



DEPARTMENT OF THE NAVY
NAVAL RESEARCH LABORATORY
4555 OVERLOOK AVE SW
WASHINGTON DC 20375-5320

IN REPLY REFER TO
3900
Ser 6130/1622
70 FEB 2016

From: Commanding Officer, Naval Research Laboratory
To: Commandant, United States Coast Guard (CG-OES-3 R. Bergner)

Subj: REVIEW OF A REQUEST FOR APPROVAL OF AN ALTERNATIVE METHOD FOR BALLAST
WATER TESTING (46 CFR 162.060-10(B)(1)): TROJAN MARINEX'S METHOD FOR
ASSESSING ORGANISMS $\geq 10 \mu\text{M}$ AND $< 50 \mu\text{M}$

Encl: (1) Two copies of subject report

1. Enclosure (1), entitled "Review of a Request for Approval of an Alternative Method for Ballast Water Testing (46 CFR 162.060-10(b)(1)): Trojan Marinex's Method for Assessing Organisms $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$," provides a technical review of a request to use an alternative method to quantify organisms $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$ (nominally protists).
2. This work was sponsored by and guidance was provided by the United States Coast Guard Office of Environmental Standards (CG-OES-3).
3. The NRL points of contact are Lisa Drake, Code 6136, 305-293-4215, e-mail: lisa.drake@nrl.navy.mil; and Edward Lemieux, Code 6130, 202-404-2123, email: edward.lemieux@nrl.navy.mil.


BARRY J. SPARGO
By direction

Copy to:
USCG HQ (CG-OES-3 R. Everett)
USCG RDC (Environment and Waterways G. Roderick)

REPORT DOCUMENTATION PAGE*Form Approved
OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY)		2. REPORT TYPE		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code)



Naval Research Laboratory
4555 Overlook Ave, S.W.
Washington, DC 20375-5320



**Center for Corrosion
Science and Engineering**

Review of a Request for Approval of an Alternative Method for Ballast Water Testing (46 CFR 162.060-10(b)(1)): Trojan Marinex's Method for Assessing Organisms $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$

Lisa A. Drake¹, Timothy P. Wier², Evan W.J. Parson³, Jonathan F. Grant⁴, and Matthew R. First⁵

Chemistry Division, Naval Research Laboratory, Washington, DC 20375

Funded by the U.S. U.S. Coast Guard Environmental Standards Division (CG-OES-3)
MIPRs HSCG23-14-X-MMS010 and HSCG23-15-X-MMS020, Tasks 5.7.1 and 2.71, respectively
Technical Representative – Richard Everett
Contractor's Technical Representative – Regina Bergner

¹Code 6136, NRL Key West, FL 33041

²Excet, Inc., Springfield, VA 22150

³Vision Point Systems, Fairfax, VA 22030

⁴Battenkill Technologies, Inc., Manchester Center, VT 05255

⁵Code 6136, NRL Washington, DC 20375

FOR OFFICIAL USE ONLY

DISTRIBUTION STATEMENT F.

Further dissemination only as directed by USCG Environmental Standards Division (OES-3)
03 FEB 2016

Encl (1) to NRL Ltr Rpt
3900
6130/1622

This page left blank intentionally

EXECUTIVE SUMMARY

In response to international and national actions to reduce the transport and delivery of aquatic nuisance species (ANS) in ballast water, an industry of commercial ballast water management systems (BWMS) using a variety of technical approaches has developed over the past decade. To gain U.S. Type Approval (TA) for these systems, they must be tested according to regulations in 46 CFR 162.060—Ballast Water Management Systems, which incorporate by reference the Environmental Protection Agency (EPA) Environmental Technology Verification (ETV) Program “Generic Protocol for the Verification of Ballast Water Treatment Technology” (U.S. EPA 2010). Within the ETV Protocol, procedures and conditions for land-based verification testing of BWMS are described.

Trojan Marinex submitted a series of documents (Proposal) to the U.S. Coast Guard (USCG) requesting approval under regulation 162.060–10(b)(1) to use an Alternative Method to quantify concentrations of living organisms ≥ 10 and < 50 μm (nominally protists, which includes autotrophic and heterotrophic organisms, as discussed below). The currently Required Method prescribed in the ETV Protocol is a microscopy-based approach combining cellular movement and two fluorescent markers, fluorescein diacetate (FDA) and 5-chloromethylfluorescein diacetate (CMFDA). Here, organisms in the ≥ 10 and < 50 μm size class—which includes autotrophs (in this case, photosynthetic organisms known as phytoplankton or microalgae) and heterotrophs (organisms, such as amoeba, that derive energy from other organisms)—are quantified. The Required Method uses movement and the two fluorescent markers to assess a cell’s active enzymes (specifically, esterases) to measure if an organism is *living* (or dead, if no enzyme activity is visible and no cellular movement is detected). The proposed Alternative Method consists of two parts: (1) the Autotroph Method, which uses the Most Probable Number (MPN) method to quantify photosynthetic organisms, and (2) the Heterotroph Method, which uses direct microscope counts to quantify heterotrophic organisms (identified by movement and absence of chl *a* fluorescence). The MPN method is a culturing technique adapted from the field of food microbiology for use in evaluating phytoplankton communities, and it uses a 14-day (d) grow-out period to quantify *viable* (= able to reproduce) organisms. This approach is in contrast to the Required Method, which quantifies *living* organisms.

The Ballast Water Science and Technology Program at the Naval Research Laboratory (NRL) was tasked by USCG to review the Proposal and provide their technical opinion. This report (Review) is the outcome of that work. The following paragraphs summarize how the Proposal addressed the requisite elements of a 10(b)(1) request—specifically its practicability and applicability argument, completeness of description, and equivalency with the Required Method—as well as the applicability of the Alternative Method to compliance testing.

- The Proposal partly addressed the practicability criterion by stating the Required Method used to quantify living organisms is not practicable for ultraviolet (UV)-based BWMS. The Proposal asserted that, as systems are currently designed, it is not practicable to increase the UV dose to kill organisms (so that the Required Method would show cells were dead following UV treatment), thus obviating the need for the Alternative Method, which uses a 14-d grow-out period to show if cells are viable. Using a summary of data submitted to a peer-reviewed journal (the data are unpublished at present), the Proposal

argued that: a UV dose sufficient to reduce cell numbers by 100 fold and damage cells' enzyme systems to render them dead following treatment—the foundation of the Required Method—would necessitate a 10-fold increase in the UV dose. Such an increase would incur a concomitant 10-fold increase in the footprint of the BWMS and its electrical power requirements. The argument, however, contained no supporting data (e.g., the engineering and economic calculations to show the increase in power and corresponding number of BWMS that would be needed to use the Required Method were lacking). This information, as well as the data and dose-response curves (demonstrating the US dose required to kill organisms in this size class) from the unpublished study, should be provided.

- As a fundamental principal, the Proposal contended that the goal of the USCG ballast water regulations is to prevent the spread of potentially invasive organisms. Hence, a treatment that renders organisms unable to reproduce (rendering organisms “non-viable”, such as using UV treatment as currently applied by the BWMS manufactured by Trojan Marinex) meets the intent of the regulation. This reasoning addresses the “applicability” aspect of the 10(b)(1) request. The authors of this Review concur that such an outcome meets the intent, although not the letter, of the USCG final rule.

- The method described in the Proposal was, for the most part, clear and understandable. Field and laboratory experiments were summarized (and the raw data were provided) to support the conclusions made in the Proposal regarding development of the method. Further, the sequence of experiments was clear as the proposers gathered more information to refine the MPN portion of the Alternative Method. This Review identified several concerns in this portion of the evaluation; most stem from a lack of data to validate steps in the method or a lack of standardization. The items and their potential outcomes are listed below in decreasing order of importance.
 - The threshold for determining growth in an MPN tube is quite low, and it will be dependent on the fluorometer's calibration. These items should be specified and standardized in the Proposal.
 - The filter set described in the Heterotrophic Method is not optimized to detect chl *a*, the target molecule.
 - Additional data are required to determine if filters should be removed from the MPN tubes.
 - From calculations prepared for this Review, it appears the number of MPN tubes used in testing should be increased; this point should be investigated.
 - MPN values may be inflated if organisms <10 μm are inadvertently added to the MPN tubes as may occur if filters are left in the dilution tubes; this potential should be investigated.
 - The confidence intervals (CIs) around the MPN results should be explained in the Alternative Method and should be reported with all results; for that matter, all data using the Required Method should also include CIs.
 - Similarly, the means to combine the CIs for the Autotroph and Heterotroph Methods should be provided.

- The Alternative Method allows the Testing Organization (TO) to determine the approach in which samples are collected, and that should not be the case: the Alternative Method should stipulate how samples are collected.
 - The means to categorize organisms by size (as part of the Heterotroph Method) should be stated explicitly in the Alternative Method.
 - The number of samples analyzed for autotrophs and heterotrophs in tests of the Alternative Method's Initial Precision and Accuracy (IPA) and Ongoing Precision and Accuracy (OPA) should be increased from one to at least three. Likewise, an upper limit of the Factor of Agreement (FOA) between the Required and Alternative Methods should be defined in the Proposal.
- The Proposal did not provide adequate justification that the proposed Alternative Method is equivalent to the Required Method for three reasons:
- (1) The equivalence between dead and non-viable organisms was not demonstrated.
 - The equivalence between dead and non-viable organisms was not demonstrated using empirical data, and it should be provided to show, first, that dead and non-viable cells are equivalent, and second, that non-viable cells remain so. These issues could be addressed in an experiment using phytoplankton that are treated to induce immediate mortality without the possibility of repair (such as electrochlorination) or treated by UV radiation, and comparing the outcomes using the Required and Autotroph (MPN) Methods over a period >14 d (to assess the capacity of cells to undergo repair). This longer assessment period is important for organisms that are capable of repair but do not reproduce to detectable levels within the span of the 14-d incubation identified in the Alternative Method. These results could be augmented or replaced with similar experiments already described in the peer-reviewed literature.
 - (2) The equivalence between the precision of the Required and Alternative Methods was not demonstrated.
 - Laboratory experiments showed the precision was better in the Required Method than in the Alternative Method: the coefficients of variation (CVs) were 1-26% (average = 13%) for the Required Method and 20-135% (average = 50%) for the Autotroph (MPN) method. In all but one case, the CV of the Required Method was better (lower) than the CV of the Autotroph (MPN) Method. Thus, using a metric typically used to compare methods, precision, the Alternative Method was shown—using the data provided in the Proposal—to be less precise than the Required Method. Further, empirical measurements of the FOA (using ambient communities) between the two methods varied widely among different experiments. At times, the Alternative Method yielded a count at least five times greater than the Required Method. This result is surprising, since the water samples in these trials were not treated with UV radiation; in untreated samples, the Required and Alternative Methods would be expected to have similar results (FOA of ~1). This result could be due to the inclusion of organisms <10 μm in the samples that are not removed by

the filtration through the 10-µm pre-filter step prior to analysis with the Alternative Method.

- In addition, in the Proposal, the Alternative Heterotrophic Method was not compared to the Required Method (it was instead compared to a microscopy-based method that was different from the epifluorescence-microscopy method outlined in the Required Method). For a legitimate comparison, the Heterotroph and Required Methods should be compared.

(3) Statistical questions remain unresolved.

