**Review report: OECD External Review September, 2015**

**AOP 25:** [Aromatase inhibition leading to reproductive dysfunction (in fish)](https://aopkb.org/aopwiki/index.php/Aop%3A25)

**Content**

1. Introduction
	1. Material for the review
	2. Background of the AOP
2. Synthesis of the main issues on reviewers comments
3. Summary of the teleconference
4. Summary of the revisions
5. Outcome of the external review

Annex 1: Table of reviewers

Annex 2: Notes of the teleconference

Annex 3. Detailed comments and authors responses

1. **Introduction**

 **1.1 Material for the review:**

Associated wiki page: [https://aopkb.org/aopwiki/index.php/Aop:25](https://aopkb.org/aopwiki/index.php/Aop%3A25)

* 1. **Background of the AOP**

This adverse outcome pathway details the linkage between inhibition of gonadal aromatase activity in females and reproductive dysfunction, as measured through the adverse effect of reduced cumulative fecundity and spawning. Initial development of this AOP draws heavily on evidence collected using repeat-spawning fish species. Cumulative fecundity is the most apical endpoint considered in the OECD 229 Fish Short Term Reproduction Assay. The OECD 229 assay serves as screening assay for endocrine disruption and associated reproductive impairment (OECD 2012). Cumulative fecundity is one of several variables known to be of demographic significance in forecasting fish population trends. Therefore, this AOP has utility in supporting the application of measures of aromatase, or in silico predictions of the ability to inhibit aromatase, as a means to identify chemicals with known potential to adversely affect fish populations.

1. **Synthesis of the main issues on reviewers comments**

There were one set of general comments on the AOP25 by the reviewers. Some detailed comments referred to the snapshots were extracted from the wiki pages. An overview of reviewers' comments is organized below according to the charge questions.

All the reviewers considered the AOP 25 well written and scientifically correct. It was proposed that the MIE and key events should be presented in a causal order: also the naming of key events should be consistent.

All the reviewers recommend this to be submitted to the WNT and TFHA.

General / other comments:

* In general, it would be helpful to number the MIE, key events and AOP as follows: 1. Aromatase, Inhibition 2. 17beta-estradiol synthesis by ovarian granulosa cells, Reduction 3. Plasma 17beta-estradiol concentrations, Reduction 4. Vitellogenin synthesis in liver, Reduction 5. Plasma vitellogenin concentrations, Reduction 6. Vitellogenin uptake into oocytes and oocyte growth/development, Reduction 7. Cumulative fecundity and spawning, Reduction 8. Population trajectory, Decrease 9. Community food-web alterations. And the order of the KEs and KERs both in the tables and in the text should be organized accordingly in causal order.
* Also the KE naming should be consistent.

Response from AOP author:

The KEs and KERS have now been reordered in a causal order and the naming of KEs is now consistent. Regards to the taxonomic applicability to invertebrates is still too challenging to answer, the invertebrate system differs totally from vertebrates, also the food web alterations are still too far reaching to be covered in the AOP.

Charge question 1: *Check if the AOP incorporates the critical scientific literature and if the scientific content of the AOP reflects the current scientific knowledge on this specific topic*

Reviewers’ main comments

* The AOP is well described and incorporates the critical and recent scientific literature on the topic.
* For better clarity and traceability, it would be helpful to include more frequent references to studies that are mentioned in the AOP. For instance, references could be added to most of the introductory sections on the level of biological organization and the weight of evidence table.
* The format and the navigation through the AOP should be clarified. The colour code generated for the AOP diagram miss a legend in case there is any meaning behind it.
* To increase understanding and application of this AOP in the regulatory field, following points need to be clarified and further elucidated. Three chemicals, fadrozole, letrozole, and prochloraz, were listed as chemical initiators. In the KER “17beta-estradiol synthesis by ovarian granulosa cells, Reduction Directly Leads to Plasma 17beta-estradiol concentrations, Reduction” the evidence of ketoconazole was included. For example ketoconazole should be included in the list of chemical initiators, and also other similar chemicals where there is adequate data.
* Interrenal, instead of adrenal, should be used in fish.
* The AOP concept was invented in order to cover the need for an objective framework to integrate and interpret results from novel test methods and their prediction models for regulatory decision making. It further intents to reduce the need for animal testing by integrating alternative methods. Thus, during AOP development, all available data, including in chemico, in vitro, and in vivo tests, should be integrated to strengthen the overall AOP and to develop alternative predictive methods. For example, the MIE should be supported by in chemico studies demonstrating the interaction of the stressor and the affected molecule (e.g. protein). The MIE can also be supported by SAR or QSAR methods. Early KEs and KERs at the molecular level could be assessed and supported by data obtained by in vitro other alternative methods. Downstream KEs and KERs at the tissue or organism level should be supported by data obtained in the target tissue/ organism, which is in this case adult female fish.

There were several more specific comments referring to the snapshot with page numbering. These are presented in Annex 3.

Response from AOP author:

Most of the reviewer’s comments are taken into account by the author and the AOP is revised accordingly. The detailed responses are available at the AOP wiki and in Annex 3.

Charge question 2: *Verify the weight of evidence judgement/scoring provided by AOP developers for KEs, KERs and the overall AOP*

Reviewers' main comments

* In the weight of evidence table, section on support for essentiality of KEs, there are no calls (high, moderate or strong) for the individual KEs. These should be added, and under the bullet point “vitellogenin production in liver (transcription, translation), reduction”, it says that there is strong support the essentiality of this event. However, the endpoint was not specifically examined in stop/reversibility studies. Therefore, there is only indirect evidence. I suggest the essentiality call for this KE should be changed to “moderate”.
* It was also noted that why are there no calls for the quantitative understanding of KERs according to Annex 2 of the AOP handbook.
* It was suggested that the table of KEs should be organized according to the order of KEs. In the subsection entitled "Consistency" of the overall assessment of the AOP, the statement of the phrase "...the cumulative fecundity endpoint can be less sensitive than key events measured at lower levels of biological organization” is unclear and not always true. As some of KEs of lower levels may be measured in vitro, sensitivity comparison may not be meaningful.
* The supporting literature for a connection of the VTG content in plasma and cumulative fecundity. Based on the literature provided in this AOP, the AOP developer correctly judged this connection as moderate. However, biological plausibility might be increased if more data on other fish species, like zebrafish and medaka, would be included. No scientific evidence is available for the direct connection of reduced cumulative fecundity to a declined population trajectory. Thus, on this basis, judging the WoE for this KER as moderate is appropriate. However, data presented here rely on one population model only, working with “ideal” fathead minnow populations. Inconsistency: The KER Plasma vitellogenin concentrations, reduced directly leads to vitellogenin uptake into oocytes and oocyte growth/ development, reduction is weighted as strong in the Table in Annex 1 and as moderate in the summary provided on the web page. This should be changed to strong, as plasma VTG is the only available VTG source.
* Judgement of empirical support for most of the KERs is appropriate. Here, the linearity of concentration-response and temporal concordance should be weighted. WoE for the KER Plasma 17β-estradiol concentration, reduction directly leads to vitellogenin production in liver (transcription, translation), reduced is judged as weak. I would suggest changing this to moderate, as there is empirical support for a direct dependence of E2 concentration and VTG concentration at the molecular level, determined by mRNA analyses or microarray studies, as vtg transcription is regulated by an ERE in its promoter region (please review current literature on alternative test systems). The WoE for the overall AOP was scored as strong for fathead minnow, and moderate for zebrafish and medaka. I agree with this judgement on the basis of the literature provided. However, as mentioned earlier, the overall AOP could also be scored as strong for the other two fish species if more literature on these species will be included.

