Contract n° 018486

NORMAN

Network of reference laboratories and related organisations for monitoring and bio monitoring of emerging environmental pollutants

Co-ordination action
Priority 6.3 – Global Change and Ecosystems

Deliverable C1.2:

Validation and harmonisation of analytical methodology for research laboratories: Oestrogens in sewage treatment effluents

Due date of deliverable: January 2008
Actual submission date: October 2008

Start date of the project: 1st September 2005 Duration: 3 years

Work package leader: **UK Environment Agency**
Leader for this task: **UK Environment Agency**

Contact person:
Anne O'Neill
Tel: +44 (0) 1491 828422
Fax: +44 (0) 1491 834703
e-mail: anne.oneill@environment-agency.gov.uk

Revision: final version

Dissemination level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

CONTENTS

1 Introduction

This report describes the outcome of the CASE 1 Work Package Tasks 2 and 3, production of a report on the analytical chemistry, bioassays, biomarker assays and intra laboratory study (C1.2) and the summary, feedback and recommendations to the NETWORK, VALIDATION and SEARCH work packages (C1.3).

The overall aim of Case Study 1 was to assess the performance of a selection of methods for detecting the presence of oestrogenic substances in sewage treatment work effluents by validating against the recovery of spiked endocrine disrupting chemicals in waters and effluents and the results of *in vivo* assays for reproductive effects in fish. The methods were selected to cover three different levels:

Type 1 methods – *In vitro* bioassays that detect oestrogenic activity (for example, YES, ER-CALUX, MELN, E-Screen)

Type 2 methods – *In vivo* bioassays that detect oestrogenic activity, where the biomarker product of the bioassay is the measured parameter (for example vitellogenin, vitellogenin mRNA, vitelline envelope protein) as opposed to somatic changes (e.g. intersex)

Type 3 methods – *In vitro* direct measurements assays for measurement of quantities of target compounds, ethynyl- oestradiol, 17 ß- oestradiol, estriol, estrone, (for example ELISA for 17 ß-oestradiol).

From the review of existing research methods (Deliverable C1.1), (Wegener, 2006), methods were ranked for each type based upon the V1 criteria tier-based method selection which assessed scientific basis (Tier 1), trueness (Tier 2) and calibration and traceability (Tier 3). The methods which were ranked highly and accepted the invitation to join the NORMAN programme for each method type were as follows:

Type 1 method selected: E-Screen (human mammary carcinoma cell proliferation assay) applied to rivers, ponds, wetlands and municipal wastewater effluent (Shappell, 2006).

Type 2 method selected: Direct homologous quantitative sandwich monoclonal ELISA for fathead minnow vitellogenin in blood plasma (Eidem *et al.*, 2006)

Type 3 method selected: Monoclonal ELISA detecting 17 ß-oestradiol in municipal wastewaters following SPE extraction (Hirobe *et al.*, 2004).

This report provides an evaluation of the selected methods performance against the validation protocols prepared by the Project NORMAN VALIDATION Group at the research laboratory level (Leonards and Schwesig, 2007). Each method is outlined in terms of the modules described in deliverable V1.1a which are:

Module A: Test method definition, documentation and general requirements.

Module B: Applicability domain and pre-validation

Module C: Intra-laboratory performance

For module C, the parameters identified were not followed exactly as outlined in Validation protocol. Instead, the Validation and Case 1 work groups had outlined a specific programme of testing for the generation of validation data for Type 1, 2, and 3 methods. These set protocols looked at different assessment stages to validate the methods. These were:

Stage 1: Assessment of Accuracy, Precision, Linearity and Range.

Stage 2: Assessment of Negative Response and Selectivity

Stage 3: Assessment of Specificity and Discriminative Ability in Environmental Matrices

Stage 4: Assessment of Relative Potency of oestrogenic Compounds

Not all method types were validated for all stages or all performance characteristics depending on what validation criteria the method was being tested against. Consequently, intra-laboratory performance has been reported against these difference stages rather than splitting up the data into the module C parameters.

This document therefore provides an assessment of the methods performance characteristics, advantages and limitations, identifies the influences which may change these characteristics, and overall, if the method is fit for purpose.

2 Contributors

Once the research groups had agreed to participate in CASE 1, a workshop was held in London, UK in January 2007 to assess the suitability of the validation protocols and rationale for each method types. The workshop was held by CASE 1 work package leader Rachel Benstead (UKEA). Contributions were made from the VALIDATION group (Pim Leonards, IVM and Dean Leverett, UKEA), CASE 1 method participants Nancy Shappell USDA-ARS (Type 1 methods), Anders Goksoyr and Sven Kristiansen, Biosense (Type 2 methods and representing the work by Japan-EnviroChemicals for Type 3 methods). Furthermore, the validation protocols and suitability of methods were peer reviewed both by internally (Rakesh Kanda, STL) and by an external reviewer (Mark Crane, Watts Crane Associates) in order to ensure the developed protocols were scientifically sound before the methods laboratories undertook the validation assessments. Fish plasma, effluent samples and spiked oestrogen samples were sent out to laboratory participants (method laboratories) and for chemical analysis (UKEA National Laboratory Service) in February 2007 by the lead laboratory, UKEA (Rachel Benstead). Results were collated by UKEA (Anne O'Neill) in June 2007 and IVM (Jan–Willem Wegener) provided an evaluation of the results against the validation protocol rationales. A second workshop was held in Bergen, Norway in October 2007 (Anne O'Neill, UKEA) with contributions from the CASE 1 group, participating research laboratories and from the VALIDATION group. This meeting provided the opportunity to review the performance of the research methods against the validation protocols and to discuss the feedback and recommendations that would be made to the NORMAN work packages. The outcomes from this work are described in detail in this document.

3 Type 1 method: *In vitro* bioassays that detect oestrogenic activity

3.1 Module A: Test method definition, documentation and general requirements

1. External requirements
    a. Aim and task
    The bioassay selected for type 1 is being considered as a method for the *in vitro* detection of oestrogenic activity.

    b. Requirements and specifications
    This is achieved by the quantification of oestrogenic activity in an environmental sample based upon the proliferative effect of oestrogens on their target cells. The assay compares cell numbers achieved by similar inocula of human breast cancer oestrogen-sensitive MCF-7 cells of those exposed in environmental samples with ones in the absence of oestrogens (negative controls) and those in the presence of 17 ß –oestradiol (E2) (standards) (Soto and Sonnenschein, 1995). The methods standard curve ranges from 0.0272 to 272 ng/L but it is recommended that the analyst work with dilutions that fall within the linear range of 0.272 to 2.72 ng/L. This is achieved by running samples in a series of dilutions from 0.1x to 30 fold concentrated from original environmental samples. There appear to be no set requirements for the correlation coefficient (R2) of the standard curve or published single lab-validation or inter-lab validation studies. The method has undergone some validation assessment by Soto and Sonnenschein, (1995) where various xenobiotics were screened for oestrogenic activity, cumulative effects were assessed and the reliability of this assay was compared to other animal bioassays used to measure oestrogenicity.

2. Title of the method
    E-Screen (human mammary carcinoma cell proliferation assay) applied to rivers, ponds, wetlands and municipal wastewater effluent.

3. Beginning and end of validation procedure
    E2, EE2, 4-Nonyl phenol and cholesterol spiked samples alongside blanks and treated effluent samples were sent to the participants by the UKEA in February 2007. Results for the four validation stages were received by the UKEA in June 2007. Owing to some unexplained differences in some of the stage results by the methods laboratory compared to those expected by the lead laboratory, additional 20 ng E2 spikes were distributed alongside positive effluent samples for stage 3 in November 2007. Data from the re-runs were received in December 2007.

4. Responsible party
    Nancy Shappell (shappeln@fargo.ars.usda.gov)

USDA-ARS Animal Metabolism Unit, Bio-sciences Research Laboratory, 1065 Albrecht Boulevard, Fargo, North Dakota 58105, USA.

5. Scientific basis of the method

The ability of oestrogens to induce the proliferation of cells of the female genital tract enables an accurate prediction of oestrogenic activity on the basis of chemical structure (Soto and Sonnenschein, 1995). The E-screen assay was developed by Soto and Sonnenschein, (1995) to assess the oestrogenicity of environmental chemicals using the proliferative effect of oestrogens on their target cells as an endpoint. The method is advantageous in that the assay can detect oestrogenic activity in environmental samples without any information as to the chemical structure of the compounds present (Shappell, 2006). However, it is unable to discriminate environmental oestrogens from endogenous ones (Soto and Sonnenschein, 1995). The E-screen assay has been validated to some extent by Soto and Sonnenschein (1995), who identified several chemicals which had not been previously identified as chemicals with oestrogenic properties.

6. Method definition

   a. Method description / SOP

      The method used for the Type 3 validation is that described in Shappell (2006).

   b. Experimental setup

      In brief, the assay allows for the determination of oestrogenic activity of environmental chemicals using the proliferation effect of oestrogens on their target cells as an end point. Oestrogen sensitive MCF-7 BOS cells derived from a human mammary carcinoma were cultured in Dulbecco's modified Eagle's medium with sodium bicarbonate supplemented with 5% heat-inactivated defined foetal bovine serum. Cells were harvested with trypsin and EDTA phosphate buffered saline before counting. Cells were plated at 5 x 10^3 cells $well^{-1}$ in 200 µl DMEM with 5 % HI-FBS. One day later, media was removed and replaced with sample extract suspended in white media (DMEM with 10% CD-FBS charcoal-dextran stripped, 1m*M* sodium pyruvate, 4 m*M* L-glutamine, 15m*M* HEPES and 100 U ml^{-1} penicillin). For this validation study the sample extracts were diluted 1:100 with the CD-FBS and then further diluted to produce a range of dilutions (1:5, 1:10, 1:50, 1:100, 1:500, 1:1000 and 1:2000). A standard dose response curve was run alongside using E2 and samples of a 1:10 dilution up to 1:10000 dilution were typically within the standard range. Proliferation was tested by incubating sample with E2 as a screen for toxicity which was indicated when proliferation was less than the sum of the sample response plus E2 and reported as a percentage of expected E2 response (100% = no toxicity). The E2 receptor dependence of the proliferation response was evaluated by co-incubation with 5 x 10^{-9} M of E2-receptor antagonist. Cells were then incubated for 5 days, then fixed with trichloroacetic acid and the cell protein was stained with sulforhodamine B. Plates were read at A_{490} and E2 Equivalents extrapolated from the standard curve.

c. Sample preparation and pre-treatment
Sample preparation for the E-screen method is achieved through solid phase extraction. This method is described in more detail in Shappell, (2006). Initially, all glassware was washed and solvent rinsed and heated for 4 hours at 450^{o}C. Water samples were brought to room temperature, shaken and any large particulates allowed to settle for 1 hour before transfer of the 250 ml to a glass separatory funnel. Samples were not filtered to allow fine particles to be included in the extraction. To maximise retention of all organic compounds, not just those with chemical properties resembling oestrogens an Oasis Hydrophilic-Lipophilic Balance Sorbent glass solid-phase exchange cartridge (HLB, 200mg packing; Waters Milford, MA) were used. Cartridges were activated using a solvent series to elute hydrophilic and hydrophobic compounds (two 3 ml rinses of acetone, methanol, acetonitrile, ethyl acetate, methylene chloride, *tert*-butyl methyl ether and nanopure water). Samples were loaded onto the cartridges and vacuum applied (26-56 cm Hg) and the retained material eluted using 3 ml of the aforementioned solvents. The eluted material was then taken to dryness under N_2 at 37^{o}C. Sample extracts were redissolved first with 80 µl of nanopure water to aid resuspension, followed by the addition of 3 ml ethanol. This was transferred to a sterile vial and dried again using the conditions described above. For this validation study, the final sample resuspension was 160 µl of white media (see section 3.1 (6b) for description), and so was 1562.5 times the original concentration.

d. Sample measurement
The samples and standards were measured using a microplate reader with the capability of reading absorbance at a wavelength of 490nm.

e. Endpoint measurement
The endpoint was determined as the concentration (µg/L) of E2 equivalent within the sample extrapolated from the standard curve. The original dilutions during sample extraction and during the E-Screen assay must be included when calculating the E2 equivalents in the whole sample.

7. Requirements on devices, reagents, organisms, experimental conditions
   a. Instruments/devices
   A microplate reader was required which is capable of reading a wavelength of 490nm. Moreover, equipment required for the sample extraction phase includes a glass separatory funnel, an Oasis Hydrophilic-Lipophilic Balance Sorbent glass solid-phase exchange cartridge and a humidity chamber for cell culture.

   b. Environmental conditions
   Environmental samples must be brought up to room temperature before extraction. The assay was performed at room temperature, however, during cell culture, cells must be maintained at 37^{o}C, 6% CO_2 to 94% air under saturating humidity.

c. Test organisms
This type of validation method does not involve any test organisms as it is an *in vitro* bioassay for detecting oestrogenic activity. The method does however, use cell cultures of oestrogen-sensitive cells derived from a human mammary carcinoma as described in section 3.1(6b).

d. Reagents
Reagents required for sample extraction include acetone, methanol, acetonitrile, ethyl acetate and methylene chloride and *tert*-butyl methyl ether of HPLC or GC grade. For the cell culture, oestrogen sensitive MCF-7 BOS cells derived from a human mammary carcinoma are required which the method laboratory obtained from the laboratory of Dr. Ana Soto and Dr. Carlos Sonnenschein, Tufts University School of Medicine, Boston, MA. Furthermore, Dulbecco's modified Eagle's medium, (DMEM, MP Biomedical, Irvine, CA), 5% heat-inactivated defined foetal bovine serum (HI-FBS, HyClone, Logan, UT), sodium bicarbonate, trypsin, glucose, and EDTA phosphate buffered saline (all purchased from Sigma Aldrich) were required for the cell culture. DMEM and HI-FBS were required for the E-Screen Assay itself alongside sodium pyruvate and L-glutamine (Invitrogen, Carlsbad, CA), HEPES, penicillin, E2-receptor antagonist ICI 182,780 (Tocris, Ellisville), trichloroacetic acid, and sulforhodamine B (SRB).

e. Medium / matrix
The E-screen method has been used to evaluate oestrogenic activity in a number of environmental matrices including rivers, ponds and wetlands influenced by various agricultural practices and municipal wastewater effluents (Shappell *et al.* 2007; Shappell, 2006). It can also be used for other environmental water samples e.g. effluent samples, but has not been validated to any degree for such applications. There are no specific requirements of the physico-chemical parameters of the matrix but samples must be initially extracted by solid phase extraction in order to maximise retention of all organic compounds (see section 3.1(6c)).

3.2 Module B: Applicability domain and pre-validation

1. Target parameters
The method endpoint is the concentration of E2 equivalent in the sample after dilution has been accounted for. This is determined by calculating the sample E2 concentration using the absorbance intensity extrapolated from the standard curve.

2. Matrix and samples
   a. Type of matrix
   The matrix in which the validation study was performed was both spiked reverse osmosis filtered water samples and treated sewage effluent samples. To ensure all effluent samples processed by the methods laboratory reflected

as near as possible the levels of oestrogenicity at which the lead laboratory had intended to distribute, samples were preserved by freezing (section 3.2(2d)).

b. Sampling

For the stage 1 assessment, the lead laboratory prepared 6 independent stock solutions of 17 ß –oestradiol (Sigma, Poole, UK) at 100 mg/L in methanol (Analysis Grade >99.9% pure, Fisher, UK). Only three different batches of 17 ß –oestradiol were used to prepare the 6 independent stocks as the lead laboratory was limited by the availability of the chemical. From each stock, a 100 µg/L dilution in methanol was produced. The standards (100, 50, 10, 5, 1, and 0.5 ng of 17 ß –oestradiol) were prepared from the six 100 µg/L stocks. The methods laboratory requested for the spikes to be suitable for a 250 ml sample and so were prepared as follows: a 0.25 ml, and 0.125 ml aliquot was taken from each 100 µg/L stock to produce standards of 100 and 50 ng of 17 ß –oestradiol respectively. A further 1 ml aliquot was taken from each 100 µg/L stock and added to 9 ml methanol to make a 10 ng/ml solution from which 0.25 ml was taken for the 10 ng/L standard. Additionally, 0.1ml of the each of the six 100 µg/L stocks were taken and added to 9 ml methanol to make a 1 ng/ml stock from which the 1 ng/L standards were produced by preparing aliquots of 0.25 ml. All vials were sealed and numbered (1-36), only the lead laboratory knew the relationship of the sample numbers to the concentrations in the vials.

For the chemical analysis, 6 secondary stocks of 100 ng/L (standards 1-6) from the 6 independent stocks were sent for analysis. Furthermore, E2 spiked samples of 0.5, 1, 5, 10, 50 and 100 ng/L (samples 1-6) were taken from standard 3 and sent alongside control blanks (see section 3.3.1 method).

For the Stage 2 assessment, the lead laboratory prepared 12 measured 1 litre volumes of reverse osmosis filtered water in glass bottles and 12 vials containing 100 µl methanol. The bottles and vials were numbered sequentially and labelled as ‘blanks’. Two of the methanol vials were spiked with 1.0 ng/L of E2 and two with 10 ng/L cholesterol (Fisher Scientific, UK), with only the leading laboratory knowing the relationship of the sample numbers to the content of the vials. For this stage, the UKEA also distributed 8 glass bottles containing negative effluent to the methods laboratory. This effluent was from a treated sewage effluent source which had been left for some time for the E2 to degrade.

The lead laboratory for Stage 3 obtained 12 litres of sewage effluent (positive samples) from a source considered to be of moderate oestrogenic activity (approx. 5-10 ng/L E2 equivalent). This was pooled and mixed thoroughly and then divided into 12 measured 1 Litre volumes in plastic bottles. These were frozen and sent alongside 6 vials containing 100 µl methanol spiked with 20 ng/L of 17 ß –oestradiol. Originally the lead laboratory proposed to add a 25 ng/L spike but this was impractical due to the graduation of the pipettes

available at the time. Only the lead laboratory knew the concentration of the spiked vials.

For Stage 4, the lead laboratory prepared 6 vials of 100 µl methanol spiked with 10 ng/l ethynyl-estradiol and 6 vials of 100µl methanol spiked with 10 µg/l 4-nonyl-phenol. They were sent to the methods laboratory alongside 12 measured 1 litre volumes of reverse osmosis filtered water in glass bottles. The spiked vials were identified, but only the lead laboratory knew the concentration of the spikes.

During the preparation of the samples for the validation study the following health and safety considerations were taken into account. Stocks and spike solutions were prepared in a fume cabinet and safety glasses and protective clothing were worn. This was to avoid exposure to 17 ß –oestradiol, ethynyl-oestradiol, nonyl-phenol, cholesterol and the carrier, methanol. For stages 2 and 3, the environmental samples containing treated sewage effluent at unidentified levels of oestrogenicity were treated as toxic. This meant both the lead and method laboratory handled the samples in the fume cabinet and wore safety glasses and protective clothing. Samples were disposed of using the correct disposal route.

c. Sample characteristics
The 17 ß –oestradiol, ethynyl-estradiol, nonyl-phenol and cholesterol spikes used in stages 1 to 4 were prepared by the lead laboratory and were of a known purity (E2 ≥ 98%; EE2 ≥ 98%, NP ≥ 98%; Cholesterol > 95%). E2, EE2 and NP were purchased from Sigma, Poole UK whilst cholesterol and the methanol carrier was obtained from Fisher Scientific, UK. The methanol was >99.9% pure. The treated sewage effluent samples used in stages 2 and 3 were acquired from the BEMO study (Environment Agency, 2007), where the final effluent steam had been sampled from a sewage treatment works in the UK.

d. Sample stability and preservation, including transport
Stage 1 samples were stored at 4^{o}C prior to sending. The 17 ß –oestradiol in methanol vials were sent in cool boxes to the participant alongside 36 measured 250 ml volumes of reverse osmosis water in glass bottles instead of 1 litre volumes to reduce shipping costs. In order to simulate the effects of transporting samples aboard, the samples sent for chemical analysis in the UK were kept in cool boxes for 3 days before dispatch.

Stage 2 samples were again stored at 4^{o}C prior to sending. The 12 measured 1 litre volumes of reverse osmosis filtered water in glass bottles was sent in cool boxes alongside the glass vials containing either E2 in methanol, cholesterol in methanol or methanol alone. The negative effluent in glass bottles was not fixed in any way as the lead laboratory wanted it to degraded by the time it arrived at the methods laboratory.

All 12 measured 1 litre samples of positive effluent in plastic bottles were frozen by the lead laboratory before sending to avoid biodegradation of the sewage. The frozen samples were sent in cool boxes with cool packs and were still frozen on arrival at the methods laboratory. The methods laboratory had requested the lead laboratory used HDPE (high density polyethylene) bottles. The lead laboratory on this occasion had been unable to do this, so the oestrogenic effects of the plastic bottles on the samples is unclear as discussed in section 3.3.2. The six glass vials containing the E2 spikes in methanol for stage 3 were sent alongside the effluent samples in cool boxes.

All spiked methanol samples (either EE2 or NP) and corresponding volumes of reverse osmosis water were prepared in glass bottles and stored at 4^{o}C prior to sending. The samples were sent in cool boxes to the participant.

e. Availability of the organisms
This is not applicable for this method as it involves the direct detection of oestrogenic activity.

3.3 Module C: Intra-laboratory performance

3.3.1 Stage 1: Assessment of Accuracy, Precision, Sensitivity, Linearity and Range

Method

The lead laboratory prepared the E2 spiked samples as outlined in section 3.2(2b), which were sent alongside the water samples for dilution. The methods laboratory stored the samples at 4^{o}C until they were required for analysis. Immediately prior to analysis the methods laboratory added a sample vial to each bottle of water ensuring that the vial was thoroughly rinsed into the bottle. The validation method instructed that no further dilution of any single concentration would be required. The samples were then prepared and pre-treated as outlined in section 3.1(6c). As the UKEA also sent spare spiked vials, the methods laboratory ran the 100 ng/L pre-extraction samples through LCMS-MS though this was not part of the validation procedure. Again the spiked vial was washed with 250 ml of water, taken to dryness and resuspended in 50 µl of 1:1 ACN H_2O containing 20 pg/µl of d4E2 as an internal standard. This standard was outside the standard curve so it was diluted 1:5 with 1:1 ACN H_2O containing 20 pg/µl of d4E2.