- Regarding the use of the Autotroph (MPN) Method, most critically, the percentage of non-culturable taxa is not accounted for in the calculation of the number of viable cells. This percentage should be known at each test facility (TF) using the Alternative Method, and upon the advice of statisticians, somehow incorporated into the calculations of cell densities. It is unclear how this factor would be incorporated into tests for shipboard verification of BWMS, because it would be nearly impossible to characterize the culturable and non-culturable taxa in every shipping port. Likewise, it is unclear how such “false negatives” would be incorporated into the Required Method during shipboard trials (that is, how organisms that are living but not moving or fluorescing would be addressed).
 - Other statistical concerns are more easily addressed: the MPN estimates are provided without measures of uncertainty (e.g., CIs), and they should be reported with all results. Additionally, at times, MPN estimates may be undefined, e.g., results may be reported as “greater than” (>) or “less than” (<) the detection limit, and in these instances, it is unclear how these results should be used, such as when calculating the sample mean. Finally, the process for combining the uncertainty in the autotroph concentration (based on probability from the MPN Method) and the uncertainty in the heterotrophic concentration (based on cell counts) is unclear (i.e., how the CIs for both numbers are used—are they combined in a straightforward fashion, by adding them together, or are additional calculations required?). These omissions should be rectified.
- For a number of reasons, it seems impracticable that the Alternative Method (specifically, the MPN portion of it) could be used for compliance testing to determine ships’ adherence to the ballast water discharge standard. If the Alternative Method was used for TA testing of BWMS and another method was used for compliance testing, it would be necessary to determine the correspondence between the two methods. At this point, it appears that variable fluorescence of photoautotrophs in unfiltered (whole) water samples may be used to determine compliance with the discharge standard. It would be prudent to begin a validation study to determine how the Alternative Method and variable fluorescence method relate to one another.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	3
TABLE OF CONTENTS	7
LIST OF FIGURES	9
LIST OF TABLES	9
ACRONYMS, ABBREVIATIONS, AND SYMBOLS	10
1 INTRODUCTION	12
1.1 Basis for Approval of Alternative Methods Changes	13
1.2 Goals and Objectives	14
2 JUSTIFICATION THE REQUIREMENT IS NOT PRACTICABLE OR APPLICABLE.....	14
3 PROPOSAL DOCUMENTS AND DESCRIPTION OF THE ALTERNATIVE METHOD	16
3.1 Documents Submitted	16
3.2 Overview of Alternative Method	20
3.2.1 Sample Collection	20
3.2.2 Autotroph Method—Most Probable Number (MPN) Approach	20
3.2.3 Heterotroph Method—Epifluorescence Microscopy	21
3.2.4 Determining the Number of Total Living Organisms	22
3.3 Potential Biases and Drawbacks of the Alternative Method	22
3.3.1 Drawbacks—Sampling	23
3.3.2 Drawbacks—Autotrophs	24
3.3.3 Drawbacks—Heterotrophs	25
3.3.4 Drawbacks—Data Analysis	26
3.4 The degree to which the Method Description Provides a Clear Presentation of Test Procedures.....	26
3.5 Validation of the Alternative Method	26
3.5.1 Considerations in Validating any Method.....	27
3.5.2 Appraisal of the Alternative Method Validation	28
3.5.3 Results of the Validation of the Alternative Method	32
3.6 Validation of the Required (FDA/CMFDA) Method	39
3.7 Statistical Review of the Alternative Method.....	41
3.7.1 Statistical Justification for the MPN Method.....	41
4 EQUIVALENCY OF THE PROPOSED METHOD WITH THE REQUIRED METHOD	45
4.1 Ability of the Alternative Method to Measure the Same Proximate Aspect or Function of the BWMS	46
4.2 Comparison of Uncertainties	48
4.2.1 Loss and Mortality.....	49
4.2.2 Inefficient Labeling	50

4.2.3	Detection of Motility	50
4.2.4	Detection of Chlorophyll <i>a</i>	50
4.2.5	Size Classification	51
4.2.6	Uncertainty of the Alternative Autotroph (MPN) method	51
4.2.7	Propagation of Uncertainty	52
4.3	Summary of Equivalency Determination.....	52
4.3.1	Equivalency—Dead to Non-Viable Organisms	52
4.3.2	Equivalency—Precision of the Required and Alternative Methods	53
4.3.1	Equivalency—Agreement of the Required and Alternative Methods.....	53
4.3.2	Equivalency—Applicability and Use of the Most Probable Number Method.....	53
5	COMPATIBILITY BETWEEN TYPE APPROVAL TESTING AND COMPLIANCE ASSESSMENT	54
6	SUMMARY	55
6.1	Practicability and Applicability	55
6.2	Full Method Description	56
6.3	Justification of Equivalence.....	56
6.4	The Alternative Method’s Use in Compliance Testing	58
	ACKNOWLEDGEMENTS	58
	REFERENCES.....	59
	APPENDIX 1—MPN Statistical Theory	63

LIST OF FIGURES

Figure 1. Theoretical comparison between two methods.	27
Figure 2. Comparison of the measurements of cell concentration of cultured algae.....	38
Figure 3. Comparison of the coefficient of variation of cell concentrations of cultured algae	39

LIST OF TABLES

Table 1. Ballast water discharge standards.	12
Table 2. Documents submitted as part of the Proposal.....	17
Table 3. Potential interferences, biases, and limitations as listed in the Alternative Method	23
Table 4. Specifications of two optical filter sets optimized for fluorescein detection.....	25
Table 5. Initial acceptance criteria for the Autotroph Method.....	30
Table 6. Documents in the Proposal describing the validation experiments.	33
Table 7. Factor of agreement between the Required Method and the Autotroph Method	36
Table 8. Agreement and precision of the Required and Autotroph Methods	37
Table 9. Summary of the key assumptions of the Most Probable Number method.	42
Table 10. Results of a theoretical matrix of MPN tubes.....	45
Table 11. Variables used to calculate concentrations of living organisms	48

ACRONYMS, ABBREVIATIONS, AND SYMBOLS

Term	Definition
<i>A</i>	Aliquot volume
ANS	Aquatic nuisance species
BP	Band pass
BWM	Ballast water management
BWMS	Ballast water management system
C	Celsius
<i>C</i>	Concentrated sample volume
CFR	Code of Federal Regulations
cfu	Colony forming unit
chl <i>a</i>	Chlorophyll <i>a</i>
CI	Confidence interval
CMFDA	Chloromethylfluorescein diacetate
CV	Coefficient of variation
d	Day
<i>D</i>	Dilution factor
DHI	Danish Hydrological Institute
DNA	Deoxyribonucleic acid
EPA	United States Environmental Protection Agency
ETV	Environmental Technology Verification Program
FDA	Food and Drug Administration or Fluorescein diacetate
FOA	Factor of Agreement
FR	Final Rule
G8	Guidelines for Approval of Ballast Water Management Systems
GSI	Great Ships Initiative
<i>I</i>	Individual count
IL	Independent Laboratory
IMO	International Maritime Organization
IPA	Initial Precision and Accuracy
ISO	International Organization for Standardization
L	Liter
LP	Long pass
<i>m</i>	Number of dilutions

Term	Definition
m ³	Cubic meter
min	Minute
mL	Milliliter
MLML	Moss Landing Marine Laboratories
MPN	Most Probable Number
MSC	Marine Safety Center (United States Coast Guard)
<i>n</i>	Number of subsamples
NIVA	Norwegian Institute of Water Research
NRL	Naval Research Laboratory
NRLKW	Naval Research Laboratory, Key West Florida
OPA	Ongoing Precision and Accuracy
<i>P</i>	Population concentration
PAR	Photosynthetically active radiation
<i>S</i>	Sample volume
SD	Standard deviation
SE	Standard error
SOP	Standard operating procedure
SP	Short pass
RNA	Ribonucleic acid
TA	Type Approval
TF	Test facility
TO	Testing Organization
USCG	United States Coast Guard
UV	Ultraviolet
UV-B	Ultraviolet radiation with wavelength of 290-320 nm
UV-C	Ultraviolet radiation with wavelength of 100-290 nm
<i>v</i>	Volume of subsample
δ	Density of organisms
μm	Micrometer

1 INTRODUCTION

Actions have been taken to reduce the transport of aquatic nuisance species (ANS) in ships' ballast water. At the largest—international—scale, the International Maritime Organization (IMO) adopted the Ballast Water Management Convention (BWM Convention; IMO 2004), which has yet to be ratified sufficiently to enter into force. At a national—US—scale, several statutorily authorized executive actions governing ballast water discharges were promulgated by the U.S. Coast Guard (USCG) and the Environmental Protection Agency (EPA) between 1990 and 2013 (e.g., USCG 2012, U.S. EPA 2013). Both the IMO and US actions aim to limit the number of organisms discharged in ballast water that could potentially establish new populations, and they have similar discharge standards (Table 1); most ships will use a ballast water management system (BWMS) to meet the numeric limits.

Table 1. Ballast water discharge standards.

Organization and standard	Organisms $\geq 50 \mu\text{m}$ in minimum dimension ^A	Organisms $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$ in minimum dimension ^B	Toxigenic <i>Vibrio cholerae</i> ^C	<i>Escherichia coli</i>	Intestinal enterococci
US Discharge Standard	$< 10 \text{ m}^{-3}$ (“living”)	$< 10 \text{ mL}^{-1}$ (“living”)	$< 1 \text{ cfu}$ 100 mL^{-1}	$< 250 \text{ cfu}$ 100 mL^{-1}	$< 100 \text{ cfu}$ 100 mL^{-1}
IMO Regulation D-2 Ballast Water Performance Standard	$< 10 \text{ m}^{-3}$ (“viable”) ^D	$< 10 \text{ mL}^{-1}$ (“viable”) ^D	$< 1 \text{ cfu}$ 100 mL^{-1} or $< 1 \text{ cfu g}^{-1}$ (wet weight zoopl.)	$< 250 \text{ cfu}$ 100 mL^{-1}	$< 100 \text{ cfu}$ 100 mL^{-1}

^ANominally zooplankton. ^BNominally protists. ^CSerotypes O1 and O139. ^DWhile the discharge standard in the Ballast Water Management Convention is framed in terms of “viable” organisms, the Guidelines for Approval of Ballast Water Management Systems (G8) specify that for purpose of testing the efficacy of BWMS, “viable” means “living”; cfu = colony forming unit, IMO = International Maritime Organization, and zoopl. = zooplankton.

Notably, the US standard applies to organisms that are “living” (which may include organisms unable to reproduce) rather than “viable” (able to reproduce). This distinction is due to (1) the inability to culture all potential organisms in the laboratory to determine their viability, and (2) the more protective and conservative criterion set by requiring cells to be *dead* rather than *non-viable* (e.g., Federal Register 2012). While the BWM Convention is framed in terms of “viable” organisms in regulation D-2, the IMO Guidelines for Approval of Ballast Water Management Systems (G8)—which were finalized after the BWM Convention was adopted—specify that for purposes of testing the efficacy of BWMS, “viable” is defined “organisms and any life stages thereof that are living” (IMO 2008). That said, in practice, viability assessments, (rather than live-dead assessments) have been made in Type Approval (TA) testing of BWMS under the BWM Convention. Here, the Most Probable Number (MPN) technique, developed for single

strains of bacteria, is used to quantify organisms in the $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$ size class (nominally protists). The bacteria for which the MPN technique was developed are readily culturable under laboratory conditions, whereas not all species present in the diverse community of ambient protists can be grown in the laboratory, which presents a difficulty, which is discussed below.

In the decade following the adoption of the BWM Convention, an industry for commercial BWMS has developed. In the US, under the USCG regulations, BWMS must be type approved by the USCG, and the process for granting TA for BWMS requires testing (following a rigorous quality management system) in accordance with standard, published protocols. The USCG Marine Safety Center (MSC) is responsible for reviewing, accepting, and approving applications for TA. Under this process, BWMS manufacturers are required to work with a USCG-accepted Independent Laboratory (IL), which will conduct land-based, shipboard, and environmental testing, as well as submit the TA test reports in support of the manufacturer's TA application to MSC.