#### Response from AOP author:

Most of the comments could be taken into account and the detailed responses are available at the AOP wiki.

Charge question 3: *What would be the regulatory applicability of this AOP in your opinion?*

Reviewers' answers and comments

* Since the AOP covers a wide range of endpoints from aromatase inhibition to population, it is highly probable that it will be applicable for mechanistic tests as part of an IATA and in the revision or development of guidance documents for testing of endocrine disrupters, e.g. the OECD Guidance Document 150 on Standardized Test Guidelines for Evaluating Chemicals for Endocrine Disruption.
* Nevertheless, it seems that for this AOP support for essentiality of the "final outcome" i.e. reduction of cumulative, fecundity and spawning is **moderate only**. Therefore it might be difficult to take a regulatory decision based on this AOP only for ED's activities if looking at apical endpoint.
* Identification of Endocrine Disrupting Chemicals (EDCs) is needed under several pieces of European Union (EU) legislation. Currently, the regulatory identification of EDCs is mainly based on the general consensus on the WHO definition, which consists of three essential elements, i.e. chemical-induced adverse effects (adversity), chemical specific endocrine modes/mechanisms of action (MOAs) and the causal relationship (causality) between adverse effects and endocrine MOAs. AOPs cover all essential elements for identification of EDCs and show the complex biology of adversity and MOAs. These will help regulators understand the complexity of identification of EDCs.
* Current regulatory tests focus on EATS pathways. In contrast, AOPs include not only EATS pathways but also other pathways, e.g. PPARs, RXR, that are essential to development, growth, and reproduction. Within each AOP, different targets at molecular, cellular, organ/tissue and individual levels could be identified and the adverse outcome would be predicated. Such information would be of help for prioritizing chemicals, for grouping chemicals and for developing an integrated testing strategy.
* By providing evidence that there is a robust relationship between aromatase inhibition and reduced egg production of fish (which is an endpoint directly used in risk assessment), this AOP will enhance the group-wise hazard assessment of aromatase inhibitors. The strong point is probably the taxonomic applicability. If my understanding of the published literature is correct, aromatase is sufficiently conserved to conclude that a compound which has been identified as aromatase inhibitor in fish species 1 will also be an aromatase inhibitor in fish species 2.
* The more problematic question relates to the quantitative relationships: how much do effect concentrations of aromatase inhibition differ between fish species (how much does this depend on species-specific toxicokinetic parameters or is it mainly a question of evolutionary conservation of the aromatase sequence?).
* In regulation, there is the need for timely and robust decision making. Thus, regulatory toxicity testing has to become more cost-effective and efficient. This could be achieved by directing the testing resources, focusing on chemicals of highest concern, limiting testing to the most probable hazards, and targeting most vulnerable species. Specific AOPs should provide causal links between the MIE and the AO of regulatory concern via well-established KEs and KERs, and should facilitate regulatory decision making.
* Endocrine disruptor testing is of high regulatory concern. It is a tiered approach, which includes QSAR methods and high-throughput screening (HTS) assays at Tier 1 for prioritization, and reproductive/developmental studies at Tier 2 for those chemicals which are identified as potential endocrine disruptors. The here provided AOP could be integrated at Tier 1, as a supporting tool for integration of already available data for a given substance. These data available could be compared to the here described AOP. If KEs and KERs are overlapping, the chemical should be further tested. As the biological plausibility for the KERs is predominantly weighted as strong, presence of one KE or KER is very likely triggering the adverse outcome of reduced cumulative fecundity. Furthermore, based on data presented in this AOP, HTS assays could be developed, to allow prioritization of substances for which no or only limited data are available.
* This AOP will be very helpful interpreting data obtained with fathead minnow. However, the AOP should be also applicable to zebrafish and medaka. It could be easily used for interpretation of data from these fish species if available literature is integrated. If it should be used as analysis tool for HTS data, a measureable KE (e.g. mRNA data) at the molecular biological level should be included.

#### Response from AOP author:

It seemed at the teleconference that most of the comments could be taken into account and the AOP can be revised accordingly.

Charge question 4: *Overall Assessment of the AOP - Would you recommend this AOP to be submitted to the Working group of the National Coordinators for the Test Guidelines Programme (WNT) and the Task Force on Hazard Assessment (TFHA) for endorsement?*

Reviewers' main comments

* Yes, the AOP is well enough developed and would be ready for submission to the WNT and TFHA after revision according to this external review process.
* The current AOP focuses only on female fish. As the majority of test guidelines include both males and females, it is important to include certain information/statement over male fish so that regulators can get an overall picture.

#### Response from AOP author:

All the remarks above represent minor comments, pointing to possible enhancements of specific aspects of the AOP. And it should be noted that all the AOPs should be considered living documents, which will improve when the science is progressing.

1. **Summary of the teleconference**

A teleconference was organized on the AOP23 reviews back to back with AOP25 teleconference 23rd October 2015. At the time of this there were no any written responses to the reviews available. And there was a general agreement on the value of the AOP25 and that it should proceed after EAGMST review and discussion to the WNT and TFHA. It was also noted that a single AOP should be simple and robust and it should be then well interlinked to other possible close AOPs, as in the case of AOP23 and AOP25, which describe linking aromatase inhibition, androgen receptor agonism, estrogen receptor antagonism, or steroidogenesis inhibition, to impaired reproduction in small repeat-spawning fish species.

The notes of the teleconference are presented in Annex 2.

1. **Summary of the author’s responses**

The responses from the author indicated that the comments can be taken into account.

1. **Outcome of the external review**

All of the five reviews were uploaded at Wiki in due time (end of September early October 2015). A teleconference was organized on the AOP23 reviews back to back with AOP25 teleconference 23rd October 2015. At the time of this there were no any written responses to the reviews available (see Annex 3). Some general issues and some more detailed comments were discussed at the teleconference. And there was a general agreement on the value of the AOP25 and that it should proceed after EAGMST review and discussion to the WNT and TFHA. It was also noted that a single AOP should be simple and robust and it should be then well interlinked to other possible close AOPs, as in the case of AOP23 and AOP25, which describe linking aromatase inhibition, androgen receptor agonism, estrogen receptor antagonism, or steroidogenesis inhibition, to impaired reproduction in small repeat-spawning fish species.

Some of the author’s written responses were available at the time of this review manager report 14th December 2015, but the responses are under development and the AOP revision is proceeding.

As a conclusion for the future the review process time-lines should be obeyed, and it is crucial that the author’s responses are available at the time of the teleconference in order to ensure productive and transparent discussion at the teleconference.

