NERC Open Research Archive sharing our research

Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL

Article (refereed) - postprint

Ellis, Tim; Margiotta-Casaluci, Luigi; Pottinger, Tom G.; Morris, Steve; Reese, R. Allan; Sumpter, John P.; Scott, Alexander P. 2020.
Immunoassays are not immune to errors: examples from two studies of steroid output from freshwater trout farms.

Crown Copyright © 2019
This manuscript version is made available under the CC BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

This version available http://nora.nerc.ac.uk/525269/

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at http://nora.nerc.ac.uk/policies.html#access

NOTICE: this is an unedited manuscript accepted for publication. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before publication in its final form. During the production process errors may be discovered which could affect the content. A definitive version was subsequently published in ***General and Comparative Endocrinology*** **(2020), 285: 113226. https://doi.org/10.1016/j.ygcen.2019.113226**

www.elsevier.com/

Contact CEH NORA team at
noraceh@ceh.ac.uk

The NERC and CEH trademarks and logos ('the Trademarks') are registered trademarks of NERC in the UK and other countries, and may not be used without the prior written consent of the Trademark owner.

ACCEPTED MANUSCRIPT

Immunoassays are not immune to errors: examples from two studies of steroid output from freshwater trout farms.

Authors: Tim Ellis[a, *], Luigi Margiotta-Casaluci[b], Tom G. Pottinger[c], Steve Morris[a], R. Allan Reese[a], John P. Sumpter[b], Alexander P. Scott[a]

Affiliations:

a. Centre for Environment, Fisheries and Aquaculture Science (Cefas), Weymouth Laboratory, Barrack Road, The Nothe, Weymouth, Dorset, DT4 8UB, UK
b. Brunel University London, Institute for Environment, Health and Societies, Uxbridge, Middlesex, UB8 3PH, UK.
c. Centre for Ecology & Hydrology, Fish Ecophysiology, Lake Ecosystem Group, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster, LA1 4AP, UK

* Corresponding author: Tim Ellis, Cefas Weymouth Laboratory, Barrack Road, The Nothe, Weymouth, Dorset, DT4 8UB, UK. E: tim.ellis@cefas.co.uk

Running title: Errors in immunoassays

©Crown Copyright 2019

Highlights:

1. We repeated a study that reported abnormally high steroid levels in river water.
2. In the repeated study, 2 of the 4 laboratories made traceable calculation errors.
3. We found steroid levels 2-3 orders of magnitude lower than in the original study.
4. We attributed the high steroid levels in the original study to calculation errors.
5. Current evidence indicates that trout farms are a minor source of river steroids.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

A "reproducibility crisis" is widespread across scientific disciplines, where results and conclusions of studies are not supported by subsequent investigation. Here we provide a steroid immunoassay example where human errors generated unreproducible results and conclusions. Our study was triggered by a scientific report citing abnormally high concentrations (means of 4 to 79 ng L^{-1}) of three natural sex steroids [11-ketotestosterone (11-KT), testosterone (T) and oestradiol (E2)] in water samples collected from two UK rivers over 4 years (2002-6). Furthermore, the data suggested that trout farms were a major source because reported steroid concentrations were 1.3 to 6 times higher downstream than upstream. We hypothesised that the reported levels were erroneous due to substances co-extracted from the water causing matrix effects (i.e. "false positives") during measurement by enzyme-linked immunoassay (EIA). Thus, in collaboration with three other groups (including the one that had conducted the 2002-6 study), we carried out field sampling and assaying to examine this hypothesis. Water samples were collected in 2010 from the same sites and prepared for assay using an analogous method [C18 solid phase extraction (SPE) followed by extract clean-up with aminopropyl SPE]. Additional quality control ("spiked" and "blank") samples were processed. Water extracts were assayed for steroids using radioimmunoassay (RIA) as well as EIA. Although there were statistically significant differences between EIA and RIA (and laboratories), there was no indication of matrix effects in the EIAs. Both the EIAs and RIAs (uncorrected for recovery) measured all three natural steroids at <0.6 ng L^{-1} in all river water samples, indicating that the trout farms were not a significant source of natural steroids. The differences between the two studies were considerable: E2 and T concentrations were ca. 100-fold lower and 11-KT ca. 1000-fold lower than those reported in the 2002-6 study. In the absence of evidence for any marked changes in husbandry practice (e.g. stock, diet) or environmental conditions (e.g. water flow rate) between the study periods, we concluded that calculation errors were probably made in the first (2002-6) study associated with confusion between extract and water sample concentrations. The second (2010) study also had several identified examples of calculation error (use of an incorrect standard curve; extrapolation below the minimum standard; confusion of assay dilutions during result work-up; failure to correct for loss during extraction) and an example of sample contamination. Similar and further errors have been noted in other studies. It must be recognised that assays do not provide absolute measurements and are prone to a variety of errors, so published steroid levels should be viewed with caution until independently confirmed.

Keywords : enzyme-linked immunoassay; radioimmunoassay; steroid; extraction; calculation error; reproducibility crisis.

ACCEPTED MANUSCRIPT

1. Introduction

A “reproducibility crisis” is thought to afflict all scientific disciplines, where results and conclusions of studies are not supported by subsequent investigation (Baker, 2016). The publication of unrepeatable research claims is attributed to multiple causes, ranging from inappropriate experimental and statistical methods, through to selective reporting and even data fabrication (Begley, 2013; Ioannidis, 2005). Here we provide an example of how human error generated unreproducible environmental steroid concentrations. We propose that such human errors may be common and account for much of the wide variation in environmental steroid concentrations reported in the scientific literature.

Steroids enter water bodies from natural and anthropogenic sources, e.g. vertebrate wildlife, farmed livestock and human sewage, e.g.(Barel-Cohen et al., 2006; Kolodziej et al., 2004; Zhao et al., 2010). Steroids in water, no matter what their origin, can potentially affect the physiology and behaviour of aquatic vertebrates. Natural steroids in water bodies are typically at far lower concentrations than within aquatic vertebrates themselves. However, if concentrations of some of these natural steroids are present in the tens of ng L^{-1} range, they are highly likely to affect internal hormone systems and reproductive function (e.g. vitellogenin induction in male and immature fish; intersex), processes commonly termed “endocrine disruption” (Balaam et al., 2010; Feswick et al., 2014; Hanselman et al., 2004; Tyler et al., 1998; Zhao et al., 2010). Steroids have also been suggested to act as pheromones in fish (Stacey, 2015), so anthropogenic inputs could potentially further affect wild fish behaviour and physiology. Valid reporting of environmental steroid concentrations is therefore key to identifying sources and potential impacts.

The trigger for the present study was that a scientific report published on a UK Government website (Anon, 2006) reported mean concentrations of the natural steroids 11-ketotestosterone (11-KT), testosterone (T) and oestradiol (E2; oestradiol-17β), ranging between 4 and 79 ng L^{-1} in two rivers (the R. Test and R. Avon) in southern England (Wiltshire and Hampshire) in samples collected between 2002 and 2006. The results implicated trout farms as the main source of the steroids, as mean concentrations collected in water samples downstream were between 130% to 600% of reference “upstream” concentrations. The findings were reported in the national media (Mole, 2008) and discussed by the governmental environmental protection agency, environmental NGOs and trout farming trade body.

Focussing only on E2, the river water concentrations reported in the 2002-6 study (Anon, 2006) ranged from 2 to 120 ng L^{-1} (mean: 17.4; median: 8.5 ng L^{-1}; CV: 116%). If these concentrations were correct, they would have triggered widespread induction of vitellogenin production (i.e. estrogenic

ACCEPTED MANUSCRIPT

endocrine disruption) in wild fish. However, these concentrations were considerably higher than those measured in surface water (e.g. (Yao et al., 2018) mean 0.56; median 0.26; 90th percentile 1.6 ng L^{-1}) or estimated in wastewater treatment plant effluent (e.g. (An et al., 2018) 0.28-0.36 ng L^{-1}), the primary source of unwanted oestrogens in the environment. Another puzzling observation was that 11-KT (a key androgen in male teleost fish) appeared to be more abundant in river water than T or E2. This steroid is only synthesised by reproductively mature males, yet on one of the farms the trout were all-female (and predominantly immature).

Based on the improbability of there being such high concentrations of steroids in river water, a second study was arranged to test the hypothesis that the assay methodology was to blame – specifically that the Enzyme-Immunoassay (EIA) kits (frequently referred to as Enzyme-Linked Immuno-Sorbent Assay, ELISA, kits) were yielding "false positive" results due to "matrix effects" (i.e. non-steroid substances in the water interfering with the enzyme component of the kits). Steroids in environmental water samples typically have to be extracted and concentrated before assay and (Hanselman et al., 2004) suggested that humic substances in water that are simultaneously extracted may cause matrix effects, the extent of which depends upon the type of enzyme used in the EIA. Several EIA kits for E2 have been found to generate values 2 to 65 times higher than analytical techniques such as liquid chromatography with tandem mass spectrometry (LC-MS/MS) when used on environmental water samples (Hirobe et al., 2004). EIAs have also been reported to give values up to 3-fold higher than radioimmunoassay (RIA) when measuring steroids in blood plasma or whole body extracts (Sink et al., 2008).

In order to examine the hypothesis that "matrix effects" were to blame for the high steroid concentrations reported, we collected river water samples from the same sites as the original study and processed them as much as possible in the same way. We then divided the concentrated sample extracts into several aliquots that were distributed between four laboratories, so they could be immunoassayed ("blind") by EIA kits (as used in the original study) and RIA. Comparisons of assay results between studies and between laboratories highlighted that, rather than EIA kits being prone to matrix effects, results from both EIA and RIA are prone to a variety of human errors. The two studies exemplify how readily human errors (particularly during the various calculation stages) occur with immunoassay data, and that these errors may not be identified and corrected unless compared to independent results.

ACCEPTED MANUSCRIPT

MATERIALS & METHODS

Sampling

The 2002-6 study (Anon, 2006) reported river steroid levels at sampling points located close to flow-through trout farms on two UK rivers, the R. Test in Hampshire (50° 59' 28" N, 1° 30' 18" W) and R. Avon in Wiltshire (51° 14' 30" N, 1° 47' 26" W). The same locations have been used in other studies examining potential environmental impacts of trout farms (Waring et al., 2012). In the current study we collected water samples from these same locations in 2010 (referred to henceforth as the "2010 study").

The trout farm on the R. Test is an intensive farm, on-growing diploid all-female rainbow trout (*Oncorhynchus mykiss*) in concrete raceways for the table. The approximate size range of fish on site is 5 to 800 g and fish are harvested before becoming reproductively mature, although occasional fish are noted with ovarian development. On-site water treatment is restricted to supplemental oxygenation and swirl concentrators to reduce the suspended solids load in the outflow discharged to a side channel from/to the main river (Fig 1A).

The trout farm on the R. Avon is a low-intensity farm growing diploid, mixed sex brown trout (*Salmo trutta*) in earth ponds for restocking angling waters. Fish range between approximately 1 g and 1 kg (small tanks are in occasional use for fry <1 g) and reproductively mature fish are present. There is no on-site water treatment, and the outflow is discharged to a side channel from/to the main river (Fig 1B).

In the 2002-6 study, water samples were collected from sampling points in side-channels of the main river, one "downstream" of where the trout farm effluent was discharged and the other from a nearby independent reference channel, termed "upstream". These sampling points were included in the current (2010) study and additional samples were collected from supplementary sampling points directly in the fish farm inflow and outflow channels (Fig. 1). At the R. Avon trout farm which had three separate outflows, outflow samples were collected from the discharge channel from the pond containing the bulk of the fish on site.

Water was collected from each sampling site on five separate occasions between January and June 2010. River water was collected in rinsed 25 L high density polyethylene drums by wading into the channels. Potential contamination (from residues, disturbed sediment, sampling personnel) was avoided by rinsing the drum with river water prior to sample collection, filling the drum by pointing upstream into the flow, and wearing of gloves. Water samples were then returned to the laboratory for processing and extraction.