The requirements for the TA application and verification testing are codified in the US Code of Federal Regulations (CFR; 46 CFR 162.060). Testing and test documentation must be in accordance with regulatory references for quality management (International Organization for Standardization, ISO 17025:2005(E)) (ISO 2005). Specifically, land-based testing will proceed according to the “Generic Protocol for the Verification of Ballast Water Treatment Technology” (U.S. EPA 2010). The Protocol was developed—with the input of stakeholders and technical experts—under the Environmental Technology Verification (ETV) Program in a joint effort between the USCG and the EPA. The draft protocol was tested at the Naval Research Laboratory, Key West, FL (NRLKW) (Lemieux et al. 2008), modified as needed, finalized (henceforth, ETV Protocol), and, later, incorporated by reference into the USCG final rule (FR) for the “Standards for Living Organisms in Ships' Ballast Water Discharged in US Waters” (USCG 2012). A shipboard protocol is now being drafted, and it may ultimately be incorporated into the existing ETV Protocol. In addition to land-based and shipboard verification testing under well-documented biological and water quality conditions, USCG regulations require all BWMS components, the marine suitability and safety of the BWMS, and its manufacturing processes to be assessed.

1.1 Basis for Approval of Alternative Methods Changes

The procedures used in sampling and analysis for US TA testing of BWMS are complicated, not only due to the engineering aspects (relatively large volumes of water are used, e.g., 200 m³ or greater), but also due to the biological aspects: a system's biological efficacy must be carefully measured to accurately determine adherence to a very strict numerical standard, e.g., < 10 organisms $\geq 50 \mu\text{m}$ in 1 m³ of ballast water. Accordingly, deviations in testing are allowable through regulation 162.060–10(b)(1), with requests addressed to and approved by USCG MSC. Here, this formal process allows for *pre-approval* for any deviations from the evaluations, inspections, or tests prescribed by the regulations. The request for a deviation must be justified and an Alternative approach validated:

“46 CFR § 162.060–10 Approval procedures. (b)(1) If an evaluation, inspection, or test required by this section is not practicable or applicable, a

manufacturer may submit a written request ... for approval of Alternatives as equivalent to the requirements in this section. The request must include the manufacturer's justification for any proposed changes and contain full descriptions of any proposed Alternative tests.”

1.2 Goals and Objectives

On 05 FEB 2015, Trojan Marinex submitted a 10(b)(1) request (Proposal, a series of documents) to USCG to use an Alternative Method to quantify living organisms in the $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$ size class. On 23 APR 2015, the Ballast Water Science and Technology Program at NRL was tasked by the USCG to review the Proposal and provide a technical evaluation on the following aspects of the application:

1. The justification that the requirement in 162.060 is not practicable or applicable,
2. The degree to which the method description provides a clear presentation of test procedures,
3. The justification that the proposed Alternative Method is equivalent to the Required Method, and
4. The potential for conflict or incompatibility between TA testing and compliance testing; that is, if efficacy during TA testing is based on viability and compliance assessment during vessel inspections is based on relationships between concentrations of living photosynthetic organisms and fluorescence of their photosynthetic pigment systems.

This report (Review) addresses those goals by first considering the elements of practicability and applicability. Next, the Alternative Method and its supporting documents are considered, followed by a discussion of the validation of the Alternative and Required Methods. Afterwards, the two methods' equivalency is discussed. Finally, the compatibility between the Alternative Method's use in TA and compliance testing is considered.

2 JUSTIFICATION THE REQUIREMENT IS NOT PRACTICABLE OR APPLICABLE

The foundation for a 10(b)(1) request is that a requirement in 162.060 is not “practicable or applicable”. That is, the requirement is (1) not possible to practice or perform or (2) not relevant, suitable, or appropriate for use in evaluating the performance of the BWMS. In this section, the practicability of using the Required Method to assess ballast water treated by ultraviolet (UV) radiation is addressed, followed by the Required Method's applicability in this instance.

The Required Method prescribed in the ETV Protocol quantifies living organisms in the $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$ size class, which includes autotrophs (here, photosynthetic organisms known as phytoplankton or microalgae) and heterotrophs (organisms, such as amoeba, that derive energy from other organisms). The assay quantifies living organisms (at the time of interrogation) and does not assess reproductive ability. This method combines two molecular, membrane-permeable markers, fluorescein diacetate (FDA) and 5-chloromethylfluorescein diacetate (CMFDA or CellTracker™ Green [Invitrogen]). These markers are widely used in many fields of biology to identify living cells. Initially as the Required Method was developed, trials were

conducted using the FDA and CMFDA markers singly; later, the markers were combined because they have similar emission spectra and could be combined and used simultaneously to stain a larger taxonomic range of organisms than either marker used singly would cover. When the colorless markers enter a living cell, non-specific esterases in the cell hydrolyze the probes to create a non-permeable product (i.e., it is trapped within the cell that has an intact cell membrane). The product fluoresces green when excited with blue light; the green signal is visible using an epifluorescence microscope. When an organism exhibits this green fluorescence or it is motile (even if it does not show green fluorescence), it is scored as “living”. Organisms that are motile (and thus are clearly living) that do not fluoresce green represent “false negative” results: they are living, so they should fluoresce, but they do not; regardless, due to their motility, they are enumerated as living cells. Another form of false negative results is organisms that are living but are not motile and do not fluoresce green—they are not scored as living cells. This error represents a shortcoming in the assay and is analogous to organisms that do not grow in MPN assays.

The argument is made by the Proposal that the Required Method will not “evaluate” organisms in the $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$ size class that are alive but treated with UV that “prevents cell replication (and thus precludes invaders from colonizing)” (Miller 2015a). Hence, the Proposal infers (but does not directly state) the Required Method is not applicable to assess UV treatment efficacy. Instead, the Proposal recommends an Alternative Method that combines two approaches to quantify organisms in this size class: (1) the Autotroph Method, which uses the MPN method to quantify photosynthetic organisms, and (2) the Heterotroph Method, which uses direct microscope counts to quantify heterotrophic organisms (identified by movement and absence of chlorophyll *a* [chl *a*] fluorescence). The Autotroph Method uses an MPN culturing technique with a 14-day (d) grow-out time to assess the number of *viable* organisms, that is, the number able to reproduce. This approach is in contrast to the Required Method, which assesses the number of *living* organisms.

The idea that the Required Method is not applicable to UV treatment is really an issue of practicability in achieving specific endpoints when using UV as a treatment. The UV doses *commonly used* in BWMS are not great enough to induce mortality during the 1-5 d hold time following treatment during TA testing. Increased UV doses would not only prevent reproduction or cellular division, but also induce mortality that is detected following treatment using the Required Method. If so, cell membranes would be compromised and cytosolic enzymes would be denatured or lost, and the fluorescent markers would not be hydrolyzed within cells. Thus, the Required Method would indicate treatment efficacy in meeting the “live” discharge standard because the dead cells would not fluoresce green. An Alternative Method would not be needed.

An assertion was made in the Proposal that it is not practicable to increase the UV dose so cells are dead (as determined by the Required Method) rather than not viable (as determined by the Alternative Method) (Miller 2015d). Using summary data (no dose-response curves were provided) from an unpublished study using 12 species of algae, the Proposal argued that a UV dose sufficient to damage cells’ non-specific esterases—the foundation of the Required Method—and reduce concentrations of algae by 100-fold (from 1000 cells mL^{-1} to 10 cells mL^{-1}) would be “extremely high”. On average, a 10-fold increase in UV dose was needed to kill cells using the Required Method compared to the dose needed to show cells were non-viable with the

Alternative method. It should be noted that in three of the 12 algal species, the study found no threshold at which cells were dead when scored using the Required method. In the remaining nine species, the average increase in dose required (i.e., 10-fold) was skewed by two species, which required an 18- to 34-fold increase. Seven of the 12 species required a dose <10-fold greater than that used with the Alternative Method to kill them. The doses were not specified. The Proposal further asserted such an increase would incur a concomitant 10-fold increase in the footprint of the BWMS and its electrical power requirements. Thus, requiring ballast water treated with UV BWMS to meet the discharge standard for organisms $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$ (i.e., < 10 living organisms mL^{-1}), as assessed using the Required Method, would not be practicable for this technology, essentially excluding UV treatment from the BWMS market. The Proposal, however, contained no supporting data to support the assertion of 10-fold increases in space or energy requirement. For example, calculations illustrating the increase in power and footprint that would be needed to meet the “live” criterion assessed by the Required Method were not included. The Proposal did not state why the practicability argument pertains to the autotrophic portion of the ballast water community (which is assessed with a viability method) but not the heterotrophic portion (which is assessed with a live-dead method).

As a fundamental principal, the Proposal contends that the goal of the USCG ballast water regulations is to prevent the spread of potentially invasive organisms. Therefore, a treatment that renders organisms unable to reproduce (rendering organisms “non-viable”, such as using UV treatment as currently applied by the BWMS manufactured by Trojan Marinex) meets the intent of the regulation. The authors of this Review concur that such an outcome meets the intent, although not the letter, of the USCG final rule. Further, any Alternative methods must be shown to be equivalent to the Required Method; the purpose of this report is to assess equivalency.

3 PROPOSAL DOCUMENTS AND DESCRIPTION OF THE ALTERNATIVE METHOD

3.1 Documents Submitted

Eighteen documents were submitted as part of the Proposal (Table 2). The first document in the table (1511138 – Request for Approval of an Alternative Method) provides an overview of the request and the initially submitted supporting documents; six documents were received after the initial submission.

Table 2. Documents submitted as part of the Proposal.

#	Title and Authors	Type of Submission
1	1511138 - Request for Approval of Alternative Method (Ref 46 CFR 162.060-10(b)(1)) Supplemental Information 02 MAR 2015 (Miller 2015a)	Correspondence: cover letter accompanying supplemental information (this correspondence references the letter accompanying the original request for approval [dated 05 FEB 2015, Reference 1511138], which was not provided to the authors of this Review)
2	1511138 - Request for Approval of Alternative Method (Ref 46 CFR 162.060-10(b)(1)) Updated Documentation 06 MAY 2015 (Miller 2015b)	Correspondence: cover letter accompanying the updated Alternative Method Description (see row 4 of this table)
3	An Alternative Method for Determining the Number of Living Organisms in the 10-50 μm Size Class for Ballast Water Management System Test Samples (Trojan Marinex 2015a)	Alternative Method Description: detailed protocol describing the Alternative Method
4	An Alternative Method for Determining the Number of Living Organisms in the 10-50 μm Size Class for Ballast Water Management System Test Samples, 05 MAY 2015 (Trojan Marinex 2015b)	Alternative Method Description: updated detailed protocol describing the Alternative Method, which was the version used for this Review; at the same time, an additional version was submitted concurrently with the “track changes” feature in Microsoft Word; this document was not assigned a separate document number
5	Evaluating the MPN Dilution-Culture Method for the Enumeration of Viable Phytoplankton Cells (Petri 2015a)	Supporting information: paper that describes the MPN dilution culture method
6	Flow Cytometric Analysis of the Relative Abundance of Heterotrophs and Autotrophs in the Regulated 10-50 μm Size Class (Maurer and Welschmeyer 2015a)	Supporting information: presentation that describes the relative abundance of heterotrophs and autotrophs at several US West Coast locations and at DHI Denmark