To provide an indication of the concentration of E2 present in the spiked samples, aliquots of the six standards (100 ng/L) from the six independent stocks and E2 spiked samples (0.5, 1, 5, 10, 50 and 100 ng/L) from standard 3 alongside control blanks were sent for chemical analysis to the National Laboratory Service (part of the UKEA). The method used was LC-MS/TOF (Environment Agency, 2008), where the aqueous sample was filtered and then spiked with deuterated steroid internal standards and extracted using a solid phase extraction cartridge. Steroids were then desorbed with ethyl acetate and the extract cleaned

up using normal phase chromatography followed by a further clean up using gel permeation chromatography (GPC). The cleaned extract was analysed by high performance liquid chromatography using negative ion atmospheric photo-ionisation mode with mass spectrometer time of flight (LC-MS/TOF) detection. The method has been shown to be suitable for soluble steroid oestrogens (estrone, 17 ß –oestradiol and 17 α-ethynyl oestradiol) in waste water and surface water samples. Stocks and standards were stored by the UKEA for 3 days at 4^{o}C in the dark to simulate the effects of the samples that were dispatched to methods laboratory types 1 and 3.

Rationale

This procedure provides a blinded study generating data that assesses recovery and therefore accuracy of 17 ß –oestradiol from 6 independent stock sources. It also allows the assessment of precision at 6 different concentrations by means of repeat analysis from the same source. To be considered acceptable at this level of validation, the mean recovery from each standard must be +/- 30% and the CV from each concentration must be <50%

The concentration range has been chosen to reflect the ability of the method to detect 17 ß oestradiol at and below the UK proposed PNEC (1 ng/L), while remaining above the limit of detection of the chemical analysis (0.3 ng/L) at the lowest end. The highest concentration distributed was similar to that which can be measured in highly oestrogenic effluents in the UK. Therefore the method should be able to reliably detect the lowest concentration to show sufficient sensitivity, and the highest concentration to demonstrate sufficient range to be considered validated at this level.

The results from each of the 6 stock solutions allows the construction of calibration curves with calculated margins of error over the range provided. These should adhere to a recognised curvilinear model to be considered validated at this level.

Results

Table 3.1 provides the levels of E2 in each of the spiked samples alongside the % CV and % recovery rates. The recovery rates are based upon the nominal spiked E2 values which were verified by chemical analysis. The accuracy was calculated as the average value of six recoveries. The precision was calculated as the coefficient of variance of the E2 concentration from the 6 samples per spike. The range was established as the spike concentration range for which the acceptable criteria was met.

It was decided during the workshop in Bergen that nominal chemical spikes would be used to calculate recovery rates as the chemistry analysis was only obtained for one spike range from standard 3 rather than for each of the six stocks so n=1. This was predominantly due to the cost of chemical analysis if all spiked samples had been analysed. All chemistry results were within 75 % to 104 % of the nominal values (see Table 3.2).

Table 3.1. E2 equivalents for Type 1 spiked water samples and recovery rates.

Spiked E2 concentration ng/L	Sample no.	Result ng/l	Average ng/l	StDev	CV%	Nominal E2 Chemistry ng/l	Recovery %	Mean % recovery	StDev recovery
0.5	5	0.21	0.25	0.03	10.49	0.50	42.00	50.33	5.28
	12	0.26					52.00		
	20	0.26					52.00		
	25	0.25					50.00		
	26	0.24					48.00		
	31	0.29					58.00		
1	3	0.46	0.49	0.07	13.29	1.00	46.00	49.00	6.51
	9	0.42					42.00		
	15	0.53					53.00		
	18	0.42					42.00		
	22	0.54					54.00		
	29	0.57					57.00		
5	1	2.39	2.58	0.20	7.76	5.00	47.80	51.50	4.00
	2	2.70					54.00		
	6	2.46					49.20		
	11	2.45					49.00		
	32	2.53					50.60		
	33	2.92					58.40		
10	10	4.85	4.96	0.64	12.98	10.00	48.50	49.63	6.44
	13	4.52					45.20		
	16	4.65					46.50		
	19	5.01					50.10		
	21	6.22					62.20		
	30	4.53					45.30		
50	4	27.8	24.82	2.05	8.25	50.00	55.60	49.63	4.10
	7	24.4					48.80		
	17	22.4					44.80		
	23	25.4					50.80		
	24	22.8					45.60		
	27	26.1					52.20		
100	8	42.8	49.23	3.49	7.09	100.00	42.80	49.23	3.49
	14	51.8					51.80		
	28	49.4					49.40		
	34	48.1					48.10		
	35	51.6					51.60		
	36	51.7					51.70		

Table 3.2. Chemistry analysis for standards and spiked samples for Type 1 and 3 methods

Standard / Sample no.	Expected E2 (ng/L)	17 beta E2 (ng/L)	% recovery
Standard 1	100	79	79
Standard 2	100	71	71
Standard 3	100	74	74
Standard 4	100	81	81
Standard 5	100	81	81
Standard 6	100	64	64
Control blank	Below detection limit	<0.3	
Sample 1	100	88.7	88.7
Sample 2	50	45.9	91.8
Sample 3	10	8.29	82.9
Sample 4	5	3.75	75
Sample 5	1	0.836	83.6
Sample 6	0.5	0.523	104.6

Figure 3.1 illustrates the recovery rates of the spiked samples against the expected concentrations. All mean recovery rates for 0.5-100 ng/L are all below the 70-130% recovery limits. The recovery rates using this method are well below the limits of the protocol rationale and do not correspond with the chemistry results obtained by the lead laboratory. It was discussed in the workshop in Bergen whether there had been an issue with the initial spiking of the samples by the lead laboratory or whether the method laboratory had further manipulated the spikes before analysis. The type 1 methods laboratory did run its own chemical analysis of E2 present in the samples alongside the E-screen assay using LCMS-MS. They reported for the theoretical 100 ng/L E2 samples concentrations of E2 ranging from 41.3 to 56.4 ng/L. We were unable to identify if an error had occurred as the same spikes had been sent to type 3 method laboratory and for chemical analysis who had both reported E2 values close to those expected. The methods laboratory could not identify a source of error so we have had to report our initial findings.

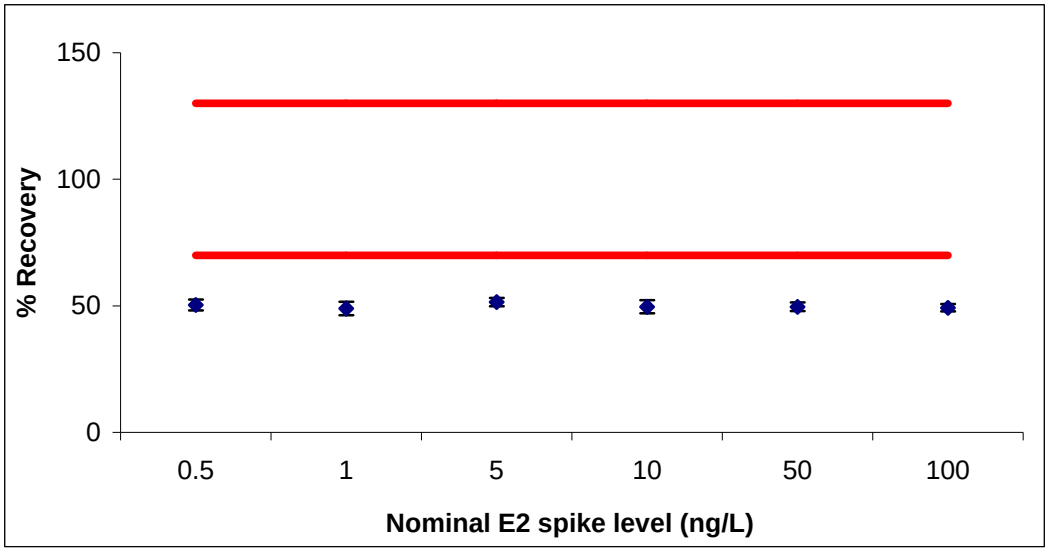

Fig. 3.1. Accuracy of Type 1 method measured by percentage recovery of the E2 spiked water samples. The horizontal lines indicate the recovery acceptance criteria (70-130%). The error bars indicate the standard error from the mean (n-6).

The type 1 method is shown to be acceptable for precision as at all the spiked concentrations the % CV is well below 50% (7.09-13.29 %) (Fig. 3.2).

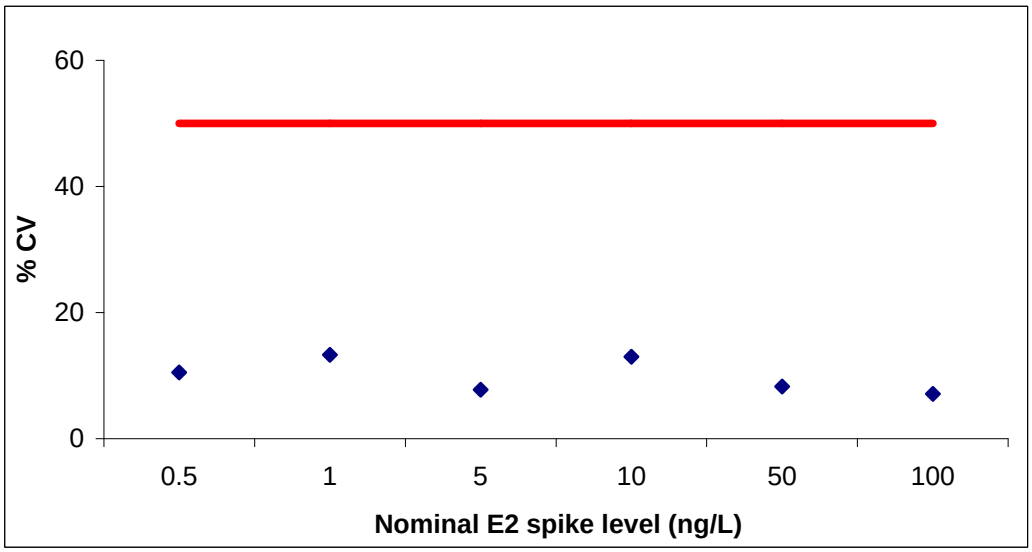

Fig. 3.2. Precision of Type 1 method in water samples spiked with E2 represented by coefficience of variance (% CV). The horizontal line indicates the precision acceptance criteria (<50%) (n=6).

Linearity was assessed by the construction of a calibration curve with calculated margins of error for the six analyses overall (Fig. 3.3) and as six separate analyses (Fig. 3.4a-f) in order to understand if there was any variation in curve shape.

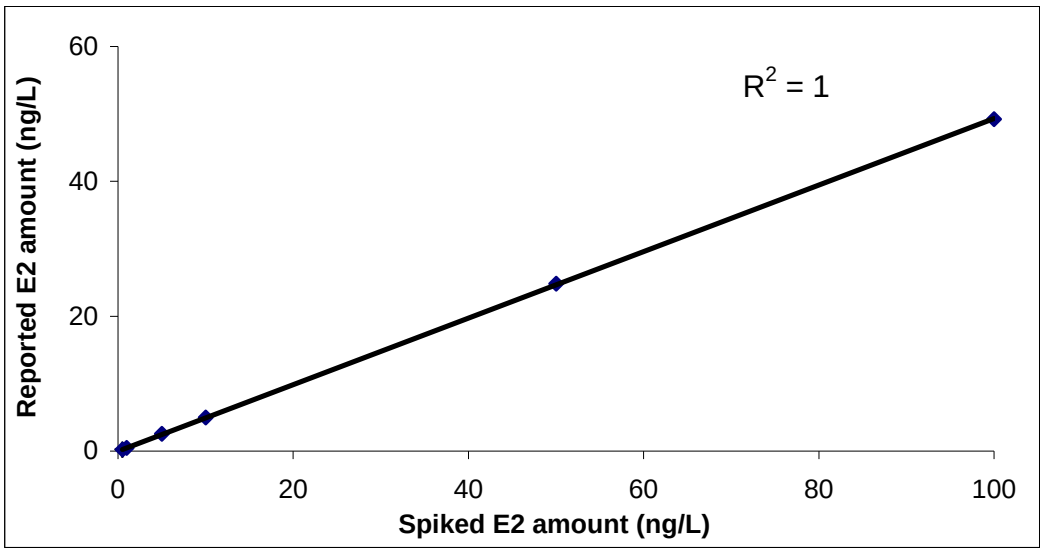

Fig. 3.3. Linearity of Type 1 method. Line represents the regression line through diamonds and forced through zero. The correlation coefficient (R2) indicates the proximity of the data to a perfect line. Error bars indicate the standard error from the mean (n=6).

When regression analysis was applied using a linear model to describe the relationship between the spiked E2 response and the reported E2 response, a statistically significant relationship at the 99% confidence level was reported (P = 0.0001).

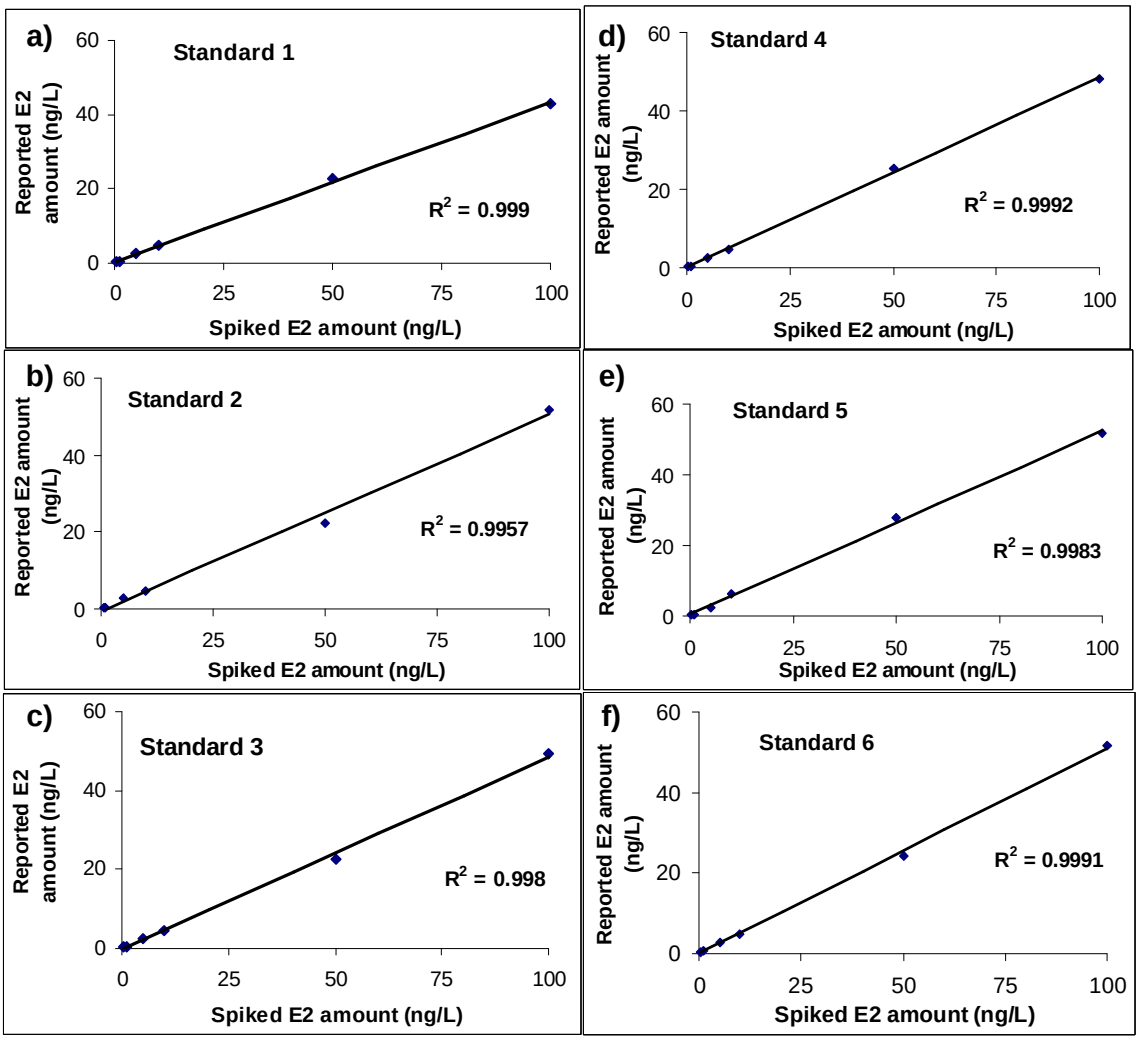

Fig. 3.4. Linearity of the individual standard analyses for Type 1 method. Line represents the regression line through diamonds and forced through zero. The correlation coefficient (R^2) indicates the proximity of the data to a perfect line.

The linear regression lines show little deviation between the six analyses with the R^2 ranging from 0.9957 to 0.9992. The validation protocol does not set a correlation coefficient above which the method would be considered validated, but as all R^2's are very close to 1 the method has shown good linearity.

Linearity for this method was also assessed by plotting the response (e.g. signal divided by the concentration) as a function of the concentration for each of the separate analyses (Fig. 3.5). The observed line should be horizontal and an ideal response factor of 1 for all concentrations if the nominal E2 spike and the measured E2 were of similar values. The linear range of the method was taken as 95 to 105 % of the horizontal line result calculated from the mean of the six analyses. Figure 3.5 illustrates that the Type 1 method gave response factors for all six analyses well below the expected value of 1, and all analyses

showed some variability predominantly falling above or below the horizontal line response. Consequently, the method at the lower concentrations does not completely demonstrate sensitivity for this validation study.

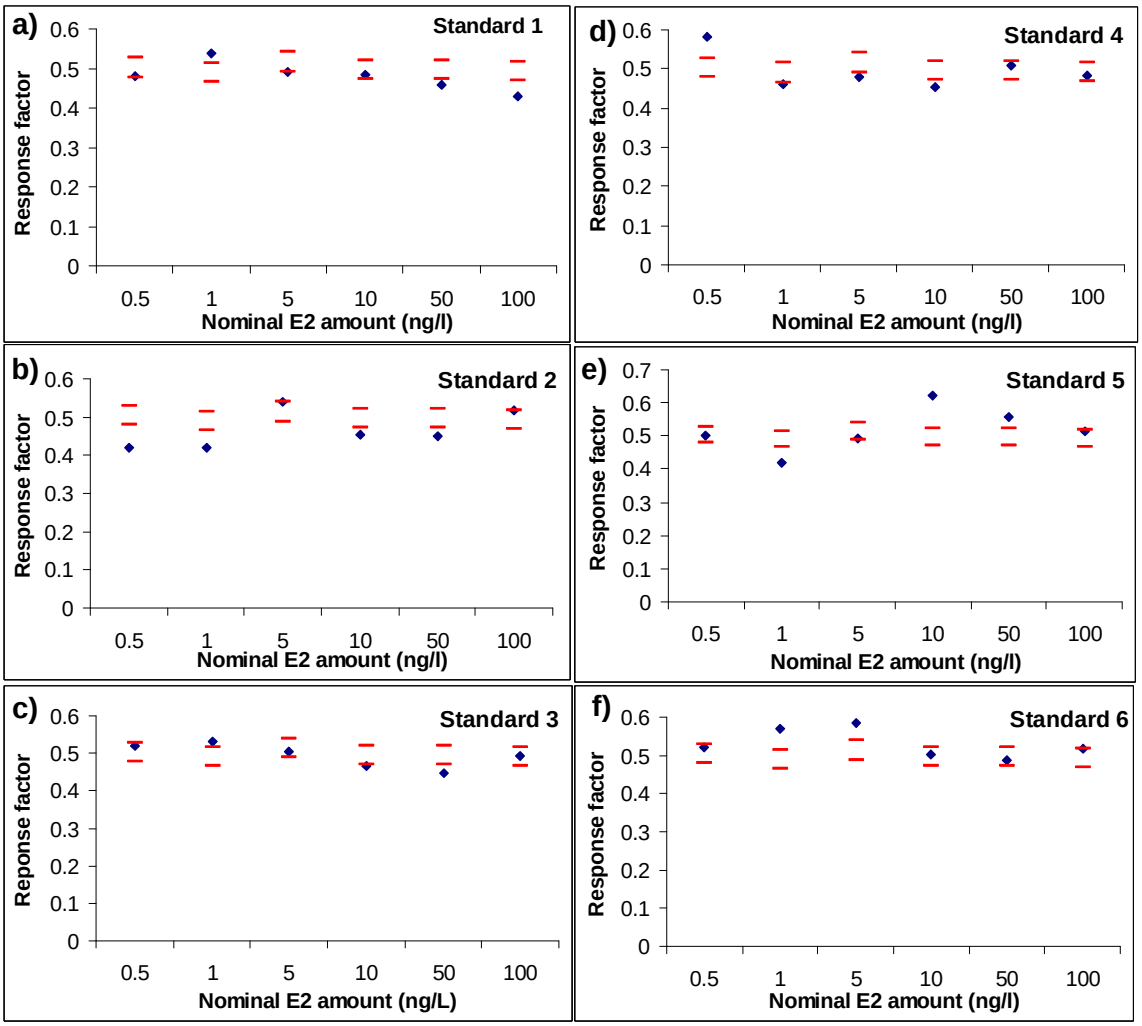

Fig. 3.5. Response factors of the type 1 method to E2 spiked water samples. The horizontal bars represent the response factor criteria which is 95% and 105% of the mean.

Though this method has shown good linearity and precision, the method on this occasion was unable to display accuracy in its recovery rate, though the reason for this has remained unexplained. Therefore, no acceptable working range can be formulated for the Type 1 method.

