EXHIBIT

F

White Paper Update

Discussion and Recommendations Related to *Arbacia punctulata* Whole Effluent Toxicity Testing Using Combined Effluent from the Bayamón, Puerto Nuevo, and Bacardí Wastewater Treatment Plants

Prepared for

Bacardi Corporation

And

Puerto Rico Aqueduct and Sewer Authority

October 2011

CH2MHILL®

Contents

WHITE PAPER UPDATE

Discussion and Recommendations Related to *Arbacia punctulata* Whole Effluent Toxicity Testing Using Combined Effluent from the Bayamón, Puerto Nuevo, and Bacardí Wastewater Treatment Plants

Introduction

The wastewater treatment system (WWTS) at the Bacardi Corporation (Bacardi) rum distillery in Cataño, Puerto Rico, shares an ocean outfall with the Puerto Rico Aqueduct and Sewer Authority (PRASA) Bayamón and Puerto Nuevo Regional Wastewater Treatment Plants (RWWTPs). The combined effluent of the three facilities is discharged more than 0.5 miles offshore at a depth of 140 feet (ft) below mean sea level into dynamic ocean waters through a high-rate (>100:1 dilution) diffuser.

This document is an update of a white paper prepared in May 2007 for the Bacardi Corporation (Bacardí) concerning the most appropriate way to evaluate whole effluent toxicity (WET) test results for the sea urchin *Arbacia punctulata* (CH2M HILL, 2007). The original version of this white paper was used to support Bacardi's request that the 25 percent Inhibition Concentration[1] (IC_{25}) calculation instead of the No Observed Effect Concentration (NOEC) or Lowest Observed Effect Concentration (LOEC) be used to evaluate compliance with the numerical Chronic Toxicity Unit (TUc) limits provided by the Puerto Rico Environmental Quality Board (EQB) *Interim Mixing Zone and Bioassay Guidelines* (April 1988), as incorporated by reference in the Puerto Rico Water Quality Standards Regulation (PRWQSR).

On May 20, 2008, the U.S. Environmental Protection Agency (EPA) issued final National Pollutant Discharge Elimination System (NPDES) permits to each of the three facilities.[2] Each permit incorporated the following effluent toxicity limitation language:

> b. Effluent Limitation:
>
> **No single IC25 test result for any species or effect in the combined discharge shall be less than 1.00%.**
>
> Results shall be reported as the IC25 percentage effluent of the combined discharge. This permit requires additional toxicity testing if a chronic toxicity effluent limit is violated. The permittee shall notify EPA in writing within fourteen days of the permittee's receipt of results violating this effluent limitation.

EPA, in response to comments from Bacardi and PRASA on the draft NPDES renewal permits issued on July 1, 2011, states the following:

> ***The Definitions Section of the 2010 Puerto Rico Water Quality Standards Regulation (PRWQSR) defines the Criteria Continuous Concentration (CCC as "the EPA national water quality criteria recommendation for the highest instream concentration of a toxicant or an effluent to which organisms can be exposed indefinitely without causing an unacceptable effect. It is equal to CCC =1.0TUc."***

[1]The IC_{25} is the percent concentration of a test solution that results in a 25% inhibition of a measurable biological response – in this case fertilization success of *Arbacia* eggs.

[2]The effective date of the permit (EDP) for the final permit all three facilities was July 1, 2008.

Also included in the Definitions Section is the calculation defining a chronic toxicity unit (TUc), which is the "... reciprocal of the effluent concentration that causes no observable effect on the test organisms by the end of the chronic exposure period, obtained during a chronic toxicity test, as defined by the following equation:

TUc = 100/NOEC
(The NOEC value should be expressed in terms of the percent (%) of the effluent in the dilution water)."

EPA is apparently trying to make the case that the NOEC should be used to define the TUc. However, it is noted that EQB reviewed the May 2007 toxicity evaluation white paper and did not object to incorporating the IC_{25} as the toxicity compliance evaluation criterion in the draft permits, nor has EQB taken issue with any of the quarterly or annual toxicity compliance results presented per the requirements of the final permits.

Problems Associated with Use of Hypothesis Testing Endpoints when Interpreting WET Test Results

This White Paper discusses the following key issues:

- Problems associated with use of hypothesis testing endpoints (NOEC) when interpreting WET test results
  - High Number of False Positive Results
  - Lack of Percent Minimum Significant Difference (PMSD) Upper and Lower Bounds for *Arbacia*
  - Variability Among Tested Concentrations
- Lack of demonstrable reason for EPA to have changed from the IC_{25} used in the previous permit for these evaluations; and
- The technical basis on which Bacardi is requesting that the IC_{25} be reinstituted as the effluent toxicity limitation.

Arbacia is a species for which conventional statistically based hypothesis testing alone typically fails to provide biologically meaningful results with respect to identifying toxicity for the purposes of permit compliance reporting. The problem stems largely from the very low variability in the control test fertilization responses. Because of this low variability, a very small difference between test dilutions and controls may be found to be statistically significant and interpreted as "toxic", when instead the results may lie within the range of the normal biological variability that is considered to be acceptable for the control replicates.

EPA (1991) and other subsequent EPA documents that address statistical variability, WET test analysis methodology, and NPDES compliance reporting provide insight and interpretive guidance that support a broader and more flexible evaluation of *Arbacia* WET test results than relying only on statistical hypothesis testing. In fact, EPA WET test evaluation guidance (EPA, 1991; EPA 2000a, 2000b) consistently recommends point estimation methods in preference to statistical hypothesis testing.

High Number of False Positive Results

The NOEC is based on whether there is a statistical difference between the measured effects in control and experimental populations of a single treatment or concentration. In the case of *Arbacia*, the measured effect is fertilization success. The test protocols employ high sperm-to-egg ratios and thereby often result in extremely low variability within the control population replicates. Thus, if there is even a very small change (for example, a statistically significant change of 1 percent) in the effects measurement in the experimental dilution series, the test is counted as a failure to comply with criteria when the TUc is calculated. This is true though the basis for "success" within the control population testing includes fertilization rates as low as 70 percent per the EPA protocol for this species.

In other words, a 1 percent difference between the test series fertilization success and that of the control population can result in a "failure" though a lack of fertilization in up to 30 percent of the control population is rated as a "success." The result is often a false positive that indicates "toxicity" according to the evaluation protocol –though there is not a biologically meaningful result. Therefore, effluent treatments with a fertilization success rate that is considered valid for a control can be determined to fail if significantly different from the control.

This is not a desired result and the NOEC should not be used to evaluate test results when the control population variability is low. Instead, EPA and various state guidance points to use of the IC_{25}, or some other measure of biological significance that represents a point estimate along an established dose-response curve to indicate where biologically meaningful toxicity effects begin to occur within the test series dilutions. For this and other reasons, EPA and numerous state regulatory agencies guidance documents (for example, EPA, 1991; EPA, 2000a; Oregon Department of Environmental Quality [ODEQ], 2005; New Jersey Administrative Code [NJAC] 7:14A-13.14; New York Department of Environmental Conservation [NYDEC], 2007; Washington Department of Ecology [WDEC], 2008) support the use of point estimates (such as the IC_{25}) for determining effluent toxicity rather than the NOEC.

Lack of PMSD Upper and Lower Bounds for *Arbacia*

Another consideration for dealing with the ability to detect small differences between treatments and controls has been adopted by EPA (2002a; 2002b). The PMSD "represents the smallest difference between the control mean and a treatment mean that leads to the statistical rejection of the null hypothesis (i.e., no toxicity)..." (EPA, 2000a). EPA (2000a) recommends that regulatory authorities implement both the lower and upper PMSD bound approach to minimize within-test variability when using hypothesis testing approaches to report a NOEC. Consideration of the lower PMSD when determining a hypothesis test result (for example, NOEC or LOEC) bound also helps to avoid penalizing laboratories that achieve unusually high precision. Lower PMSD bounds represent a practical limit to the sensitivity of the test method that few laboratories are able to achieve, and below which NOECs or LOECs are not considered toxic (that is, significantly different from the control). For example, the lower bound of the PMSDs established via the EPA interlaboratory testing program was 11 percent for inland silverside minnows (*Menidia beryllina*) and mysid shrimp (*Mysidopsis bahia*). Therefore, using this evaluation technique, 10 percent reductions in survival or growth rates are not considered significant, regardless of the statistical results.