#	Title and Authors	Type of Submission
7	MPN Assay – Analyses of Algal Regrowth for Performance Evaluation of Ballast Water Management Systems Primary Validation (DHI 2014)	Supporting information: paper that describes the validation of the MPN assay at DHI Denmark
8	MPN Method Development Experiments 1 to 3 Inter-Lab Comparison of the MPN Dilution-Culture Method and Fluorescein-Based Staining Methods for the Enumeration of Viable or Living Phytoplankton Cells (Petri 2015b)	Supporting information: paper that describes experimental results and analyses of BWMS samples obtained from a test at DHI Denmark and then analyzed at three locations: DHI Denmark, MLML, and BallastTech-NIVA AS
9	MPN Method Development Report Experiment 4 (Miller et al. 2015a)	Supporting information: paper that describes experimental results and analyses of samples that were collected at three laboratories: DHI Denmark, MLML, and NIVA. Samples were analyzed to determine the MPN results with different growth media, temperature, and filtration
10	MPN Method Development Report Experiment 5 (Miller et al. 2015b)	Supporting information: paper that describes with experimental results and analyses of samples collected at DHI and NIVA. The Autotroph Method, Heterotroph Method, and Required Method were used to analyze the samples
11	MPN Method Development Experiment 6 Generating Method Performance Data for the Alternative Method for Analyzing 10-50 µm Organisms in the ETV Generic Protocol for the Verification of Ballast Water Treatment Technology (Miller and Petri 2015)	Supporting information: paper that describes experimental results and analyses of samples for a workshop in London, Canada. This evaluation used standard test organisms, <i>Tetraselmis spp.</i> (phytoplankton) and <i>Brachionus spp.</i> (a rotifer, a heterotroph), to assess the Autotroph and Heterotroph Methods at various cell concentrations and compared the results to those obtained using the Required Method. The goal was to determine the precision and accuracy of the Alternative Method

#	Title and Authors	Type of Submission
12	On the use of the serial dilution culture method to enumerate viable phytoplankton in natural communities of plankton subjected to ballast water treatment (Cullen and MacIntyre 2015)	Supporting information: article that was accepted by the Journal of Applied Phycology that describes the use of MPN assay for assessing viability of organisms; the final version of this document was later sent but not assigned a separate document number
13	Rationale for the Use of Most Probable Number (MPN) Technique in the Evaluation of UV-based Ballast Water Management Systems (Maurer and Welschmeyer 2015b)	Supporting information: paper that analyzes data from GBF over a 2-year period to compare the FDA counting technique to the MPN assay that was used during testing of UV-treated ballast water samples at GBF
14	Toward Best Practices for Assessing the Effectiveness of Ultraviolet Radiation for Treatment of Phytoplankton in Ballast Water (MacIntyre et al. 2015)	Supporting information: presentation that describes false positives that may be encountered when the markers used in the Required Methods are compared to the MPN growth assay
15	1511138—Request for Approval of Alternative Method (Ref 46 CFR 162.060-10(b)(1) Supplemental Information 03 JUNE 2015 (Miller 2015c)	Correspondence: letter that provides an overview of the publication by Cullen and MacIntyre et al. (2015)
16	1511138—Request for Approval of Alternative Method (Ref 46 CFR 162.060-10(b)(1) Supplemental Information 30 JUNE 2015 (Miller 2015d)	Correspondence: letter that explains why the Required Method is not practicable
17	1511138 - Request for Approval of Alternative Method (Ref 46 CFR 162.060-10(b)(1) Supplemental Information 23 JULY 2015 (Miller 2015e)	Correspondence: letter that explains the equivalent environmental protection of the Alternative Method from a quantitative and experimental perspective
18	1511138 - Request for Approval of Alternative Method (Ref 46 CFR 162.060-10(b)(1) Supplemental Information 26 JULY (Miller 2015f)	Correspondence: letter with an overall summary of the accuracy of the proposed Alternative Method

DHI = Danish Hydrological Institute, ETV = Environmental Technology Verification, FDA = fluorescein diacetate, GBF = Golden Bear Facility, MLML = Moss Landing Marine Laboratories, MPN = most probable number, and NIVA = Norwegian Institute of Water Research

3.2 Overview of Alternative Method

The Proposal describes a two-part Method that evaluates autotrophs and heterotrophs in the ≥ 10 μm and < 50 μm size class (it is assumed that organisms ≥ 50 μm are removed in the BWMS prior to treatment with UV, and thus, this size class is not considered in the Proposal or this Review). The terms “Autotroph Method” and “Heterotroph Method” were used throughout the Proposal, and for clarity, this Review will employ consistent terminology; the sum of the two methods, i.e., the overall method, is deemed the Alternative Method. Note that the Alternative Method sometimes referred to the largest size class as “10-50 μm ”, which this Review considers equivalent to “ ≥ 10 μm and < 50 μm ”, or it discussed organisms > 50 μm , which this Review considers equivalent to ≥ 50 μm (following the size classes delineated in Table 1).

3.2.1 Sample Collection

In the Alternative Method, triplicate subsamples are collected and analyzed. The Alternative Method allows the testing organization (TO) some latitude in procedure. For example, to determine how the triplicate subsamples are collected, the available choices are: “...either 3 replicate samples can be collected with a subsample taken from each, or a single sample can be taken with 3 replicate subsamples taken” (Trojan Marinex 2015b). This process is discussed further below (Section 3.3.1 Drawbacks—Sampling).

3.2.2 Autotroph Method—Most Probable Number (MPN) Approach

The Autotroph Method uses a chl *a*-based MPN culture dilution technique to evaluate the concentration of reproductive, photoautotrophic organisms (phytoplankton). While autotrophs may derive energy from other sources than oxygenic photosynthesis, such as chemoautotrophy, the most abundant autotrophs expected to be encountered in ballast water are photoautotrophs (phytoplankton or microalgae). The methods in the Proposal were developed to evaluate UV treatment, although, in theory, the methods should be applicable to water treated by any type of BWMS. Regardless, the Proposal discussed the damage rendered by UV treatment to organisms’ deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), which can prevent the organism’s ability to reproduce. The Proposal also contended that organisms incapable of reproduction were the same (from the standpoint of the risk for biological invasions) as organisms that are scored as “dead” in the Required Method. The Proposal considered “viable cells” the same as the cells scored as “living” using the Required Method, and in several instances, interchanged the term “viable” with “living”. This distinction is critical because the current U.S. Ballast Water Discharge Standard sets a limit on the allowed concentration of *living* organisms, and this Proposal would equate living, non-reproducing cells to dead cells. This point is discussed further below (Section 4.1 Ability of the Alternative Method to Measure the Same Proximate Aspect or Function of the BWMS).

In the Autotroph Method, a whole water sample of ballast water is collected, filtered on a 10- μm filter to remove microorganisms < 10 μm in size, and the material retained on the filter is serially diluted. An array of culture tubes are filled with media (Guillard’s for marine water, and presumably [although not stated], also for brackish water; Bold Modified Basal media for

freshwater). Three dilutions are created, and each dilution has 5 replicate culture tubes (thus, a total of 15 MPN tubes is used). The initial fluorescence of the chl *a* is measured in each tube using a fluorometer, the tubes are incubated in continuous irradiance (nominally 50-150 $\mu\text{mol photon m}^{-2} \text{ s}^{-1}$ photosynthetically active radiation [PAR]) for 14 days (d), and then the fluorescence is measured again.

The tubes are scored as + (showing population growth [growth]) or - (showing no growth) based on an increase (or the lack of an increase) in fluorescence; an increase in fluorescence from the start of the incubation period to the end of the incubation period shows an increase in the amount of chl *a*, which corresponds to an increase in the number of cells. Note that the term “growth” is used in this Review, and it applies to the growth of the population, not an increase in the size of an individual organism. In addition, growth refers to culturability over the duration of the MPN assay and does not imply the organisms must be maintained in perpetuity.

The threshold for growth is set at “4 times the standard deviation of the fluorescence of a set of method blanks”, with the justification that it would “provide the easiest (minimal fluorescence increases) scoring of growth, leading to higher MPN values, and thus the most environmental protection when used to evaluate equipment performance.” (Trojan Marinex 2015b). This idea should be formalized with data, e.g., an experiment to demonstrate this threshold is appropriate. Note that for the purposes of this Review, we assume that the fluorometer is properly calibrated and yields stable, reproducible readings across its range of detection. The scoring pattern is then entered into an MPN “calculator” (e.g., <http://www.i2workout.com/mcuriale/mpn/index.html>) to determine the most probable concentration of viable autotrophs in the original sample. In the Proposal, three MPN calculators were evaluated, with no difference among their outputs of cell concentration (Trojan Marinex 2015b). The confidence intervals (CI) were different, but the Proposal dismissed this difference, as the CIs are not used in the Alternative Method (this idea is discussed in Section 3.7.1.2 Confidence Intervals in the Autotroph Method).

3.2.3 Heterotroph Method—Epifluorescence Microscopy

Because heterotrophs do not convert inorganic carbon to organic compounds using photosynthesis, and therefore do not contain chl *a*, the Autotroph (MPN) Method is not applicable to this group of organisms. The proposed Heterotroph Method uses a microscopic technique to evaluate if an organism is a living heterotroph by assessing two criteria, motility and chl *a* autofluorescence. Moving cells or organisms are clearly living and are scored as such. Obligate heterotrophs lack chl *a* (the primary light-harvesting pigment in photosynthesis). When excited with blue light, chl *a* fluoresces red, naturally exhibiting red autofluorescence; no fluorescent markers are added to the sample. Thus, using this criterion, organisms *lacking* chl *a* do not exhibit red autofluorescence and are heterotrophs.

In the Heterotroph Method, a 1-mL subsample is loaded into a Sedgewick Rafter counting chamber, which is examined on an epifluorescence microscope, and living heterotrophs are identified as organisms that *do* show motility but *do not* show red autofluorescence (i.e., they lack chl *a*). Notably, this method differs from the one used in the Required Method that is used to enumerate living heterotrophs in this size class of organisms. In the Required Method, samples exhibiting green fluorescence from FDA and CMFDA markers, motility, or both

approaches are scored as living. The Required Method is used to enumerate all living organisms (both autotrophs and heterotrophs) in the size class. In addition to the methodological differences used to quantify heterotrophs using the Required and Alternative Methods, the methods also differ in their fundamental approaches. The Required Method solely enumerates living organisms (autotrophs and heterotrophs), whereas the Alternative Method enumerates living organisms (heterotrophs) and viable organisms (autotrophs). It is unclear (and unexplained in the Proposal) why it would be appropriate to enumerate the living organisms in the heterotrophic portion of the protist community when UV treatment (at the doses currently used in ballast water applications) is intended to render cells non-viable. Following the argument that non-viable cells are equivalent to dead cells, the viability of the heterotrophs would be determined as well.

3.2.4 Determining the Number of Total Living Organisms

Under the Alternative Method, the mean of three subsamples analyzed by the Autotroph Method is added to the mean of three subsamples analyzed by the Heterotroph Method to obtain the total living organism concentration in the $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$ size class. This value is used to evaluate the performance of a BWMS compared to the ballast water discharge standard, or to quantify organisms in the uptake water or control discharge. The combination of these values is discussed below (Section 3.3.4 Drawbacks—Data Analysis).

3.3 Potential Biases and Drawbacks of the Alternative Method

The Proposal includes a discussion of possible interferences, biases, limitations, and mitigating strategies that may affect the data collected using the Alternative Method. Therein, these were grouped by autotrophs and heterotrophs, and for this Review, they are excerpted and summarized (Table 3). The biases seem more impactful in the Autotroph Method than the Heterotroph Method, but since both carry potential risk for bias, both parts of the Alternative Method will be considered. Below (in this section and others), these items, and additional concerns, are discussed.

Table 3. Potential interferences, biases, and limitations as listed verbatim in the Alternative Method (Trojan Marinex 2015b); additional text added for this Review is indicated in square brackets.