ACCEPTED MANUSCRIPT

Additional samples were collected or prepared in the laboratory to assess the performance of the extraction and assay procedures:

- Two large samples (one from each river) were sub-divided to generate three replicate samples to assess the precision (variability) associated with water steroid extraction and measurement;
- Two deionised water samples were included as "blank" samples;
- Two deionised water samples and two river water samples (one from each river) were "spiked" with a known amount of steroids (dissolved in 1 mL methanol) to assess the recovery efficiency and the performance of the assays; and
- Two river water samples (one from each river) had 1 mL methanol added as a "solvent control" for the steroid spiked samples.

The spiking of river and deionised water samples was conducted by a researcher not involved in the sample collection, extractions or assays. A stock solution was prepared containing a mixture of 11KT (0.498 ng μL^{-1}), T (0.501 ng μL^{-1}), E2 (0.503 ng μL^{-1}) and the synthetic steroid 17α-ethynylestradiol (EE2, 0.024 ng μL^{-1}). From this, exactly 1 mL was added to a measured water sample volume of 10 L to provide equivalent nominal concentrations of 49.8 ng(11KT) L^{-1} , 50.1 ng(T) L^{-1} , 50.3 ng(E2) L^{-1}, and 2.4 ng(EE2) L^{-1}. These concentrations were similar to those reported by Anon (2006) for the natural steroids, and a factor of 10 higher for EE2. The stock solution containing the mixture of steroids was retained for inclusion in the assays, along with a sample of un-spiked methanol.

Water sample processing and extraction

With the exception of being carried out on a larger scale (i.e. processing 12 L water to generate eight 1 L replicate samples, as opposed to single 1 L water samples), the extraction procedure was the same as that in the 2002-6 study. This was based upon the procedure advocated by (Rubio et al., 2004). Upon return to the laboratory, suspended solids were removed from the water samples by vacuum filtration through glass microfiber filters (1.2 µm, 150 mm diameter). The samples were then acidified (2 ml L^{-1} HCl (36%)) and stored overnight in the dark at 4°C. Steroids were extracted from the water samples using C18 solid phase extraction cartridges (SPEC; Varian Bond Elut; 1 g, 6 mL). Each SPEC was preconditioned with 5 mL methanol and rinsed with 10 mL of deionised water. For each water sample, eight sub-samples of approximately 1 L were extracted by pumping at around 10 mL min^{-1} through conditioned SPEC; the water passed through the SPEC was collected to determine the exact volume gravimetrically. The SPEC were washed with 5 mL de-ionised water, dried under vacuum for 1 min, washed with 5 mL hexane, and eluted with 5 mL dichloromethane into glass

ACCEPTED MANUSCRIPT

tubes. The dichloromethane was evaporated under nitrogen at 37°C, and the near dried residue re-dissolved in 1 mL of methanol and stored at -20˚C.

The eight subsample extracts of each sample were then combined and cleaned by passage through an aminopropyl SPEC (Varian Mega Bond Elut, 10 g, 60 mL) preconditioned with 50 mL methanol. The sample methanol (8 x 1 mL) and an additional 42 mL methanol rinse were passed through the aminopropyl SPEC under gravity at <3 mL min^{-1}. The collected methanol (50 mL) of each sample was reduced to approximately 6 mL under partial vacuum at 43°C using a rotary evaporator, transferred to graduated tubes, and made up to 8 mL (equating to 1 mL for each litre of river water originally extracted). The samples were then sub-divided into eight replicate 1 mL methanol aliquots. Methanol sample extracts were stored in capped, glass vials and held at -20°C between stages.

The sampling and processing produced eight replicate sets of 1 mL methanol aliquots of 44 samples for assay. The 44 samples compromised 42 extracted water samples (34 river water samples, 2 steroid spiked river water samples, 2 river water samples to which 1 mL methanol had been added before filtration, 2 deionised water samples, 2 steroid spiked deionised water samples), 1 sample of the methanol spiking solution, and 1 sample of the methanol used to prepare the spiking solution. Six sets of the samples were distributed for assaying (three sets by EIA and three sets by RIA) with two sets held in reserve in case of mishap.

Assaying

The samples were assayed by EIA and RIA at four independent laboratories in the UK by researchers with previous experience of immunoassays. Two laboratories conducted both EIA and RIA, whilst one laboratory conducted just EIA and one laboratory conducted just RIA. Prior to assay, the participating laboratories evaporated the methanol in the sample aliquots and reconstituted them in buffer used for assay. The EIAs were conducted using commercial kits based upon 96 well plates (Table 1). The RIAs were based upon established in-house assay techniques which varied between participants (Table 1). One of the RIA laboratories received radiolabels and antibodies from the lead laboratory, but RIA standards were sourced independently. The ranges of steroid concentration in the standard curves are provided (Table 1).

The lead laboratory undertook trial assays (EIA and RIA) prior to sample distribution to other participating laboratories to ensure that the limited amounts of material provided (1 mL extract per sample, equivalent to 1 L water sample, being provided for assay of 11KT, T, E2 and EE2) was

ACCEPTED MANUSCRIPT

sufficient. The other participants conducted the assays blind, i.e. samples were coded with a number that did not reveal the origin/nature of the sample. The lead laboratory provided generic advice on the range of suitable dilutions to minimise wastage of the limited samples and assay materials. There was no interaction between the participating laboratories, each acting independently.

Data analysis and statistical methods

Unless otherwise stated, results were $\log_{10}$ transformed before analysis.

Assessing potential differences between measurements derived from the two methods (EIA and RIA), conducted at different laboratories, is not a trivial task, and was assessed for each steroid by a variety of methods:

- Visual examination of box-and-whisker plots of the data for the individual laboratories and assay methods for each steroid.
- 3 factor ANOVA of assay method, participant [nested within assay method] and sample (Minitab General Linear Model). Individual values that fell below the lowest standards were treated as missing values.
- Averaging of the three values for each sample derived from each method (if individual measurements fell below the lowest standard, a value of 0 was assumed; if all three measurements fell below the lowest standard, then the sample was treated as a missing value) for correlation and regression. As the values are $\log_{10}$ transformed, departure of the intercept from 0 indicated a systematic difference between methods, and departure of the slope from 1 indicated a difference between methods that changes with steroid concentration.
- Assessment following methodology recommended by (Bland and Altman, 1986) (who pointed out flaws in simple correlation), i.e. averaging the three values derived from each method (as above), and regression of the difference between the methods (EIA – RIA) against the average ([EIA+RIA]/2). If there were no differences between measurements from the two methods, all points would lie on the line of equality (y=0; as the measurements are $\log_{10}$ transformed, this difference equates to a ratio of 1 for the untransformed values); if there were consistent differences, point lie above (or below) the line of equality; if the magnitude of the difference changes with concentration, there would be a positive or negative trend.

To assess the variability within each assay method, the coefficient of variation (CV, standard deviation / mean) for the three replicate measurements of each steroid in each sample was

ACCEPTED MANUSCRIPT

calculated (untransformed values used; values < minimum standard were treated as missing values). The CVs for the two assay methods were then compared by paired t-test (Minitab) for each steroid.

Accuracy is the degree of closeness of measurements of a quantity to its actual (true) value. The assessment of accuracy was limited to a single sample (methanol spike) for which a nominal expected value for each steroid was known. Accuracy was assessed for each assay method by expressing measured values as a percentage of expected values, and T-tests were used to compare to 100%.

Steroid concentration in the original water sample (ng L^{-1}) was calculated from the median amount (in ng) measured in the aliquots (≡ ng mL^{-1} as aliquot volumes were 1 mL) and the volume (in L) of the water sample that the aliquot related to (mean 1.030 L, range 0.868 – 1.129 L). In the absence of evidence that one assay technique gave more reliable or accurate results than the other, or one participant's results were more reliable than another's, the median extract concentration from the 6 values (≥3 values for EE2 due to many values below the EIA lowest standard) was used. The median, rather than the mean, was used intentionally to reduce the effect of possible outliers.

The recovery (%) of the steroids from the water samples was assessed by comparing the median steroid concentration in the spiked methanol stocks to the concentrations in the extracts of the spiked water samples (2 deionised, 1 R. Test, 1 R. Avon) after subtraction of the amounts measured in un-spiked water samples. During rotary evaporation of the extract, accidental transfer between two of these samples occurred (denoted as D and E) without any loss, which was noted at the time of occurrence. Two factor ANOVA (Minitab) was used to assess the effects of steroid (11KT, T, E2 and EE2) and water (deionised v river water) on recovery.

The reproducibility of the extraction (and assay) methodology was assessed for the two river water samples that were processed and assayed in triplicate, and expressed as the coefficient of variation (CV). The CV values (2 per steroid) were compared between the four steroids by One-Way ANOVA (Minitab).

The river steroid concentrations (ng L^{-1}, $\log_{10}$ transformed) were assessed for effects of river (R. Test v R. Avon), season (Winter = Jan/Feb; Spring = April-June) and sampling station (Reference, Downstream, Fish farm inflow, Fish farm outflow) for each of the four steroids by 3-way ANOVA (General Linear Model, Minitab). Of the 44 samples assayed, the 10 quality control samples were excluded from this analysis, i.e. blank methanol (n=1), spike methanol (n=1), deionised water (n=2),

ACCEPTED MANUSCRIPT

river water samples with blank methanol added (n=2), spiked deionised water (n=2) and spiked river water samples (n=2). Due to limited degrees of freedom for the analyses, sampling occasion was not included and neither was the 3-way interaction; 2-way interactions were tested in sequential analyses and dropped when non-significant. Multiple testing (i.e. 4 separate analyses for the different steroids measured in the same samples) does increase the probability of Type I errors, i.e. rejecting the null hypothesis when it is correct (Petrie and Sabin, 2000). As probability correction methods are considered conservative, the approach was adopted of judging the probability levels across all four analyses.

The 3-way ANOVA method lacks sensitivity for assessing whether fish farms do increase river steroid levels as effects may be masked, e.g. by differences between sampling occasions and rivers. Therefore, an additional paired t-test of the steroid levels in matched (by sampling occasion) fish farm inflow and outflow samples was conducted when judged appropriate.

RESULTS

EIA and RIA measurements of steroid concentrations in sample extract aliquots

The participating laboratories reported to the lead laboratory the concentrations of each of the four steroids measured in the coded 1 mL aliquots (ng mL^{-1}) and associated water samples (ng L^{-1}). Therefore, three measurements for each sample for each assay technique were available, making anomalous results apparent for some samples:

- One participant reported EE2 EIA measurements for 5 samples that were derived by extrapolation below the minimum standard. The other two laboratories had reported n.d. (not detected) or <0.05 ng mL^{-1} (the lowest standard concentration) for these aliquots.
- One participant reported EE2 EIA measurements for the six spiked samples that were between 3 and 10 fold lower than the other participants. It was established that the absorbance data had been entered into the same homemade spreadsheet that had been used in the 2002-6 study and then calculated using the slope and intercept values derived from a previous standard curve of unknown origin. Further confusion was created by the fact that the EE2 kit standards were in picograms (pg), whereas the standard curve was labelled ng L^{-1}.
- One participant reported EE2 RIA measurements for the six spiked samples that were 10 to 25 fold higher than the other participants. It was subsequently established that these samples had been assayed at various dilutions (as suggested by the lead laboratory), then between the assay printout and the calculation spreadsheet the sample order had been transposed, resulting in incorrect dilution factors being applied during calculation. The error was traceable from the contradictions between the measured radioactivity and the initially assigned dilutions.

ACCEPTED MANUSCRIPT

These anomalous results were corrected before further analysis.

The measured steroid concentrations in the river water (and blank) sample aliquots are presented separately to those from the spiked samples due to the 10 to 100-fold difference in values (Fig 2). The plots indicate differences between the assay methods and the laboratories which are supported by ANOVA (Table 2).