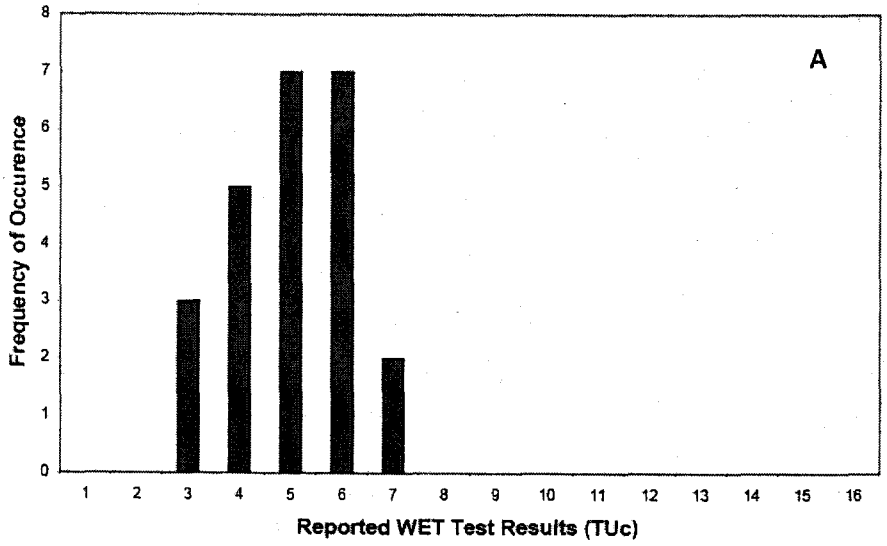
PMSD upper and lower bounds were not established by EPA for *Arbacia*. However, it is reasonable to preserve the intent of this EPA methodology by adopting a PMSD equal to the most sensitive value determined for another invertebrate WET test species, such as the 11 percent lower PMSD for mysid shrimp when evaluating *Arbacia* WET test data.

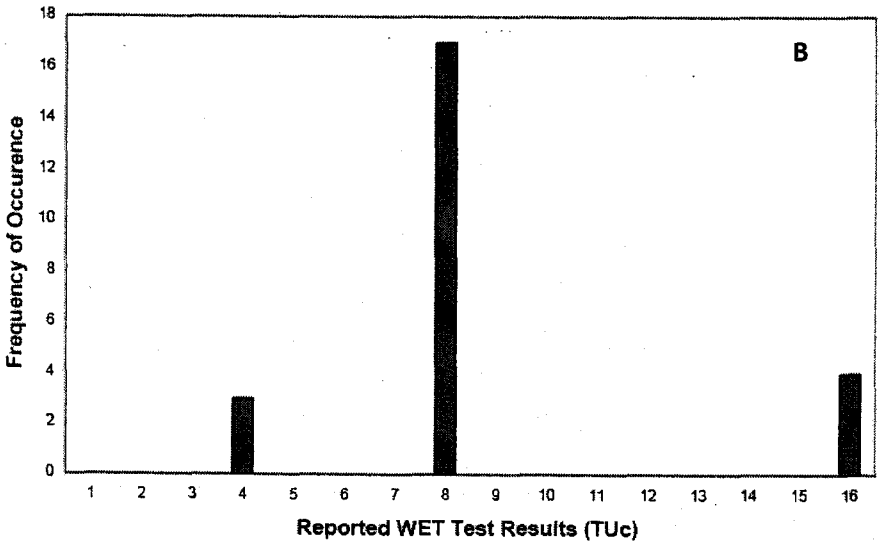
Variability Among Tested Concentrations

Calculating hypothesis testing endpoints, such as the NOEC, uses only data from a single tested concentration when making comparisons to controls. The variability among tested concentrations is not considered. In contrast, point estimates of toxicity, such as the IC_{25}, use all of the concentration-response data and provide a more robust estimate of toxicity. Exhibit 1 illustrates similar IC_{25} and NOEC effects for chronic WET tests with the fathead minnow (*Pimephales promelas*) where the incidences of false positive toxicity were lower for the IC_{25}. It could be argued that using the IC_{25} generally produces a "better" (less toxic) result. However, this is largely because of a more rigorous identification of the point at which there is no "toxicity" using the point estimate than is available using the NOEC approach. The availability of multiple endpoints in toxicity testing demonstrates that there is no single result in WET testing, and that the selection of a robust measure that meets both regulatory and discharger needs can be achieved.

Several fundamental problems have been identified by EPA (EPA, 2000b) in the use of hypothesis testing endpoints when interpreting WET test results. Both the NOEC and IC_{25} toxicity endpoints are based on statistical models that assume a monotonic concentration-response (that is, a steadily increasing effect as effluent concentrations increase). This is not always the case for toxicity test results, and these endpoints can be compromised under different conditions. EPA guidance provides ten examples of possible concentration-response curves and their appropriate interpretation (EPA, 2000b).

EXHIBIT 1
Fathead Minnow Growth in IC_{25} in EPA's Effluent Sample using (A) the IC_{25} and (B) the NOEC Toxicity Endpoints
Source: National Association of Clean Water Agencies (NACWA), 2006





The ten possible concentration-response curves are:

1. Ideal concentration-response relationship
2. All or nothing response
3. Stimulatory response at low concentrations and detrimental effects at higher concentrations
4. Stimulation at low concentrations but no significant effect at higher concentrations
5. Interrupted concentration-response: significant effect bracketed by non-significant effects
6. Interrupted concentration-response: non-significant effects bracketed by significant effects
7. Significant effects only at highest concentration
8. Significant effects at all test concentrations but flat concentration-response curve
9. Significant effects at all test concentrations with a sloped concentration-response curve
10. Inverse concentration-response relationship

Four of these examples (5, 6, 8, and 9) yield results where the NOEC should be interpreted with caution or is considered severely compromised. In these cases, it is important to follow the guidance to determine the validity of the test and toxicity endpoints. In contrast, only two examples (examples 2 and 8) are not necessarily valid for point estimates and should be similarly interpreted with caution.

Criticisms of the use of the NOEC continue in the scientific literature. An evaluation of toxicity from a pulp mill effluent (Chapman et al., 1996) determined by 50 percent effective (EC_{50}) concentrations and NOECs provided the following conclusions:

- The NOEC is not a good estimate of the no-effect concentration.
- The NOEC is highly variable between tests and can lead to contradictory results.
- EC_{50} or other point estimates are more consistent, more reliable, and have less variable estimates than NOECs and can be compared between tests.

Moore and Caux (1997) also compared results obtained via exposure-response curves and comparison of treatment groups to determine No Observed Effect Levels (NOELs) for the same data sets. They demonstrated that NOELs typically correspond to point estimates of toxicity ranging from the EC_{10} to the EC_{30}. The concentration-response curves also demonstrated a more transparent and accurate estimate of no-effect exposures, and the confidence intervals provided a measure of the uncertainty of the estimates. In other words, estimates of the no-effect concentration can be better estimated using point estimates than the statistical hypotheses testing employed by the NOEC.

Landis and Chapman (2011) noted that point estimates of toxicity have no standard error or deviation and have no context of effects at higher or lower exposures. However, point estimates of toxicity do express statistical variability and uncertainty of the data, and provide information on the slope of the response that is much more useful for interpreting concentration-response relationships than do NOEC estimates.

Lack of Demonstrable Reason for EPA to have Changed from the IC_{25}

After considerable previous correspondence with EPA and EQB concerning the appropriate compliance measure to use for *Arbacia*, in a June 22, 2007, meeting between Bacardí and EPA staff and their respective attorneys, it was agreed that the IC_{25} calculation would be applied to flow-proportionally blended samples from the Bacardí, Bayamón, and Puerto Nuevo effluent streams to assess chronic effluent toxicity compliance for *Arbacia* per the conditions of the upcoming permit. However, EPA indicated that it would also require bioassays in each of the individual effluent streams, and that toxicity identification evaluation/toxicity reduction evaluation (TIE/TRE)

action would be initiated if the blended compliance sample failed and the individual samples demonstrated unacceptable toxicity.