3.3.2 Stage 2: Assessment of Negative Response and Selectivity

Method

For this stage, the lead laboratory provided the methods laboratory with bottles and vials labelled as blanks prepared as outlined in section 3.2(2b). The methods laboratory stored the samples at 4^{o}C until they were required for analysis. Immediately prior to analysis, the methods validation laboratory added the relevant vial of methanol (some containing spikes of 1.0 ng/L E2 or 10 ng/L cholesterol) to the corresponding bottle of water, ensuring that the vial was thoroughly rinsed using water from that bottle. Cholesterol was utilised as a secondary spiking compound in order to assess the capability of the method to distinguish between E2 and other likely interfering compounds. Many compounds e.g. estrone and EE2, could not be used for this stage as the method assesses oestrogenicity of an environmental sample as a whole and therefore does not distinguish between specific oestrogenic compounds. Therefore, a non-oestrogenic organic compound was required, testosterone would have been more effective but for reasons of practicality cholesterol was selected. Additionally, the methods laboratory analysed the degraded treated sewage effluent (negative effluent) which was prepared as described in section 3.2(2b). These samples were run alongside the positive effluent samples in Stage 3 rather than separately at this stage. The samples were then extracted and run as per the methods laboratory protocol.

Rationale

This stage was conducted to provide a blind study to assess the ability of the method to distinguish between negative and positive samples. To be considered acceptable at this level of validation, the method must be able to identify all of the samples which were spiked with 17 ß –oestradiol. The ability of the methods laboratory to determine accurately the concentration of E2 and cholesterol spikes is not relevant for this stage and consequently additional samples were not sent for chemical analysis.

A negative response to those samples spiked with cholesterol will demonstrate the selectivity of the method against similar compounds and therefore all other samples apart from those spiked with E2 should return a result below the limit of detection of the method to be considered validated at this level. A negative result to the treated sewage effluent (negative effluent) will demonstrate the selectivity of the method in environmental matrices and should return a result below those obtained for similar samples in stage 3 (positive effluent) to be considered validated at this level.

Results

The results show that the method was able to distinguish between the samples spiked with E2 (samples 5 and 7) and those spiked with cholesterol (Table 3.3 and Figure 3.6). However, some of the blanks reported E2 levels at or above those which were spiked with E2 on two occasions. This effect is unexplained as in this instance all samples were sent in glass bottles unlike for stage 3 where the methods laboratory identified that blank samples which had been transported in plastic bottles contained some oestrogenic activity. We can

only assume that there was either some oestrogenic contamination of the blanks during their preparation by the lead laboratory, though a similar problem was not identified for the stage 2 blanks for the type 3 method or that they were contaminated by the methods laboratory. The E2 equivalents for the E2 spikes were greatly above the methods limit of detection of 0.064 ng/L.

For the negative effluent, E2 equivalents were reported above the limit of detection for all samples (0.295-0.558 ng/L). These concentrations were similar to those reported for the sample samples using the Type 3 method and therefore show that the effluent samples had not degraded as much as had been expected by the lead laboratory. In this instance this was not considered a problem for the validation study as the concentrations were well below those reported by the type 1 and 3 methods laboratories for the positive effluents in stage 3. The significance of this negative effluent concentration in comparison to those of the positive effluent is discussed in section 3.3.3.

Table 3.3. E2 equivalents in spiked water samples and negative effluents for Type 1 method

Sample	Sample no.	E2 Result ng/l	E2 Average ng/l	StDev	CV%
Blanks	1	<0.064	0.61	1.32	46.21
	2	<0.064			
	3	0.177			
	6	3.84			
	9	0.55			
	10	0.005			
	11	<0.064			
	12	0.104			
E2	5	0.52	0.53	0.01	2.54
	7	0.539			
Cholesterol	4	<0.043			
	8	<0.064			
Negative effluent		0.348	0.39	0.10	26.27
		0.558			
		0.295			
		0.297			
		0.403			
		0.468			

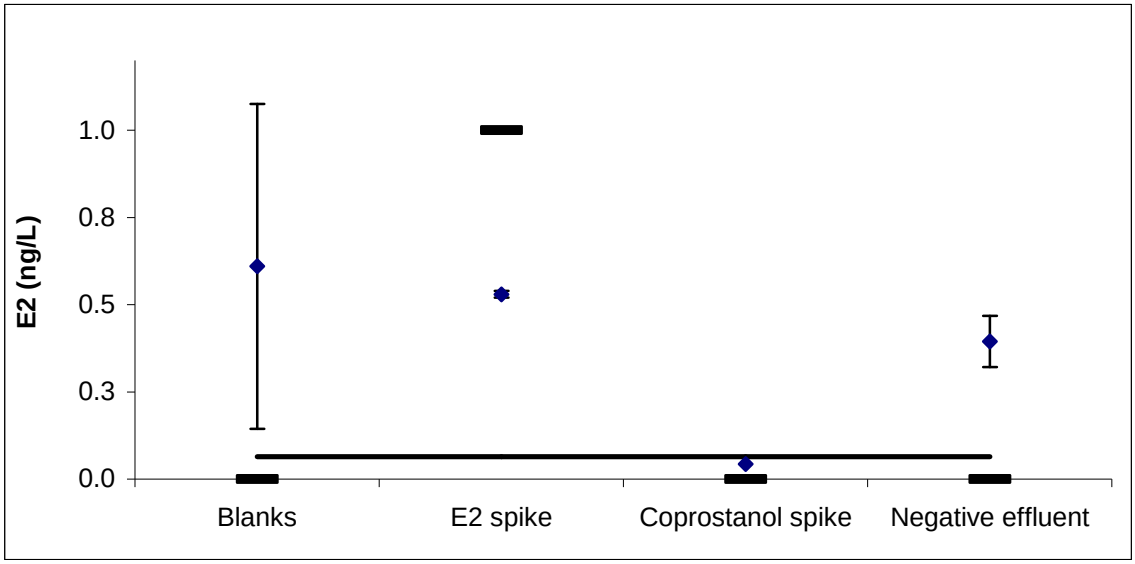

Fig. 3.6. Negative response of Type 3 method. Assays ability to discriminate between spiked and unspiked samples. The horizontal line indicates the methods limit of detection whilst the horizontal bars denote the level at which each sample was spiked with E2. Error bars indicate the standard error from the mean.

3.3.3 Stage 3: Assessment of Specificity in Environmental Matrices

Method

The lead laboratory provided 12 bottles of treated sewage effluent (positive sample) and six vials of methanol spiked with E2 as described in section 3.2(2b). It was advised that the preparation and analysis of this stage was performed immediately on receipt of the bottles as the oestrogenic activity of the treated sewage effluent was not stable. In this instance the methods laboratory had asked for the samples to be frozen before dispatch and thereby avoiding this issue. Immediately prior to analysis, the methods validation laboratory added one vial of the spike to six of the bottles of treated sewage effluent to create six samples of treated effluent sewage matrices that reliably contain 17 ß –oestradiol (spiked positive effluent).

Rationale

The study was performed to generate data to demonstrate the robust ability of the method to detect and discriminate between concentrations of 17 ß –oestradiol in environmental matrices. To be considered acceptable at this level of validation the COV from each sample must be <50% (n=6).

The accuracy of the result is not taken into account for the two environmental samples, as chemical analysis of the sample cannot be provided due to sample changes during shipping. However, the concentration of 17 ß –oestradiol spike can be inferred by the differences between the results of the spiked and unspiked environmental sample, demonstrating the

specificity of the method in test conditions. Therefore, to be considered acceptable at this level of validation, the mean recovery of the spike should be between 70-130% (n=6).

Results

The type 1 method has demonstrated that it is able to detect and discriminate between samples spiked with 17 ß –oestradiol in an environmental matrices. The coefficience of variance for the positive effluent sample replicates and the spiked positive effluent replicates was 7.9% and 9.4% respectively (Table 3.4 and Fig 3.7). The recovery of the spike (20 ng/L) was shown to be satisfactory when the six recoveries were calculated as a mean however, they ranged from 66.39 to 85.10 % and so one value would be considered as outside the acceptance criteria of 70-130 %.

As mentioned in section 3.1(3), the lead laboratory resent positive effluent samples and spikes to the methods laboratory in November 2007, as the methods laboratory had used different filtration methods for the sample extraction and it was thought that this would influence the E2 response for each sample. Consequently the methods laboratory only reported E2 concentrations for 5 samples for the positive effluent (samples frit and glass wool filtered pre extraction). For the positive effluents with spikes, three of the previous results were discounted and three new samples run so all samples were frit and glass wool filtered, then spiked with E2 prior to the column.

Table 3.4. Selectivity of Type 1 method in environmental matrices. Assays ability to discriminate between spiked and unspiked samples (n=6).

Sample	E2 Result ng/l	E2 Average ng/l	StDev	CV%	% Recovery	Mean % Recovery	StDev Recovery
Positive effluent	3.46	3.35	0.27	7.9			
	3.53						
	3.51						
	3.35						
	2.89						
Positive effluent (spiked E2)	17.0	17.77	1.68	9.44	72.81	76.12	7.19
	19.1				81.81		
	15.5				66.39		
	16.56				70.93		
	18.6				79.66		
	19.87				85.10		

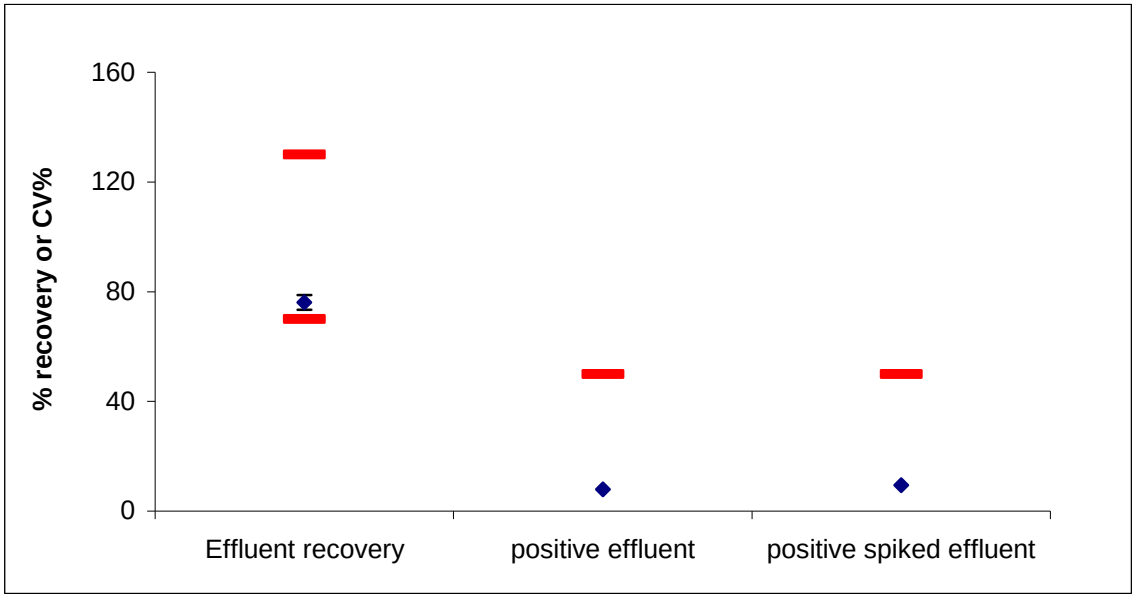

Fig. 3.7. Specificity of Type 3 method in environmental matrices measured by the recovery of the E2 spike and the precision of the environmental samples. The horizontal bars indicate the recovery acceptance criteria (70-130%) and the precision acceptance criteria (<50%) (n=6). Error bars indicate the standard error from the mean.

Comparison of the E2 responses in the negative effluent samples from stage 2 and the positive effluent samples in stage 3 demonstrates that the oestrogenicity of the negative effluent sample had degraded to some extent before being analysed. The E2 response for the negative effluent ranged from 0.295-0.558 ng/L, whilst the positive effluent E2 response ranged from 2.89-3.53 ng/L(Fig. 3.8).

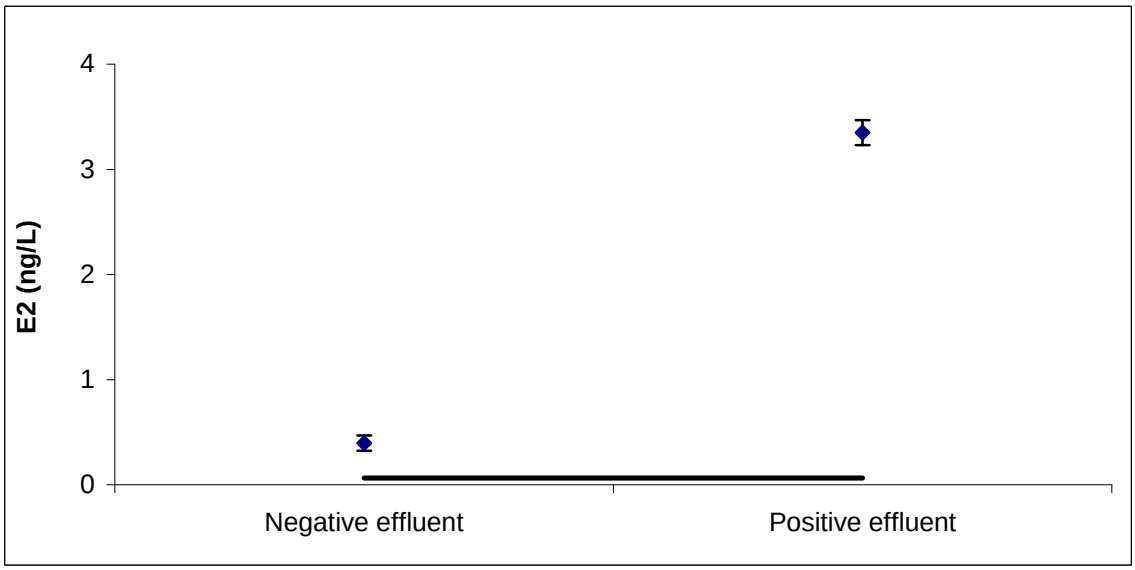

Fig. 3.8. Selectivity of Type 3 method in environmental matrices demonstrated by the E2 equivalents in the negative and positive effluent samples. The horizontal line indicates the limit of detection of the method. Error bars indicate the standard error from the mean.

3.3.4 Stage 4: Assessment of Relative Potency of Oestrogenic Compounds

Method

The methods laboratory was send vials containing methanol spiked with EE2 or nonylphenol and corresponding bottles of reverse osmosis filtered water as described in section 3.2(2b). Immediately prior to analysis, one vial of methanol was added to each bottle of water, ensuring that the vial was thoroughly rinsed into the bottle using only water from that bottle. Analysis of the sample, including extraction was instructed to take place immediately after mixing of the vial and water sample.

Rationale

Given the relative potencies of the oestrogenic compounds used, the method should detect the ethnyl-estradiol spiked samples with a relative value close to the top of the range of 17 ß –oestradiol used to generate the calibration curve. The method should detect the nonyl-phenol spiked samples with a relative value close to the bottom of the range of 17 ß – oestradiol used to generate the calibration curve.

To be considered acceptable at this level of validation, the mean result of the samples spiked with 10 ng/l ethynyl-estradiol (n=6) must not be lower than the mean result for those spiked with 10 ng/l 17 ß –oestradiol (n=6) as the potency of ethynl-estradiol is at least x10 greater. In a similar manner, the mean result of the samples spiked with 10 µg/l nonyl-phenol (n=6) must not exceed the mean result for those spiked with 10 ng/l 17 ß –oestradiol (n=6), as the potency of nonyl-phenol is at least x10,000 less.

Results

The type 1 method has demonstrated that it is able to return E2 equivalents for EE2 which are above those for E2 whilst NP returns E2 equivalents below those for E2 (Fig. 3.9). The acceptance criteria for relative potency are therefore met for the type 1 method. The difference between ethynyl-oestradiol and oestradiol is small and is not statistically significant amongst the medians at the 95% confidence level (Kruskal-Wallis non-parametric ANOVA *P* = 0.056604). The rationale for this method was that the difference in potency would be at least a factor of 10. However, for the E-screen method it has previously been reported that EE2 and E2 have the same relative proliferate potency (RPP) (Soto and Sonnenschein, 1995). That is the ratio between the minimal concentration of oestrodiol needed for maximal yield and the minimal dose of the xenoestrogen needed to produce a similar effect. Therefore, the results gained for the E-screen method in this validation study are as expected. In the same study, 4-nonylphenol was shown to be approximately 27,000 times less potent than E2 for the E- screen. In this validation study it was only shown to be only 5 times less potent. It must be noted that the RPP's can vary depending on the isomer used and the grade of the chemical (Soto and Sonnenschein, 1995). It is also unclear the exact

spike that the lead laboratory sent for EE2 and NP as this was not verified by chemical analysis.

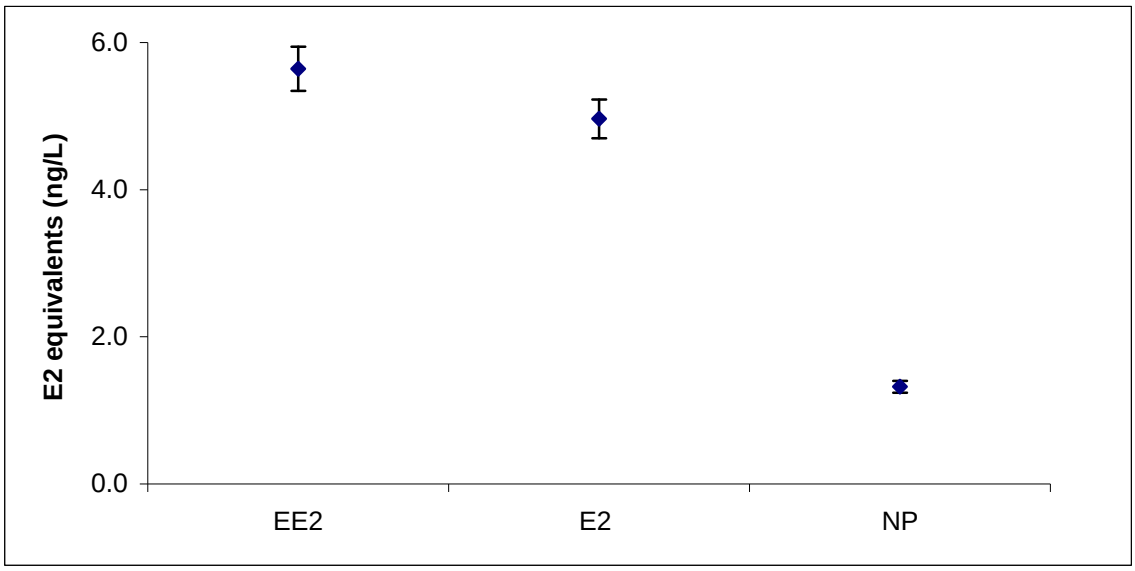

Fig. 3.9. Oestrogenic potency of type 1 method. Error bars indicate the standard error of the mean (n=6).

It must be noted that this method was not really fit for purpose for assessing the relative potencies of oestrogenic compounds as compounds other than (anti-)-estrogens have been reported to stimulate or inhibit cell growth, thereby over- or underestimating the estrogenic response (Murk *et al.*, 2002). Therefore, there is a need to look again at the criteria for method selection for this stage. To identify the relative potencies of specific oestrogenic compounds a method such as a reporter gene effects assay based on an estrogenic response (e.g. YES assay or ER-CALUX) which have previously been identified as possibilities to quantify estrogenic potencies of complex environmental matrices may prove valuable (Murk *et al.* 2002).

4 Type 2 method: *In vivo* bioassays that detect oestrogenic activity

4.1 Module A: Test method definition, documentation and general requirements

1. External requirements
    a. Aim and task
    The bioassay selected for type 2 is being considered as a method for the *in vivo* detection of oestrogenic activity

    b. Requirements and specifications
    For this to be achieved, the egg yolk precursor vitellogenin (Vtg) is detected in the blood and tissue samples of the fathead minnow (*Pimephales promelas*). Vtg normally occurs in sexually active female oviparous fish, but can be induced to occur in males in response to oestrogenic substances (EPA, 2003). The ELISA (enzyme-linked immunosorbent assay) technique is sensitive laboratory method widely used to detect and quantify antigens or antibodies in a variety of biological samples (Goksøyr *et al.*, 2003). The assay has a standard curve working range of 0.1 to 25 ng fathead minnow Vtg/ml, Data points with NBS (Non-specific binding) corrected absorbance lower the 0.010 are not included in the working range. The correlation coefficient (R_2) must be higher than 0.990 otherwise data points must be excluded. Only sample dilutions with absorbance values that fall within the standard curve working range should be used. Previous measurements of uncertainty during the validation of the method (Eidem *et al.*, 2006) demonstrated the assay has a between day repeatability precision of 9.9 % CV (single lab-validation) and a between-lab reproducibility precision of 18.6% CV (inter-lab validation).

2. Title of the method
    Direct homologous quantitative sandwich monoclonal ELISA for fathead minnow vitellogenin in blood plasma

3. Beginning and end of validation procedure
    Frozen plasma samples were sent to the participants by the UKEA in May 2007. Some of the samples had thawed on arrival and a further batch was sent in June 2007 to assess the effects of thawing on the Vtg responses. Results for the three validation stages were received by the UKEA at the end of June 2007. Results for the comparison of Vtg responses for the frozen and thawed samples were received in December 2007.

4. Responsible party
    Anders Goksøyr (anders.goksoyr@mbi.uib.no)
    Biosense Laboratories AS, Thormohlengst 55, 5008 Bergen, Norway.