Although there were very strong correlations ($r \geq 0.98$; $p < 0.001$) between the EIA and the RIA results for all four steroids, the regression intercept differed to the origin in all cases (Figure 3, Table 3), indicating systematic differences between EIA and RIA values. Furthermore, the regression slope differed from unity for 11KT and E2, indicating that differences changed with steroid concentration.

The Bland-Altman approach (Table 3; Figure 4) also corroborated systematic differences between the two assay methods for all four steroids (the intercept differed from 0 for all four steroids), and that the difference changed with steroid concentration for 11KT and E2 (i.e. the slope was significant). The average of the differences between the logged EIA and RIA values were -0.43, -0.19, -0.06, -0.28 for 11KT, T, E2 and EE2 respectively; these equate to average ratios (EIA value: RIA value not transformed) of 37%, 65%, 116% and 53%.

The inter-laboratory variability associated with two assay methods differed between the four steroids (Table 4): RIA produced more variable results than EIA for 11KT, but less variable results for T and E2. There were insufficient EIA measurements for EE2 to enable a valid comparison.

Accuracies ranged between 42% and 128% (Figure 5) for 11KT, T and E2, with means between 74% and 97% which were not different to 100% for either assay method (Table 5). Accuracies for EE2 were notably higher (86-332%; Figure 5), but the means (159%, 248% for EIA and RIA respectively) were again not different to 100% (Table 5), probably due to the variation and limited number of observations.

Steroid concentrations in water samples

Within this (2010) study, all participating laboratories correctly transformed their measured extract concentration to a water steroid concentration using the equivalent water volume. As explained above, a median extract concentration from the 6 values (≥3 values for EE2 due to many values below the EIA lowest standard) was used to calculate the steroid concentration (ng L^{-1}) in the original water sample.

ACCEPTED MANUSCRIPT

Reproducibility for the extraction and assay methods, assessed from the two large water samples each sub-divided, processed and extracted into three replicate samples, was judged as good for 11KT, T, and E2 which had similar CVs in the range of 3 - 11% (Table 6). [Please note that these CV values were derived from the medians of the ≤6 values from RIA and EIA]. However, variation was much higher ($P<0.001$) for EE2 (Table 6; CVs: 58%, 67%) than the other three steroids.

Calculated recoveries derived from the spiked water samples varied between 25% and 82% (Table 7). Recovery differed between the individual steroids ($p<0.001$; 11KT<T<E2=EE2) but was not affected by the origin of the water ($p=0.329$; deionised= river water). Calculated water steroid concentrations were not corrected for recovery, due to the limited recovery data, the inherent variability between samples (and it being compromised by accidental methanol transfer between two samples). Furthermore, this decision facilitates comparison with the 2002-6 study data which were not corrected for recovery efficiency.

In the river water samples, T was present at the highest concentrations (0.13-0.58 ng L^{-1}), followed by E2 (0.14-0.30 ng L^{-1}), EE2 (0.03-0.22 ng L^{-1}) and 11KT (0.01-0.03 ng L^{-1}) (Table 8). This ranking would remain if the concentrations were to be corrected for recovery efficiency which was lower for T than for E2. When assessed for effects of river, season and sampling station, only season was significant ($P<0.001$ for 11KT and T, and approaching significance ($P<0.01$) for E2 and EE2; Table 9), with higher concentrations in the spring than winter samples (Table 8). River and sampling station were not significant (Table 9). Nevertheless, sampling station did approach significance for T ($P<0.1$) and the data (Table 8) suggested higher T concentrations in the fish farm outflow and downstream samples on the R. Test. An increase in water testosterone by the R. Test fish farm was supported by a paired t-test of inflow v outflow concentrations ($n=4$, $T= -5.00$, $p = 0.015$).

DISCUSSION

We carried out the 2010 study to test the hypothesis that substances co-extracted with steroids from river water caused matrix effects (i.e. “false positives”) during steroid measurement by EIA; it was not designed to validate steroid extraction and measurement methodologies nor as an inter-laboratory comparison (Heath et al., 2010; McMaster et al., 2001). The matrix effect hypothesis was based upon abnormally high river steroid concentrations reported in the prior 2002-6 study (Anon, 2006) . However, concentrations measured by EIA in our 2010 study were low and similar (37-116%) to those measured by RIA. There was therefore no evidence for matrix effects in any of the 38 river water sample extracts, collected from two rivers on five occasions over a 6 month period. We have therefore rejected the matrix effect hypothesis as an explanation for the abnormally high river steroid concentrations reported by (Anon, 2006).

ACCEPTED MANUSCRIPT

Calculation error to explain the unreproducible steroid concentrations in the 2002-6 study

River steroid concentrations in samples collected from the same points at the same times of year were far lower in the 2010 study than the 2002-6 study (Anon, 2006): levels of 11KT, T, E2 and EE2 were respectively 1300, 70, 90 and 3 times lower. The rank order was therefore also different, with concentrations of 11KT being lowest in the 2010 study, but highest in the 2002-6 study.

Water quality data for a variety of parameters (concentrations of chloride, ammonia, nitrate, nitrite, total oxidised nitrogen, dissolved oxygen, suspended solids; biological oxygen demand; pH; temperature; turbidity) were provided by the Environment Agency for sampling stations around the two sites. The only difference evident between the two periods was at the R. Test site where nitrite concentrations decreased (both upstream of the trout farm and in its outlet), likely reflecting reduced inputs to the river above the site.

Various alternate hypotheses were considered that could potentially explain the difference in water steroid concentrations between the two studies, i.e. potential changes in:

- field factors - sources of steroids (e.g. sewage treatment works, wildlife, farm livestock), river flows and dilution, concentrations of interfering chemicals, timing of sampling, contamination by sampling personnel.
- laboratory factors – quality of EIA kits, sampling and storage procedures, extraction efficiency, contamination in the laboratory.

However, there was no evidence of major changes or differences to support any of these hypotheses. We therefore conclude that calculation error was the probable cause of the high concentrations reported in the 2002-6 study. This conclusion was based upon calculation errors being made by two out of the four laboratories in the present study and the lack of an alternative explanation. Calculation errors have previously been reported: three out of seven laboratories in a similar blind study (Feswick et al., 2014), and two out of seven in another study (McMaster et al., 2001).

As in the 2010 study, water samples in the 2002-6 study were concentrated 1000 times, i.e. 1 L water was reduced to a 1 mL extract. It is the extract that is subjected to assay (not the water itself). If the real river water concentrations in the 2002-6 study extracts were similar to those in the 2010 study then, for steroid concentrations to fall on the accurate part of the standard curves, the workers in the earlier study would likely have chosen to pre-dilute the extracts by 1:10 for the T and E2 EIAs, but used without any dilution for the 11KT EIA. The output values from calculations using the standard curves normally represent the concentrations in the assayed extract aliquot, not the

ACCEPTED MANUSCRIPT

concentrations in the water sample. In order to obtain the water concentrations, aliquot concentrations need to be converted, in the above case divided by 100 for T and E2 and 1000 for 11-KT. We propose that this conversion step was missed out in the 2002-6 study because if retrospectively applied, water concentrations correspond to those found in the present 2010 study.

This type of error (i.e. failure to convert extract concentrations into water concentrations) was also recorded in the study by (Feswick et al., 2014). Unfortunately, although this seems the most likely explanation for the difference between studies, we were unable to confirm it as no data processing spreadsheets or written "lab book" records (e.g. of sample extract dilutions made up for assay) had been retained from the 2002-6 study. The failure to retain paper and electronic records has been recognised as a mistake. Retention of records is now one of the requirements of the UK's Joint Code of Practice for Research (Defra, 2015).

Additional errors to explain other unreproducible conclusions of the 2002-6 study

The 2002-6 study (Anon, 2006) concluded that: the trout farms were a major source of steroids with 11KT, T and E2 being 1.3 to 6 times higher downstream than upstream of both subject trout farms; and river steroid levels were higher in winter than spring. The 2010 study data contradicted these deductions, indicating that: the trout farms were not a major source of steroids; and river steroid concentrations were higher in spring than in winter.

There was no evidence from farm records (on fish stocks and management) and regulatory body data (on water quality changes between the fish farm inflow and outflow) for any marked changes in fish farm practices between the two study periods that could account for the observed differences. Seasonal river flows and dilution are now recognised as a major factor affecting river steroid levels (Johnson, 2010). Mean gauged daily flow data were obtained for the nearest gauging stations (Centre for Ecology and Hydrology's National River Flow Archive http://www.ceh.ac.uk/data/nrfa/data/retrievals.html and supplied by Richard J Williams, CEH). River flow rates were significantly lower on the dates of the spring sampling than the winter sampling in both 2010 (GLM ANOVA $p=0.003$; Least Squares Mean 11.1 v 6.9 $m^3 s^{-1}$) and 2002-6 (GLM ANOVA $p=0.041$; Least Squares Mean 8.1 v 6.2 $m^3 s^{-1}$).

Other errors that might have led to different data, and therefore conclusions, between studies include:

- inter-assay variation: if compared samples were assayed on different EIA plates;

ACCEPTED MANUSCRIPT

- intra-assay variation: if compared samples were separated within different areas of the same EIA plate(s) that produced differing results (e.g. due to localised evaporation);
- poor parallelism: if compared samples were assayed at different dilutions, and values fell on different (sensitive v insensitive/extreme) parts of the standard curve;
- calculation errors: if compared samples were assayed at different dilutions and the necessary correction was omitted from calculations; and
- bias (unconscious or conscious): if divergent assay results from sample replicates were selected based upon expectation.

As above, these potential errors could not be explored due to the lack of records.

Further potential sources of error in steroid studies

The production of steroid information involves various successive stages - sampling, sample storage, sample extraction, sample clean-up, assaying, calculation, statistical analysis, interpretation and presentation. Researchers and users of assay results need to be alert to potential sources of error at all stages.

Sampling: The samples taken in the two studies described here are point samples, with the implicit assumption of homogeneity in time and space. The similarity between measurements at the different sampling points and over time in the 2010 study indicate that this assumption is fair for these river systems, but would need to be confirmed for other sites.

Storage: There are inevitable delays between sampling and assay during which samples could deteriorate. Practices are typically introduced to prevent deterioration, e.g. cold / freeze storage of water / extract samples. Studies typically do not include any assessment of losses during storage, indicating that practices adopted assume negligible loss, which merits confirmation.

Extraction & clean-up: Environmental samples typically require extraction and clean-up before assay. Within the 2010 study, samples were cleaned before and after C18 extraction, by GFC filtration and amino-propyl SPE respectively. The limited amount of data indicates that considerable steroid loss occurred during these processes which differed between steroids: 70% for 11KT, 50% for T and 30% for E2 and EE2. Extraction efficiency was not reported in the 2002-6 study. One question that remains to be answered is why the extraction efficiency of 11-KT (and to a lesser extent T) was worse than that of the oestrogens. A small experiment (unpublished) suggested that the losses occurred during the aminopropyl clean-up stage (i.e. unlike the oestrogens, the androgens tended to

ACCEPTED MANUSCRIPT

bind to the aminopropyl rather than pass through). It must be pointed out that the aminopropyl clean-up stage was never recommended for androgens (only for E2). It was only used in the 2010 study because it was used in the 2002-6 study.

Contamination: Contamination can occur at any stage during sample collection, processing (extraction and clean-up) and assaying. Sediment disturbance during sampling was recognised (and avoided) as a potential source of contamination in the 2010 study. Steroids are released from human skin which could contaminate water samples during collection, processing and assay (Ellis et al., 2013). Human contamination can also occur via exhaled breath and saliva and will depend upon the physiological state of technicians (Abraham, 1975; Peppler and Stone, 1979). Cross-contamination within and between samples and standards can also occur. Although appropriate procedures and practices should prevent such contamination, accidents do occur as illustrated during the 2010 study:

- cross-contamination of samples was noted due to accidental transfer during rotary evaporation of methanol extracts; and
- contamination of the blank methanol sample with E2 and EE2 (but not 11KT or T) – it was not possible to determine how this occurred but was likely to be due to a contaminated pipette tip.

