The 21-day androgenised female stickleback endocrine screening assay

Defra funded research

July 2009.

Ioanna Katsiadaki, BVMD, PhD, OVS, MRCVS

ioanna.katsiadaki@cefas.co.uk

Contractor: Cefas Weymouth laboratory

Barrack Road, The Nothe

Weymouth, Dorset

DT4 8UB

Tel: 0044 1305 206600
Fax: 0044 1305 206601

Contributors: Dr Tom Pottinger, CEH Lancaster (Lab 3)

Professor Ian Mayer, Bergen University (Lab 4)

Dr Cecile Jolly, Bergen University

Dr Steve Morris, Cefas Weymouth (Lab 2)

Dr Alex Scott, Cefas Weymouth

Dr Matthew Sanders, Cefas Weymouth

Mr Mark Hurst, Cefas Burnham (Lab 1)

Executive summary

1. The current OECD test guidelines for screening chemicals with potential endocrine disrupting activity (TG 229 and TG 230) cannot clearly identify androgen antagonists due to the lack of a suitable end-point in the 3 core species involved, the fathead minnow, the medaka and the zebrafish.

2. The reported antiandrogenic activity in the aquatic environment based on the Yeast Androgen Screen (YAS), a suitable *in vitro* test, is substantial. The evidence of high levels of antiandrogens in the environment emphasises the need to develop a suitable assay for chemical screening.

3. The 3-spined stickleback posses a unique trait, the presence of an androgen regulated protein in their kidney, spiggin, which can be deployed in the detection of environmental androgens and antiandrogens.

4. There are two *in vivo* tests using the stickleback that have the potential to identify compounds with antiandrogenic activity, the stickleback-breeding test and the androgenised female stickleback screen.

5. The design of the androgenised female stickleback screen (AFSS) is better suited as an endocrine screen due to its simplicity and reproducibility in any laboratory. The fish are simultaneously treated with a model androgen (dihydro-testosterone, DHT) at 5μg/L and a range of concentrations of the putative antiandrogen. Any antiandrogenic activity is detected by the degree of reduction/inhibition of spiggin induction by the DHT treatment.

6. This report also addresses a retrospective validation of a large dataset using the AFSS produced over a period of 7 years with the participation of four laboratories (Lab 1, 2, 3 and 4). The *in vivo* exposures included testing of four antiandrogenic compounds using kidney spiggin levels as an end-point. Spiggin was measured by a validated ELISA method.

7. Independent statistical analysis of the dataset revealed that the AFSS can unambiguously and reproducibly detect antiandrogens. All four tested compounds (Flutamide, Fenitrothion, Vinclozolin and Linuron) were characterised as antiandrogens on the basis of the AFSS.

8. We also report a small *in vivo* test, conducted in Lab 2 which investigated the requirement to employ solvent carrier during the AFSS and the suitability of potassium permanganate as an negative control substance (as per the OECD draft guideline)

9. We found that there is a need to employ solvent vehicle during the AFSS as the presence of water in the stock solutions or mixing vessels can result in a high degradation rate of DHT.

10. The use of potassium permanganate (PP) as a negative control substance at 400μg/L is not advised. PP was toxic to the fish, it affected food palatability and its use as an oxidising agent for steroid synthesis compromises to a large extend its potential as a substance with no endocrine activity.

Background and objectives

11. The so far validated test guidelines for the detection of endocrine disrupting chemicals (EDCs) include TG229 and TG230. This project provides an assessment of a potentially complementary test guideline using the stickleback (*Gasterosteus aculeatus*) and results of a validation of the protocols used.

12. The main shortfall of TG230 and to a certain extent of TG229 is the lack of ability to detect antiandrogens (OECD report, 2006), an important class of EDCs. In 2003 and 2004, two nationwide surveys revealed a significant antiandrogenic activity present in UK final sewage effluents (Johnson *et al*, 2007). More recently, significant antiandrogenic activity was reported in water and sediment samples from a river in Italy (Urbatzka *et al*, 2007) and in the produced water from oil platforms in the North Sea (Tollefsen *et al*, 2007). All three studies used the same, well-established *in vitro* method, namely the yeast androgen screen (YAS), in which the human androgen receptor is incorporated into the yeast. The results are expressed in flutamide equivalents (FL eq), flutamide being the most commonly used model antiandrogen. Although the number of reports is so far small, the actual antiandrogenic activity was very significant; up to 1230 μg/L FL eq in the UK final sewage effluent, up to 4200 μg/L FL eq in the river Lambro and an extraordinary 8000 μg/L FL eq in the oil platform produced water. The newly emerged problem of high antiandrogenic activity in the aquatic environment and recent suggestions that antiandrogens are contributing to wild fish sexual disruption in UK rivers (Jobling *et al*, 2009) has highlighted the inability of the fish screen to detect this class of EDCs.

Milestones

13. The first milestone of this project was to collate all relevant data generated by four laboratories (Lab 1, Lab 2, Lab 3 and Lab 4) over the past 7 years (2001-2008) on the antiandrogenic potential of chemicals as assessed by the reduction/inhibition of spiggin, the stickleback kidney glue protein, using androgen-stimulated female fish. Spiggin is to date the only known androgen induced protein in fish and has been extensively used as a biomarker of androgenic and antiandrogenic chemicals (Katsiadaki *et al*, 2002a; Hahlbeck *et al*, 2004; Katsiadaki *et al*, 2006; Jolly *et al*, 2006; Andersson *et al*, 2007; Bjorkblom *et al*, 2007; Allen *et al*, 2008; Sanchez *et al*, 2008).

14. Spiggin is produced in the kidney of male sticklebacks under androgen stimulation during their breeding season. It is stored in the urinary bladder from which is excreted and used as a cement to build up a nest in which the female lays her eggs. Under natural conditions it should not be present in the kidneys of female fish (similarly to vitellogenin that should not be present in the blood of male fish), hence it is a specific biomarker for androgens. There is more than one way in which spiggin can be estimated or measured in the stickleback kidney. The first published method (Borg *et al*, 1993) for obtaining an estimate of spiggin production in the stickleback kidney employed histological measurements of the kidney secondary proximal epithelium height (KEH). The second method is an enzyme linked immunosorbent assay (ELISA), which was first developed by Cefas scientists in 1999 (Katsiadaki *et al*, 1999). This method has been validated (Katsiadaki *et al*, 2002a) and is still the only widely used assay for spiggin. More recently Sanchez *et al*, (2008) described an ELISA for spiggin using an antibody against a peptide sequence of spiggin. In addition, spiggin mRNA levels can be detected in the stickleback kidney by real time PCR; we developed the first published protocol for this method in collaboration with Japanese co-workers (Nagae *et al*, 2007) but more methods are coming to light (Hogan *et al*, 2009).

15. The bulk of the existing data were produced during the course of the EU-funded project Endocrine disrupters: exploring novel endpoints, exposure, low-dose-and mixture-effects in humans, aquatic wildlife and laboratory animals (EDEN), that was particularly focusing on the effects of chemical mixtures. A sub-set of the data for exposures to single chemicals rather than mixtures has been analysed in this report.

16. The second milestone was to undertake independent statistical analysis of the existing data to assess whether or not the main end-point of the assay (kidney spiggin levels) has the

17. In support of the third milestone we generated of a small set of data anticipated as essential for the validation process under the OECD guidelines. These data were focused on two separate items:

- To investigate the effect of carrier solvent on the end-point employed by the assay. The OECD guideline for the fish screen explicitly states that the use of solvent should be avoided all together and if it absolutely necessary to employ a solvent for the *in vivo* exposures then the level of solvent should not exceed 0.01% at the fish aquaria. However all existing data employed methanol as a carrier solvent for administering the chemicals to the experimental aquaria reaching a final concentration of 0.1%.
- Investigate the effect of a non-endocrine active substance on the end-point employed by the assay. Currently the OECD guideline suggests that either potassium permanganate and/or n-Octanol are suitable negative control substances. We investigated the suitability of potassium permanganate (PP) for this purpose.

18. The final milestone was the production of the current report, which is a detailed validation report on the methodology and analyses conducted. It was prepared in order to be peer reviewed (as proposed in the guidance document on the validation and international acceptance of new or updated test methods for hazard assessment, No 34).

19. A brief overview of the stickleback tests in respect to the OECD guidelines for detecting endocrine disruptors is provided to set the context of this work for the reader. A detailed description of the stickleback antiandrogen tests is also provided.

Historical overview of the stickleback tests under OECD guidelines

20. An intercalibration using the three-spined stickleback (*Gasterosteus aculeatus*) was conducted in 2004 to assess the potential of this species for inclusion in an OECD Test Guideline "The Fish Screening Assay for Endocrine Active Substances"

21. Two biomarker endpoints, vitellogenin induction in males and spiggin induction in females, were shown to be suitably relevant, sensitive and reproducible for potent oestrogens and androgens (Allen *et al*, 2008).

22. The OECD Validation management Group requested that a second intercalibration be carried out to assess the suitability of these endpoints for weak oestrogens, aromatase

23. The results from the phase 1B intercalibration exercise indicated that the endpoints available in the three core species (i.e. Japanese medaka, fathead minnow and zebrafish) could not detect a significant class of endocrine disruptors, namely antiandrogens.

24. Unfortunately, the results of the second ring test involving the stickleback were not as conclusive, in that detection of antiandrogenic activity using intact male sticklebacks and spiggin inhibition as an end-point was not achieved by two of the three participating laboratories. The Cefas Weymouth laboratory however was very successful in conducting these tests, providing firm evidence on the antiandrogenic activity of flutamide (a pharmaceutical used in the treatment of prostate cancer) and fenitrothion (an organophosphate pesticide) in fish (Sebire *et al*, 2008; Sebire *et al*, 2009).

25. Although there were a number of confounding factors (i.e. disease outbreaks, inappropriate reproductive status of fish, inappropriate sample preparation), it was generally accepted that the test design that was employed during the second intercalibration exercise was not particularly robust and perhaps too labour intensive for a screen.

26. We presented these results, along with some data that we generated over the past 7 years within Cefas (Labs 1 and 2), CEH Lancaster (Lab 3) and Bergen University (Lab 4) using an alternative test for antiandrogens, to the OECD Validation management Group on January the 16th, 2008.

27. The OECD Validation management Group suggested that a Standard Protocol Submission Form (SPSF) form is submitted to the UK representative for consideration in May 2008. The work was commissioned in early July 2008.

28. The 'new' test is using only female fish that are androgenised by a low concentration of a model androgen (Dihydrotestosterone, DHT at 5µg/L) and are simultaneously exposed to the putative antiandrogen.

29. A total number of 20 exposures using this design had already been conducted under the EU funded project EDEN in 3 laboratories. The design is similar to the OECD guideline (with some minor differences) and so the data could be used retrospectively to obtain validation status. Four publications are under preparation from the EDEN work.

30. The main shortfalls of the previously conducted work were related to the use of a solvent as a vehicle for DHT, the lack of data on an endocrine negative substance and the lack of a detailed report that could be submitted for peer review.

31. A small test was conducted at Lab 2 in August 2008 to address the need for solvent use when administering the androgen DHT (alone and in combination with the model antiandrogen flutamide). In addition we investigated the suitability of PP an endocrine negative control compound.

32. In August 2008, we received a grant from the National Centre of 3Rs (Reduction, Replacement and Refinement in animal experimental procedures), which allows us to participate in the validation of the Fish Sexual Development test (FSDT). The stickleback is therefore included in this validation work of the FSDT and work is underway.

Description of the stickleback anti-androgen assays

The androgenised female stickleback screen (AFSS)

33. The androgenised female stickleback screen was first described in a Cefas report to Defra (AE1150) in 2003 as a unique *in vivo* test for the detection of xenobiotics with androgen antagonistic activity. The principle of the assay is simple: Female sticklebacks are simultaneously exposed to concentration of an androgen able to induce spiggin production in their kidneys (DHT at 5μg/L or 17α-methyl testosterone at 0.5μg/L) and a range of concentrations of the putative antiandrogen. Positive controls (groups of females treated with the androgen only), negative controls (groups of females treated with the putative antiandrogen only at the highest concentration tested) as well as water and solvent controls are employed.

34. By definition the test is relevant only for female fish, however, male fish are often misidentified as females during selection from the holding tanks. This is a commonly observed problem besides the fact that sticklebacks display strong sexual dimorphism during breeding. Some mails fail to mature in the holding tanks (almost certainly due to social hierarchies established) and hence do not display strong secondary sexual characters such as the distinct red coloration of the lower jaw and blue irises. There are two ways of dealing with this issue:

a) to allow for some level of sex misidentification when designing the test by increasing the number of fish are used. Indeed during the majority of the exposures reported here we applied this rule and increased the number of fish from 10 in each

b) to positively identify male fish from the stick population and gradually remove them from the holding tanks in order to ensure an all-female population. Guidance on doing this is provided in the draft test guideline that accompanies this report. This is the preferred way as it is more ethical and reduces the number of fish to half.

35. The *in vivo* antiandrogenic potential of the chemical is assessed by reduction or inhibition of spiggin induction by the model androgen. The duration of the assay is 21 days, in line with TG 230. It is possible that the AFSS could be reduced to 14 days or less if a higher concentration of DHT is used, however all the validation work was undertaken using this DHT concentration and this duration, so further studies are needed to confirm this.

36. Spiggin is measured in the kidneys of the fish by ELISA as described previously (Katsiadaki *et al*, 2002a). No other core end-point is employed by the AFSS.

37. Although the androgenised female stickleback screen was originally designed to specifically detect antiandrogens, it can also clearly detect chemicals with androgenic activity, through the induction of spiggin kidney content in the negative control groups relative to the control fish. In addition exposure to androgenic chemicals will result in a higher spiggin content in all treated fish relative to the positive control (DHT) groups

38. Annex 1 provides the details of the exposure conditions for all flow-through experiments that are reported here and Annex 2 a list of all 12 studies.

39. During the test development, we employed a semi-static system (not reported here) for administering five suspected antiandrogens, namely Fenitrothion (FN), Linuron (LN), Procymidon (PR), Vinclozolin (VZ) and p.p'-DDE. In addition, Flutamide (FL) was used as a model antiandrogen in both flow through and semi-static exposures. The exposures took place at Lab 1 and Lab 2 between 2002 and 2004. Using this system we provided the first evidence for *in vivo* antiandrogenic potential of LN and FN and confirmed the antiandrogenic properties of FL, PR, VZ and p.p'-DDE (Katsiadaki *et al*, 2006). A very smooth dose response was obtained with FL (figure 1).

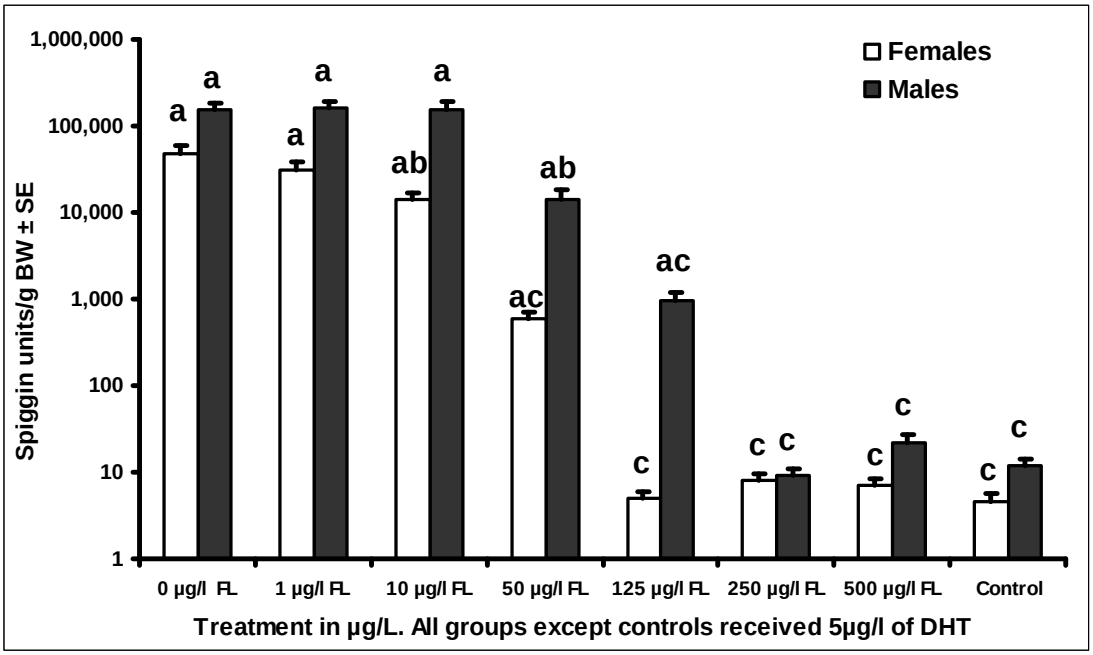


Figure 1: Inhibition of DHT-induced spiggin production by flutamide in a dose response manner (Katsiadaki *et al*, 2006. EHP 114 (Suppl 1):115-121).

40. Follow-up experiments were conducted as part of the AE1149 (Defra SD funding) and the EU-funded project EDEN between 2004 and 2007 in three laboratories, Lab 2, Lab 3 and Lab 4. An example of the typical data generated during these exposures, employing FN is illustrated in figure 2. Independent statistical analysis of all the data produced in four different laboratories over a 7-year period using the androgenised female stickleback screen is provided in this report.

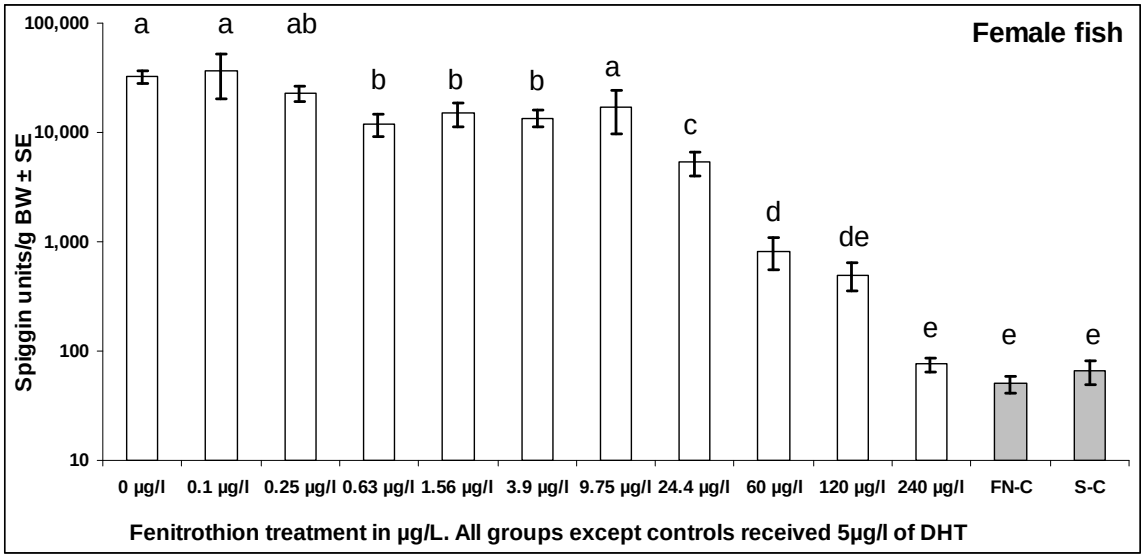


Figure 2: Inhibition of DHT-induced spiggin production by Fenitrothion in a dose response manner.

41. Although this assay has provided an excellent mechanistic model, unique in the fish world, for the detection of antiandrogenic xenobiotics it does not allow for a full assessment of the impacts that these chemicals might have on fish reproduction and particularly on fish populations. In addition, the test was regarded by the OECD group as an exposure to a chemical mixture due to the simultaneous use of androgen and a putative antiandrogen.

42. We therefore identified the need for a new model system where the effects of antiandrogens could be assessed using intact male and female sticklebacks (rather than androgenised females). For this we reviewed the well-described reproductive physiology of the male stickleback and designed an appropriate breeding test that would be more suited to the revised OECD guideline for an endocrine screen.

43. The initial intention within the OECD VMG-Eco was to develop a fish screen to detect EDCs that would employ fish in non-spawning status. Validation data during the phase 1a used this version of the draft guideline. However, this intention changed over the years and already in phase 1b the fish used were in an active spawning state. We therefore had to modify the stickleback screen in line with the new requirements and designed the stickleback-breeding test (see Annex 3).