5. Scientific basis of the method

   The measurement of Vtg has become an accepted routine screening test for oestrogenic and anti-androgenic effects of endocrine disrupting chemicals in fish (EPA, 2003; Environment Agency, 2007). This Enzyme-Linked Immunosorbent Assay (ELISA) has been specifically developed for the fathead minnow, though the methods validation laboratory has produced ELISA kits for Carp (Vtg standard for carp and fathead minnow), Medaka (Vtg standard specifically for Medaka) and Rainbow trout (Vtg standard specifically for Rainbow trout) (www.biosense.com). The fathead minnow Vtg ELISA kit with a homologous antibody was developed through single and inter-lab validation studies (Eidem *et al.*, 2006) and compared to existing methods (Biosense Carp Vtg ELISA and Competitor Fathead minnow ELISA kit) (Eidem *et al.*, 2005).

6. Method definition

   a. Method description / SOP

      The method used for the Type 2 validation is described in the protocol produced by Biosense Laboratories AS 'Fathead Minnow Vitellogenin ELISA Kit, Product No. V01018401'. This can be obtained from Biosense Laboratories AS, Thormohlengst 55, 5008 Bergen, Norway.

   b. Experimental setup

      The ELISA utilises the specific binding between antibodies and vitellogenin to quantify Vtg in plasma samples from fathead minnow. The test kit microplate wells are pre-coated with a specific Capture antibody that binds to the Vtg present in the standards and samples added to the wells. A different Vtg-specific detecting antibody, labelled with the enzyme horseradish peroxidise (HRP), is added to create a sandwich of Vtg and antibody. The enzyme activity is determined by adding the substrate 3,3',5,5'-tetramethyl-benzidine (TMB) that gives a colouration, whose intensity is directly proportional to the amount of Vtg present.

   c. Sample preparation and pre-treatment

      As compounds in the sample matrix may interfere non-specifically with the assay, usually leading to an underestimation of Vtg at low level sample dilutions, the minimum dilutions to avoid these matrix effects are 1:50 for plasma and 1:100 for whole body homogenates. As Vtg levels in experimental studies can vary widely, the samples require at least three different dilutions in order to fall within the linear part of the standard curve. For plasma, the method recommends a 1:50, a 1:5000 and a 1:500000 dilution. For this validation study, plasma samples were initially diluted with buffer by 1:1815 for stage 1, and 1:1650 for stage 2. For stage 3, samples were analysed in two separate rounds which was necessary as many of the samples in analysis 1 were not sufficiently diluted to get absorbance values that were within the standard curve working range. Plasma samples were diluted 1:50, 1:5000 and 1:500 000 for analysis 1 and 1:5000, 1:500 000 and 1:50 000 000 for analysis 2.

d. Sample measurement
The samples and standards were measured using a microplate reader with the capability of reading absorbance at a wavelength of 450nm.

e. Endpoint measurement
The endpoint was determined as the concentration of Vtg in the plasma samples (ng fathead minnow Vtg/ml). The mean absorbance value for the non specific binding (NSB) wells (unspecific background signal) is subtracted from the absorbance values for the standard and sample dilutions to give the NSB-corrected values. The standard curve is plotted, (absorbance values against standard Vtg concentrations) and the equation from the regression analysis used to calculate the Vtg concentrations in the original samples. These are then multiplied to take into the account the original dilution.

7. Requirements on devices, reagents, organisms, experimental conditions
   a. Instruments/devices
   A microplate reader was required, capable of reading a wavelength of 450nm. A microplate washer is also recommended though this procedure can be achieved manually.

   b. Environmental conditions
   As Vtg is an unstable molecule, all standards and sample dilutions should be prepared and stored on ice. Frozen samples must be thawed on ice and buffers and substrate solutions must be kept cold. Plates should be incubated at room temperature (20-25^{o}C), whilst the development stage must occur in the dark at room temperature.

   c. Test organisms
   The method has shown to be a sensitive biomarker for endocrine disrupting effects in male and female small fish species including the fathead minnow (*Pimephales promelas*), zebrafish *(Danio rerio*) and Japanese medaka (*Oryzias latipes*). These test species are ideal for reproductive toxicity tests in that they are a small size at maturity, have relatively short generation times, have asynchronous spawning and their overall ease of culture (Goksøyr *et al*., 2003). The validation method itself does not involve test organisms but they are required to provide the source of Vtg.

   d. Reagents
   The majority of reagents are provided as part of the Biosense test kit and includes a vial of dilution buffer (x5), PBS/Tween tablets, detecting antibody, concentrated 500x, TBS substrate and fathead minnow Vtg standard (purified, lypohilised Vtg from fathead minnow). Additionally, a 0.3M solution of H_2SO_4 is required to stop the reaction.

   e. Medium / matrix
   Vtg can be obtained from whole body homogenates, plasma samples, liver, ascites fluid and egg yolk (Goksøyr *et al*., 2003). For each source the state

and stability of the Vtg protein is different. The liver cell contains immature Vtg that has not undergone full post-transitional modifications, as well mature Vtg ready for secretion. The egg contains the lipovitellin form processed after uptake and the whole body homogenate contains a mixture of all of these (unless ovaries and/or liver have been removed), in addition to a high level of proteolytic activity that may act to degrade the protein during preparation.

4.2 Module B: Applicability domain and pre-validation

1. Target parameters
   The method endpoint is the concentration of the Vtg in the sample whether plasma or whole body homogenate after the dilution has been accounted for. This is determined by comparison of the sample Vtg concentration to the standard curve.

2. Matrix and samples
   a. Type of matrix
      The matrix in which the validation study was performed was in fathead minnow plasma. Plasma contains Vtg in its circulatory form at high concentrations though taking blood samples from small fish is problematic and can be contaminated by other body fluids leading to differences in composition between samples.

   b. Sampling
      In this validation study, the fathead minnow were sacrificed using a lethal dose (500 mg/L) of MS222 (3-aminobenzoic acid ethyl ester, methanesulfonate salt). Blood was collected by cardiac puncture, using a heparinised syringe (1000 Units heparin/mL), then centrifuged (7,000g; 5 min 15^{o}C) and the plasma removed (Environment Agency, 2007). Samples were pooled by the lead laboratory for stage 1 in order to be divided by the methods laboratory into 36 aliquots. The lead laboratory were unable to provide plasma samples for stage 2 as is detailed in section 4.3.2 (method).

      For stage 3, the lead laboratory provided 6 aliquots of plasma from male fish exposed to 18 ± 2.0 ng/L ethinyl-estradiol for 21 days (EE2 male). This time period was though sufficient enough to induce vitellogenin induction. A further 18 aliquots were provided that had been exposed to three effluents (6 aliquots per effluent) of varying degrees of oestrogenicity (High, Medium and Low males). The UKEA also sent 6 aliquots from female fish exposed to 18 ± 2.0ng/L EE2 (EE2 female) and 6 aliquots from female fish exposed to a highly oestrogenic effluent (High female). All samples were randomly numbered and labelled 1-36 so the method laboratory did not know the relationship of the sample numbers to the aliquots. For each sample type, the lead laboratory had initially pooled replicates and then sub-divided the samples into the six replicates. Additionally, the UKEA was expected to provide 6 aliquots of plasma from female fish that had not been exposed to any oestrogenic compound. Owing to the small number of plasma samples available from the

experimental study from which the samples were acquired, the unexposed females sent for stage 1 were used for analysis (unexposed female).

As all the plasma samples obtained for the validation study contained unknown quantities of oestrogens, all samples were treated as toxic environmental samples. This meant both the lead and method laboratory wore protective clothing and safety glasses and all samples were disposed of through the correct waste disposal route.

c. Sample characteristics
The Vtg fathead minnow plasma samples were obtained from a previous study (BEMO, 'Biological effects measures in fish – applications to treated sewage effluent') (Environment Agency, 2007). Male and female fish were exposed to treated sewage effluents from various sewage treatment works around the UK and some were exposed to EE2. Samples were chosen for this validation study based upon the Vtg concentrations reported in the BEMO study and assigned either as a high (1000,000 Vtg ng/ml), medium (1,000 Vtg ng/ml) or low (100 Vtg ng/ml) estrodial effluent.

d. Sample stability and preservation, including transport
To avoid degradation of Vtg in the samples, protease inhibitors can be used, samples kept on ice and preparation times reduced. In this validation study, the samples were stored at -80^{o}C at the lead laboratory. Samples were transferred to the methods validation laboratory on dry ice. Some samples had thawed on arrival so a further set was sent in order to assess the effects of thawing.

e. Availability of the organisms
For this validation method, plasma samples were obtained from a previous study and consequently there were no issues over the availability of the organisms. Fathead minnow are widely used in aquatic toxicity studies due to the relative ease of husbandry and the large number of offspring they produce making them an ideal species for such studies such as this. The fish for the BEMO study were either bred at the Brixham Environmental Laboratory, UK or at the University of Exeter, UK. Obtaining samples from a previous study for this validation method however provided other difficulties, including limiting the scope of the validation study for stages 1 and 2, as well as restricting the available amount of sample for stage 3 which affected the reproducibility of the method.

4.3 Module C: Intra-laboratory performance

4.3.1 Stage 1: Assessment of Accuracy, Precision, Linearity and Range

Method

The UKEA provided a pool of fish blood plasma with a low level of vitellogenin of sufficient volume to be divided into 36 aliquots of a size suitable for the assay when diluted with sample buffer. However, it was found that the Vtg concentration was 400 ng Vtg /ml plasma which would have meant the samples would have had to be diluted 1:25 000 in order to be suitable for the ELISA. Consequently, the method validation laboratory used there own plasma from another validation study which involved less dilution with sample buffer. All plasma samples were analysed at a total plasma dilution of 1:1815 to ensure the final concentration did not exceed the limit of detection (0.009 ng/ml). At this plasma dilution there was no detectable level of Vtg in the blank samples. The Vtg spiking of the plasma samples was achieved with three different batches of lyophilised Vtg. Analysis was conducted 6 times in total over three different days with two consecutive analyses each day. To ensure blind analysis, the samples were labelled and spiked by one analyst and analysed by another. Samples were spiked with a pure lyophilized vitellogenin standard over a concentration range expected to display the whole response of the assay after dilution. The spiked concentrations selected were 0.2, 0.5, 2.5, 6.25, and 12.5 ng/mL. A blank for all six analyses were run alongside the spiked samples. The samples were analysed using the Biosense FHM Vtg ELISA kit.

Rationale

This study allows the assessment of the accuracy of the assay's response to known concentrations of vitellogenin. This is calculated by subtracting the background levels in the plasma samples that have not been spiked with the lyophilized standard. Furthermore, it allows the assessment of precision at 6 different concentrations by means of repeat analysis (n=6 in each case). To be considered acceptable at this level of validation, the mean recovery at each concentration must be +/- 30% and the CV from each concentration must be <50%.

The results from each of the 6 analyses allows the construction of calculated calibration curves. These should adhere to a recognised curvilinear model to be considered validated at this level.

Results

The levels of Vtg in each spiked sample alongside the %CV and % recovery rates are displayed in Table 4.1. Vtg concentrations in the unspiked samples were below the detection limit (N.D = 0.009 ng/ml) for all six replicates. The accuracy was calculated as the average value of six recoveries. The precision was calculated as the coefficient of variance of the six recovery experiments. The range was established as the spike concentration range for which the acceptable criteria were met.

Table 4.1. Vtg spikes in plasma samples for Type 2 method and recovery rates.

Spike Vtg concentration (ng/ml)	Replicate	Vtg concentration (ng/ml)	Average Vtg (ng/ml)	StDev	%CV	% Recovery	Mean % recovery	StDev % recovery
0	1	N.D						
	2	N.D						
	3	N.D						
	4	N.D						
	5	N.D						
	6	N.D						
0.2	1	0.21	0.22	0.07	32.06	105.00	107.50	34.46
	2	0.20				100.00		
	3	0.12				60.00		
	4	0.33				165.00		
	5	0.24				120.00		
	6	0.19				95.00		
0.5	1	0.54	0.55	0.17	30.06	108.00	110.33	33.16
	2	0.54				108.00		
	3	0.33				66.00		
	4	0.79				158.00		
	5	0.68				136.00		
	6	0.43				86.00		
2.5	1	2.92	2.85	0.57	19.87	116.80	114.13	22.68
	2	2.98				119.20		
	3	1.88				75.20		
	4	3.49				139.60		
	5	3.26				130.40		
	6	2.59				103.60		
6.25	1	6.94	6.67	0.96	14.35	111.04	106.75	15.32
	2	7.08				113.28		
	3	4.86				77.76		
	4	7.19				115.04		
	5	7.53				120.48		
	6	6.43				102.88		
12.5	1	14.15	11.91	1.84	15.46	113.20	95.31	14.74
	2	12.17				97.36		
	3	9.18				73.44		
	4	11.95				95.60		
	5	13.50				108.00		
	6	10.53				84.24		

Figure 4.1 illustrates the recovery rates of the spiked samples against the expected concentrations. All mean recovery rates for 0.2-12.5 ng/ml are within the 80-100% recovery limits. Overall, the method appears to produce marginally higher recovery rates than the initial sample spikes but this is within the limits of the protocol rationale.

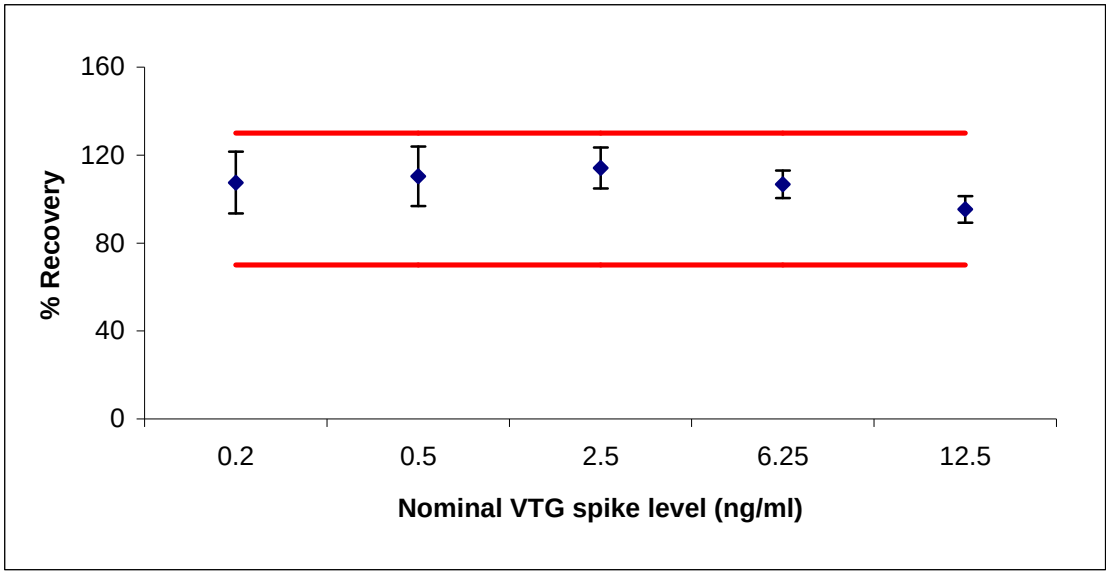

Fig. 4.1. Accuracy of Type 2 method measured by percentage recovery of the Vtg spiked plasma samples. The horizontal lines indicate the recovery acceptance criteria (70-130%). Error bars indicate the standard error from the mean (n=6).

This method is shown to be acceptable for precision as at all spiked concentrations the %CV is well below 50% (14.35-32.06%) (Fig. 4.2). As expected, the precision is shown to be improved at the higher spiked concentrations (2.5-12.5 ng/ml).

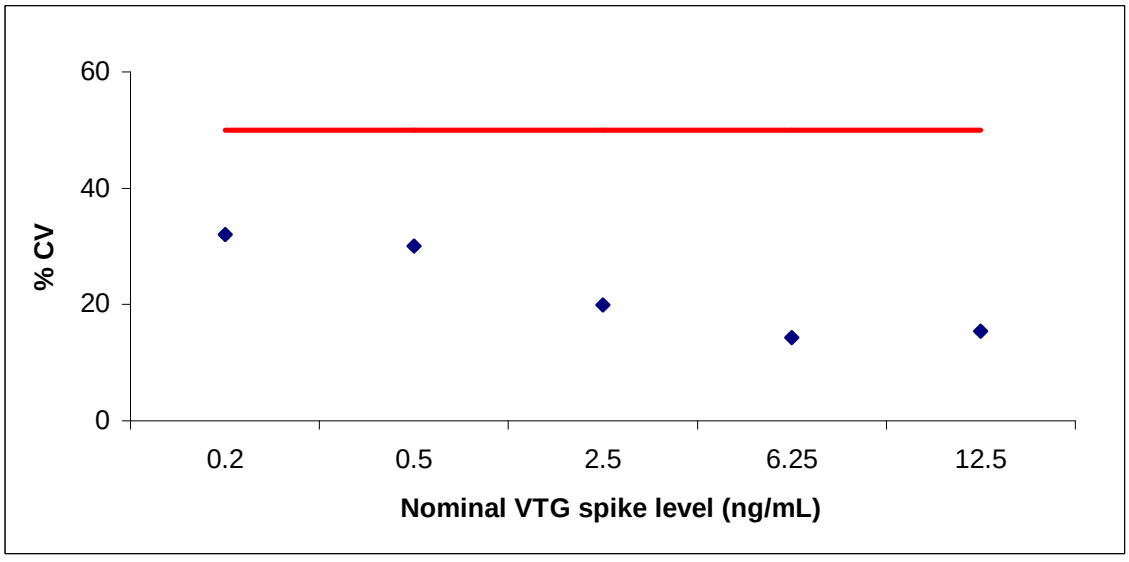

Fig. 4.2. Precision of Type 2 method in plasma samples spiked with Vtg represented by Coefficience of variance (CV %). The horizontal line indicates the precision acceptance criteria (<50%) (n=6).

Linearity was assessed not only by construction of a calibration curve with calculated margins of error for the six analyses overall (Fig. 4.3) but also as the six separate analyses (Figs. 4.4a-f) in order to understand if there is any variation in curve shape.

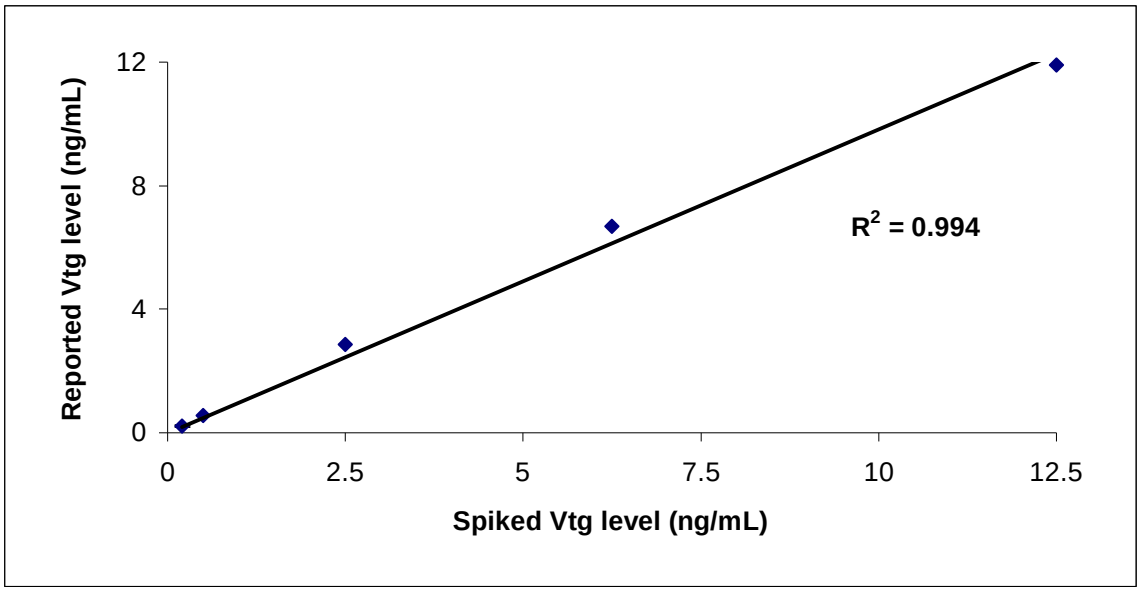

Fig. 4.3. Linearity of Type 2 method. Line represents the regression line through diamonds and forced through zero. The correlation coefficient (R^2) indicates the proximity of the data to a perfect line. Error bars indicate the standard error from the mean (n=6)

When regression analysis was applied using a linear model to describe the relationship between the spiked Vtg level and the reported Vtg level, a statistically significant relationship at the 99% confidence level was reported (P = 0.0001).