Assay: The 2010 study demonstrated consistent differences in measured (after correction for identified calculation errors) values between laboratories using the same technique, and between assay techniques. These differences also represent potential sources of error. Examination of the causes of such differences would require more intensive research. The 2010 study did demonstrate that mistakes can potentially be made in production of standards: the ca. 200% accuracy value for EE2 indicates a mistake in the preparation of the "standard" solution used for spiking.

Both EIA and RIA methods registered false, albeit low, concentrations for extracts of deionised water. False positives for over-diluted plasma sample extracts are thought to explain some odd low plasma cortisol values (<0.5 ng mL^{-1}) reported by (Wong et al., 2008). This reflects the problem of the unreliability of sample values that fall within the standard curve, but on flatter sections. The CVs of the measurements for the various RIAs and EIAs (Table 4) illustrate this issue. Water extracts assayed with the 11-KT RIAs had a higher CV than when assayed with the EIA, due to a combination of low overall amounts of steroid and differences in the sensitivity of the two assays. In the RIA, the samples were close to the lowest standard and thus fell on the visually flatter inaccurate part of the (sigmoid) standard curve, while in the EIA they fell on the steep accurate part. Although the T and

ACCEPTED MANUSCRIPT

E2 RIAs were also less sensitive than their EIA counterparts, the steroids were present in higher amounts (which meant that their values fell on the accurate part of the standard curve). In the case of EE2, both assays were relatively insensitive and there was little or none of this steroid in any of the extracts (thus again generating high CVs).

Whether interpolated values from the flatter parts of standard curves are used, rejected, or samples are re-assayed after further dilution /concentration is dependent upon the judgement of the operator and their constraints. This is true for in-house RIAs, and commercial kit EIAs for which guidance booklets may advise users to use such values with caution rather than with confidence. As illustrated by our 2010 study, users of EIAs and RIAs may not sufficiently consider the sensitivity of the assay for their particular samples/extracts. More attention needs to be given to determining the Limit of Blank (LoB, the highest apparent analyte concentration of blank samples), Limit of Detection (LoD, the lowest analyte concentration distinguishable from the LoB) and Limit of Quantitation (LoQ, the lowest analyte concentration that can be quantified) (Armbruster and Pry, 2008), as well to values lying at the opposite end of a sigmoid standard curve.

There is also the question of whether assay immunoactivity truly represents the target steroid. Antibodies are rarely, if ever, 100% specific for the target steroid; they all cross-react with other closely-related steroids to a greater or lesser extent (Abraham, 1975) and may also have their affinity altered by other compounds in the extracts (i.e. the “matrix effect”). In the 2010 study, it was planned to partially characterise the steroids by chromatography (i.e. checking that the steroid immunoactivity behaves chromatographically like its standard); however, this approach was abandoned when the very low amount of steroid in the water samples was revealed.

Calculation: It appears that the calculation stage (involving spreadsheets) is where errors are most likely to occur. The 2002-6 and 2010 studies provide examples of:

- transposition of samples between the assay and data work-up;
- using the wrong standard curve;
- extrapolation beyond the lowest standard, i.e. reporting of data derived from below the minimum of the standard curve. Aliquot steroid concentrations should only be estimated by interpolation within the standard curve if radioactivity / optical density measurements fall between the minimum and maximum standards;
- failure to correct values produced by the assay spreadsheets for the water concentration step;
- and failure to account for loss during extraction and clean-up (which, in the case of 11-KT led to a 3 to 4-fold underestimation of concentrations in the river water)

ACCEPTED MANUSCRIPT

(Feswick et al., 2014) reported several similar calculation and loss errors.

Statistical analysis: The 2010 study has illustrated how different statistical tests can indicate different conclusions, e.g. the 3-factor ANOVA versus paired t-test of the impact of fish farms on river T levels. Another statistical issue is that EIA and RIA commonly produce values that fall below the limits of the assay, and the handling of such values is problematic. The values selected to represent such samples (e.g. 0 ng L^{-1} or a nominal low concentration) could affect the results of statistical analysis and researchers need to be aware or use appropriate statistical methods (e.g. non-parametric). An alternative approach, of excluding such values from analyses, e.g. (Sundh et al., 2010), facilitates statistical testing but averages will be artificially elevated.

Presentation and interpretation: Mistakes can be made during presentation of results. For example, in (Santulli et al., 1999), the units for plasma cortisol concentration differed for the same numerical values between the results table (µmol/100 mL) and text (µg/100 mL) which represents a 362-fold difference in concentration. The 2002-6 study (Anon, 2006) concluded that there were significant steroid outputs from trout farms based on samples from the reference and downstream stations. For brevity the reference stations were termed "upstream" and the textual description indicated that they were on the same channels as the downstream stations and located on the main rivers. However, the reference and downstream sampling stations were in fact located on different channels (and therefore potentially differed in inputs other than from the trout farm) and the samples were collected from side channels rather than the main river (Figure 1). Such brief text descriptions can mislead and obscure relevant information that affects reader interpretation of reported results.

Gross errors can also occur from malfunction of measurement equipment and use of faulty or expired reagents (Feswick et al., 2014), although such sources of error are likely to be recognised by experienced researchers. Additional minor errors will inevitably creep into assay results associated with:

- imprecision in pipetting of sample extracts, standards and dilutions; and
- use of radioactivity or optical density measurements to estimate, via the standard curve, the amount of steroid in the assayed aliquot; this value is then scaled-up to estimate the amount of steroid in the original sample using a multiplication factor.

It is these smaller errors that affect the accuracy and precision of the techniques.

ACCEPTED MANUSCRIPT

The potential for errors in immunoassays does not negate the value of their measurements - it just illustrates that results should be viewed with caution unless there is strict quality control and until confirmed independently. Steroid immunoassays remain a key tool in clinical laboratories, where they are used routinely by highly experienced technicians. Research use should include safeguards such as: assay of all samples at the same time and with the same set of reagents to minimise inter-assay variation; typical assay validation (e.g. examination of cross-reactivity, accuracy, inter and intra-variation, parallelism) and quality control samples (spikes, blanks, replicates) to document extraction / recovery efficiency and reproducibility (Abraham, 1975); determination of LoB, LoD and LoQ for the particular samples/extracts (Armbruster and Pry, 2008); "blind" assaying to eliminate the possibility of bias in selecting reported data; **independent checking of results**, especially important for inexperienced personnel; **retention** of raw assay data and associated calculations used to derive final steroid concentration values in case of queries. (Wudy et al., 2018) have recently emphasised the importance of expertise and thorough training of staff in all steroid assays to ensure quality results.

It is notable that in two inter-laboratory comparisons: (McMaster et al., 2001) reported highly variable T and E results for the same samples measured by RIA, 0.6-23.1 ng mL^{-1} and 26-317 pg mL^{-1} respectively; (Heath et al., 2010) identified extreme outliers (i.e. likely errors) in measurements of water steroids by gas and liquid chromatography with mass spectrometry (GC-MS and LC-MS/MS), and indicated that these arose because not all participating laboratories were proficient in the techniques. Hannah et al., 2009 compiled a database of EE2 concentrations recorded in surface waters from 16 countries (n= 1,652) and they ranged from < 0.1 ng L^{-1} to 273 ng L^{-1}. Any values higher than 1 ng L^{-1} (which was 16% of them) were deemed to be improbable based on the amounts of EE2 known to be used in those countries. One implication of all these studies (including the present one) is that ca. 30% of published steroid values are possibly wrong (and arguably 10% are grossly wrong). Unfortunately, there is no easy way to determine which studies these are.

Should we be concerned about sex steroid outputs from freshwater trout farms?

In a key publication on future food security, (Godfray et al., 2010) highlighted the need to expand aquaculture, while being alert to potential harms to the environment via effluents. The R. Test and R. Avon provide important chalk stream habitats in southern England for native brown trout and Atlantic salmon (*Salmo salar*) (Ikediashi et al., 2018). Concerns have therefore been raised about the environmental impacts of trout farm effluents on such migratory fish in these rivers, e.g. (Mole, 2008). Fish farm effluents contain particulate and dissolved wastes that can exert sedimentation, deoxygenation and eutrophication pressures on the receiving water body. Such environmental

ACCEPTED MANUSCRIPT

pressures have been well documented and regulatory controls of fish farm effluents are now in place to mitigate associated environmental impacts (Bergheim and Brinker, 2003). However, concerns have also been raised that fish farm effluents may also contain other (unregulated) biogenic substances, such as steroids, that may affect the behaviour and physiology of wildlife downstream (Kamps and Neill, 1999). Such theoretical concerns delayed the development of a fish farm in Denmark when the "precautionary principle" of environmental protection was applied (Anon, 2015).

Fish farms are certainly a potential source of steroid input to natural water bodies: (Mota et al., 2014) reported steroids accumulating in the water of recirculating aquaculture systems, growing a variety of (non-salmonid) fish species. These steroids may be:

- endogenous, produced *de novo* by the fish stock and released into the water via branchial, urinary or biliary routes, in free and conjugated forms (Vermeirssen and Scott, 1996). Sex steroid production is expected to be associated with maturing and mature fish, and be seasonal (Baynes and Scott, 1985; Scott et al., 1980);
- exogenous, originating from treatments used to produce sex-reversed broodstock, e.g. (Feist et al., 1995); and
- exogenous, originating from feeds. Several studies have shown that formulated diets fed to cultivated fish contain residues of vertebrate steroids (Davis et al., 2009; Feist and Schreck, 1990; Pelissero et al., 1989; Pelissero and Sumpter, 1992; Sower and Iwamoto, 1985) or show steroidal activity (Beresford et al., 2011). The steroids could be released from uneaten food, faeces or by fish after absorption via the gut.

Kolodziej et al., 2004 measured steroids (estrone, T and androstenedione) in US rainbow/steelhead trout farm effluents at concentrations (0.1 - 0.8 ng L^{-1}) similar to (natural) levels measured in a wild salmonid spawning ground. Two studies in Israel have reported higher steroid concentrations (1-10 ng L^{-1} of total estrogen (E1+E2)) in fish pond effluent (Barel-Cohen et al., 2006; Shore et al., 2004). However, the fish species was not mentioned and is considered unlikely to be salmonid, as farmed production in Israel is minor compared to that of carps, tilapias and mullet (FAO, 2018).

There was no evidence from our 2010 study that either farm elevated the levels of 11KT and E2 in the rivers; nor was there evidence that the R. Avon fish farm increased T levels. However, the intensive rainbow trout farm on the R. Test probably did marginally increase the T concentration in the outflow and therefore also in the downstream channel. The R. Test fish farm did not produce

ACCEPTED MANUSCRIPT

broodstock, and although this farm held only female rainbow trout, females do produce noteworthy amounts of T (Scott et al., 1980; Sumpter et al., 1984). The testosterone could therefore have originated from the fish themselves, and/or the feed.

The important question to answer is whether the T output pressure from the farm would result in an environmental impact, which will depend upon dilution in the receiving waters, the sensitivity of resident species, and the rate of degradation. Local concentrations in the outflow of the farm were <1 ng L^{-1} (even if corrected for 50% recovery), and these would then be diluted downstream. T output from the farm is therefore unlikely to have had a direct endocrinological effect on wild fish. Salmonids are known to be highly sensitive to water-borne olfactory signals (Selset and Døving, 1980). It has been reported that Atlantic salmon parr can sense T which acts as a behavioural attractant at concentrations as low as 10^{-14} M (Moore and Scott, 1991), equivalent to 0.003 ng L^{-1} (Moore, 1991). However, this evidence of pheromonal activity (and that for a conjugated progestogen) (Moore and Scott, 1992) has recently been highlighted as needing independent replication (Stacey, 2015). The available evidence therefore indicates that flow-through trout farms are a minor contributor to river sex steroid levels relative to other sources, and measured levels are unlikely to impact on wildlife.