All WET test results reported to EPA and EQB by Bacardi and PRASA under the conditions of the previous permits for the three facilities include a presentation of NOEC, LOEC, IC_{25}, and TUc calculations. Exhibit 2 is an excerpt from an August 2011 quarterly test report summarizing the WET test results.

EXHIBIT 2
Typical Chronic WET Test Results Summary

Chronic Definitive Bioassays Using the Sea Urchin (*Arbacia punctulata*)

Introduction

Hydrosphere Research[1] conducted chronic definitive whole effluent toxicity (WET) tests using the sea urchin (*Arbacia punctulata*) for the Bacardi Corporation wastewater treatment plant (WWTP) as well as for the Puerto Rico Aqueduct and Sewer Authority (PRASA) Bayamón and Puerto Nuevo Regional WWTPs. The tests were conducted on samples from each facility individually and also on a salinity-adjusted, flow-proportioned composite sample from the effluent of each of the three plants. The tests were conducted on August 25, 2011.

Summary of Test Results

Exhibit 1 summarizes the test results. Test data and further discussion are provided in the Results and Discussion section.

EXHIBIT 1
Summary of Chronic Test Results

Species	Sample ID	NOEC	LOEC	IC_{25}	TUc
Arbacia punctulata	Combined discharge	27.0%	9.0%	35.9%	2.79
Arbacia punctulata	Bacardí WWTP	0.27%	0.09%	1.11%	90.1
Arbacia punctulata	Bayamón RWWTP	8.1%	0.09%	>24.3%	<4.12
Arbacia punctulata	Puerto Nuevo RWWTP	48.6%	>48.6%	>48.6%	<2.06

Notes:
NOEC=no observed effect concentration
LOEC=lowest observed effect concentration
IC_{25}=inhibition concentration (estimate of the concentration that would cause a 25-percent reduction in test organism growth or fecundity)
TUc=toxic unit chronic (100%/IC_{25})

The current National Pollutant Discharge Elimination System (NPDES) permits for the Bayamón, Puerto Nuevo, and Bacardí wastewater treatment plants stipulate that **"No single IC25 test result for any species or effect in the combined discharge shall be less than 1.00%."** The combined discharge clearly meets that condition for this series of bioassay tests.

Methods and Materials

Test Methods

All chronic tests were performed according to: *Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Marine and Estuarine Organisms*, Third Edition (2002); EPA 821-R-02-014.

Additional guidance was provided by:

- *Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications Under the National Pollutant Discharge Elimination System Program*, (EPA June 2000), EPA 833-R-00-003.
- *Method Guidance and Recommendations for Whole Effluent Toxicity (WET) Testing* (40 CFR Part 136), (EPA July 2000), EPA 821-B-00-004.

[1]The results and methodology from the Hydrosphere Research sea urchin test are summarized in this report. However, more details from this testing are provided in the Hydrosphere Research report (see Appendix A).

GJV31031158104.4.DOCX/112630001
WBG092011062050DFB

1

To date, there have been no failures of the combined effluent to meet chronic toxicity criteria for *Arbacia* when applying the IC_{25} and using the appropriate TUc of 102. Further, after more than a decade of intensive and extensive sampling around the joint outfall, it has been clearly and repeatedly demonstrated that there are no effects associated with this discharge in the vicinity of the outfall on:

- Fish or benthic invertebrate populations
- Water column concentrations of toxic organic or inorganic constituents
- Fish tissue accumulations of toxic organic or inorganic constituents
- Sediment accumulations of toxic organic or inorganic constituents
- Phytoplankton concentrations

In other words, in more than 10 years of monitoring, there have been no measurable ecological responses to this discharge, which further validates the toxicity conclusions based on the biologically-relevant IC_{25} evaluation of effluent toxicity to the most sensitive stage in an organism's life cycle (that is, fertilization).

Nonetheless, on September 27, 2011, Bacardi and PRASA received renewed final NPDES permits for all three facilities.[3] In each permit, the effluent limitation for toxicity was changed to read as follows:

> b. Effluent Limitation:
> **No test result for any species or effect in the combined discharge shall be greater than 83.32 TUc.**

The 2011 permit is not clear with respect to why 83.32 was used as the TUc limit when the critical initial dilution (CID) for this outfall supports a TUc of 102. Further, although the specific measure to be applied to determine compliance with WET limitations is not stated for any of the test species, subsequent conversations with the EPA permit writer have indicated that, under the new permits, the NOEC rather than the IC_{25} will be used to evaluate compliance with effluent toxicity criteria (O'Brien, 2011).

It is not clear why EPA changed both the effluent limitation for TUc and the compliance evaluation protocol, but in the process, it apparently failed to take into account—or even acknowledge—the following:

- That a mixing zone for effluent toxicity can be granted under the PRWQSR based on the numerical results of WET test calculations and the verified critical initial dilution (CID)[4] of the effluent as it rises through the water column. Using this approach, and the data and reports available to EPA, the TUc should be stated as 102, not 83.32.
- That the results of the previous white paper (CH2M HILL, 2007) clearly indicate the problems associated with applying the NOEC to the *Arbacia* to calculate a TUc.
- That there is specific EPA and state guidance that recommends using a point estimate such as the IC_{25} to evaluate toxicity where the NOEC fails to adequately address biologically significant responses to potential toxicants.
- The agreement reached in the June 22, 2007, meeting on the basis of the arguments put forward in the previous white paper that the IC_{25} is a more appropriate criterion by which to evaluate effluent chronic effluent toxicity for *Arbacia* than is the NOEC.

Application of the NOEC will Result in Excessive and Unnecessary Effort

Under the effluent toxicity limitations provided in the 2008 NPDES permits for the Bacardi and PRASA facilities, application of the IC_{25} for compliance evaluations led to the conclusion that the toxicity of the combined effluent consistently complied with PRWQSR criteria for all three test species. There were only two toxicity "failures"

[3] with EDPs of December 1, 2011

[4] The CID is the lowest postulated initial dilution based on very conservative model inputs. Actual field-verified initial dilutions generally exceed the CID by a factor of at least 2 or 3.

under the IC_{25} evaluation protocol, the most recent being in May 2007 (Exhibit 3). None of the test failures was a consequence of *Arbacia* test results. However, as discussed above, applying the NOEC to evaluate the effluent toxicity compliance would have resulted result in false positive results and reporting of unacceptable toxicity where there were actually no meaningful biological responses to the effluent concentrations tested.

This was pointed out in the following tables (excerpted from the Bacardi comments on the July 2011 draft NPDES permit, Appendix B), in which it is clear that using the inappropriate NOEC as the chronic toxicity compliance measure and a TUc of 83.32 would have resulted in "failures" in many of the samples tested, all of which would have been related to *Arbacia* test results.