44. The stickleback-breeding test can provide meaningful data for environmental risk assessment and address the risk posed by antiandrogenic chemicals to fish populations but it presents some drawbacks:

- It requires for all male fish to be in a non-breeding status prior to the test. If the reproductive status is variable, the level of endogenous androgens will also be variable and the statistical power of the test to detect significant decreases in spiggin content due to chemical exposure would be compromised. At the same time, the female fish need to be in an active breeding status in order to resume spawning within the 21-days of the duration of the test. The above require long-term planning and careful husbandry regimes that can only be achieved easily in laboratories with experience in keeping this species.
- Even though adequate photoperiodic regimes can largely control the stickleback reproductive status, in practice we see that some individuals attain sexual maturation in winter holding conditions. We have so far tested the system a few times and found that is it much easier to synchronise and control the reproductive system of male fish

45. Considering these drawbacks, (a) the fact that it requires a considerable amount of planning in order to pre-condition the fish and (b) the risk of failure when it takes place during the natural breeding season of the species, (between March and July in Northern Europe), it is debatable how useful the stickleback breeding test will be as a rapid screen for endocrine disrupters.

46. The results of an international ring test using this system (see report to Defra CPEC75) reflected these risks. Only one out of the three laboratories was able to demonstrate the antiandrogenic properties of flutamide by means of spiggin reduction/inhibition in male fish that were brought into breeding by photoperiod stimulation. Since the robustness and reproducibility of the spiggin assay itself was confirmed during the first international ring test (Allen *et al*, 2008), the apparent failure to obtain uniform results during the second ring test was presumably due to the inability of the two participating laboratories to control the male reproductive status.

47. To this end we revisited the possibility of reviewing the existing data produced by the androgenised female stickleback test in view of validating a new guideline that is complementary to the TGs 229 and 230 at Level 3 of the Conceptual Framework for the validation of tests sensitive to endocrine disrupters. The intended use will be specifically to screen chemicals that are thought to posses antiandrogenic activity hence prior information either based on their structure (i.e. QSAR) or *in vitro* assays (i.e. AR Calux, YAS screen) should indicate such a potential. The test we propose can confirm the antiandrogenic activity *in vivo* using androgenised female sticklebacks.

The effect of solvent and potassium permanganate on spiggin expression in the female stickleback screen.

Experimental design and methodology

48. A small-scale experiment was set up in view of answering two fundamental questions: a) is the use of solvent for administration of DHT as a model androgen necessary and b) is potassium permanganate (PP, $KMnO_4$) a suitable endocrine negative control compound in this test (i.e. does it display antiandrogenic activity)?

323 49. The already existing data included a large set of data from 'control' animals, including
324 fish that were exposed to solvent alone (methanol) and water controls and 'negative
325 control' animals where the putative antiandrogen was administered alone (with not DHT).
326 In this respect and considering all 3Rs in ethical experimentation we took the decision of
327 not including the same control groups (water, solvent, flutamide) in this test. Expert
328 statisticians supported this view and requested the inclusion of 3 replicate groups for each
329 treatment level since the two hypotheses were tested for the first time and there were no
330 other available data. Table 1 provided the details of the experimental design.
331 50. Methanol would not have been our solvent of choice with the current state of
332 knowledge (Hutchinson *et al*, 2006), but in order to make results comparable with the
333 previous datasets, we followed the same principles used in previous exposures. As
334 mentioned above, the test conditions were very similar to TG230 (flow through system,
335 loading, photoperiod, water changes in 24 hours, water sampling, etc). Methanol however,
336 wherever used reached a final concentration in the aquaria water of 0.1%, which is 10 fold
337 higher than the recommended OECD concentration.
338 51. After 21 days of exposure the fish were terminally anaesthetised and placed in liquid
339 nitrogen until dissection. The kidneys were dissected out and placed in 200µl of a strong
340 urea buffer for digestion and subsequent spiggin analysis according to Katsiadaki *et al*,
341 2002a.

342 **Table 1: Experimental design of the small-scale test.**

Tank	Treatment	No of fish
1	**DHT at 5µg/L (dissolved in water)**	20
2	**DHT at 5µg/L (dissolved in methanol)**	20
3	**DHT at 5µg/L and Flutamide at 250 µg/L (dissolved in water)**	20
4	**DHT at 5µg/L and Flutamide at 250 µg/L (dissolved in methanol)**	20
5	**DHT at 5µg/L and Potassium permanganate at 400 µg/L (dissolved in water)**	20
6	**DHT at 5µg/L (dissolved in water)**	20
7	**DHT at 5µg/L (dissolved in methanol)**	20
8	**DHT at 5µg/L and Flutamide at 250 µg/L (dissolved in water)**	20
9	**DHT at 5µg/L and Flutamide at 250 µg/L (dissolved in methanol)**	20
10	**DHT at 5µg/L and Potassium permanganate at 400 µg/L (dissolved in water)**	20
11	**DHT at 5µg/L (dissolved in water)**	20
12	**DHT at 5µg/L (dissolved in methanol)**	20
13	**DHT at 5µg/L and Flutamide at 250 µg/L (dissolved in water)**	20
14	**DHT at 5µg/L and Flutamide at 250 µg/L (dissolved in methanol)**	20
15	**DHT at 5µg/L and Potassium permanganate at 400 µg/L (dissolved in water)**	20
	Grand total	300

343

52. The chemical concentrations tested were selected as follows: DHT has been used at this concentration throughout the EDEN and previous exposures after fine-tuning of the antiandrogen test (Katsiadaki *et al*, 2006). FL at 250µg/L has been shown to result in a total inhibition of spiggin induction by DHT at 5µg/L (Katsiadaki *et al*, 2006; figure 1). PP was used at 400µg/L according to the OECD draft guideline for the fish screen as an endocrine negative compound.

53. Water samples (0.5-1L) were collected each week from each tank and filtered through a 0.45 µm filter (Pall Life Sciences, Portsmouth, UK) and a C18 solid phase extraction cartridge (SPEs, Sep-pak Plus; Waters Ltd., Watford, UK), as described by Ellis *et al*, (2004). Extracts were divided into two aliquots, one of which was used for the measurement of DHT and the other for flutamide.

54. The analytical method used for the verification of flutamide in the water has been previously described (Katsiadaki *et al*, 2006). For DHT we used an established radio-immunoassay (RIA), according to the same general procedure as used for the measurement of other steroids in our laboratory (Scott *et al*, 1994). Briefly, the extracts were dried under nitrogen gas at 45°C, reconstituted in 1 ml RIA buffer (0.5M phosphate buffer containing 0.2% bovine serum albumin, 0.8% sodium chloride, 0.03% EDTA and 0.01% sodium azide) and stored at -20°C. The recovery rate of DHT from the water was tested by means of 'spiking' tank water samples with known amounts of DHT. The recovery rate in all cases lay between 60% and 65%. Thus all measured levels of DHT were multiplied by a factor of 1.6 to correct for losses that occurred during extraction.

55. Statistical analysis: One way ANOVA was used to assess differences between female spiggin levels in the different groups at $p<0.001$, followed by pairwise comparison using t tests. All spiggin data were logarithmically transformed before analysis.

Results and discussion

56. Upon dissection, the sex of each fish was identified by visual evaluation of the gonads. Few male fish were present in the experimental populations. This is commonly the case as explained in paragraph 34, but since the majority of the male fish in this experiment were not in breeding status we included their responses to the treatments. In reality, more than two thirds of the fish in each tank/treatment were correctly identified as females, which was above our target number. Table 2 provides the details on fish sex and mortalities in each treatment.

Table 2: Fish sex and mortalities during the small-scale test.

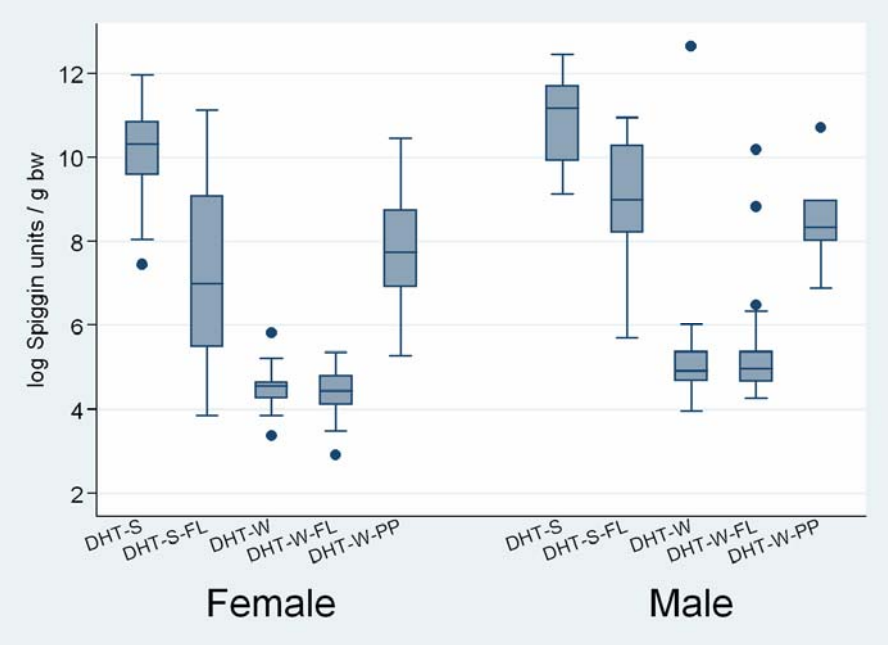
	No of females	No of males	Mortalities	% Mortality
DHT-Solvent	46	13	1	1.67
DHT-FL-Solvent	51	9	0	0
DHT-Water	46	14	0	0
DHT-FL-Water	41	17	2	3.33
DHT-PP-Water	38	5	17	28.33

57. As table 2 displays the mortalities were low (as expected) with the exception of the PP treatments, where high mortality rates were observed during the second week of exposure. Since the severity limit of the experiment was set as low, we informed the Home Office Inspector who advised us to reduce the concentration from 400µg/L to 200µg/L during the final week of exposure. This measure precluded further mortalities.

58. Another observation we made during the exposures was that the level of feeding in the PP treated groups was impaired. Appetite was present since fish tried to feed but it appeared as if the chemical altered the palatability of the food since in most cases initial intake by the fish resulted in vomiting.

59. The responses of fish to the treatments by means of kidney spiggin content are displayed in figure 3. The answer to the first question of our investigation of the necessity of using solvent for androgen administration is clearly yes it is required. DHT putatively dissolved in water does not have any androgenic properties, as it did not induce any spiggin in the female fish. In addition, the male fish spiggin levels showed a huge variation, which is typical in fish that received no androgen treatment as it reflects the social hierarchies in the exposure tank, where depending on the loading only 1 or 2 males are reproductively active, the rest remain suppressed.

60. Although the DHT stock solutions were made in water and after sonication for 30 minutes the compound was apparently dissolved, the results suggest that the androgen either came out of solution when diluted further with aquaria water in the tank or it was quickly degraded in water but not in solvent (presumable due to bacteria action). The results of chemical analysis in the aquaria water as means of all replicate tanks that were sampled in at least 3 occasions each (one each week) are displayed in table 3. The lack of DHT recovery in the aquaria water is in full agreement with the biological responses and suggests that either DHT solubility or stability in water or both are very low, affecting the expected biological responses.



406 **Figure 3: Spiggin responses of female and male sticklebacks after exposure to**
407 **DHT alone or in combination with FL or PP. Stock solutions of DHT were made**
408 **either in methanol (DHT-S) or in water (DHT-W).**
409
410 61. One-way ANOVA of ln(spiggin) for females shows means differ at $p<0.0001$. Selected
411 t-tests then show that:
412 • DHT-W group has the same mean as DHT-W-FL ($p>0.7$), which was expected as we
413 observed no spiggin induction in the DHT group
414 • DHT-S group has higher spiggin levels in comparison to DHT-S-FL ($p<0.0001$), which
415 was expected as we are aware of the antiandrogenic potential of flutamide, so it yet
416 another confirmation that the assay can detect antiandrogens
417 • DHT-W-PP has significantly lower spiggin levels than DHT-S ($p<0.0001$), and
418 significantly higher spiggin levels that DHT-FL-S (($p<0.0001$).
419
420 **Table 3: Analytical verification of test compounds in aquaria water.**

	DHT (µg/L) nominal	DHT measured	Fl (µg/L) nominal	Fl (µg/L) measured
DHT-Solvent	5	3.38	0	0
DHT-FL-Solvent	5	3.49	250	180
DHT-Water	5	0.51	0	0
DHT-FL-Water	5	0.89	250	228
DHT-PP-Water	5	3.84	0	0

421

62. The fish responses to the DHT and DHT plus flutamide treatments when a solvent was used were as expected. DHT induced spiggin in female fish and this induction was reduced/inhibited by flutamide.

63. Based on our previous data we would expect FL at the nominal concentration of 250µg/L to totally inhibit spiggin induction by DHT at the nominal concentration of 5µg/L. However, we only observed a significant reduction (and not total inhibition), a fact that can be partly explained by the levels of flutamide recovered in the aquaria, which were 72% of the nominal concentrations (see table 3).

64. In addition, we used a batch of flutamide that was also used during the EDEN exposures (it is common practice when comparable data need to be generated), which means that the compound was at least 4 to 5 years old. As there is no expiry date for this product stated by the supplier, we assumed that the chemical is stable over long period of time. However, when we conducted the manufacturer they explicitly stated that they would guarantee activity for only 1 year after purchase. To this end, it seems logical to assume that the flutamide lot we used was beyond its acceptable self-life and hence likely to have reduced biological activity.

65. Nevertheless besides this caveat the results were conclusive in that DHT is androgenic in fish when administered with a solvent vehicle and that flutamide moderates strongly this androgenic response by antagonism.

66. The answer to the second question of our investigations was less straightforward. Since we obtained firm evidence that administering DHT in water did not lead to the expected fish androgenisation (as it does when administered in solvent) we would expect to see no spiggin induction in the DHT-W-PP treated groups. There was however a strong response in all three replicate tanks, with spiggin levels being significantly higher in comparison to the spiggin levels of female fish treated with DHT in water alone and significantly lower in comparison to those of female fish treated with DHT in methanol alone (figures 3 and 4).

67. Our first reaction to the unexpected spiggin induction was to look at the possibility of PP altering the DHT structure since it is a highly oxidising agent, employed in the chemical synthesis of steroids. The oxidising capacity of PP is used in steroid chemistry to hydroxylate double carbon-carbon bonds, and has been used as an oxidising agent to convert the double bond at position 4-5 of testosterone to dihydroxy alcohols in a method to measure DHT in serum (Shiraishi *et al*, 2008). However, unlike testosterone the

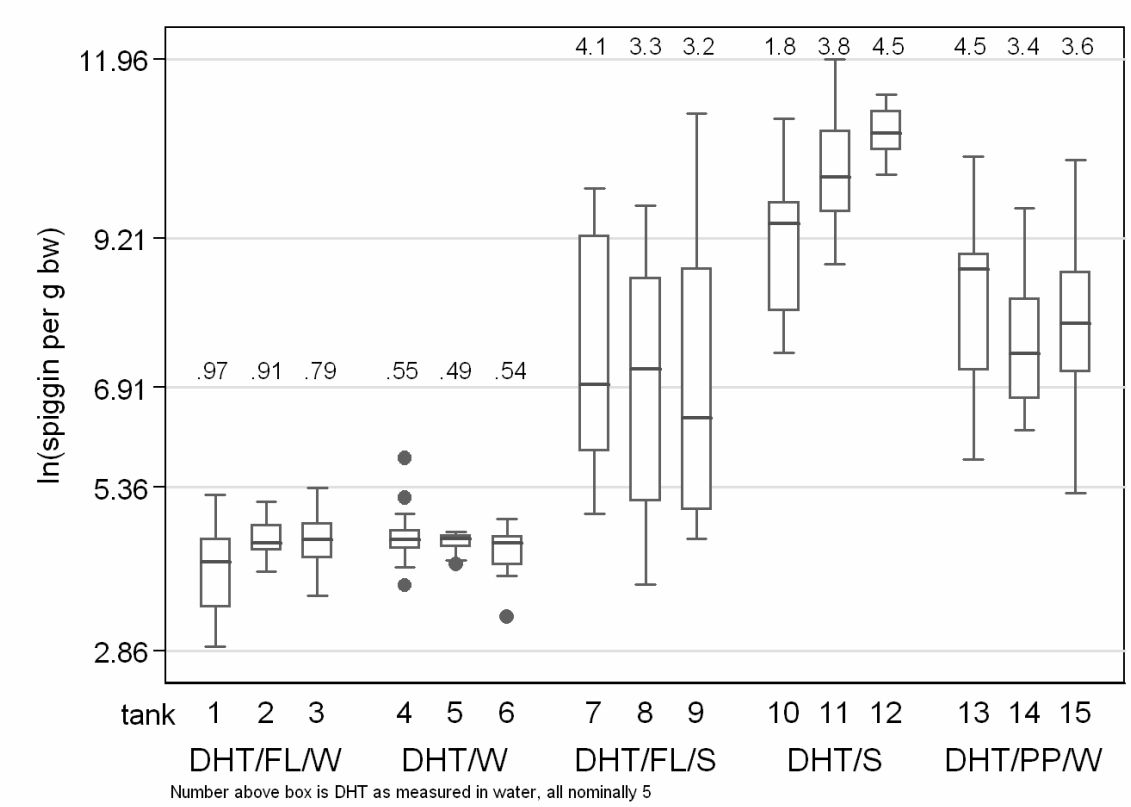


Figure 4: Spiggin responses of female and male sticklebacks after exposure to DHT alone or in combination with FL or PP in individual tanks (data are the same as in figure 3 but separated by tank).

68. Furthermore, the DHT concentrations in the PP treatments were close to the nominal and similar to those obtained from aquaria where methanol was used for DHT administration (table 3). We therefore propose that the presence of PP in the aquaria water resulted in a greater stability of the androgen in water maintaining its biological activity.

69. The lack of spiggin response and inability to detect DHT in the aquaria of fish treated with DHT-W was attributed to the low solubility and stability of DHT in water. The antibacterial properties of PP would favour the explanation that this was a demonstration of the steroid's low stability and high degradation by bacterial in water (rather than low solubility). This 'fixing' of DHT in water by PP is most likely responsible for the androgenic properties of this mixture and the resulting spiggin induction. The spiggin response in the DHT-W-PP treated groups was lower to that of DHT-S groups, although

70. The alleged low stability of DHT in water might be responsible for the generally low androgenic activity of sewage effluents (Thomas *et al*, 2002) besides the fact that men excrete larger amounts of androgens in the urine than the levels of oestrogens excreted by women. The apparent ability of PP to 'fix' the androgen in aqueous solutions merits further investigations as it could provide (if the same properties are present at lower PP concentrations) an alternative way of stabilising androgens for aquatic *in vivo* exposures.

Conclusions

71. There is a need to employ solvent vehicle during *in vivo* exposures of fish to DHT, a model androgen. Although we did not specifically set out to determine the level at which the use of carrier solvent is necessary, our results suggest that the presence of water in the stock solutions or mixing vessels would result in a high degradation rate of DHT.

72. The use of PP as an endocrine negative substance at 400µg/L is not supported. Firstly PP was toxic to the fish, secondly it affected food uptake by the fish, thirdly its use as an oxidising agent for steroid synthesis are compromising its potential as an endocrine negative substance to a large extend.

Statistical analysis of all existing data to date.

Background information on the datasets

73. As stated above a large dataset comprising of 20 exposures undertaken in 4 different laboratories over a period of 7 years was available. Of these, 8 exposures were excluded because 4 were conducted in a semi-static system (not flow-through) and 4 were a multi-component mixture of all antiandrogens, hence not suitable for the purpose of this exercise.

74. The sub-set of data analysed comprises of 12 full dose response exposures that took place in four different laboratories, the majority of which (10 exposures) were part of the EU-funded EDEN programme. These tests included a total number of over 1600 independent measurements of spiggin in female fish that were treated with DHT at 5µg/L and range of antiandrogen concentrations.

