**0.42 ****	**0.98 ****	**0.78 ****
	n	28	--	33	10	31
MELN	p	**0.02 ****	**< 0.01 ****	--	**< 0.01 ****	0.28
	R^2	0.02	**0.42 ****	--	0.03	**0.29 ****
	n	29	33	--	11	32
KBluc	p	**< 0.01 ****	**0.03 ***	**< 0.01 ****	--	**< 0.01 ****
	R^2	0.40	**0.98 ****	0.03	--	**0.90 ****
	n	9	10	11	--	11
E-Screen [a]	p	0.93	**< 0.01 ****	0.28	**< 0.01 ****	**< 0.01 **** [b]
	R^2	**0.50 ****	**0.78 ****	**0.29 ****	**0.90 ****	**0.73 **** [b]
	n	29	31	32	11	32 [b]

Notes: Numbers in bold are statistically significant. The * indicates $p < 0.05$, ** indicates $p < 0.01$. Analysis does not include nondetects. The statistical significance of the coefficient of determination (R^2) was calculated from its t-value using a two-tailed t-distribution.
(a) Data from multiple assays combined.
(b) Value is for comparison of the same assay but performed in two different laboratories (eg YES 1 vs YES 2).

Table 14. Measured and Predicted Increase in Estrogenic Activity (Ng/L Eeq) Due to Spiking of Environmental Samples.

Assay	Groundwater (GW)		River water (RW)		Treated sewage (ST)		Raw sewage (SR)	
	Meas	Pred	Meas	Pred	Meas	Pred	Meas	Pred
YES 1	22.3 ± 2.04	87.0	23.0 ± 3.39	87.0	14.3 ± 14.0	87.0	30.5 ± 11.6	218
YES 2	63.3 ± 20.5	157	52.1 ± 19.7	157	47.6 ± 15.9	157	58.2 ± 12.8	391
ER-CALUX	86.1 ± 9.41	96.2	99.7 ± 50.4	96.2	48.7 ± 62.5	96.2	229 ± 61.3	240
MELN	26.0 ± 0.48	64.7	29.0 ± 10.3	64.7	15.8 ± 10.6	64.7	-24.4 ± 22.0	162
KBluc	70.0 ± 37.1	26.5 [a]	83.8 ± 3.55	26.5 [a]	34.3 ± 43.7	26.5 [a]	149 ± 38.5	66.3 [a]
E-Screen 1	23.8 ± 5.94	33.9	34.0 ± 18.5	33.9	13.6 ± 17.1	33.9	34.5 ± 0.07	84.8
E-Screen 2	45.1 ± 10.1	38.2	41.0 ± 6.08	38.2	23.4 ± 28.0	38.2	108 ± 29.6	95.6

Notes: n = 6 except for the KBluc assay where n = 2. Meas = measured; Pred = predicted assuming 100% recovery of the nominal concentrations (Table 3) and based on relative potencies given in Table 1.
(a) The relative potency of estriol for the KBluc assay is currently unavailable (Table 1) and the predicted estrogenicity value is thus knowingly low, as it does not include the estriol contribution.

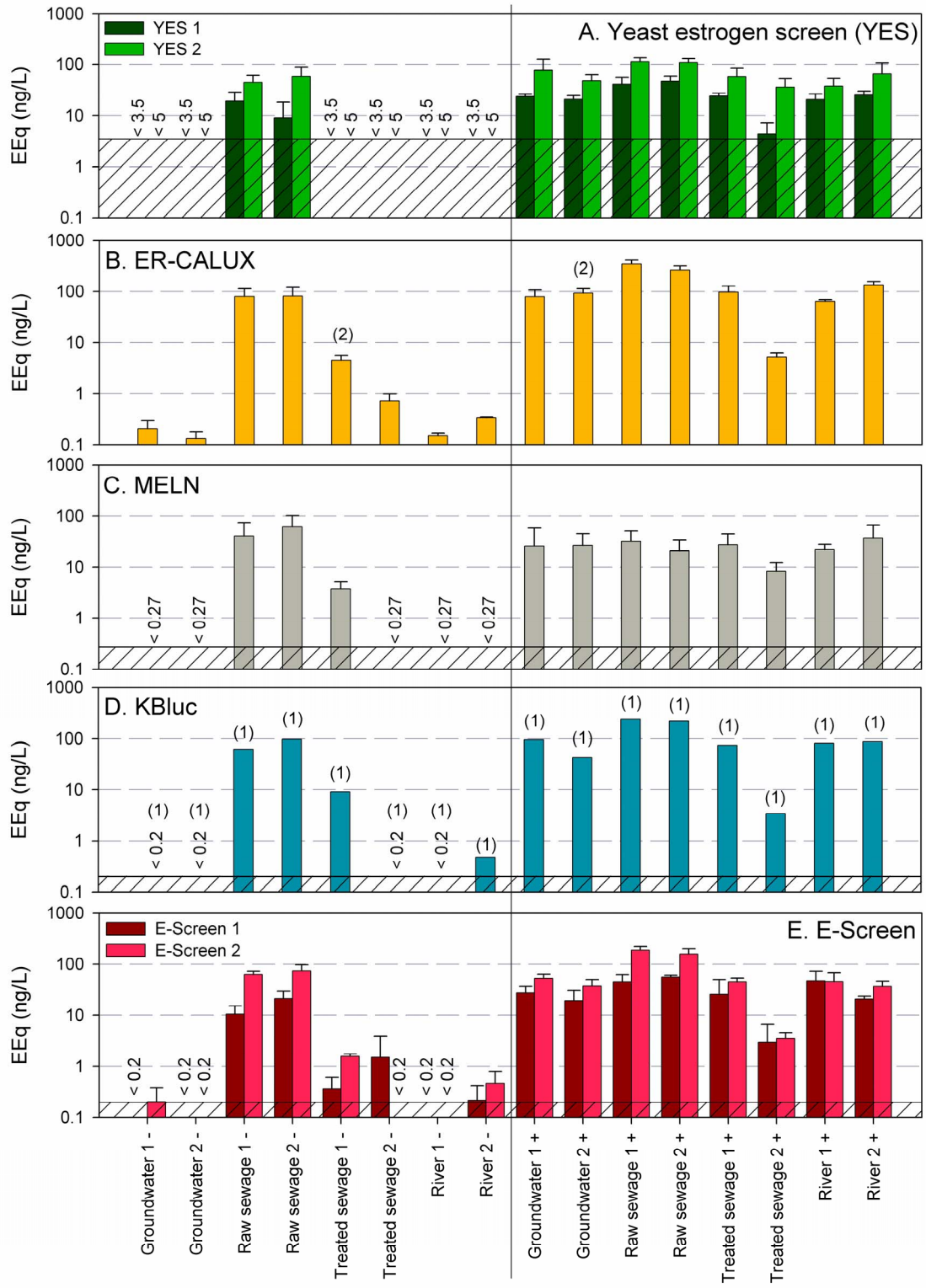


Figure 4. Estrogenicity (Expressed as Estradiol Equivalents, Eeq) in the Environmental Samples as Determined by Bioassay.

Notes: n = 3 independent samples unless otherwise indicated by a number in brackets above the bars. Non spiked samples are on the left, and spiked samples are on the right. The cross-haired area indicates data below the method quantification limit. Error bars represent standard deviation. Data includes nondetects at half the quantification limit.

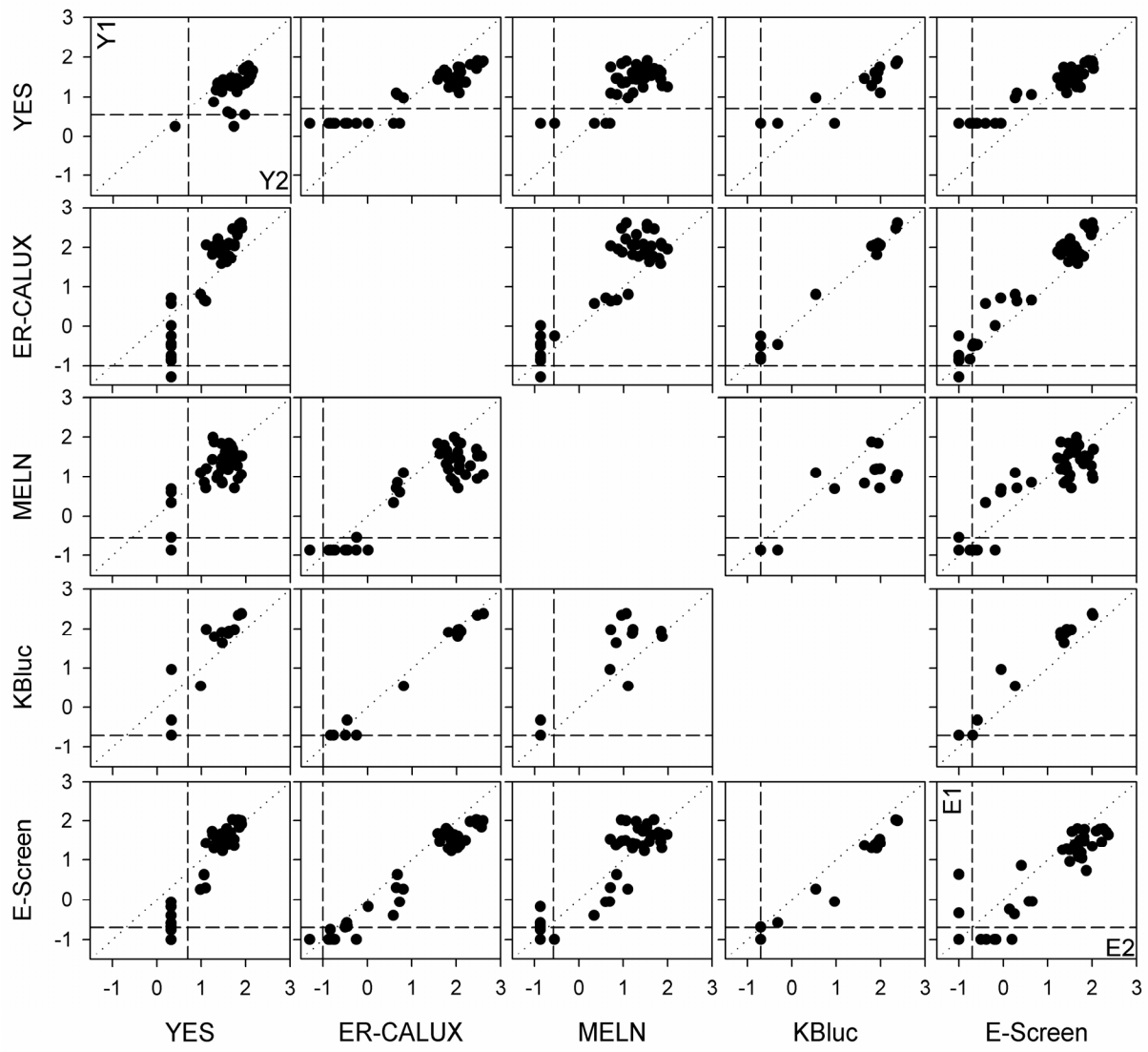


Figure 5. Pairwise Comparisons of Activity (Log Eeq) Between the Different Bioassays with Environmental Samples (Including Nondetects).

Notes: The dotted line is the isometric line. Dashed lines mark the method quantification limit for that assay. Nondetects were assigned a value equal to half the quantification limit. For the YES and E-Screen assays, the dataset was combined for comparisons with other assays but the two assay datasets are compared on their relevant lines (and identified as Y1 and Y2 for the YES datasets 1 and 2, and E1 and E2 for the E-Screen datasets 1 and 2, respectively).

3.2.3 Comparison of Bioassays

To evaluate the performance of the bioassays irrespective of the type of sample, a survival analysis (also known as Kaplan-Meier analysis) was carried out using GraphPad Prism 5 (GraphPad Software Inc., San Diego, CA, USA). This type of analysis allows the statistical

comparison of right-censored datasets, and thus the left-censored dataset from all bioassays in this study was transformed into right-censored data using the following equation:

$$x_t = \frac{1}{\log(x)} + 4$$

where x_t = transformed data
x = original data

The analysis shows that once the censored data is taken into account, most of the bioassays generate remarkably similar results (Table 16 and Figure 6C), with again the exception of the MELN. The survival curve also shows that the ER-CALUX provides data over the largest range (due to its low quantification limit) and that all assays (with the exception of the MELN) exhibit a very comparable response with low to mid-estrogenicity samples (from 0.2 to 20 ng/L).

YES assay:
When all dataset is combined (irrespective of sample source and inclusive of nondetects), the responses in two YES assays were not statistically different (Mantel-Cox test, $p = 0.23$, Table 16). The first YES assay however appears slightly shifted to the left on the survival curve (Figure 6A), suggesting that overall the first YES assay reported slightly lower estrogenicity than the second YES assay. The large proportion of nondetects (46.4 and 45.2% for the first and second YES assays, respectively) is clearly apparent from the survival curve. When combined, the YES assay is not significantly different from the other assays (Mantel-Cox test, $p > 0.09$, Table 16), although much more censored (45.8% nondetects; Table 15).

ER-CALUX assay:
The ER-CALUX assay is significantly different from the MELN assay (Mantel-Cox test, $p = 0.01$,Table 16) but similar to all other assays (Mantel-Cox test, $p > 0.24$ excluding MELN assay, Table 16). The ER-CALUX assay has the lowest censorship with only 9.9% of nondetects (Table 15).

MELN assay:
The MELN assay was significantly different when compared to all except the YES assay (Mantel-Cox test, $p < 0.05$ except for YES assay where $p = 0.09$, Table 16). In fact, the results with that particular assay are skewed to the left of the survival curve (Figure 6C), suggesting that estrogenicity determined by the MELN is consistently lower than that determined by other assays. Of the entire dataset, 35.7% were nondetects (Table 15).

KBluc assay:
Like the ER-CALUX and the E-Screen, the KBluc assay was significantly different from the MELN assay (Mantel-Cox test, $p = 0.02$, Table 16) but similar to all other assays (Mantel-Cox test, $p > 0.33$ excluding MELN assay, Table 16). Only 13.3% of the data were nondetects, although it should again be emphasized that the KBluc dataset is less comprehensive than that obtained with other assays (sample size $n = 30$ for the KBluc vs. $n > 81$ with the all other assays, Table 15).