EXHIBIT 3
Bioassay Test Results for the Bayamón/Puerto Nuevo/Bacardí Flow-weighted Effluent Composite
Comments on the Draft NPDES Permit WEL Limitation for the Bacardi WWTS

Date	Organism	Percent Effluent	
		Chronic NOEC	Chronic IC_{25}
September 2005	*Mysidopsis bahia*	6.25	0.68
	Cyprinodon variegatus	25	40.50
	Arbacia punctulata	Organism Not Available	N/A
February 2006	*Mysidopsis bahia*	6.25	3.04
	Cyprinodon variegatus	25	29.2
	Arbacia punctulata	Not definitive	7.25
March 2006	*Mysidopsis bahia*	3.13	2.72
	Cyprinodon variegatus	25	51.8
	Arbacia punctulata	6	7.31
April 2006	*Mysidopsis bahia*	12.5	13.1
	Cyprinodon variegatus	25	34
	Arbacia punctulata	3	5
September 2006	*Mysidopsis bahia*	12.5	20
	Cyprinodon variegatus	50	59.6
	Arbacia punctulata	<0.78	1.68
November 2006	*Mysidopsis bahia*	6.25	8.6
	Cyprinodon variegatus	50	56.3
	Arbacia punctulata (Nov 4)	<0.78	1.7
	Arbacia punctulata (Nov 7	1.56	4
April 2007	*Mysidopsis bahia*	10.7	2.96
	Cyprinodon variegatus	10.7	30.3
	Arbacia punctulata (Apr 17)	0.29	3.09
	Arbacia punctulata (Apr 19)	<0.09	2.12
	Arbacia punctulata (Apr 21)	<0.09	4.47
May 2007	*Mysidopsis bahia*	Not definitive	0.49
	Cyprinodon variegatus	10.7	18.1
	Arbacia punctulata (May 1)	0.09	4.92
	Arbacia punctulata (May 3)	0.96	14.8
	Arbacia punctulata (May 5)	0.032	14.4

EXHIBIT 3
Bioassay Test Results for the Bayamón/Puerto Nuevo/Bacardí Flow-weighted Effluent Composite
Comments on the Draft NPDES Permit WEL Limitation for the Bacardi WWTS

Date	Organism	Percent Effluent	
		Chronic NOEC	Chronic IC_{25}
May 2007	*Mysidopsis bahia*	10.7	17.9
	Cyprinodon variegatus	10.7	18.2
	Arbacia punctulata (May 15)	0.09	4.88
	Arbacia punctulata (May 17)	0.96	3.01
	Arbacia punctulata (May 19)	0.29	5.23
May/June 2007	*Mysidopsis bahia*	10.7	0.21
	Cyprinodon variegatus	10.7	24.2
	Arbacia punctulata (May 31)	3.2	5.91
September 2008	*Mysidopsis bahia*	8.00	7.20
	Cyprinodon variegatus	16.0	>16.0
	Arbacia punctulata	0.96	4.15
December 2008	*Arbacia punctulata*	3.20	5.57
February 2009	*Arbacia punctulata*	9.00	13.5
June 2009	*Arbacia punctulata*	3.00	9.51
August 2009	*Arbacia punctulata*	1.00	4.34
November 2009	*Mysidopsis bahia*	16.0	14.5
	Cyprinodon variegatus	16.0	>16.0
	Arbacia punctulata	3.00	4.31
March 2010	*Arbacia punctulata*	3.00	4.68
May 2010	*Arbacia punctulata*	9.00	13.96
September 2010	*Arbacia punctulata*	3.00	12.9
November 2010	*Mysidopsis bahia*	16.0	>16.0
	Cyprinodon variegatus	16.0	.16.0
	Arbacia punctulata	1.00	13.4
March 2011	*Arbacia punctulata*	9.00	13.9
May 2011	*Arbacia punctulata*	3.00	5.25

Note:
Shaded entries indicate IC_{25} <1.2% effluent.

There is no controversy with respect to use of *Arbacia* as a test organism (it is on the EPA-approved list of sensitive species to be used for WET testing) or to the protocols under which the tests are conducted (which are the most recent published EPA protocols). The purpose of this update is to provide additional information relevant to the question of the most appropriate WET test statistic to apply to evaluate compliance of the effluent with the numerical criteria provided by the PRWQSR.

Basis for Replacing NOEC-based Effluent Toxicity Limitation with IC_{25}-based Effluent Toxicity Limitation

The WQCs that were incorporated in each of the 2011 NPDES permits require acute and chronic WET tests (bioassays) using the sheepshead minnow (*Cyprinodon variegatus*) and a mysid shrimp (*Mysidopsis bahia*), as well

as chronic toxicity tests for the sea urchin *Arbacia punctulata*[5] using EPA- and EQB-approved WET test protocols. Three levels of testing are performed:

1. Annual tests on all three species on flow-proportional combined effluent
2. Annual tests on all three species on individual effluents from all three facilities
3. Quarterly tests on *Arbacia* only on individual effluents from all three facilities

The results of the annual flow-proportional combined effluent tests are used to evaluate compliance with NPDES permit requirements. The results from the other two test series are used to help identify the causative facility if there are failures in the combined effluent tests and serve as the basis for initiating the TIE/TRE process. Per the combined effluent compliance test protocols, WET compliance evaluations are made on the basis of blended effluent from the three facilities. The tests are performed on flow-proportional 24-hr composite samples taken from the three effluents. More specifically, a 24-hr composite sample is obtained from each facility, with flow during the compositing period at each plant recorded. Next, the three effluent samples are sent to the bioassay laboratory with instructions about how to combine the samples in a proportional fashion based on these flows. Finally, flow-proportional composite samples are then used for WET testing and data evaluation.

The flow-proportional composite approach allows for an evaluation of whatever synergisms and/or antagonisms may be present in the three effluents in relation to the relative toxicity of the mixed effluent that is ultimately discharged to the marine environment. The WET test results are used to evaluate whether receiving water toxicity requirements will be met at the edge of the small permitted mixing zone that is established around the outfall diffuser. All test results are reported in terms of statistical hypothesis testing (NOEC and LOEC) and as point estimates (IC_{25}).

In brief, the hypothesis-testing method proposed for compliance evaluation in the 2011 NPDES permits relies on a NOEC that is based on the statistical difference in variances between the control and test populations of the organisms tested for each of the tested effluent concentrations (or treatments). Therefore, the NOEC does not consider the entire dose-response relationship. The IC_{25} point estimate method uses the entire WET test data set to estimate (through interpolation) a sub-lethal biological response endpoint on a dose-response curve. Thus, the two methods may result in numerically different estimates of chronic endpoints and the IC_{25} produces a more meaningful estimate of actual biological responses.

An evaluation of published EPA guidance for WET test data interpretation indicates that the point estimation technique is preferred for purposes of regulatory compliance evaluations. This white paper is intended to clarify the most appropriate method to use for interpreting *Arbacia* test results with respect to both past and future WET test data obtained from the Bacardí, Bayamón, and Puerto Nuevo wastewater treatment plants. It discusses how NOECs derived from hypothesis testing frequently lead to "false positive" toxicity indications, summarizes key issues, presents case-specific data with respect to WET test findings and conclusions, questions whether statistical hypothesis testing should be used to evaluate the results of the chronic definitive bioassays conducted using *Arbacia*, and offers recommendations for what are considered to be appropriate WET test data evaluation methods when using *Arbacia* as a test organism.

[5]There are no acute toxicity protocols for *Arbacia punctulata*.

Alternative EPA-Approved *Arbacia* WET Test Data Evaluation Methods

As noted above, using statistical hypothesis testing to evaluate *Arbacia* WET test data is liable to introduce Type I (false positive) errors because the percent fertilization variance within the control group replicates is normally very small. Thus, even a very small difference between the control group replicates and the effluent test group replicates would be calculated as statistically different from the variance for the control group, indicating an "effect" that is interpreted as "toxicity." This can either make it difficult to define a NOEC (as in the indeterminate <0.78 values in Exhibit 4) or may define a NOEC at an artificially low concentration that results in reported false positives for toxicity, and possibly erroneous findings of noncompliance with NPDES permit limits that in turn trigger unnecessary advanced testing and TIE/TRE investigations.

EPA has carefully addressed these and other issues related to toxicological data interpretation in several of its guidance documents. For example, in its 1991 *Technical Support Document for Water Quality-Based Toxics Control* (TSD; USEPA 1991), EPA compared results from hypothesis testing and point estimate endpoints such as the IC_{25} and concluded that:

> "Comparisons of both types of data indicate that a NOEC derived using the IC_{25} is the approximate analogue of a NOEC derived using hypothesis testing. For the above reasons, if possible, **the IC_{25} is the preferred statistical method for determining the NOEC.**" (emphasis added)

Moreover, EPA (2000a) specifically addresses effluent toxicity variability and states the following (on p. 6-4):

> "EPA recommends that point estimates be used to estimate effluent variability, to determine the need for limits, and to set permit limits. This is recommended whether the self-monitoring test results will be determined using hypothesis tests or point estimates. Point estimates have less analytical variability than NOECs using current experimental designs... **Point estimates make the best use of the whole effluent toxicity (WET) test data for purposes of estimating the coefficient of variation, long term average, and relative percent factors and calculating the permit limit.**" (emphasis added)

An EPA sponsored review committee was formed several years ago to assess this issue. The committee found that in the case of a species with low control variability, such as that exhibited by *Arbacia*, using only the NOEC-derived from statistical hypothesis testing is problematic and may not be an effective approach for monitoring toxicity compliance and reporting. As a result of these issues, EPA Region 1 modified the hypothesis testing approach to include the species test acceptability criteria (TAC) for determining permit compliance. This approach provides a more biologically relevant reporting endpoint for compliance evaluation.[6] The basis of the biological significance evaluation is that the TAC for control fertilization rate (>70% fertilization) is applied *in combination with* the statistical hypothesis testing results to determine the "biologically significant" effects concentrations (as opposed to only statistically-derived effects concentrations).