75. These exposures were conducted in a similar (but not identical) manner (see Annex 1 for details). The main differences between them were in relation to the origin and size of

512 76. Hence some degree of variability was expected in relation to biological responses and
513 analytical measurements, stemming from one or more of these factors. The assay requires
514 at least two analytical measurements to be made at each time point (with the exception of
515 positive and negative control groups where only one chemical is administered); one is the
516 DHT concentration and the other the test chemical.
517 77. Since the amount of DHT in the water increases spiggin content in female kidneys in a
518 dose response manner (Katsiadaki *et al*, 2002a), then the degree of antagonism by the
519 putative antiandrogens would be directly linked to the amount of DHT present in these very
520 same tanks (i.e. not the positive control tanks alone). For example, assuming that in one
521 particular experiment, the average spiggin levels of the positive control groups were
522 60,000-units/g body weight and the measured DHT concentration (average of 3 weekly
523 measurements) to which this response is attributed was established as 4.3μg/L. Assuming
524 also that at the same experiment, the nominal concentration of 50μg/L flutamide was also
525 tested (this is always in combination with the nominal concentration of DHT at 5μg/L) and
526 this gave an average spiggin level of 62,000-units/g body weight. The first reaction is to
527 conclude that flutamide had no effect at this concentration. This can be proved to be
528 incorrect when one examines the actual concentrations. It is possible for example that the
529 actual DHT concentration in the flutamide treated group was 6.4μg/L in this instance which
530 would have resulted in significantly higher spiggin induction in comparison to the positive
531 control group, and hence potentially to a significant reduction by flutamide. Similarly, the
532 flutamide actual concentration could be as low as 10% of the nominal, so to state that
533 flutamide has no effect at 50μg/L would be also incorrect. We therefore suggest that the
534 best way of expressing the data was to use the ratio of the antiandrogen over the androgen,
535 as in this way both important determinants of the responses are present.
536 78. Another important issue we considered was whether or not to analyse the data on the
537 basis of nominal or measured test compound concentrations. Measured concentrations

79. It should be mentioned that at least two variables were kept standard throughout the generation of the reported data. All spiggin measurements were conducted at Lab 2 employing the same specific reagents (i.e. standard and anti-spiggin serum), all analytical measurements for DHT were also conducted in Lab 2 (by Dr Alex Scott and his team) and all the analytical measurements of the antiandrogens tested were conducted by Dr Steve Morris in either Lab 1 or Lab 2.

80. Only few caveats were associated with the above methodologies: a) the standard curve (SC) of the spiggin assay was modified in 2003 after recommendations from the EDEN statistician so the initial SC included data points from 1 to 1000 units and the modified version (which was used thereafter for all measurements) ranged from 0.02 to 200 units, b) the label for DHT was recently replaced (the original stock was exhausted during the EDEN analysis), c) at least three different instruments were used for flutamide analysis alone as a result of equipment replacements in the Cefas analytical suite.

81. For brevity the details of water extraction, analytical methods and spiggin analysis protocols are not included in this report but they are available upon request. The ELISA for spiggin is also attached at the draft test guideline accompanying this report.

Analysis of control data

82. The first task was to analyse all 'control' data including water control, solvent control, and negative control (where only the test compound was used without any DHT) and investigate any differences between the different groups, the different Labs and the time that exposure and analysis took place (date).

83. Spiggin data were normalised using a natural log transformation and comparisons between spiggin levels within each factor of interest were made using a one-way ANOVA. All spiggin data were expressed as spiggin units per g of body weight but in some figures this is abbreviated to ln spiggin. Where a significant difference was detected, pairwise comparisons were made using t-tests (adjusted for multiple comparisons by the Bonferroni method). Differences were deemed significant if the adjusted p-value was less than 0.05. All analysis was conducted in Stata (version 10.0) and R (version 2.7). The results are displayed in figure 5, whilst table 4 provides the details.

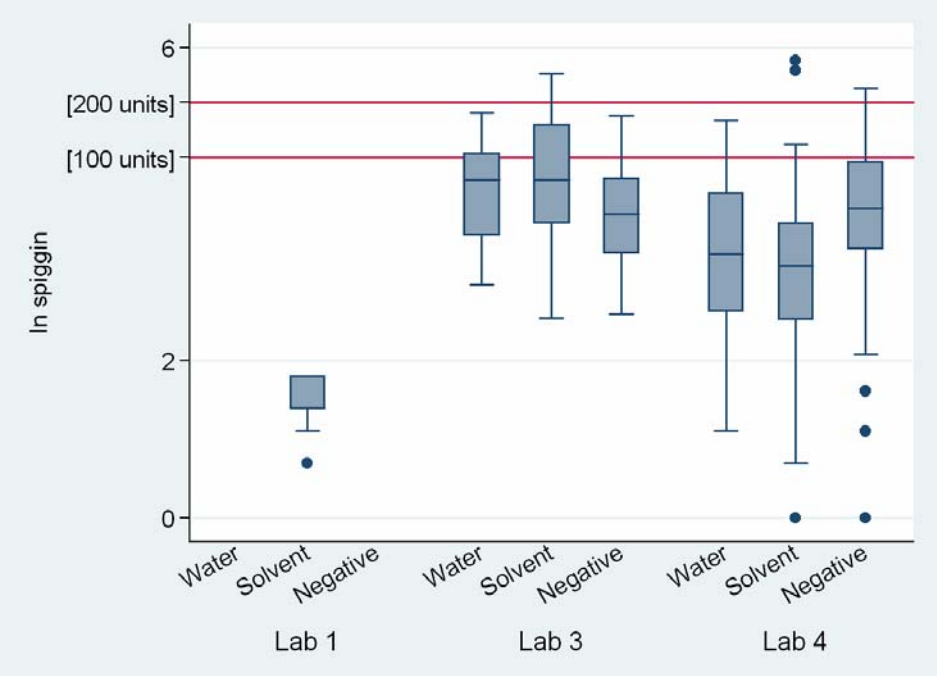


Figure 5: Water, solvent and negative control spiggin data in each laboratory (all studies).

84. Both water, solvent and negative control groups registered spiggin values that are within the baseline levels of the assay (i.e. less than 200units/g body weight), indicating that the assay is robust and reproducible in terms of recognising female fish that were not treated with an androgen. Importantly there were no differences between water and solvent control groups within each laboratory exposure. The study conducted by Lab 2 did not include any control data as explained in paragraph 49.

591 **Table 4: Details of solvent, water, negative and positive controls in each lab.**

Site	Solvent control				
	Mean Spiggin	SD	N	SE	Mean ln spiggin
Lab 1	4.66	1.12	19	0.26	1.54
Lab 2	No solvent controls were used				
Lab 3	76.96	62.42	22	13.31	4.34
Lab 4	43.08	60.96	60	7.87	3.76
Average	42				3.22
SD	36				1.48
SE	21				0.85
CV	0.87				0.46
	Water control				
	Mean Spiggin	SD	N	SE	Mean ln spiggin
Lab 1	No water controls were used				
Lab 2	No water controls were used				
Lab 3	77.42	42.50	18	10.02	4.35
Lab 4	42.56	38.46	71	4.56	3.75
Average	60				4.05
SD	25				0.42
SE	17				0.21
CV	0.41				0.10
	Negative control				
	Mean Spiggin	SD	N	SE	Mean ln spiggin
Lab 1	No negative controls were used				
Lab 2	No negative controls were used				
Lab 3	57.52	36.84	21	8.04	4.05
Lab 4	65.99	51.03	63	6.43	4.19
Average	62				4.12
SD	6				0.10
SE	4				0.05
CV	0.10				0.02
	All baseline data				
Average	52.60				3.71
SD	26.16				1.74
SE	10				1
CV	0.50				0.47
	Positive (DHT) control				
	Mean Spiggin	SD	N	SE	Mean ln spiggin
Lab 1	55,367	46,047	19	10,564	10.92
Lab 2	38,666	32,215	46	4,750	10.56
Lab 3	33,715	21,289	19	4,884	10.43
Lab 4	46,268	31,936	75	3,688	10.74
Average	43,504				10.66
SD	9,444				0.22
SE	4,722				0.11
CV	0.22				0.02

85. However, statistical analysis indicated differences in the spiggin levels of 'control' fish depending on the site that the exposures were conducted. Specifically the study conducted by Lab 1 in 2001 was registering lower values in comparison to the Lab 3 and Lab 4 studies and in addition the Lab 4 fish were registering lower values than the Lab 3 fish.

86. The latter differences (between Lab 3 and 4) could be attributed to random biological variability or could be related to genuine differences between the fish populations used for the exposures. Lab 3 for example used as a regular source of sticklebacks for exposure a freshwater site near Reading whilst the fish used by Lab 4 were of marine origin collected from fiords in Norway.

87. The differences between both Lab 3 and Lab 4 to the Lab 1 data are most likely due to the change in the spiggin assay standard curve in 2003 as mentioned earlier. This change was requested by the EDEN statistician who needed a higher resolution at the low end of the assay, in order to model low dose effects. The different standard curve employed in the Lab 1 study was underestimating slightly the spiggin units in control females.

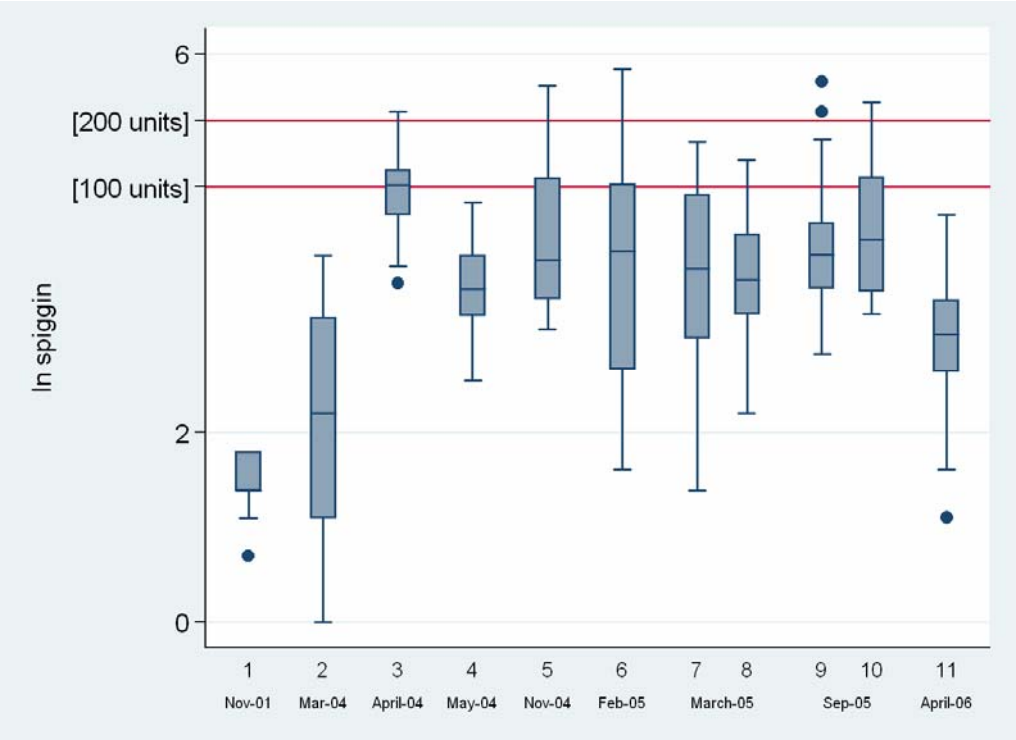


Figure 6. Water, solvent and negative control spiggin data from all exposures marked by study number.

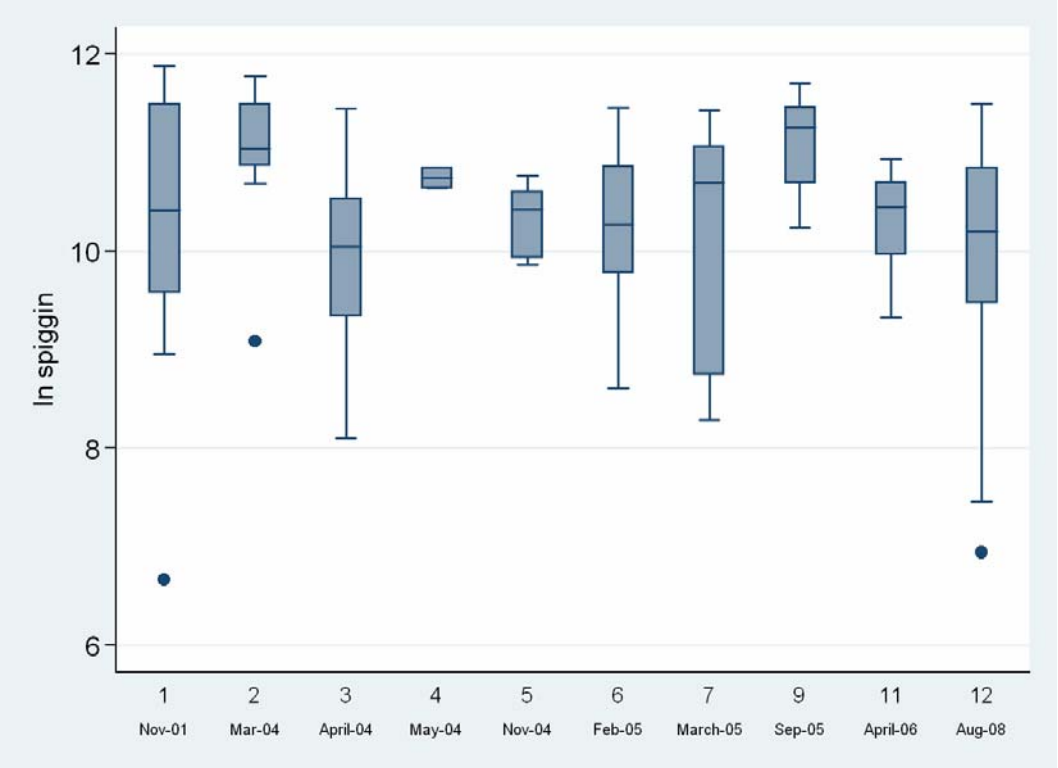
88. Lastly, statistical analysis of all baseline control data (water, solvent and negative) pointed out significant differences between dates that the exposure and analysis took place (figure 6). The differences involved studies 1, 2, 11 (which registered lower baseline spiggin levels) and study 3 (which registered higher baseline spiggin levels). The reason for these differences is unclear, however it is arguable that it is of any biological significance, since all measurements were within the defined baseline levels

89. Taking everything into account it is fully justified to claim that the control female spiggin data are consistently within the baseline (background) limits of the assay (below 200 units/g body weight) and that there were no differences between solvent (methanol) and water control data.

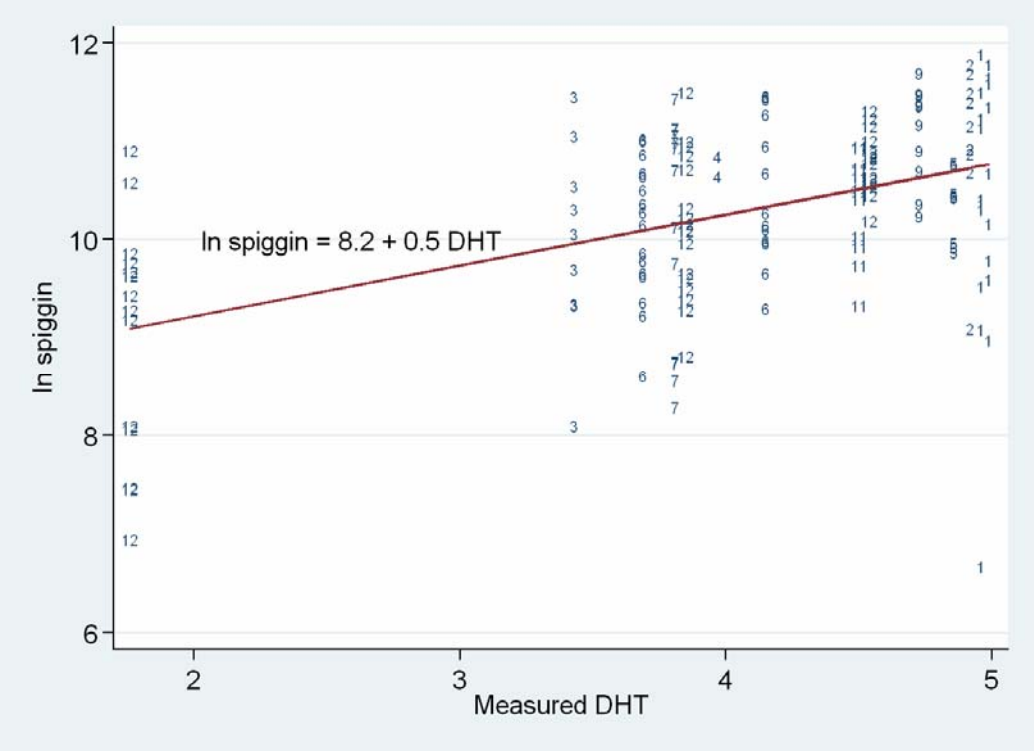
90. Analysis of positive control data: The DHT concentration was always nominally 5μg/L, but the measured values varied from 1.76 to 4.99μg/L (mean 4.13μg/L). Excluding the 1.76 value (in one of the replicate tanks from the Lab 2 study), the range was 3.43 to 4.99. The spiggin values varied from 782 to 156,354. These were converted to the ln scale. Although the results presented here refer to 12 full dose experiments, in 2 occasions (studies 8 and 10) two chemicals were tested simultaneously sharing the control fish data, hence the number of studies compared in terms of spiggin responses to the DHT treatment were 10. The number of observations in each studied varied from 2 to 45 but were around 10-15 in most cases.

91. Comparing the spiggin values between dates suggests differences in variance rather than level and AVOVA demonstrated that the visible variation is more related to number of observations. The overall significance at 4% is so marginal that one would accept the system as stable. There is certainly no support for a trend over time (learning effect, drift of standards), whilst excluding study 9, conducted at Lab 4 in September 2005, the differences between studies are completely non-significant (figure 7).

92. The relationship of the measured (ln) spiggin response to the actual DHT concentration is apparently linear over the observed range. If DHT concentrations varied over orders of magnitude the relationship would be more likely ln-ln. Figure 8 presents the same data but fitted with a regression line. In conclusion, the spiggin response after DHT exposure is stable over time and between laboratories but needs to be adjusted for the actual DHT burden. Points marked by study number.



642
643
644 **Figure 7: Positive control (DHT at nominal concentration of μg/L) spiggin**
645 **responses from all exposures, marked by study number.**
646



647
648 **Figure 8: Positive control (DHT at measured concentrations in μg/L) spiggin**
649 **responses from all exposures marked by study number.**
650

Analysis of Flutamide data

93. Antiandrogens in general exert their effects by occupying androgen receptors (AR) without activating them (e.g. acting as an antagonist). Two classes of antiandrogens are currently recognised: the steroidal derivatives (which posses mixed agonistic and antagonistic androgenic activity) and the non-steroidal derivatives or 'pure' antiandrogens, exemplified by flutamide (2-Methyl-N(4-nitro-3-[triflouromethylphenylpropanamide], henceforth referred to as FL).

94. FL and its derivatives are the main representatives of the latter category and have been extensively studied due to their proven clinical efficacy in the treatment of prostate cancer (Singh *et al*, 2000). They are potent AR antagonists that compete with androgens for binding to the AR and prevent AR DNA binding and transcription of androgen-dependent genes. For this reason FL is used as a model, reference antiandrogen in *in vitro* systems that employ ARs, such as the YAS and was extensively tested in the development and validation of the androgenised female stickleback screen.

95. A total number of 5 full dose response studies with over 900 individual spiggin measurements were analysed. Figure 9 displays the results separated by study and Figure 10 displays the same data separated by site of exposure. Both figures use the nominal ratio between FL and DHT. Since DHT was always used at a nominal concentration of 5µg/L the actual concentration of FL can be quickly estimated by multiplying the ratio FL/DHT featuring on the X-axis by a factor of 5.

96. As figures 9 and 10 and statistical analysis reveal, although the responses were highly variable, flutamide when used at a ratio above 2 (hence at 10µg/L or above), significantly reduced the spiggin levels induced by DHT alone ($p<0.01$) and when used at a ratio above 20 (i.e. 100µg/L or above) the reduction in spiggin levels was highly significant ($p<0.001$).

97. Statistical analysis of the Lab 2 dataset (study 12) revealed that as expected FL significantly reduced the spiggin levels induced by DHT (only the solvent data were used for comparisons as the DHT in water did not induced spiggin for the reasons discussed in paragraph 59).