E-Screen assay:
The responses in the two E-Screen assays were not statistically different (Mantel-Cox test, $p = 0.52$, Table 16). At estrogenicities above 0.5 ng/L however the first E-Screen assay appears

slightly shifted to the left on the survival curve (Figure 6B), suggesting perhaps that induction in the second E-Screen assay was stronger at the higher estrogenicities. When the data are combined, the E-Screen assay was not significantly different from any of the other assays except the MELN assay (Mantel-Cox test, $p > 0.24$ except for MELN assay where $p = 0.04$, Table 16). Of the entire dataset, 26.8% were nondetects (Table 15).

Table 15. Kaplan-Meier Descriptives for Bioassay Data.

Assay	Median EEq (ng/L)	Censored data [(a)] (n_0)	Uncensored data (n_1)	Sample size (n)
YES 1	5.52	39	45	84
YES 2	14.8	38	46	84
Combined YES	11.4	77	91	168
ER-CALUX	5.30	8	73	81
MELN	4.10	30	54	84
KBluc	8.18	4	26	30
E-Screen 1	7.28	22	62	84
E-Screen 2	12.4	23	61	84
Combined E-Screen	8.57	45	123	168

Notes: Nondetects included in the survival analysis as "non-events".
(a) Censored data = number of non-detects.

Table 16. Significance (P-Value) of a Pairwise Comparisons (Log-Rank Mantel-Cox Test) for Differences In Survival Curve (360hFigure 6) Among the Different Assays.

Assay	Combined YES [(a)]	ER-CALUX	MELN	KBluc	Combined E-Screen [(a)]
Combined YES [(a)]	0.23 [(b)]	0.71	0.09	0.83	0.94
ER-CALUX	0.71	--	**0.01** *	0.81	0.24
MELN	0.09	**0.01** *	--	**0.02** *	**0.04** *
KBluc	0.83	0.81	**0.02** *	--	0.33
Combined E-Screen [(a)]	0.94	0.24	**0.04** *	0.33	0.52 [(b)]

Notes: A significant value ($p < 0.05$) indicates that the survival curves are statistically different between the two assays. Nondetects included in the survival analysis as "non-events".
(a) Data from multiple assays combined.
(b) Value is for comparison of the same assay but performed in two different laboratories (eg YES 1 vs YES 2).

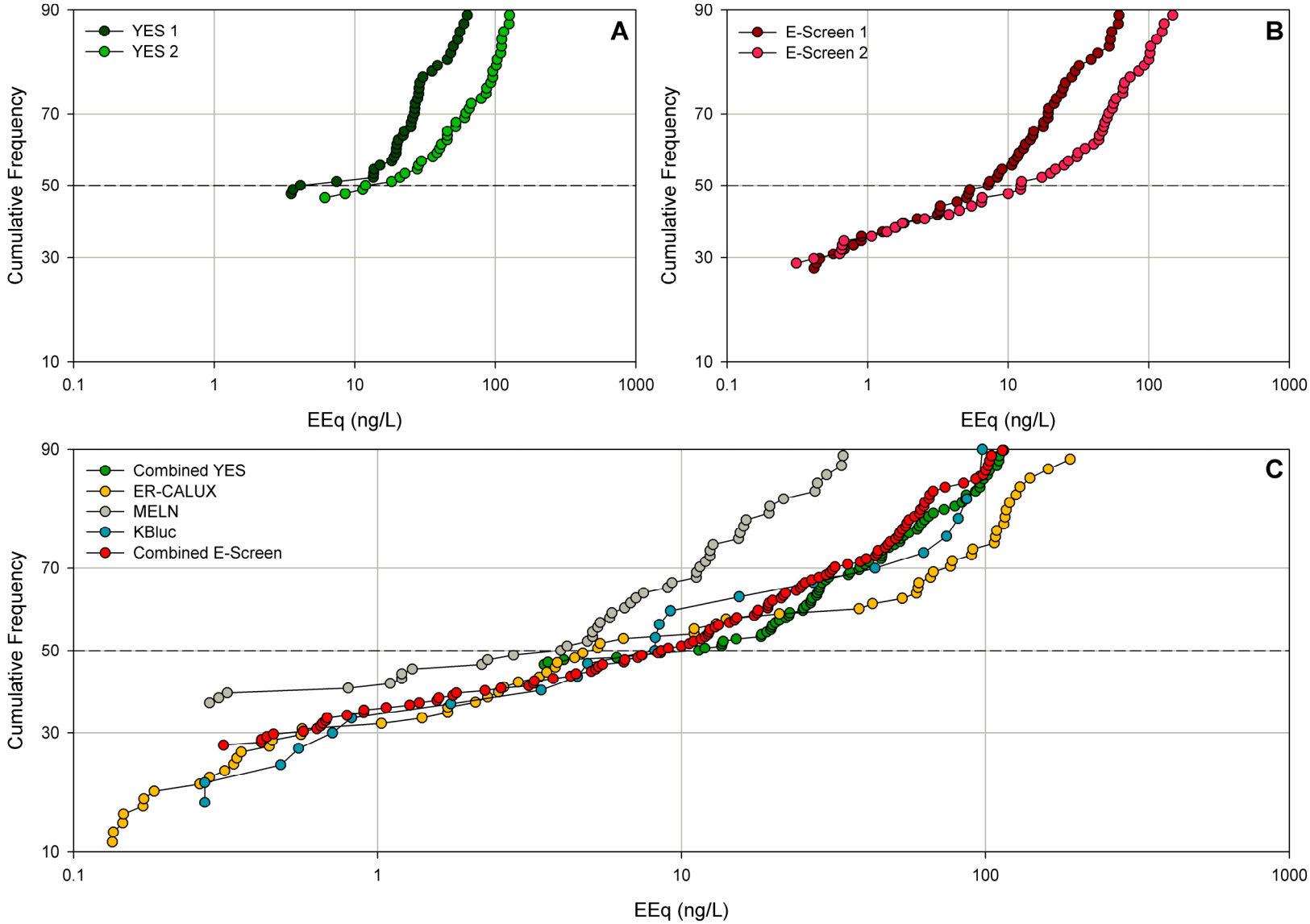


Figure 6. Survival Graph for Bioassay Data. A) YES; B) E-Screen; C) All Assays.

3.3 Comparison Between Chemical and Bioassay Data

By multiplying the concentration of each chemical as determined by standard chemical methods (Table 9 and Table 10) with the relative potency for each individual compound (most of which was calculated from responses with individual compounds in the artificial samples, and reported in Table 1), it is possible to calculate a predicted total estrogenicity for each sample. This is calculated as:

$$Pred = \sum (RP \times conc)$$

where RP = relative potency, and
$conc$ = concentration of each chemical

The ratio of the estrogenicity directly measured in the bioassay (*Meas*) divided by the estrogenicity predicted from chemical measurements alone (*Pred*) then provides an estimate of how well the two sets of data correlate. When the measured or predicted estrogenicity was below the MQL of the assay, the value was set at the MQL. Thus, if both bioassay and chemical methods identify a sample as below the quantification limit (BQL), the ratio of *Meas/Pred* is 1.

If the ratio *Meas/Pred* is less than 1, it indicates that the estrogenicity measured by the bioassay is lower than would have been expected from chemical measurements alone. This could be due to the presence of other (unmeasured) chemicals in the mixture that have an anti-estrogenic effect in the bioassay, complex interaction between the different chemicals in the mixtures (e.g. non-additive effects), or poor bioassay performance due to matrix interference.

On the other hand, a ratio of *Meas/Pred* greater than 1 indicates that the estrogenicity measured in the bioassay is greater than would have been predicted from chemical measurements alone. This could be due to the presence of other chemicals in the mixture that have estrogenic activity but have not been measured by chemical analysis, or complex interactions between the different chemicals in the mixture (e.g. synergistic effects).

Finally if the ratio *Meas/Pred* is close to 1, this indicates that the estrogenicity measured in the bioassay can be accurately predicted from the chemical analysis alone.

3.3.1 Artificial Samples

Overall there was good agreement between the activity measured by the E-Screen and the ER-CALUX assays and that predicted by chemical analysis with the artificial samples; the ratio of Meas/Pred was close to 1 for most samples in these bioassays (Figure 7). The activity in the YES and particularly the MELN assays was lower than would have been predicted by chemical analysis (Figure 7), possibly suggesting media interference or non-additive effects.

YES assay:
In the YES assay (Figure 7A), both assays performed very similarly. In the first YES assay, the Meas/Pred ratio ranged from 0.08 to 1.00 (average ± SD of 0.53 ± 0.44). For the second YES assay, the ratio ranged from 0.15 to 1.00 (0.63 ± 0.39).

ER-CALUX assay:
In the ER-CALUX assay (Figure 7B) the measured estrogenicity was usually slightly higher in the artificial samples than predicted from the chemical measurements (ratio excluding the hormones mix ranging from 1.00 to 11.8), except in the hormone mix where it was slightly lower (0.20). The possible contribution of 4-t-octylphenol could not be integrated into the calculation of predicted estrogenicity (Pred) in the ER-CALUX assay for the industrial and combo mixes due to a lack of potency data for that chemical in that assay (the individual compound was cytotoxic in the artificial samples stage, see Section 33.2.1). Therefore the Pred estrogenicity in those two mixtures could be underestimated, resulting in an artificially high Meas/Pred ratio (Figure 7B).

MELN assay:
In the MELN assay (Figure 7C), the measured estrogenic activity was generally lower than predicted from the chemical analysis (ratio ranging from 0.02 to 2.22, average ± SD of 0.72 ± 0.93).

KBluc assay:
Artificial mixtures were not analyzed in the KBluc assay, and this data is therefore not available for that assay.

E-Screen assay:
Finally, measured estrogenicity was close to predicted by chemical analysis for both E-Screen assays (Figure 7D), slightly lower in the first (ratio ranging from 0.23 to 1.81, average ± SD of 0.75 ± 0.67) and slightly higher in the second (ranging from 0.46 to 3.42, average ± SD of 1.21 ± 1.25).

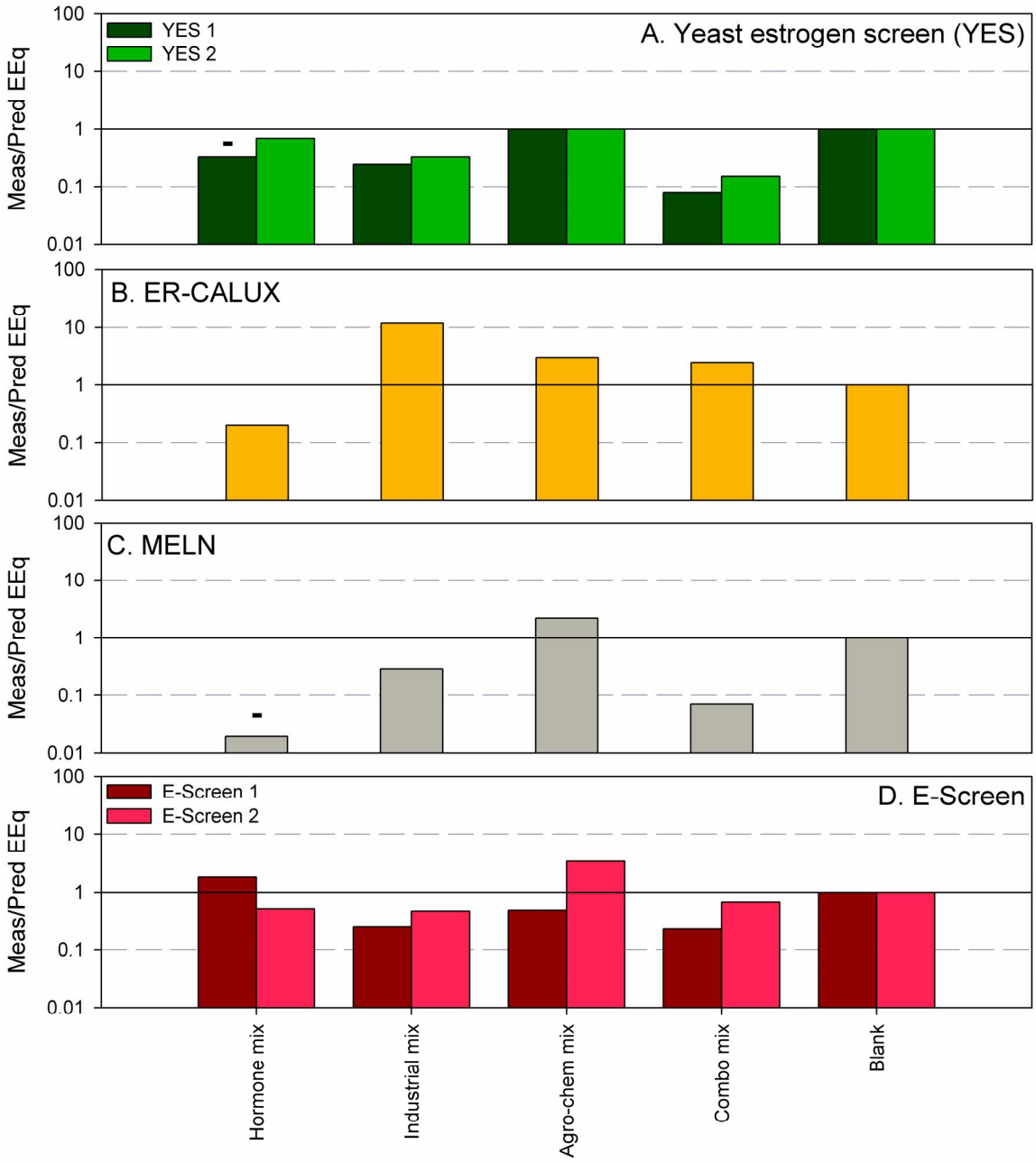


Figure 7. Ratio of Measured Over Predicted Estradiol Equivalents (Eeq) In Artificial Samples.