[Description]	Bias on Number of Living 10-50 μm Organisms	Proposed Magnitude [and Mitigation Strategy from this Review]
Autotrophs [Autotroph Method]		
Inclusion of >50 μm Autotrophs	Overestimate	Small, <3%
Inclusion of <10 μm Autotrophs	Overestimate	[Not listed]
Removal of >10 μm Autotrophs	Underestimate	Can be minimized [with gentle filtration]
Filtration mortality of >10 μm Autotrophs	Underestimate	Can be minimized [with gentle filtration]
Non-Growing Autotrophs	Underestimate	Small, near 0% by abundance [when detailed taxonomy was used]
Slow-Growing Autotrophs	Underestimate	Small
Chain-Forming Autotrophs	Under- and Overestimate	[Can be reduced when chains are “shortened by disruptions like pumping and tank agitation”*]
Grazing by Heterotrophs	Underestimate	Small
Heterotrophs [Heterotroph Method]		
Inclusion of >50 μm Heterotrophs	Overestimate	Can be minimized [as they will be excluded in microscopic counts]
Inclusion of <10 μm Heterotrophs	Overestimate	Can be minimized [as they will be excluded in microscopic counts]
Removal of >10 μm Heterotrophs	Underestimate	Can be minimized [with gentle filtration]
Filtration mortality of >10 μm Heterotrophs	Underestimate	Can be minimized
Assessment of Live-Dead Status	Overestimate	Large for UV treatment

*Cullen and MacIntyre (2015) recommend “relatively gentle mixing” of 100 inversions to disrupt chains or colonies.

3.3.1 Drawbacks—Sampling

Recall that the sampling scheme allowed for two options: “...either 3 replicate samples can be collected with a subsample taken from each, or a single sample can be taken with 3 replicate subsamples taken” (Trojan Marinex 2015b).” As above, it would be preferable if only one sampling scheme was put forward. More worrisome, this choice can greatly affect the outcome of the sampling. The latter approach results in “pseudoreplication” (Hurlbert 1984), in which the

“replicates” are actually “subsamples” that violate the assumption of independence, which is the foundation of most commonly used statistical analyses. That is, the statistical analyses would be flawed, yielding results that were wrong. Thus, the first approach (three replicate samples) should be used for statistical comparisons, e.g., between the control and treatment tanks upon discharge. This advice does not apply to sample tubes used in the MPN approach; in this case, water for all dilution tubes (i.e., an array consisting of 3 dilutions, each with 5 replicates) should be drawn from the same population, so here, tubes are inoculated with water from a single, original sample. Regardless, the procedures should be unambiguous in the Proposal.

3.3.2 Drawbacks—Autotrophs

A small number of steps within the Alternative Method invite ambiguity. Others are not fully validated. These potential drawbacks are discussed below.

3.3.2.1 Drawbacks—Autotrophs: Filtering

Samples for the Autotroph (MPN) Method are filtered onto 10- μm filters, and if organisms $<10\ \mu\text{m}$ are inadvertently retained on the filter, estimates of organism numbers will be artificially inflated. In effect, the final MPN number would include organisms that are regulated by USCG ($\geq 10\ \mu\text{m}$ and $<50\ \mu\text{m}$) as well as those that are not regulated by USCG ($<10\ \mu\text{m}$). This question could be addressed by filtering cultures or ambient organisms that are $<10\ \mu\text{m}$ and using microscopy to measure the organisms retained on the filter.

Organisms $\geq 50\ \mu\text{m}$ could also be retained on the filter, again, artificially inflating the MPN estimate. The Proposal indicates this bias would be small, with a note that phytoplankton $\geq 50\ \mu\text{m}$ tend to be rare. While this is generally true, it would be prudent for each test facility (TF) using this approach to determine the fraction of relatively large phytoplankton at the test facility. In shipboard testing, a whole water sample could be collected and the cells measured by microscopy to ensure the community of microalgae did not include cells $\geq 50\ \mu\text{m}$.

According to the Alternative Method, the filter may or may not be left in the MPN tube during the grow-out period. This is a concern because it is unclear if leaving a filter in the tube affects the fluorescence reading (as discussed above) or if the filter itself serves as a barrier, reducing the fluorescence reading. If this issue has been addressed, it was not included in the Proposal. Regardless, the Alternative Method should be updated to provide unambiguous direction, and if the direction is to leave the filter in the tube, data showing that the practice does not affect the results should be presented.

3.3.2.2 Drawbacks—Autotrophs: Measurements

To determine if an MPN tube has growth (and is scored as positive) a threshold was set in the Proposal as four times the standard deviation (SD) of method blanks. In order for this threshold to be uniformly applied across laboratories, the fluorometers would need to be calibrated following the same, standardized procedure and demonstrate consistency in measuring the threshold value. This specification is not included in the proposed method, and it should be.

3.3.3 Drawbacks—Heterotrophs

The Heterotroph Method relies partly on the ability to detect the red autofluorescence of chl *a*-containing organisms using epifluorescence microscopy. The basis of epifluorescence microscopy is that a series of optical filters is arranged to select visible wavelengths to illuminate organisms in the field of view. Fundamentally, fluorescing molecules will emit light at a different (longer) wavelength than the wavelength used to excite them, and thus, using a given set of filters for excitation and emission, fluorescing biomolecules can be excited and detected.

For chl *a* (or any target fluorescent substance), excitation and emission filters should correspond to the maximum excitation and emission wavelengths (here, 430 nm and 662 nm, respectively, for chl *a*). However, in the Heterotroph Method, the optical filter set is optimized to detect the fluorescence from fluorescein, not chl *a*; the guidance was developed to accommodate commonly used filter sets, so this one was chosen because green-fluorescing markers, such as fluorescein, are very common (B. Petri, pers. com.). According to the wavelengths of the filter set specified in the Heterotroph Method (B-2E/C), however, the red autofluorescence of chl *a* would *not* be visible (Table 4). This arrangement could lead to an overestimate of the number of heterotrophs, as all organisms detected would appear to lack chl *a* and, therefore, if they were motile, they would be scored as heterotrophs. It is unknown if this method was used in TA applications submitted to USCG. If it was, that would indicate the overestimation of heterotrophs was small enough not to exceed the discharge standard, since TA applications indicate the discharge standard is not exceeded.

Table 4. Specifications of two optical filter sets optimized for fluorescein detection. The filter set names are specific to Nikon microscopes, but other, similar configurations are available.

Filter Set	Excitation Filter	Dichromatic Mirror	Emission Filter	Red Fluorescence
B-2E/C*	465-495 nm BP	505 nm LP	515-555 nm BP	Not visible
B-2A	450-490 nm BP	500 nm LP	515 nm LP**	Visible

*The filter specified in the Alternative Method. **Wavelengths longer than 515 nm, such as 662 nm (the peak emission for chl *a*), can be detected. BP = Band Pass, a filter that transmits wavelengths of light within the range specified, dichromatic mirror = a type of LP filter that reflects light of wavelengths shorter than the nominal value and transmits longer wavelengths, LP = Long Pass, a filter that attenuates wavelengths of light shorter than a nominal value to transmit longer wavelengths, nm = nanometer, and SP = Short Pass, an optical filter that attenuates wavelengths of light longer than a nominal value so that shorter wavelengths are transmitted).

Encouragingly, filter sets are available that detect fluorescein *and* may also allow chl *a* fluorescence to be detected. One such filter set (B-2A) is listed in Table 4 for comparison to the filter set identified in the Alternative Method. While the excitation filter in B-2A is not optimal (it is not well aligned with the peak excitation of chl *a*, 430 nm), the emission filter would allow red fluorescence to be detected. If this filter set (B-2A in Table 4) does not allow sufficient detection of chl *a*, other, chl *a*-specific, filter sets exist, and they could be specified.

Another concern is the procedure used to detect non-fluorescing organisms to quantify living heterotrophic organisms. Recall that epifluorescence microscopy is used, and if cells are moving and do not show red autofluorescence (i.e., they do not contain chl *a*), organisms are scored as living heterotrophs. The Alternative Method states that organisms are detected by movement, but if they do not exhibit fluorescence and are not viewed with another light source (one is not stipulated in the Proposal), organisms (whether moving or stationary) would be, at best, dimly illuminated by light and difficult to see. Perhaps a step to initially detect moving organisms (via brightfield or phase-contrast microscopy) is implicit in the method, but this point should be explicitly stated and clearly described.

3.3.4 Drawbacks—Data Analysis

From the materials provided, it is unclear how the uncertainties around the Autotroph and Heterotroph Methods are applied. Each will have an estimate of the mean's variance. If they are to be added together, there is no statistical justification provided for doing so. These instructions should be provided in the Alternative Method.

3.4 The degree to which the Method Description Provides a Clear Presentation of Test Procedures

Overall, the Alternative Method is clear and understandable. In several instances, however, it allows the TO to make potentially impactful decisions. It is best, in the opinion of this Review's authors, if a method prescribes steps to allow as little differential implementation as possible. Following that idea, in some instances (enumerated above), data are needed to justify the use of a particular procedure, e.g., leaving the filter in the MPN tube for analysis. These data should be provided.

3.5 Validation of the Alternative Method

In this section, first, the means to validate *any* method will be discussed. Next, the validation that was completed by the proposers for the Alternative Method will be described. Then, for completeness, the validation of the Required Method will be summarized. Finally, the statistical foundation of the Alternative Method will be reviewed.

Before the validation is discussed, it is useful to consider two parameters that are commonly used to assess any method: precision and accuracy. If two methods are compared, the relationship between them can be illustrated in a conceptual figure (Figure 1), with the key metrics of precision (Figure 1A) and accuracy (Figure 1B) of the method relative to a known value (or a reference method). In a comparison between the Alternative and the Required Methods, the *true* cell concentration is not—and cannot be—known, so methods can be compared based upon their precision (the variation around the mean value of each method) and the difference between the two means (a proxy for accuracy, since the true value is not known) (Figure 1C). Ideally, the two methods will have a similarly small variation around the mean. Notably, the Alternative and Required Methods measure different metrics: the Alternative Method measures non-photosynthetic organisms by movement and autofluorescence, and it measures viable

phytoplankton amenable to culturing, whereas the Required Method measures phytoplankton and non-photosynthetic organisms in this size class by the examining active enzyme systems.

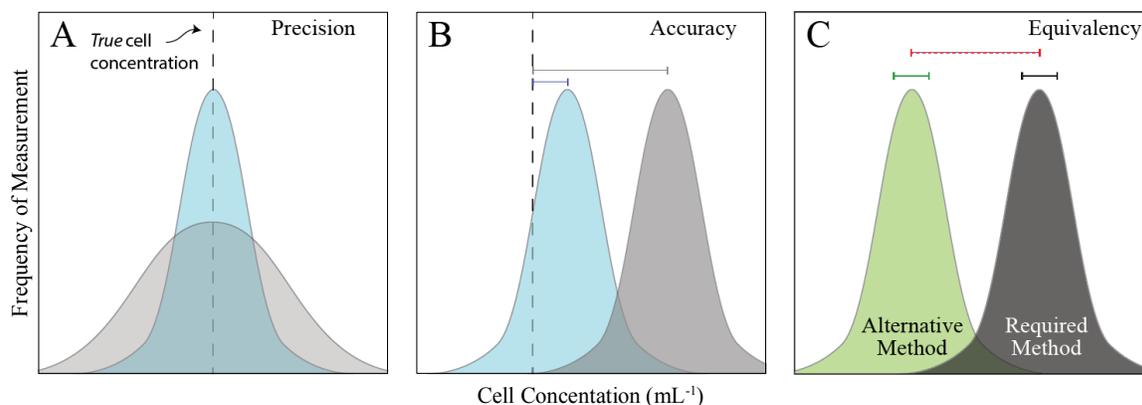


Figure 1. Theoretical comparison between two methods. The panels show conceptual frequency distributions of measurements of cell concentrations. The two methods can be compared based upon their precision (i.e., the spread of the data used to calculate the mean value, Panel A) and their accuracy (i.e., the correctness of the estimated mean value relative to the true value, Panel B). The true cell concentration is shown as a dotted, vertical line. The two methods may have the same measurements of the mean cell concentration, but the precision (variation around the mean) may differ, as shown in Panel A. The two methods may have the same precision, but the mean values may differ in accuracy (Panel B). In a comparison between the Alternative and the Required Methods, the *true* cell concentration is not known, so methods can be compared based upon their precision (the variation within a set of readings, shown as the solid error bars above the measurement sets) and the difference between their means, shown as the red dotted line (panel C).