**Annex 1**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Affiliation** | **representing** | **e-mail** |
| Christopher Fassbender | ICAPO/PETA, Germany | research/academia | ChristopherF@peta.de |
| François Busquet | Universität Konstanz, Germany | research/academia | caat-eu-policy@uni-konstanz.de |
| ZhiChao Dang | Rijksinstituut voor Volksgezondheid en Milieu, Research, NL | regulator | zhicao.dang@rivm.nl |
| Elke Muth-Köhne | Fraunhofer Institute, Germany | research/academia | elke.muth-koehne@ime.fraunhofer.de |
| Helmuth Segner | Universität Bern, Zentrum fur Fisch und Wildtiermedizin, Switzerland | research/academia | helmut.segner@vetsuisse.unibe.ch |

**Annex 2**

**Notes of the 23rd October 2015 joint teleconference on AOP 23 and AOP 25**

Participants:

Author of both AOPs: Dan Villeneuve (villeneuve.dan@epa.com )

Review manager: Jukka Ahtiainen (mikris.ahtiainen@gmail.com )

- Christopher Fassbender (ICAPO, ChristopherF@peta.de )

- Ionna Katsiadaki (UK, ionna.katsiadaki@cefas.co.uk )

- François Busquet (DE, caat-eu-policy@uni-konstanz.de )

- ZhiChao Dang (NL, zhicao.dang@rivm.nl )

- Elke Muth-Koehne (DE, elke.muth-koehne@ime.fraunhofer.de )

- Helmut Segner (Switzerland, helmut.segner@vetsuisse.unibe.ch )

-

Project 1.12: AOP linking aromatase inhibition, androgen receptor agonism, estrogen receptor antagonism, or steroidogenesis inhibition, to impaired reproduction in small repeat-spawning fish species

As the AOP 23 [Androgen receptor agonism leading to reproductive dysfunction](https://aopkb.org/aopwiki/index.php/Aop%3A23) and AOP25 [Aromatase inhibition leading to reproductive dysfunction (in fish)](https://aopkb.org/aopwiki/index.php/Aop%3A25) have interlinkages and the reviewer groups have a certain overlap it was decided to have these teleconferences back to back.

* Unfortunately at the time of this there were no any written responses to the reviews available. Some general issues and some more detailed comments were discussed at the teleconference.
* And there was a general agreement on the value of the AOP23 and that it should proceed after EAGMST review and discussion to the WNT and TFHA.
* It was also noted that a single AOP should be simple and robust and it should be then well interlinked to other possible close AOPs, as in the case of AOP23 and AOP24, which describe linking aromatase inhibition, androgen receptor agonism, estrogen receptor antagonism, or steroidogenesis inhibition, to impaired reproduction in small repeat-spawning fish species.
* Some of the technical comments were discussed more in detail.

**Annex 3. Detailed comments by the reviewers and authors responses by December 14th 2015**

**General / other comments**

**Reviewer 1:**

• In general, it would be helpful to number the MIE, key events and AOP as follows: 1. Aromatase, Inhibition 2. 17beta-estradiol synthesis by ovarian granulosa cells, Reduction 3. Plasma 17beta-estradiol concentrations, Reduction 4. Vitellogenin synthesis in liver, Reduction 5. Plasma vitellogenin concentrations, Reduction 6. Vitellogenin uptake into oocytes and oocyte growth/development, Reduction 7. Cumulative fecundity and spawning, Reduction 8. Population trajectory, Decrease 9. Community food-web alterations

• The order of KEs both in the table on page 3 and the following pages is confusing (KEs 3, 4, 6, 2, 7, 5, 8 according to the numbering system suggested above) and should be brought in line with the causal order. Analogously, the order of KERs both in the table on page 9 as well on the subsequent pages is confusing (KERs 1, 2, 3, 7, 6, 5, 4 according to the numbering system suggested above) and should be brought in line with the causal order.

• Some KE titles in the text diverge from KE titles in the AOP diagram; e.g. “vitellogenin synthesis” in the text vs. “vitellogenin expression and production” in the AOP diagram, or “population trajectory, decrease” in the text vs. “population trajectory, decline” in the AOP diagram. For clarity, this should be in line.

• Page 8, KE 5, Plasma vitellogenin concentrations, reduction: Under “evidence supporting taxonomic applicability”, the situation in invertebrates could be explained more in detail.

• Pages 2-9: In the description of the KEs, the Adverse Outcome (community food-web alterations) is not discussed. Should this be added?

• Page 10, section on quantitative understanding of the linkage of KER 1-2 (according to the numbering system suggested above): The last bullet point should not be a bullet point but normal text because it summarizes the list of bullet points on models of ovarian steroidogenesis (“These may be adaptable…”).

• Page 13, section on “evidence supporting taxonomic applicability of KER 2-3 (according to the numbering system suggested above): The list of species needs further elaboration.

• The same font should be used in all of the weight of evidence tables.

**Response from AOP author:**

* Key event and KER tables were re-ordered to match the sequence depicted in the AOP diagram.
* The AOP diagram has been updated to match the KE titles represented in the KE table. Several attempts were made to replace the old diagram with the revised version, but these were unsuccessful. The authors will seek the assistance of the Wiki developers.
* Page 8, KE5 (i.e., Event:221) - the text was revised to clarify that while many invertebrates synthesize vitellogenin, invertebrate vitellogenins are transported via hemolymph rather than plasma.
* Although included in the AOP diagram, potential community food web alterations were not included as a KE in the AOP description. The nature of the food web alteration would be both site and species dependent, and thus is difficult to generalize as a KE. Consequently, we chose to leave it out of the AOP description. The community level of organization and this KE were removed from the revised AOP diagram.
* KER 1-2 (i.e., Relationship:45) - the bullet was removed from the last statement in the "quantitative understanding of the linkage" section
* KER 2-3 (i.e., Relationship:5) - the text in the evidence supporting taxonomic applicability section was revised to include a more general and mechanistic basis, rather than simply listing the general classes or organisms covered in the taxonomic relevance widget table above.
* Differing fonts in the weight of evidence tables do not represent a barrier to use or understanding. No change was made in response to this comment.

**Charge question 1:*Check if the AOP incorporates the critical scientific literature and if the scientific content of the AOP reflects the current scientific knowledge on this specific topic***

**Responses**

**Reviewer 1:**

The AOP incorporates the critical and recent scientific literature on the topic. However, for better clarity and traceability, it would be helpful to include more frequent references to studies that are mentioned in the AOP. For instance, references could be added to most of the introductory sections on the level of biological organization and the weight of evidence table.

**Reviewer 2:**

I have no comment about the content but rather about the format and the navigation through all these information. The color code generated for the AOP diagram miss a legend in case there is any meaning behind it. Otherwise, please mention absence of signification all the tables that are summarising key events etc.. Can be manipulated in an ascending or descending way with no added value so far. It could be excluded maybe?

**Reviewer 3:**

The AOP is well described and included the critical literature, which reflects the current understanding in reproductive impairments resulted from the exposure to aromatase inhibitors in small model fish. To increase understanding and application of this AOP in the regulatory field, following points need to be clarified and further elucidated. Three chemicals, fadrozole, letrozole, and prochloraz, were listed as chemical initiators. In the KER “17beta-estradiol synthesis by ovarian granulosa cells, Reduction Directly Leads to Plasma 17beta-estradiol concentrations, Reduction” the evidence of ketoconazole was included. Should ketoconazole be included in the list of chemical initiators? Should other similar chemicals be included?