Notes: Predicted EEq *Pred* was calculated as $Pred = \sum(RP \times conc)$, where *RP* is the relative potency (Table 1) and *conc* is the concentration as determined by analytical chemistry (Table 10). When available, relative potency obtained in that particular assay in this project was used (calculated from samples A01 – A07), otherwise the average of the relative potencies for that assay type published in the literature was used (Table 1). A "+" or "-" above the bar indicates "more than" or "less than", respectively. Note that the KBluc data for artificial mixes is not available, and hence not included on this graph.

3.3.2 Environmental Samples

In critically analyzing the results of the *Meas/Pred* ratio with the environmental samples, it is crucial to understand the contributing factors in spiked and non spiked samples and to analyze them separately. Indeed, because most of the activity in the spiked samples is expected to be due to the 7 spiked chemicals (E1, E2, E3, EE2, BA, NP, BBP), the *Meas/Pred* ratio of the spiked samples is more representative of the ability of the assay to integrate with the chemical analysis. On the other hand, estrogenic activity with the non spiked samples may also be due to chemicals that were not measured by chemical analysis. As such, a ratio higher than 1 in the non spiked samples is not necessarily an indication of abnormal bioassay behavior, but could simply be a measure of estrogenic chemicals that were not detected or not measured (ie., not looked for) by chemical analysis.

Overall there was a good agreement between the measured estrogenicity in the ER-CALUX and E-Screen assays and the predicted estrogenic activity from the chemical measurements, with *Meas/Pred* ratio hovering around 1 (Figure 8). There was a fair agreement between measured and predicted with the YES and the MELN assays, although the *Meas/Pred* ratio in the spiked samples dropped markedly below 1, which might indicate either media interference or non-additive effects in those assays. The *Meas/Pred* ratio in the KBluc assay was artificially high because estriol could not be incorporated into the *Pred* value (as its potency in the KBluc assay is currently unknown).

YES assay:
The data from the two YES assays (Figure 8A) show good agreement between the estrogenic activity measured in the bioassays and predicted from the chemical analysis for most samples *Meas/Pred* ratio was often below 1.

With the non spiked samples, there was a very good agreement between the bioassay data and the chemical data with *Meas/Pred* ratio ranged from 0.19 to 1.00 (average ± SD of 0.75 ± 0.35) in the first assay and from 0.35 to 1.39 (average ± SD of 0.94 ± 0.33) in the second assay. It should be emphasized however that this positive result is biased by to the high MQL of the YES assays in this study (3.5 and 5 ng/L). This results in an artificially-high proportion of *Meas/Pred* ratio of 1, as the chemistry confirms that bioassay measurement should indeed be below the quantification limit of the assay (this is the case for both groundwater samples as well as the treated sewage 2 and river 1 samples). When these data are excluded, the average ratio (± SD) becomes 0.50 ± 0.34 and 0.89 ± 0.49 for the first and second YES assays, respectively.

With the spiked samples, the *Meas/Pred* ratio was much lower and ranged from 0.23 to 0.54 (average ± SD of 0.33 ± 0.10) in the first assay and from 0.33 to 1.52 (average ± SD of 0.74 ± 0.39) in the second assay. Combined with similar conclusions for the artificial samples, this suggests that the YES assay may not fully integrate the effect of estrogenic chemicals at high concentrations (as is the case in the spiked samples).

ER-CALUX assay:
In the ER-CALUX assay (Figure 8B), there was good agreement between *Meas* and *Pred*, with slightly higher estrogenic activity measured in the bioassay than predicted from chemical analysis with the sewage samples.

With the non spiked samples, the ratio ranged from 0.34 to 7.40 (average ± SD of 2.06 ± 2.30), with the highest values in the treated sewage samples. This suggests the presence of estrogenic compounds not detected by chemical analysis, particularly in treated sewage. However the values for the raw sewage samples are lower than expected, which suggests possible matrix interference with raw sewage in that assay.

In spiked samples, the ratio ranged from 0.52 to 1.78 (average ± SD of 1.11 ± 0.39). This indicates that in the spiked samples, the measured activity in the bioassay is very close to the predicted activity – not an unexpected results as most of the activity in the spiked samples should be due known (and measured) chemicals. This suggests that ratio higher than 1 in the non spiked samples indeed identify the presence of estrogenic compounds not detected by chemical analysis.

MELN assay:
In the MELN assay (Figure 8C), the measured activity was generally lower than predicted from the chemical analysis (similar to what was observed in the artificial samples). In fact, the MELN results are very similar to those reported for the YES assay (Figure 8A).

With non spiked chemicals, there was a good agreement between the *Meas/Pred* ratio, which ranged from 0.12 to 1.00 (average ± SD of 0.80 ± 0.32). However there is a significant bias introduced by the large proportion of non-detect samples, where there is good agreement between the bioassay and chemical analysis that these should be below quantification and thus a *Meas/Pred* ratio of 1 (both groundwater samples as well as the treated sewage 2 and river 1 samples). When these data are excluded, the average ratio (± SD) becomes 0.61 ± 0.37.

With the spiked samples, the *Meas/Pred* ratio was much lower and ranged from 0.16 to 0.76 (average ± SD of 0.40 ± 0.19). As was hypothesized for the YES assay results (Figure 8A, right), this may suggest that the MELN assay does not fully integrate the effect of estrogenic chemicals at high concentrations.

KBluc assay:
For the KBluc assay (Figure 8D), the ratio of *Meas/Pred* estrogenicity is usually higher than 1. Rather than being indicative of the undetected estrogenic chemicals this is most likely due to the lack of integration of the effect of estriol in the predicted value (*Pred*), which was not included because its potency in the KBluc assay is unknown (Table 1). It is nevertheless possible to cautiously evaluate this data to obtain a rough estimate of the agreement between chemical and KBluc data, keeping in mind that a ratio above 1 can often be explained by estriol.

With the non spiked samples, the *Meas/Pred* ratio ranged from 1.00 to 12.0 (average ± SD of 3.13 ± 3.78). This is somewhat biased by the good agreement in non-detect samples (groundwater 1 and 2, treated sewage 2 and river 1 samples). When these are excluded from the dataset, the average ratio (± SD) becomes 5.26 ± 4.62. Most of the "missing" predicted activity is likely to be due to estriol.

With the spiked samples, the ratio ranged from 0.41 to 5.94 (average ± SD of 3.52 ± 2.13). It is reassuring to find that the ratio is higher than 1 in 6 out of 8 samples, as would be expected considering that estriol was added in the spike and the predicted estrogenicity value (*Pred*)

excludes estriol (and is thus knowingly low). Assuming a relative potency for estriol of 1 (or log RP 0.00) as is the case in the ER-CALUX assay (Table 1), the average *Meas/Pred* ratio (± SD) becomes 1.06 ± 0.51, suggesting that the relative potency of estriol in the KBluc assay is likely to be similar to that in the ER-CALUX.

E-Screen assay:
Finally the E-Screen assay (Figure 8E) showed a good agreement between measured and predicted estrogenicity.

In the non spiked samples, the ratio of *Meas/Pred* ranged from 0.13 to 2.98 (average ± SD of 0.78 ± 0.93) in the first and from 0.38 to 1.89 (average ± SD of 1.00 ± 0.58) in the second E-Screen assay. Measured estrogenicity in the raw sewage samples was higher than predicted in the second E-Screen assay (similar to the ER-CALUX assay), but not in the first E-Screen assay.

With the spiked samples, the ratio ranged from 0.33 to 2.11 (average ± SD of 0.87 ± 0.53) in the first and from 0.37 to 3.22 (average ± SD of 1.60 ± 0.86) in the second E-Screen assay. This good agreement is not unexpected,as most of the activity in the spiked samples should be due to the identified (and measured) spiked chemicals.

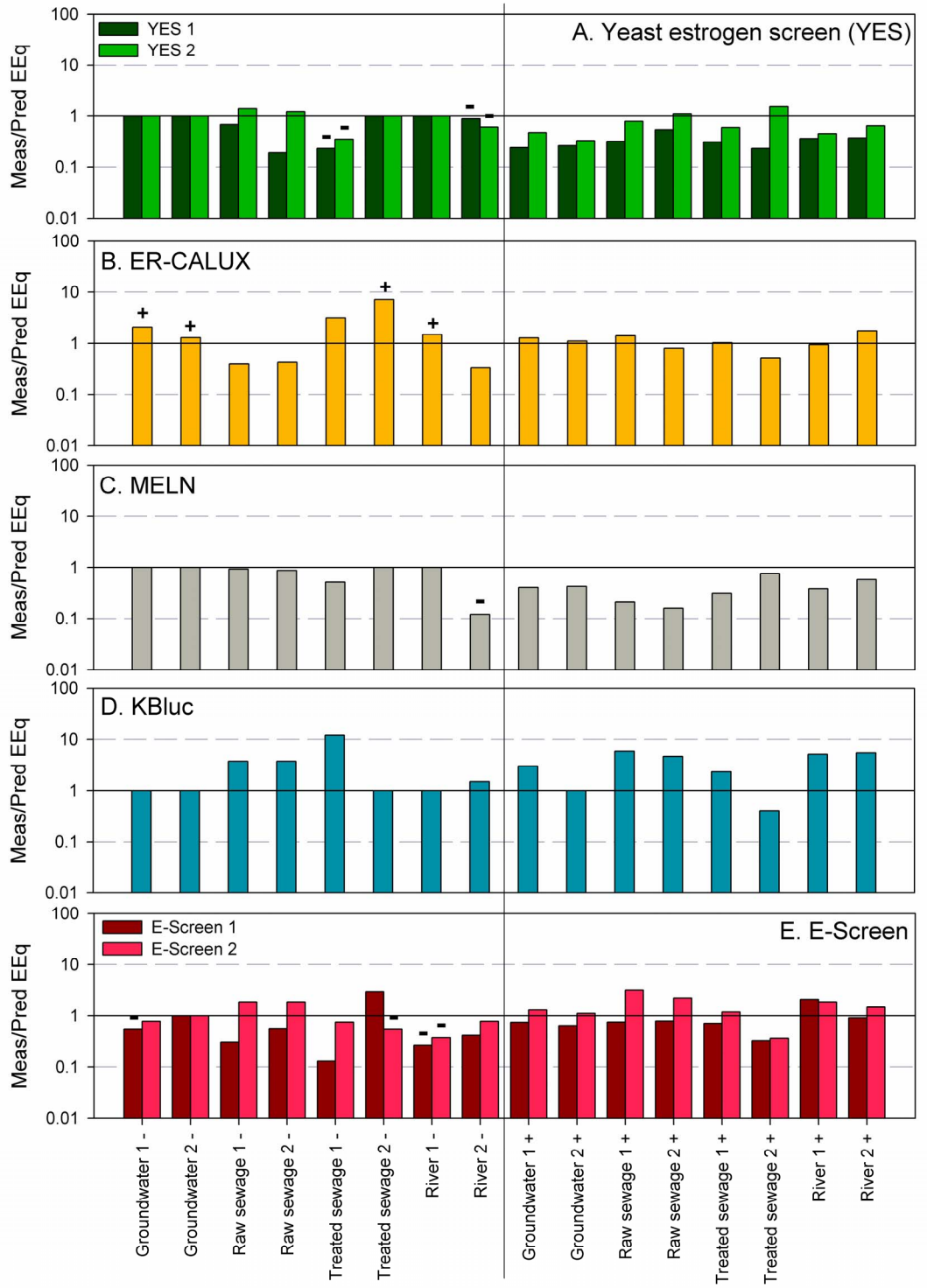


Figure 8. Ratio of Measured Over Predicted Estradiol Equivalents (Eeq) in Environmental Samples.

Notes: Predicted EEq *Pred* was calculated as $Pred = \sum(RP \times conc)$, where *RP* is the relative potency (Table 1) and *conc* is the concentration as determined by analytical chemistry (Table 10). When available, relative potency obtained in that particular assay in this project was used (calculated from samples A01 – A07), otherwise the average of the relative potencies for that assay type published in the literature was used (Table 1). A "+" or "–" above the bar indicates "more than" or "less than", respectively. Non spiked samples are on the left, and spiked samples are on the right.

3.4 Estrogenicity of the Environmental Samples

The combined bioassay and chemical data can provide a powerful measure and interpretation of estrogenic pollution in environmental samples. The following section combined the two datasets to describe the estrogenic pollution in the eight environmental samples tested in this study.

Groundwater:
Both groundwater sampled displayed very little estrogenic activity, undetectable in most assays except the ER-CALUX and the E-Screen assays (Figure 4). The groundwater sampled from the shallow aquifer (GW1) was slightly more estrogenic (0.211 ng/L EEq in the ER-CALUX assay, 0.205 ng/L in the second E-Screen assay; Figure 4) than that sampled from the deeper aquifer (GW2; 0.133 ng/L in the ER-CALUX, < 0.2 ng/L in the E-Screen assay; Figure 4). Chemical analysis confirms that the shallower water may contain more contaminants, with low concentrations of octylphenol and nonylphenol detected in the shallow groundwater but not in the deeper water (Table 9).

Sewage:
Both raw sewage samples (SR1 and SR2) were highly estrogenic in all bioassays (Figure 4). Almost all organic contaminants measured by chemical analysis were present in raw sewage (Table 10). Treatment greatly reduced the estrogenic activity at both treatment plants, although the treated sewage from the first treatment plant generally displayed higher estrogenic activity than that from the second plant (Figure 4). This trend is confirmed by chemical analysis, with all except one organic pollutants (estrone) measured lower in the treated sewage compared to the raw sewage, and lower in treated sewage from plant 2 vs. plant 1 (Table 9 and and Table 10). This suggests that the removal efficacy for estrogenic pollutants was better at the second treatment plant compared to the first plant. In fact, the average removal efficacy for estrogenic activity across all bioassays was 89.7% at STP 1 vs. 95.4% at STP2 (calculated from Table 19 using ½ the quantification limit for nondetects). The first STP is much larger and treats approximately 58,000 m^3/d vs. 26,000 at the second STP, and also receives a higher proportion of industrial waste—a fact that is anecdotally confirmed by chemical analysis with industrial estrogen mimics such as OP and NP are present in higher concentrations in raw sewage 1 and natural hormones in higher concentrations in raw sewage 2. Both these factors may play a role in the differences in removal efficacies between the two plants.