It must be recognised that the free sex steroids that are measured by EIA and RIA are likely to be minor components of total sex steroid released in trout farm effluent, with conjugated steroids likely to dominate (Mota et al., 2014; Vermeirssen and Scott, 1996). The amounts of conjugated sex steroids released from trout farms (and other sources), their behavioural dynamics (persistence and de-conjugation), and biological activity are unknown. Trout farm effluents are also likely to contain free and conjugated corticosteroids (Mota et al., 2014; Pottinger et al., 1992), damage-released chemical alarm cues and handling-related semiochemicals (Barcellos et al., 2011; Leduc et al., 2004), and prostaglandins. [F-type prostaglandin is thought to act as a pheromone in salmonids (Moore et al., 2002; Moore and Waring, 1996; Stacey, 2015). F-type prostaglandins were reported in the 2002-6 Study (Anon, 2006) at 1-3 ng L^{-1} downstream of the subject trout farms; however, as this compound was measured by EIA in the same way as the sex steroids, one must be cautious about these values.] Nevertheless, no evidence documenting release and impacts of such biogenic substances in fish farm effluents appears to have emerged since (Kamps and Neill, 1999) highlighted the theoretical concern.

Conclusions

Immunoassay results differed between EIA and RIA techniques and laboratories. However, such differences were minor relative to the errors that were introduced by miscalculation.

ACCEPTED MANUSCRIPT

Unexpectedly, calculation errors in immunoassay results appear to be common, albeit easily avoidable, so results should be viewed with caution unless there is strict quality control and independent replication. Although trout farms effluents represent a theoretical source of environmental steroids (and other biogenic compounds), there is currently no evidence that these have an adverse impact on wildlife.

Acknowledgements

This work was funded by Defra under project CB0427 Accuracy of methods of sex steroid determination. Thanks to: the farmers and land-owners for their cooperation and discussions; Dr Colin Waring and staff at the University of Portsmouth (School of Biological Sciences, Institute of Marine Sciences Laboratories) for practical assistance; Richard Williams (Centre for Ecology and Hydrology) and the National River Flow Archive http://www.ceh.ac.uk/data/nrfa/data/retrievals.html) for supplying river flow data; John Elliott (Environment Agency, EA) for provision of maps of the R Test river system; the EA National Requests Team for supplying water quality data for EA sampling points at the field sites; and Ioanna Katsiadaki for critical review. The literature examples of errors in assay results were simply from previous incidental encounters. Any references to products or services do not imply or constitute endorsement or recommendation.

References

Abraham, G.E., 1975. Radioimmunoassay of steroids in biological fluids. J. Steroid Biochem. 6, 261–270. https://doi.org/10.1530/acta.0.065S320

An, J., Choi, K., Yang, S., Nam, K., 2018. Estimation of human-origin estrone and 17 β -estradiol concentrations in the Han River , Seoul , South Korea and its uncertainty-based ecological risk characterization. Sci. Total Environ. 633, 1148–1155. https://doi.org/10.1016/j.scitotenv.2018.03.248

Anon, 2015. Evidence gathering questionnaire for the fitness check of the Nature Directives [WWW Document]. URL http://ec.europa.eu/environment/nature/legislation/fitness_check/evidence_gathering/docs/EU Level Organisations/FEAP/EU Organisation - FEAP - EGQ.pdf?_sm_au_=iVVZ5Vt6HQj60KHr (accessed 7.12.18).

Anon, 2006. SF0241 Impact of intensive in-river aquaculture on wild salmonids [WWW Document]. URL http://randd.defra.gov.uk/Document.aspx?Document=SF0241_7578_FRP.doc (accessed 7.12.18).

Armbruster, D.A., Pry, T., 2008. Limit of blank, limit of detection and limit of quantitation. Clin. Biochem. Rev. 29, 49–52.

Baker, M., 2016. Is there a reproducibility crisis? Nature 533, 452–454.

ACCEPTED MANUSCRIPT

https://doi.org/10.1038/533452A

Balaam, J.L., Grover, D., Johnson, A.C., Jürgens, M., Readman, J., Smith, A.J., White, S., Williams, R., Zhou, J.L., 2010. The use of modelling to predict levels of estrogens in a river catchment: How does modelled data compare with chemical analysis and in vitro yeast assay results? Sci. Total Environ. 408, 4826–4832. https://doi.org/10.1016/J.SCITOTENV.2010.07.019

Barcellos, L.J.G., Volpato, G.L., Barreto, R.E., Coldebella, I., Ferreira, D., 2011. Chemical communication of handling stress in fish. Physiol. Behav. 103, 372–375. https://doi.org/10.1016/J.PHYSBEH.2011.03.009

Barel-Cohen, K., Shore, L.S., Shemesh, M., Wenzel, A., Mueller, J., Kronfeld-Schor, N., 2006. Monitoring of natural and synthetic hormones in a polluted river. J. Environ. Manage. 78, 16–23. https://doi.org/10.1016/J.JENVMAN.2005.04.006

Baynes, S.M., Scott, A.P., 1985. Seasonal variations in parameters of milt production and in plasma concentration of sex steroids of male rainbow trout (Salmo gairdneri). Gen. Comp. Endocrinol. 57, 150–160. https://doi.org/10.1016/0016-6480(85)90211-4

Begley, C.G., 2013. Six red flags for suspect work. Nature 497, 433–434. https://doi.org/10.1038/497433a

Beresford, N., Brian, J. V., Runnalls, T.J., Sumpter, J.P., Jobling, S., 2011. Estrogenic activity of tropical fish food can alter baseline vitellogenin concentrations in male fathead minnow (Pimephales promelas). Environ. Toxicol. Chem. 30, 1139–1145. https://doi.org/10.1002/etc.479

Bergheim, A., Brinker, A., 2003. Effluent treatment for flow through systems and European Environmental Regulations. Aquac. Eng. 27, 61–77. https://doi.org/10.1016/S0144-8609(02)00041-9

Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 327, 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Davis, L.K., Fox, B.K., Lim, C., Hiramatsu, N., Sullivan, C. V., Hirano, T., Grau, E.G., 2009. Induction of vitellogenin production in male tilapia (Oreochromis mossambicus) by commercial fish diets. Comp. Biochem. Physiol. Part A Mol. Integr. Physiol. 154, 249–254. https://doi.org/10.1016/J.CBPA.2009.06.009

Defra, 2015. Joint Code of Practice for Research [WWW Document]. https://doi.org/http://www.defra.gov.uk/science/documents/QACoP_V8.pdf

Ellis, T., Sanders, M.B., Scott, A.P., 2013. Non-invasive monitoring of steroids in fishes. Wien. Tierarztl. Monatsschr. 100, 255–269. https://doi.org/urn:nbn:nl:ui:11-dbi/52820484b78fe

FAO, 2018. FishStatJ - software for fishery statistical time series [WWW Document]. URL http://www.fao.org/fishery/statistics/software/fishstatj/en (accessed 7.12.18).

Feist, G., Schreck, C.B., 1990. Hormonal content of commercial fish diets and of young coho salmon (Oncorhynchus kisutch) fed these diets. Aquaculture 86, 63–75. https://doi.org/10.1016/0044-8486(90)90222-9

Feist, G., Yeoh, C.-G., Fitzpatrick, M.S., Schreck, C.B., 1995. The production of functional sex-reversed male rainbow trout with 17α-methyltestosterone and 11 β-hydroxyandrostenedione. Aquaculture 131, 145–152. https://doi.org/10.1016/0044-8486(94)00336-M

ACCEPTED MANUSCRIPT

Feswick, A., Ankley, G.T., Denslow, N., Ellestad, L.E., Fuzzen, M., Jensen, K.M., Kroll, K., Lister, A., Maclatchy, D.L., Mcmaster, M.E., Orlando, E.F., Servos, M.R., Tetreault, G.R., Van Den Heuvel, M.R., Munkittrick, K.R., 2014. An inter-laboratory study on the variability in measured concentrations of 17β-estradiol, testosterone, and 11-ketotestosterone in white sucker: Implications and recommendations. Environ. Toxicol. Chem. 33, 847–857. https://doi.org/10.1002/etc.2503

Godfray, H.C.., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J., Pretty, J., Robinson, S., Thomas, S., Toulmin, C., 2010. Food security: the challenge of feeding 9 billion people. Science (80-.). 327, 812–818. https://doi.org/10.4337/9780857939388

Hannah, R., D'Aco, V.J., Anderson, P.D., Buzby, M.E., Caldwell, D.J., Cunningham, V.L., Ericson, J.F., Johnson, A.C., Parke, N.J., Samuelian, J.H., Sumpter, J.P., 2009. Exposure assessment of 17α-ethinylestradiol in surface waters of the United States and Europe. Environ. Toxicol. Chem. 28, 2725–2732.

Hanselman, T.A., Graetz, D.A., Wilkie, A.C., 2004. Comparison of three enzyme immunoassays for measuring 17β-Estradiol in flushed dairy manure wastewater. J. Environ. Qual. 33, 1919–1923. https://doi.org/10.2134/jeq2004.1919

Heath, E., Kosjek, T., Andersen, H.R., Holten Lützhøft, H.-C., Adolfson Erici, M., Coquery, M., Düring, R.-A., Gans, O., Guignard, C., Karlsson, P., Manciot, F., Moldovan, Z., Patureau, D., Cruceru, L., Sacher, F., Ledin, A., 2010. Inter-laboratory exercise on steroid estrogens in aqueous samples. Environ. Pollut. 158, 658–662. https://doi.org/10.1016/J.ENVPOL.2009.10.029

Hirobe, M., Rubio, F., Fujita, M., Shiraishi, H., 2004. Development of ELISAs for quantification of surfactants, endocrine disruptors and estrogens, and their application for environmental and biological sample analysis [WWW Document]. Present. 2004 Natl. Monit. Conf. Present. Results, Chattanooga, TN - May 17 – 20,. URL https://www.researchgate.net/profile/Fernando_Rubio/publication/268047918_Development_of_ELISAs_for_quantification_of_surfactants_endocrine_disruptors_and_estrogens_and_their_application_for_environmental_and_biological_sample_analysis/links/54b7ae570cf2b (accessed 7.12.18).

Ikediashi, C., Paris, J.R., King, R.A., Beaumont, W.R.C., Ibbotson, A., Stevens, J.R., 2018. Atlantic salmon *Salmo salar* in the chalk streams of England are genetically unique. J. Fish Biol. 92, 621–641. https://doi.org/10.1111/jfb.13538

Ioannidis, J.P.A., 2005. Why most published research findings are false. PLoS Med. 2, 0696–0701. https://doi.org/10.1371/journal.pmed.0020124

Johnson, A.C., 2010. Natural variations in flow are critical in determining concentrations of point source contaminants in rivers: An estrogen example. Environ. Sci. Technol. 44, 7865–7870. https://doi.org/10.1021/es101799j

Kamps, R.H., Neill, W.H., 1999. Aquacultural effluents: Directive signals to the system downstream? J. Chem. Ecol. 25, 2041–2050. https://doi.org/10.1023/A:1021032722170

Kolodziej, E.P., Harter, T., Sedlak, D.L., 2004. Dairy wastewater, aquaculture, and spawning fish as sources of steroid hormones in the aquatic environment. Environ. Sci. Technol. 38, 6377–6384. https://doi.org/10.1021/es049585d

Leduc, A.O.H.C., Kelly, J.M., E. Brown, G., 2004. Detection of conspecific alarm cues by juvenile salmonids under neutral and weakly acidic conditions: laboratory and field tests. Oecologia

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

139, 318–324. https://doi.org/10.1007/s00442-004-1492-8

McMaster, M.E., Jardine, J.J., Ankley, G.T., Benson, W.H., Greeley, M.S., Gross, T.S., Guillette, L.J., MacLatchy, D.L., Orlando, E.F., Van Der Kraak, G.J., Munkittrick, K.R., 2001. An interlaboratory study on the use of steroid hormones in examining endocrine disruption. Environ. Toxicol. Chem. 20, 2081–2087. https://doi.org/10.1002/etc.5620200930

Mole, G., 2008. Environment Agency accused of flouting EU laws over fish farms [WWW Document]. Telegr. URL https://www.telegraph.co.uk/news/earth/earthnews/3353750/Environment-Agency-accused-of-flouting-EU-laws-over-fish-farms.html (accessed 7.12.18).