In the methods described, OECD test guideline 456 H295R Steroidogenesis Assay was not included. Should it be included in the AOP?

Fecundity is an important apical endpoint for fish full/partial life cycle tests, and for TG240 (MEOGT). In addition to TG229, it is important to add these long term toxicity tests.

Interrenal, instead of adrenal, is used in fish. Should the term interrenal be included in the AOP?

**Reviewer 4:**

The AOP incorporates the critical scientific literature and reflects the current scientific knowledge. The text well identifies areas with high confidence and more controversial areas, for instance, it depicts that the “linkage between vitellogenin uptake into oocytes” and “oocyte growth” is only correlative. The text is written in a differentiated way, although I have comments on mainly two points:• While I found it a strength that the possible role of compensatory responses is addressed which may confound concordance of dose-response relationships, I suggest that still more consideration is given to such “caveats”. Through this, the user will learn on the application domain of AOP as well as on potential pitfalls. Of course, there must be a compromise between the discussion of potential confounders on the one hand and keeping the AOP description reasonably short and concise on the other hand (an AOP cannot be a literature review), but I feel that the current version of AOP could still tolerate certain additions here and there (see examples below). • A second comment deals with possible “branching” points or potential deviations from the linear effect propagation as suggested in the AOP. I understand that an AOP must be linear avoiding to get lost among all the possible side-directions and deviations, still it might be helpful for the user to be informed on sites within AOP where such deviations from the linear pathway may be possible (see examples below). In these cases I recommend to extend a little bit and to include at least a very short hint (1, 2 sentences at max) on possible branching events, misfits etc. (see examples below). Apart from these comments, I find that the author found a good balance between the contrasting demands and wishes, still, at certain

Specific comments:

Page 2, “level of biological organization”. The text says that “inhibition of aromatase activity results in decreased rate of E2 production…” – what of course is correct. However, reduced estrogen synthesis may also arise from the inhibition of other enzymes in steroidogenesis, like StAR or other CYP enzymes. For instance, one group of well-known aromatase inhibitors are the agricultural and pharmaceutical fungicide azoles; they are designed to CYP51 of funghi and have the most potent inhibitory activity on that CYP enzyme, but have some affinity to other CYP enzymes as well. Thus, azole inhibitory effects are not restricted to aromatase, but include – to a varying degree - the other steroidogenic CYP enzymes. This fact does not have any implications for the validity of the aromatase AOP, still I feel it would be good to mention shortly, as background information, reduced E2 synthesis can also result from inhibition of other steroidogenic enzymes. This would inform the user that there may be compounds which reduce E2 synthesis without showing inhibitor activity in the aromatase assay.

***Response: I fully agree with the reviewer's point. However, the potential for impacts on other steroidogenic enzymes to cause decreased E2 synthesis will be covered in the context of the broader AOP network being described in the knowledgebase. It is expected that key event 2 in this AOP (i.e., Event:3 in the knowledgebase) will link to additional AOPs in the KB that will include MIEs related to the inhibition of other key enzymes involved in steroidogenesis. See also the earlier response regarding the use of evidence related to ketoconazole.***

Page 2, first line “Measurement/detection”: probably an “ of” needs to be added?: “.. production of tritiated..”

**Response: "of" was added as recommended by the reviewer.**''

Page 2, “Name”: I agree that CYP19 orthologs are generally not found in invertebrates, meaning that adverse effects of aromatase inhibitors should primarily occur in vertebrates. However, referring to the comment above that the inhibitory action of aromatase inhibitors such as the azole fungicides is not exclusive for CYP19 but can affect other CYP enzymes as well, these compounds may interfere with CYP-dependent hormone synthesis of invertebrates, for instance, ecdysteroid synthesis (in fact, there are publications reporting this). Also the imposex phenomenon in molluscs has been attributed to an inhibitory action of organotins on molluscan aromatase, but this view has been challenged more recently. The AOP text as it is worded now is in my view too narrow in excluding the endocrine system of invertebrates as potential target of aromatase inhibitors.

***Response: I agree with the reviewers' point that chemicals identified as aromatase inhibitors may also impact other steroidogenic cyps, including, potentially those involved in ecdysteroid synthesis in invertebrates. However, the MIE is specific to aromatase. That is the MIE of concern for the current AOP. Impacts on steroidogenic cyps in invertebrates will be covered in the context of an AOP network within the KB in which those targets are considered explicitly as the MIEs. No change was made in response to this comment.***

Page 2, “name”. In the last sentence, mentioning is given to the dualism of piscine aromatase, i.e. the existence of a brain and a gonad form. Apart from this, little consideration is given to this dualism and its possible implications for the reproductive output. My suggestion is to extend a little bit on this. A first question in this context is whether gonad and brain aromatase show similar sensitivity towards aromatase inhibitors (according to Hinfray et al. 2006, Comp Biochem Physiol C 144:252-262, the differences appear to be modest). A second aspect to be considered is the consequences of brain and aromatase inhibition. This AOP focuses on the gonadal form, and shows how its inhibition leads to reduced fecundity. However, also inhibition of brain aromatase may contribute to reduced reproductive output , as the brain aromatase is involved in the regulation of the hypothalamic-hypophysis axis and may influence FSH/LH release from the pituitary, what eventually translates into reduced fecundity. The presence of such a mechanism does not mean that the gonad-based AOP is invalid, the brain effect probably would occur in parallel (thus, we have here a branching point: the MIE “aromatase inhibition” proceeds through a gonadal and a brain pathway merging again in the apical outcome, fecundity). Nevertheless, I think it is important for the user to have this background information. The brain aromatase story is essential enough to be shortly addressed in the AOP, perhaps just one, two sentences plus reference to a literature review e.g., Cheshenko et al. 2008, Gen Comp Endocrinol 155: 31-62.

***Response: The "site of action" under "how this key event works" explicitly defines this key event in terms of ovarian granulosa cells. A separate KE/MIE would describe impacts on brain aromatase activity, which would lead to different downstream key events potentially associated with neurosteroid regulation of the HPG axis. Nonetheless, as per the reviewer's suggestion, a few additional sentences were added to clarify that both isoforms may be expressed in fish ovaries, but given the dominant expression of the cyp19a1a isoform and evidence that both isoforms appear to be similarly sensitive to aromatase inhibitors, in defining/describing/and measuring this key event, it is not really necessary to distinguish between isoforms. Total activity can be considered.***

Page 3, Key events: I wonder why the sequence of the KEs in the list on p. 3 differs from that in the AOP diagram. Mechanistically, the diagram makes sense, while the order of KE in the list p. 3 is confusing. When reading the AOP, jumping from e.g., vitellogenin transcription forward to vitellogenin uptake into the oocytes and then back to vitellogenin plasma levels is confusing. Any specific reason for this order of KE descriptions ?