River water:
The river water samples also displayed limited estrogenic activity, detectable only in the most sensitive bioassays (ER-CALUX, KBluc and E-Screen assays; Figure 4). The water sampled from Oxleigh Creek (RW2) was slightly more estrogenic (0.345 ng/L EEq in the ER-CALUX assay, 0.48 ng/L in the KBluc assay, and an average of 0.343 ng/L in the E-Screen assays; Figure 4) than that sampled from the Brisbane River (RW1; 0.150 ng/L EEq in the ER-CALUX assay, and < 0.2 ng/L in the KBluc and both E-Screen assays; Figure 4). Chemical analysis in this case is more ambiguous, with higher concentrations of nonylphenol but lower concentrations of octylphenol in RW2 vs. RW1 (Table 9). There appears however to be a small amount of estrone in RW2, detectable in some (but not all) of the samples by chemical analysis (Table 10), which may explain the higher estrogenic activity in RW2 detected in most assays.

3.5 Further Data Mining (Variability, MQL and Analysis)

3.5.1 Data Variability

To determine if bioassay-derived data were more variable than that derived by chemical analysis, the average coefficient of variation for all environmental samples were plotted for individual biooassays and compared with the average coefficient of variation from the chemical analysis (Figure 9). There was no significant difference between the coefficient of variation of any of the bioassays with the chemical measurement (one-way ANOVA, $p > 0.05$). This suggests that bioassay-derived data are not more variable than chemical analysis when dealing with low levels of estrogenic chemicals.

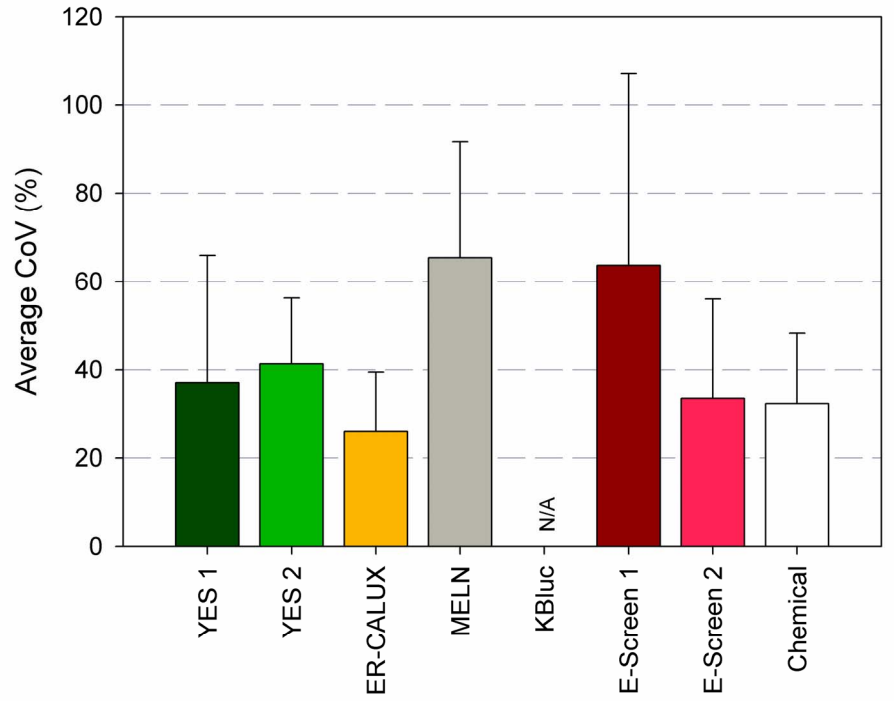


Figure 9. Average Coefficient of Variation (Cov) ± SD for All Environmental Samples.

Notes: N/A = not available: only 1 independent sample was analysed in the KBluc assay (albeit in triplicate) and it is therefore not possible to calculate and average CoV for that assay.

3.5.2 Method Quantification Limits

Most of the bioassays had much lower quantification limits than the quantification limit reported for estradiol using chemical analysis methods, with the exception of the YES assays (Figure 10). The ER-CALUX assay has the lowest method quantification limit of all assays tested (0.1 ng/L EEq, corresponding to a quantification limit on a full sample of 8.0 pg/L), and the other mammalian assays were all in a very similar range from 0.2 to 0.27 ng/L EEq (16 and 22 pg/L on full sample, respectively). The YES assay had the highest method quantification limit, in the same range as chemical analysis (3.5 and 5.0 ng/L EEq for the YES assays, 5 ng/L 17β-estradiol for the chemical analysis; corresponding to 280 and 400 pg/L on a full sample,

respectively). Yeast assays are known to be less sensitive than mammalian assays, with EC_{50} for estradiol in the YES assay around 200 – 700 pM compared with 1 – 10 pM in the mammalian bioassays (Table 1). This is likely due to the thickness of the yeast cell wall and the presence of only the ERα isomer of the estrogen receptor (stably transfected into the yeast cells), whereas mammalian cells used in the assays in this project endogenously express both isomers of the estrogen receptor.

It should be highlighted that the method quantification limits in this study (Figure 2 and Figure 4) are relatively high due to the large number of analyses and participating laboratories (each of which received 1 aliquot of every sample for each assay to be performed). Indeed each sample was split in 12 aliquots, thus decreasing the method quantification limit by a factor of 12 (as indeed each laboratory only received approximately $1/12^{th}$ of each sample, or the equivalent of 200 mL from the original 2.5 L). On whole (2.5 L) samples, the method quantification limits would be much lower for all assays (including chemical analysis) (Figure 10), in agreement with literature values.

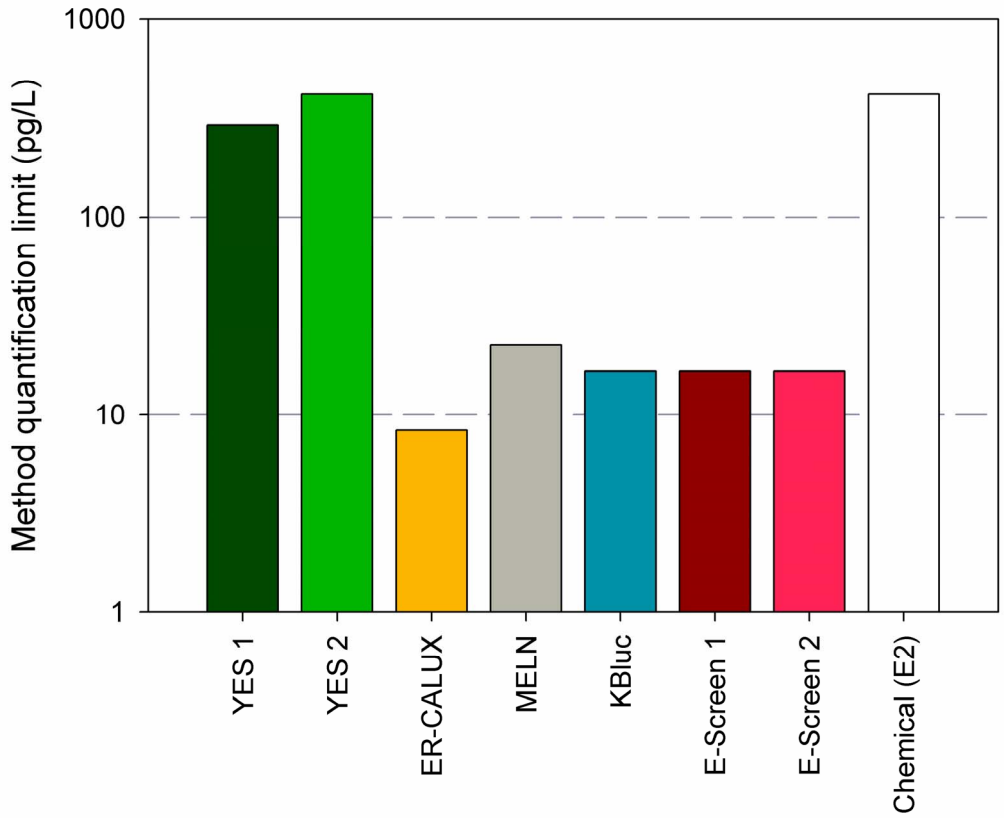


Figure 10. Method Quantification Limits on Whole (2.5 L) Samples.
Notes: the reported MQL for chemical analysis is for 17β-estradiol.

3.5.3 Notes on Data Analysis

Estimating estradiol equivalent concentration of environmental samples:

One essential consideration that was highlighted during this study is the importance of method standardization and particularly for data analysis. While there is currently no widely

accepted data analysis technique that fits all the bioassays used in this study, the data output of these bioassays is very similar: plotting the response in the assay (whether β-galactosidase activity in the YES assay, or luminescence in the ER-CALUX or the KBluc assays, or the number of viable cells in the E-Screen assay) vs the amount (or concentration, or dilution) of sample incubated yields a sigmoid curve similar to the one drawn in Figure 11. Most estrogenic compounds are cytotoxic at high concentrations, and the response curve generally drops sharply when high amounts of sample are incubated (Figure 11, right). In transcriptional activation assays, this drop at high ligand concentration is also sometimes attributed to "quenching" of the limited number of available cofactors. But with lower amounts, a typical dose-response sigmoid curve is usually obtained (Figure 11, left). This is one of the reasons why it is very important to run a dose response rather than only relying on the most concentrated sample.

A sigmoid equation is usually fitted to the data in the dose-response range (the data in the cytotoxicity range is dropped from the analysis), either using specific curve-fitting software such as Prism (Graphpad, San Diego, CA, USA) or using non-linear regression tools in statistical programs such as Solver in Excel (Microsoft Corp, Redmond, WA, USA) or Sigmaplot (SPSS Inc, Chicago, IL, USA). In most cases, the main purpose of this regression exercise is to extract the x-value where 50% of the effect is measured. This point, called the EC_{50}, is the most reliable point to extract a single value from a sigmoid curve because the confidence intervals are narrowest at that point and, as the inflexion point of the curve, it is not affected by its slope.

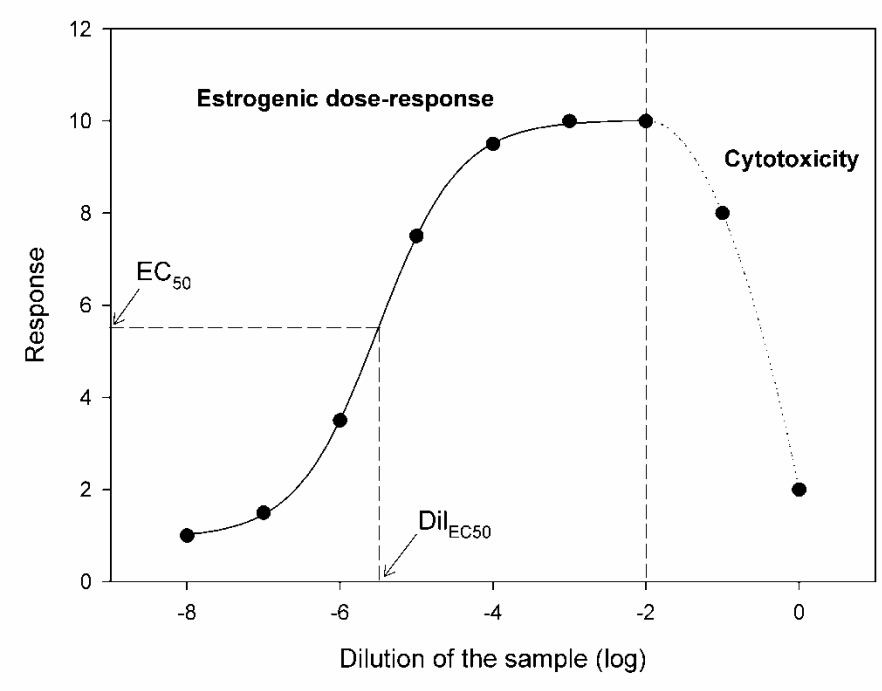


Figure 11. Typical Response in a Bioassay.

Most bioassay analyses vary in how they relate the EC_{50} of the sample to an environmental concentration, in estradiol equivalents (EEq); this is dependent on which units have been used on the x-axis (dilution, equivalent volume, or equivalent concentration directly). The paragraph below suggests a method using dilution that may be used in any circumstances.

First, the terms used have to be exactly defined:
CF_X = concentration factor of the extraction process (eg solid-phase extraction)
e.g.: 2.5L were extracted by SPE and reconstituted in 500µL
$CF_X = 2.5 / 0.0005 = 5000$
Dil = dilution used in the assay
e.g.: 10µL of the reconstituted sample were added to 190µL of buffer in the first well
Dil = 10 / (190+10) = 0.05
Dil_{EC50} = dilution in the assay at EC50, estimated from non-linear regression of response *vs.* Dil
e.g.: from the response curve above (Figure 11), the Dil_{EC50} would be approx $10^{-5.5}$.
EEq = Estradiol equivalent concentration

Estradiol equivalency can then be estimated from the dilution using the following equation:

$$\text{EEq (ng/L)} = \frac{E2_{EC50}}{Dil_{EC50} \times CF_X}$$

where $E2_{EC50}$ = concentration of estradiol at EC_{50} from the standard curve (in ng/L)
Dil_{EC50} = dilution at EC_{50} in the sample curve (unitless)
CF_X = concentration factor of the extraction process (unitless)

Example:
Sample extracted: 200 mL
Reconstitution volume: 40 µL
The concentration of estradiol at EC_{50} in a standard curve: 0.8 ng/L
Dilution required to reach EC_{50} in the sample: 0.0005 (ie 2000× dilution factor)

Then $CF_X = \frac{0.2}{0.00004} = 5000$, and $\text{EEq} = \frac{0.8}{0.0005 \times 5000} = 0.32$ ng/L.