Moore, A., 1991. Behavioural and physiological responses of precocious male Atlantic salmon (Salmo salar L.) to testosterone., in: Scott, A., Sumpter, J., Kime, D., Rolfe, M. (Eds.), Proceedings of the 4th International Symposium on Reproductive Physiology of Fish. FishSymp91, Sheffield, pp. 194–196.

Moore, A., Olsén, K.H., Lower, N., Kindahl, H., 2002. The role of F-series prostaglandins as reproductive priming pheromones in the brown trout. J. Fish Biol. 60, 613–624. https://doi.org/10.1111/j.1095-8649.2002.tb01688.x

Moore, A., Scott, A.P., 1992. 17 alpha,20 beta-dihydroxy-4-pregnen-3-one 20-sulphate is a potent odorant in precocious male Atlantic salmon (Salmo salar L.) parr which have been pre-exposed to the urine of ovulated females. Proceedings. Biol. Sci. 249, 205–9. https://doi.org/10.1098/rspb.1992.0105

Moore, A., Scott, A.P., 1991. Testosterone is a potent odorant in precocious male Atlantic salmon (Salmo-Salar L) parr. Philos. Trans. R. Soc. London Ser. B-Biological Sci. 332, 241–244. https://doi.org/10.1098/rstb.1991.0052

Moore, A., Waring, C.P., 1996. Electrophysiological and endocrinological evidence that F-series prostaglandins function as priming pheromones in mature male Atlantic salmon (Salmo salar) parr. J. Exp. Biol. 199, 2307–2316.

Mota, V.C., Martins, C.I.M., Eding, E.H., Canário, A.V.M., Verreth, J.A.J., 2014. Steroids accumulate in the rearing water of commercial recirculating aquaculture systems. Aquac. Eng. https://doi.org/10.1016/j.aquaeng.2014.07.004

Pelissero, C., Cuisset, B., Le Menn, F., 1989. The influence of sex steroids in commercial fish meals and fish diets on plasma concentration of estrogens and vitellogenin in cultured Siberian sturgeon *Acipenser baeri*. Aquat. Living Resour. 2, 161–168. https://doi.org/10.1051/alr:1989019

Pelissero, C., Sumpter, J., 1992. Steroids and “steroid-like” substances in fish diets. Aquaculture 107, 283–301.

Peppler, R.D., Stone, S.C., 1979. Effect of oral contraceptive used by technician on sensitivity of progesterone radioimmunoassay. Horm. Res. 11, 56–90.

Petrie, A., Sabin, C., 2000. Medical statistics at a glance, 1st ed. Blackwell Science Ltd, Oxford.

Pottinger, T.G., Moran, T.A., Cranwell, P.A., 1992. The biliary accumulation of corticosteroids in rainbow trout,Oncorhynchus mykiss, during acute and chronic stress. Fish Physiol. Biochem. 10, 55–66. https://doi.org/10.1007/BF00004654

Rubio, F., Ramsey, K., Stackelberg, P., Meyer, M., 2004. Monitoring of selected herbicides,

ACCEPTED MANUSCRIPT

antibiotics, steroids, and industrial chemicals in water by ELISA [WWW Document]. Natl. Monit. Conf. Present. Results. URL https://acwi.gov/monitoring/conference/2004/conference_agenda_links/power_points_etc/03_ConcurrentSessionA/24_Rm13_RubioFernando.pdf (accessed 7.12.18).

Santulli, A., Modica, A., Messina, C., Ceffa, L., Curatolo, A., Rivas, G., Fabi, G., D'Amelio, V., 1999. Biochemical responses of European sea bass (Dicentrarchus labrax L.) to the stress induced by off shore experimental seismic prospecting. Mar. Pollut. Bull. 38, 1105–1114. https://doi.org/10.1016/S0025-326X(99)00136-8

Scott, A.P., Bye, V.J., Baynes, S.M., 1980. Seasonal variations in sex steroids of female rainbow trout (Salmo gairdneri Richardson). J. Fish Biol. 17, 587–592. https://doi.org/10.1111/j.1095-8649.1980.tb02790.x

Selset, R., Døving, K.B., 1980. Behaviour of mature anadromous char (Salmo alpinus L.) towards odorants produced by smolts of their own population. Acta Physiol. Scand. 108, 113–122. https://doi.org/10.1111/j.1748-1716.1980.tb06508.x

Shore, L.S., Reichmann, O., Shemesh, M., Wenzel, A., Litaor, M.I., 2004. Washout of accumulated testosterone in a watershed. Sci. Total Environ. 332, 193–202. https://doi.org/10.1016/J.SCITOTENV.2004.04.009

Sink, T.D., Lochmann, R.T., Fecteau, K.A., 2008. Validation, use, and disadvantages of enzyme-linked immunosorbent assay kits for detection of cortisol in channel catfish, largemouth bass, red pacu, and golden shiners. Fish Physiol. Biochem. 34, 95–101. https://doi.org/10.1007/s10695-007-9150-9

Sower, S.A., Iwamoto, R.N., 1985. The identification of the sex steroid, testosterone, in various commercial salmon diets. Aquaculture 49, 11–17.

Stacey, N., 2015. Hormonally derived pheromones in teleost fishes, in: Sorensen, P., Wisenden, B. (Eds.), Fish Pheromones and Related Cues. John Wiley & Sons, Inc., pp. 33–88. https://doi.org/https://doi.org/10.1002/9781118794739.ch3

Sumpter, J.P., Scott, A.P., Baynes, S.M., Witthames, P.R., 1984. Early stages of the reproductive cycle in virgin female rainbow trout (Salmo gairdneri Richardson). Aquaculture 43, 235–242. https://doi.org/10.1016/0044-8486(84)90025-5

Sundh, H., Kvamme, B., Fridell, F., Olsen, R., Ellis, T., Taranger, G., Sundell, K., 2010. Intestinal barrier function of Atlantic salmon (Salmo salar L.) post smolts is reduced by common sea cage environments and suggested as a possible physiological welfare indicator. BMC Physiol. 10, 22. https://doi.org/10.1186/1472-6793-10-22

Tyler, C.R., Jobling, S., Sumpter, J.P., 1998. Endocrine disruption in wildlife: A critical review of the evidence. Crit. Rev. Toxicol. 28, 319–361. https://doi.org/10.1080/10408449891344236

Vermeirssen, E.L.M., Scott, A.P., 1996. Excretion of free and conjugated steroids in rainbow trout (Oncorhynchus mykiss): Evidence for branchial excretion of the maturation-inducing steroid, 17,20β-dihydroxy-4-pregnen-3-one. Gen. Comp. Endocrinol. 101, 180–194. https://doi.org/10.1006/gcen.1996.0020

Waring, C.P., Moore, A., Best, J.H., Crooks, N., Crooks, L.E., 2012. Do trout farm effluents affect Atlantic salmon smolts? Preliminary studies using caged salmon smolts. Aquaculture 362–363, 209–215. https://doi.org/10.1016/J.AQUACULTURE.2011.11.046

ACCEPTED MANUSCRIPT

Wong, S.C., Dykstra, M., Campbell, J.M., Earley, R.L., 2008. Measuring water-borne cortisol in convict cichlids (Amatitlania nigrofasciata): Is the procedure a stressor? Behaviour 145, 1283–1305. https://doi.org/10.1163/156853908785765863

Wudy, S.A., Schuler, G., Sánchez-Guijo, A., Hartmann, M.F., 2018. The art of measuring steroids: Principles and practice of current hormonal steroid analysis. J. Steroid Biochem. Mol. Biol. 179, 88–103. https://doi.org/10.1016/j.jsbmb.2017.09.003

Yao, B., Li, R., Yan, S., Chan, S., Song, W., 2018. Occurrence and estrogenic activity of steroid hormones in Chinese streams : A nationwide study based on a combination of chemical and biological tools. Environ. Int. 118, 1–8. https://doi.org/10.1016/j.envint.2018.05.026

Zhao, S., Zhang, P., Melcer, M.E., Molina, J.F., 2010. Estrogens in streams associated with a concentrated animal feeding operation in upstate New York, USA. Chemosphere 79, 420–425. https://doi.org/10.1016/J.CHEMOSPHERE.2010.01.060

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1: Details of standard curves of the enzyme-linked immunoassays (EIAs) and radioimmunoassays (RIAs) used in the 2010 study.

Steroid	Assay	Manufacturer/Participant	No of standards	Minimum (ng mL^{-1})	Maximum (ng mL^{-1})
11-KT	EIA	Cayman Chemical Company	8	0.00078	0.1
	RIA	Participant 1	9	0.01	2.5
		Participant 2	7	0.02	1.25
		Participant 3	8	0.06	8.0
T	EIA	Cayman Chemical Company	8	0.0039	0.5
	RIA	Participant 1	9	0.02	5.0
		Participant 2	7	0.02	1.25
		Participant 3	8	0.06	8.0
E2	EIA	Cayman Chemical Company	8	0.0066	4.0
	RIA	Participant 1	9	0.005	1.25
		Participant 2	8	0.01	1.25
		Participant 3	8	0.06	8.0
EE2	EIA	Tokiwa Chemical Industries	4	0.05	3.0
	RIA	Participant 1	9	0.02	5.0
		Participant 2	7	0.02	1.25
		Participant 3	8	0.06	8.0

Table 2: Effects of assay method and laboratory on steroid measurements. ANOVAs assessing factors affecting measurements of steroid concentration in sample aliquots. Please note that values below the lowest standard were treated as missing values in these analyses.

	Log (11KT)			Log (T)			Log (E2)			Log (EE2)		
	df	F	p	df	F	p	df	F	p	df	F	p
Assay method	1	188	<0.001	1	231	<0.001	1	38	<0.001	1	6.8	0.011
Participant lab [assay method]	4	45	<0.001	4	27	<0.001	4	68	<0.001	4	45	<0.001
Sample	43	143	<0.001	43	272	<0.001	43	673	<0.001	43	24	<0.001
Error	184			215			215			104		
Total	232			263			263			152		

ACCEPTED MANUSCRIPT

Table 3: Comparison of EIA and RIA measurements: statistical comparisons of values.

Steroid			11KT	T	E2	EE2
Correlation: log(EIA mean) v log (RIA mean)	n		44	44	44	7
	r		0.983	0.990	0.998	0.978
	p		<0.001	<0.001	<0.001	<0.001
Regression: log(EIA mean) v log (RIA mean)	Slope	Value (SE)	1.15770 (0.03320)	1.01973 (0.02196)	0.95754 (0.01018)	1.0959 (0.1035)
		P, ≠ 1	<0.001	>0.20	<0.001	>0.20
	Intercept	Value (SE)	-0.26704 (0.04368)	-0.18233 (0.01687)	0.042867 (0.009241)	-0.29735 (0.08122)
		P, ≠ 0	<0.001	<0.001	<0.001	0.015
Bland-Altman; Regression: Log(EIA mean) – log (RIA mean) v average [Log(EIA mean) & log (RIA mean)]	Slope	Value	0.16443	0.02932	-0.04106	0.11451
		P, ≠ 0	<0.001	0.179	<0.001	0.273
	Intercept	Value	-0.22468	-0.17656	0.044856	-0.28582
		P, ≠ 0	<0.001	<0.001	<0.001	0.012

Table 4: Inter-laboratory variation within EIA and RIA measurements. Coefficients of variation for the two methods for the four steroids, compared by paired t-test

Steroid		11KT	T	E2	EE2
EIA CV	mean	0.218 (n=44)	0.321 (n=44)	0.269 (n=44)	0.330 (n=6)
RIA CV	mean	0.572 (n=44)	0.160 (n=44)	0.152 (n=44)	0.814 (n=44)
Paired t-test	T-value	-7.88	5.69	6.21	-1.19
	p	<0.001	<0.001	<0.001	0.289

Table 5: Accuracy of EIA and RIA: Comparison of measured concentrations with nominal values for the prepared spiking solution.