98. The results from the Lab 1 exposure indicated that when FL was used at a ratio to DHT equal or above 2 significantly reduced the spiggin levels induced by flutamide at $p<5.1\ 10^{-6}$ (i.e. highly significant), hence the Lowest observed effect concentration (LOEC) for FL was determined by these data to be as low as 10µg/L

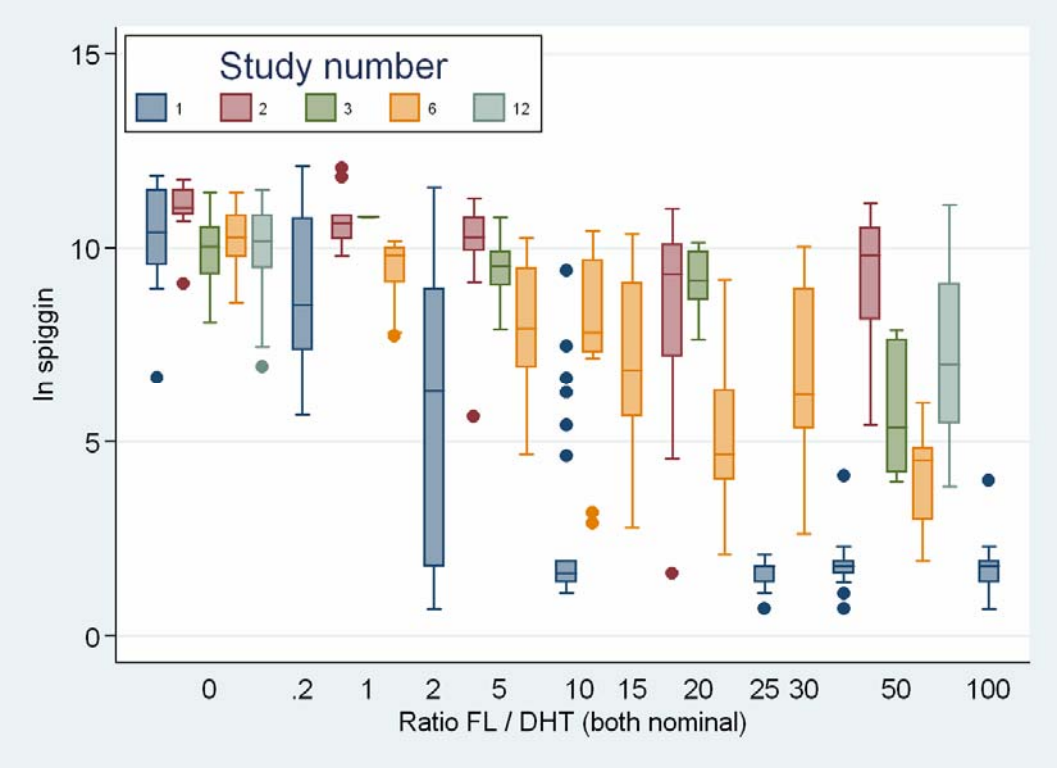


Figure 9: Spiggin responses to flutamide from all studies plotted against the ratio of nominal FL/DHT concentrations.

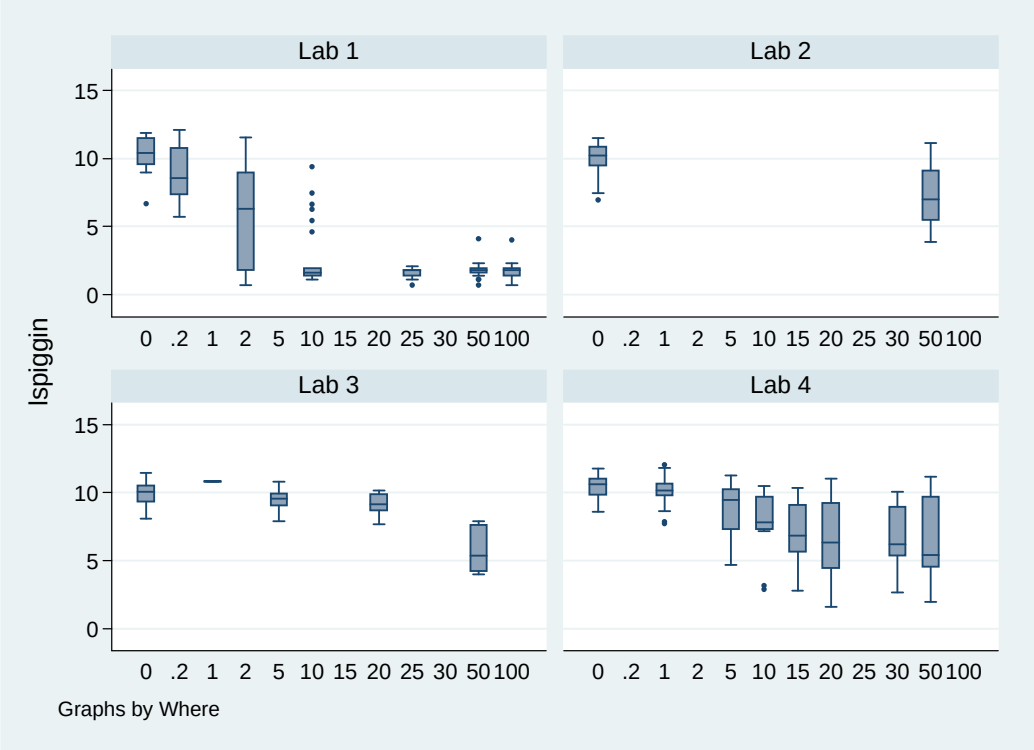


Figure 10: Spiggin responses to flutamide from each site plotted against the ratio of nominal FL/DHT concentrations.

99. The Lab 3 data were very different to the Lab 1 study in that they indicated significant reduction of spiggin by FL only when the ratio was 50 (hence the LOEC for flutamide for the Lab 3 data was 250µg/L).

100. The data obtained from the Lab 4 exposures were in between those of Lab 1 and Lab 3, providing significant reduction of spiggin by FL at a ratio equal or above 15 ($p<0.001$) giving a LOEC for FL of 60µg/L.

101. All the above data however, refer to the nominal concentrations of the chemicals and any significant differences between the responses could be simply related to the actual concentrations achieved during the exposures. Figure 11 presents all the FL data (marked by site) expressed as ratio of measured FL over measured DHT and figure 12 present the same data separated by site.

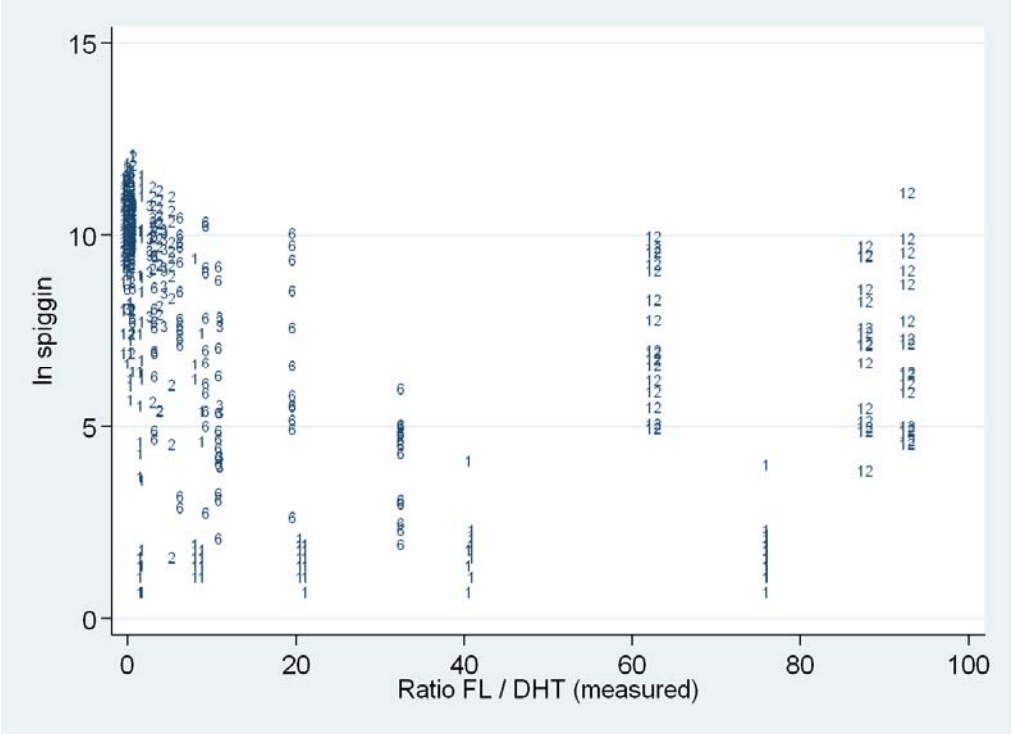


Figure 11: Spiggin responses to flutamide from all sites plotted against the ratio of measured FL/DHT concentrations.

102. The spiggin responses are much improved when using the measured concentrations of FL and DHT (as opposed to the nominal concentrations), with the exception of the Lab 2 data (study 12). One explanation for the anomalous Lab 2 data is the fact that the same batch of FL was used as for the EDEN exposures and because the "use by" date had been exceeded.

103. For this analysis of measured FL concentrations a different statistical tool was employed. The concentrations of test compounds measured are continuous (i.e. not discrete as in the nominal concentrations) and so a standard linear regression was run for each of the sites using logarithmically transformed spiggin values.

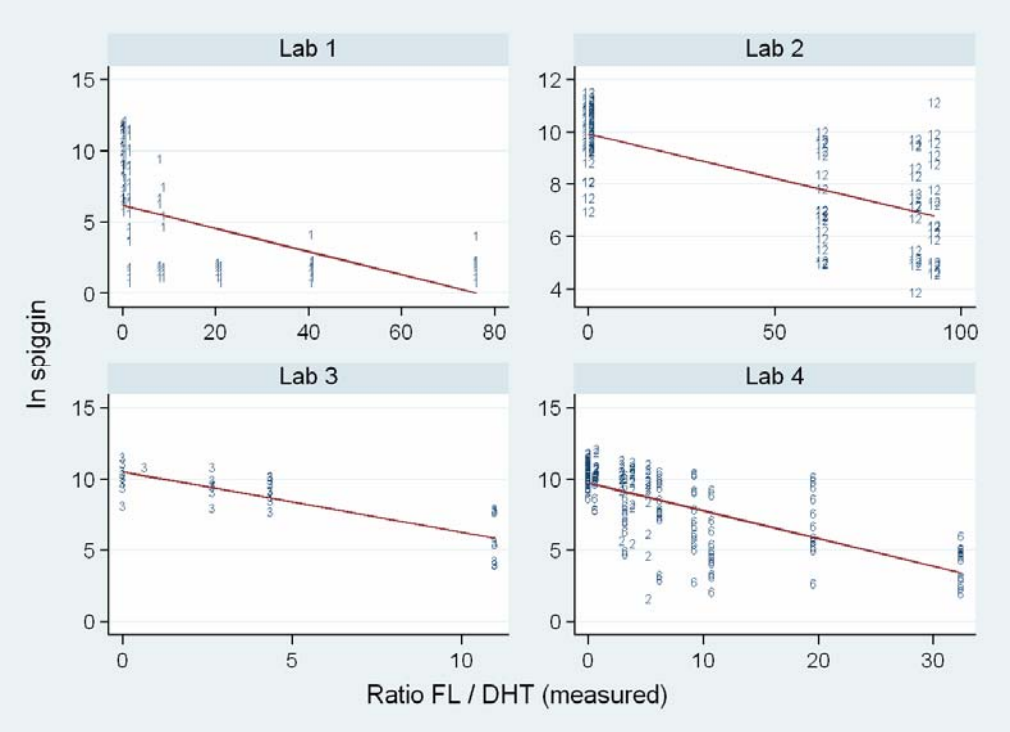


Figure 12: Spiggin responses to flutamide from each site plotted against the ratio of measured FL/DHT concentrations.

104. As figure 12 displays within each exposure site a dose response was observed showing inhibition of spiggin induction by FL in every single experiment. The measured concentrations of FL were on average 60% of the nominals (18-87%).

105. The main findings of the linear model applied in all FL exposures in respect to the spiggin data indicate that the intercept of all experimental data for each site are similar (9.7, 9.9, 10.4) with the exception of the Lab 1 exposure where the intercept was lower (6.3). We attribute this difference to the different standard curve in the spiggin assay employed as discussed earlier (paragraph 87).

106. Another important feature of this analysis is the slope of the fitted curve. In all cases this was negative, hence FL antagonised DHT in a dose response manner. The result was very similar between each site over the same range of the ratio's FL/DHT (0-30) with the exception of the Lab 2 exposure where the slope was flatter (-0.034). The reason for this

107. Using the measured chemical concentrations improved the regression line for each experimental site underlining the importance of chemical analysis of aquaria water as required by the OECD test guidelines. Direct chemical analysis harmonises any experimental differences of protocol or chemical stability, which might affect the result of the assay.

108. In order to evaluate whether the deviations between nominal and measured concentrations in each experiment affected the interpretation of biological responses, we plotted nominal versus measured concentrations for each site where the exposures took place (figure 13). As figure 13 displays, a very good agreement between nominal and measured concentrations existed for the Lab 1 and Lab 3 exposures, a good agreement was evident in the Lab 4 data (although at least one exposure deviated from the straight line) and a rather poor agreement was shown for the Lab 2 exposure, forming a contributing factor to the generally anomalous response recorded at this experiment.

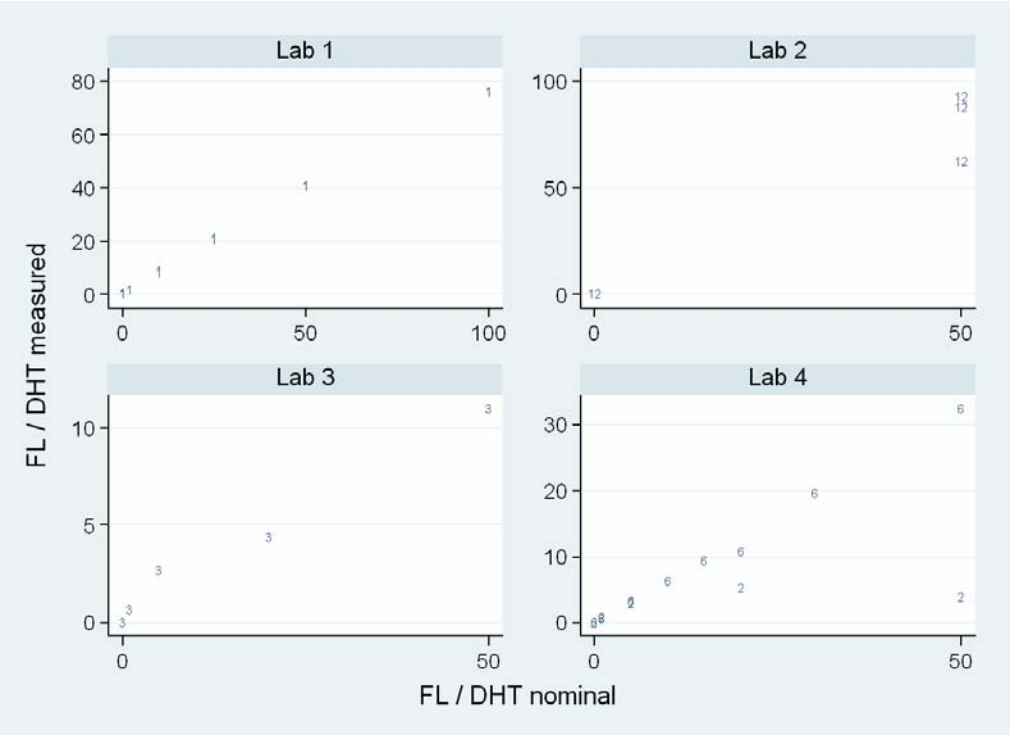


Figure 13: Measured versus nominal FL/DHT concentrations from each site.

Analysis of Fenitrothion data

109. Fenitrothion [0,0-dimethyl-0-(4-nitro-m-tolyl)phosphorothioate, henceforth referred to as FN], is an insecticide (organo-phosphate, OP) that has been widely used since 1959 to control insects in agriculture and for fly, mosquito, and cockroach control in public health programs. Since 2001 FN has not been approved for plant protection products in the UK, but it only registered for use as an insecticide for non-agricultural use in UK (Advisory Committee on Pesticides, 2006). Recently, the Commission of the European Communities decided on the withdrawal of FN for plant protection products to be implemented by the 25 November 2008 (2007/379/EC).

110. The 48-hour acute LC50 values for carp range between 2.0ppm (µg/L) and 4.1ppm. One source states that aerial spraying of FN at 2 or 3 oz/acre, on New Brunswick forests has been reported to have no deleterious effect on fish in streams in the treated area. The sublethal effects of FN exposure on fish include morpho-anatomical changes, behavioural changes, biochemical changes (inhibition of acetyl cholinesterase activity), respiratory effects and effects on growth. However, none of the published risk assessments has included potential endocrine disrupting activity.

111. FN has structural similarities with FL and the environmental antiandrogenic herbicide Linuron (Tamura *et al,* 2001). One study on FN found it to be an AR antagonist both *in vitro* (reporter gene assay) and *in vivo* using the Hershberger rat assay (Tamura *et al,* 2001). Other *in vivo* studies on FN, however, have been inconclusive about its antiandrogenic activity (Sunami *et al*, 2000; Sohoni *et al*, 2001).

112. The androgenised female stickleback assay employing a semi-static system of exposure provided the first evidence on its antiandrogenic properties in fish (Katsiadaki *et al,* 2006), whilst the stickleback breeding test revealed that FN does not only inhibit spiggin induction in the male but also affects the expression of typical male reproductive behaviour (Sebire *et al*, 2009).

113. Two full dose response studies (4 and 5), both conducted at Lab 3 with over 260 individual spiggin measurements were analysed. As explained in paragraph 78 we decided to analyse and present the data both in terms of nominal and aquaria measured test compound concentrations.

114. Figure 14 displays the results from both FN studies on the basis of the nominal ratio between FN and DHT. Since DHT was always used at a nominal concentration of 5µg/L

115. According to figure 14, the NOEC for FN was estimated as 25µg/L and the LOEC as 60µg/L (although in one occasion the concentration of 24.4µg/L significantly reduced spiggin induction by DHT).

116. Figure 15 presents the same spiggin data as figure 14 but the responses are expressed as a ratio of measured concentrations of FN over DHT whilst figure 16 shows the same data with a fitted linear regression line. This ratio was logarithmically transformed to provide a better linear fit.

117. To summarise, FN significantly reduced and totally inhibited spiggin induction by DHT at 60µg/L and 240µg/L nominal concentrations respectively suggesting that FN has equal to or even higher antiandrogen potential than FL. The concentrations of measured FN were very different between the two experiments. In the first experiment the recovered concentrations were low ranging between 11 and 24% of the nominals but in the second experiment recovery was improved and the range was between 60 and 74% of the nominals.

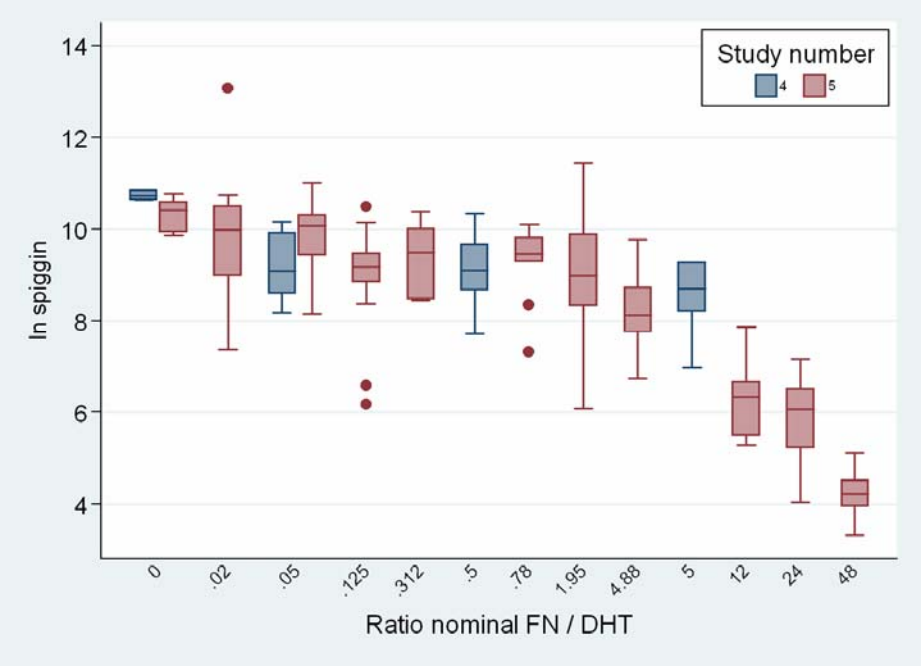


Figure 14: Spiggin responses to Fenitrothion plotted against the ratio of nominal FN/DHT concentrations.