ODEQ (2005) relied on the EPA TSD in stating that "the concentration of effluent that results in 25% inhibition of the parameter used to determine a chronic effect (e.g., growth) be the benchmark for determining whether the results of chronic tests indicate toxicity." Oregon similarly promotes the reliability of point estimate techniques, which "are the preferred statistical methods in calculating endpoints for effluent toxicity tests" for chronic toxicity. This is the result of the inability to estimate test precision when using hypothesis testing for deriving a NOEC, thus making the inherent variability of bioassay results difficult to address statistically. EPA compared data from the point estimate and NOEC approaches and concluded that the "IC_{25} is approximately the analogue of an NOEC derived using hypothesis testing."(EPA, 1991).

[6]Documentation is provided at the following web page (http://www.epa.gov/region1/npdes/epa_attach.html) under the link **Marine Chronic Test Procedure and Protocol.**

EXHIBIT 4

Summary of *Arbacia* Bioassay Tests Conducted to Date with Combined Bacardí/Bayamón/Puerto Nuevo Effluent, showing Comparison of NOECs with IC_{25} Point Estimates of Chronic Toxicity

Test Date	Control % Fertilization	Hypothesis-based NOEC	% Fertilization	IC25	% Fertilization	Biologically-based NOEC	% Fertilization	Biologically-based LOEC	% Fertilization
8/29/06	90.3	<0.78	68.8	1.68	67.7	<0.78	N/A	0.78	68.8
11/7/06	97.8	<0.78	91.1	1.67	73.3	1.56	77.8	3.13	18.4
11/9/06	95.1	1.56	88.6	3.97	71.3	3.13	79.9	6.25	48.9
4/17/07	94.0	0.29	93.0	3.09	70.5	3.2	70.2	10.7	8.6
4/19/07	95.2	<0.09	90.8	2.12	71.6	0.96	77.9	3.2	65.4
4/21/07	96.3	<0.09	91.5	4.47	72.3	3.2	85.4	10.7	7.4
5/1/07	94.3	0.09	91.1	4.92	70.7	3.2	87.5	10.7	14.2
5/3/07	92.8	0.96	92.4	14.8	69.6	10.7	83.3	35.5	1.0
5/5/07	90.5	3.2	87.8	14.4	67.9	3.2	87.8	10.7	79.0
5/15/07	93.3	0.09	89.8	4.88	70	3.2	83.0	10.7	24.8
5/17/07	91.3	0.96	85.5	3.01	68.5	0.96	85.5	3.2	67.1
5/19/07	93.8	0.29	92	5.23	70.4	3.2	88.5	10.7	21.5
9/9/08	95.8	0.96	98.0	4.15	73.5	3.2	82.5	10.7	7.5
12/2/08	93.7	3.2	87.8	5.57	70.3	3.2	87.8	10.7	38.3
2/12/09	95.3	9.0	87.8	7.41	71.5	9.0	87.8	27.0	22.5
6/23/09	96.8	3.0	96.0	10.5	72.6	9.0	75.2	27.0	18.2
8/18/09	99.5	1.0	99.3	4.34	74.6	3.0	94.0	9.0	7.5
11/10/09	99.0	3.0	99.0	4.31	74.3	3.0	99.0	9.0	20.8
3/2/10	81.7	3.0	92.0	4.68	61.3	3.0	92.0	9.0	16.0
5/11/10	85.8	9.0	98.3	14.0	64.4	9.0	98.3	27.0	8.8
9/9/10	96.8	3.0	95.8	12.9	72.6	3.0	95.8	27.0	39.5
11/9/10	90.0	1.0	88.0	13.4	67.5	9.0	82.5	27.0	21.2
3/8/11	99.0	9.0	98.5	13.9	74.3	9.0	98.5	27.0	9.8
5/14/11	98.5	3.0	99.2	5.25	73.9	3.0	99.2	9.0	33.0
8/25/11	98.2	3.0	90.5	35.9	73.7	27.0	90.8	81.0	19.0

Notes:

Mean Control Fertilization = 94.2%

Shaded cells exceed the NOEC limit of <0.96 percent effluent.

Other jurisdictions have adopted the IC_{25} as the basis of effluent limitations, including New Jersey (N.J.A.C. 7:14A-13.14) and New York (NYDEC, 2007), or similarly adjusted their approach to prevent incorrectly identifying an effluent sample as toxic. WDEC (2008), for example, controls Type 1 errors (false positives) in hypothesis testing when differences in test organisms response are small. Specifically, alpha (the maximum Type 1 error rate) "will be lowered from 0.05 to 0.01 if a 10% difference in an acute test is significant or a 20% difference in a chronic test is significant."

For its part, the PRWQSR defines chronic toxicity testing and evaluation as follows:

> Chronic Bioassay
> Toxicity test designed to determine if the response to a stimulus such as, a total effluent, a specific substances, or combination of these has sufficient severity to induce a long-term effect that could linger for up to one-tenth of the life span of the organism. A chronic effect could be lethality, growth rate reduction, reproduction rate reduction, etc. A chronic bioassay shall be performed according to procedures described in "Mixing Zone and Bioassay Guidelines", approved by the Board.
>
> Chronic Effect
> Organism response to a stimulus, detected during a chronic bioassay that comprises a stimulus that lingers or continues for a relatively long period of time, which could be of the order of one-tenth of the life span of the organism used in the test. A chronic effect could imply lethality, growth rate reduction, reduced reproduction rate, etc.
>
> Chronic Toxic Unit
> The reciprocal of the effluent dilution that causes no unacceptable effect on the test organisms by the end of the chronic exposure period, obtained during a chronic bioassay, as defined by the following equation:
>
> $$TUc = \frac{100}{NOEC}$$
>
> (The NOEC value should be expressed in terms of the percent (%) of the effluent in the dilution water).

It is noted that, although the PRWQSR chronic toxicity definition refers to a NOEC, it does not refer to a specific method by which a NOEC is to be obtained. It is further noted that the PRWQSR refers to the *Puerto Rico Mixing Zone and Bioassay Guidelines*, which are defined as follows:

> Technical guidelines developed by the Board which describe procedures, methods, models, techniques and organisms to be used to calculate the initial dilution; perform chronic and acute bioassays; to collect field data, or to establish the natural background concentration value, as required to verify compliance with inherent mixing zone conditions. These Guidelines are based on the following EPA publication: "Technical Support Document for Water Quality Based Toxics Control" and Users Guide to the Conduct and Interpretation of Complex Effluent Toxicity Tests at Estuarine/Marine Sites".[7] The guidelines will be revised, as necessary, in accordance with updated versions of these documents or other documents released by EPA which directly impact the guidelines in effect at the time of publication of the final document.

There are several alternative EPA-approved methods that are available to evaluate compliance with toxicity criteria that do not rely solely on statistical hypothesis testing. These include biological significance evaluation (as described above used by EPA Region 1), IC_{25} point estimate evaluation, and test variability evaluation.[8] Of the three, the first two are in more common use for *Arbacia* fertilization tests. These are simply WET test data evaluation alternatives; **they are not WET test protocol alternatives**. The following subsections discuss these alternative methods. It is noted that these data evaluation alternatives should also be applicable to other Puerto Rico NPDES permits that use *Arbacia* as a test organism.