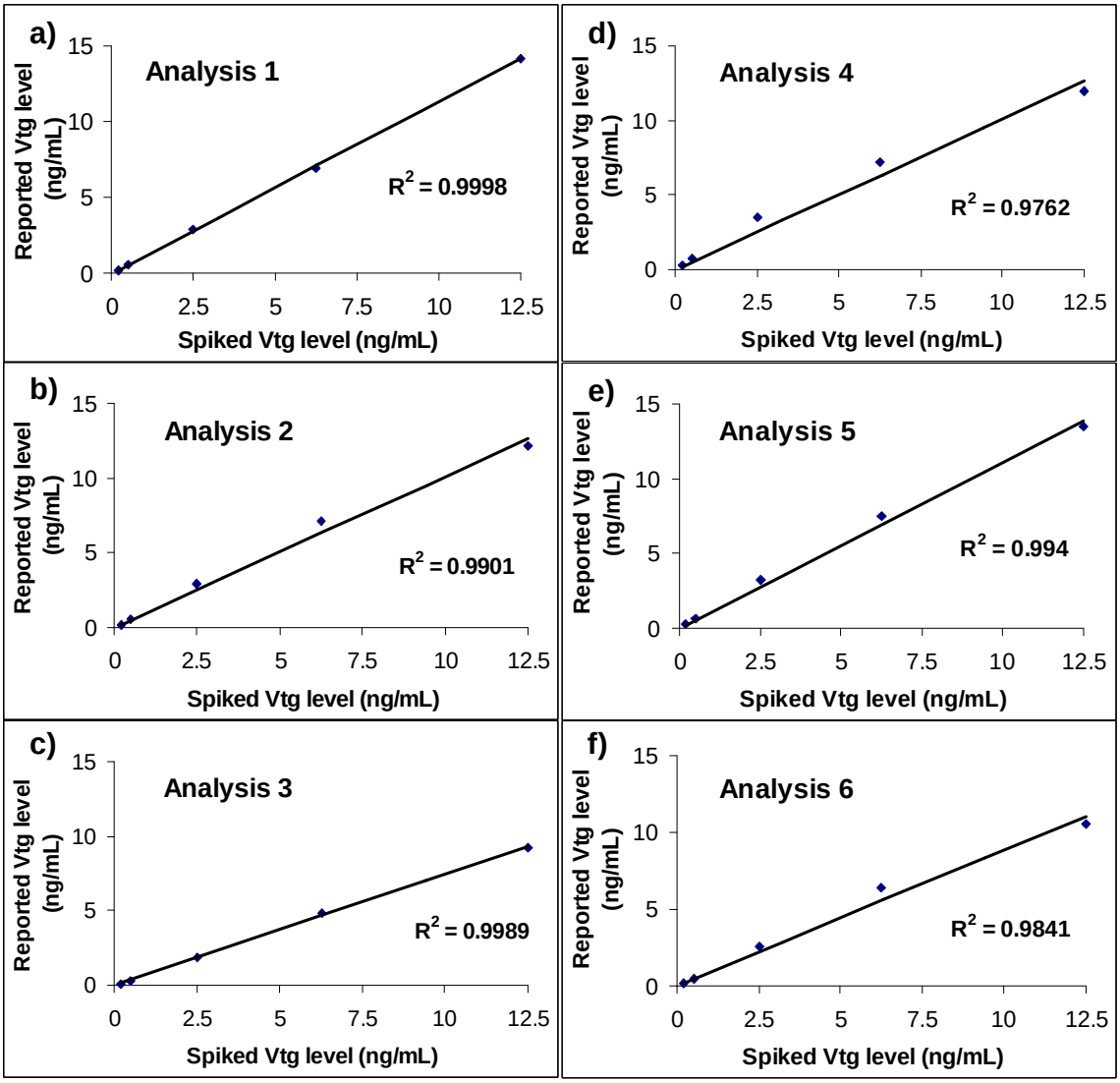

Fig. 4.4. Linearity of the individual standard analyses for Type 2 Method. Line represents the regression line through diamonds and forced through zero. The correlation coefficient (R^2) indicates the proximity of the data to a perfect line.

There were some deviations from the linear regression line for some of the six analyses, with the R^2 ranging from 0.9762 to 0.9998. The validation protocol rationale did not set a correlation coefficient above which the method would be considered validated. However, taking into the account the validation laboratories own criteria (R_2 > 0.990), two of the six analysis would be outside the methods limits. In this instance points would be omitted that deviate from the line.

Linearity for this method was also assessed by plotting the response (e.g. signal divided by the concentration) as a function of the concentration for each of the separate analyses (Fig. 4.5). The observed line should be horizontal and an ideal response factor of 1 for all concentrations if the nominal Vtg spike and the measured Vtg were of similar values. The

linear range of the method was taken as 95 to 105 % of the horizontal line result calculated from the mean of the six analyses. Figure 4.5 illustrates that though the Type 2 method gave responses for all six analyses around the response factor of 1, only analyses 1 and 2 produced results which fell near to the horizontal line response, thereby showing that though the method displayed a good linearity for 4 of the six analyses they have been shown to vary considerably from one analysis to the next.

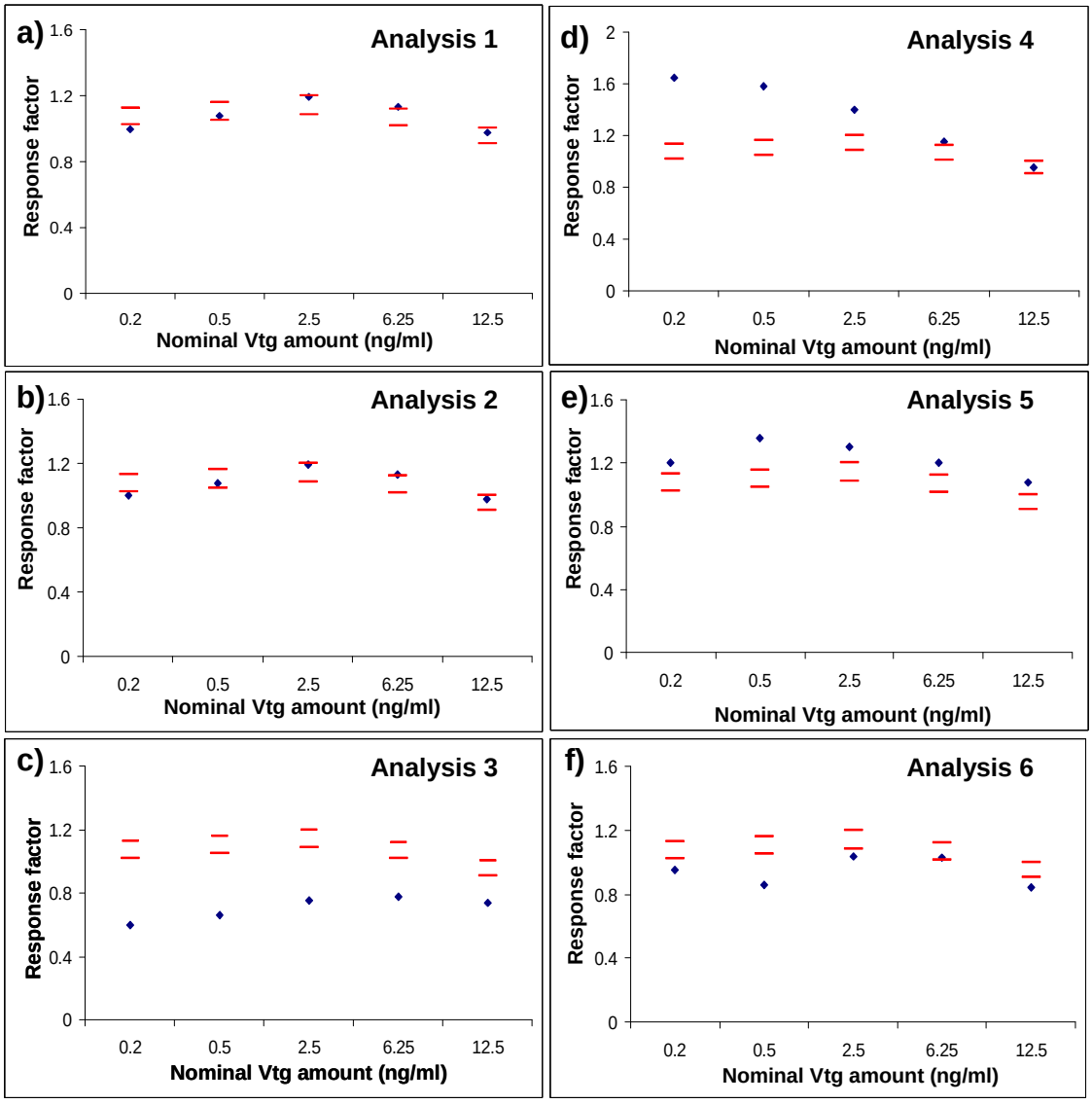

Fig. 4.5. Response factors of the type 2 method to E2 spiked water samples. The horizontal bars represent the response factor criteria which is 95% and 105% of the mean.

As this method has displayed accuracy in its recovery rate, good precision and linearity, an acceptable working range of 0.2 to 12.5 ng/mL can be formulated.

4.3.2 Stage 2: Assessment of Sensitivity, Negative Response and Selectivity

Method

For this stage, the UKEA were to provide the type 2 laboratory with twelve volumes of plasma from male fish that had not been exposed to oestrogens. The lead laboratory was to provide 6 samples of unexposed plasma, 3 samples of plasma from the same male fish that had been contaminated with an unrelated Fathead Minnow blood protein and 3 volumes of plasma from male fish with a low level of vitellogenin. Unfortunately, as all plasma samples were obtained from a previous experimental study, the UKEA were unable to do this. Instead, the validation laboratory ensured blind analysis by one analyst spiking 6 of the samples with a low level of vitellogenin (0.1 ng Vtg/mL) and another analyst performing the assay. All plasma samples were analysed at a 1:1650 dilution.

Rationale

The rationale behind this stage is to provide a study which performed blindly will generate data to assess the ability of the method to distinguish between negative and positive samples. To be considered acceptable, the method has to be able to discriminate between those samples spiked with vitellogenin at a low level and the unspiked samples. The accuracy of the result is not relevant for this study. The selectivity of the method against similar proteins could not be assessed on this occasion as during the first workshop in London a protein could not be identified for which the UKEA would be able to obtain samples. Furthermore, for this method you would not use an antibody that is cross referenced with anything else.

Results

The results show that the method was able to identify the six samples which had been spiked with low levels of vitellogenin (Table 4.2). The six samples which were unspiked all returned values below that of the methods limit of detection (0.0123 ng Vtg/mL).

Table 4.2. Negative response of Type 2 method. Assays ability to discriminate between spiked and unspiked samples (n=6).

Sample no.	Vtg (ng/ml)
1	0.1
2	0.1
3	0.1
4	<0.0123
5	<0.0123
6	<0.0123
7	0.1
8	0.1
9	<0.0123
10	<0.0123
11	0.1
12	<0.0123

Though on this occasion the assay could not be conducted independently of the method laboratory or in accordance to the original protocol, the method did display an ability to discriminate between samples spiked with low concentrations of vitellogenin and those which were unspiked.

4.3.3 Stage 3: Assessment of Sensitivity and Discriminative Ability

Method

The lead laboratory prepared the plasma samples for this stage as described in section 4.2(2b). In brief, the samples comprised of EE2 males (plasma from males exposed to EE2), high, medium and low males (plasma from males exposed to effluents of varying degrees of oestrogenicity), EE2 females (plasma from females exposed to EE2), high females (plasma from females exposed to an effluent of high oestrogenicity) and unexposed females (plasma from females not exposed to oestrogenic effluent). All aliquots were diluted as required by the method validation laboratory before analysis.

Although the samples were shipped on dry ice by the lead laboratory they had unfortunately defrosted upon arrival at the method validation laboratory. Vitellogenin samples are sensitive to repeated freezing and thawing and it is known to rapidly degrade when not frozen. As it was unknown what effect this would have on the thawed samples a further two pooled samples were sent that were known to have similar Vtg concentrations in the plasma as the EE2 male and female. These samples were still frozen on arrival and will be referred to as EE2 male frozen and EE2 female frozen.

Rationale

By conducting this stage, the study will generate data to demonstrate the robust ability of the method to detect and discriminate between the plasma of fish exposed to highly oestrogenic compounds and effluents and those not exposed. In the case of male fish, levels of vitellogenin close to maximal (1 mg/ml) should be observed in exposed fish, where none was observed in the plasma of unexposed fish in stage 2. This procedure will also demonstrate the ability of the method to rank plasma samples from male fish according to the degree of oestrogen exposure.

In order for the method to be considered acceptable at this level of validation, the CV for each treatment must be <50% (n=6). The mean result from the male fish exposed to 10 ng/L must be above 750 ng/mL and the mean results for the male fish exposed to the three effluents must fall in the expected increments (high>medium>low oestrogenicity). The mean results from the female fish exposed to 10 ng/L EE2 and to the highly oestrogenic effluent must be above that of the unexposed female fish.

Results

Samples were analysed in two separate rounds. This was necessary as many of the samples during the first analysis were not sufficiently diluted enough to get absorbance values that were within the standard curve working range. Unfortunately, this meant that the amount of sample available for the second analysis was limited (1 µl). Samples were diluted in dilution buffer by 1:50, 1:5000 and 1:500 000 in analysis 1 and 1:5000, 1:500 000 and 1:50 000000 in analysis 2 (Table 4.3). The limit of detection for this stage was 2.44 ng/ml and all samples were above the limit of detection.

Table 4.3. Vtg concentrations in the exposure groups for type 2 method

Sample	Sample ID	Analysis 1 (Vtg µg/ml)	Analysis 2 (Vtg µg/ml)	Mean (Vtg µg/ml)	Treatment mean	StDev	%CV
High male	2		200	200	2983	2372	79.50
	7		700	700			
	21		6000	6000			
	23		5000	5000			
	30		2000	2000			
	35		4000	4000			
Medium male	3	2000	4000	3000	2500	316	12.65
	9	2000	3000	2500			
	19	200	3000	2500			
	20	2000	3000	2500			
	26	2000	2000	2000			
	34	2000	3000	2500			
Low male	5	6	6	6	7	1	20.61
	8	5	5	5			
	10	6		6			
	14	9	7	8			
	29	9	8	8.5			
	32	6	**4000**	6			
High female	1		200	200	2550	1357	53.21
	4		1100	1100			
	12		1000	1000			
	24		4000	4000			
	28		5000	5000			
	36		4000	4000			
EE2 Male	6		44000	44000	39500	5683	14.39
	13		37000	37000			
	16		33000	33000			
	18		34000	34000			
	25		42000	42000			
	33		47000	47000			
EE2 female	11	**9**			38600	10455	27.08
	15		26000	26000			
	17		39000	39000			
	22		55000	55000			
	27		37000	37000			
	31		36000	36000			
Unexposed female				1900	1967	121	6.16
				1900			
				2100			
				2100			
				1800			
				2000			
EE2 male frozen				29441	25414	2498	9.83
				26876			
				23546			
				25991			
				22984			
				23648			
EE2 female frozen				13128	15143	2440	16.11
				15107			
				15781			
				11662			
				16817			
				18366			

In table 4.3, two Vtg concentrations are highlighted in bold (sample 11 and assay 2 for sample 32). Though these values have been reported they have been omitted from the mean Vtg concentrations as the assay was limited by the volume of plasma available for dilution. Samples 2 and 7 were identified as possible outliers for the High male but statistical analysis demonstrated there was not statistical difference between these values and the four other sample Vtg concentrations reported for High male (Dixon's Q Test: Sample 2 Q value = 0.310; Sample 7 Q value = 0.245 < Critical value of Q 0.717 P= 0.05).

For the type 2 method, the results for the male fish rank in order of the effluent oestrogenicity levels (high>medium>low) (Fig. 4.6), although the difference between the fish exposed to high and medium oestrogenic effluent is statistically not significant amongst the medians at the 95% confidence level (Kruskal-Wallis non-parametric ANOVA *P*= 0.934925). This is due to the variability in Vtg concentrations between the samples for High male. The male fish exposed the high and medium oestrogenicity effluents were predominantly higher than the rationale expectant maximal Vtg concentration of 1000 µg Vtg/mL. For the male fish exposed to EE2, the vitellogenin plasma concentrations (mean 39500 µg Vtg/mL), are considerably higher than the rationale lower limit of 0.75 µg Vtg/mL. This may in part be due to the EE2 exposure concentration to the fathead minnow being almost double of that expected in the rationale (18 ± 2.0ng/L EE2 instead of 10ng/L EE2).

As expected, the results for the ethynyl- estradiol exposed and the highly oestrogenic effluent exposed female fish are above that of the unexposed female fish (Fig. 4.6). However, there is not a statistically significant difference at the 95% confidence level between the medians for the High female and Unexposed female (Kruskal-Wallis ANOVA, *P*= 1.0). This is because there is some variability in the Vtg concentration for the different samples for High females. Sample 1 for the high female samples is not considered an outlier according to Dixon's Q test (Q value = 0.167 < Critical value of Q 0.621 P= 0.05) and was therefore included in the mean.

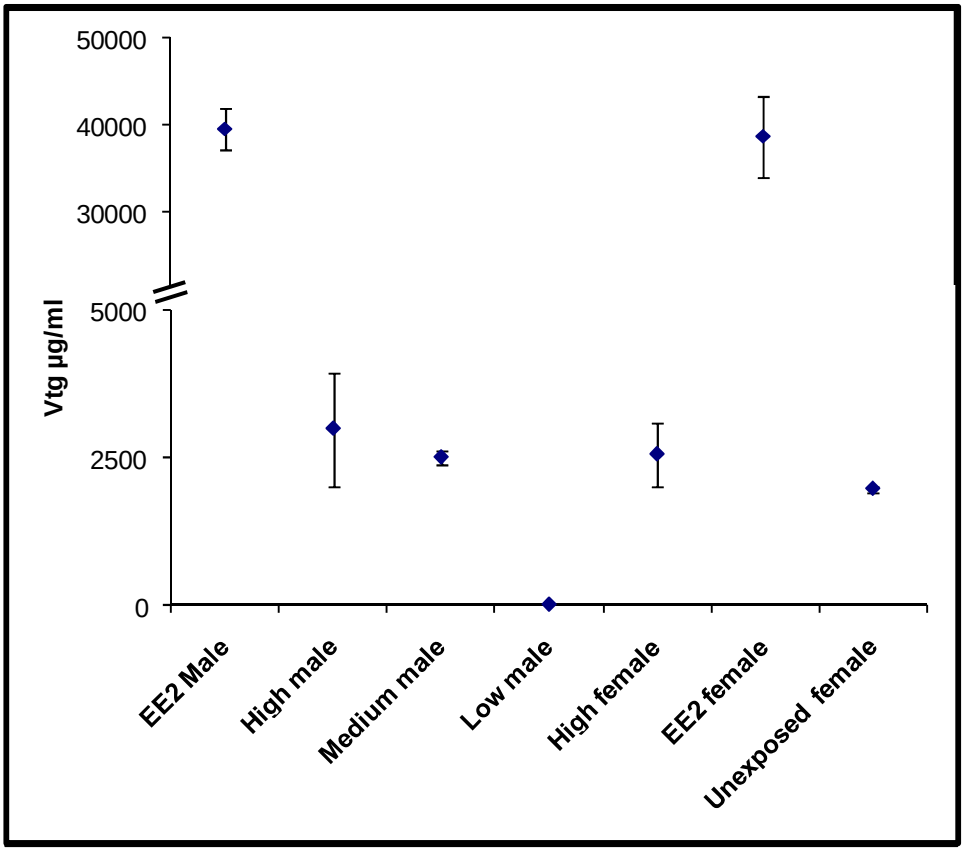

Fig. 4.6. Discriminative ability of Type 2 method. Vtg plasma concentrations in male and female fathead minnow from exposure groups of different levels of oestrogenic effluents and EE2. Error bars indicate the standard error from the mean (n=6).

According to the rationale, the CV for each sample group must be less than 50% (Fig. 4.7). In this instance High male and High female samples had CV's above the 50% limit (79.50 % and 53.21 % respectively). The elevated CV's are due to the variability of samples 2 and 7 for High male compared to the other four samples and sample 1 for High female compared to the other 5 samples. As the samples for stage 3 were pooled for each sample type and then sub-divided into the six samples this variability has to be accepted.

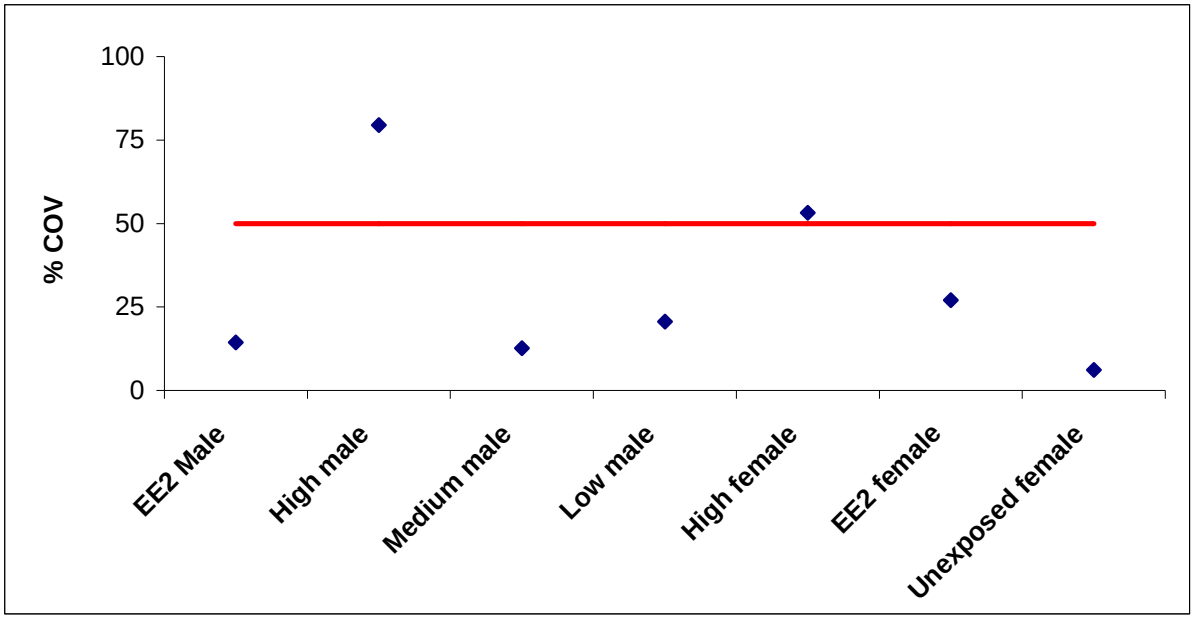

Fig. 4.7. Precision of Type 2 method in different exposure groups represented by Coefficience of variance (CV %). The horizontal line indicates the precision acceptance criteria (<50%) (n=6).

Comparison of the mean Vtg plasma concentrations of the pooled thawed and frozen samples for EE2 exposed males and females indicated that the thawing process in this instance had little effect of the Vtg plasma concentrations observed (Fig. 4.8). It was found that the thawed samples for both EE2 males and females were above those levels determined in the corresponding frozen male and female samples. It cannot be said for certain that both thawed and frozen samples initially had similar Vtg concentrations as the sample had to be taken from different exposure groups but as the thawed samples returned higher values the effects of thawing must be considered as minimal.