Limitations of this approach and alternatives:

The most significant limitation of this approach is that calculating the EC_{50} requires knowledge of the EC_{100} (the maximal response). This is not always known. In some cases, cytotoxicity interferes with the dose-response curve and the maximal activity (EC_{100}) is not reached. In other rare cases, chemicals or samples may also induce a "supra-maximal" response, where more than 100% of the activity is induced. One way around this problem is to estimate the EC_{100} from the estradiol standard curve when it cannot be estimated from the sample response curve.

An alternative method, sometimes referred to as "first-response" calculation, is based on using the first concentration that induces a statistically significant response above the baseline in the bioassay and relating that response to an equivalent response in the estradiol standard curve. There are however significant limitations of this method, 1) this approach is that it is based on a LOEL (lowest observable effects level), which is determined by the experimental conditions

(which doses were used) rather than an actual activity of the sample; and 2) it may result in false positives because it does not take into account the amplitude of the induced effect.

Another alternative attempts to combine the advantages of the two methods described above by calculating the EC_{10} (concentration that yield 10% of the effect) and use that for comparison with the EC_{10} of the standard. This method however still requires an estimate of the EC100, and thus suffers from the same limitations as the EC_{50} method.

As became obvious during the project, each analysis method can result in very significant differences in the reported activity from the bioassays. While the EC_{50} method has its problems, it provides a standardized method that is applicable to analyze most bioassay results.

Special recommendations for the YES assay:

In the YES assay, two endpoints are measured during the assay: turbidity (a measure of cytotoxicity) at an absorbance of 620-670nm and galactosidase activity (an indirect measure of estrogenicity) at 540-570nm. There are several methods recommended to correct galactosidase activity by turbidity, and they can result in widely different results especially when turbidity varies due to cytotoxicity. The most widely-used method for adjusting galactosidase activity uses the following equation:

$$\text{Corrected Abs540}_{TEST} = \text{Abs540}_{TEST} - (\text{Abs620}_{TEST} - \text{median (Abs620}_{CONTROL}))$$

Many other methods are in use however [28, 39], and as a result the corrected values can vary quite significantly depending on which method is used. There is therefore an obvious need for a standardization of this correction step. Although the above formula at first appears mathematically sound, it attempts to subtract turbidity (cell number) from galactosidase activity, which is not biologically-relevant. An approach that recognizes that two endpoints are being measured (cytotoxicity and estrogenicity) and analyses them separately would be more meaningful. The following approach was used to analyze YES assay results in this study:

- Step 1: First analyze cytotoxicity. A concentration was considered cytotoxic when its Abs620 reading was below the average – 3 standard deviations of the control Abs620. Estrogenicity data for any concentration identified as cytotoxic were then ignored in the following steps.

- Step 2: Calculate the induction ratio (IR) at each concentration tested by correcting galactosidase activity by turbidity (only for concentrations that were not cytotoxic) using the following equation:

- $IR_{TEST} = (\text{Abs540}_{TEST} / \text{Abs620}_{TEST}) / [\text{ mean (Abs540}_{CONTROL} / \text{Abs620}_{CONTROL})]$

- Step 3: Plot IR *vs.* Dil for all concentrations tested (excluding any concentrations that were determined as cytotoxic in step 1) and determine the EEq as described in the previous section.

3.5.4 Notes on Data Censoring and Statistical Analyses

Statistical analysis of a dataset containing a high number of nondetects (also known as censored dataset) is a recognized difficult issue in environmental statistics [40]. Excluding the nondetects, as was done to calculate p-values of the paired t-tests and the Pearson product-quotient correlation coefficient (R^2), produces a strong upward bias and means that comparisons are being made between the high concentrations only. This is not an optimal situation. However, it was preferred over using the standard technique of substituting an arbitrary value of half the quantification limit for nondetects, which would have resulted in artificial conclusions of correlation between the dataset. Survival analysis (or Kaplan-Meier analysis), a statistical technique for estimating censored data in the industrial and medical fields, was used in this report to compare the dataset generated from the different bioassays, irrespective of the source of the sample (artificial or enrivonmental). While this technique can be very powerful, it also comes with significant limitations that need to be recognized. Firstly, these techniques are generally more adequate for handling right-censored data instead of left-censored data. However, this problem can and was overcome by using mathematical tricks to transform left-censored into right-censored data. More importantly however, while these techniques allow the fabrication of data based on maximum likelihood estimation which is superior to simple substitution [40] and have been used to correlate datasets where one of the sets was left-censored [41], they cannot be applied to test for paired comparisons when both datasets are censored (as is the case with all datasets in this report). We therefore opted to perform paired comparisons on the censored datasets excluding nondetects but to discuss the data trends including nondetects, in particular whether there is good agreement in nondetects between the datasets (hence the usefulness of Figure 3 and Figure 5). Finally, when discussing the performance of the assay irrespective of the sample tested, a survival analysis was used as the loss of dataset pairing then becomes less significant (Figure 6).

3.5.5 Notes on Recovery Efficiency

Recoveries in this study were relatively low (ranging from 12 to 100%), but this was at least in some part attributable to the early spiking. Indeed, the spike compounds were added immediately after sample collection before any pre-treatment. As most of the spiked compounds have a log K_{OW} between 3 and 6, it is likely that most would preferably associate with particles and sludge in the water sample, thus resulting in poor recovery. Nevertheless early spiking was preferred in this study as it was more relevant to include all possible matrix effects and properly determine the ability of the methods to measure estrogenicity in the water phase (and thus exclude compounds bound to particulates and sludge).

This however raises the question of how bioassay results should be reported when recoveries are known. For example, if a bioassay determines that a sample has an EEq of 50 ng/L but we know that the extraction efficiency was 25%, should the scientist report a putative value of 4× that, i.e., 200 ng/L? Although simple in principle, the issue is complicated by the fact that although bioassay results are expressed as estradiol equivalents, the measured activity is in fact due to all estrogenic compounds in the sample, which may very have different extraction efficiencies than that of estradiol. We therefore chose not to correct EEq by the recovery efficiency (a common trend in the literature), but it should be recognized that the estrogenicity of the original (un-extracted) sample may thus be higher than reported – although as stated before it is likely that most of the "lost" compounds are in fact associated with sludge and other particulates, and that the EEq is an accurate measure of estrogenicity in the water phase.

4.0 Conclusions and Future Directions

Overall the ER-CALUX and E-Screen assays were robust and reliable, with variability similar to chemical analysis and significantly lower quantification limits suggesting that they would be well-suited as preliminary screening tools for environmental monitoring. The KBluc assay likewise appears well-suited to that purpose, although conclusions for this assays are based on a limited dataset. The relatively high detection and quantification limits of the YES assay and the apparently diminished response of the MELN assay at high concentrations suggest that these assays may not be well-suited to this purpose.

The results of this study show that bioassay techniques are now sufficiently advanced that they can be used either as a cost-effective first-pass detection system or in combination with standard analytical methods to measure estrogenic pollutants in environmental waters. Each assay has its advantages and limitations, and the notion of "fit-for-purpose" is critical in determining what bioassay to use in a particular project. Standardization of bioassay data analysis was identified as a crucial step forward towards accurate bioassay-derived estrogenicity measurements.

4.1 Summary

Overall, all the bioassays assessed in this study performed well. The ability of the bioassays to detect estrogenic chemicals based on their effect and not their chemistry in a high-throughput form means that bioassays can play a major role as an initial screening tool, to screen large numbers of environmental samples for estrogenic activity (first-pass screening). Combined with chemical data this provides a comprehensive assessment of the water samples enabling identification of hazards and identification and assessment of critical control points (HACCP) as adopted by the World Health Organization and Australian Drinking Water Framework [42, 43].

While there were differences in the amplitude of the response, the data trends were very similar between the different assays (Section 3.2). Yet there were subtle differences in estrogenicity as reported by the bioassays that could be due to the mechanism of action of the estrogenic chemicals. For example, reporter gene assays would detect estrogenicity only through receptor-mediated genomic events, while the E-Screen assay would also detect non-genomic estrogenic effects (e.g., non-genomic effects via cell surface receptors). Other differences could be due to receptor isomer-specific responses. For example, the YES assay contains only the α isomer of the estrogen receptor, while the other assays in this project contain both α and β isomers. In the end, each assay has advantages and limitations: the YES assay is relatively quick and comparatively inexpensive, but this comes at the expense of a lesser sensitivity; the ER-CALUX assay is very sensitive, precise and relatively quick, but requires advanced laboratories and highly qualified personnel; the MELN assay is sensitive and relatively quick, but also requires advanced laboratories and highly qualified personnel and may be sensitive to interference from complex matrices; the KBluc assay appears similar to the ER-CALUX assay albeit slightly less sensitive, but it should be acknowledged that fewer samples were analyzed in the KBluc assay in this project and more research is needed to confirm these conclusions; while the E-Screen assay is sensitive but takes more time to perform, appears slightly more variable, and also requires advanced laboratories and highly qualified personnel.

A simple performance evaluation of the different assays used in this study (Table 17) shows that the ER-CALUX assay performed best and that the YES and the E-Screen assays both performed well. Because of the poor correlation between measured and predicted estrogenicity and its high coefficient of variation, the MELN assay was ranked the lowest. This is not to say that the MELN assay is inferior, but rather that it is not yet a widely used assay, especially compared with the ER-CALUX, the YES and the E-Screen assays, and its performance may increase as it matures. For the KBluc assay, only a limited set of data is currently available, so it is not possible to reliably compare it with the other assays. However the current data suggests that this assay will most likely favorably compare with the other assays used in this study and may prove to be a useful tool in the future.

Table 17. Overall Performance of the Tested Bioassays in this Study.

Assay	Lab EEq [(a)]	Field EEq [(b)]	Likeness to other assays [(c)]	CoV [(d)]	MQL [(e)]	Lab *Meas/Pred* [(f)]	Field *Meas/Pred* [(g)]	Overall
YES	++	+	++	+++	-	++	+	+½
ER-CALUX	++	+++	+++	+++	+++	(++)[(h)]	+++	++¾
MELN	++	+	-	+	+++	-	+	+¼
KBluc	(++)[(i)]	(++)[(i)]	+++	N/A	+++	N/A	(++)[(i)]	(++½)
E-Screen	++	++	+++	++	+++	++	++	++¼

Notes: "-" = poor performance, "+" = fair performance, "++" = good performance, "+++" = excellent performance. N/A = data not available.
(a) Artificial samples estrogenic activity, data in Figure 2.
(b) Environmental samples estrogenic activity, data in Figure 4.
(c) Comparison with other assays irrespective of sample source, data in Figure 6.
(d) Coefficient of variation, data in Figure 9.
(e) Method quantification limit, data in Figure 10.
(f) Ratio of measured over predicted estrogenicity for artificial samples, data in Figure 7.
(g) Ratio of measured over predicted estrogenicity for environmental samples, data in Figure 8.
(h) Octylphenol could not be included in the *Pred* value because it was cytotoxic on its own.
(i) Value is in brackets to indicate a lower precision due to less samples being analyzed in this assay.

In most cases, the estrogenic activity in all bioassays could be related to the chemical concentrations determined by chemical analysis and their relative potencies (Section 3.3). The generally-good agreement between bioassay-derived estrogenic activity and predicted activity from the chemical analysis in the non spiked environmental samples suggests that the most potent estrogenic pollutants in the Australian environment have been identified and were measured by chemical analysis (17β-estradiol, estrone, estriol, ethinylestradiol, nonylphenol, bisphenol A, 4-*t*-octylphenol, and benzyl butyl phthalate). Whether this is true for other environmental water matrices and/or other countries remains to be shown. For example, p,p'-DDT may be present a much higher concentrations in South Africa than in Australia, and despite its low relative estrogenic potency it may significantly contribute to estrogenicity in environmental waters there.

4.2 Conclusions

In conclusion, the ER-CALUX and E-Screen assays tested in this study successfully detected estrogenicity in environmental water samples, even at very low levels. The estrogenic activity measured in these bioassays could be related to predicted estrogenic activity based on chemical analysis, suggesting that any of these two bioassays could be used as initial screening tools to detect estrogenicity in environmental water samples. The KBluc assay also appeared to perform well, and in fact very similar to the ER-CALUX, but the dataset obtained in this study is more limited for that assay than for the others, and any conclusion based on the KBluc assay must therefore be appraised critically. The YES assay performed well with highly polluted samples but its relatively high quantification limit meant that it was unable to quantify low-level estrogenicity in less polluted samples such as groundwater, river water and even treated sewage. With artificial samples, the performance of the YES assay was also significantly affected by octylphenol. The MELN assay tested in this study also performed well, correctly identifying low and high estrogenic activity. However, possible susceptibility to matrix interference from complex matrices (such as sewage) needs to be further investigated before results from the MELN assay can be correlated accurately with chemical analysis. Table 18 below provides a matrix summarizing the advantages and disadvantages of the different bioassays tested in this study.

Table 18. Strength and Weakness of all Five Assays Tested in this Study.

Assay	YES	ER-CALUX	MELN	KBluc	E-Screen
Analysis of model compounds	+++	+++	++	(+++)	+++
Analysis of environmental samples	-	+++	+	(++)	+++
Ease of use	++	+	+	+ [(a)]	+
Simple training	++	-	-	- [(a)]	-
Low cost	+++	-	+	+	+
Sensitivity	-	+++	++	++	++
Robustness	- [(b)]	++	++	++	++
Reproducibility	++	+++	+	(?)	++
Maturity (widespread use)	+++	++	+	+	+++
High-throughput screening	+++	+++	+++	+++	+++
Quick results	++	++	++	++	-

Notes: "-" = below average, "+" = fair, "++" = good, "+++" = excellent.
(a) This project in fact intended to run the KBluc assay in two independent laboratories, but one laboratory was not able to get the KBluc assay running within pre-defined QA/QC parameters in time before the project deadline.
(b) Octylphenol creeping across the assay plate appears to be a significant problem in this assay, resulting in variable results when octylphenol is present at high concentrations in the samples.

This table clearly highlights that assay has its advantages and limitations, and the notion of "fit-for-purpose" is critical in determining what bioassay to use in a particular project. For example, the YES assay may be suitable for testing sewage where its poor sensitivity may not be a liability but its low cost is clearly an advantage; when testing estrogenicity in drinking water however, a more sensitive (but also more expensive) bioassay such as the ER-CALUX or the E-Screen bioassays may be more appropriate. In the end, any of the five bioassays tested may be suitable for testing estrogenic activity in environmental samples as long as their limitations are clearly understood by the researcher.