***Response: There was no specific reason for the order of KEs in the list. This is simply a function of the way they were ordered in the wiki as they were created and/or revised. The order of the KEs and KERs in the tables has been revised and now appears in the same sequence as presented in the AOP diagram. We agree that presenting the KEs in order makes the tables much easier to follow.***

p. 3, “level of biological organization, individual”: The first sentence – for the first time – mentions that E2 is synthesized not exclusively in the gonads (before, only gonadal E2 production has been mentioned) is already confusing because the reader my wonder whether there is a specific reason to mention here the extragonadal E2 sources. In addition, I am not sure whether those extragonadal tissues indeed contribute to plasma E2 levels. In my understanding – but I may be wrong - the brain E2 has primarily a paracrine, no endocrine activity. If so, then the statement “E2 synthesized by the gonads and other steroidogenic tissues is transported to other tissues” is misleading. My suggestion is to remove the “other steroidogenic tissues”.

***Response: Event:219 was revised to remove reference to other steroidogenic tissues as suggested by the reviewer. The "biological plausibility" section of Relationship:5 was also revised to acknowledge that estradiol can be synthesized in other tissues, but generally plays a paracrine role only. Extragonadal the contribution of extragonadal synthesis to circulating estradiol concentrations can generally be considered negligible***

Page 4 “How it is measured or detected”: When talking of RT-PCR detection, is it necessary to mention that there exist vitellogenin isoforms, which can show differential sensitivity to E2 induction?

***Response: The developers are not aware of specific references that define the sensitivity differences of the different isoforms. It is our opinion that mention of vitellogenin isoforms showing differential sensitivity to E2 is more relevant to the quantitative understanding of the linkage between event:219 and event:285. Consequently, the following statement was added to the "quantitative understanding of the linkage" section of relationship:252. "There are multiple isoforms of vitellogenin. The sensitivity and inducibility of each of those isoforms may vary somewhat. Consequently, response-response relationships may vary somewhat depending on the specific isoform for which QPCR primers or antibodies were developed"***

Page 5 “vitellogenin uptake…”: Actually, this KE does not measure vitellogenin “uptake” into the oocytes but vitellogenin “accumulation” in the oocytes, since the suggested measurement method – histology – is not able to detect the uptake process but only the accumulation of vitellogenin being present in the oocyte. This is a detail, but for the sake of consistency it may be more accurate to talk of “vitellogenin accumulation into oocytes..”

***Response: I agree in principle with the reviewer's point. I will work with the AOP-wiki development team to try to change the title of Event:309 to "vitellogenin accumulation into oocytes..." as suggested. Because this page and title are currently linked to multiple AOPs, I am not sure how to make the change directly without potentially disrupting the links to other pages.***

Apart from this linguistic issue: Elsewhere in the AOP, it is mentioned that the concordance between vitellogenin accumulation in the oocytes and spawning success sometimes fails. The AOP describes oocyte maturation and spawning mainly as a function of vitellogenin accumulation. However, the regulation of oocyte maturation and spawning is clearly more complex (e.g. Clelland and Peng, 2009, Mol Cell Endcorinol 312: 42 -52), and adding a sentence with such a statement will help to understand why misfits between vitellogenin and spawning success can occur.

***Response: A sentence indicating that other factors are known to be involved in the regulation of oocyte maturation and spawning was added to the biological plausibility section of the weight of evidence summary on the AOP:25 page.***

Page 6, “How is it measured”: A formal thing - in this para, references are given by numbers ([3], etc.), in other paragraphs names and years are cited. Why this difference ? and where is the reference list with the numbered references?

***Response: The numbers were actually links to the reference documents/websites that were embedded in the wiki text. That is why no accompanying reference was provided. More descriptive links and an accompanying reference in the reference section were provided. However, the two methods documents do not fit a standard reference format***

Page 7, taxonomic applicability: does the “Taxonomic applicability” refer to the process of “egg laying”, or does it refer to “E2 regulation of egg laying”?

***Response: The KE description for Event:78 was developed to be a stand-alone entity in the wiki which does not make direct reference to up-stream or downstream events in an AOP. This allows this KE node to be shared among multiple AOPs in order to facilitate the assembly of AOP networks. Consequently, the taxonomic applicability in this case applies to all egg laying animals, not just those whose egg-laying is estrogen regulated. No change was made in response to this comment.***

Page 14, “how is plasma vitellogenin measured?”: The Korte et al. reference is perhaps not the most appropriate for the purpose here, because it addresses the vitellogenin ELISA more as a side aspect. I would cite here papers which focus specifically on the vitellogenin ELISA, e.g., on Holbech et al., 2001, Comp Biochem Physiol C 130: 119-131, Fenske et al. 2001, Comp Biochem Physiol C129: 117-132, etc.

***Response: The Holbech et al. and Fenske et al. references were added as per the reviewer's suggestion.***

p.8, population trajectory: In my view it is mandatory to refer here to the role of life histories for translating reduced fecundity into reduced population growth. The response of demographic parameters to reduced egg numbers can differ strongly between different life histories, and this aspect must be considered both for the extrapolation of the fecundity data to population growth within one and the same species, as well as for the extrapolation from one species to another species. Generally, I found the parts on the population trajectories are somewhat short and superficial (there is too much “it is plausible that lower fecundity will translate in lower population growth”). Extending on this discussion will add credibility to the KER; currently, it is too much “, ad could still gain from including some population dynamics theory.

***Response: A few sentences regarding the importance of life history and reproductive strategies in influencing the nature of the relationship between cumulative fecundity and spawning and population trajectory were added to the "uncertainties and inconsistencies" section of Relationship:94. The primary developer of this AOP is not an expert in population dynamics theory and is not in position to contribute intelligently on the topic without additional study. The developers invite the reviewer and/or the broader community to contribute additional relevant text to address this comment.***

**Reviewer 5:**

The AOP concept was invented in order to cover the need for an objective framework to integrate and interpret results from novel test methods and their prediction models for regulatory decision making. It further intents to reduce the need for animal testing by integrating alternative methods. Thus, during AOP development, all available data, including in chemico, in vitro, and in vivo tests, should be integrated to strengthen the overall AOP and to develop alternative predictive methods. For example, the MIE should be supported by in chemico studies demonstrating the interaction of the stressor and the affected molecule (e.g. protein). The MIE can also be supported by SAR or QSAR methods. Early KEs and KERs at the molecular level could be assessed und supported by data obtained by in vitro other alternative methods. Downstream KEs and KERs at the tissue or organism level should be supported by data obtained in the target tissue/ organism, which is in this case adult female fish.

Generally:

Literature provided by the AOP developer is limited regarding in chemico or in vitro data, or data obtained in alternative test systems like the fish embryo or transgenic fish. Furthermore, for supporting early KEs also FSDT studies with aromatase inhibitors might be included, as long as identical KERs are addressed.

The developer also only sporadically included literature on other fish species like zebrafish and medaka, which are also important test species for risk assessment. Empirical support for aromatase inhibition should include studies on zebrafish (e.g. Sun et al., 2010, Dang et al., 2015) and medaka (e.g. Zhang et al., 2008; Sun et al., 2007, 2011). These studies provide evidence for a concentration-dependent effect on e.g. plasma E2 levels, vtg concentration, oocyte growth, or fecundity. Furthermore, the study of Sun et al., 2007 with medaka also provides evidence for decreased survival of the F1-generation, which would provide support for the AO at the population level. This reference, for example, is included in the overall assessment of the AOP, demonstrating concentration-response of cumulative fecundity. However, it is not included in the section describing the individual KERs, and the test species is not mentioned.