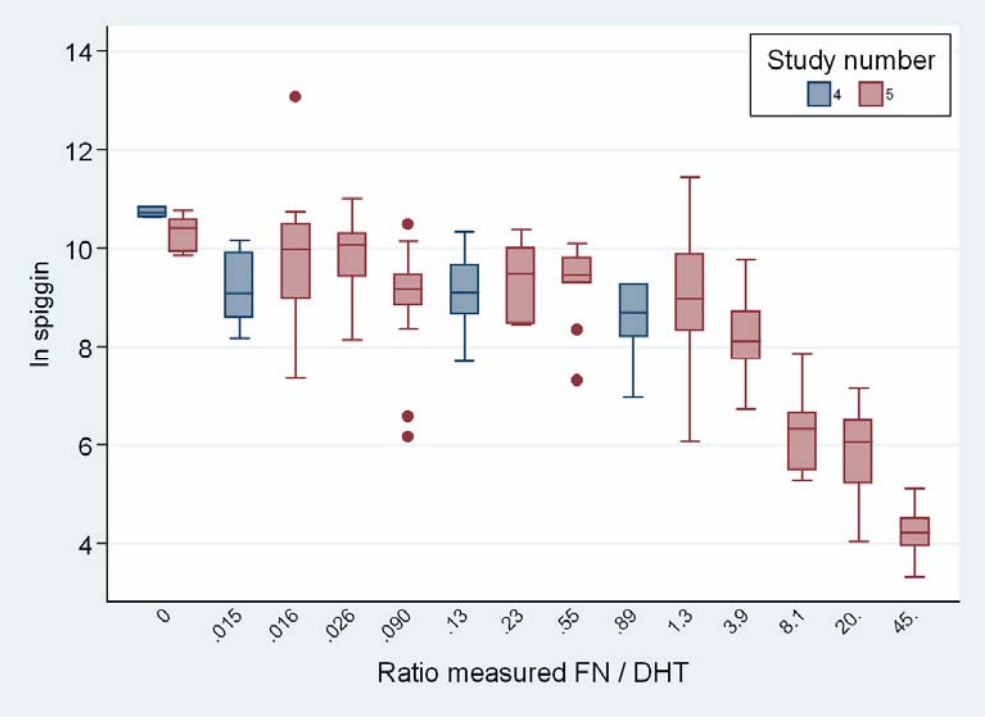


Figure 15: Spiggin responses to Fenitrothion plotted against the ln ratio of measured FN/DHT concentrations.

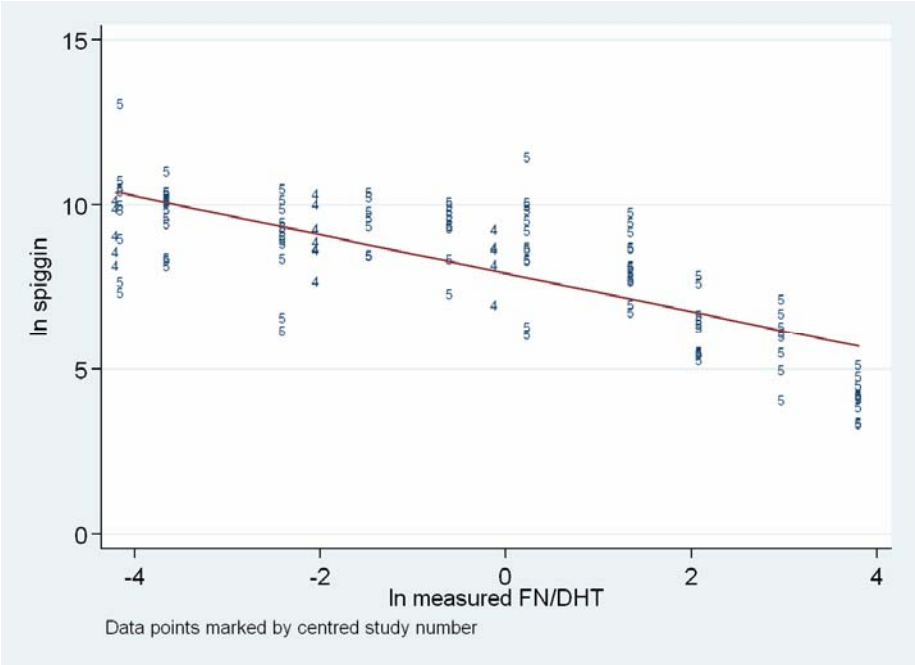


Figure 16: Spiggin responses to Fenitrothion plotted against the ln ratio of measured FN/DHT concentrations.

809

810 ***Analysis of Linuron data***

811 118. Linuron (LN) is a phenyl urea herbicide applied to suppress broadleaf and grassy
812 weeds. It is registered in the UK and other parts of the EU, having recently undergone an
813 extensive re registration review. It is also registered in the USA and is listed as showing
814 "no reproductive effects". Its use rate is 950 g/Ha in the EU with a soil persistence of T/2
815 of 30-150 days.

816 119. It is soluble in water (63.8 mg/l at pH 7, 20°C) and may be transported into aquatic
817 environments by run off. It is not readily broken down in water. Given its use rate and
818 physical properties one could expect that it might achieve substantial concentrations in the
819 aquatic environment.

820 120. LN is slightly toxic to fish and aquatic invertebrate species. The reported LC50 for
821 LN in trout and bluegill is 16ppm. The median threshold levels, i.e. levels at which
822 adverse, sublethal effects were apparent in 50% of the test animals, are greater than 40ppm
823 in crawfish and tadpoles exposed over a 48-hour period.

824 121. Linuron is a weak competitive AR antagonist *in vitro*, induces a positive response in
825 the immature and adult rat Hershberger assay, and suppresses androgen-dependent gene
826 expression (Cook *et al*, 1993; Lambright *et al*, 2000; McIntyre *et al*, 2000). Using the
827 androgenised female stickleback assay in a semi-static system of exposure we provided the
828 first evidence of the antiandrogenic properties of LN in fish (Katsiadaki *et al,* 2006).

829 122. Three dose response studies (8, 10 and 11), all conducted at Lab 4 with over 280
830 individual spiggin measurements were analysed. Results were analysed and presented both
831 in terms of nominal and aquaria measured test compound concentrations.

832 123. Figure 17 displays the results from the LN studies on the basis of the nominal ratio
833 between LN and DHT. Since DHT was always used at a nominal concentration of 5µg/L
834 the actual concentration of LN can be quickly estimated by multiplying the ratio LN/DHT
835 featuring on the X-axis by a factor of 5.

836 124. According to figure 17, the NOEC for LN was estimated as 200µg/L and the LOEC
837 as 250µg/L. Figure 18 presents the same spiggin data as figure 17 but the responses are
838 expressed as a ratio of measured concentrations of LN over DHT whilst Figure 19 shows
839 the same data with a fitted linear regression line.

840 125. The measured concentrations of LN ranged between 21 and 77% of nominal.

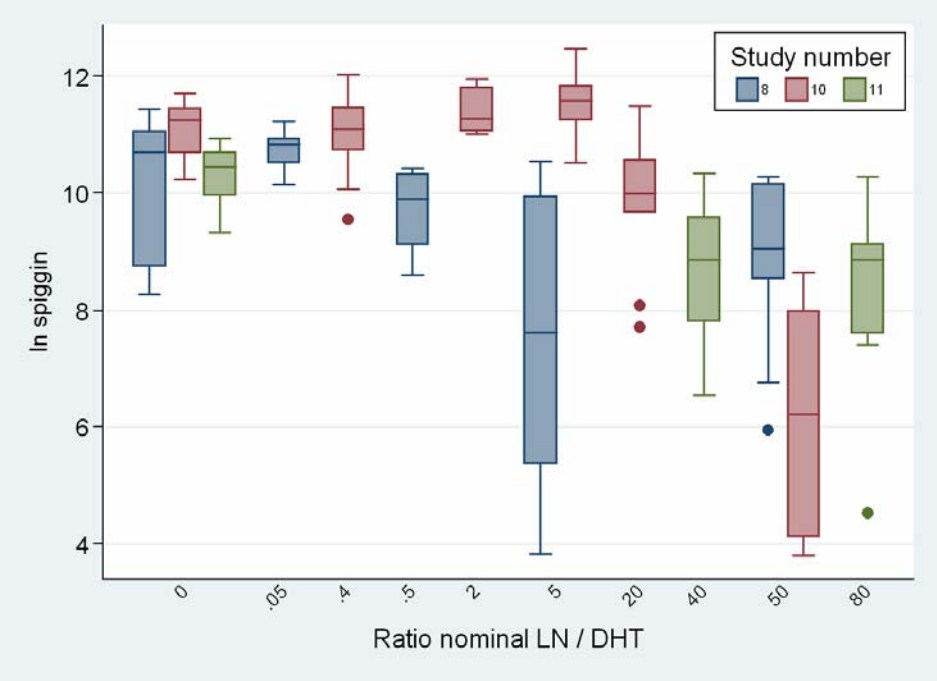


Figure 17: Spiggin responses to Linuron plotted against the ratio of nominal LN/DHT concentrations.

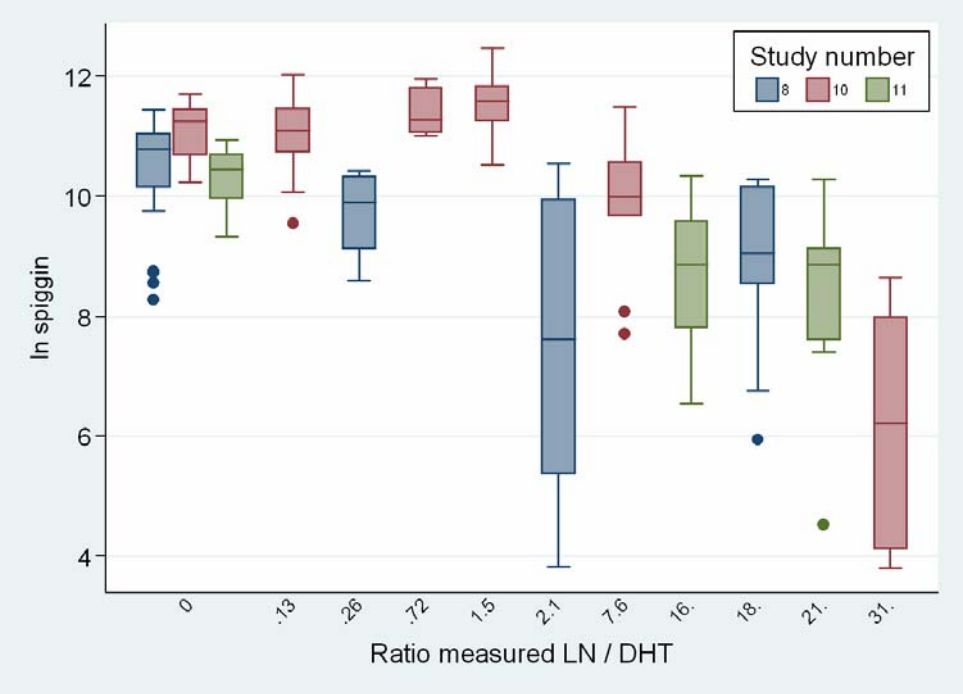


Figure 18: Spiggin responses to Linuron plotted against the ln ratio of measured LN/DHT concentrations.

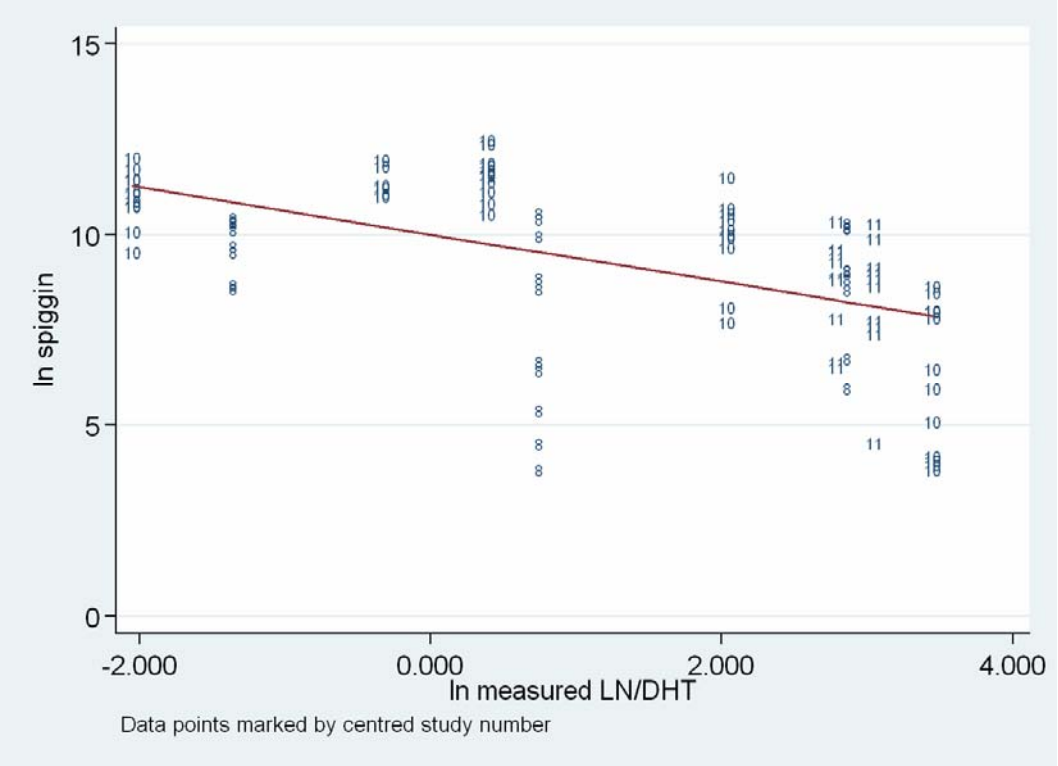


Figure 19: Spiggin responses to Linuron plotted against the ln ratio of measured LN/DHT concentrations.

Analysis of Vinclozolin data

126. Vinclozolin (VZ) is a dicarboximide fungicide used on oilseed rape and peas in the UK and on vines, fruit and vegetables globally. VZ is registered in the EU but banned in the USA due to adverse reproductive effects. A series of mechanistic studies were conducted to define the antiandrogenic properties of VZ by the US Environment Protection Agency. The results of these studies showed that VZ elicits the antiandrogenic effects by binding to androgen sensitive organs. Its status is being reviewed in many countries of the world. Its solubility in water is 3.4 mg/l at 20°C.

127. Vinclozolin is only moderately toxic to freshwater fish. The acute LC50 (96-hour) for the compound is 130ppm in guppies and 52.2ppm in trout.

128. Several mammalian studies have demonstrated both *in vivo* and *in vitro* androgen antagonism. Kelce *et al*, (1994) and Gray *et al* (1994) identified the first environmental antiandrogens that share the same mechanism of endocrine disruption with FL (interfering at the AR level) as M1 and M2, metabolites of the fungicide vinclozolin. Vinclozolin itself has poor affinity for the mammalian AR (Kelce *et al*, 1994). However, *in vivo* vinclozolin

129. In fish, a relevant study in fathead minnows (a species used widely for testing) failed to demonstrate antiandrogenicity (Makynen *et al*, 2002). A second study on guppies indicated some effects on behaviour and spermatogenesis (Bayley *et al*, 2003). We have also confirmed that VZ is a strong androgen antagonist using the AFSS in a semi static system (Katsiadaki *et al*, 2006).

130. Two full dose response studies (7 and 9), both conducted at Lab 4 with over 210 individual spiggin measurements were analysed. Results were analysed and presented both in terms of nominal and aquaria measured test compound concentrations.

131. Figure 20 displays the pooled results from the two VZ studies on the basis of the nominal ratio between VZ and DHT. Since DHT was always used at a nominal concentration of 5µg/L the actual concentration of VZ can be quickly estimated by multiplying the ratio VZ/DHT featuring on the X-axis by a factor of 5.

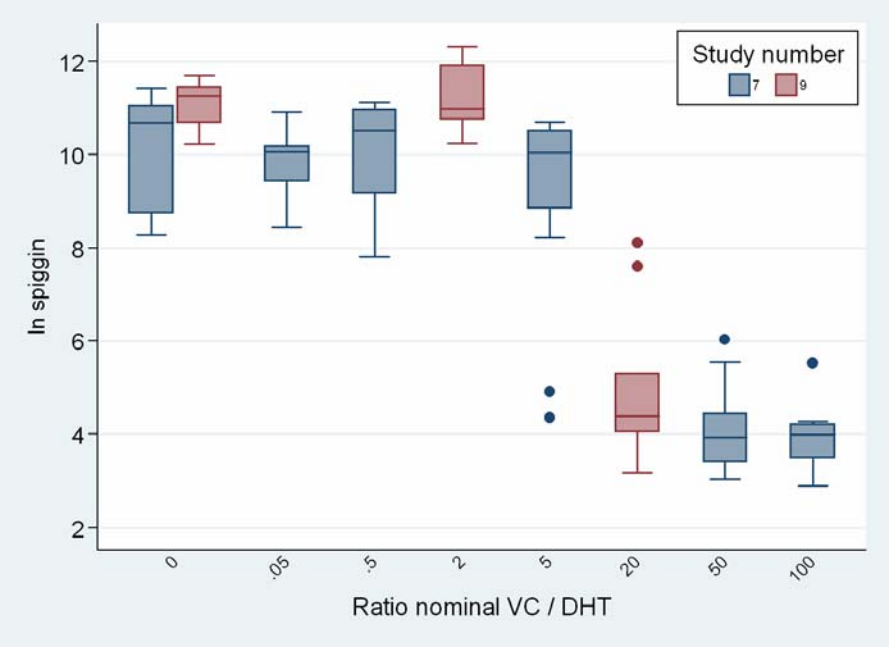


Figure 20: Spiggin responses to Vinclozolin plotted against the ratio of nominal LN/DHT concentrations.

132. As shown in figure 20, the NOEC for LN was estimated as 25µg/L and the LOEC as 100µg/L. Figure 21 presents the same spiggin data as figure 20 but the responses are expressed as a ratio of measured concentrations of VZ over DHT whilst figure 22 shows the same data with a fitted linear regression line. This ratio was logarithmically transformed to provide a better linear fit.

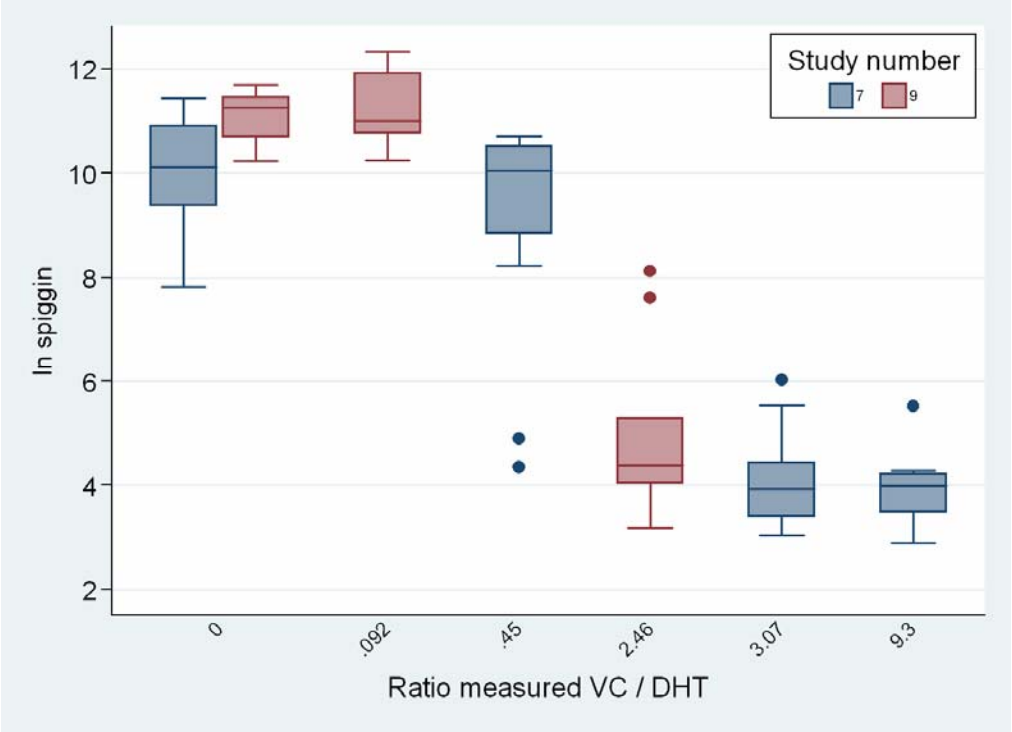


Figure 21: Spiggin responses to Vinclozolin plotted against the ln ratio of measured LN/DHT concentrations.