---

[7] It is noted that the most recent version of the Puerto Rico Mixing Zone and Bioassay Guidelines is a 1989 draft that predates the 2001 EPA Technical Support Document, and that advances in methods and technology in the last 17 years are therefore not reflected in the Guidelines. However, the Guidelines explicitly provide EQB with the ability to approve alternative methods.

[8] Test variability evaluation is discussed in Appendix A to this white paper.

Biological Significance Evaluation

EPA Region 1 has recognized that evaluation of *Arbacia* fertilization tests using statistical hypothesis testing often results in putative statistically-based "toxicity effects" at effluent concentrations that are much lower than likely biological effects. When the fertilization success in the control group replicates varies by only small percentages, a statistically significant difference between the control and a test group could be interpreted as a "toxic" response, without respect to biological significance.

The chronic WET method for *Arbacia* (EPA 2002a) stipulates that fertilization rates for the control group of replicates should be greater than 70 percent. For the purposes of evaluating permit compliance, if a test group yield fertilizations rates greater than 70 percent (that is, are within the range of acceptable control group fertilization), but are shown to be statistically different from the control using hypothesis testing, those test group concentrations *are not* considered different from the control for the purposes of assessing toxicity (that is, they are not biologically significant; see biologically-based NOEC data in Exhibit 4). In a test where that occurs, the NOEC concentration corresponds to the highest test group concentration that has a fertilization rate = greater than or equal to 70 percent, without regard to whether it is statistically different from the control using hypothesis testing.

This combined hypothesis testing/biological significance method for *Arbacia* WET test data compliance evaluation is considered by EPA Region 1 to be a reliable approach and is preferred over the sole use of statistical hypothesis testing. Therefore, it is believed that evaluating the biological significance results for the Bayamón/Puerto Nuevo/Bacardí discharge system WET tests using this approach is a practical and acceptable means by which to evaluate compliance with toxicity criteria for *Arbacia.* This approach could replace statistical hypothesis testing alone as per EPA Region 1 data evaluation protocols.

PMSD

EPA (2002a) states that, "In determining hypothesis test results (for example, NOEC or LOEC), a test concentration shall not be considered toxic (that is, significantly different from the control) if the relative difference from the control is less than the lower PMSD bound". For inland silverside and mysid shrimp, for example, the lower bounds of the PMSDs were 11 percent, and, therefore, 10 percent reductions in survival or growth rates would not be considered significant, regardless of the statistical results.

The ranges of upper and lower PMSDs shown in Exhibit 5 indicate the labs should generally be able to detect a 28 to 47 percent difference from controls and not less than a 9 to 13 percent difference. EPA WET testing guidance (EPA, 2002a and 2002b) states that treatments with significant differences that are smaller than the lower PMSD will not be considered to "fail" compliance evaluations. In other words, even if a statistically significant difference between a treatment and its control is found below the lower PMSD, it will not be considered different from the control for identifying the NOEC.

As noted above, PMSD upper and lower bounds were not established by EPA for *Arbacia*. However, it is reasonable to adopt a PMSD equal to the most sensitive determined for another invertebrate WET test species, such as the 11 percent lower PMSD for mysid shrimp when evaluating *Arbacia* WET test data.

EXHIBIT 5
Variability Criteria (upper and lower PMSD bounds) for Sublethal Hypothesis Testing Endpoints Submitted under NPDES Permits

Test Species	Guidance	Chronic WET Lower PMSD	Chronic WET Upper PMSD
Inland silverside minnow	USEPA 2002a	11%	28%
Mysidopsis bahia	USEPA 2002a	11%	37%
Fathead minnow	USEPA 2002b	12%	30%
Ceriodaphnia dubia	USEPA 2002b	13%	47%
Selenastrum capricornutum	USEPA 2002b	9.1%	29%

IC_{25} Evaluation

Exhibit 4 also shows the IC_{25} point estimates for the *Arbacia* WET tests that have been conducted to date for the Bayamón/Puerto Nuevo/Bacardí discharge system. The IC_{25} is a commonly used, widely accepted point estimation technique that is calculated to estimate chronic toxicity thresholds. The IC_{25} method uses all of the WET test data as opposed to statistical hypothesis testing, which does not. As seen in Exhibit 4, if IC_{25} values were used to evaluate the data, all *Arbacia* chronic WET test results would have met permit compliance requirements of no chronic toxicity at the edge of the mixing zone at concentrations less than the 0.96 percent compliance targets.

In the preamble to its *Final Rule for Guidelines Establishing Test Procedure for the Analysis of Pollutants; Whole Effluent Toxicity Test Methods*, Fed. Reg. 69951-69972 (November 19, 2002), EPA states in two separate discussions:

> "**EPA recommends the use of point estimation techniques over hypothesis testing approaches for calculating endpoints for effluent toxicity tests under NPDES Permitting Program.**" (emphasis added)

(*Id.* at 69957 and 69958.) This statement is reiterated in EPA (2002a). On Page 44, section 9, EPA states:

> "**NOTE: For the NPDES Permit Program, the point estimation techniques are the preferred statistical methods in calculating end points for effluent toxicity tests.**" (emphasis in original).

Therefore, it is believed that evaluating the IC_{25} point estimate for the Bayamón/Puerto Nuevo/Bacardí discharge system bioassays (or other NPDES permits requiring *Arbacia* testing) not only represents a more reliable alternative than the NOEC with which to evaluate permit compliance relating to *Arbacia* test data, it is the preferred method of evaluation based on specific guidance from EPA and a number of state environmental regulatory agencies.

As demonstrated by the EPA interlaboratory variability study (EPA, 2001), the point estimate toxicity endpoints pose lower risk of false positives than the NOEC. A single sample with low toxicity tested at multiple labs had a 28 percent false positive rate[9] relative to the median response with *Ceriodaphnia* chronic results (Exhibit 6). Extrapolating these false positive rates to the Bacardí/Bayamón/Puerto Nuevo monitoring frequency would suggest that one false positive toxicity result may be expected in each species with quarterly testing when using the NOEC. A high rate of false positives would cause unnecessary use of resources for accelerated testing and TIE/TRE evaluations.

Summary of *Arbacia* WET Test Results to Date

For the existing WQCs and NPDES permits for the Bacardí distillery and the Bayamón and Puerto Nuevo RWWTPs, which were issued by EQB and EPA, respectively, in 2008, the CID and compliance TUc are 104, equating to an acceptable NOEC of ≥0.96 percent.

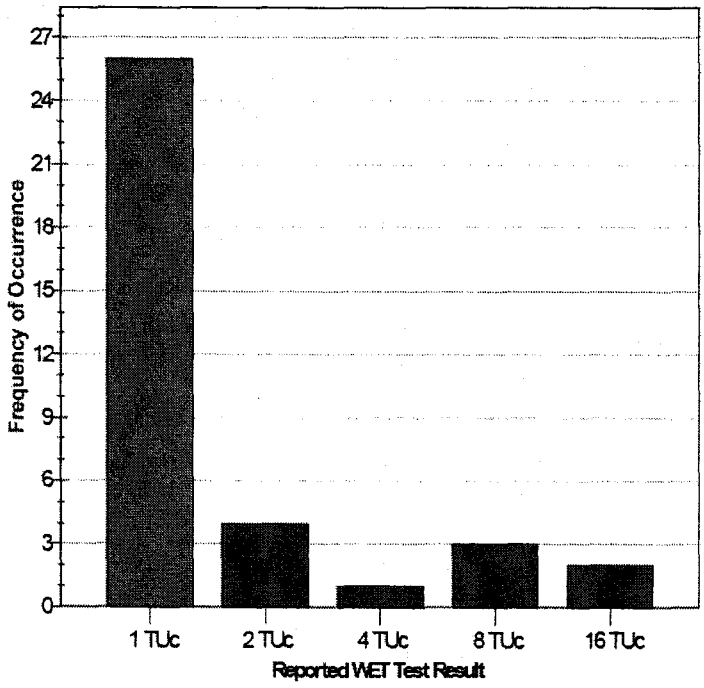
Using the newly-proposed NOEC of ≥1.02 percent for the existing permit and statistical hypothesis testing to assess compliance, there would have been 8 "failures" in the combined effluent testing since August 2006. These NOEC-based toxicity interpretations rely exclusively on statistical hypothesis testing to determine the NOEC (using Bonferroni's T-test), which is directly correlated to the degree of statistical variance in controls. Because this variance may be very small among control replicates, T-test results are purely statistically-based (that is, based on statistical variance alone without respect to actual biological responses) and therefore are prone to "false positive" or Type I errors.