3.5.1 Considerations in Validating any Method

The validation of a method requires considering numerous parameters, such as those outlined in a widely used paper describing the validation of chemical assays (Green 1996). While this framework applies to chemical methods, it identifies attributes of a method that are useful in any application. Note that while the Proposal addresses many of these elements explicitly, it also includes additional parameters, in particular, the Factor of Agreement (FOA)—the approach defined in the Proposal that measures the agreement between the Required and the Alternative Methods:

- **Accuracy:** The difference between estimated and actual values; since standards with known concentrations of living organisms are not available, accuracy is determined by a comparison estimates of the Alternative Method to the Required Method
- **Precision:** The variation among replicate readings or analyses; several factors contribute to the overall precision of a method:

- ***Volumetric precision***: The precision of the device used to aliquot, transfer, or inject the sample
- ***Inter-replicate precision***: A measure of variation among readings of multiple subsamples from a single sample
- ***Inter-analyst or inter-instrument precision***: A measure of the variation among readings of subsamples among different analysts or different instruments
- ***Reproducibility***: A measure of variation in the readings of a single sample across multiple laboratories or locations
- ***Specificity***: A measure of whether the assay is completely inclusive of the target population (in this case, living organisms $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$) and exclusive of non-target organisms
- ***Linearity***: A measure whether the response of an assay (i.e., its estimates of concentrations) vary proportionally to the actual organism concentrations or aliquot volumes
- ***Detection range***: A numerical range indicating the upper and lower limits of detection for the assays; since organism $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$ can be concentrated or diluted prior to analysis, the detection range can span several orders of magnitude; some specific factors for consideration are described below:
 - ***Detection limit***: The lowest detectable concentration; for a direct microscope count, the detection limit is one single organism. In the Alternative Method, organisms are concentrated on a mesh filter prior to preparing an array of dilutions, each with replicate tubes. In this case, the limit of detection hinges upon the volume concentrated on the filter and the set of dilutions
 - ***Quantitation limit***: The lowest concentration that can be accurately and precisely measured
- ***Sensitivity (Robustness)***: The intensity of change in an assay's response to small changes in the assay conditions; for example, what is the difference in concentration estimates if a sample is incubated for eight minutes (min) (rather than 10 min) with fluorescent labels? Or does a change in the incubation light regime affect the final result due to consequent effects on autotroph reproductive rates?

3.5.2 Appraisal of the Alternative Method Validation

In this Review, the Alternative Method used to quantify living organisms $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$ is evaluated by the guidance for analytical method validation (Green 1996). Here, the Alternative Method is compared to the Required Method. Both approaches face similar complexities in the analysis of living organisms (which differ from the features and attributes of chemical assays appraised by Green [1996]). This Review first considers the acceptance criteria listed in the Proposal. Next, the results from the Alternative Method validation are considered (beginning in Section 3.5.3 Results of the Validation of the Alternative Method)

3.5.2.1 Establishment of Acceptance Criteria

A key step in method validation is to establish minimum acceptance criteria to evaluate the performance of the method (Green 1996). The criteria used to gauge whether the Alternative Method is valid are defined and described in Trojan Marinex (2015b) (in the following sections, experiment [Exp.] numbers, or table numbers from [Trojan Marinex 2015b] are cited where appropriate). Because the Alternative Method combines two parts (the Autotroph Method and Heterotroph Method), they are discussed separately below.

3.5.2.2 Acceptance Criteria—Autotroph (MPN) Method

To determine the Initial Precision and Accuracy (IPA; this term was defined and used in the Proposal) of the Required and Alternative Methods, three subsamples of each of three concentrations of the cultured phytoplankter *Tetraselmis suecica* were analyzed by the proposers using both the Required and the Autotroph (MPN) Methods (Trojan Marinex 2015b). Since all organisms were phytoplankton (autotrophs; no heterotrophs were used in this case) and presumably the phytoplankton were capable of growth (since they were in culture, under favorable light and nutrient conditions), there was no need to perform the Heterotroph Method in addition to the Autotroph Method to determine the total number of organisms.

To determine accuracy, the Proposal showed it should be calculated as the mean of the three subsamples using the Autotroph approach divided by the mean concentration determined using the Required Method, multiplied by 100 (the Required Method was considered the benchmark for this comparison). The value (converted to a percent) would be 100% if both methods yielded the exact same concentration. Because the “true” concentration is unknown, this metric actually shows the agreement between the Required and Alternative Methods (rather than accuracy, which is the term used in the Proposal). For consistency with the Proposal, “accuracy” will be used in this Review.

Precision for the Autotroph Method was calculated in the Proposal as the coefficient of variation (CV). Specifically, the SD of three subsamples was divided by the mean of the three subsamples and multiplied by 100. This calculation is typically used to quantify precision.

3.5.2.2.1 Acceptance Criteria—Autotroph (MPN) Method—Initial Precision and Accuracy Values

The acceptance criteria for accuracy and precision defined in the Proposal were self-derived, based on empirical work conducted at the Danish Hydrological Institute (DHI) and the Norwegian Institute of Water Research (NIVA) in support of the Alternative Method development (Trojan Marinex 2015b) (Table 5 in this Review; Miller and Petri 2015). According to the Proposal, these criteria will be used to determine the IPA, and if the thresholds are met, analyses of field samples can proceed by organizations using the Alternative Method. For example, at a concentration of 10 *T. suecica* cells mL⁻¹, the CV must be less than 48% and the accuracy must fall within 61 and 168%. The accuracy and precision values in Table 5 would be used to determine acceptable precision and accuracy; in this manner, these criteria serve as data quality indicators. This Review suggests the number of samples per concentration should be increased from one to three (and the subsamples per concentration decreased from three to one), because, with few exceptions, at least three independent measurements are taken in

scientific endeavors to provide a reasonable estimation of the uncertainty around the result. Additionally, the OPA measurements should also be applied to ambient communities, since they are the communities used in BWMS testing, and it is expected that the community composition will change temporally (see Section 3.5.2.4 Factor of Agreement (FOA)).

Table 5. Initial acceptance criteria for the Autotroph Method as defined and listed in the Proposal for the comparison of the Alternative Method and the Required Method. This table copies the data in Trojan Marinex 2015b (Section 10.2.2.1.4, Table 4 therein). CV = coefficient of variation.

Concentration of <i>Tetraselmis suecica</i> (mL ⁻¹)	Precision (CV)	Accuracy*
10	<48%	61 – 168%
100	<55%	109 – 166%
1000	<63%	107 – 161%

*As defined in Section 3.5.2.2 Acceptance Criteria—Autotroph (MPN) Method of this Review: (mean concentration using the Autotroph approach)/(mean concentration using the Required Method) * 100.

Here, precision is dependent upon sample concentration. Unexpectedly, measurements of higher concentrations (100 and 1000 mL⁻¹) are *less* precise (the CV is higher) than measurements of the lowest concentration (10 mL⁻¹). This is surprising, as CV (a measure of precision) is sensitive to small mean values, and as the mean approaches zero, CV approaches infinity.

Regarding the accuracy of the acceptance criteria, for high concentrations (100 and 1000 mL⁻¹), it appears that MPN estimates are expected to be greater than estimates from the Required Method (as the accuracy range is entirely >100%). Note that these criteria were established by conducting empirical experiments, and the criteria reflect the empirical findings from a study the proposers conducted using a single, cultured alga. Such conditions should yield highly accurate and precise measurements. Presumably, this is the justification for using these values as acceptance criteria.

3.5.2.2.2 Acceptance Criteria—Autotroph (MPN) Method—Other Precision and Accuracy Values

For context, an empirical study using cultured organisms and the Required Method was conducted at NRLKW (this previous work was unrelated to the Proposal or this Review). The precision (CV) of microscope counts of an organism $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$, the dinoflagellate *Prorocentrum micans*, was 19% for samples with organism concentrations of approximately 10 mL⁻¹, and it was 12% for samples with concentrations of approximately 1000 mL⁻¹ (First et al., in prep.). Thus, the NRLKW study had better precision. Even given that biological data tend to be “messy”, the precision criteria stipulated in the Autotrophic Method are uncomfortably broad (<48 to <63%). Particularly in the best-case scenario, such as using cultured organisms in controlled, laboratory conditions (Table 5), a better precision should be achievable. Other organizations set more stringent thresholds for the precision of ambient organisms $\geq 10 \mu\text{m}$ and

<50 μm : <20% (the Great Ships Initiative [GSI] test facility, which has a long history of ballast-water testing [Drake et al. 2012]) or <25% (the Golden Bear test facility [Drake et al. 2012]).

Regarding accuracy, a threshold for this parameter is not set by test facilities, but in water quality parameters (dissolved and particulate organic carbon), GSI set their accuracy in recovering “spikes” (known concentrations of a given parameter) at 75-125% (Drake et al. 2012). Of course, making measurements of carbon is more straightforward than measuring living organisms. Regardless, it would be desirable for the accuracy in the Autotrophic Method (61-168%, Table 5) to be lower (e.g., 80-120%).

3.5.2.2.3 Acceptance Criteria—Autotroph (MPN) Method—Ongoing Precision and Accuracy

The Proposal indicates that Ongoing Precision and Accuracy (OPA) should be measured at the start of the BWMS testing season and every six months. One sample of *T. suecica* at a concentration of 1000 cells mL^{-1} is prepared, and three subsamples are analyzed according to procedures for the Initial Precision and Accuracy. The Proposal lists acceptance criteria as precision (CV) <63% and accuracy of 107-161%.

3.5.2.3 Acceptance Criteria—Heterotroph Method

To establish Initial Precision and Accuracy for the Heterotroph portion of the Alternative Method, 10 subsamples of a sample containing the rotifer *Brachionus plicatilis* (10 mL^{-1}) were analyzed using the Alternative Method and a comparison method: stereomicroscopy, i.e., a microscope employing white light (Trojan Marinex 2015b). Note that the latter (comparative) approach differs from the Required Method, which uses the two fluorescent markers and an epifluorescence microscope). In the Proposal, accuracy and precision were calculated as described for the Autotroph Method.

3.5.2.3.1 Acceptance Criteria—Heterotroph Method—Initial Precision and Accuracy Values and Ongoing Precision and Accuracy

For heterotrophic analysis, the Initial Precision and Accuracy acceptance criterion was set for precision <46%, and the acceptance criterion for accuracy was set at 85-95% (Miller 2015b, Table 4 therein). In the Proposal, these values were presented, and they are based on the results from NIVA (Miller and Petri 2015), which were updated in the final Alternative Method. The precision is greater than desirable (<25%, see Section 3.5.2.2.2 *Acceptance Criteria—Autotroph (MPN) Method—Other Precision and Accuracy Values*), but the range of accuracy (agreement) is acceptable (<20%), although it does not allow for values >100%. Notably, the comparison of the Alternative Heterotrophic Method was not to the Required Method, so while these data are encouraging, it is difficult to interpret them in the context of the Required Method. For a robust comparison, the Heterotroph and Required Methods should be compared.

Similar to the Autotroph Method, OPA of the Heterotroph Method was indicated. One sample with 10 subsamples is to be analyzed at the start of testing and at six-month intervals, using the

accuracy and precision thresholds in the preceding paragraph. This Review suggests the number of samples should be increased to three samples.