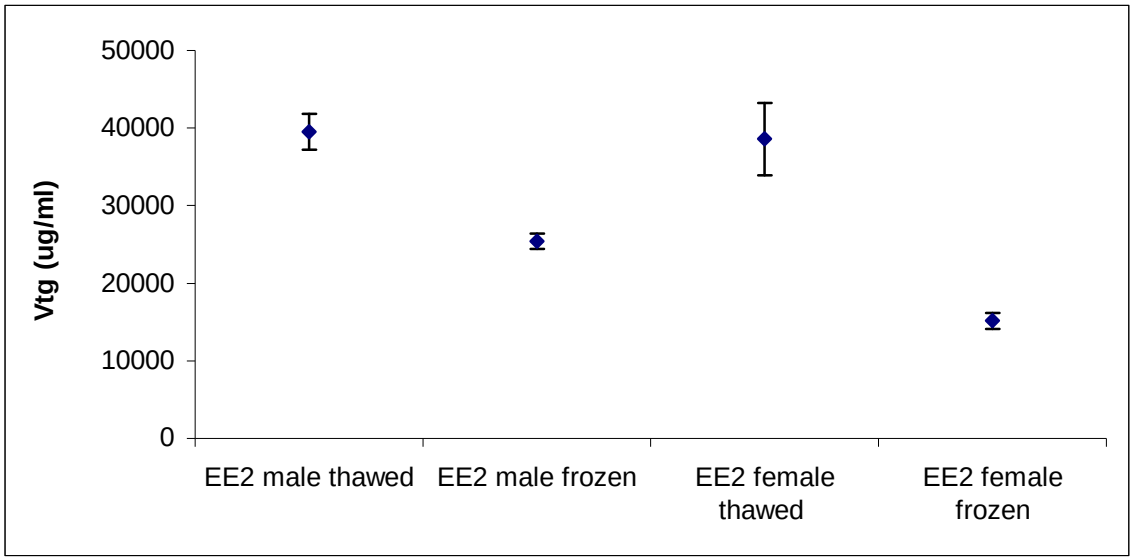

Fig. 4.8. Comparison of Vtg concentrations of thawed and frozen plasma samples (n=6)

For stage 3, the oestrogenically exposed and EE2 exposed plasma samples were obtained from a previous study funded by the UKEA based upon the US EPA Fathead Minnow Pair Breeding Assay that has been shown to exhibit reproductive inhibition using various known oestrogenic EDC's. The BEMO study, 'Biological effects measures in fish – applications to treated sewage effluent' (Environment Agency, 2007) used a carp ELISA (Tyler *et al.*, 1999) in order to analyse the Vtg blood plasma concentrations of the fathead minnow. The samples selected by the UKEA for analysis at stage 3 as effluents exhibiting high, medium and low oestrogenicity were based upon the ELISA results obtained from this previous study. Consequently, Table 4.4 provides a comparison of the mean Vtg plasma concentrations from each exposure group from the BEMO study and stage 3. Recoveries are calculated as the mean of each of the six Vtg concentrations for Stage 3 divided by the corresponding exposure Vtg mean from the BEMO study and reported as a percentage.

Table: 4.4. Comparison of stage 3 Vtg plasma concentrations with those from the BEMO study

Sample	Vtg (µg/ml) Biosense (fat head minnow ELISA) Type 2			Vtg (µg/ml) BEMO (carp ELISA)			Mean % recovery	StDev % recovery
	Total mean	StDev	%CV	Total mean	StDev	%CV		
High male	2983	2372	79.50	142.20	201.71	141.85	2098	1668
Medium male	2500	316	12.65	24.56	6.95	28.29	10179	1288
Low male	6.58	1.36	20.61	0.36	0.29	78.2	1829	377
High female	2550	1357	53.21	728.67	265.49	36.43	350	276
EE2 Male	39500	5683	14.39	423.40	39.87	9.42	9329	1342
EE2 female	38600	10455	27.08	353.25	273.99	77.56	10927	5187

The recovery rates for stage 3 are consistently high when based upon the Vtg plasma concentrations from the BEMO study. The methods validation laboratory met the criteria requirements for their standard curve run alongside the samples (R^2 = 0.9987 higher than the method criteria of 0.990) and therefore the results obtained for the samples must be accepted. The differences in measured Vtg plasma concentrations are not considered unusual in Vtg ELISA studies and often result from laboratories using different ELISA methods (EPA, 2003; Goksøyr, *et al.*, 2003; Jensen and Ankley, 2006). This ranges from exposure protocols, sampling procedures, Vtg analysis, data handling and quality assurance (Goksøyr, *et al.*, 2003). In this instance as the plasma samples were obtained from one study, the differences in plasma Vtg concentrations are most likely to be due to differences in data handling and the ELISA itself. Differences in the ELISA include numerous antibodies, binding antigens and standards (EPA, 2003; Jensen and Ankley, 2006). The methods validation laboratory used an ELISA kit based on a homologous fathead minnow antibody, whereas for the BEMO study utilised a CARP Vtg ELISA developed to quantify Vtg in fathead minnow (Tyler *et al.*, 1999). The latter uses monoclonal and polyclonal carp Vtg-specific antibodies, as well as carp Vtg standard and is considered heterologous for the fathead minnow though the antibodies show excellent cross-reactivity with fathead minnow Vtg. Therefore the methods using different Vitellogenin standards for quantification of fathead minnow Vtg is a factor likely to have had a considerable influence on Vtg plasma concentrations. Furthermore, it must be noted that the results from the BEMO study are based upon individual plasma samples whereas the samples for Stage 3 were pooled for each exposure group and then sub-divided into the six samples. In conclusion, though the absolute Vtg concentrations from the two studies differ due to factors such as Vtg standards, both studies have shown that they can differentiate between control and oestrogenically exposed groups of fathead minnow.

5 Type 3 method: *In vitro* direct measurement assays for measurement of quantities of oestrogenic target compounds

5.1 Module A: Test method definition, documentation and general requirements

1. External requirements
    a. Aim and task
    The direct measurement assay selected for type 3 is being considered as a method for the *in vitro* direct detection of 17 ß –oestradiol.

    b. Requirements and specifications
    This is achieved by the recognition of 17 ß –oestradiol (E2) by specific monoclonal antibodies. The E2 monoclonal antibody binds exclusively with E2 and does not show cross-reaction with other chemicals of similar structures. The assay has previously shown good reproducibility with the coefficient of variance largely under 10% (Abraxis, 2006) with uncertainty of the ELISA largely based upon pipetting error of the viscous antiserum solution and the slope of the calibration curve (Choi *et al.*, 2004). The assay has a standard curve working range of 0.05 to 1 µg/L. Only samples which fall within the standard curve working range should be reported. Samples below this range must be concentrated with solid phase extraction prior to analysing whilst those above the range must be diluted with 10 % methanol. There appear to be no set requirements for the correlation coefficient (R^2) of the standard curve or published single lab-validation or inter-lab validation studies for this specific test kit.

2. Title of the method
    Monoclonal ELISA detecting 17 ß-oestradiol in municipal wastewaters following SPE extraction.

3. Beginning and end of validation procedure
    E2 and E1 spiked samples alongside blanks and treated sewage effluent samples were sent to the participants by the UKEA in February 2007. Results for the three validation stages were received by the UKEA at the end of March 2007.

4. Responsible party
    Masato Hirobe
    Japan EnviroChemicals Ltd, 2-17-85 Juso-Honmachi Yodogawa-ku, Osaka 532-0024, Japan.

    Since this work, the Ecologiena® ELISA test kits are now manufactured by Tokiwa Chemical Co. Tokyo and distributed in Europe by Biosense Laboratories AS, Bergen, Norway.

5. Scientific basis of the method

   The competitive enzyme-linked immunosorbent assay (ELISA) have been widely adopted to quantify a small molecular antigen with a single epitope and enables detection of chemicals at very low levels without complex sample preparation. In this instance, it has been developed for the quantitative analysis of 17 ß-estradiol in water (groundwater, surface water and well water), and has not been validated for other applications. The methods laboratory have also previously provided kits for detecting the oestrogenic substances Estrone www.abraxiskits.com/moreinfo/PN590061USER.pdf (Japan EnviroChemicals Ltd, 2004a) and Ethynyl oestradiol www.abraxiskits.com/moreinfo/PN590051USER.pdf (Japan EnviroChemicals Ltd, 2004b). The E2 ELISA kit has been found not to display cross-reaction with other oestrogens and hormones with similar structures. The highest percentage cross-reactivity of the E2 ELISA test kit with another compound was 1.3% with estrone (Abraxis, 2006). Furthermore, comparison of the test kit against LC/MS/MS for E2 concentrations in primary and secondary effluent and aeration from an effluent tank displayed some similarities though all recoveries using the test kit were higher and ranged from 110-217 %. Four other test kits for E2 were run alongside the comparison with LC/MS/MS all reporting even higher concentrations of E2 in the effluent samples in comparison to the chemical analysis.

6. Method definition

   a. Method description / SOP

      The method used for the Type 3 validation is described in the protocol produced by Japan EnviroChemicals, Ltd, 'Ecologiena® 17ß –Estradiol (E2) ELISA kit (Microplate) User's Guide'. This can be obtained from www.abraxiskits.com/moreinfo/PN590062USER.pdf (Japan EnviroChemicals Ltd, 2004c).

   b. Experimental setup

      In brief, the test is based upon the recognition of E2 by specific monoclonal antibodies. The sample, containing E2, is premixed with an E2-enzyme conjugate solution and added to microplate wells, where they compete for a limited number of binding sites of specific antibodies immobilized on the surface of the wells. When the E2 concentration in the sample is higher relative to the E2-enzyme conjugate, the E2 will predominantly bind to the antibody, this is reversed when the E2-enzyme conjugate is more concentrated. Unbound E2 and excess E2-enzyme conjugates are then washed out and the presence of E2 is detected by the addition of a chromogenic solution (colour solution). The enzyme labelled E2 bound to the E2 antibody in the plate catalyses the conversion of the substrate to a coloured product. After the 30 minute incubation period, the reaction is stopped by the addition of a diluted acid (stop solution). The higher the E2 concentration in the sample, the less antigen-enzyme conjugate is bound to the antibody binding sites and therefore a lower absorbance is obtained.

c. Sample preparation and pre-treatment
For this method, pre-treatment for female steroid hormones was achieved using a Graphitized Carbon Black (GCB) Column. Raw water samples containing sediment or suspended matter were filtered through 1µl pore diameter glass fibre filters (Advantec Co., Toyo Roshi, Japan) using a vacuum pump. Prior to extraction, a 100-fold synthetic sewage concentrate (10 g glucose; 10 g polypeptone; 10 g KH_2PO_4; make up volume to 1 L with distilled water; adjusted to pH 7) was added to the samples at a ratio of 1:100. This is because graphitized carbon column usually gives lower recovery rates and synthetic sewage will improve the recovery rates during the solid phase extraction (SPE) step. The GCB Column (GL-Pak Carbograph 1,000 mg / 12 ml 20/PK, GL Science Ltd. Toyko, Japan) was then pre-conditioned. The GCB columns was rinsed in the following order: 10 ml dichloromethane (HPLC grade), 10 ml methanol, 20 ml distilled water (5 ml/min for each). For sample loading, adapter caps were attached to a GCB column and connected with a Teflon coated tube and loaded (5 ml/min). The columns were subsequently washed with 10 ml of 100% methanol (5 ml/min). For the elution of E2, 10 ml of a methanol : DCM solution (ratio 5:5) was eluted (2 ml/min). The eluant was collected in a centrifuge tube containing 10 µl of 1% BSA (0.1g BSA in 10 ml distilled water), added as a keeper substrate prior to elution. The solvent was evaporated with nitrogen gas, temperature controlled (40-50^{o}C) in order to accelerate evaporation. Samples were then reconstituted to 10% v/v methanol.

d. Sample measurement
The samples and standards were measured using a microplate reader with the capability of reading absorbance at a wavelength of 450 nm.

e. Endpoint measurement
The endpoint was determined as the concentration (µg/L) of E2 within the sample. Absorbance of the samples and standards must be read within 15 minutes of the reaction being stopped. A standard curve is prepared (E2 concentration against optical density plotted on Log-Log graph paper) based on at least duplicate standards. The assay must be performed within the range of 0.05µg/L and 1µg/L. Samples with a concentration above 1 µg/L must be diluted with 10% methanol and re-tested. Due to the extraction step, sample ELISA mean values (µg./L) must be divided by their concentration factor.

7. Requirements on devices, reagents, organisms, experimental conditions
   a. Instruments/devices
   A microplate reader is required that is capable of reading a wavelength of 450nm. Additionally, equipment is required for the solid phase extraction (pre-treatment e.g. vacuum manifold, diaphragm vacuum pump, N_2 purge equipment).

b. Environmental conditions
The ELISA kit reagents must be stored at 2-8^{o}C, but as the assay is performed at room temperature, the analyst must ensure all reagents have reached room temperature before proceeding.

c. Test organisms
This type of validation method does not involve test organisms as it is an *in vitro* direct measurement assay of oestrogenic compounds in wastewaters.

d. Reagents
The majority of reagents are provided as part of the Ecologiena® E2 ELISA kit and includes vials of the E2 standards (0, 0.05, 0.15, 0.4 and 1.0 µg/L), antigen-enzyme conjugate powder, buffer solution, wash solution (6-fold concentration), colour solution and stop solution). Methanol (HPLC grade), Dichloromethane (HPLC grade), BSA (bovine serum albumin) and GCB columns are required for the sample pre-treatment stage.

e. Medium / matrix
The E2 ELISA test kit has been developed for the quantitative analysis of 17 ß-estradiol in water (groundwater, surface water and well water). It also can be used for other environmental water samples e.g. effluent samples, but has not been validated for such applications. There are no specific requirements for the physico-chemical parameters of the matrix. The GCB (graphitized carbon black) sample pre-treatment stage is essential in order to extract the steroid hormones from the raw water sample (section 5.1(6c)).

5.2 Module B: Applicability domain and pre-validation

1. Target parameters
The method endpoint is the concentration of E2 in the sample after the dilution has been accounted for. This is determined by calculating the sample E2 concentration using the absorbance intensity obtained from the standard curve.

2. Matrix and samples
   a. Type of matrix
   The matrix in which the validation study was performed was both spiked reverse osmosis filtered water samples and treated sewage effluent samples. To ensure all samples processed by the methods laboratory reflected as near as possible the concentrations at which the lead laboratory had intended to distribute, samples were preserved (section 5.2(2d)) and pre-treated (section 5.1(6c)). Without these steps the samples would vary greatly in composition and no longer represent the initial spiking regime.

   b. Sampling
   For the stage 1 assessment, the lead laboratory prepared 6 independent stock solutions of 17 ß –oestradiol (Sigma, Poole, UK) at 100 mg/L in

methanol (Analysis Grade >99.9% pure, Fisher, UK). Only three different batches of 17 ß –oestradiol were used to prepare the 6 independent stocks as the lead laboratory was limited by the availability of the chemical. From each stock, a 100 µg/L dilution in methanol was produced. The standards (100, 50, 10, 5, 1, and 0.5 ng of 17 ß –oestradiol) were prepared from the six 100 µg/L stocks. The methods laboratory requested for the spikes to be suitable for a 500 ml sample and so were prepared as follows: a 0.5 ml and 0.25 ml aliquot was taken from each 100 µg/L stock to produce standards of 100 and 50 ng of 17 ß –oestradiol respectively. A further 1 ml aliquot was taken from each 100 µg/L stock and added to 9 ml methanol to make a 10 ng/ml solution from which 0.5 ml was taken for the 10 ng/L standard. Additionally, 0.1ml of the each of the six 100 µg/L stocks were taken and added to 9 ml methanol to make a 1 ng/ml stock from which the 1 ng/L standards were produced by preparing aliquots of 0.5 ml. All vials were sealed and numbered (1-36), only the lead laboratory knew the relationship of the sample numbers to the concentrations in the vials.

For the chemical analysis, 6 secondary stocks of 100 ng/L (standards 1-6) from the 6 independent stocks were sent for analysis. Furthermore, E2 spiked samples of 0.5, 1, 5, 10, 50 and 100 ng/L (samples 1-6) were taken from standard 3 and sent alongside control blanks (see section 3.3.1 method).

For the Stage 2 assessment, the lead laboratory prepared 12 measured 1 litre volumes of reverse osmosis filtered water in glass bottles and 12 vials containing 100 µl methanol. The bottles and vials were numbered sequentially and labelled as ‘blanks’. Two of the methanol vials were spiked with 1.0 ng/L of E2 and two with 10 ng/L estrone (Sigma, Poole, UK), with only the leading laboratory knowing the relationship of the sample numbers to the content of the vials. For this stage, the UKEA also distributed 6 bottles containing negative effluent to the methods laboratory. This effluent was from a treated sewage effluent source which had been left for some time for the E2 to degrade.

The lead laboratory for Stage 3 obtained 12 litres of sewage effluent (positive samples) from a source considered to be of moderate oestrogenic activity (approx. 5-10 ng/L E2 equivalent). This was pooled and mixed thoroughly and then divided into 12 measured 1 Litre volumes in glass bottles. This was preserved as described in section 5.2.2d. These were sent alongside 6 vials containing 100 µl methanol spiked with 20 ng/L of 17 ß –oestradiol. Originally the lead laboratory proposed to add a 25 ng/L spike but this was impractical due to the graduation of the pipettes available at the time. Only the lead laboratory knew the concentration of the spiked vials.

During the preparation of the samples for the validation study the following health and safety considerations were taken into account. Stock solutions and spike solutions were prepared in a fume cabinet and safety glasses and protective clothing were worn. This was to avoid exposure to the oestrogenic

substances (17 ß –oestradiol and estrone) and the carrier, methanol. For stages 2 and 3, the environmental samples containing treated sewage effluent at unidentified levels of oestrogenicity were treated as toxic. This meant both the lead and method laboratory handled the samples in the fume cabinet and wore safety glasses and protective clothing. Samples were disposed of using the correct disposal route.

c. Sample characteristics
The 17 ß –oestradiol and estrone spikes used in stages 1 to 3 were prepared by the lead laboratory from compounds purchased from Sigma, Poole UK. All compounds were of a known purity (E2 ≥ 98%; E1 ≥ 99%). The methanol carrier was obtained from Fisher Scientific, UK and was 99.9+ % pure. The treated sewage effluent samples used in stages 2 and 3 were acquired from the BEMO study (Environment Agency, 2007), where the final effluent stream had been sampled from a sewage treatment works in the UK.

d. Sample stability and preservation, including transport
Stage 1 samples were stored at 4^{o}C prior to sending. The 17 ß –oestradiol in methanol vials were sent in cool boxes to the participant alongside 36 measured 500 ml volumes of reverse osmosis water in glass bottles instead of 1 litre volumes to reduce shipping costs. In order to simulate the effects of sending the samples aboard, the samples sent for chemical analysis in the UK were kept in cool boxes for 3 days before dispatch.

Stage 2 samples were again stored at 4^{o}C prior to sending. The 12 measured 1 litre volumes of reverse osmosis filtered water in glass bottles was sent in cool boxes alongside the glass vials containing either E2 in methanol, E1 in methanol or methanol alone. The negative effluent in glass bottles was not fixed in any way as the lead laboratory as wanted it to degraded by the time it arrived at the methods laboratory.

All 12 measured 1 litre samples of positive effluent in glass bottles were preserved by the lead laboratory before sending to avoid biodegradation of the sewage. To every 1 litre of sewage, 1 ml HCL and 0.25 g $CuSO_4$ was added. The positive effluent samples were sent in cool boxes along with the six glass vials containing the E2 spikes in methanol.

e. Availability of the organisms
This is not applicable for this method as it involves the direct measurement of oestrogenic substances

5.3 Module C: Intra-laboratory performance

5.3.1 Stage 1: Assessment of Accuracy, Precision, Sensitivity, Linearity and Range

Method

The lead laboratory prepared the E2 spiked samples as outlined in section 5.2(2b) which were sent alongside the water samples for dilution. The methods laboratory stored the samples at 4^{o}C until they were required for analysis. Immediately prior to analysis, the laboratory added a sample vial to each bottle of water, ensuring that the vial was thoroughly rinsed using only water from that bottle. The validation method suggested that no further dilution of any single concentration would be required. Samples were prepared and pre-treated as outlined in section 5.1(6c). The samples were originally analysed at an identical concentration factor of 150-fold and therefore all samples were analysed without further dilution as per the lead laboratory's instruction. However, the method laboratory found that this meant for the samples spiked with 50 and 100 ng/L E2, the concentration of E2 present in the sample could only be reported as > 6.7 ng/L giving a quantification range of only 0.5-10 ng/L. Consequently, appropriate concentration values were adopted for the spiked samples (150 concentration factor for 0.5-1 ng/L; 15 concentration factor for 5-10 ng/L; 1.5 concentration factor for 50-100 ng/L). The values reported here are using the adjusted concentration factors. The method validation laboratory ran the samples using ELISA assays for E2, EE2 and E1 though only the E2 ELISA had been required for the NORMAN project and only those results are presented here.

Rationale

This procedure provides a blinded study generating data that assesses recovery and therefore accuracy of 17 ß –oestradiol from 6 independent stock sources. It also allows the assessment of precision at 6 different concentrations by means of repeat analysis from the same source. To be considered acceptable at this level of validation, the mean recovery from each standard must be +/- 30% and the CV from each concentration must be <50%.

The concentration range has been chosen to reflect the ability of the method to detect 17 ß –oestradiol at and below the UK proposed PNEC (1 ng/L), while remaining above the limit of detection of the chemical analysis (0.3 ng/L) at the lowest end. The highest concentration dispatched is similar to that which can be measured in highly oestrogenic effluents in the UK. Therefore, the method should be able to reliably detect the lowest concentration to show sufficient sensitivity, and the highest concentration to demonstrate sufficient range to be considered validated at this level.

The results from each of the concentrations from the 6 stock solutions allows the construction of calibration curves with calculated margins of error over the range provided. These should adhere to a recognised curvilinear model to be considered validated at this level.