This is shown in Exhibit 4 where 8 of 25 tests exceed the NOEC limit (that is, NOEC <0.96 percent effluent) if evaluated by statistical hypothesis testing, but where using alternative EPA-approved (and preferred) data evaluation techniques (that is, the IC_{25} and biological significance testing) leads to the conclusion that there is no unacceptable toxicity indicated at the compliance TUc (0.96 percent combined effluent concentration).

[9]The false positive rate in this example is defined as the number of tests with a reported result above the central tendency of the data, divided by the total number of tests. USEPA and some state NPDES agencies may use a false positive definition that does not consider all of the data above the point of central tendency to be false positives. However, using Exhibit 2 as an example, the difference between 1 TUc and 2 TUc will typically be significant for a discharger.

EXHIBIT 6

***Ceriodaphnia* Reproduction (NOEC) in USEPA's Reference Toxicant Sample**



28% False Positive

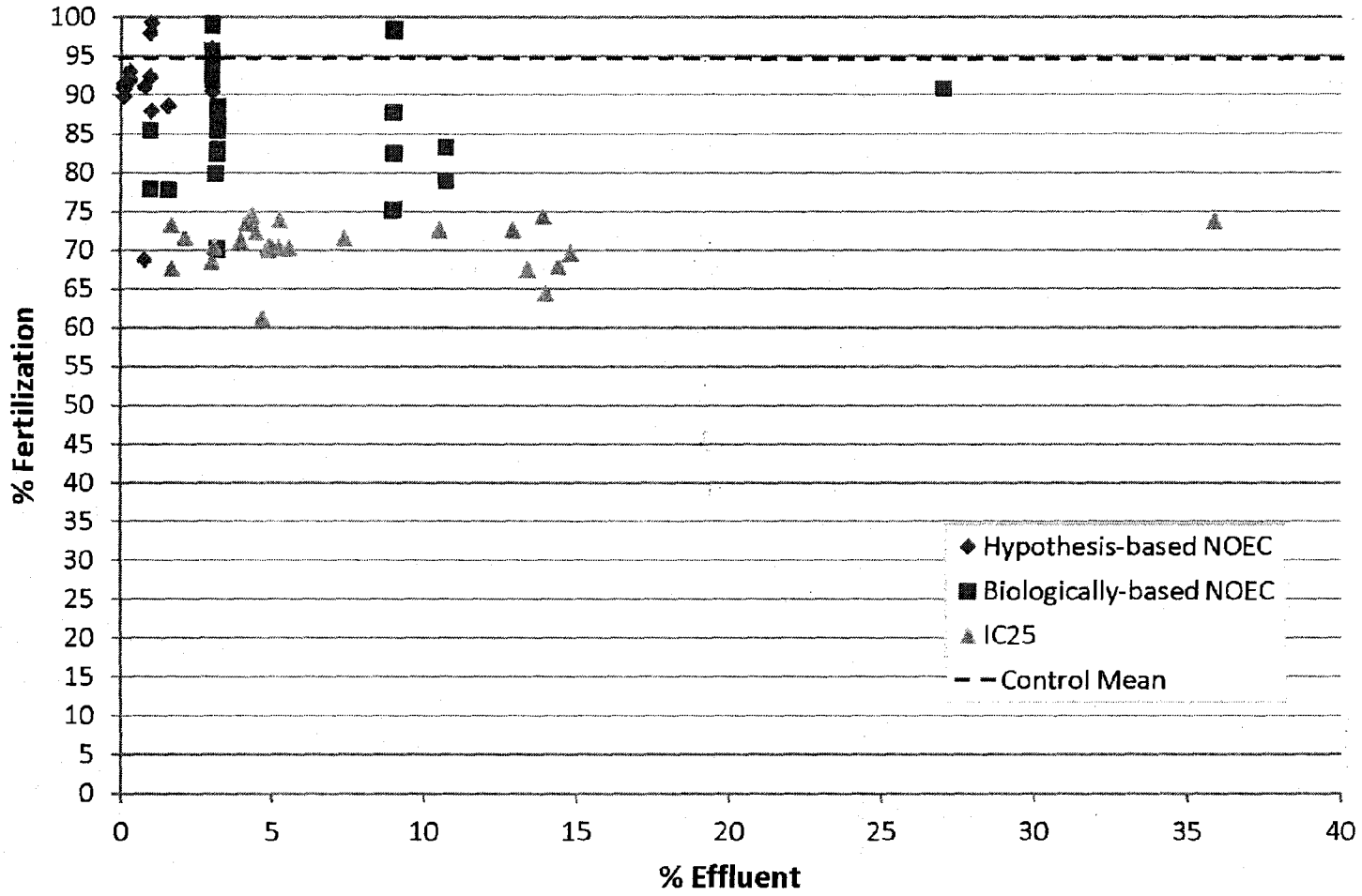
In addition to the hypothesis testing-based NOECs, Exhibit 4 shows biologically-based NOEC values. In these cases, if a test group yields fertilization rates greater than 70 percent (that is, are within the range of acceptable control group fertilization), but are statistically different from the control using hypothesis testing, those test group concentrations *are not* considered different from the control for the purposes of assessing toxicity (that is, they are not biologically significant). These are based on an EPA test acceptability criterion that does not allow for a test to be considered valid if control fertilization rates are less than 70 percent (USEPA, 2002a). A minimum PMSD of 10 percent was also considered as a threshold for significant results.

Exhibit 4 also shows point estimates of chronic toxicity based on the IC_{25}, which is commonly used and widely accepted by EPA and other regulatory agencies (for example, ODEQ, 2005; N.J.A.C. 7:14A-13.14; NYDEC, 2007; WDEC, 2008) as a comparable value of the chronic toxicity threshold. In the case of the biological significance and the IC_{25} toxicity evaluations, all but one of the values are ≥0.96 percent effluent, suggesting that few tests indicate potentially unacceptable levels of toxicity. It is noted once again that unlike NOEC-based compliance evaluations, IC_{25} point estimates allow the use of all of the WET test response data to determine, through linear interpolation, the point at which the toxicity response is equal to the target value (that is, a 25-percent inhibition of fertilization).

Exhibit 7 is a scatter plot showing percent fertilization and percent effluent for hypothesis-based NOECs, IC_{25} point estimates, and biologically-based NOEC values. It is clear that the only data points that appear to indicate non-compliant toxicity (that is, are to the left of the 0.96 percent compliance target for the new WQC) are NOEC values derived from statistical hypothesis testing. Both the IC_{25} point estimates and the biologically-based NOEC data points do not provide evidence of unacceptable (non-compliant) effluent toxicity.

EXHIBIT 7

Percent Fertilization and Percent Effluent for Hypothesis-Based NOECs and IC_{25} Point Estimates for Arbacia WET tests (2006-2011)



Conclusions and Recommendations

Arbacia is a species for which conventional statistically-based hypothesis testing alone typically fails to provide biologically meaningful results with respect to identifying toxicity for the purposes of permit compliance reporting. The problem stems largely from the very low variability in the control test fertilization responses. Because of this low variability, a very small difference between test dilutions and controls may be found to be statistically significant and interpreted as "toxic", when instead the results may lie within the range of the normal biological variability that is considered to be acceptable for the control replicates.

EPA (1991) and other subsequent EPA documents that address statistical variability, WET test analysis methodology, the lack of a PMSD, and NPDES compliance reporting provide insight and interpretive guidance that support a broader and more flexible evaluation of *Arbacia* WET test results than relying only on statistical hypothesis testing. In fact, EPA WET test evaluation guidance consistently recommends point estimation methods in preference to statistical hypothesis testing.