3.5.2.4 Factor of Agreement (FOA)

The Alternative Method stipulates an FOA between the Required and Alternative Methods for field samples (not laboratory cultures, which were used for the Initial and Ongoing Precision and Accuracy determinations). Note that this parameter was not defined by Green (1996) but was developed for the Alternative Method to assess treated discharge samples. These comparisons are similar to those used in the acceptance criteria, but the FOA is used to measure agreement of the methods in field samples—which have more variability than laboratory cultures. The FOA is calculated as the mean of the Alternative Method divided by the mean of the Required Method (but *not* multiplied by 100). Here, the Required Method is used as the benchmark. For both uptake and discharge control samples, the FOA (as defined in the Proposal) should be ≥ 0.5 , which indicates that the Alternative Method estimate is at least 50% of the estimate of the Required Method. If that threshold is met, the analyses can continue; thus, although not explicitly stated, it appears this analysis would be conducted at all test facilities.

There is no upper limit on the FOA specification, and in the Proposal, there was no justification for the value of ≥ 0.5 , other than the Required and Alternative Methods differ in their means to assess organisms and “a general expectation is that the methods should agree within a factor of two (2), with no expectation of the direction of bias. As more comparative data is collected, this factor of agreement can be refined.” (Miller 2015b). Thus, the FOA is a safety factor with a wide allowance for differences, and given the breadth of acceptable agreement between the methods, it is unclear how valuable this assessment could be. Values >1 (i.e., $>100\%$) indicate the Alternative Method yielded an estimate greater than the Required Method, which for treated discharge, would be a conservative estimate. However, since accurate concentrations are necessary to assure that the test criteria are met for the uptake and untreated discharge, an upper limit of the FOA should be defined in the Proposal.

3.5.3 Results of the Validation of the Alternative Method

The empirical experiments used to measure the accuracy, precision, and reproducibility of the Alternative Method were described within the Proposal documents (see Table 2 and Table 6 of this Review). Note that the Proposal was not framed in the format developed by Green (1996), but the work done to support the Proposal and the resulting data do allow an examination of the parameters identified by Green (1996). **Accuracy** was addressed by comparing the results combined from the Autotroph and Heterotroph Methods to the Required Method for both ballast water samples (Petri 2015b [Exp. 1-3]) and target concentrations of cultured organisms (Miller and Petri 2015 [Exp. 6]) (although, as noted above, because the true answer is unknown, these manipulations point to agreement between methods rather than accuracy). **Precision** was measured by assessing the variation among replicate readings, and **reproducibility** was addressed by conducting comparisons at different locations by different analysts. Both precision and reproducibility were addressed in all six experiments. **Specificity** was addressed by evaluating the presence and growth of culture-amenable species relative to the presence of non-culture amenable species (the specificity also affects accuracy and precision in that species that are not amenable to culture will show lower values of accuracy and precision). Changes in culturing

conditions (e.g., temperature and growth media) were used to determine the *sensitivity* of the method (Miller et al. 2015a [Exp. 4]; Miller et al. 2015b [Exp. 5]).

Table 6. Documents in the Proposal describing the validation experiments.

Parameter*	Document Title	Experiment Description	Experimental Goals [Implied in the Proposal]	Identifier
Accuracy, precision, reproducibility	MPN Method Development Experiments 1-3.pdf	Inter-laboratory comparison of the MPN dilution-culture method and fluorescein-based staining methods for the enumeration of viable or living phytoplankton cells	Measure accuracy, precision, reproducibility of ballast water samples	Petri 2015a (Exp. 1-3)
Robustness, specificity	MPN Method Development Experiment 4.pdf	MPN Method development report	Evaluate different culturing conditions (temperature and media) and growth efficiency	Miller et al. 2015a (Exp. 4)
Accuracy, specificity	MPN Method Development Experiment 5.pdf	MPN Method development report	Categorize culture-amenable species to measure the method specificity; measure accuracy by comparing the MPN Method to the Required Method	Miller et al. 2015b (Exp. 5)
Accuracy, precision, reproducibility	MPN Method Development Experiment 6.pdf	Generating method performance data for the Alternative Method for analyzing 10-50 µm organisms in the ETV generic protocol for the verification of ballast water treatment technology	Measure accuracy, precision, reproducibility of cultured organisms	Miller and Petri 2015 (Exp. 6)

*From Green (1996). MPN = Most Probable Number, ETV = Environmental Technology Verification, and Exp. = Experiment.

3.5.3.1 Method Accuracy and Precision

Frequently, results of the Autotroph (MPN) Method were reported in the Proposal as greater or less than the limit of detection. This occurs in instances where all replicates of the different dilutions showed either growth or no growth. In this case, values may be, for example, <0.18 cells mL⁻¹ (when *no* replicates exhibit growth) or >1600 cells mL⁻¹ (when *all* replicates show growth). These values are based upon the number of dilutions and number of replicate tubes.

When no tubes at any dilution show growth, the value $0.18 \text{ cell mL}^{-1}$ represents the concentration when only one of five tubes in the lowest dilution showed growth. Thus, no growth in any tube is below the limit of detection ($<0.18 \text{ cells mL}^{-1}$). Likewise, when all tubes show growth, the value $>1600 \text{ cells mL}^{-1}$ indicates that cell concentration exceeds the upper quantitation limit. In these instances, the values are considered by the Proposal to be categorical so they can be used for calculations (Petri 2015b [Exp. 1-3]). This is an unusual approach, and using this categorical result complicates mathematical operations, such as adding the results of the Autotrophic and Heterotrophic Methods.

The results from the comparison of the two methods differed when untreated or UV-treated ballast water was assessed: in untreated (control) samples, all estimates by the Autotrophic Method were greater than estimates of the Required Method (>1 order of magnitude, in several instances), but in UV-treated samples, all MPN estimates were lower than estimates of the Required Method (Petri 2015b [Exp. 1-3]). Note that this work was done prior to the finalization of the Alternative Method, and the methodologies used at the three laboratories were not standardized at the time. Nonetheless, the results illustrate the different outcomes of the Autotroph (MPN) approach and the Required Method.

Possibly, overestimations by the Autotroph (MPN) Method in untreated samples were due to the inclusion of organisms $<10 \mu\text{m}$ in the samples that are not removed by the filtration through the $10\text{-}\mu\text{m}$ pre-filter; these organisms would be excluded in the Required Method, as the microscopist would not include them in the count based on organism size. On the other hand, this overestimation should also affect the UV-treated samples, although this assessment is complicated by the fact that the Required Method will enumerate all living organisms, even those rendered non-viable by UV treatment. Indeed, the observation that the MPN numbers were lower than the Required Method in the UV-treated samples was expected according to Petri (2015b), as the Required Method can detect living cells that are not able to reproduce. Additionally, non-culturable organisms in the Autotrophic Method would decrease the counts relative to the Required Method. Note that if the two methods worked perfectly, the Required Method would *always* yield a concentration equal to or greater than the Alternative Method: the two methods should measure the same populations of living heterotrophic organisms, i.e., those displaying movement without chl *a* autofluorescence (in the Alternative Method) or those displaying FDA/CMFDA fluorescence, or movement, or both (in the Required Method). The two methods, however, measure different populations of autotrophic organisms (living vs. viable). The population of autotrophic organisms capable of reproduction will be $<100\%$ of the total number of living organisms (from the data discussed in the next Section, 3.5.3.2 Method Specificity), and, therefore, the Required Method (assuming no false negatives occur) will always yield a concentration greater than or equal to the Alternative Method. This difference between the two methods should be greatest in samples treated by UV radiation.

3.5.3.2 Method Specificity

The specificity of the Autotrophic Method was determined by the ability of phototrophic organisms ≥ 10 and $<50 \mu\text{m}$ to exhibit measurable population growth under laboratory conditions. At two test sites (DHI and NIVA), an inventory was conducted of organisms at the

test sites that were (1) found in ambient water samples, and (2) capable of demonstrating growth under laboratory conditions.

The results varied between the two locations, where 20-44% or 56-89% of species in the ambient untreated samples consistently demonstrated growth (i.e., they grew in *all* tests) using the MPN procedure at DHI and NIVA, respectively (Miller et al. 2015a [Exp. 4]). However, both test sites reported higher numbers of species that were capable of growth in cultures, when *any single test* was considered, that is, the percentages of species that have been able to grow in *at least one test* were 80-89% and 70-94% at DHI and NIVA, respectively; this higher number represents the historical record, that is, if a species could grow in *any* test over the history of testing (no reference to the extent of the historical testing was provided, e.g., 10 samples per year over 5 years). When the number of *individuals* that could grow was considered (the number was extrapolated from relative abundances within a sample), the trends were similar, with DHI reporting lower percentages than NIVA: 37-43% and 95-100% of individuals were capable of growth in all tests at DHI and NIVA, respectively, and 66-70% and 97-100% of individuals were capable of growth in any test at DHI and NIVA, respectively (Miller et al. 2015a [Exp. 4]). The difference between the facilities' results was ascribed to their methods for assessing the taxonomic diversity in the MPN arrays, as DHI (where the percent culturable was lower) analyzed one MPN tube (the middle dilution), whereas NIVA analyzed all dilutions in the MPN array. Regardless, the ability of organisms to be cultured is central to the usefulness of the MPN Method. If cells do not grow, and that lack of growth is attributed—incorrectly—to ballast water treatment, then the efficacy of the BWMS is overestimated. This issue is addressed below (Section 3.7.1.1 Key Assumptions of the MPN Method).

3.5.3.3 Agreement between the Required and Alternative Methods

Empirical measurements of FOA were reported in the Proposal from ambient samples collected and analyzed at two test facilities (Table 7 in this Review; data from Miller et al. 2015b [Exp. 5], Table 5 and the text within). In these trials, two culture media types were used (Guillard's and Keller) and two temperatures were used (10° and 20°C), with two samples of ambient water analyzed at each facility for each of the four possible media and temperature combinations. This discussion (Table 7) will focus on the results using Guillard's media, since that media is recommended in the Proposal for marine facilities, such as those participating in these trials. The FOA ranged from 0.47 to 33.88, and in 7 of 8 samples, it was >0.5. In 3 of 8 samples, the FOA was >5, meaning the Alternative Method yielded a count at least five times greater than the Required Method. This result is unexpected, since the water samples were not treated with UV radiation; in untreated samples, the Required and Alternative Methods would be expected to have similar results (FOA of ~1). It may be that the conditions (temperature, media) were not fully optimized, but the ambient community will likely change (at least seasonally), so selecting one set of conditions would be difficult, as these data show.

The highest FOA value, 33.88, occurred when the ambient temperature (8°C) was similar to the incubation temperature (10°C). This result could be due to the inclusion of organisms <10 µm in the samples that were not removed by the filtration through the 10-µm pre-filter (Section 3.3.2.1 Drawbacks—Autotrophs: Filtering). Conversely, the second-highest FOA value, 9.30, occurred

when the ambient temperature (8°C) was less than half of the incubation temperature (20°C). Regardless, these results point to the need for each TF to validate the use of the media and incubation temperature, although it is unclear how the media could be validated for shipboard testing, given the variety of potential phytoplankton. In shipboard testing, the incubation temperature could be close to the ambient water temperature, as stipulated in the Alternative Method.

Table 7. Factor of Agreement between the Required Method and the Autotroph Method (data from Miller et al. 2015b [Exp. 5], data from Table 5 therein; the column of ambient water temperature is from the text of Miller et al. 2015b [Exp. 5]).

Laboratory	Sample	Ambient Water Temperature (°C)	Guillard's 20°C	Guillard's 10°C
DHI	5:1	19	5.59	0.53
	5:2	10	0.47	0.64
NIVA	5:1	16	0.51	0.57
	5:2	8	9.30	33.88

DHI = Danish Hydrological Institute and NIVA = Norwegian Institute of Water Research.