Steroid		11KT	T	E2	EE2
EIA	Mean Accuracy	74%	75%	93%	159%
	T-test mean ≠ 1	T=-2.49, p=0.130	T=-4.24, p=0.051	T=-0.25, p=0.825	T=1.58, p=0.255
RIA	Mean Accuracy	84%	92%	97%	248%
	T-test mean ≠ 1	T=-1.97, p=0.188	T=-1.38, p=0.301	T=-0.19, p=0.867	T=3.02, p=0.094
T-test EIA ≠ RIA		T=-0.75, p=0.506	T=-1.98, p=0.143	T=-0.13, P=0.907	T=-1.45, p=0.244

Table 6: The reproducibility of the extraction (and assay) methodology expressed as the coefficient of variation of replicate samples (n=3).

	11KT	T	E2	EE2
R. Test	7%	11%	7%	67%
R. Avon	6%	3%	8%	58%

ACCEPTED MANUSCRIPT

Table 7: Recovery efficiency of the method to extract steroids from water samples into methanol aliquots for assay. Please note that accidental transfer of methanol extract occurred between samples D and E during rotary evaporation; however, as no methanol was lost, although this will increase the apparent variation in recovery, the mean remains valid.

Steroid		11KT	T	E2	EE2
Sample	B: De-ionised	27%	48%	79%	77%
	C: R Test	26%	44%	65%	62%
	D: De-ionised	25%	40%	50%	61%
	E: R Avon	37%	55%	82%	68%
Mean		29%	47%	69%	67%

Table 8: Median measured river steroid concentrations (ng L^{-1}), uncorrected for recovery efficiency.

Steroid			11KT		T		E2		EE2	
River	Season	Sampling site	mean	(n)	mean	(n)	mean	(n)	mean	(n)
R. Test	Winter	Reference "upstream"	0.01	(2)	0.15	(2)	0.19	(2)	0.07	(2)
		Trout farm inflow	0.01	(2)	0.14	(2)	0.16	(2)	0.06	(2)
		Trout farm outflow	0.02	(2)	0.25	(2)	0.15	(2)	0.04	(2)
		Downstream	0.01	(2)	0.20	(2)	0.17	(2)	0.05	(2)
	Spring	Reference "upstream"	0.01	(2)	0.28	(2)	0.17	(2)	0.04	(2)
		Trout farm inflow	0.01	(2)	0.32	(2)	0.19	(2)	0.07	(2)
		Trout farm outflow	0.02	(4)	0.47	(4)	0.19	(4)	0.11	(4)
		Downstream	0.02	(3)	0.42	(3)	0.22	(3)	0.05	(3)
R. Avon	Winter	Reference "upstream"	0.01	(2)	0.20	(2)	0.18	(2)	0.04	(2)
		Trout farm inflow	No samples collected							
		Trout farm outflow								
		Downstream	0.01	(2)	0.17	(2)	0.16	(2)	0.04	(2)
	Spring	Reference "upstream"	0.02	(2)	0.34	(2)	0.20	(2)	0.07	(2)
		Trout farm inflow	0.02	(2)	0.38	(2)	0.20	(2)	0.09	(2)
		Trout farm outflow	0.02	(2)	0.36	(2)	0.21	(2)	0.08	(2)
		Downstream	0.02	(5)	0.29	(5)	0.21	(5)	0.09	(5)

Table 9: Results of ANOVA assessing the effects of river, season and sampling station on median measured river steroid concentrations (ng L-1), uncorrected for recovery efficiency (n=34; spiked, de-ionised, and blank-methanol data excluded).

Steroid	11KT	T	E2	EE2
River	p=0.078	p=0.409	p=0.718	p=0.393
Season	p<0.001	p<0.001	p=0.081	p=0.087
Sampling station	p=0.271	p=0.081	p=0.965	p=0.674

ACCEPTED MANUSCRIPT

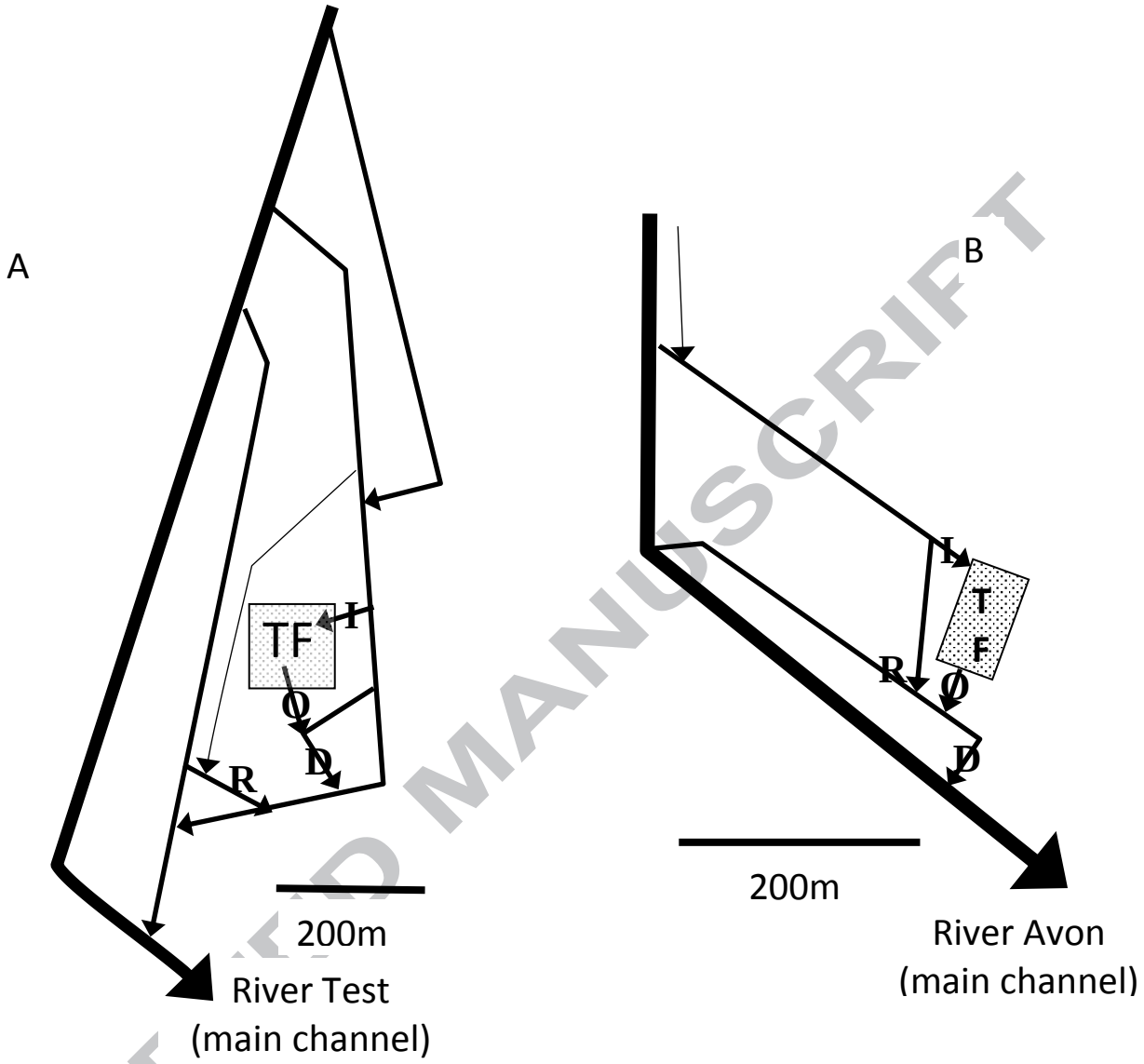

Figure 1: Diagrammatic representations of locations of sampling sites (I= inflow, O=outflow, R=reference, D= downstream,) on the R Test (A) and R Avon (B) in relation to the trout farms (TF) and river channels (→). Main channel of rivers labelled, with arrows indicating direction of flow.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

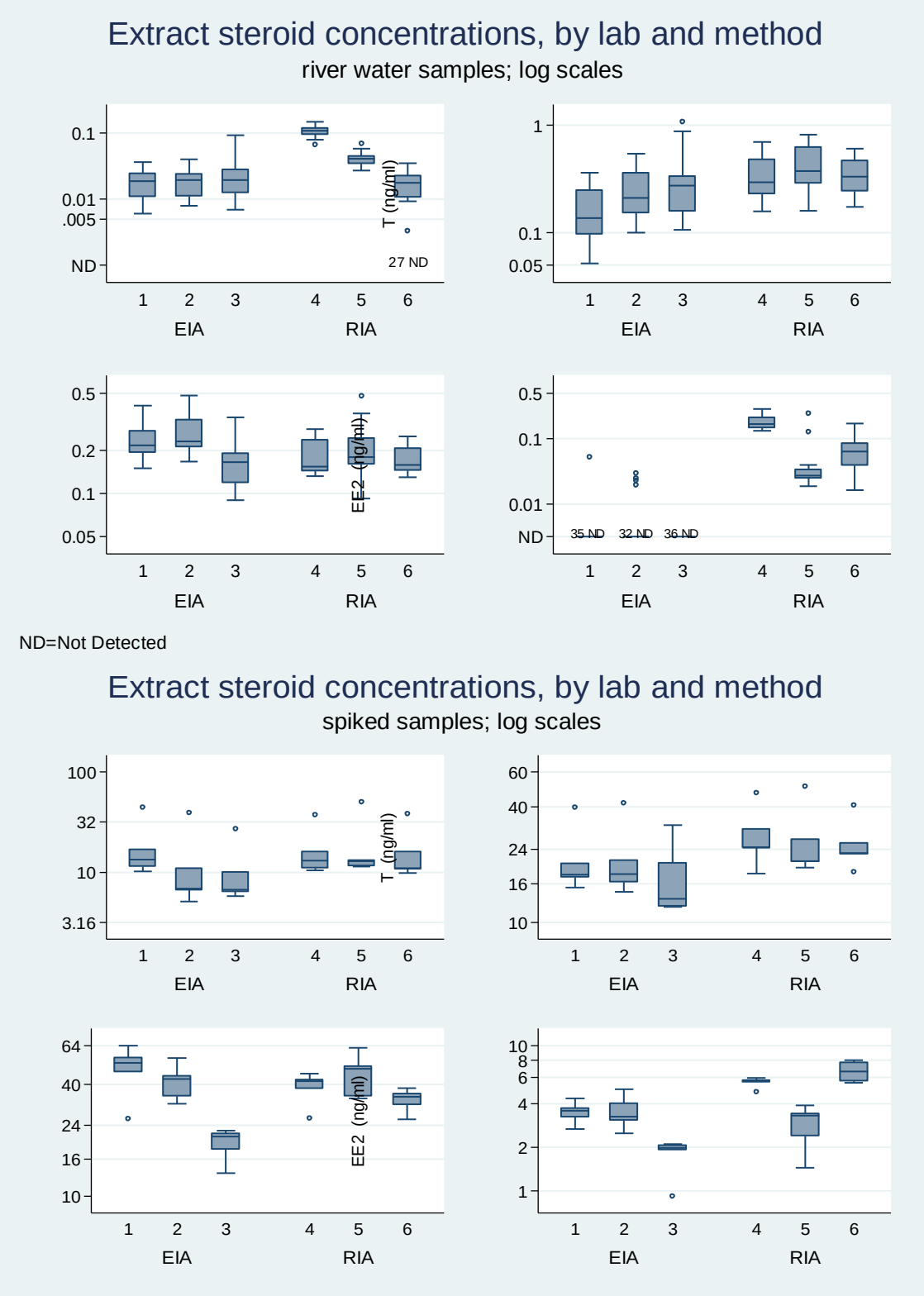

Figure 2: Comparison of EIA and RIA measurements: Box and whisker plots showing distributions of measured concentrations of steroid (11-KT, T, E2 and EE2) in replicate aliquots by EIA and RIA by participant laboratories (denoted by 1-3 for EIA, 4-6 for RIA). Separate plots for river water samples (top) and spiked samples (bottom) on log scales. Plots show quartiles as box, median as horizontal line within box, whiskers to expected range “adjacent values”, and outliers outside the whiskers.