Results

Table 5.1 provides the levels of E2 in each of the spiked samples alongside the % CV and % recovery rates. The recovery rates are based upon the nominal spiked E2 values which were verified by chemical analysis (see section 3.3.1 results). The accuracy was calculated as the average value of six recoveries. The precision was calculated as the coefficient of variance of the E2 concentrations from the 6 samples per spike. The range was established as the spike concentration range for which the acceptable criteria were met.

It was decided during the workshop in Bergen that nominal chemical spikes would be used to calculate recovery rates as the chemistry analysis was only obtained for one spike range from standard 3 rather than for each of the six stocks so n=1. This was predominantly due to the cost of chemical analysis if all spiked samples had been analysed. All chemistry results were within 75 % to 104 % of the nominal values (see Table 3.2).

Table 5.1. E2 responses for Type 3 spiked water samples and recovery rates.

Spiked E2 concentration ng/L	Sample no.	E2 concentration ng/l	Average E2 ng/l	StDev	%CV	Nominal E2 Chemistry ng/l	Recovery %	Mean % recovery	StDev % recovery
0.5	5	0.38	0.49	0.11	21.73	0.50	76.00	98.33	21.37
	12	0.38					76.00		
	20	0.61					122.00		
	25	0.59					118.00		
	26	0.43					86.00		
	31	0.56					112.00		
1	3	1.02	0.87	0.16	18.96	1.00	102.00	86.83	16.46
	9	1.06					106.00		
	15	0.96					96.00		
	18	0.75					75.00		
	22	0.75					75.00		
	29	0.67					67.00		
5	1	4.2	4.15	0.24	5.85	5.00	84.00	83.00	4.86
	2	3.8					76.00		
	6	4.1					82.00		
	11	4.5					90.00		
	32	4.3					86.00		
	33	4					80.00		
10	10	8.4	8.10	0.35	4.28	10.00	84.00	81.00	3.46
	13	8					80.00		
	16	8.3					83.00		
	19	8.4					84.00		
	21	8					80.00		
	30	7.5					75.00		
50	4	41	41.83	1.83	4.39	50.00	82.00	83.67	3.67
	7	41					82.00		
	17	45					90.00		
	23	41					82.00		
	24	43					86.00		
	27	40					80.00		
100	8	96	86.33	7.12	8.24	100.00	96.00	86.33	7.12
	14	88					88.00		
	28	93					93.00		
	34	82					82.00		
	35	80					80.00		
	36	79					79.00		

Figure 5.1 illustrates the recovery rates of the spiked samples against the expected concentrations. All mean recovery rates for 0.5-100 ng/L are within the 70-130% recovery limits. Overall, the method appears to produce slightly lower recovery rates than the initial spikes but this is within the limits of the protocol rationale and also corresponds with the chemical analysis of the spiked samples were verified concentrations were below those of the initial spike (see section 3.3.1 results).

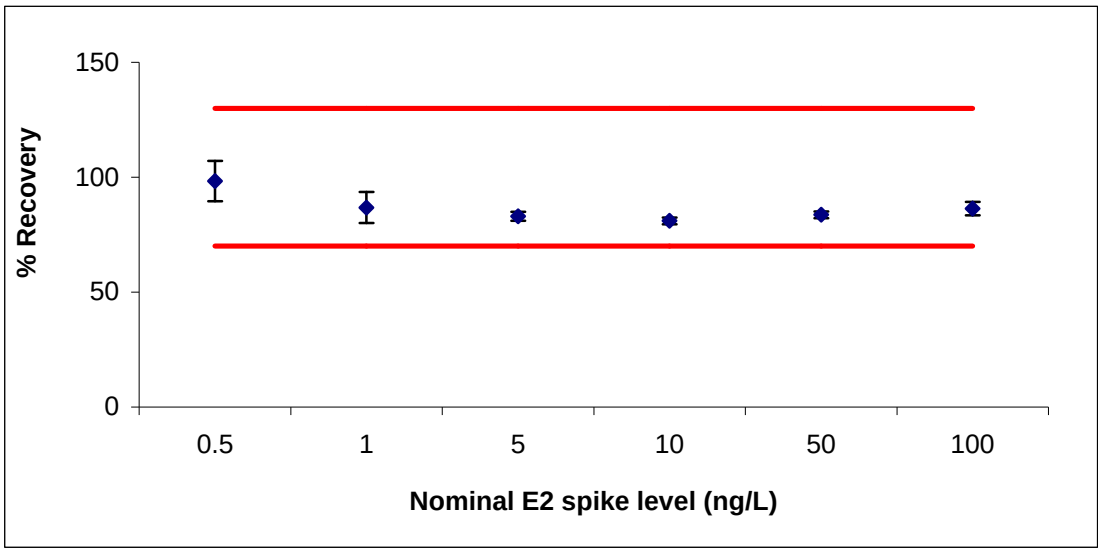

Fig. 5.1. Accuracy of Type 3 method measured by percentage recovery of the E2 spiked water samples. The horizontal lines indicate the recovery acceptance criteria (70-130%). The error bars indicate the standard error from the mean (n=6).

The Type 3 method is shown to be acceptable for precision as at all spiked concentrations the % CV is well below 50% (4.28-21.73%) (Fig. 5.2). As expected the precision is shown to be enhanced at the higher spiked concentrations.

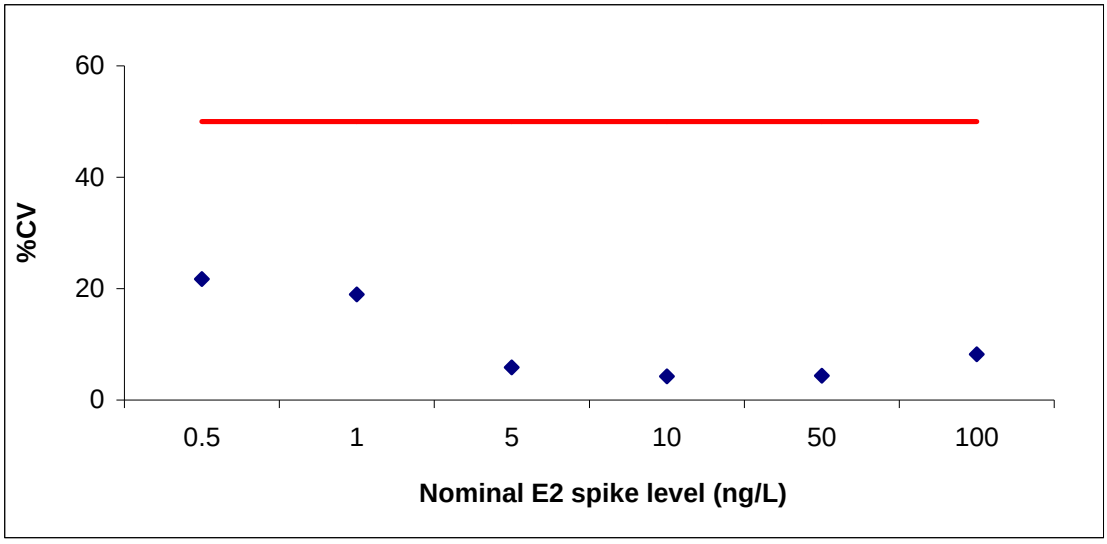

Fig. 5.2. Precision of Type 3 method in water samples spiked with E2 represented by coefficience of variance (% CV). The horizontal line indicates the precision acceptance criteria (<50%) (n=6).

Linearity was assessed by the construction of a calibration curve with calculated margins of error for the six analyses overall (Fig. 5.3) and as the six separate analyses (Fig. 5.4a-f) in order to understand if there was any variation in curve shape.

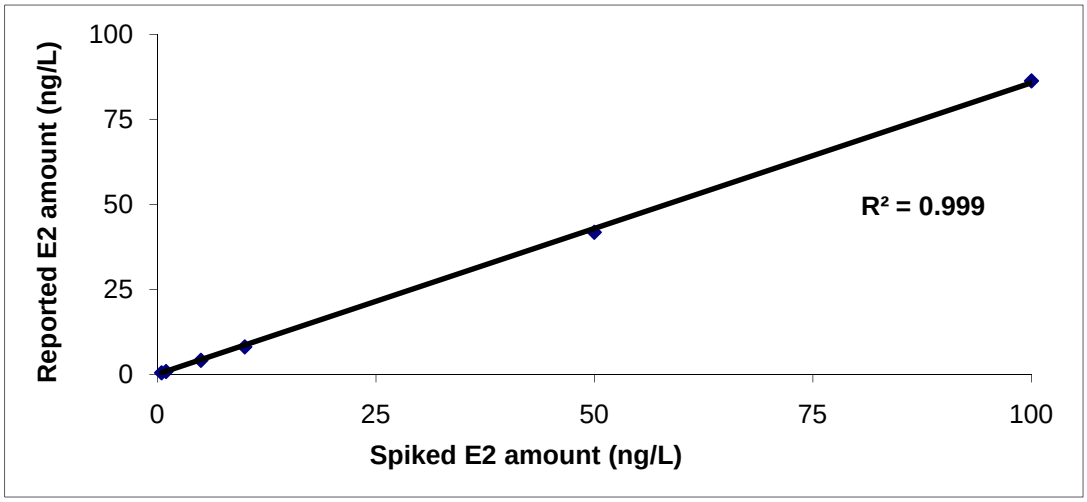

Fig. 5.3. Linearity of Type 3 method. Line represents the regression line through diamonds and forced through zero. The correlation coefficient (R^2) indicates the proximity of the data to a perfect line. Error bars indicate the standard error of the mean (n=6).

When regression analysis was applied using a linear model to describe the relationship between the spiked E2 response and the reported E2 response, a statistically significant relationship at the 99% confidence level was reported (*P* = 0.0001).

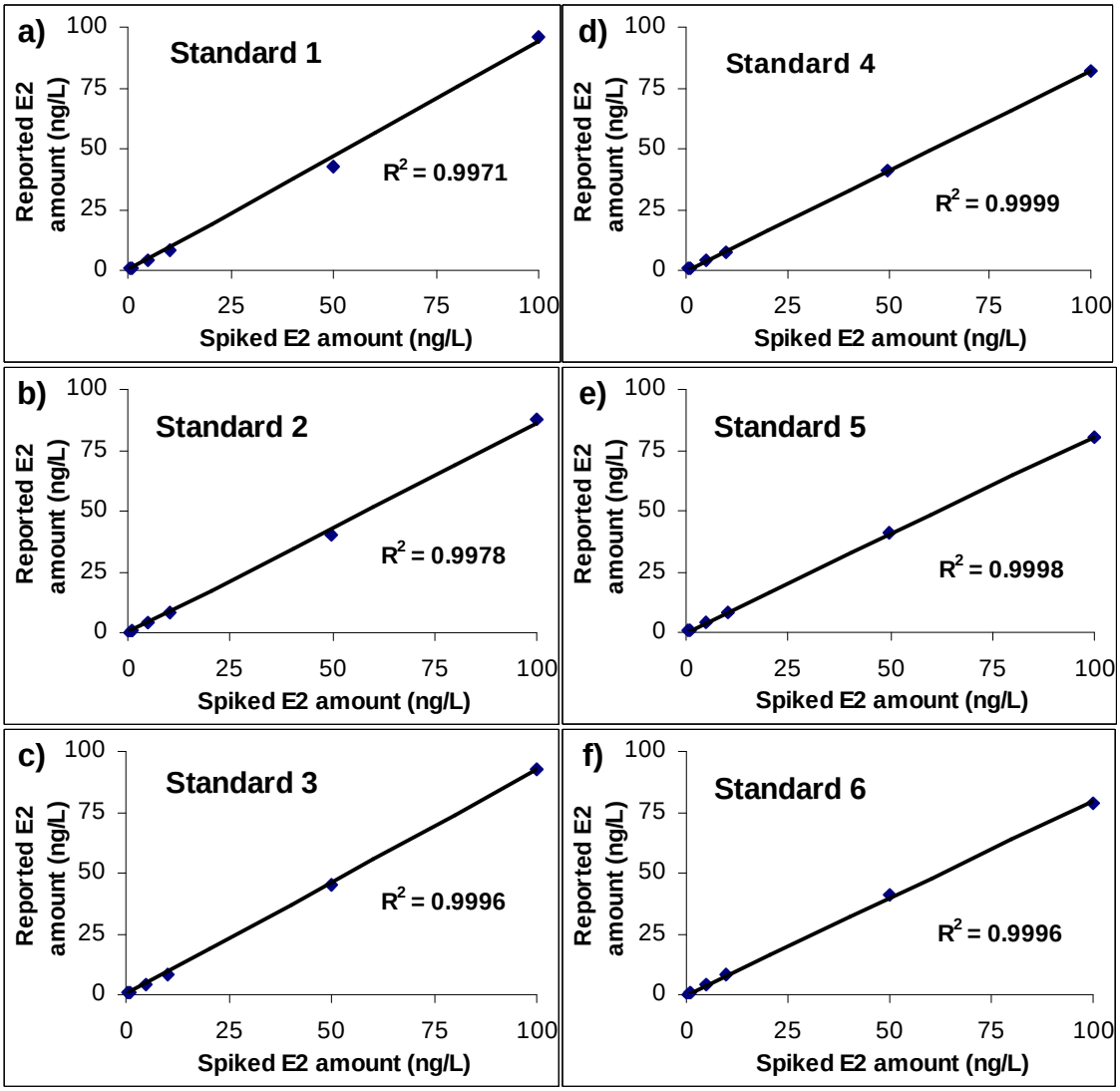

Fig. 5.4. Linearity of the individual standard analyses for Type 3 Method. Line represents the regression line through diamonds and forced through zero. The correlation coefficient (R^2) indicates the proximity of the data to a perfect line.

The regression lines for the six analyses displayed little deviation from the linear regression line, with the R^2 ranging from 0.9971 to 0.9999. The validation protocol rationale did not set a correlation coefficient above which the method would be considered validated, but as all R^2's are very close to 1 the method has shown good linearity.

Linearity for this method was also assessed by plotting the response (e.g. signal divided by the concentration) as a function of the concentration for each of the separate analyses (Fig. 5.5). The observed line should be horizontal and an ideal response factor of 1 for all concentrations if the nominal E2 spike and the measured E2 were of similar values. The linear range of the method was taken as 95 to 105 % of the horizontal line result calculated from the mean of the six analyses. Figure 5.5 illustrates that though the Type 3 method gave responses for all six analyses around the response factor of 1, all analyses for the 0.5 and 1.0 ng/l spikes fell above or below the horizontal line response, showing that there is some difficultly in replicating results at the lower spike concentrations. Consequently, the method does not completely demonstrate sensitivity at the lower concentrations but as the coefficience of variance for 0.5 and 1.0 ng/l were below 50% and the recovery rates for both concentrations were above 70% the method can still be considered sensitive enough for this validation study.

As this method has displayed accuracy in its recovery rate, good precision and linearity, an acceptable working range of 0.5 to 100 ng/l of E2 can be formulated for Type 3 method.

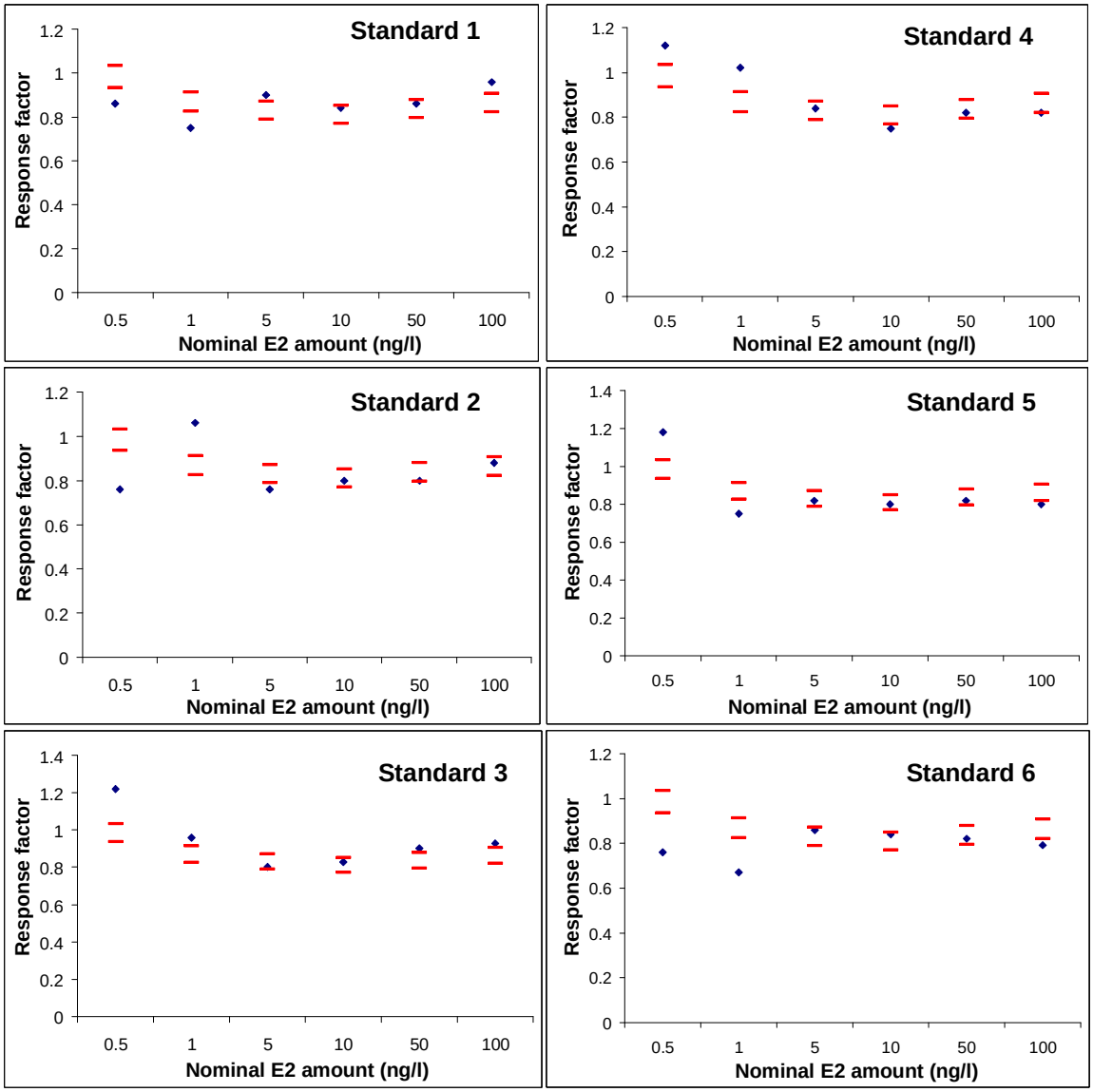

Fig. 5.5. Response factors of the type 3 method to E2 spiked water samples. The horizontal bars represent the response factor criteria which is 95% and 105% of the mean.

5.3.2 Stage 2: Assessment of Negative Response and Selectivity

Method

For this stage, the lead laboratory provided the methods laboratory with bottles and vials labelled as blanks prepared as outlined in section 5.2(2b). The methods laboratory stored the samples at 4°C until they were required for analysis. Immediately prior to analysis, the methods validation laboratory added the relevant vial of methanol (some containing spikes of 1.0 ng/L E2 or 10 ng/L E1) to the corresponding bottle of water, ensuring that the vial was thoroughly rinsed using water from that bottle. Estrone was utilised as a secondary spiking compound in order to assess the capability of the method to distinguish between E2 and other likely interfering compounds. Additionally, the methods laboratory analysed the degraded treated sewage effluent (negative effluent) which was prepared as described in

section 5.2(2b). A sample concentration factor of 150-fold was adopted for analysis of the spiked samples and the negative effluent. Samples were run using the ELISA assay for E2. In addition, the methods laboratory ran the E1 ELISA assay for the spiked samples as it was known some of the blanks had been spiked with E1 but the laboratory did not know which ones.

Rationale

This stage was conducted to provide a blind study to assess the ability of the method to distinguish between negative and positive samples. To be considered acceptable at this level of validation, the method must be able to identify all of the samples that were spiked with 17 ß –oestradiol. The ability of the methods laboratory to determine accurately the concentration of E1 and E2 spikes is not relevant for this stage and consequently additional samples were not sent for chemical analysis.

A negative response to those samples spiked with estrone will demonstrate the selectivity of the method against similar compounds and therefore all other samples apart from those spiked with E2 should return a result below the limit of detection of the method to be considered validated at this level. A negative result to the treated sewage effluent (negative effluent) will demonstrate the selectivity of the method in environmental matrices and should return a result below those obtained for similar samples in stage 3 (positive effluent) to be considered validated as this level.

Results

The results show that the method was able to distinguish between the samples spiked with E2 (samples 5 and 7) and those spike with either E1 or the blanks (Table 5.2). The E1 ELISA which was also conducted by the methods laboratory identified correctly the E1 spikes as samples 4 and 8 returning estrone concentrations of 6.6 and 7.1 ng/L respectively. The method did demonstrate that for samples spiked with E1, the E2 ELISA reported E2 equivalents at or below the limit of detection (0.33 ng/L). However, for the blanks samples (no oestrogenic spike), E2 concentrations above the limit of detection were reported for five of the eight samples. Thus indicating that perhaps the limit of detection is too low for this method if such a large proportion of blanks are detected above the limit of detection. The assay has previously shown that the antibody binds exclusively with E2 and does not show cross-reaction with other chemicals of similar structures. The highest cross-reactivity of the E2 ELISA kit for other oestrogens and hormones was 16.0% for 16-keto E2 and E2-3-glucronide and 1.3% for estrone (Abraxis, 2006). Figure 5.6 provides an illustration of the 17 ß –oestradiol concentrations returned for each of the samples in comparison to the levels spiked and the limit of detection.

For the six negative effluent (degraded treated sewage) samples, concentrations of E2 ranged from 0.44 to 0.48 ng/L, above the limit of detection indicating there was some oestrogenicity remaining (Fig. 5.6). The significance of this negative effluent E2 concentrations in comparison to those of the positive effluent is discussed in section 5.5.3 (results).