4.3 Limitations of *in vitro* Bioassays

It should be emphasized that although this work clearly shows that *in vitro* bioassays for estrogenicity are sufficiently advanced to be used as screening tools for estrogenic contaminants in water, there is still much work to be done to link effects *in vitro* to whole organism *in vivo* responses. Currently, *in vitro* bioassays are useful indicators of potential *in vivo* effects and can be critical in determining mechanisms of effects, but only *in vivo* exposures can conclusively determine whole organism effects. It should also be understood that while *in vitro* bioassays can provide some measure of interaction of mixtures, they can only do so for chemicals that have similar modes of action. In other words, *in vitro* bioassays cannot integrate the whole range of possible interaction of mixtures *in vivo*, where multiple organs might be involved to produce the synergistic effect. Nevertheless, *in vitro* bioassays methods provide a reliable additional tool for measuring exposure and are currently the best available alternative to *in vivo* experiments.

4.4 Future Directions

The concepts and approach used in this study to test the usefulness of several bioassays to detect and quantify estrogenicity in environmental water samples is a robust method to test the usefulness of bioassays for other endocrine effects (such as androgenicity) or even other endpoints relevant to human health, such as carcinogenicity, immunotoxicity, neurotoxicity, etc. A combined research approach could eventually develop a battery of bioassays to detect multiple endpoints relevant to human health to screen large numbers of samples for biologically-active contaminants.

Although not designed to address the issue of inter-laboratory variability in detail, this study showed slight variability in results in the same bioassay from different laboratories. A more thorough study, testing the same assay with the same samples in multiple laboratories (perhaps as many as six or seven) could help identify critical steps of standard operating protocols (SOPs). It is not always possible to completely standardize SOPs, with issues of sourcing of chemical reagents, reliability of support systems, and variations in measuring equipment between different laboratories in different countries sometimes unavoidable. Identifying the critical steps of SOPs may help determine when these issues are irrelevant and when efforts should be made to adhere to the original SOP to avoid compromising data quality.

Finally, while the bioassays evaluated in this study can measure estrogenic activity in environmental samples, more work is needed to identify trigger levels for further investigation. Indeed, as some of these bioassays can detect estrogenic activity as low as few pg/L of estradiol equivalents, it becomes important to clearly understand and communicate levels of associated risks. As our detection limits improve, it is crucial to understand that we will be able to detect lower and lower levels of pollution, and that there are thresholds of toxicological concerns below which pollution may be acceptable. In the case of estrogenic activity for example, based on studies with fish [37, 38, 44], a level of 1 ng/L estradiol equivalent may be acceptable.

5.0 References

1. GWRC. *In vitro bioassays to detect estrogenicity in environmental waters - Literature review*; Water Environment Research Foundation: Alexandria, VA, USA, 2006.
2. GWRC. *Round Robin of Analytical Chemical Methods for EDCs in Water and Sludge*; Global Water Research Coalition: London, UK, 2006.
3. Jobling, S.; Tyler, C.R. Endocrine disruption in wild freshwater fish. *Pure Appl Chem* **2003**, 75, 2219-2234.
4. WHO/IPCS Global assessment of the state-of-the-science of endocrine disruptors, ed. Damstra, T.; Barlow, S., et al. 2002: World Health Organization and International Program on Chemical Safety. 180 pp.
5. Andersen, H.R.; Andersson, A.M.; Arnold, S.F.; et al. Comparison of short-term estrogenicity tests for identification of hormone-disrupting chemicals. *Environ Health Perspect* **1999**, 107, 89-108.
6. Fang, H.; Tong, W.D.; Perkins, R.; et al. Quantitative comparisons of *in vitro* assays for estrogenic activities. *Environ Health Perspect* **2000**, 108, 723-729.
7. Pillon, A.; Boussioux, A.M.; Escande, A.; et al. Binding of estrogenic compounds to recombinant estrogen receptor-alpha: Application to environmental analysis. *Environ Health Perspect* **2005**, 113, 278-284.
8. Murk, A.J.; Legler, J.; van Lipzig, M.M.H.; et al. Detection of estrogenic potency in wastewater and surface water with three *in vitro* bioassays. *Environ Toxicol Chem* **2002**, 21, 16-23.
9. Van den Belt, K.; Berckmans, P.; Vangenechten, C.; et al. Comparative study on the in vitro in vivo estrogenic potencies of 17 beta-estradiol, estrone, 17 alpha-ethynylestradiol and nonylphenol. *Aquatic Toxicology* **2004**, 66, 183-195.
10. Rutishauser, B.V.; Pesonen, M.; Escher, B.I.; et al. Comparative analysis of estrogenic activity in sewage treatment plant effluents involving three in vitro assays and chemical analysis of steroids. *Environ Toxicol Chem* **2004**, 23, 857-864.
11. Folmar, L.; Hemmer, M.; Denslow, N.; et al. A comparison of the estrogenic potencies of estradiol, ethynylestradiol, diethylstilbestrol, nonylphenol and methoxychlor *in vivo* and *in vitro*. *Aquatic Toxicology* **2002**, 60, 101-110.
12. Legler, J.; van den Brink, C.E.; Brouwer, A.; et al. Development of a stably transfected estrogen receptor-mediated luciferase reporter gene assay in the human T47D breast cancer cell line. *Toxicol Sci* **1999**, 48, 55-66.
13. Gutendorf, B.; Westendorf, J. Comparison of an array of *in vitro* assays for the assessment of the estrogenic potential of natural and synthetic estrogens, phytoestrogens and xenoestrogens. *Toxicology* **2001**, 166, 79-89.
14. Leusch, F.D.L.; van den Heuvel, M.R.; Chapman, H.F.; et al. Development of methods for extraction and in vitro quantification of estrogenic and androgenic activity of wastewater samples. *Comparative Biochemistry and Physiology Part C* **2006**, 143, 117-126.
15. Drewes, J.E.; Hemming, J.; Ladenburger, S.J.; et al. An assessment of endocrine disrupting activity changes during wastewater treatment through the use of bioassays and chemical measurements. *Water Environ Res* **2005**, 77, 12-23.

16. Körner, W.; Spengler, P.; Bolz, U.; et al. Substances with estrogenic activity in effluents of sewage treatment plants in southwestern Germany. 2. Biological analysis. *Environ Toxicol Chem* **2001**, 20, 2142-2151.
17. Segner, H.; Navas, J.M.; Schafers, C.; Wenzel, A. Potencies of estrogenic compounds in in vitro screening assays and in life cycle tests with zebrafish in vivo. *Ecotoxicol Environ Saf* **2003**, 54, 315-322.
18. De Boever, P.; Demare, W.; Vanderperren, E.; et al. Optimization of a yeast estrogen screen and its applicability to study the release of estrogenic isoflavones from a soygerm powder. *Environ Health Perspect* **2001**, 109, 691-697.
19. Meerts, I.A.T.M.; Letcher, R.J.; Hoving, S.; et al. In vitro estrogenicity of polybrominated diphenyl ethers, hydroxylated PBDEs, and polybrominated bisphenol A compounds. *Environ Health Perspect* **2001**, 109, 399-407.
20. Körner, W.; Hanf, V.; Schuller, W.; et al. Validation and application of a rapid in vitro assay for assessing the estrogenic potency of halogenated phenolic chemicals. *Chemosphere* **1998**, 37, 2395-2407.
21. Andersen, H.R.; Vinggaard, A.M.; Rasmussen, T.H.; et al. Effects of currently used pesticides in assays for estrogenicity, androgenicity, and aromatase activity *in vitro*. *Toxicol Appl Pharmacol* **2002**, 179, 1-12.
22. Legler, J.; Dennekamp, M.; Vethaak, A.D.; et al. Detection of estrogenic activity in sediment-associated compounds using *in vitro* reporter gene assays. *Sci Total Environ* **2002**, 293, 69-83.
23. Wilson, V.S.; Bobseine, K.; Gray, L.E. Development and characterization of a cell line that stably expresses an estrogen-responsive luciferase reporter for the detection of estrogen receptor agonist and antagonists. *Toxicol Sci* **2004**, 81, 69-77.
24. Cargouet, M.; Perdiz, D.; Mouatassim-Souali, A.; et al. Assessment of river contamination by estrogenic compounds in Paris area (France). *Sci Total Environ* **2004**, 324, 55-66.
25. Pliskova, M.; Vondracek, J.; Canton, R.F.; et al. Impact of polychlorinated Biphenyls contamination on estrogenic activity in human male serum. *Environ Health Perspect* **2005**, 113, 1277-1284.
26. Aerni, H.R.; Kobler, B.; Rutishauser, B.V.; et al. Combined biological and chemical assessment of estrogenic activities in wastewater treatment plant effluents. *Anal Bioanal Chem* **2004**, 378, 688-696.
27. Legler, J.; Jonas, A.; Lahr, J.; et al. Biological measurement of estrogenic activity in urine and bile conjugates with the in vitro ER-CALUX reporter gene assay. *Environ Toxicol Chem* **2002**, 21, 473-479.
28. Routledge, E.J.; Sumpter, J.P. Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen. *Environ Toxicol Chem* **1996**, 15, 241-248.
29. Balaguer, P.; François, F.; Comunale, F.; et al. Reporter cell lines to study the estrogenic effects of xenoestrogens. *Sci Total Environ* **1999**, 233, 47-56.
30. Fenet, H.; Gomez, E.; Pillon, A.; et al. Estrogenic activity in water and sediments of a French river: Contribution of alkylphenols. *Arch Environ Contam Toxicol* **2003**, 44, 1-6.
31. Soto, A.M.; Sonnenschein, C.; Chung, K.L.; et al. The E-Screen assay as a tool to identify estrogens: an update on estrogenic environmental pollutants. *Environ Health Perspect* **1995**, 103, 113-122.

32. Körner, W.; Hanf, V.; Schuller, W.; et al. Development of a sensitive E-screen assay for quantitative analysis of estrogenic activity in municipal sewage plant effluents. *Sci Total Environ* **1999**, 225, 33-48.
33. Rasmussen, T.H.; Nielsen, J.B. Critical parameters in the MCF-7 cell proliferation bioassay (E-Screen). *Biomarkers* **2002**, 7, 322-336.
34. Matsuoka, S.; Kikuchi, M.; Kimura, S.; et al. Determination of estrogenic substances in the water of Muko River using in vitro assays, and the degradation of natural estrogens by aquatic bacteria. *J Health Sci* **2005**, 51, 178-184.
35. Villalobos, M.; Olea, N.; Brotons, J.A.; et al. The E-Screen assay: a comparison of different MCF7 cell stocks. *Environ Health Perspect* **1995**, 103, 844-850.
36. Kinnberg, K. *Evaluation of in vitro assays for determination of estrogenic activity in the environment*; 43; Danish Environmental Protection Agency: Copenhagen, Denmark, 2003.
37. Routledge, E.J.; Sheahan, D.; Desbrow, C.; et al. Identification of estrogenic chemicals in STW effluent. 2. *In vivo* responses in trout and roach. *Environmental Science and Technology* **1998**, 32, 1559-1565.
38. Thorpe, K.L.; Cummings, R.I.; Hutchinson, T.H.; et al. Relative potencies and combination effects of steroidal estrogens in fish. *Environmental Science and Technology* **2003**, 37, 1142-1149.
39. Silva, E.; Lopez-Espinosa, M.J.; Molina-Molina, J.M.; et al. Lack of activity of cadmium in in vitro estrogenicity assays. *Toxicol Appl Pharmacol* **2006**, 216, 20-28.
40. Helsel, D.R. Fabricating data: How substituting values for nondetects can ruin results, and what can be done about it. *Chemosphere* **2006**, 65, 2434-2439.
41. Lyles, R.H.; Fan, D.; Chuachoowong, R. Correlation coefficient estimation involving a left censored laboratory variable. *Stat Med* **2001**, 20, 2921-2933.
42. WHO Guidelines for drinking water quality. Third edition ed. 2004, Geneva, Switzerland.
43. NHMRC/NRMMC Australian drinking water guidelines. 2004.
44. Metcalfe, C.D.; Metcalfe, T.L.; Kiparissis, Y.; et al. Estrogenic potency of chemicals detected in sewage treatment plant effluents as determined by *in vivo* assays with Japanese medaka (*Oryzias latipes*). *Environ Toxicol Chem* **2001**, 20, 297-308.

Appendix I

Extra Data Tables

Table 19. Bioassay Results (in estradiol equivalents, EEq ± SD).