There are clearly problems inherent with using statistical hypothesis testing to evaluate toxicity data from *Arbacia* fertilization tests. EPA provides toxicity test evaluation guidance that explicitly recommends point estimate techniques as preferred alternatives to statistical hypothesis testing. Further, the PRWQSR and the associated Puerto Rico *Mixing Zone and Bioassay Guidelines* provide the flexibility to use alternative, EPA-approved approaches to compliance evaluations as they become available.

It is believed that a review of alternative methods for evaluating *Arbacia* test data and incorporating more appropriate agency-approved methods in new NPDES permits is warranted. Based on the above analysis, it is suggested that Bacardi (and PRASA) request that EPA and EQB consider the following options as the basis for toxicity compliance evaluations for WET tests using *Arbacia*:

1. Use the IC_{25} point estimate methodology as the definitive toxicity evaluation.
2. Adopt the EPA Region 1 test acceptability criterion, using biological significance (that is, the biologically significant NOEC where there is less than 70 percent fertilization in *Arbacia*) in combination with statistical hypothesis testing, and greater than 10 percent difference from control (corresponding to a minimum PMSD of 11 percent).
3. Use both a biological-significance-based NOEC with a minimum PMSD of 11 percent and IC_{25} point estimates to determine effluent toxicity using *Arbacia* data.

Options 1 or 2 are preferred, as they follow clear EPA guidance, and have already proven acceptable to EPA for use in NPDES permits for *Arbacia* WET test evaluation. Option 1 has also proven to be acceptable to EQB, as it was approved for use in the 2008 NPDES permits for the Bacardí, Bayamón, and Puerto Nuevo facilities in light of the flexibility offered by the Puerto Rico *Mixing Zone and Bioassay Guidelines*. However, Option 3 is also acceptable and is consistent with EPA guidance concerning evaluation of acceptable whole effluent toxicity.

Works Cited

CH2M HILL. 2007. *Discussion and Recommendations Related to Arbacia punctulata Whole Effluent Toxicity Testing Using Combined Effluent from the Bayamón, Puerto Nuevo and Bacardí Wastewater Treatment Plants*. Prepared for Bacardi Corporation. May 2007.

Chapman, P.M., R.S. Caldwell, and P.F. Chapman. 1996. Letter to the Editor; A Warning: NOECs are Inappropriate for Regulator Use. *Environ. Toxicol. Chem.* 15: 77-79.

Environmental Quality Board. 2001. Intent to Issue Water Quality Certificates to Define and Authorize a Mixing Zone and Approve Compliance Plans. NPDES Permit No. PR0000591. Authorization to Bacardi Corporation to discharge under the NPDES System.

Federal Register. *Final Rule for Guidelines Establishing Test Procedure for the Analysis of Pollutants; Whole Effluent Toxicity Test Methods*, Fed. Reg. 69951-69972 (November 19, 2002)

Landis, W.G. and P.M. Chapman. 2011. Well Past Time to Stop Using NOELs and LOELs. *Integr. Environ. Assess. Manag.* 7: 6-8.

Moore, D.R.J. and P. Caux. 1997. Estimating low toxic effects. *Environ. Toxicol. Chem.* 16:794–801.

National Association of Clean Water Agencies. 2006. Whole Effluent Toxicity (WET) NPDES Permit Testing and Limitations for Public Agencies White Paper. January 2006.

New York Department of Environmental Conservation. 2007. *Acute and Chronic Toxicity Testing in the SPDES Permit Program. Technical and Operational Guidance Series 1.3.2.* Division of Water.

O'Brien, Karen/EPA. 2011. Personal communication with Julio Torruella/Bacardi. October 14, 2011.

Oregon Department of Environmental Quality. 2005. *Reasonable Potential Analysis for Toxic Pollutants*. Internal Management Directive. September 2005.

Puerto Rico Environmental Quality Board. *Interim Mixing Zone and Bioassay Guidelines*. April 1988.

U.S. Environmental Protection Agency. 2001. Final Report: Interlaboratory Variability Study of EPA Short-Term Chronic and Acute Whole Effluent Toxicity Test Methods, Vol. 1 ("Interlab Variability Study"), EPA 821-B-01-004.

U.S. Environmental Protection Agency. 2000a. *Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications under the National Pollutant Discharge Elimination System.* EPA 833-R-00-003, June.

U.S. Environmental Protection Agency. 2000b. *Method Guidance and Recommendations for Whole Effluent Toxicity (WET) Testing* (40 CFR Part 136). Office of Water. EPA 821-B-00-004. July.

U.S. Environmental Protection Agency. 2002a. *Short-Term Methods For Estimating the Chronic Toxicity of Effluents and Receiving Water to Marine and Estuarine Organisms* (EPA-821-R-02-014, Third Edition).

U.S. Environmental Protection Agency. 2002b. *Short-Term Methods For Estimating the Chronic Toxicity of Effluents and Receiving Water to Freshwater Organisms* (EPA-821-R-02-013, Fourth Edition).

U.S. Environmental Protection Agency. 1991. *Technical Support Document for Water Quality-Based Toxics Control* (TSD), EPA/505/2-90/001. USEPA Office of Water.

Washington Department of Ecology. 2008. *Laboratory Guidance and Whole Effluent Toxicity Test Review Criteria*. Publication No. WQ-R-95-80.

Appendix A
Test Variability Evaluation when using Hypothesis Testing Methods

APPENDIX A

Test Variability Evaluation when using Hypothesis Testing Methods

In the Preamble to its Final Rule, 67 Fed. Reg. at 69968, EPA states:

> "....to reduce the within-test variability and to increase statistical sensitivity when test endpoints are expressed using hypothesis testing rather than the preferred point estimation techniques, variability criteria must be applied as a test review step when NPDES permits require sublethal hypothesis testing endpoints (i.e., NOEC or LOEC) and the effluent has been determined to have no toxicity at the permitted receiving water concentration."

(67 Fed. Reg. at 69967 (emphasis added).) For tests for which in-test variability assessment is required, EPA defines this variability term as the percent minimum significant difference (PMSD). The Preamble to the EPA Final Rule states:

> "Within-test variability, measured as the percent minimum significant difference (PMSD), must be calculated and compared to upper bounds established for test PMSDs. Under this new requirement, tests conducted under NPDES permits that fail to meet the variability criteria (i.e., PMSD upper bound) and show "no toxicity" at the permitted receiving water concentration (i.e., no significant difference from the control at the receiving water concentration or above) are considered invalid and must be repeated on a newly collected sample."

(*Id.*) The EPA Final Rule did not include specific language requiring mandatory application of variability criteria for *Arbacia* fertilization tests, although a number of species with similar control test variability characteristics were defined. The Preamble to that Final Rule indicates that for the chronic methods that were not evaluated in the WET Interlaboratory Variability Study, EPA does not have sufficient data to support the implementation of mandatory variability criteria at this time.

Important to the issue of test variability, especially in the case of the *Arbacia* fertilization tests, are the following statements by EPA in the Preamble to the Final Rule:

> "Lower bounds on the PMSD are also applied, such that test concentrations shall not be considered toxic (i.e., significantly different from the control) if the relative difference from the control is less than the lower PMSD bound."

(*Id.* at 69957.) and

> "According to the proposed approach, any test treatment with a percentage difference from the control (i.e., [mean control response-- mean treatment response]/ mean control response * 100) that is greater than the upper PMSD bound would be considered as significantly different; and any test treatment with a percentage difference from the control that is less than the lower PMSD bound would not be considered as significantly different."

(*Id.* at 69958.)

Because EPA, at the time of issuing its Final Rule, did not have sufficient data from an Interlaboratory Variability Study to develop variability criteria and PMSD bounds for the *Arbacia* fertilization test, there are no existing criteria with which to examine test variability. While test variability might prove to be an acceptable WET test data evaluation option for *Arbacia*, using it would require constructing a database that is not currently available. It is not believed that this approach is compatible with the current Bacardí and PRASA permit renewal schedules and it is further noted that there are other EPA-approved alternatives that are both appropriate and already in use for NPDES permit toxicity compliance evaluations for *Arbacia*.