Generally, the AO Cumulative fecundity is likely more reliable with zebrafish, as higher egg numbers could be achieved, and zebrafish are less sensitive to physical disturbances. Including data on zebrafish and medaka would thus increase applicability of this AOP for regulatory purposes (change from moderate to strong for these fish species). As the title of the AOP is not restricted to fathead minnow, this is highly recommended.

Several of the KERs for this AOP are identical e.g. to AOP:23: Androgen receptor agonism leading to reproductive dysfunction, written by the same AOP developer. As individual KEs/ KERs should written in such a way that they could be shared by other AOPs, data supporting this KE/ KER might be also shared. Thus, the here discussed AOP might also include data on androgen receptor agonists as long as the KE /KER is identical. This is applicable for most of the KERs except for the first, Aromatase inhibition directly leads to 17β-estradiol synthesis by ovarian granulosa cells, reduction.

I would further suggest screening the literature on molecular data for aromatase inhibition, like proteome data, microarray analyses and qPCR data. Aromatase inhibition very likely results in downregulation of genes possessing an ERE in their promoter regions, via the KE of reduced plasma E2 concentrations. This could be also included as an additional, measureable KE.

Specifically:

Page 8: Plasma vitellogenin concentrations, reduction. Fenske et al., 2001, and Holbech et al., 2001 should be included as references for this method.

***Response: The suggested references were added to the "How it is Measured or Detected" section of the KE page for Event:221.***

Page 14: I would suggest to include a description of vtg transcriptional regulation by the estrogen receptor-responsive element (ERE) in the promoter region, for the KER Plasma 17β-estradiol concentration, reduction directly leads to vitellogenin production in liver (transcription, translation), reduced. This would significantly increase the biological plausibility of this KER.

***Response: A sentence regarding the presence of EREs in the promoter region of VTG genes and supporting references were added as suggested by the reviewer.***

Furthermore, transgenic fish (ERE:GFP; Gorelick et al., 2011) will probably be useful as alternative method for EDC testing and application of the AOP for regulatory purposes (for data review for substances of concern). They are already proven to be very sensitive to estrogenic substances but might also be applicable in challenging experiments with estrogens and aromatase inhibitors. Even though transgenic fish are not likely to be included in standard ecotoxicological tests and test guidelines, data obtained with these test systems could be very helpful in prioritization.

Page 16ff: Include Kramer et al., 2011. This reference also describes applicability of the AOPs regulatory purposes in general.

Page 20f: The AOP developer states that transcription and translation occurs prior to protein synthesis. This is not correct, as transcription and translation are both parts of protein synthesis. Correct: transcription occurs prior to translation. Please provide literature for the proposed time-lag between transcription and plasma VTG concentrations.

Minor comments:

The order of description on page 9ff of the snapshot is not considering the order of KEs and KERs of the AOP.

**Response from AOP author:**

**To reviewer 1:**

I don't quite follow what the reviewer is asking for.

**To reviewer 2:**

A color coded legend was added to the revised AOP diagram, indicating that green KEs are MIEs, and red KEs are AOs. The tables have been reorganized so that KEs are listed in sequence.

**To reviewer 3:**

Regarding chemical initiators: Chemical initiators are associated with MIE pages and pertain to interactions at the MIE only. Studies with ketoconazole, a steroidogenesis inhibitor that is not necessarily specific to aromatase, provide evidence that supports the relationship between measures of ovarian steroid synthesis and decreases in circulating E2 concentrations. However, because ketoconazole is not necessarily specific for aromatase and has actions on other steroidogenic cytochrome P450 enzymes at concentrations lower than those at which it impacts aromatase, we felt it was appropriate to exclude it as a chemical initiator for this particular pathway. There are additional chemicals that likely to act fairly potently and specifically at this molecular initiating event, for example anastrozole. However, as most of the evidence compiled to date draws on the chemicals listed, those were the ones included. The current list of chemical initiators is not comprehensive.

OECD test guideline 456, H295R steroidogenesis assay was not included as a method for measuring the MIE (Event:36) because it is not specific to aromatase. Inhibition of other steroidogenic enzymes can impair E2 and T production in the assay. The test guideline is broadly suitable for detecting reductions in E2 synthesis, and is listed as a method for detecting that key event (Event:36).

Reference to OECD test guideline 240 was added to the "how it is measured or detected" and "regulatory examples using this adverse outcome" sections of the description of Event: 78.

Reference to adrenal tissue was found in the biological plausibility section of Relationship:5. The term adrenal was replaced with interrenal as recommended by the reviewer.

**To reviewer 4:**

Please see responses inserted below each specific comment provided by reviewer 4.

**To reviewer 5:**

**Charge question 2:*Verify the weight of evidence judgment/scoring provided by AOP developers for KEs, KERs and the overall AOP***

**Responses**

**Reviewer 1:**

a) In the weight of evidence table, section on support for essentiality of KEs, there are no calls (high, moderate or strong) for the individual KEs. Why? b) In the weight of evidence table, under the bullet point “vitellogenin production in liver (transcription, translation), reduction”, it says that there is strong support the essentiality of this event. However, the endpoint was not specifically examined in stop/reversibility studies. Therefore, there is only indirect evidence. I suggest the essentiality call for this KE should be changed to “moderate”.c) Why are there no calls for the quantitative understanding of KERs according to Annex 2 of the AOP handbook?

**Reviewer 2:**

here again, I have no comment about the content which is fine with me but rather about the format. For non-AOP expert entering in the AOP-wiki world, I was a bit lost with "temporal concordance" or "essentiality calls". By adding some more details, this could become much clearer...to me at least.

**Reviewer 3:**

The weight of evidence table in the annex is very useful, which helps understand the biological plausibility and empirical support. In the section of Empirical Support for Linkage, results of DEHP/MEHP in mammals have been included. Did they have similar mechanisms to those of fadrozole, prochloraz in fish?

In the section of Empirical support for linkage, some lines of evidence were based on an increase in 17β-estradiol concentration resulting in an enhancement in VTG in male fish. Such evidence is different from the title of reduction and female of AOPs.

The table of the KEs was not presented according to the orders of KEs, which may cause some confusing. It is suggested that the table of KEs should be organized according to the order of KEs. In the subsection entitled "Consistency" of the overall assessment of the AOP, the statement of the phrase "...the cumulative fecundity endpoint can be less sensitive than key events measured at lower levels of biological organization” is unclear and not always true. As some of KEs of lower levels may be measured in vitro, sensitivity comparison may not be meaningful.

**Reviewer 4:**

The WOE evaluation of the specific KERs and the overall AOP is based on a critical evaluation of the existing data, although there is some bias towards the small laboratory fish species. Nevertheless, from the data as presented, a grading of most KERs of this AOP to be “strong”, with only two “moderate” KERs, is well defendable. The weakest KER, at least as presented, is the link between fecundity and population growth. Here, the part on “uncertainties” needs more consideration. There are good examples available in the literature which provide insight going beyond the cited statement of Wester et al. that “egg production may not translate in a simple manner to population reductions”. For instance, reference is given to the experimental lake study in Canada, and this study is a very nice example how fish species experiencing the same estrogenic exposure can differ in their population-level responses (for instance, Kidd et al., Proc Roy Soc London B).