133. As figure 21 displays the measured concentrations of VZ were much lower to the nominal concentrations (6-10%). The low recovery of VZ is almost certainly related to its low stability and quick hydrolysis to the bioactive metabolites M1 and M2, which have affinity for the AR whilst the parent compound does not (Kelce *et al*, 1994, Wong *et al*, 1995).

134. Unfortunately due to financial and time constrains, the only analysis that was undertaken was of the parent compound, and hence it appears that VZ is extremely potent when the measured concentrations are reported. To this end, we suggest that for VZ the best way of expressing the data would be on the basis of the nominal concentrations tested.

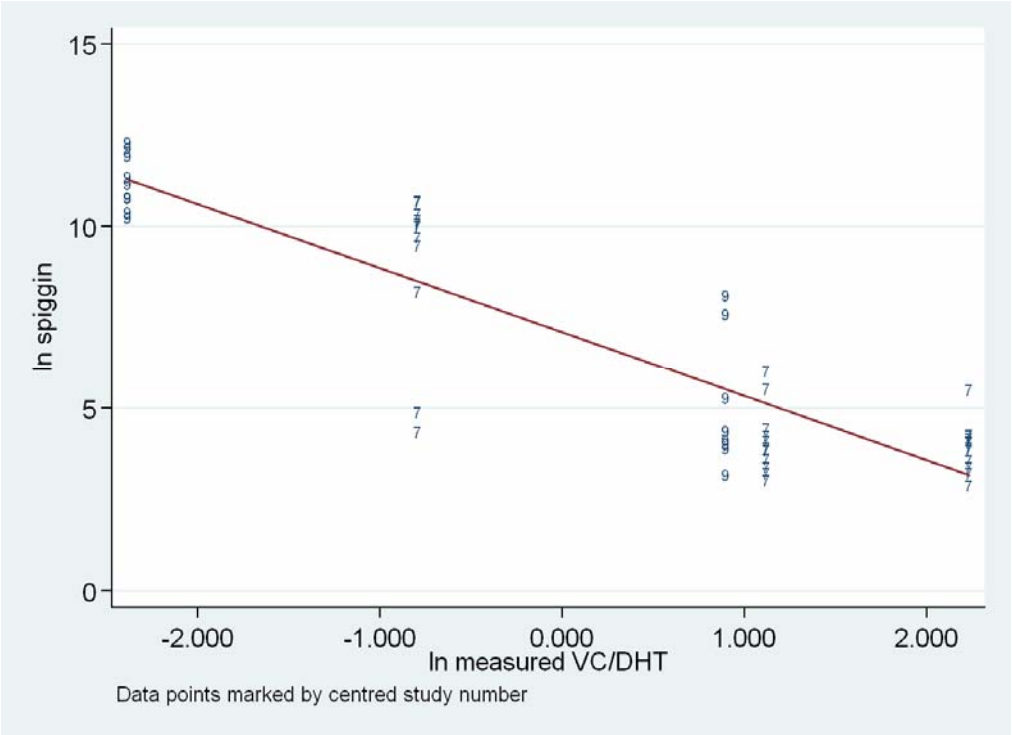


Figure 22: Spiggin responses to Vinclozolin plotted against the ln ratio of measured LN/DHT concentrations.

General discussion

135. Statistical analysis of a large dataset using the androgenised female stickleback test has revealed that the assay is robust and can detect antiandrogens unambiguously. Although the fish responses to antiandrogen administration via the water by means of kidney spiggin levels were variable, they never failed to detect the antiandrogenic activity of the tested compounds and followed a dose response manner in each set of data.

136. In many ways, the results of this analysis were surprisingly uniform even between experiments because several confounding factors had the potential to adversely affect the data interpretations. Some of these factors are listed below:

- The origin of the fish used for the exposures was not standardised, the vast majority were collected from the wild ranging from populations fully adapted to freshwater to fully adapted to marine environment.
- The size of the fish used was not uniform, their weight ranging from 0.37 to 4.27g.
- The methodology employing in water extraction for analytical verification of compounds changed over the period of experimentations.

- The analytical instruments (GC-MS for FN, LC-MS for VZ, FL, LN) employed for antiandrogen concentrations were also different over the 7-year period.
- The standard curve employed in the spiggin assay was altered after the first set of data were analysed.

137. Despite all the above counteracting factors a strong association between antiandrogen concentration in the water and spiggin levels in the kidneys was always present, indicating that the assay is robust.

138. The slopes of the regression analysis were always negative and relatively similar between the four compounds tested. Figure 23 presents all the data except the Lab 2 study, which was considered anomalous. The only slope that was significantly different was that of VZ, due to the very low measurements of the compound in the water.

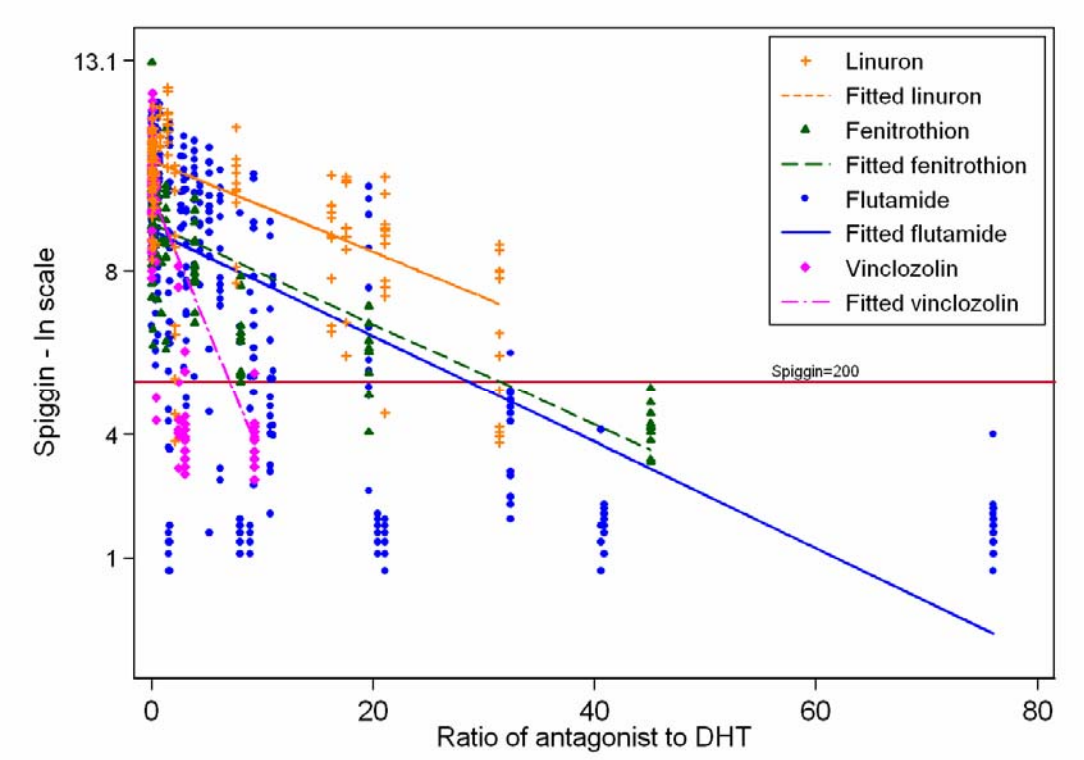


Figure 23: Pooled spiggin responses from all experiments and all compounds expect the Lab 2 study.

139. The intercepts for each compound were also very close adding confidence to the stability of the assay and the specific reagents employed over several years of use.

140. With regard to the power of the assay, figure 24 shows the same data as before re-arranged to make the following points:

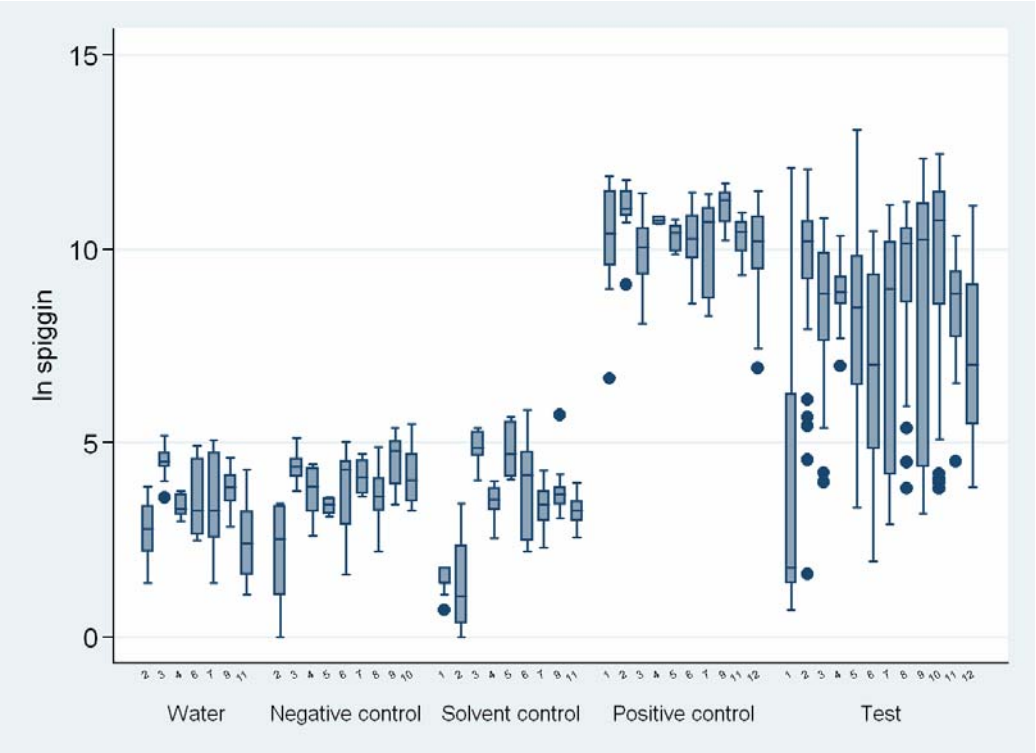


Figure 24: All spiggin data analysed separated by study number and grouped by treatment (various controls and test where FL, LN, FN and VZ were present at various concentrations.

141. The various control groups run at each lab over an extended period show that the baseline spiggin value (200, ln=5.3) is exceeded rarely and the variability is very consistent between studies. In fact, 90% of the control data registered less than 100 spiggin units/g body weight.

142. Most importantly there was no difference in terms of spiggin levels between water and solvent control, indicating that the use of solvent (methanol in this case) even at concentrations as high as 0.1% in the final aquaria water does not affect the spiggin responses.

143. The water and solvent groups are accepted as normal by the Shapiro-Wilks test, both separately and jointly. The negative control group are more uniform than normal but do not extend over a wider range. Taking the water, solvent and negative control groups combined, the mean ln spiggin is 3.44 and the SD 1.18. For a fitted normal distribution, approximately 5% of observations would exceed 200 and fewer than 1 in 1000 would exceed 500 (figure 25).

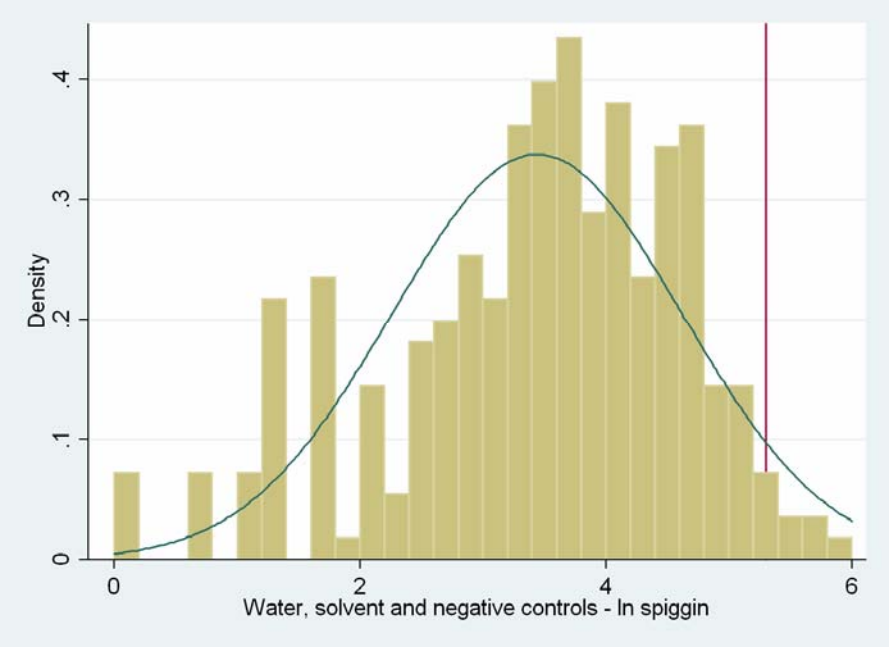


Figure 25: Fitted normal distribution of all baseline spiggin values.

144. The positive control group observations are left-skewed and truncated at approximately 160,000 (ln=12). The mean of 10.3 with SD .97 indicates that the probability of random overlap with an androgenised female is effectively zero, despite the variation in actual concentrations of DHT.

145. The measurement error of the response is thus demonstrated as comparable at very high and very low values of the response.

146. The presence of outliers in several positive control groups shows that the main barrier to precision is that some fish may fail to respond substantially to the DHT stimulus.

147. The robustness of observing an anti-androgen effect at all sites with all test chemicals indicates that the test is highly reliable. The apparently discordant studies (see details in paragraph 136) nevertheless detected an effect, with the few discrepancies potentially explained by unfamiliarity and alteration of the procedures involved in addition to the variable origin of fish.

148. Hence we might conclude that if a laboratory can demonstrate competence in maintaining precise concentrations of the test chemicals, the fish used originate from a reputable source and are correctly identified as females, the number of fish used in the AFSS can be reduced.

149. To guard against outlier responses and tank effects, our suggestion would be to use five female fish in two replicate tanks over a range of five test concentrations.

993 150. The mechanistic basis of the test is straightforward and is controlled by exogenous
994 androgens so the reproductive status of the fish does not interfere with the outcome of the
995 assay.
996 151. The only alternative *in vivo* test that is currently used for the detection of
997 antiandrogens is utilising rats and was described by Hershberger back in 1953. The basis
998 of the mammalian assay is that sexually mature male rats undergo regression of five
999 androgen-sensitive tissues (ventral prostate, seminal vesicles plus coagulating glands, glans
1000 penis, levator-ani and bulbocavernosus muscle complex, and Cowper's glands) when
1001 castrated. These tissues are restored to their original weight upon treatment with
1002 testosterone, and that growth can be blocked by the concomitant administration of an
1003 antiandrogen.
1004 152. In many ways the principle of the proposed test is similar to the rat Hershberger assay
1005 but does not require the use of male mammals nor surgical removal of reproductive organs,
1006 which can be expensive and time consuming.
1007 153. The need for a more rapid test for antiandrogens is reflected by the increasing number
1008 of alternative tests such as the weanling male rat assay (Ashby and Lefevre, 1997); the
1009 intact young male rat assay (O'Connor *et al*, 1999); the peripubertal intact male assay
1010 (Stocker *et al*, 2000); the use of androgen-stimulated immature intact male rats (Ashby *et*
1011 *al*, 2002); *in utero* exposure (Schultz *et al*, 2001) and the use of gonadotrophin release
1012 hormone-inhibited rats (Nellemann *et al*, 2003). Although the need for surgical castration
1013 is not required in these alternative tests nevertheless, the majority of these assays rely on
1014 the same principle, changes in the reproductive or accessory sex gland weights upon
1015 treatment. This end-point has received criticism because chemical treatment may affect
1016 growth rate, thus the relationship between body weight and / or accessory gland weight is
1017 problematic (Marty *et al*, 2003).
1018 154. In addition, the spiggin assay has a much higher resolution (up to 100,000 spiggin
1019 units difference between treatment levels), in comparison to the Hershberger test where
1020 only small differences in organ weights between treated and non-treated rats are found.
1021 155. The *in vivo* responses reported here are in full agreement with *in vitro* data using the
1022 stickleback kidney primary culture and part of the datasets analysed in this report have
1023 been peer reviewed (Jolly *et al*, 2009). This was expected as all the information on the
1024 mode of action of tested compounds indicates that they act via receptor (AR) binding.
1025 156. The *in vivo* test described here however can detect more modes or sites of action. For
1026 example, we have reported before the antiandrogenic potential of oestrogens by means of

157. During the EDEN project we modelled this response using the androgenised female stickleback assay and confirmed that high concentrations of oestrogens (E_1, E_2, EE_2, NP) also inhibit spiggin induction by DHT (i.e. they have a separate antiandrogenic effect, in addition to the well characterised oestrogenic responses).

158. This effect of oestrogens was also observed *in vitro* (Jolly *et al*, 2009) but was not as pronounced as it was *in vivo* and it required much higher concentrations to reach the point of total inhibition of spiggin by DHT.

159. This implies that the antiandrogenic effect of oestrogens is not mediated via their well known affinity for the ARs (Sohoni and Sumpter, 1998) present in the stickleback kidney but goes beyond this, acting either via membrane steroid receptors (see or Loomis and Thomas, 2000) or via other mechanism(s) involving feedback control of sex steroid levels to gonadotrophins (Schultz *et al*, 2001).

160. One of the great advantages of the stickleback as a model organism is that the simultaneous assessment of an androgen and oestrogen end point (we have developed an homologous ELISA for stickleback VTG) can provide vital clues on the mechanisms responsible for endocrine adverse effects.

161. We applied the VTG ELISA to detect oestrogenicity of the test compounds and found that none of the environmental antiandrogens tested induced VTG in male fish or increased the VTG content in female fish (results not shown). As the two protein markers are produced and stored in different organs (kidney for spiggin, plasma or liver for VTG), a single fish can be analysed for androgenic, antiandrogenic and oestrogenic activity. This reduces the number of test organisms needed, which is of great importance from an ethical and economic perspective.

162. There are several other reasons why the stickleback is an ideal European bio-indicator species as listed elsewhere (Katsiadaki *et al*, 2002b; Katsiadaki *et al*, 2007). Importantly, the stickleback is a true sentinel and an ideal model species, as its entire genome is sequenced and molecular analysis tools are available (i.e. cDNA microarrays).

163. In view of the numerous clinical implications for human health, the high antiandrogenic activity detected in the aquatic environment and the fact that the only reliable antiandrogen bioassay requires castrated rats, an *in vivo* test using intact fish in order to screen and identify antiandrogenic chemicals is highly desirable.

164. We propose that the stickleback assays can fill this gap. The AFSS is better suited as an *in vivo* screen following *in silico* and *in vitro* data that suggest antiandrogenic activity.

165. The stickleback-breeding test can be also employed as a higher tier test to aid risk assessment as it can provide vital information on the effect of antiandrogen at fish population level.

References

Advisory Committee on Pesticides, 2006. Evaluation on fenitrothion. Department for Environment, Food and Rural Affairs, Pesticides Safety Directorate, Issue No 225, York, 324 pp

Allen, Y.T, Katsiadaki, I., Pottinger, T.G., Jolly, C., Matthiessen, P., Mayer, I., Smith, A., Scott, A.P., Eccles, P., Sanders, M.B., Pulman, K.G.T. and Feist, S., 2008. Intercalibration exercise using a stickleback endocrine disrupter screening assay. Environmental Toxicology and Chemistry 27(2):404-412.

Andersson, C., Katsiadaki, I. Lundstedt-Enkel, K., and Örberg J., 2007. Effects of 17α-ethynylestradiol on EROD activity, spiggin and vitellogenin in three-spined stickleback (*Gasterosteus aculeatus*). Aquatic Toxicology 83:33-42.

Ashby, J., and Lefevre, P.A., 1997. The Weanling male rat as an assay for endocrine disruption: Preliminary observations. Regulatory Toxicology and Pharmacology 26(3):330-337.

