REVIEWER #1

General comments:

1. Since the entire document refers to nonhuman animals, use the term "sex" not "gender". In addition, use "dam" instead of "mother". Finally, use "that" instead of "who" when referring to nonhuman animals (e.g., last paragraph of p. D-24).

2. When referring to a specific PND, be consistent with either a space before the day or no space.

3. Rodent behavior can be substantially altered by poor maternal care and potential neurotoxicants can alter maternal care. While this is addressed in Module E, those assessments/data are not required for inclusion. However, the reviewer should be aware of such potential.

4. Module B provides some description of effect size, but this should be strengthened and encouraged for inclusion in reports.

5. Each module should have some specific information about testing for statistical outliers.

6. All modules should be written in U.S. English format (i.e., labor, not labour as on p. D-5).

*7. There are small grammatical/punctuation errors throughout and these should be caught by a spell/grammar check. A couple of these examples are: p. B-9, in the sentence beginning "Testing time in relation to", the word "directed" should be "directly"; p. D-17 "The order of start point **is** varied . . ."; p. E-4, the sentence beginning "The level of confidence in the overall conclusion" needs a period at the end.*

SPECIFIC QUESTIONS:
Does the document provide sufficient guidance to assist regulatory scientists in reviewing reports to determine whether critical details regarding procedure, study design, results (including summary and individual data for all relevant parameters), and statistical evaluation are included in the reports for studies conducted under the EPA or OECD DNT Guidelines? If not, why not?

Module A: This module is probably the most difficult to adequately review. Observational and functional evaluations vary greatly from lab to lab. However, this module does describe the most important points: lack of standardization, necessity of treatment-blind testers, etc. There are some areas that could use a bit more specific details. The document states that tests should progress from the least to the most invasive (p. A-6); however, if a regulatory scientist is not a behavioral expert, they may not be able to determine if this was done. The document states that pups must not be separated from the dam "for very long", but an actual duration (i.e., 15-20 minutes?) would be more helpful. I agree that it is best to use the same subjects at each testing time.

Module B: This module is a very well-written portion. I would suggest adding a reference to Figure 1 describing number of rats and strain as well as if the same subjects were tested at each age. Height of the vertical positioning of photocell sensors (to detect rearing) should be required for inclusion in reports, since this will vary by age of the animal.

Module C: For readability, add the red dashed line to the female graph of Figure 3. One important procedural detail missing is that the report must describe the size of the chambers used - that is, the typically Plexiglas restrainers used to hold each subject. Most companies sell different sizes for different sizes of animals (weanling, adult, etc.). If a weanling is tested in an adult-sized chamber, it will likely be able to move freely around the chamber. Also, many laboratories include "blank" trials in which the same dependent variables are collected but no auditory stimulus is presented. These are important to verify that the subjects are not "startling" without the auditory stimulus. Similar to the comment for Module A, knowledge of the testing history for the subjects is essential. Auditory startle testing is obviously stressful for the subject and it would be important to know what testing occurred prior and afterwards. On Figure 2, it is not clear what the solid lines represent. On p. C-17, in the "Treatment-by-Trial blocks-by-sex" paragraph, I believe the phrase should read: contract interaction tests, simple effects tests, or specific contrasts. One thing that may be noted or described in reports is that the initial trial typically has the highest startle response, often much more than even the second trial. When averaging the first 10 trials, this extremely high response contributes to much of the reported increased response on the first trial block.

Module D: In general, this module is very specific and detailed which provides useful information for the reviewer. The examples (e.g., Table 3) are excellent (but see my comment about Figure 4 below). Only a few minor edits for this module are suggested. In the second column of Table 1, perhaps the first row can state M or T, instead of not noting the T maze at all. In the first column, there is a hyphen between T maze in one row but no hyphen in the other. Table 2, last row has a checkmark that needs to be moved down. In Figure 1, it would be helpful if a small platform diagram is placed in one arm of each maze with an indication that this is the "goal" or "end" arm. Reviewers unfamiliar with the water-based tasks would likely appreciate a temperature range for the water to ascertain if the reported methodology is appropriate. Additionally, it has been shown that rats can use urine trails from a previous subject to navigate in the Morris water maze (Means et al., 1992) so it is good practice to gently stir the water between subjects (whether a letter maze, Morris water maze or Cincinnati/Biel maze). Some laboratories using the Morris water maze place a circular curtain around the maze and it is the curtain that contains the extra-maze cues (not all laboratories place the cues on the room walls). Figure 2 is blurry in my copy. The distance from the water maze wall for the platform must be described. If the platform is too close to the wall, it is an easier task to solve since rats will tend to engage in thigmomaxis on the first few trials. If swimming near the wall allows them to find the platform, this is not the intended learning. The inter-trial interval must be constant for animals run in squads. If one rat locates the platform quickly on its second trial, the next rat cannot begin its second trial until the inter-trial interval has ended. Thus, even with automated videotracking systems, a stopwatch or timer is typically used for the inter-trial interval.