Sample ID	Description	YES 1 (ng/L)	YES 2 (ng/L)	ER-CALUX (ng/L)	MELN (ng/L)	T47D-KBluc (ng/L)	E-Screen 1 (ng/L)	E-Screen 2 (ng/L)
A01	Artificial - Estradiol	**< 3.5**	**< 5**	0.445 ± 0.007 (2)	0.833 ± 0.643	0.630 ± 0.113 (2)	1.94 ± 2.71	0.287 ± 0.323
A02	Artificial - Nonylphenol	63.3 ± 18.0	262 ± 185	28.3 ± 27.5	2.29 ± 3.65	6.66 ± 2.50 (2)	9.31 ± 4.65	16.5 ± 3.93
A03	Artificial - Octylphenol	3,457 ± 720	2,758 ± 478	**Cytotoxic**	7.77 ± 6.65	91.7 ± 91.2 (2)	972 ± 315	669 ± 33.1
A04	Artificial - Benzy Butyl Phthalate	58.3 ± 17.7	**< 5**	15.0 ± 5.29	5.27 ± 8.78	10.0 ± 7.73 (2)	6.40 ± 2.79	11.6 ± 1.37
A05	Artificial - TBBPA	**cytotoxic**	**< 5**	0.197 ± 0.127	0.467 ± 0.635	0.270 ± 0.000 (2)	3.60 ± 6.06	**< 0.2**
A06	Artificial - p,p'-DDT	**< 3.5**	**< 5**	1.90 ± 0.346	2.73 ± 2.87	1.28 ± 0.651 (2)	0.777 ± 0.913	36.6 ± 57.7
A07	Artificial - Tamoxifen	14.6 ± 13.4	6.85 ± 4.76	2.57 ± 1.04	2.10 ± 2.88	8.18 ± 0.028 (2)	9.77 ± 13.6	10.5 ± 12.9
A08	Artificial - Hormone mix	**< 3.5**	7.47 ± 4.53	2.73 ± 0.777	**< 0.27**	N/A	8.57 ± 9.27	2.40 ± 2.68
A09	Artificial - Industrial mix	279 ± 346	321 ± 374	153 ± 32.1	3.27 ± 2.62	N/A	79.2 ± 57.2	105 ± 22.2
A10	Artificial - Agro-chemical mix	**< 3.5**	**< 5**	3.97 ± 1.40	7.03 ± 4.90	N/A	0.723 ± 0.591	31.0 ± 53.6
A11	Artificial - Combo mix	102 ± 74.2	162 ± 96.8	223 ± 35.1	4.57 ± 6.81	N/A	83.9 ± 64.1	186 ± 96.0
A12	Artificial - Blank	**< 3.5**	**< 5**	**< 0.1**	**< 0.27**	N/A	**< 0.2**	**< 0.2**
GW1-	Shallow aquifer (non spiked)	**< 3.5**	**< 5**	0.211 ± 0.093	**< 0.27**	**< 0.2** (1)	**< 0.2**	0.205 ± 0.181
GW1+	Shallow aquifer (spiked)	23.8 ± 2.56	77.8 ± 48.6	79.7 ± 28.0	25.6 ± 31.9	95.8 (1)	28.0 ± 9.46	52.5 ± 11.2
GW2-	Deep aquifer (non spiked)	**< 3.5**	**< 5**	0.133 ± 0.044	**< 0.27**	**< 0.2** (1)	**< 0.2**	**< 0.2**
GW2+	Deep aquifer (spiked)	20.9 ± 3.78	48.7 ± 14.9	92.9 ± 21.4 (2)	26.3 ± 17.9	43.3 (1)	19.6 ± 11.6	38.0 ± 12.0
SR1-	Raw sewage 1 (non spiked)	19.2 ± 9.74	45.1 ± 16.9	80.1 ± 33.6	40.0 ± 33.4	62.6 (1)	10.5 ± 4.95	62.7 ± 9.54
SR1+	Raw sewage 1 (spiked)	41.5 ± 14.9	112 ± 22.5	352 ± 60.7	31.4 ± 18.8	239 (1)	45.0 ± 17.5	191 ± 35.4
SR2-	Raw sewage 2 (non spiked)	8.95 ± 9.29	59.6 ± 29.0	81.5 ± 39.5	60.6 ± 40.9	97.7 (1)	21.6 ± 8.54	73.6 ± 23.6
SR2+	Raw sewage 2 (spiked)	47.7 ± 12.1	109 ± 21.5	267 ± 52.9	20.8 ± 12.6	220 (1)	56.0 ± 4.66	160 ± 43.4
ST1-	Treated sewage 1 (non spiked)	**< 3.5**	**< 5**	4.57 ± 1.03 (2)	3.70 ± 1.37	9.19 (1)	0.367 ± 0.242	1.57 ± 0.199
ST1+	Treated sewage 1 (spiked)	24.3 ± 3.38	58.8 ± 25.8	97.4 ± 32.2	27.0 ± 16.5	74.4 (1)	26.0 ± 23.6	44.9 ± 8.39
ST2-	Treated sewage 2 (non spiked)	**< 3.5**	**< 5**	0.718 ± 0.270	**< 0.27**	**< 0.2** (1)	1.51 ± 2.43	**< 0.2**
ST2+	Treated sewage 2 (spiked)	4.42 ± 2.85	36.4 ± 17.2	5.21 ± 1.08	8.29 ± 3.94	3.45 (1)	2.99 ± 3.62	3.61 ± 0.987
RW1-	River 1 (non spiked)	**< 3.5**	**< 5**	0.150 ± 0.018	**< 0.27**	**< 0.2** (1)	**< 0.2**	**< 0.2**
RW1+	River 1 (spiked)	20.7 ± 5.93	38.2 ± 15.9	64.2 ± 4.33	21.7 ± 5.73	81.3 (1)	47.1 ± 25.4	45.3 ± 22.9
RW2-	River 2 (non spiked)	**< 3.5**	**< 5**	0.345 ± 0.010	**< 0.27**	0.48 (1)	0.218 ± 0.205	0.469 ± 0.320
RW2+	River 2 (spiked)	25.5 ± 4.93	66.0 ± 40.7	136 ± 22.1	36.3 ± 30.2	86.8 (1)	21.1 ± 3.06	37.1 ± 9.15

Note that the data in this table is presented in Figure 2 and Figure 4 in the text.
(1) Bioassay performed on one independent sample only.
(2) Bioassay performed on two independent samples only.

N/A = KBluc analysis not available.

Table 20. Chemical Data (1 of 3) for Industrial Chemicals.

Sample ID	Description	Lab 1 4t-OP (µg/L)	Lab 2 4t-OP (µg/L)	Lab 1 NP (µg/L)	Lab 2 NP (µg/L)	Lab 1 Bisphenol A (µg/L)	Lab 2 Bisphenol A (µg/L)	Lab 1 BBP (µg/L)	Lab 2 BBP (µg/L)	Lab 2 TBBPA (µg/L)
A01	Artificial - Estradiol	< 5		< 5		< 5		< 50		
A02	Artificial - Nonylphenol	< 5		273 ± 187	238 ± 20.2	< 5		< 50		
A03	Artificial - Octylphenol	1,494 ± 118	2,767 ± 375	< 5		< 5		< 50		
A04	Artificial - Benzy Butyl Phthalate	< 5		< 5		< 5		11,741 ± 14,432	3,983 ± 535	
A05	Artificial - TBBPA	< 5		< 5		< 5		< 50		1,483 ± 839
A06	Artificial - p,p'-DDT	5.97 ± 6.00		< 5		< 5		< 50		
A07	Artificial - Tamoxifen	< 5		< 5		< 5		< 50		
A08	Artificial - Hormone mix	< 5		< 5		< 5		< 50		
A09	Artificial - Industrial mix	284 ± 161	1,017 ± 231	60.5 ± 28.1	78.3 ± 2.89	36.3 ± 17.9	39.8 ± 3.01	2,039 ± 1,358	600 ± 86.6	450 ± 180
A10	Artificial - Agro-chemical mix	< 5		< 5		< 5		< 50		
A11	Artificial - Combo mix	402 ± 69.3	1,067 ± 144	72.4 ± 6.56	96.7 ± 2.89	47.7 ± 8.44	45.2 ± 2.25	1,290 ± 580	650 ± 218	192 ± 14.4
A12	Artificial - Blank	< 5	< 0.025	< 5	< 0.125	< 5	< 0.025	< 50	< 5	< 5
GW1-	Shallow aquifer (non spiked)	< 5	0.583 ± 0.454	< 5	0.271 ± 0.219	< 5	< 0.025	< 50	< 5	
GW1+	Shallow aquifer (spiked)	6.27 ± 6.52	5.05 ± 8.18	98.3 ± 9.83	69.5 ± 23.7	61.7 ± 13.1	56.5 ± 6.06	1,312 ± 369	708 ± 52.0	
GW2-	Deep aquifer (non spiked)	< 5	< 0.025	< 5	< 0.125	< 5	< 0.025	< 50	< 5	
GW2+	Deep aquifer (spiked)	< 5	0.300 ± 0.100	95.4 ± 12.6	65.8 ± 29.0	51.6 ± 5.21	56.7 ± 7.64	935 ± 190	708 ± 123	
SR1-	Raw sewage 1 (non spiked)	< 12.5	2.62 ± 2.53	< 12.5	6.33 ± 1.15	5.13 ± 4.56	0.433 ± 0.029	< 125	< 12.5	
SR1+	Raw sewage 1 (spiked)	< 12.5	7.28 ± 7.35	17.9 ± 9.50	20.8 ± 2.08	51.2 ± 24.5	47.2 ± 2.25	350 ± 196	68.8 ± 54.1	
SR2-	Raw sewage 2 (non spiked)	< 12.5	0.400 ± 0.087	< 12.5	3.64 ± 5.52	< 12.5	0.633 ± 0.126	< 125	< 12.5	
SR2+	Raw sewage 2 (spiked)	< 12.5	0.450 ± 0.100	< 12.5	14.8 ± 0.764	68.8 ± 6.71	41.5 ± 3.04	182 ± 17.1	< 12.5	
ST1-	Treated sewage 1 (non spiked)	< 5	0.500 ± 0.278	< 5	1.12 ± 0.076	< 5	0.050 ± 0.000	< 50	< 5	
ST1+	Treated sewage 1 (spiked)	< 5	4.20 ± 5.47	21.9 ± 6.74	32.5 ± 6.00	45.2 ± 3.15	52.2 ± 7.01	880 ± 662	325 ± 75.0	
ST2-	Treated sewage 2 (non spiked)	< 5	0.233 ± 0.104	< 5	0.338 ± 0.239	< 5	< 0.025	< 50	< 5	
ST2+	Treated sewage 2 (spiked)	< 5	0.817 ± 0.679	11.8 ± 4.81	12.2 ± 1.04	14.7 ± 4.55	19.2 ± 1.89	792 ± 425	350 ± 100	
RW1-	River 1 (non spiked)	< 5	0.483 ± 0.029	< 5	0.221 ± 0.146	< 5	< 0.025	< 50	< 5	
RW1+	River 1 (spiked)	< 5	1.22 ± 0.321	28.1 ± 7.37	39.5 ± 2.00	50.2 ± 5.88	49.2 ± 5.20	555 ± 114	300 ± 50.0	
RW2-	River 2 (non spiked)	< 5	0.367 ± 0.153	< 5	25.3 ± 43.1	< 5	< 0.025	< 50	< 5	
RW2+	River 2 (spiked)	< 5	0.383 ± 0.126	45.8 ± 5.14	53.3 ± 2.89	47.3 ± 5.56	49.0 ± 5.41	734 ± 125	500 ± 66.1	

Values are average ± SD, n = 3 independent samples.

Table 21. Chemical Data (2 of 3) - Natural Hormones and Drugs.

ID	Description	Lab 1 Estrone (µg/L)	Lab 2 Estrone (µg/L)	Lab 1 b-Estradiol (µg/L)	Lab 2 b-Estradiol (µg/L)	Lab 1 Estriol (µg/L)	Lab 2 Estriol (µg/L)	Lab 1 Tamoxifen (µg/L)	Lab 2 Tamoxifen (µg/L)	Lab 2 EE2 (µg/L)
A01	Artificial - Estradiol	< 0.01		< 0.005	< 0.005	< 0.01		< 0.1		
A02	Artificial - Nonylphenol	< 0.01		< 0.005		< 0.01		< 0.1		
A03	Artificial - Octylphenol	< 0.01		< 0.005		< 0.01		< 0.1		
A04	Artificial - Benzy Butyl Phthalate	< 0.01		< 0.005		< 0.01		< 0.1		
A05	Artificial - TBBPA	< 0.01		< 0.005		0.028 ± 0.040		< 0.1		
A06	Artificial - p,p'-DDT	< 0.01		< 0.005		< 0.01		< 0.1		
A07	Artificial - Tamoxifen	< 0.01		< 0.005		< 0.01		60.9 ± 34.2	17.3 ± 1.53	
A08	Artificial - Hormone mix	< 0.01	0.014 ± 0.003	< 0.005	< 0.005	< 0.01	0.006 ± 0.003	< 0.1		0.006 ± 0.002
A09	Artificial - Industrial mix	< 0.01		< 0.005		< 0.01		< 0.1		
A10	Artificial - Agro-chemical mix	< 0.01		< 0.005		< 0.01		< 0.1		
A11	Artificial - Combo mix	0.032 ± 0.008	0.056 ± 0.013	0.011 ± 0.002	0.009 ± 0.001	0.028 ± 0.040	0.048 ± 0.006	103 ± 23.7	20.8 ± 2.36	0.010 ± 0.002
A12	Artificial - Blank	< 0.01	< 0.005	< 0.005	< 0.005	< 0.01	< 0.005	< 0.1	< 0.005	< 0.005
GW1-	Shallow aquifer (non spiked)	< 0.01	< 0.005	< 0.005	< 0.005	< 0.01	< 0.005	< 0.1		< 0.005
GW1+	Shallow aquifer (spiked)	< 0.01	0.075 ± 0.022	< 0.005	0.025 ± 0.005	< 0.01	0.018 ± 0.013	< 0.1		< 0.005
GW2-	Deep aquifer (non spiked)	< 0.01	< 0.005	< 0.005	< 0.005	< 0.01	< 0.005	< 0.1		< 0.005
GW2+	Deep aquifer (spiked)	0.017 ± 0.016	0.065 ± 0.005	0.012 ± 0.004	0.023 ± 0.008	0.043 ± 0.003	0.030 ± 0.009	< 0.1		0.008 ± 0.004
SR1-	Raw sewage 1 (non spiked)	0.046 ± 0.036	< 0.013	0.023 ± 0.018	< 0.013	0.232 ± 0.054	0.180 ± 0.009	< 0.25		< 0.013
SR1+	Raw sewage 1 (spiked)	0.119 ± 0.072	0.283 ± 0.014	0.055 ± 0.043	0.032 ± 0.008	0.337 ± 0.203	0.197 ± 0.010	< 0.25		< 0.013
SR2-	Raw sewage 2 (non spiked)	0.081 ± 0.021	0.127 ± 0.108	0.046 ± 0.018	< 0.013	0.460 ± 0.185	0.127 ± 0.060	< 0.25		< 0.013
SR2+	Raw sewage 2 (spiked)	0.131 ± 0.001	0.238 ± 0.006	0.088 ± 0.016	< 0.013	0.581 ± 0.117	0.142 ± 0.079	< 0.25		< 0.013
ST1-	Treated sewage 1 (non spiked)	< 0.01	0.023 ± 0.021	< 0.005	< 0.005	< 0.01	< 0.005	< 0.1		< 0.005
ST1+	Treated sewage 1 (spiked)	0.066 ± 0.011	0.143 ± 0.003	0.028 ± 0.005	0.018 ± 0.012	0.056 ± 0.048	0.055 ± 0.005	< 0.1		< 0.005
ST2-	Treated sewage 2 (non spiked)	< 0.01	< 0.005	< 0.005	< 0.005	< 0.01	< 0.005	< 0.1		< 0.005
ST2+	Treated sewage 2 (spiked)	< 0.01	< 0.005	< 0.005	0.017 ± 0.013	< 0.01	< 0.005	< 0.1		< 0.005
RW1-	River 1 (non spiked)	< 0.01	< 0.005	< 0.005	< 0.005	< 0.01	< 0.005	< 0.1		< 0.005
RW1+	River 1 (spiked)	0.026 ± 0.019	0.078 ± 0.008	0.014 ± 0.010	0.010 ± 0.000	0.055 ± 0.045	0.040 ± 0.005	< 0.1		0.006 ± 0.004
RW2-	River 2 (non spiked)	< 0.01	0.007 ± 0.007	< 0.005	< 0.005	< 0.01	< 0.005	< 0.1		< 0.005
RW2+	River 2 (spiked)	0.020 ± 0.013	0.103 ± 0.006	0.015 ± 0.011	0.007 ± 0.003	0.028 ± 0.040	0.045 ± 0.005	< 0.1		0.008 ± 0.003

Values are average ± SD, n = 3 independent samples.