ACCEPTED MANUSCRIPT

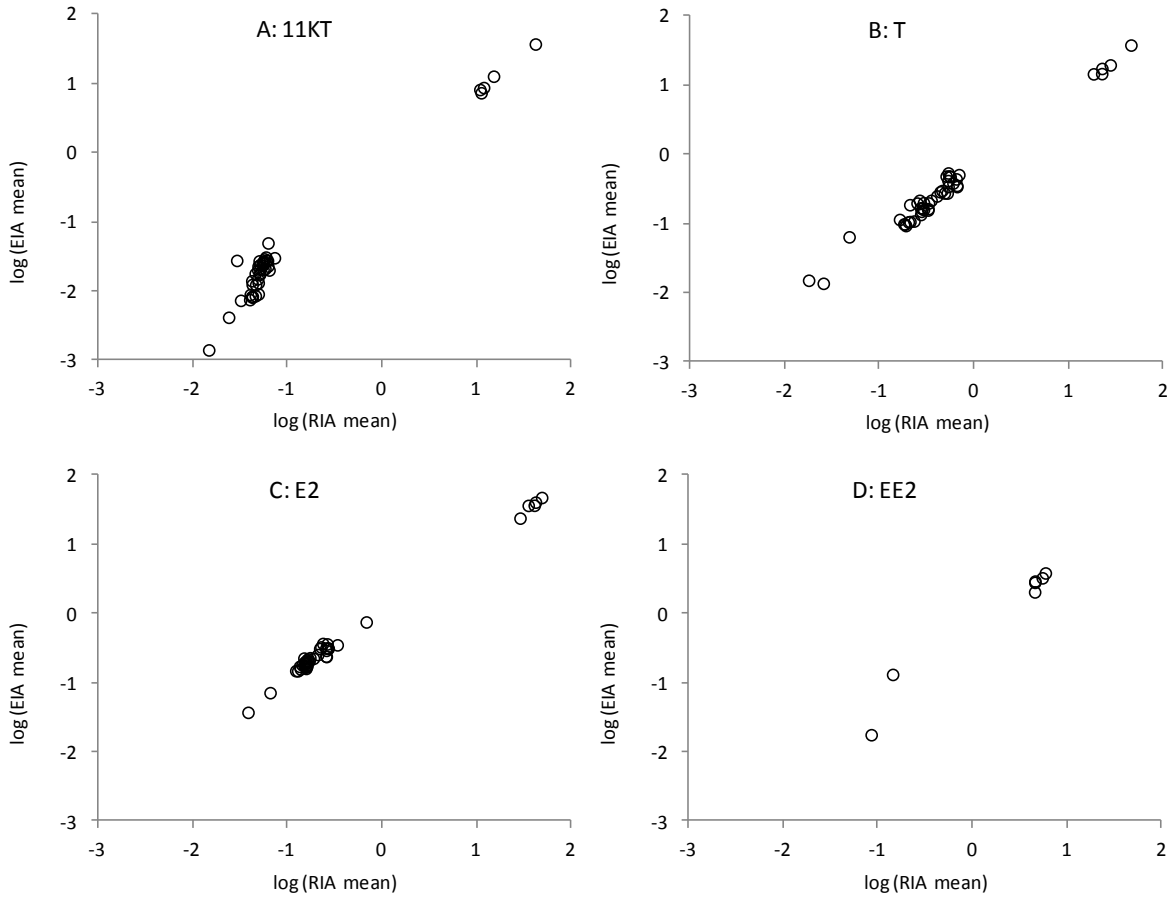

Figure 3: Comparison of EIA and RIA measurements: Plots of mean EIA values (log_{10} transformed) against mean RIA values (log_{10} transformed). The lower groups consist of the river water samples, the upper groups represent the steroid spiked samples.

ACCEPTED

ACCEPTED MANUSCRIPT

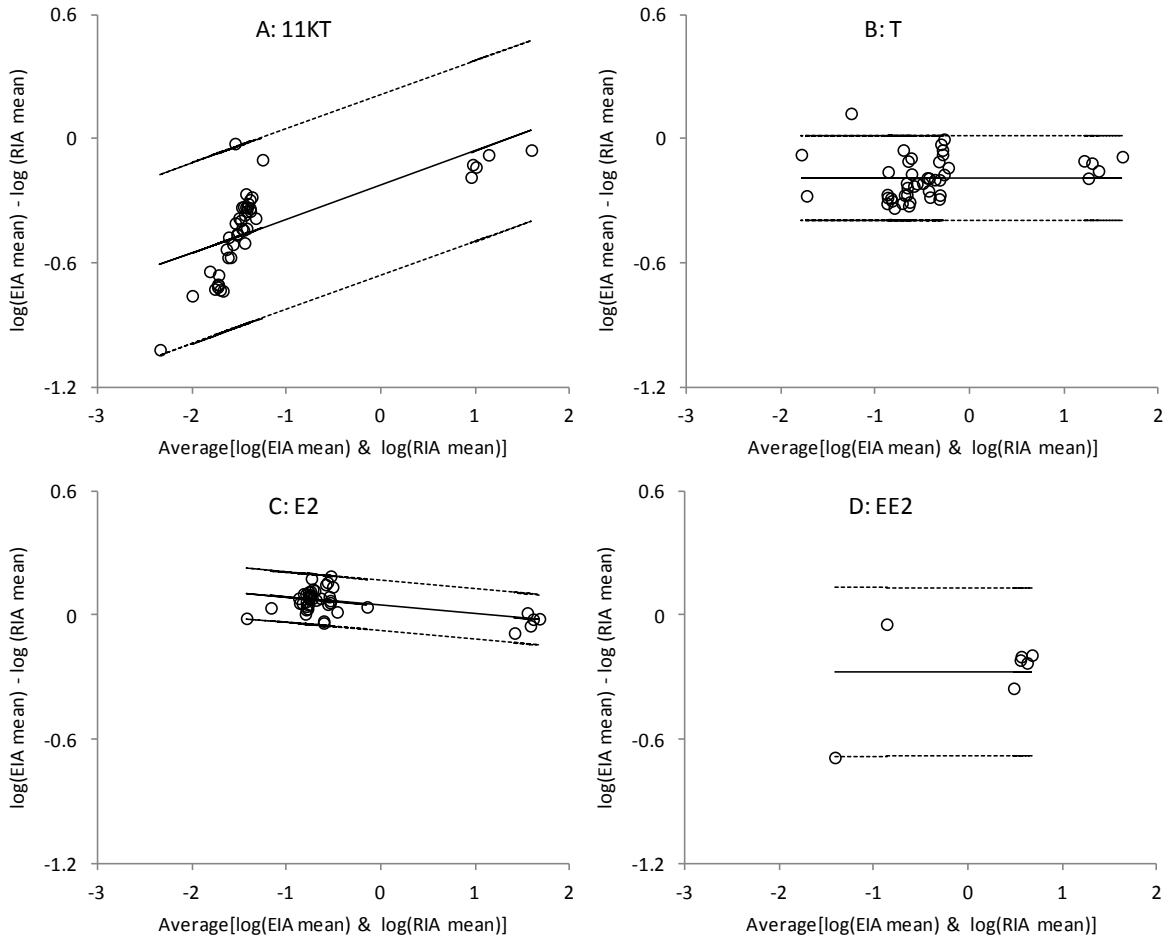

Figure 4: Comparison of EIA and RIA measurements: Bland-Altman plots (difference v average) of mean concentrations of four steroids measured by EIA and RIA.

ACCEPTED

ACCEPTED MANUSCRIPT

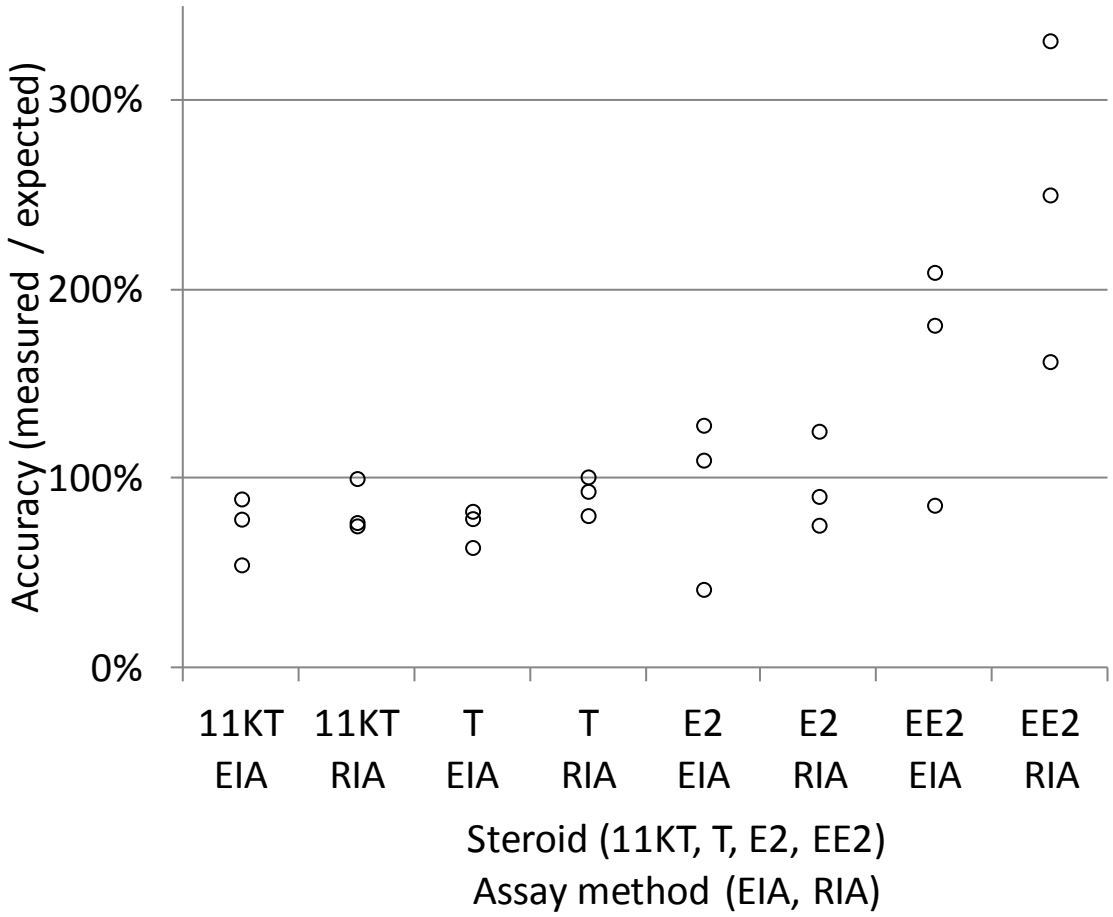

Figure 5: Accuracy: Individual measured values of steroid concentration in methanol spiking solution, expressed relative to expected concentration. n=3 for each steroid and assay method representing values from the participating laboratories.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

6. We repeated a study that reported abnormally high steroid levels in river water.
7. In the repeated study, 2 of the 4 laboratories made traceable calculation errors.
8. We found steroid levels 2-3 orders of magnitude lower than in the original study.
9. We attributed the high steroid levels in the original study to calculation errors.
10. Current evidence indicates that trout farms are a minor source of river steroids.

ACCEPTED MANUSCRIPT