Ashby, J., Owens, W., and Lefevre, P.A., 2002. Concept evaluation-Stimulated immature intact male rats as an assay for anti-androgens. Regulatory Toxicology and Pharmacology 35:280-285.

Baggerman, B., 1972. Photoperiodic responses in the stickleback and their control by a daily rhythm of photosensitivity. General and Comparative Endocrinology 2:466-476.

Baggerman, B., 1985. The role of biological rhythms in the photoperiodic regulation of seasonal breeding in the stickleback *Gasterosteus aculeatus* L. Netherlands Journal of Zoology 35:14-31.

Bayley, M., Junge, M. and Baatrup, E., 2002. Exposure of juvenile guppies to three antiandrogens causes demasculinization and a reduced sperm count in adult males. Aquatic Toxicology 56(4):227-239.

Björkblom, C., Olsson, P-E., Katsiadaki, I., Wiklund, T., 2007. Estrogen- and androgen-sensitive bioassays based on primary cell and tissue slice cultures from the three-spined stickleback (*Gasterosteus aculeatus*). Comparative Biochemistry and Physiology C 146:431-442.

Borg, B., Antonopoulou, E., Andersson, E., Carlberg, T., and Mayer, I., 1993. Effectiveness of several androgens in stimulating kidney hypertrophy, a secondary

sexual character, in castrated male three-spined sticklebacks, *Gasterosteus aculeatus*. Canadian Journal of Zoology 71:2327-2328.

Bornestaf, C. and Borg, B., 2000. Endogenous breeding cycles in male threespine sticklebacks, *Gasterosteus aculeatus*. Behaviour. 137, 921-932.

Bush, I.E., 1961. The Chromatography of steroids. (Pergamon Press).

Cook, J.C., Mullin, L.S., Frame, S.R., and Biegel, L.B., 1993. Investigation of a mechanism for Leydig cell tumorigenesis by linuron in rats. Toxicology and Applied Pharmacology 119(2):195-204.

Ellis, T.; James, J. D.; Stewart, C.; Scott, A. P., 1994. A non-invasive stress assay based upon measurement of free cortisol released into the water by rainbow trout. Journal of Fish Biology 65, 1233-1252.

Gray, L.E. Jr., Ostby, J.S., and Kelce, W.R., 1994. Developmental effects of an environmental antiandrogen: the fungicide vinclozolin alters sex differentiation in the male rat. Toxicology and Applied Pharmacology 129:46-52.

Hahlbeck, E., Katsiadaki, I., Mayer, I., Adolfsson-Erici, M., James, J. and Bengtsson, B.-E., 2004. The juvenile three-spined stickleback (*Gasterosteus aculeatus* L.) as a model organism for endocrine disruption II- kidney hypertrophy, vitellogenin and spiggin induction. Aquatic Toxicology 70, 311-326.

Hershberger L.G., Shipley E.G., and Meyer R.K., 1953. Myotrophic activity of 19-nortestosterone and other steroids determined by modified levator ani muscle method. Proceedings of the Society of Experimental Biology and Medicine 83:175–180.

Hogan, N.S., Wartman, C.A., Finley, M.A., van der Lee, J.G. and van den Heuvel, M.R., 2008. Simultaneous determination of androgenic and estrogenic endpoints in the threespine stickleback (*Gasterosteus aculeatus*) using quantitative RT-PCR. Aquatic Toxicology 90:269–276.

Hutchinson, T. H., Shillabeer, N., Winter, M. J. and Pickford, D. B., 2006. Acute and chronic effects of carrier solvents in aquatic organisms: A critical review. Aquatic Toxicology 76:69-92.

Jobling, S., Burn, R.W., Thorpe, K., Williams, R. and Tyler C., 2009. Statistical modelling suggests that anti-androgens in wastewater treatment works effluents are contributing causes of widespread sexual disruption in fish living in English Rivers. Environmental Health Perspectives *in press*, DOI:101289/EHP0800197.

Jolly, C., Katsiadaki, I., Le Belle, N., Mayer, I., and Dufour, S., 2006. Development of a stickleback kidney cell culture assay for the screening of androgenic and anti-androgenic endocrine disrupters. Aquatic Toxicology 79(2):158-166.

Jolly, C., Katsiadaki, I., Morris, S., Le Belle, N., Dufour, S., Mayer, I., Pottinger, T.G., and Scott, A.P., 2009. Detection of the anti-androgenic effect of endocrine disrupting environmental contaminants using *in vivo* and *in vitro* assays in the three-spined stickleback. Aquatic Toxicology, 92:228-239.

Johnson, I., Hetheridge, M. and Tyler, C. R., 2007. Assessment of (anti-) oestrogenic and (anti-) androgenic activities of final effluents from sewage treatment works. Science report-SC020118/SR, Environment Agency, Bristol, UK, pp. 61.

Katsiadaki, I., Scott, A.P., and Matthiessen, P., 1999. The use of the three-spined stickleback as a potential biomarker for androgenic xenobiotics. In the Proceedings of the 6th International Symposium on the Reproductive Physiology of Fish, Bergen, Norway, July 4-9, 1999, Norberg, B., Kjesbu, O.S., Taranger, G.L., Andersson, E., and Stefansson, S.O., (eds.), pp: 359-361.

Katsiadaki, I., Scott, A. P., Hurst, M. R., Matthiessen, P. and Mayer, I., 2002a. Detection of environmental androgens: a novel method based on enzyme-linked immunosorbent assay of spiggin, the stickleback (*Gasterosteus aculeatus*) glue protein. Environmental Toxicology and Chemistry 21:1946-1954.

Katsiadaki, I., Scott, A. P. and Mayer, I., 2002b. The potential of the three-spined stickleback (*Gasterosteus aculeatus* L.) as a combined biomarker for oestrogens and androgens in European waters. Marine Environmental Research. 54:725-728.

Katsiadaki, I., Morris, S., Squires, C., Hurst, M. R., James, J. D. and Scott, A. P., 2006. Use of the three-spined stickleback (*Gasterosteus aculeatus*) as a sensitive *in vivo* test for the detection of environmental antiandrogens. Environmental Health Perspectives 114 (suppl 1):115-121.

Katsiadaki, I., Sanders, M.B., Sebire, M., Nagae, M., Soyano K., and Scott, A.P., 2007. Three-spined stickleback: an emerging model in environmental endocrine disruption. Environmental Sciences 14(5):263-283.

Kelce, W.R., Monnosson, E., Gamcsik, M.P., Laws, S.C., Gray, L.E. Jr., 1994. Environmental hormone disruptors: evidence that vinclozolin developmental toxicity is mediated by antiandrogenic metabolites. Toxicology and Applied Pharmacology 126: 276-285.

Lambright, C., Ostby, J., Bobseine, K., Wilson, V., Hotchkiss, A.K., Mann, P.C., and Gray, L.E. Jr., 2000. Cellular and molecular mechanisms of action of linuron: An antiandrogenic herbicide that produces reproductive malformations in male rats. Toxicological Sciences 56:389–399.

Makynen, E.A., Kahl, M.D., Jensen, K.M., Tietge, J.E., Wells, K.L., Van Der Kraak, G. and Ankley, G.T., 2000. Effects of the mammalian antiandrogen vinclozolin on development and reproduction of the fathead minnow (*Pimephales promelas*). Aquatic Toxicology 48:461-475.

Marty, M.S., Johnson, K.A., and Carney, E.W., 2003. Effect of feed restriction on Hershberger and pubertal male assay points. Birth Defects Research B 68:363-374.

McIntyre, B.S., Barlow, N.J., Wallace, D.G., Maness, S.C., Gaido, K.W., and Foster, P.M.D., 2000. Effects of *in utero* exposure to linuron on androgen-dependent reproductive development in the male Crl:CD(SD)BR rat. Toxicology and Applied Pharmacology 167(2): 7–99.

Nellemann, C., Lefevre, P.A., and Ashby, J., 2003. Comparison of prostate gene expression and tissue weight changes as monitors of antiandrogen activity in GNRH-Inhibited rats. Birth Defects Research B 68:344-354.

O'Connor, J.C., Frame, S.R., Davis, L.G., and Cook, J.C., 1999. Detection of the environmental antiandrogen p,p'-DDE in CD and Long-Evans rats using a Tier 1 screening battery and Hershberger assay. Toxicological Sciences 51:44-53.

Oguro, C., 1957. Notes on the change in the kidneys of the three-spined stickleback, *Gasterosteus aculeatus* aculeatus L. caused by oestrogen administration. Journal of the Faculty of Sciences, Hokkaido University Series VI Zoology 13:404-407.

Organisation for Economic Cooperation and Development (OECD), 2006. Report of the validation of the 21-day fish screening assay for the detection of endocrine substances (phase 1B). Series on Testing and Assessment. Number 61. 140 pages.

Sanchez, W., Katsiadaki I., Piccini, B., Ditche J-M. and Porcher J-M., 2008. Biomarker responses in wild three-spined stickleback (*Gasterosteus aculeatus* L.) as a useful tool for freshwater biomonitoring: a multiparametric approach. Environment International, 34:490-498.

Sanchez, W., Goin, C., Brion, F., Olsson, P.E., Goksoyr, A., Porcher, J.M., 2008. A new ELISA for the three-spined stickleback (*Gasterosteus aculeatus* L.) spiggin, using antibodies against synthetic peptide. Comparative Biochemistry Physiology C 147: 129-137.

Schultz, V.D., Phillips. S., Sar, M., Foster, P.M.D, and Gaido, K.W., 2001. Altered gene profiles in fetal rat testes after *in utero* exposure to di(n-butyl)phthalate. Toxicological Sciences 64:233-242.

Scott, A. P.; MacKenzie, D. S.; Stacey, N. E., 1984. Endocrine changes during natural spawning in the white sucker, *Catostomus commersoni*. II. Steroid hormones. General and Comparative Endocrinology 56(3):349-59.

Sebire, M., Allen, Y., Bersuder P. and Katsiadaki, I., 2008. The model antiandrogen flutamide suppresses the expression of typical male stickleback reproductive behaviour. Aquatic Toxicology 90:37-47.

Sebire, M., Scott, A.P., Tyler, C.R., Cresswell, J., Hodgson, D.J., Morris, S., Sanders, M.B., Stebbing, P.D. and Katsiadaki, I., 2009. The organophosphorous pesticide, fenitrothion, acts as an anti-androgen and alters reproductive behavior of the male three-spined stickleback, *Gasterosteus aculeatus*. Ecotoxicology 18:122-133.

Sebire, M., Katsiadaki, I., Maack, G. and Tyler, C.R., (in preparation). Effect of a short-term exposure to a treated sewage effluent on the reproductive physiology and behaviour in the three-spined stickleback (*Gasterosteus aculeatus*).

Shiraishi, S., Lee, P.W.N., Leung, A, Goh, V.H.H., Swerdloff, R.S. and Wang, C., 2008. Validation of a testosterone and dihydrotestosterone liquid chromatography tandem mass spectrometry assay: Interference and comparison with established methods. Steroids 73(13):1345-1352.

Sohoni, P. and Sumpter, J.P. 1998. Several environmental oestrogens are also anti-androgens. Journal of Endocrinology 158:327-339.

Sohoni, P., Lefevre, P.A., Ashby, J., and Sumpter, J.P., 2001. Possible androgenic/anti-androgenic activity of the insecticide fenitrothion. Journal of Applied Toxicology 21:173–178.

Stocker, T.E., Parks, L.G., Gray. E., Cooper, R.L., 2000. Endocrine-disrupting chemicals: Prepubertal exposures and effects on sexual maturation and thyroid function in the male rat. A focus on the EDSTAC (Endocrine Disrupter Screening and Testing Advisory Committee) recommendations. Critical Reviews in Toxicology 30:197-252.

Sunami, O., Kunimatsu, T., Yamada, T., Yabushita, S., Sukata, T., Miyata, K., Kamita, Y., Okuno, Y., Seki, T., Nakatsuka, I., and Matsuo, M., 2000. Evaluation of a 5-day Hershberger assay using young mature male rats: methyltestosterone and p,p0-DDE, but not fenitrothion, exhibited androgenic or antiandrogenic activity *in vivo*. Journal of Toxicological Sciences 25:403–415.

Tamura, H., Maness, S.C., Reischmann, K., Dorman, D.C., Gray, L.E. and Gaido, K.W., 2001. Androgen receptor antagonism by the organophosphate insecticide fenitrothion. Toxicological Sciences 60:56-62.

Thomas, K.V., Hurst, M.R., Matthiessen, P., McHugh, M., Smith, A., and Waldock, M.J., 2002. An assessment of *in vitro* androgenic activity and the identification of environmental androgens in United Kingdom estuaries. Environmental Toxicology and Chemistry 20:1456-1461.

Tollefsen, K.E., Harman, C., Smith, A. and Thomas, K.V., 2007. Estrogen receptor (ER) agonists and androgen receptor (AR) antagonists in effluents from Norwegian North Sea oil production platforms. Marine Pollution Bulletin 54:277-283.

Urbatzka, R., van Cauwenberge, A., Maggioni, S., Viganò, L., Mandich, A., Benfenati, E., Lutz, I. and Kloas, W., 2007. Androgenic and antiandrogenic activities in water and sediment samples from the river Lambro, Italy, detected by yeast androgen screen and chemical analyses. Chemosphere 67:1080-1087.

Wong, C.I., Kelce, W.R., Sar, M., and Wilson, E.M., 1995. Androgen Receptor Antagonist versus Agonist Activities of the Fungicide Vinclozolin Relative to Hydroxyflutamide. Journal of Biological Chemistry 270(34):19998–20003.

Statistical tables

Statistical analysis of data presented in figure 3: Means and SDs for each group of females.

Treatment (plot order)	Summary of ln(spig) Mean	Std. Dev.	Freq.
DHT-S	10.2	1.05	45
DHT-S-FL	7.2	1.8	51
DHT-W	4.5	0.36	46
DHT-W-FL	4.4	0.53	41
DHT-W-PP	7.9	1.3	38
Total	6.8	2.5	221

Statistical analysis of data presented in figures 5 and 6. Water, solvent and negative control groups.

```
Compare water and solvent controls

. anova lsp labno control if ~inlist( control, 0, 1, 3) & labno~=1

                           Number of obs =     173     R-squared     =  0.1482
                           Root MSE      =  1.0646     Adj R-squared =  0.1382

                  Source |  Partial SS    df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  33.5344843     2  16.7672421      14.79     0.0000
                         |
                   labno |  33.4050698     1  33.4050698      29.47     0.0000
                 control |  .470185124     1  .470185124       0.41     0.5204
                         |
                Residual |  192.671872   170  1.13336395
              -----------+----------------------------------------------------
                   Total |  226.206356   172  1.31515323

. table labno control if ~inlist( control, 0, 1, 3) & labno~=1, c(mean lsp sd lsp
n lsp)

----------------------------------------
          |             control
    labno |           Water  Solvent control
----------+-----------------------------------
    Lab 3 |        4.101406         4.344005
          |        .6548476          .842097
          |              20               22
          |
    Lab 4 |        3.307975         3.091557
          |        1.020185          1.26965
          |              71               60
----------------------------------------
```

```
. anova lsp study if ~inlist( control, 0, 1)

                           Number of obs =     276     R-squared     =  0.5169
                           Root MSE      = .836074     Adj R-squared =  0.4987

                  Source |  Partial SS    df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  198.232195    10  19.8232195      28.36     0.0000
                         |
                   study |  198.232195    10  19.8232195      28.36     0.0000
                         |
                Residual |  185.240056   265  .699019078
              -----------+----------------------------------------------------
                   Total |  383.472251   275  1.39444455
```

All baseline controls

```
. anova lsp study  if inlist( control, 2, -1, 3) & ~inlist(study,1,2,3,11)

                           Number of obs =     172     R-squared     =  0.0530
                           Root MSE      = .842284     Adj R-squared =  0.0186

                  Source |  Partial SS    df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  6.55276211     6  1.09212702       1.54     0.1683
                         |
                   study |  6.55276211     6  1.09212702       1.54     0.1683
                         |
                Residual |  117.057977   165  .709442286
              -----------+----------------------------------------------------
                   Total |  123.610739   171   .72286982

. table study if inlist( control, -1, 2, 3), c(mean lsp sd lsp n lsp)

------------------------------------------------
group(ord |
ate where |
experimen |
t)        | mean(lspig)    sd(lspig)     N(lspig)
----------+-------------------------------------
        1 |    1.497947     .2850094           19
        2 |    2.092776     1.228135           31
        3 |    4.568988      .450674           29
        4 |    3.514693     .4949893           22
        5 |     4.08479     .8846126           12
        6 |    3.678614     1.072198           40
        7 |    3.679244     .9285469           39
        8 |    3.672297     .6891353           11
        9 |    3.979214     .6587296           38
       10 |    4.160357     .7380273           10
       11 |    2.914972     .8109522           25
------------------------------------------------
```

Statistical analysis of data presented in figure 7: Positive control groups.

```
. anova lsp study if inlist( control, 1)

                           Number of obs =     159     R-squared     =  0.1123
                           Root MSE      = .943812     Adj R-squared =  0.0587

                  Source |  Partial SS    df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |  16.7880244     9  1.86533605       2.09     0.0334
                         |
                   study |  16.7880244     9  1.86533605       2.09     0.0334
                         |
                Residual |   132.72629   149   .89078047
              -----------+----------------------------------------------------
                   Total |  149.514314   158  .946293129

. anova lsp study if inlist( control, 1) & ~inlist(study,9)

                           Number of obs =     149     R-squared     =  0.0731
                           Root MSE      = .965307     Adj R-squared =  0.0202

                  Source |  Partial SS    df       MS           F     Prob > F
              -----------+----------------------------------------------------
                   Model |   10.291207     8  1.28640088       1.38     0.2100
                         |
                   study |   10.291207     8  1.28640088       1.38     0.2100
                         |
                Residual |  130.454475   140  .931817679
              -----------+----------------------------------------------------
                   Total |  140.745682   148  .950984338
```

Statistical analysis of data presented in figure 9: Nominal flutamide data.

One-way ANOVA (not assuming homogeneity of variance)

```
> anova=with(alldata,{as.factor(FLtoDHTn)
+ oneway.test(log(Spiggin+1)~FLtoDHTn)})
```

data: log(Spiggin + 1) and FLtoDHTn
F = 661.3066, num df = 11.000, denom df = 109.971, p-value < 2.2e-16

```
> pairwise2=with(alldata,{as.factor(FLtoDHTn)
+ pairwise.t.test(log(Spiggin+1),FLtoDHTn, pool.sd=F)})
> pairwise2
```

Pairwise comparisons using t tests with non-pooled SD

data: log(Spiggin + 1) and FLtoDHTn

	0	0.2	1	2	5	10	15	20	25	30	50
0.2	0.09577	-	-	-	-	-	-	-	-	-	
1	1.00000	0.21676	-	-	-	-	-	-	-	-	
2	2.4e-05	0.01380	3.7e-05	-	-	-	-	-	-	-	
5	0.01380	1.00000	0.06112	0.00983	-	-	-	-	-	-	
10	3.2e-12	2.2e-07	3.4e-12	0.91057	3.4e-08	-	-	-	-	-	
15	0.01761	0.46487	0.02062	1.00000	0.44230	0.08283	-	-	-	-	
20	3.4e-05	0.27079	9.2e-05	1.00000	0.21676	0.00252	1.00000	-	-	-	
25	< 2e-16	2.9e-13	< 2e-16	3.4e-05	< 2e-16	0.00036	0.00022	7.9e-12	-	-	
30	0.00889	0.23679	0.00983	1.00000	0.21775	0.18647	1.00000	1.00000	0.00036	-	
50	< 2e-16	2.6e-06	< 2e-16	1.00000	1.3e-08	0.43672	0.91057	0.09609	< 2e-16	1.00000	-
100	< 2e-16	1.1e-13	< 2e-16	4.7e-05	< 2e-16	0.00062	0.00022	7.8e-12	1.00000	0.00036	< 2e-16

Statistical analysis of data presented in figure 10: Nominal flutamide data separated by site.