Table 22. Chemical Data (3 of 3) - Pesticides and Genistein.

Sample ID	Description	Lab 1 α-Endosulfan (µg/L)	Lab 1 β-Endosulfan (µg/L)	Lab 2 Endosulfan (µg/L)	Lab 1 Dieldrin (µg/L)	Lab 2 Dieldrin (µg/L)	Lab 1 p,p'-DDT (µg/L)	Lab 2 p,p'-DDT (µg/L)	Lab 2 Genistein (µg/L)
A01	Artificial - Estradiol	< 10	< 10		< 10		< 50		
A02	Artificial - Nonylphenol	< 10	< 10		< 10		< 50		
A03	Artificial - Octylphenol	< 10	< 10		< 10		< 50		
A04	Artificial - Benzy Butyl Phthalate	< 10	< 10		< 10		< 50		
A05	Artificial - TBBPA	< 10	< 10		< 10		< 50		
A06	Artificial - p,p'-DDT	10.7 ± 9.81	16.6 ± 20.1		15.0 ± 17.4		7,095 ± 8,663	2,000 ± 218	
A07	Artificial - Tamoxifen	< 10	< 10		< 10		< 50		
A08	Artificial - Hormone mix	< 10	< 10		< 10		< 50		
A09	Artificial - Industrial mix	< 10	< 10		< 10		< 50		
A10	Artificial - Agro-chemical mix	1,048 ± 962	1,075 ± 915	567 ± 28.9	1,122 ± 990	442 ± 40.4	1,679 ± 1,532	423 ± 48.6	4.01 ± 3.49
A11	Artificial - Combo mix	647 ± 323	681 ± 308	487 ± 18.9	805 ± 389	455 ± 22.9	1,195 ± 588	470 ± 47.7	8.50 ± 0.500
A12	Artificial - Blank	< 10	< 10	< 0.05	< 10	< 0.05	< 50	< 0.05	< 0.05
GW1-	Shallow aquifer (non spiked)	< 10	< 10		< 10		< 50		
GW1+	Shallow aquifer (spiked)	< 10	< 10		< 10		< 50		
GW2-	Deep aquifer (non spiked)	< 10	< 10		< 10		< 50		
GW2+	Deep aquifer (spiked)	< 10	< 10		< 10		< 50		
SR1-	Raw sewage 1 (non spiked)	< 25	< 25		< 25		< 125		
SR1+	Raw sewage 1 (spiked)	< 25	< 25		< 25		< 125		
SR2-	Raw sewage 2 (non spiked)	< 25	< 25		< 25		< 125		
SR2+	Raw sewage 2 (spiked)	< 25	< 25		< 25		< 125		
ST1-	Treated sewage 1 (non spiked)	< 10	< 10		< 10		< 50		
ST1+	Treated sewage 1 (spiked)	< 10	< 10		< 10		< 50		
ST2-	Treated sewage 2 (non spiked)	< 10	< 10		< 10		< 50		
ST2+	Treated sewage 2 (spiked)	< 10	< 10		< 10		< 50		
RW1-	River 1 (non spiked)	< 10	< 10		< 10		< 50		
RW1+	River 1 (spiked)	< 10	< 10		< 10		< 50		
RW2-	River 2 (non spiked)	< 10	< 10		< 10		< 50		
RW2+	River 2 (spiked)	< 10	< 10		< 10		< 50		

Values are average ± SD, n = 3 independent samples.

Table 23. Measured and Predicted Estradiol Equivalents (Eeq) for the Artificial Sample Mixes Based on Combined Chemical Data and Relative Potencies of Individual Chemicals.

Mix ID	Sample	Data type	YES 1	YES 2	ER-CALUX	MELN	KBluc	E-Screen 1	E-Screen 2
A08	Hormone mix	Measured (ng/L)	< 3.5	7.47	2.73	< 0.27	N/A	8.57	2.40
		Predicted (ng/L)	10.8	10.8	14.0	15.0	2.43 (b)	4.72	4.72
		Meas/Pred	< 0.32	0.69	0.20	< 0.02	N/A	1.81	0.51
A09	Industrial mimics mix	Measured (ng/L)	279	321	153	3.27	N/A	79.2	105
		Predicted (ng/L)	1155	1002	12.9	11.4	34.0 (b)	322	227
		Meas/Pred	0.24	0.32	11.8	0.29	N/A	0.25	0.46
A10	Agro-industrial mix	Measured (ng/L)	< 3.5	< 5	3.97	7.03	N/A	0.72	31.0
		Predicted (ng/L)	< 3.5 (a)	< 5 (a)	1.32	3.16 (a)	0.49 (b)	1.51	9.06
		Meas/Pred	1	1	3.00	2.22	N/A	0.48	3.42
A11	Combo mix	Measured (ng/L)	102	162	223	4.57	N/A	83.9	186
		Predicted (ng/L)	1221 (a)	1068 (a)	91.2	65.0 (a)	59.0 (b)	359	273
		Meas/Pred	0.08	0.15	2.48	0.07	N/A	0.23	0.68
A12	Blank	Measured (ng/L)	< 3.5	< 5	< 0.1	< 0.27	N/A	< 0.2	< 0.2
		Predicted (ng/L)	< 3.5	< 5	< 0.1	< 0.27	< 0.20 (b)	< 0.2	< 0.2
		Meas/Pred	1	1	1	1	N/A	1	1

Notes: Note that the *Meas/Pred* ratio from this table are presented in

Figure 7 in the body of the document. Predicted EEq *Pred* was calculated as $Pred = \sum (RP \times conc)$, where *RP* is the relative potency (Table 1) and *conc* is the concentration (Table 9 and Table 10). When available, relative potency obtained in that particular assay in this study (samples A01-A07) was used, otherwise the average of the relative potencies for that assay type published in the literature was used (Table 1).

In Table 23, individual compounds (A01 – A07) are not shown because the relative potency (Table 1) for those particular compound was derived from the measured EEq (Table 19), and hence the ratio of measured/predicted is always 1.

N/A = data not available for the KBluc assay.

(a) Predicted EEq may be undervalued due to missing potencies for dieldrin and endosulfan in YES and MELN assays (Table 1).

(b) Predicted EEq is undervalued due to missing potency for estriol in the KBluc assay (Table 1).

Table 24. Measured and Predicted Estradiol Equivalents (Eeq) in Field Samples Based on Combined Chemistry and Relative Potencies of Chemicals.

Mix ID	Sample	Data type	YES 1	YES 2	ER-CALUX	MELN	KBluc	E-Screen 1	E-Screen 2
GW1-	Shallow aquifer (non spiked)	Measured (ng/L)	< 3.5	< 5	0.21	< 0.27	< 0.2	< 0.2	0.20
		Predicted (ng/L)	< 3.5	< 5	< 0.1	< 0.27	< 0.2 (a)	0.36	0.26
		Meas/Pred	1	1	> 2.1	1	1	< 0.55	0.78
GW1+	Shallow aquifer (spiked)	Measured (ng/L)	23.8	77.8	79.7	25.6	95.8	28.0	52.5
		Predicted (ng/L)	98.4	164	61.9	63.2	31.5 (a)	37.8	40.5
		Meas/Pred	0.24	0.47	1.29	0.40	3.04	0.74	1.30
GW2-	Deep aquifer (non spiked)	Measured (ng/L)	< 3.5	< 5	0.13	< 0.27	< 0.2	< 0.2	< 0.2
		Predicted (ng/L)	< 3.5	< 5	< 0.1	< 0.27	< 0.2 (a)	< 0.2	< 0.2
		Meas/Pred	1	1	> 1.3	1	1	1	1
GW2+	Deep aquifer (spiked)	Measured (ng/L)	20.9	48.8	92.9	26.3	43.3	19.6	38.0
		Predicted (ng/L)	81.5	146	83.5	63.1	22.8 (a)	30.6	34.3
		Meas/Pred	0.26	0.33	1.11	0.42	1.90	0.64	1.11
SR1-	Raw sewage 1 (non spiked)	Measured (ng/L)	19.2	45.0	80.1	40.0	62.6	10.5	62.7
		Predicted (ng/L)	28.2	32.4	199	43.3	16.7 (a)	33.4	33.1
		Meas/Pred	0.68	1.39	0.40	0.93	3.75	0.31	1.89
SR1+	Raw sewage 1 (spiked)	Measured (ng/L)	41.5	112	352	31.4	239	45.0	191
		Predicted (ng/L)	130	143	249	153.2	40.2 (a)	60.0	59.4
		Meas/Pred	0.32	0.79	1.41	0.21	5.94	0.75	3.22
SR2-	Raw sewage 2 (non spiked)	Measured (ng/L)	8.95	59.6	81.5	60.6	97.7	21.6	73.6
		Predicted (ng/L)	46.9	49.8	188	69.5	25.9 (a)	38.9	38.9
		Meas/Pred	0.19	1.20	0.43	0.87	3.77	0.56	1.89
SR2+	Raw sewage 2 (spiked)	Measured (ng/L)	47.7	109	267	20.8	220	56.0	160
		Predicted (ng/L)	88.4	99.5	332	128	46.6 (a)	70.9	71.4
		Meas/Pred	0.54	1.09	0.80	0.16	4.72	0.79	2.25
ST1-	Treated sewage 1 (non spiked)	Measured (ng/L)	< 3.5	< 5	4.57	3.70	9.19	0.37	1.57
		Predicted (ng/L)	14.9	14.2	1.43	7.31	0.77 (a)	2.83	2.09
		Meas/Pred	< 0.23	< 0.35	3.19	0.51	12.0	0.13	0.75
ST1+	Treated sewage 1 (spiked)	Measured (ng/L)	24.3	58.8	97.4	27.0	74.4	26.0	44.9
		Predicted (ng/L)	79.3	97.9	94.9	86.4	31.7 (a)	36.7	37.9
		Meas/Pred	0.31	0.60	1.03	0.31	2.35	0.71	1.18
ST2-	Treated sewage 2 (non spiked)	Measured (ng/L)	< 3.5	< 5	0.72	< 0.27	< 0.2	1.51	< 0.2
		Predicted (ng/L)	< 3.5	< 5	< 0.1	< 0.27	< 0.2 (a)	0.51	0.36
		Meas/Pred	1	1	> 7.2	1	1	2.98	< 0.55
ST2+	Treated sewage 2 (spiked)	Measured (ng/L)	4.42	36.4	5.21	8.29	3.45	2.99	3.61
		Predicted (ng/L)	18.9	23.9	10.1	11.0	8.35 (a)	9.04	9.81
		Meas/Pred	0.23	1.52	0.52	0.76	0.41	0.33	0.37
RW1-	River 1 (non spiked)	Measured (ng/L)	< 3.5	< 5	0.15	< 0.27	< 0.2	< 0.2	< 0.2
		Predicted (ng/L)	< 3.5	< 5	< 0.1	< 0.27	< 0.2 (a)	0.74	0.52
		Meas/Pred	1	1	> 1.5	1	1	< 0.27	< 0.38
RW1+	River 1 (spiked)	Measured (ng/L)	20.7	38.2	64.2	21.7	81.3	47.1	45.3
		Predicted (ng/L)	56.7	84.0	67.7	56.7	15.7 (a)	22.4	23.9
		Meas/Pred	0.36	0.45	0.95	0.38	5.18	2.11	1.89
RW2-	River 2 (non spiked)	Measured (ng/L)	< 3.5	< 5	0.35	< 0.27	0.48	0.22	0.47
		Predicted (ng/L)	3.96	8.16	1.03	2.27	0.32 (a)	0.51	0.60
		Meas/Pred	< 0.88	< 0.61	0.34	< 0.12	1.50	0.42	0.78
RW2+	River 2 (spiked)	Measured (ng/L)	24.5	66.0	136	36.3	86.8	21.1	37.1
		Predicted (ng/L)	68.0	103	76.4	64.4	15.7 (a)	23.1	25.3
		Meas/Pred	0.37	0.64	1.78	0.57	5.54	0.91	1.47

Notes: Note that the *Meas/Pred* ratio from this table are presented in Figure 8 in the body of the document. Predicted EEq *Pred* was calculated as $Pred = \sum (RP \times conc)$, where *RP* is the relative potency (Table 1) and *conc* is the concentration (Table 9 and Table 10). When available, relative potency obtained in that particular assay in this study (samples A01-A07) was used, otherwise the average of the relative potencies for that assay type published in the literature was used (Table 1).

(b) Predicted EEq is undervalued due to missing potency for estriol in the KBluc assay (Table 1).

Global Water Research Coalition
c/o International Water Association
Alliance House
12 Caxton Street
London SW1H 0QS
United Kingdom
tel: +44 207 654 5545
email: gwrc@iwahq.org.uk
www.globalwaterresearchcoalition.net