A general comment on the WOE part: the various aspects contributing to WOE such as “biological plausibility” etc. are discussed, and in the table in annex 1, an excellent synthesis is given for all three WOE criteria, i.e. biological plausibility, essentiality and empirical support. However, would it not be important to have also one integrative assessment of the strength of the KERs (not separated into biological plausibility, essentiality and empirical support), as it has been done for this particular AOP in the recent publication by Becker et al. 2015, Regulat Toxicol 72:517-537. Or did I miss a table (on p. 22, it is referred to a “summary table”, but I found only the table 1 and the concordance table I the annex).

p. 15, “Evidence supporting taxonomic applicability”. The authors address the “taxonomic applicability” question by asking which taxa possess CYP19 gene(s). However, under this headline I would have expected information on whether a compound which is a CYP19 inhibitor in, e.g., mammals, is also a CYP19 inhibitor in fish. In other words: How conserved is the CYP19 protein structure and the inhibitor-CYP19 interaction? For this, it might be helpful to refer to the paper of Celander et al., 2011, Environ Toxicol Chem 30:52-63, which addresses exactly this question.

p. 15, “quantitative understanding..:”: Quantitative understanding of KERs can be derived from computer models, as described here. However, it may also be derived from comparing concentration response curves/effect thresholds between two KEs: for instance, if the chemical concentration needed to induce a reduction of circulating E2 is higher than the concentration which leads to reduced liver vitellogenin synthesis, makes a causative relationship between the two events unlikely. Thus, the question here (and for the other KERs as well) is: what does the existing literature tell on how the concentration-response curves compare between two subsequent KEs? (the answer is probably difficult due to data paucity, but where possible, it would be helpful if this aspect is addressed).

p. 17, “Uncertainties: Again, I suggest to spell out in more detail on the complex relation between fecundity and population growth. In the text it is said “it may not translate in a simple manner” – my suggestion is to make this statement stronger, actually it should be warned against the expectation of a simple linear relationship, particularly under field conditions (think for instance on the Bruker model for trout recruitment; according to this, reduced fecundity is likely to lead to reduced young-of-the-year mortalities and thus overall population recruitment may not suffer from a decline in fecundity, also probably the vulnerability of the population will increase. Again, the AOP cannot provide an in-depth discussion of these complex relationships, but it also should avoid to give rise to too simplistic expectations.

**Reviewer 5:**

The weight of evidence (WoE) table in the Annex 1 is intended to provide (1) support for the biological plausibility of KERs, (2) support for the essentiality of KE, and (3) empirical support for KERs. Based on these considerations the confidence in the overall AOP for potential regulatory application is assessed. Biological plausibility of each of the KERs is the most influential consideration for the WoE of the overall AOP, with the essentiality of the KEs as second most influential and the extent of empirical support being least influential element. Regarding the biological plausibility, I agree with the WoE judgment/scoring provided by the AOP developer. KERs that were judged as strong are well established and belong to the basic knowledge in endocrine disruptor evaluation in fish. As described in the previous charge question, WoE judgment/scoring could be even more enhanced if literature on other test systems, e.g. on other fish species like zebrafish and medaka, on related Moans leading to the same KER, and also on in vitro tests providing evidence for the supposed KER would be more extensively included. Furthermore, it would be advantageous if scientific evidence (i.e. references) would also be provided in the WoE table presented in Annex 1.

Biological plausibility of two KERs was weighted as moderate. These KERs were:

*Vitellogenin uptake into oocytes and oocyte growth/ development, reduction directly leads to cumulative fecundity and spawning, reduction. Cumulative fecundity and spawning, reduction directly leads to population trajectory, decrease.*

No supporting literature for a direct connection of the VTG content in oocytes and the fecundity of female fish could be provided. However, there is supporting literature for a connection of the VTG content in plasma and cumulative fecundity. Based on the literature provided in this AOP, the AOP developer correctly judged this connection as moderate. However, biological plausibility might be increased if more data on other fish species, like zebrafish and medaka, would be included. No scientific evidence is available for the direct connection of reduced cumulative fecundity to a declined population trajectory. Thus, on this basis, judging the WoE for this KER as moderate is appropriate. However, data presented here rely on one population model only, working with “ideal” fathead minnow populations. There are likely more population models published, which could also be cited (e.g. Kramer et al., 2011). It is worth considering also including full life cycle studies for this KER, as conclusion might also be drawn from survival success of the F1 generation (e.g. Sun et al., 2007, for medaka). This might increase the biological plausibility of this KER.

Inconsistency: The KER Plasma vitellogenin concentrations, reduced directly leads to vitellogenin uptake into oocytes and oocyte growth/ development, reduction is weighted as strong in the Table in Annex 1 and as moderate in the summary provided on the web page. This should be changed to strong, as plasma VTG is the only available VTG source.

For judgment of the essentiality of the KEs, the AOP developer does not provide a scoring of individual KEs in the table in the Annex 1. However, scoring is included in the summary of the AOP provided on the web page. Essentiality of the KEs is supported by time-course experiments with aromatase inhibitors on fathead minnow, which could reliably demonstrate the essentiality of one KE to provoke the following; in vitro studies, studies using alternative models, and studies with other fish species are not included. The developer indicated that the essentiality of the KEs is demonstrated by stop/reversibility experiments, which are however not explained nor referenced (also not in the main text). References should be included. The developer should revise this section of the table.

Judgment of empirical support for most of the KERs is appropriate. Here, the linearity of concentration-response and temporal concordance should be weighted. WoE for the KER Plasma 17β-estradiol concentration, reduction directly leads to vitellogenin production in liver (transcription, translation), reduced is judged as weak. I would suggest changing this to moderate, as there is empirical support for a direct dependence of E2 concentration and VTG concentration at the molecular level, determined by mRNA analyses or microarray studies, as vtg transcription is regulated by an ERE in its promoter region (please review current literature on alternative test systems). The WoE for the overall AOP was scored as strong for fathead minnow, and moderate for zebrafish and medaka. I agree with this judgment on the basis of the literature provided. However, as mentioned earlier, the overall AOP could also be scored as strong for the other two fish species if more literature on these species will be included.

**Charge question 3:*What would be the regulatory applicability of this AOP in your opinion?***

**Responses**

**Reviewer 1:**

Since the AOP covers a wide range of endpoints from aromatase inhibition to population relevant effects, it is highly probable that it will be applicable for mechanistic tests as part of an IATA and in the revision or development of guidance documents for testing of endocrine disrupters, e.g. the OECD Guidance Document 150 on Standardized Test Guidelines for Evaluating Chemicals for Endocrine Disruption.

**Reviewer 2:**

It covers endocrine disrupters' activities and therefore if adverse effects are observed, this could lead to ED identification of tested substances. Nevertheless, it seems that for this AOP support for essentiality of the "final outcome" i.e. reduction of cumulative, fecundity and spawning is **moderate only**. Therefore it might be difficult to take a regulatory decision based on this AOP only for ED's activities if looking at apical endpoint. However, I look very much forward how WNT will reflect when an AOP is "accepted". Would it be possible to waive a certain number of tests for some of the KE in case AOP evidence is strong? I am probably anticipating slightly question 4.. Nevertheless, It would be worth discussing how many KE are necessary to draw conclusions without testing/assessing them all. Otherwise, I do not get the point of doing AOPs and its purpose will become mainly a collaborative exercise to deliver a comprehensive, endorsed encyclopedia.