Table 5.2. E2 equivalents in spiked water samples and negative effluents for Type 2 method

Sample	Sample no.	E2 Result ng/l	E2 Average ng/l	StDev	CV%
Blanks	1	0.30	0.38	0.07	17.79
	2	0.32			
	3	0.36			
	6	0.32			
	9	0.38			
	10	0.42			
	11	0.48			
	12	0.46			
E2	5	0.82	0.82	0.01	0.87
	7	0.81			
E1	4	0.33	0.32	0.02	6.73
	8	0.30			
Negative effluent		0.44	0.45	0.02	3.32
		0.45			
		0.48			
		0.45			
		0.44			
		0.46			

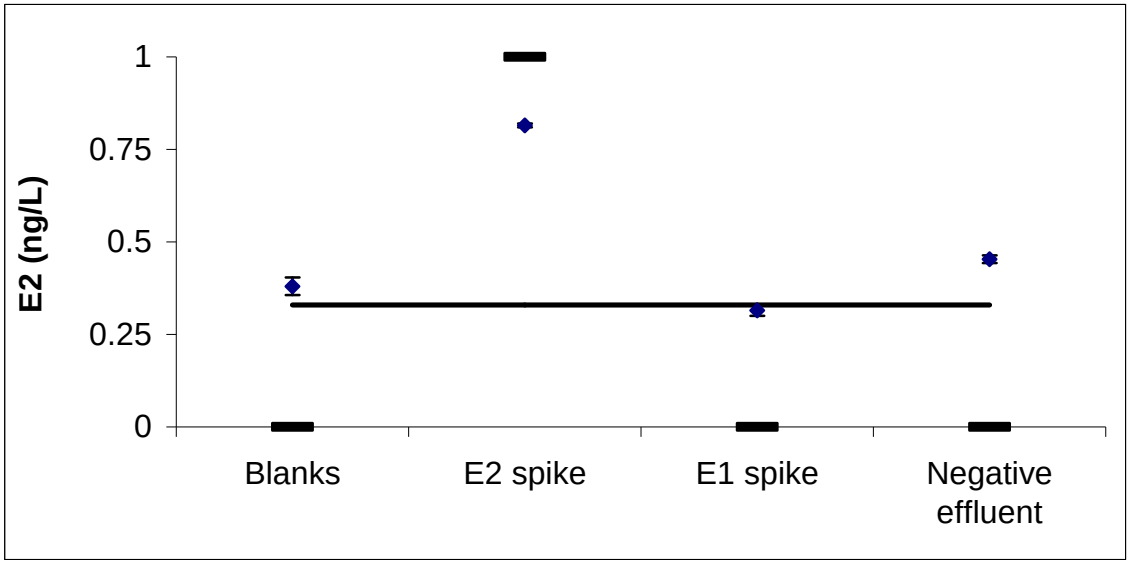

Fig. 5.6. Negative response of Type 3 method. Assays ability to discriminate between spiked and unspiked E2 samples. The horizontal line indicates the methods limit of detection whilst the horizontal bars denote the level at which each sample was spiked with E2. Error bars indicate the standard error from the mean.

5.3.3 Stage 3: Assessment of Specificity in Environmental Matrices

Method

The lead laboratory provided 12 bottles of treated sewage effluent (positive sample) and six vials of methanol spiked with E2 as described in section 5.2(2b). It was advised that the preparation and analysis of this stage was performed immediately on receipt of the bottles as the oestrogenic activity of the treated sewage effluent was not stable. Immediately prior to analysis, the methods validation laboratory added one vial of the spike to each of the six bottles of treated sewage effluent to create six samples of treated effluent sewage matrices that reliably contained 17 ß –oestradiol (spiked positive effluent). For this stage the methods laboratory used a 150-fold concentration factor for the positive effluent and a 15-fold concentration factor for the spike positive sample analysis. Again, an EE2 ELISA was run alongside the analysis for E2 but ethynyl oestradiol was not detected in any significant amount in either the positive effluent or spiked positive effluent (0.27-0.33 ng/L EE2).

Rationale

This study was performed to generate data to demonstrate the robust ability of the method to detect and discriminate between concentrations of 17 ß –oestradiol in environmental matrices. To be considered acceptable at this level of validation the COV from each sample must be <50% (n=6).

The accuracy of the result is not taken into account for the two environment samples, as chemical analysis of the sample cannot be provided due to sample changes during shipping.

However, the concentration of the 17 ß –oestradiol spike can be inferred by the difference between the results of the spiked and unspiked environmental sample, demonstrating the specificity of the method in test conditions. Therefore to be considered acceptable at this level of validation, the mean recovery of the spike should be between 70-130% (n=6).

Results

The type 3 method has demonstrated that is able to detect and discriminate between samples spiked with 17 ß –oestradiol in an environmental matricies. The coeffcience of variance for the positive effluent sample replicates and the spiked positive effluent replicates was 5.70% and 3.19% respectively (Table 5.3 and Fig. 5.7). The recovery of the E2 spike (20 ng/L) was shown to be satisfactory, as the recovery for the six samples ranged from 85.0-92.5%.

Table 5.3. Selectivity of Type 3 method in environmental matrices. Assays ability to discriminate between spiked and unspiked samples (n=6).

Sample	E2 Result ng/l	E2 Average ng/l	StDev	CV%	% Recovery	Mean % Recovery	StDev Recovery
Positive effluent	2.5	2.58	0.15	5.70			
	2.6						
	2.5						
	2.8						
	2.4						
	2.7						
Positive effluent (spiked E2)	17.0	18.12	0.58	3.19	75.28	80.22	2.56
	18.4				81.48		
	18.3				81.03		
	18.5				81.92		
	18.0				79.70		
	18.5				81.92		

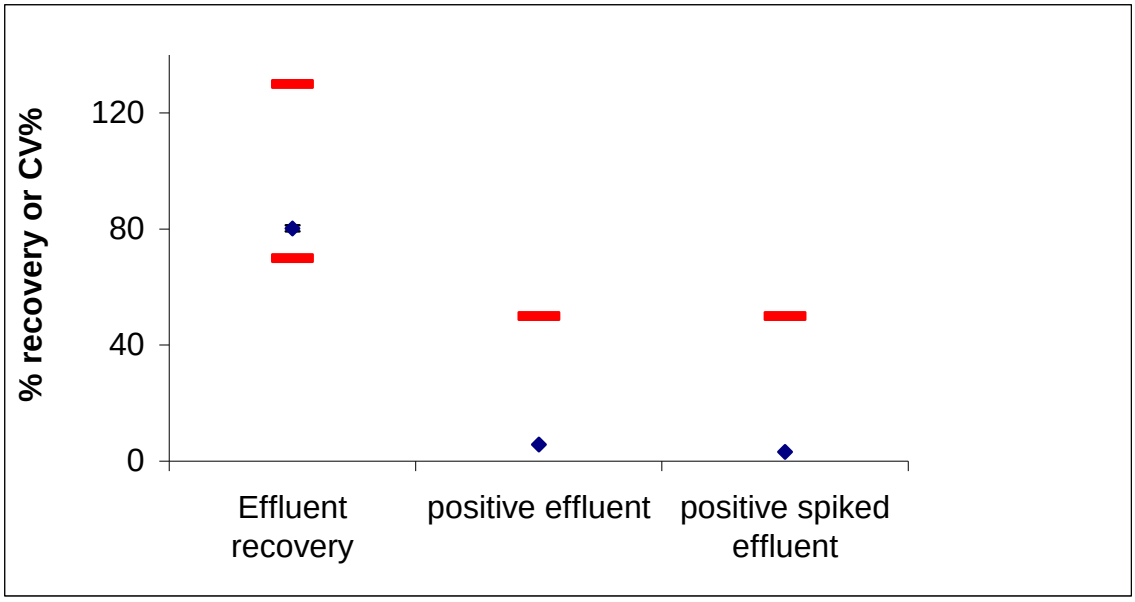

Fig. 5.7. Specificity of Type 3 method in environmental matrices measured by the recovery of the E2 spike and the precision of the environmental samples. The horizontal bars indicate the recovery acceptance criteria (70-130%) and the precision acceptance criteria (<50%). Error bars indicate the standard error from the mean (n=6).

Comparison of the E2 responses in the negative effluent samples from stage 2 and the positive effluent samples in stage 3 demonstrates that the oestrogenicity of the negative effluent sample had degraded to some extent before being analysed. The E2 response for the negative effluent ranged from 0.44-0.48 ng/L whilst the positive effluent E2 response ranged from 2.4-2.8 ng/L (Fig. 5.8).

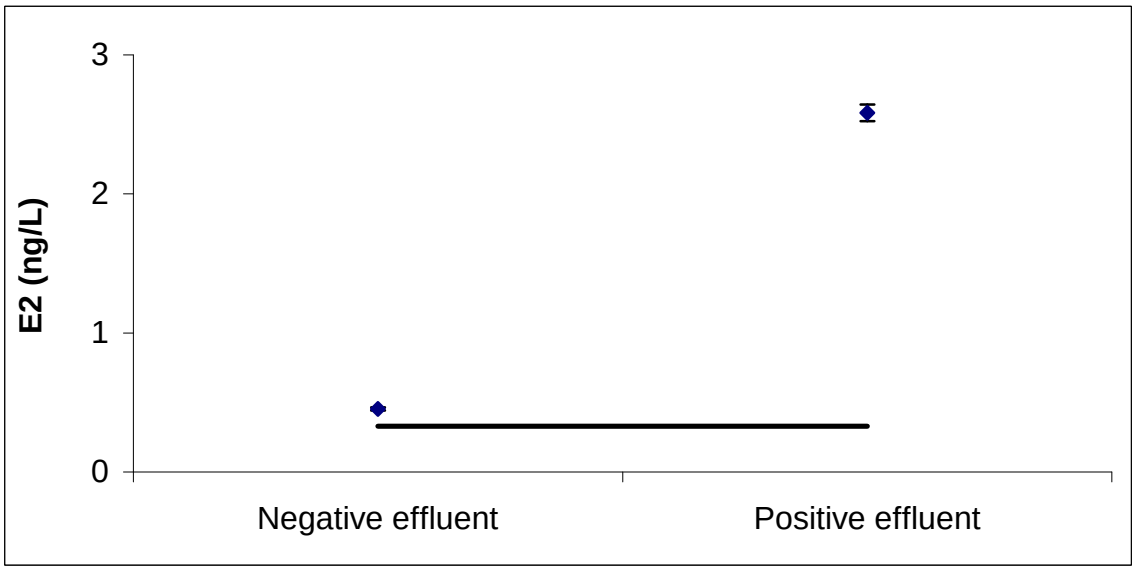

Fig. 5.7. Selectivity of Type 3 method in environmental matrices demonstrated by the E2 equivalents in the negative and positive effluent samples. The horizontal line indicates the limit of detection of the method. Error bars indicate the standard error from the mean.

6 Conclusions and Recommendations

This intra-laboratory validation study of oestrogens at the research level has provided an assessment of the methods performance characteristics, advantages and limitations and has helped to identify some important considerations when setting up a validation study. Hopefully this will provide some insight and lessons learned for any future intra-laboratory studies within the NORMAN permanent network.

In summary, the methods performed well against CASE 1's specific programme of testing for the generation of validated data (Table 6.1), with only a few exceptions. The E-screen assay (Type 1 method) suffered from an unexplainable low recovery of E2 from the water samples in stage 1 (the recoveries from the effluents was good) and from outlying values in the unspiked water samples (blanks) in stage 2. The low recovery rates at stage 1 meant that an acceptable working range for the assay could not be formulated. The type 2 method (Vitellogenin), only in part met the criteria for discriminative ability (stage 3) due to the large variability between replicates caused by the limited sample volume available. The ELISA assay (Type 3 method) did not meet the criteria for stage 2 negative response as some of the blanks returned E2 results above the limit of detection. However, as all the E2 equivalents were considered low in the blanks this may mean that the limit of detection should be raised rather than an issue with the selectivity of the method.

Though the methods selected for Validation 1 were considered as method validation at the level of research laboratories, the extent each method had previously been validated varied considerably. For example, the detection of vitellogenin in fathead minnow plasma (type 2 method) had previously been standardised undergoing both intra- and inter-laboratory validation (Eidem *et al.*, 2006), where the need for the standardization of the performance criteria and validation had already been recognised (Goksøyr *et al.*, 2003). As this assay along with the Type 3 ELISA had been developed into a test kit, this aided the application of this validation process. This means the methods laboratory is less likely to deviate for the test kit protocol thereby making it easier to understand the methods requirements and applicability as well as aiding the interpretation of the results. This is not to say methods which are not in development as a test kit should be ignored, but that the lead laboratory needs to be aware that there will be more gaps in the information available to the method laboratories protocols.

Table 6.1. Summary of the validation results for each method type at each stage. The shaded boxes indicate cases where an acceptance criteria was not met.

Stage	Performance characteristic	Type 1 method (E-Screen)	Type 2 method (Vtg)	Type 3 method (ELISA)
1	**Accuracy**	<70%	70-130%	70-130%
	Precision	<50%	<50%	<50%
	Linearity	OK	OK	OK
	Sensitivity	OK	N/A	OK
	Range	None	0.2-12.5 ng VTG /ml	0.5-100 ng E2 /L
2	**Negative response**	Unspiked>LOD	OK	Unspiked>LOD
	Selectivity	OK	Could not be tested	OK
3	**Specificity**	OK	N/A	OK
	Discriminative ability	N/A	Some difficulties in discrimination between samples	N/A
4	**Relative potency**	OK	N/A	N/A

Feedback from the methods laboratories during the workshop in Bergen gave a good insight into the advantages and disadvantages of this type of validation study. The participants felt that participating in a blind study such as this provided valuable information on the assays performance which is not assessed in the same way when producing a peer reviewed publication. Furthermore, when the acceptance criteria is not set by the method laboratory, information can be gained about the assays weaknesses as well as its strengths which may be ignored during a publication, such as the assays cross reactivity with other chemicals. However, it was recognised that there needs to be far more consideration during the selection process of whether a method it fit for purpose. For example, Type 1 method was not suitable for the assessment of relative potency of oestrogenic compounds (Stage 4) as compounds other than (anti-)-estrogens have been reported to stimulate or inhibit cell growth for this assay, thereby over- or underestimating the estrogenic response. Method selection at this validation level is problematic in that the documented methodology, usually restricted to peer reviewed publications, do not necessarily provide the detail required to fully assess the suitability of the assay. When this is the case, there is a clear need for increased interaction between the lead and method laboratories over the capabilities of the method and

the laboratories requirements before the initiation of the validation study. This point was also demonstrated by the lack of sample volume for type 2 method for stage 3 where the laboratory was unable to conduct replicates and as the sample required dilution some samples could not be repeated so where outside the methods working range. This created uncertainty in the results reported. Again this is something difficult to assess when using selection criteria based upon published papers but is an important requirement to include in future validation studies.

It was felt that the validation study would have benefited with a larger amount of money put aside for chemical analysis of the samples at several of the stages. Spiked E2 samples were analysed by LC-MS/TOF for stage 1 (Type 1 and 3 methods) but only one set of samples (0.5-100 ng/l) were analysed alongside the six stock solutions. This meant that the recoveries for the spiked samples using the assays were based upon the nominal spikes rather than the chemistry as the E2 concentrations in the other standards was unknown. If all standards had been analysed this would have provided greater confidence in the results and would have helped reduce the uncertainties of the recovery rates for Type 1 method.

For validation studies to continue to succeed and for the work to be of a high standard, it was considered important that for future studies participating methods laboratory should receive recognition for the amount of work involved and it was discussed whether this should be a financial gain. Furthermore, it necessary that the outcomes from any validation study are publicly available and include the performances of the assays as well as the lessons learned.

References

Abraxis, (2006), ECOLOGIENA for Environmental Pollutants 17 ß-Estradiol (E2) ELISA Kit Prod. N. PN590062 [online], Abraxis, Warminster, PA,. Available from: www.abraxiskits.co/moreinfo/PN590062info.pdf [Accessed 02/11/2007].

Biosense Laboratories, (2005) Fathead Minnow Vitellogenin ELISA Kit Prod. No.: V01018401 [online], Biosense Laboratories AS, Bergen. Available from: www.biosense.com/docs/FHM_(2005.1).pdf [Accessed 19/06/2007].

Choi, D.H., Katakura, Y., Matsuda, R., Hayashi, Y., Hirobe, M., Goda, Y., Ninomiya, K., and Shioya, S (2004) Validation of a method for predicting the precision, limit of detection and range of quantification in competitive ELISA, Water Research 38, 4323-4330.

Eidem, J.K., Kleivdal, H., and Goksøyr, A., (2005) Single-Laboratory Validation of a homologous ELISA to quantify Vitellogenin in the fathead minnow (*Pimephales promelas*) [online]. Biosense Laboratories AS, Bergen. Available from: www.biosense.com/docs/fhmsinglelab.pdf [Accessed 29/01/08].

Eidem, J.K., Kleivdal, H., Kroll, K., Denslow, N., van Aerle, R., Tyler, C., Panter, G., Hutchinson, T., and Goksøyr, A., (2006) Development and validation of a direct homologous quantitative sandwich ELISA for fathead minnow (*Primephales promelas*) vitellogenin, Aquatic Toxicology 78, 202-206.

Environment Agency, (2007) Biological effect measures in fish – application to treated sewage effluent, Science report SC030278/SR, Environment Agency, Bristol.

Environment Agency, (2008) The determination of steroid oestrogens in waters using chromatography and mass spectrometry, Methods for the examination of waters and associated materials, Environment Agency, Bristol, Available from: www.environment-agency.gov.uk/commondata/acrobat/steroids220_2107605.pdf [Assessed: 06/10/08].

EPA, (2003) Comparative Evaluation of Vitellogenin for Medaka and Zebra Fish, draft Final Report. EPA Contract Number 68-W-01-023, Work Assignment 2-26 [online]. U.S. Environmental Protection Agency, Columbus, Ohio. Available from: www.epa.gov/scipoly/oscpendo/pubs/edmvs/draft_final_medaka_zebrafish_jul30.pdf [Accessed: 29/01/08].

Goksøyr, A., Eidem, J.K., Kristiansen, S-I., Nilsen, B.M., (2003) On the need for a standardized set-up for validation studies of fish vitellogenin assays as an endpoint in endocrine disrupter testing and screening- a proposal [online], Biosense Laboratories AS, Bergen. Available from: www.biosense.com/docs/goksoyetal2003.pdf [Accessed 29/01/08].

Hirobe, M., Rubio, F., Fujita, M., and Shiraishi, H., (2004) Development of ELISAs for quantification of surfactants, endocrine disrupters and estrogens, and their application for environmental and biological sample analysis. Presented at the National Monitoring Conference, Chattanoga, Tennessee, USA, 17-20 May 2004.

Japan EnviroChemicals Ltd, (2004a) Ecologiena Estrone (E1) ELISA kit (Microplate) User's Guide, Product No. 590062 [online], Japan EnviroChemicals Ltd, Tokyo. Available from: www.abraxiskits.com/moreinfo/PN590061USER.pdf [Accessed 28/03/2007].

Japan EnviroChemicals Ltd, (2004b) Ecologiena Ethynlestradiol (EE2) ELISA kit (Microplate) User's Guide, Product No. 590051 [online], Japan EnviroChemicals Ltd, Tokyo. Available from: www.abraxiskits.com/moreinfo/PN590051USER.pdf [Accessed 28/03/2007].

Japan EnviroChemicals Ltd, (2004c) Ecologiena 17 ß-Estradiol (E2) ELISA kit (Microplate) User's Guide, Product No. 590062 [online], Japan EnviroChemicals Ltd, Tokyo. Available from: www.abraxiskits.com/moreinfo/PN590062USER.pdf [Accessed 28/03/2007].

Jensen, K.M., and Ankley, G.T., (2006) Evaluation of a commercial kit for measuring vitellogenin in the fathead minnow (*Pimephales promelas*), Ecotoxicology and Environmental Safety, 64, 101-105

Leonards, P., and Schwesig, D., (2007) Protocol for the development of chemical and biological monitoring methods at the research level, Version 1.0 [online], NORMAN Report, Verneuil-en-Halatte, France. Available from: www.norman-network.net/public_docs/norman_v1_v2_v3_version_01.pdf [Accessed 02/11/07].

Murk, A.J., Legler, J., Marola, M.H., van Lipzig, M.H., Meerman, J.H.N., Belfroid, A.C., Spenklink, A., van der Burg, B., Rus, G.B.J., and Vethaak, D., (2002) Detection of estrogenic potency in wastewater and surface water with three in vitro bioassays, Environmental Toxicology and Chemistry, 21, 16-23.

Shappell, N.W., (2006) Estrogenic activity in the environment: municipal wastewater effluent, river, ponds and wetlands, Journal of Environmental Quality, 35, 122-132.

Shappel, N.W., Billey, L.O., Billey, Forbes, D., Matheny, T.A., Poach, M.E., Reddy, G.B., and Hunt, P.G., (2007) Estrogenic activity and steroid hormones in swine wastewater through a lagoon constructed-wetland system, Environ. Sci. Technol., 41, 444-450.

Soto, A.M., and Sonnenschein, C., (1995) The E-screen assay as a tool to identify estrogens: An update on estrogenic environmental pollutants, Environ. Health Perspect., 103 (Supplement 7), 103, 113-122.

Tyler, C.R., van Aerle, R., Hutchinson, T.H., Maddix, S., and Trip, H., (1999) An *in vitro* testing system for endocrine disrupters in fish early life stages using induction of vitellogenin, Environmental Toxicology and Chemistry, 18, (2), 337-347.

Wegener, J-W., (2006) Overview of the methods and list of research laboratories participating in the laboratory study, Final version, NORMAN Report, Veneuil-en-Halatte, France. Restricted access – NORMAN network, [Accessed 02/11/07].