Lab 2 ANOVA:

```
> anova=with(Lab 2,{as.factor(FLtoDHTn)
+ oneway.test(log(Spiggin+1)~FLtoDHTn)})
```

One-way analysis of means (not assuming equal variances)

data: log(Spiggin + 1) and FLtoDHTn

F = 86.9491, num df = 1.000, denom df = 81.628, p-value = 1.694e-14

Lab 1 ANOVA:

```
> anova=with(Lab 1,{as.factor(FLtoDHTn)
+ oneway.test(log(Spiggin+1)~FLtoDHTn)})
```

One-way analysis of means (not assuming equal variances)

data: log(Spiggin + 1) and FLtoDHTn

F = 180.7192, num df = 6.000, denom df = 69.798, p-value < 2.2e-16

```
> pairwise=with(Lab 1,{as.factor(FLtoDHTn)
+ pairwise.t.test(log(Spiggin+1),FLtoDHTn, pool.sd=F)})
```

Pairwise comparisons using t tests with non-pooled SD

data: log(Spiggin + 1) and FLtoDHTn

```
   0      0.2    2      10     25     50
0.2 0.0355 -      -      -      -      -
2   5.1e-06 0.0037 -      -      -      -
10  < 2e-16 9.2e-13 0.0020 -      -      -
25  6.9e-16 7.9e-14 9.3e-06 0.1627 -      -
50  < 2e-16 3.7e-14 1.7e-05 0.2843 0.4196 -
100 < 2e-16 3.2e-14 1.2e-05 0.2232 0.9489 0.9489
```

Lab 3 ANOVA:

```
> anova=with(Lab 3,{as.factor(FLtoDHTn)
+ oneway.test(log(Spiggin+1)~FLtoDHTn)})
```

One-way analysis of means (not assuming equal variances)

```
data:  log(Spiggin + 1) and FLtoDHTn
F = 29.2576, num df = 4.000, denom df = 13.974, p-value = 1.172e-06
> pairwise5=with(Lab 3,{as.factor(FLtoDHTn)
+ pairwise.t.test(log(Spiggin+1),FLtoDHTn, pool.sd=F)})
> pairwise5
```

Pairwise comparisons using t tests with non-pooled SD

```
data:  log(Spiggin + 1) and FLtoDHTn

  0      1       5       20
1  0.16268 -       -       -
5  0.65630 0.09006  -       -
20 0.28877 0.00247  0.65630  -
50 0.00017 0.00017  0.00080  0.00083
```

Lab 4 ANOVA:

```
> anova=with(Lab 4,{as.factor(FLtoDHTn)
+ oneway.test(log(Spiggin+1)~FLtoDHTn)})
```

One-way analysis of means (not assuming equal variances)

```
data:  log(Spiggin + 1) and FLtoDHTn
F = 19.1353, num df = 7.000, denom df = 52.931, p-value = 1.817e-12
> pairwise=with(Lab 4,{as.factor(FLtoDHTn)
+ pairwise.t.test(log(Spiggin+1),FLtoDHTn, pool.sd=F)})
```

Pairwise comparisons using t tests with non-pooled SD

```
data:  log(Spiggin + 1) and FLtoDHTn
```

	0	1	5	10	15	20	30
1	1.00000	-	-	-	-	-	-
5	0.00781	0.09145	-	-	-	-	-
10	0.03191	0.09133	1.00000	-	-	-	-
15	0.00781	0.01788	0.64956	1.00000	-	-	-
20	2.4e-05	0.00014	0.09145	1.00000	1.00000	-	-
30	0.00352	0.00781	0.28477	1.00000	1.00000	1.00000	-
50	4.0e-06	3.0e-05	0.04529	1.00000	1.00000	1.00000	1.00000

Statistical analysis of data presented in figure 12: Measured flutamide/DHT data separated by site. The important features (intercept, slope and p values are highlighted in bold.

Lab 2 data

```
lm(formula = log(Spiggin + 1) ~ FLtoDHTm, data = Lab 2)
Residuals:
    Min      1Q  Median      3Q     Max
-3.0944 -1.0169  0.1929  0.9515  4.3288
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.946098   0.224247  44.353  < 2e-16 ***
FLtoDHTm    -0.033993   0.003813  -8.916 3.43e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.544 on 95 degrees of freedom
Multiple R-squared: 0.4556, Adjusted R-squared: 0.4498
F-statistic: 79.49 on 1 and 95 DF, p-value: 3.426e-14
> anova(lm1)
Analysis of Variance Table
Response: log(Spiggin + 1)
          Df  Sum Sq Mean Sq F value    Pr(>F)
FLtoDHTm   1 189.616 189.616  79.493 3.426e-14 ***
Residuals 95 226.604   2.385
```

Lab 1 data

```
lm(formula = log(Spiggin + 1) ~ FLtoDHTm, data = Lab 1)
Residuals:
    Min      1Q  Median      3Q     Max
-5.0376 -2.6964 -0.1649  2.0364  5.8978
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.257164   0.296896  21.075   <2e-16 ***
FLtoDHTm    -0.079081   0.008678  -9.112   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 3.072 on 179 degrees of freedom
Multiple R-squared: 0.3169,     Adjusted R-squared: 0.3131
F-statistic: 83.03 on 1 and 179 DF,  p-value: < 2.2e-16
> anova(lm2)
Analysis of Variance Table
Response: log(Spiggin + 1)
           Df  Sum Sq Mean Sq F value    Pr(>F)
FLtoDHTm    1  783.64  783.64  83.035 < 2.2e-16 ***
Residuals 179 1689.32    9.44
```

Lab 3 data

```
lm(formula = log(Spiggin + 1) ~ FLtoDHTm, data = Lab 3)
Residuals:
     Min       1Q   Median       3Q      Max
-2.39092 -0.89518  0.04979  0.78898  2.00089
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.48378    0.29667   35.34  < 2e-16 ***
FLtoDHTm    -0.41867    0.04863   -8.61 5.97e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.203 on 33 degrees of freedom
Multiple R-squared: 0.692,      Adjusted R-squared: 0.6826
F-statistic: 74.13 on 1 and 33 DF,  p-value: 5.972e-10
> anova(lm3)
Analysis of Variance Table
Response: log(Spiggin + 1)
          Df  Sum Sq Mean Sq F value    Pr(>F)
FLtoDHTm   1 107.193 107.193  74.127 5.972e-10 ***
Residuals 33  47.721   1.446
```

Lab 4 data

```
lm(formula = log(Spiggin + 1) ~ FLtoDHTm, data = Lab 4)
Residuals:
    Min      1Q  Median      3Q     Max
-6.9174 -0.8962  0.4384  1.2722  4.1447
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.72114    0.18366   52.93   <2e-16 ***
FLtoDHTm    -0.19461    0.01571  -12.38   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.94 on 175 degrees of freedom
Multiple R-squared: 0.4671,     Adjusted R-squared: 0.464
F-statistic: 153.4 on 1 and 175 DF,  p-value: < 2.2e-16
> anova(lm4)
Analysis of Variance Table
Response: log(Spiggin + 1)
           Df Sum Sq Mean Sq F value    Pr(>F)
FLtoDHTm    1 577.40  577.40  153.37 < 2.2e-16 ***
Residuals 175 658.84    3.76
```

Statistical analysis of data presented in figure 14.

```
One-way analysis of means (not assuming equal variances)
data:  log(Spiggin + 1) and FNtoDHTn
F = 91.465, num df = 12.000, denom df = 43.456, p-value < 2.2e-16
        Pairwise comparisons using t tests with non-pooled SD

data:  log(Spiggin + 1) and FNtoDHTn
      0     0.02    0.05    0.125   0.312     0.5     0.78    1.95    4.88    5       12      24
0.02  1.0       -       -       -       -       -       -       -       -       -       -       -
0.05  0.08      1.0     -       -       -       -       -       -       -       -       -       -
0.125 0.05      1.0     1.0     -       -       -       -       -       -       -       -       -
0.312 0.27      1.0     1.0     1.0     -       -       -       -       -       -       -       -
0.5   0.10      1.0     1.0     1.0     1.0     -       -       -       -       -       -       -
0.78  0.10      1.0     1.0     1.0     1.0     1.0     -       -       -       -       -       -
1.95  0.22      1.0     1.0     1.0     1.0     1.0     1.0     -       -       -       -       -
4.88  7.1e-06   0.85    0.004   1.0     0.27    1.0     0.27    1.0     -       -       -       -
5     0.091     1.0     0.85    1.0     1.0     1.0     1.0     1.0     1.0     -       -       -
12    1.3e-06   0.0066  1.7e-06 0.0006 8.0e-05  0.00027 3.6e-05 0.00593 0.00274 0.02193 -       -
24    5.9e-05   0.0026  5.7e-05 0.0005 0.00015  0.00033 0.00012 0.00230 0.00371 0.00854 1.0 -   -
48    < 2e-16   0.00031 2e-16   9.3e-08 5.0e-07 1.4e-06 8.6e-09 9.7e-06 1.1e-10 0.00067 0.00092
0.0768
```

Statistical analysis of data presented in figure 15.

```
lm(formula = log(Spiggin + 1) ~ log(FNtoDHTm + 1), data = alldata)

Residuals:
    Min      1Q  Median      3Q     Max
-3.4379 -0.6222  0.1798  0.6916  3.3709
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        9.73353    0.11408   85.32   <2e-16 ***
log(FNtoDHTm + 1) -1.36560    0.07258  -18.82   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.057 on 138 degrees of freedom
Multiple R-squared: 0.7195,     Adjusted R-squared: 0.7175
F-statistic:   354 on 1 and 138 DF,  p-value: < 2.2e-16

> anova(lm1)
Analysis of Variance Table
Response: log(Spiggin + 1)
                   Df Sum Sq Mean Sq F value    Pr(>F)
log(FNtoDHTm + 1)   1 395.71  395.71  354.05 < 2.2e-16 ***
Residuals         138 154.24    1.12
```

Statistical analysis of data presented in figure 17

```
        One-way analysis of means (not assuming equal variances)
data:  log(Spiggin + 1) and LNtoDHTn
F = 14.6157, num df = 9.000, denom df = 42.097, p-value = 2.322e-10
> pairwise1=pairwise.t.test(log(Spiggin+1),LNtoDHTn, pool.sd=F)
> pairwise1
        Pairwise comparisons using t tests with non-pooled SD
data:  log(Spiggin + 1) and LNtoDHTn

     0       0.05    0.4     0.5     2       5       20      40      50
0.05 1.00000 -       -       -       -       -       -       -       -
0.4  0.76893 1.00000 -       -       -       -       -       -       -
0.5  0.51993 0.11638 0.05054 -       -       -       -       -       -
2    0.01910 0.19978 1.00000 0.00632 -       -       -       -       -
5    1.00000 1.00000 0.69937 1.00000 0.21928 -       -       -       -
20   1.00000 0.51993 0.26831 1.00000 0.03177 1.00000 -       -       -
40   0.01989 0.00672 0.00242 0.62087 0.00055 1.00000 0.51993 -       -
50   5.3e-05 1.6e-05 5.1e-06 0.00443 9.4e-07 0.05687 0.00392 0.51993 -
80   0.03396 0.01494 0.00632 0.51993 0.00212 0.99756 0.39350 1.00000 1.00000
```

Statistical analysis of data presented in figure 18

```
lm(formula = log(Spiggin + 1) ~ log(LNtoDHTm + 1), data = alldata)
Residuals:
    Min      1Q  Median      3Q     Max
-5.8875 -0.6060  0.2607  0.9611  2.6450
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        10.7343     0.1719   62.43  < 2e-16 ***
log(LNtoDHTm + 1)  -0.8759     0.0949   -9.23 2.62e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.537 on 148 degrees of freedom
Multiple R-squared: 0.3653, Adjusted R-squared: 0.3611
F-statistic:  85.2 on 1 and 148 DF, p-value: 2.617e-16
> anova(lm1)
Analysis of Variance Table
Response: log(Spiggin + 1)
                   Df Sum Sq Mean Sq F value    Pr(>F)
log(LNtoDHTm + 1)   1 201.39  201.39  85.196 2.617e-16 ***
Residuals         148 349.86    2.36
```

Statistical analysis of data presented in figure 20

```
One-way analysis of means (not assuming equal variances)
data:  log(Spiggin + 1) and VCtoDHTn
F = 146.4974, num df = 7.000, denom df = 36.482, p-value < 2.2e-16
> pairwise1=pairwise.t.test(log(Spiggin+1),VCtoDHTn, pool.sd=F)
> pairwise1

        Pairwise comparisons using t tests with non-pooled SD
data:  log(Spiggin + 1) and VCtoDHTn

     0       0.05    0.5     2       5       20      50
0.05 0.26043 -       -       -       -       -       -
0.5  1.00000 1.00000 -       -       -       -       -
2    0.42751 0.00358 0.10259 -       -       -       -
5    0.43426 1.00000 1.00000 0.10259 -       -       -
20   5.6e-05 0.00018 5.9e-05 1.9e-05 0.00176 -       -
50   2.0e-13 1.7e-11 4.0e-10 1.6e-12 5.9e-05 1.00000 -
100  < 2e-16 8.5e-16 4.3e-10 5.5e-14 5.9e-05 0.76223 1.00000
```

Statistical analysis of data presented in figure 21

```
lm(formula = log(Spiggin + 1) ~ log(VCtoDHTm + 1), data = alldata)
Residuals:
    Min      1Q  Median      3Q     Max
-4.6666 -1.0787  0.4439  1.0923  2.8287

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        10.2289     0.1791   57.12   <2e-16 ***
log(VCtoDHTm + 1)  -3.2325     0.1795  -18.01   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.476 on 101 degrees of freedom
Multiple R-squared: 0.7625,     Adjusted R-squared: 0.7602
F-statistic: 324.3 on 1 and 101 DF,  p-value: < 2.2e-16
> anova(lm1)
Analysis of Variance Table
Response: log(Spiggin + 1)
                   Df Sum Sq Mean Sq F value    Pr(>F)
log(VCtoDHTm + 1)   1 706.76  706.76  324.30 < 2.2e-16 ***
Residuals         101 220.12    2.18
```

ANNEX 1

Exposure conditions during the AFSS data presented in the validation report

	Lab 1	Lab 2	Lab 3	Lab 4
Species	*Gasterosteus aculeatus*	*Gasterosteus aculeatus*	*Gasterosteus aculeatus*	*Gasterosteus aculeatus*
Origin	Wild (freshwater)	Lab born	Wild (freshwater)	Wild (marine)
Test type	Flow-through	Flow-through	Flow-through	Flow-through
Sex	Mixed	Female biased	Mixed	Mixed
Water temperature	15 ± 2°C	18 ± 2°C	15 ± 2°C	16 ± 2°C
Illumination quality	Wide spectrum fluorescence bulbs	Wide spectrum fluorescence bulbs	Wide spectrum fluorescence bulbs	Wide spectrum fluorescence bulbs
Photoperiod	12hd:12:hl	12hd:12:hl	12hd:12:hl	12hd:12:hl
Loading rate	<5g/L	<5g/L	<5g/L	<5g/L
Test chamber size	40L	40L	40L	40L
Test solution volume	30L	30L	30L	30L
Volume exchanges of test solutions	4.8	4.8	4.8	4.8
Age of test organisms	>10 months	>10 months	>10 months	>10 months
Approximate wet weight of adult fish (g)	0.94g	0.83g	1.62	1.73g
No. of fish per test vessel	25	20	15	15
No. of treatments	6	1	4-10	2-7
No. vessels per treatment	2	3	1	1
No. of fish per test concentration	50	60	15	15
Feeding regime	Once every two days; Frozen bloodworm/shri	Once a day; Frozen bloodworm	Once a day; Frozen bloodworm	Once a day; Frozen bloodworm/mysids
Aeration	Yes, air stone	Yes, air stone	Yes, air stone	Yes, air stone
Dilution water	De-chlorinated tap water	De-chlorinated tap water	Raw lake water	De-chlorinated tap water
Pre- exposure period	None	None	None	None
Chemical exposure duration	21d	21d	21d	21d
Biological endpoints	Wet weight spiggin	Wet weight spiggin	Wet weight spiggin	Wet eight spiggin

ANNEX 2

Study No	Lab	Date	Compound	No of treatments	Water Control	Solvent Control	Negative Control	Positive Control	No of fish
1	1	Nov 2001	FL	6		Yes		Yes	371 (200 Females)
2	4	Mar 2004	FL	4	Yes	Yes	Yes	Yes	120 (88 Females)
3	3	Apr 2004	FL	4	Yes	Yes	Yes	Yes	120 (64 Females)
4	3	May 2004	FN	4	Yes	Yes	Yes	Yes	109(43 Females)
5	3	Nov 2004	FN	10		Yes	Yes	Yes	185 (131 Females)
6	4	Feb 2005	FL	7	Yes	Yes	Yes	Yes	177 (160 Females)
7	4	Mar 2005	VZ	5	Yes	Yes	Yes	Yes	143 (113 Females)
8	4	Mar 2005	LN	4	Yes	Yes	Yes	Yes	122 (90 Females)
9	4	Sep 2005	VZ	2	Yes	Yes	Yes	Yes	67 (67 Females)
10	4	Sep 2005	LN	5	Yes	Yes	Yes	Yes	100 (99 Females)
11	4	Apr 2006	LN	2	Yes	Yes	Yes	Yes	65 (60 Females)
12	2	Aug 2008	FL	1				Yes	119 (97 Females)
								Total No	1698 (1212F)

ANNEX 3

Description of the stickleback-breeding test

The principle of the stickleback-breeding test is the same as in the androgenised-female test; the only difference is that spiggin is induced in male fish by photoperiod manipulations (employment of summer conditions, light:dark 18:6 hours and 18°C) rather than administration of DHT at 5µg/L to female fish. The breeding test requires the male fish to be in a non-breeding condition prior to the onset of exposures and the female fish to be in an advance stage of vitellogenesis so they can resume spawning within the 21-day exposure period. This is achieved by holding the male and female fish under different temperature and photoperiodic conditions for several months before the test commences.

The principle of the test and the details of the tank design have been described elsewhere (Katsiadaki *et al*, 2007). Briefly, the males are given a separate individual compartment (five in each tank) where they are provided with nest building material and gravel. The reason for this male isolation is to avoid mortalities and social hierarchies due to the aggressive behaviour that male sticklebacks demonstrate when entering a breeding cycle. The females are kept together at a sixth compartment of the tank and are in visual and chemical communication with the males throughout the test. The endpoints measured include morphometric, histological (gonadal histopathology), biochemical (VTG and spiggin) and behavioural (nest building/courtship behaviour, spawning) data. Of these, only spiggin and nest building/ courtship behaviour have a diagnostic value specific for antiandrogens and are also unique to the stickleback (i.e. not present in the other commonly used test species such as fathead minnow, zebrafish and medaka).

We developed and applied this experimental design in several *in vivo* exposures during a Defra project ABRAX (CT20051) and provided evidence for a strong antiandrogenic activity for FL (Sebire *et al*, 2008) and FN (Sebire *et al*, 2009). To our knowledge this was the first time that flutamide was shown to antagonise endogenous androgens and abolish the expression of clear androgen regulated characters such as spiggin induction and reproductive behaviour. In addition, we provided the first evidence of the harmful effect of fenitrothion on the physiology and behaviour of male fish. The lack of spiggin and the decrease in nest-building activity and courtship (zigzag) behaviours leads to significant

reduction of reproductive output of the impacted fish. The concentrations employed in our exposures were close to the reported concentrations of flutamide equivalent found in UK sewage treatment works and much less in comparison to the activities found in a European river or the oil platform produced water. The lowest observed effect concentration (LOEC) for flutamide was 100μg/L for the behavioural endpoints and 500μg/L for spiggin. For fenitrothion the LOEC for both endpoints was between 50 and 200μg/L (variation between replicates). Hence the reported high levels of antiandrogenic activity in aquatic environments is certainly as high if not higher than those causing experimental responses and the responsible chemical(s) appear to have the capacity to significantly disrupt the reproductive physiology of a common fish that is generally considered pollution-tolerant (e.g. the stickleback). The implications of this finding for the future of biodiversity (genetic or species) are enormous particularly if we take into account the increasing evidence of synergism in chemical mixtures.

Nevertheless, the effects of an effluent with allegedly high antiandrogenic activity tested under the same system were less conclusive (Sebire *et al*, *in prep*). The nest building and courtship behaviors were negatively affected by the effluent but the spiggin kidney content was not different between the control and exposed fish. Whether this was a result of a diminishing activity over time of effluent storage (revealed by the *in vitro* test applied in the aquaria effluent) or an indication of profound differences between the human and fish androgen receptors is still unknown. If the former hypothesis is correct then it means that the antiandrogenic compounds present in the effluent are not persistent and hence of little environmental impact. If the latter speculation holds true then the high antiandrogenic activities in the aquatic environment are not relevant to fish but only to humans and the observed effects on reproductive behavior of the stickleback might not be of endocrine nature.