**Reviewer 3:**

Identification of Endocrine Disrupting Chemicals (EDCs) is needed under several pieces of European Union (EU) legislation, including the Regulation on industrial chemicals (Registration, Evaluation, Authorization and restriction of Chemicals, EC 1907/2006, REACH), the Plant Protection Products Regulation (EC 1107/2009, PPPR), and the Biocides Products Regulation (528/2012, BPR). Currently, the regulatory identification of EDCs is mainly based on the general consensus on the WHO definition, which consists of three essential elements, i.e. chemical-induced adverse effects (adversity), chemical specific endocrine modes/mechanisms of action (MOAs) and the causal relationship (causality) between adverse effects and endocrine MOAs. AOPs cover all essential elements for identification of EDCs and show the complex biology of adversity and MOAs. These will help regulators understand the complexity of identification of EDCs. Besides, current regulatory tests focus on EATS pathways. In contrast, AOPs include not only EATS pathways but also other pathways, e.g. PPARs, RXR, that are essential to development, growth, and reproduction. Within each AOP, different targets at molecular, cellular, organ/tissue and individual levels could be identified and the adverse outcome would be predicated. Such information would be of help for prioritizing chemicals, for grouping chemicals and for developing an integrated testing strategy. It is important to indicate that the AOP needs extensive amount of data which might be possible for a few chemicals but will not be possible for a majority of chemicals. Current data requirements under REACH, PPPR, BPR, etc. do not cover all key events of the AOP.

**Reviewer 4:**

By providing evidence that there is a robust relationship between aromatase inhibition and reduced egg production of fish (which is an endpoint directly used in risk assessment), this AOP will enhance the group-wise hazard assessment of aromatase inhibitors. The strong point is probably the taxonomic applicability. If my understanding of the published literature is correct, aromatase is sufficiently conserved to conclude that a compound which has been identified as aromatase inhibitor in fish species 1 will also be an aromatase inhibitor in fish species 2. The more problematic question relates to the quantitative relationships: how much do effect concentrations of aromatase inhibition differ between fish species (how much does this depend on species-specific toxicokinetic parameters or is it mainly a question of evolutionary conservation of the aromatase sequence?). Is the existing database good enough to support extrapolations from in vitro aromatase inhibition assays to in vivo effect concentrations? These and other questions are addressed at several places in the AOP but I think there should be one conclusive section where such questions are comprehensively discussed, with respect to the question which parts of the AOP are ready for regulatory applicability, which ones are close to this and what is actually missing to move them further (but maybe this is included in the text and I missed it – honestly, the current AOP format is not always easy to read and to understand where exactly we are in the discussion process).

**Reviewer 5:**

In regulation, there is the need for timely and robust decision making. Thus, regulatory toxicity testing has to become more cost-effective and efficient. This could be achieved by directing the testing resources, focusing on chemicals of highest concern, limiting testing to the most probable hazards, and targeting most vulnerable species. Specific AOPs should provide causal links between the MIE and the AO of regulatory concern via well-established KEs and KERs, and should facilitate regulatory decision making. In terms of the evaluated AOP of aromatase inhibition leading to reproductive dysfunction (in fish), the causal link is between aromatase inhibition and a reduced cumulative fecundity at the organism level, or a declining trajectory at the population level, which are the adverse outcomes of regulatory concern. Applicability of an AOP in a regulatory context is only provided if there are reliable methods for a quantitative evaluation of the KEs, and if KERs are biologically plausible. The here described AOP provides sufficient evidence for methods for KE measurements, and most of the KERs are appropriately judged as strong regarding their biological plausibility (however, in most cases only for studies with fathead minnow). Thus, the principal requirements for the regulatory applicability of this AOP are provided. However, a KE assessing effects at the molecular biological level is missing (e.g. qPCR data, as mentioned in charge question 1), which would be helpful in judging data at the molecular level, which amount will be increasing in the upcoming years. Endocrine disruptor testing is of high regulatory concern. It is a tiered approach, which includes QSAR methods and high-throughput screening (HTS) assays at Tier 1 for prioritization, and reproductive/developmental studies at Tier 2 for those chemicals which are identified as potential endocrine disruptors. The here provided AOP could be integrated at Tier 1, as a supporting tool for integration of already available data for a given substance. These data available could be compared to the here described AOP. If KEs and KERs are overlapping, the chemical should be further tested. As the biological plausibility for the KERs is predominantly weighted as strong, presence of one KE or KER is very likely triggering the adverse outcome of reduced cumulative fecundity. Furthermore, based on data presented in this AOP, HTS assays could be developed, to allow prioritization of substances for which no or only limited data are available. These HTS assays should assess the effect of aromatase inhibition on easily measurable KEs, like E2 and VTG plasma content, or analyses of gene expression, e.g. in cultured hepatocytes.

Summary:

This AOP will be very helpful interpreting data obtained with fathead minnow. However, the AOP should be also applicable to zebrafish and medaka. It could be easily used for interpretation of data from these fish species if available literature is integrated. If it should be used as analysis tool for HTS data, a measureable KE (e.g. mRNA data) at the molecular biological level should be included.

**Charge question 4:*Overall Assessment of the AOP - Would you recommend this AOP to be submitted to the Working group of the National Coordinators for the Test Guidelines Programme (WNT) and the Task Force on Hazard Assessment (TFHA)for endorsement?***

**Responses**

**Reviewer 1:**

Yes, the AOP is very well developed and would be ready for submission to the WNT and TFHA after revision according to this external review process.

**Reviewer 2:**

Based on the author's comment, it seems that some evidence is missing to support the following key event "reduction of vitellogenin uptake into oocyte growth/development". It might be considered to undergo new laboratory experiment to confirm it. In fine, as I said previously in question 3, if AOP is robust, and perfectly understood, there might be no further need to identify every single KE to draw regulatory conclusion. It will therefore save time etc. As the AOP is described, I would still recommend it for submission.

**Reviewer 3:**

Yes. It is important to get the official stamp for publishing the AOP. To increase the regulatory applications of the AOP, following points are suggested: The current AOP focuses only on female fish. As the majority of test guidelines include both males and females, it is important to include certain information/statement over male fish so that regulators can get an overall picture. It is also important to indicate that the major androgen in male teleost fish is 11-ketotestosterone. In addition, it is important to specify the uncertainties or inconsistencies that are related to chemicals. This would be of great help for non-experts and risk assessors to understand confounding factors.

**Reviewer 4:**

All the remarks above represent minor comments, pointing to possible enhancements of specific aspects of the AOP but the essence of the AOP appears to be solid and strong. Thus, I recommend the AOP to be submitted to WNP and TFHA.

**Reviewer 5:**

As an AOP is intended to be a constantly developing document, I would recommend this AOP for submission, as it provides a linear and plausible relationship of aromatase inhibition and reduced cumulative fecundity, and will be helpful in aromatase inhibitor identification and prioritization. As soon as it is applied for prioritization, more and more data will be available to strengthen the individual KEs/ KERs, also for other fish species. If molecular biological or HTS methods will be developed for testing aromatase inhibition, an early KE could be included.