Initially, I was a bit confused about Figure 4 and the text interpretation. In a second reading, it appears clearer. However, I think the text could be made clearer to indicate that very little (if any) learning occurred in the high dose PTU group (left graph) while learning did occur in the high dose heptachlor group, but that learning occurred at a much slower rate than the control group. Also, since swimming speed is so important, mention here that there were no differences in swim speed so it is clear that these effects are on cognitive behavior, not motoric performance (if this is true).

P. D-23 only notes swim speed for the probe trial. Since most videotracking systems collect swim speed on all trials, swim speed needs to be detailed for all trials.

Perhaps I am missing something, but I do not see the increased complexity of Path "B" in the Cincinnati maze. Is it because in Path "B", the animal may choose to continue in a straight path or make a turn? I don't see that as more complex than making a left or right turn – it's still a two choice decision? Perhaps this could be more descriptive as to why Path "B" is more complex? If it is anticipated that there

will be increased use of the Cincinnati/Biel mazes, it would be helpful to know if the data are typically collected via videotracking or by hand.

Use of the KM method (survival curves) can be recommended for all mazes for which there is a maximum cut-off time for a trial (not just passive avoidance).

I greatly appreciate inclusion of the statement on p. D-41 "It is also possible to have significant interactions but not have a significant effect on follow-up tests".

Module E: The issue of changes in exposure levels to the dam over gestation and lactation is discussed on p. E-10-11. While the document notes that "changes in chemical intake over time should be noted", it could be strengthened by noting that where dietary exposure (food or water) is used, those data should be included in the report. Since the dams are regularly weighed and food or water intake is measured, calculations of the actual dose is simple.

Given that regulatory reviews are conducted independent of any review or interpretation presented by the study authors: does the document provide sufficient guidance to assist regulatory scientists in interpreting the data and results from regulatory studies conducted under the EPA or OECD DNT Guidelines? If not, why not?

Module A: This is where the module is the most descriptive and helpful. But it would be useful to have a stronger statement about maternal toxicity potentially influencing offspring behavior. Even a decreased gestational weight gain can potentially alter offspring behavior. Any potential toxicity in the dam must be understood to have a potential effect on the offspring's behavior. In addition, if a treatment causes delays in such landmarks as eye opening, pinna detachment, etc., behaviors that are dependent on those will appear to be affected.

Module B: I don't know anyone who includes rearing in the measure of total activity. Typically, total activity is the sum of the ambulatory and small non-ambulatory movements only. I am delighted to see on p. B-12 the absolute prohibition against doing an interval-by-interval analysis of dose effects in the absence of a significant interaction. The description of the potential effects (p. B-13-14) is excellent.

Module C: Again, I am delighted to see on p. C-13 the absolute prohibition against doing separate comparisons among treatment groups at different trial blocks without a significant interaction. With the minor edits suggested above, this portion should serve as sufficient guidance.

Module D: This was a well-written module. With the minor edits suggested above, it should provide valuable guidance.

Module E: P. E-12 recommends the use of the translating time website. But it should be emphasized that this useful site only compares on the basis of neurogenesis. The matching of ages is not based on synaptogenesis, myelination, pruning, etc.

Does the document provide the correct summary of the kinds of information to look for in submitted data, provide relevant examples, and assist in interpretation of any treatment-related changes?

Module A: With the minor edits described above, it will be as complete as it can be.

Module B: In section 6.1, one of the more common dependent variables is average speed. Most automated systems (whether photocell- or video-based) provide this measure. P. B-14 notes that Figure

1 "shows how the variability decreases with age" but this cannot really be seen without knowing the group sizes for each group.

Module C: As noted above, some reports may list startle response for the first trial separately since it is typically so much higher than even that exhibited on the second trial. The examples listed in this portion were very helpful in understanding potentially "real" treatment-related effects as well as potentially false positive treatment-related effects.

Module D: Most examples were excellent (see my one note about Figure 4 above).

Module E: On p. E-13, the sentence "Interpretation of DNT data, however, can be challenging" is quite an understatement! However, this module clearly presents the issues that must be kept in mind when evaluating DNT data.

REFERENCE CITED ABOVE:

Means LW, Alexander SR, O'Neal MF (1992) Those cheating rats: male and female rats use odor trails in a water-escape "working memory" task. Behav Neural Biol 58:144-151.