The California Association of Sanitation Agencies (CASA) appreciates the opportunity to provide comments on the Toxicity Provisions Staff Report Appendices. CASA is a statewide association representing more than 125 local agencies that provide wastewater collection, treatment, clean energy and water recycling services to millions of Californians. CASA remains committed to effective, scientifically sound, and appropriately implemented use of toxicity testing to protect California waterways. We have been an active partner in the development of these provisions since their inception and respectfully submit the following comments to maximize the utility of these appendices.

CASA has reviewed Appendix J and the statistical findings made by State Water Board staff in Appendix J appear to be accurate. CASA is also appreciative of the revisions made to improve the accuracy and clarity of the document, specifically the acknowledgement that a test can be “truly non-toxic” and still result in a failure using the Test of Significant Toxicity (TST) statistical approach. Unfortunately, the Appendix does nothing to address or alleviate our primary and still unresolved concern: the frequency at which known non-toxic samples are identified as toxic.

A statistical “false positive” as described in the Appendix is not the same thing as a “false indication of toxicity.” The data in the Appendix only addresses the former and not the latter. This is an essential distinction. CASA submitted a white paper to the Water Board and staff in 2019 entitled *Ceriodaphnia dubia Short-term Chronic Reproduction Test: Understanding the Probability of Incorrect Determinations of Toxicity in Non-toxic Samples*, which highlighted our concerns with false indications of toxicity in clearly non-toxic blank samples. Although CASA does not see any significant shortcomings in the statistical analysis as conducted in Appendix J, the analysis unfortunately failed to take into account several important factors that limit its applicability as a decision-making tool for development of statewide toxicity regulations.

Moreover, the occurrences of statistical false positives that can be ameliorated through careful laboratory controls on within test variability as described in Appendix J do not appear to be related to actual false indications of toxicity in non-toxic blank samples. Instead, a significant proportion of laboratories conducting toxicity tests on these known non-toxic samples seem to identify unacceptably large effects (>25%) at an unacceptably high frequency.

CASA looks forward to working alongside State Water Board staff and other stakeholders to address and hopefully resolve this fundamental issue related to false indications of toxicity through development and implementation of the State Water Board’s *Ceriodaphnia* toxicity study.

Given the above, we suggest that modifications be made to these Appendices regarding how the results of the analysis are communicated and summarized, to ensure that the results are used appropriately in decision-making relating to the proposed statewide toxicity regulations. Specifically:
• Appendix J is often vague as to whether the statements being made apply to population versus sample level statistics or altogether non-statistical terms.
  o The definition of a “false positive” in Appendix J is statistically based. However, CASA is more concerned with “false determinations of toxicity” where the observed effects in a known non-toxic sample are greater than the Regulatory Management Decision (RMD) of 25%. **We recommend clarification by using the term “statistical false positive” for the State Water Board’s “false positive” to distinguish between the two.**
  o Appendix J fails to recognize that the “true toxicity” of the samples used in this post-hoc analysis was unknown. Consequently, these must be sample level statistics. However, this isn’t made clear, which could lead to the misinterpretation that these were truly non-toxic samples.

• Appendix J relies upon long-term medians of within-test variability using the control treatments’ coefficient of variation (CV). This fails to acknowledge that half of a laboratory’s tests had higher CVs and that the corresponding number of additional replicates to achieve an acceptable statistical false positive rate will often be much higher. The end result of this underestimation of additional replicates may be significantly higher efforts and costs to achieve compliance.

• Figure J-4 incorrectly identifies 2014 as the point at which the TST was required in the Sanitation Districts of Los Angeles County’s (LACSD) permits. Two (2) of this permittee’s eight facilities incorporated the TST in January 2015 and the other six (6) in mid to late 2015. As a result, the key finding that LACSD’s municipal laboratory performance increased upon use of the TST is not supported.

• It is unknown which states and/or municipalities’ data were used to develop the control CV percentiles in the US EPA 2010 TST Technical Document. This could have the unintentional consequences of artificially underestimating variation and, ultimately, on the setting of appropriate alpha values. Multiple states’ (e.g., South Carolina, North Carolina and Hawaii) whole effluent toxicity programs have a test acceptability criterion that invalidates tests with a control CV exceeding 40%. If data from these states, and any others with upper CV test acceptability criteria, were included in the analysis and calculation of the CV percentiles, these national percentiles would be incorrectly identified as lower than they actually were. Figure J-4 demonstrates this phenomenon has a high likelihood of occurrence in a municipal monitoring program. Distribution of this data set would aid in addressing this underlying issue.

• The simulations performed by Fox et al. assume that each test is independent (i.e., the likelihood of misidentifying toxicity in one test has no impact on future tests). While this is true statistically, it is unclear whether this is accurate in practice. Downturns in culture happen to every lab. These dips are often not identified until tests are concluded and can take weeks to recover from. Poor culture health tends to be reflected in poor test performance (e.g., high CVs and lower than normal control reproduction) and are often observed in clusters, as opposed to randomly distributed. If a false determination of toxicity is made in one test and it is caused or exacerbated by poor culture health, it would be expected to occur in multiple tests initiated concurrently and/or within several days of each other, and which may have no relation to “toxicity.”

We appreciate the opportunity to provide comments on Appendix J and look forward to continuing to work with the Water Board to develop a final version of the toxicity provisions and help scope and implement the forthcoming *Ceriodaphnia dubia* study.

Sincerely,

Adam D. Link  
Executive Director