Accuracy and precision were assessed by measuring three concentrations (10, 100, 1000 mL⁻¹) of cultured algae (*Tetraselmis suecica*) with the Required and Autotrophic (MPN) Methods (Miller and Petri 2015 [Exp. 6, Table 2]). Samples were analyzed at two locations: NIVA (with two samples, labeled NIVA-1 and NIVA-2) and DHI. The raw data were provided in Miller and Petri (2015), and for this Review, the CV, a measure of precision, was calculated (Table 8). All concentrations are reported in cells of *T. suecica* per mL. Also for this Review, the difference between the two methods (in %) was calculated as the absolute difference in concentration between the Alternative and Required Methods divided by the concentration reported by the Required Method, multiplied by 100. This value is another indication of the agreement of the two methods. The percent differences in concentrations between the methods ranged from 7 to 413%, although the difference of 413% was observed in a sample with the lowest (and most difficult to measure) target concentration, 10 cells mL⁻¹ (DHI). In this instance, the Required and Autotroph (MPN) method reported mean concentrations of 7 and 36 cells mL⁻¹, respectively. The CVs ranged from 1-26% (average = 13%) and from 20-135% (average = 50%) for the Required Method and Autotroph (MPN) Method, respectively (Table 8). The means of the CVs were significantly different ($p < 0.05$, using a non-parametric Mann-Whitney U-test [SigmaPlot, V11, Systat Software, San Jose, CA]). In all but one case, the CV of the Required Method was lower than the CV of the Autotroph (MPN) Method, meaning the Required Method had greater precision.

Table 8. Agreement and precision of the Required and Autotroph (Most Probable Number, MPN) Methods using *Tetraselmis suecica* (raw data from Miller and Petri 2015 [Exp. 6]). All values were rounded to the nearest integer. The “Difference between the Mean Values” was calculated for this Review by subtracting the Required Method Concentration mean from the Autotroph (MPN) Method Concentration mean, taking the absolute value of that number, dividing it by the Required Method Concentration mean, and multiplying it by 100.

Location and Sample	Target Conc. (mL ⁻¹)	Required Method Concentration			Autotroph (MPN) Method Concentration			Difference between the Mean Values (%)
		Mean (mL ⁻¹)	SD (mL ⁻¹)	CV	Mean (mL ⁻¹)	SD (mL ⁻¹)	CV	
NIVA-1	1000	698.0	8.7	1%	863.3	209.8	24%	24%
NIVA-1	100	60.7	2.5	4%	66.0	15.4	23%	9%
NIVA-1	10	14.0	3.6	26%	8.6	4.1	48%	39%
NIVA-2	1000	931.0	66.6	7%	993.3	630.1	63%	7%
NIVA-2	100	109.3	12.1	11%	129.7	46.4	36%	19%
NIVA-2	10	5.3	1.2	23%	8.9	1.8	20%	68%
DHI	1000	776.7	110.6	14%	1,030.0	467.7	45%	33%
DHI	100	66.3	5.0	8%	109.7	60.5	55%	65%
DHI	10	7.0	1.7	24%	35.9	48.6	135%	413%

Conc. = concentration, CV = coefficient of variation, DHI = Danish Hydrological Institute, NIVA = the Norwegian Institute of Water Research, and SD = standard deviation.

The accuracy and precision were also determined for the Heterotroph Method. Accuracy was measured in comparison to a stereomicroscope count, and precision was measured as the variability among replicate readings. The percent difference between the stereomicroscope method and the Heterotroph Method was low (5%), and the CVs were comparable (23 and 30%, respectively; data not shown; Miller and Petri 2015 [Exp. 6, Table 3]). Neither method yielded concentrations that were consistently higher, thus indicating no systematic, or directional, bias. Notably, the comparison of the Heterotroph Method was not to the Required Method, which uses epifluorescence microscopy.

The data in Table 8 are plotted in Figure 2 and Figure 3, which were prepared for this Review. Figure 2 plots the mean cell concentration for the Required and Alternative Methods. Linear regression analysis was used to measure the strength of the relationship between the two approaches, and the best-fit line was plotted. In this case, there was a significant, positive, linear relationship between the two methods ($R^2 = 0.98$, $p < 0.05$). In Figure 3, the coefficients of variation (CV) of both methods were plotted. Here, no relationship was evident, although the CV for the Required Method was lower (<30%) than that of the Alternative Method (approximately 20-65%). In both figures, the statistical analyses were conducted using SigmaPlot. In sum, there was good agreement between the methods, but the precision of the Alternative Method was greater than that of the Required Method. Note that the term “Alternative Method” is not used in either figure, as the experiment’s results (from Table 8,

describing measurements of one cultured phytoplankter) were reported using the Autotroph Method only (not a combination of the Autotroph and Heterotroph Method).

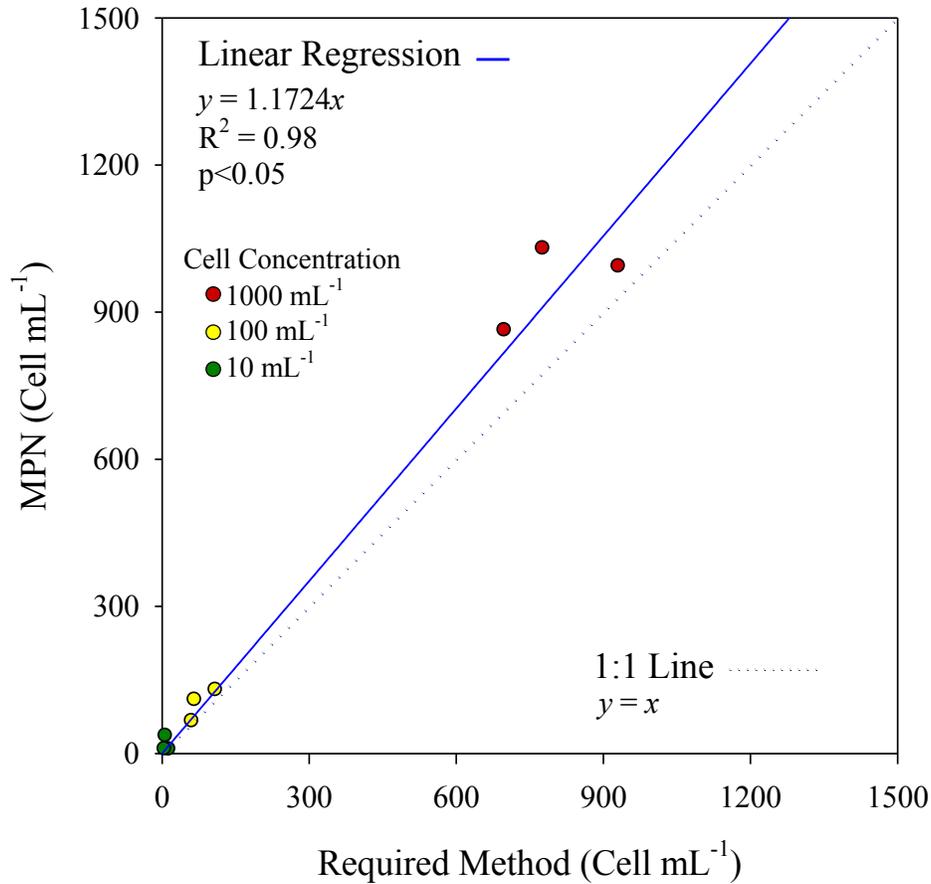


Figure 2. Comparison of the measurements of cell concentration of cultured algae performed using the Required Method and the Autotroph (Most Probable Number [MPN]) Method. Data are from Miller and Petri (2015 [Exp. 6]). The dotted line indicates a 1:1 (perfect) relationship between the methods.

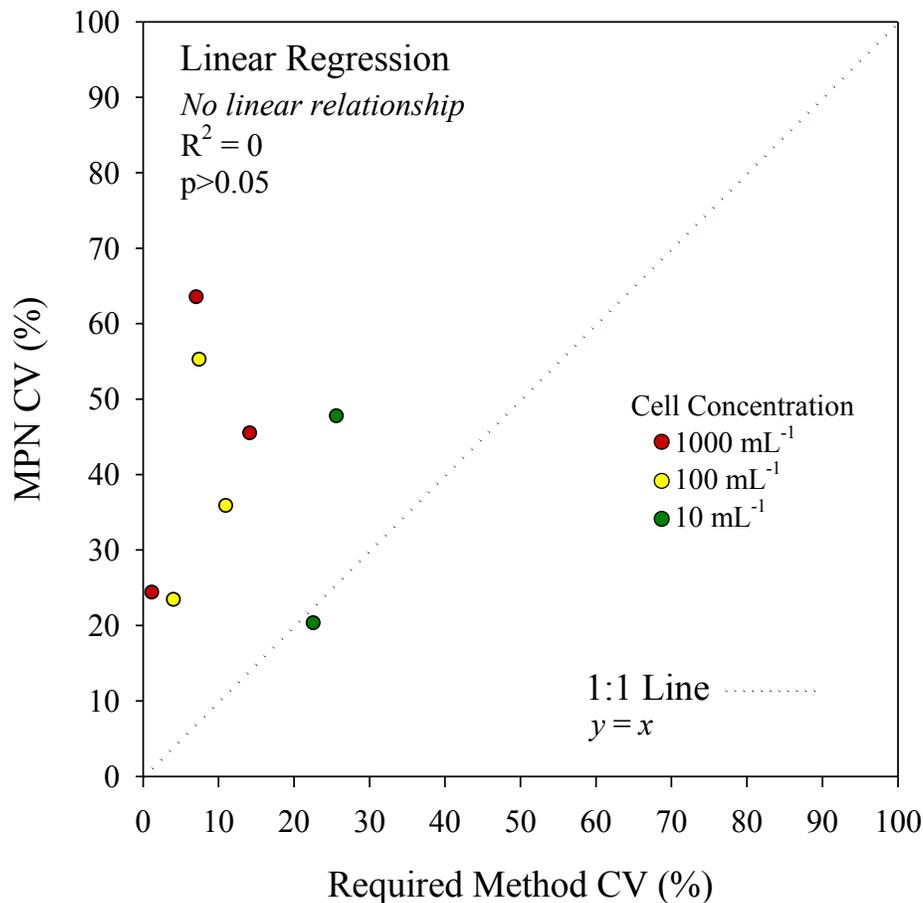


Figure 3. Comparison of the coefficient of variation (CV) of cell concentrations of cultured algae performed using the Required Method and the Autotroph (Most Probable Number [MPN]) Method. The dotted line indicates a 1:1 (perfect) relationship between the methods.

3.5.3.4 Method Reproducibility and Sensitivity

Validation tests were performed at two locations, and in the analysis of cultured organisms (Miller et al. 2015b [Exp. 5]) yielded results that were comparable with a FOA of approximately 0.5 (see Table 7 in this Review). However, in other experiments, different outcomes—in one case, with a FOA of >30—were realized based upon the culturing temperature and growth media. This result points to the need for site-specific validation.

3.6 Validation of the Required (FDA/CMFDA) Method

During the initial development of the Required Method at NRLKW, the Required Method was used to quantify samples of natural assemblages of protists collected from ambient seawater, as well as cultures of protists purchased from commercial vendors (Reed Mariculture; Campbell, California and the Provasoli-Guillard National Center for Culture of Marine Phytoplankton; West Boothbay Harbor, Maine). Cultures purchased from vendors and maintained in the NRLKW laboratory until use included the motile green flagellate *Tetraselmis* sp. (strain PLY

429); the non-motile pennate diatoms *Melosira octogona* (strain CCMP483) and *Amphora* sp.; the motile pennate diatom *Cylindrotheca closterium* (strain CCMP340); and the motile dinoflagellate *Prorocentrum hoffmannianum* (strain CCM2804). This variety of phytoplankton was chosen, as they represent a range in cell size (very close to the 10 µm threshold and larger), motility (some algae had that capability, some did not), and morphology (some were single-celled, and others were chain-formers). Various combinations of stain concentration and staining time were used, and heat killing was used to generate negative controls.

After studies at NRLKW showed good results (low false positive and false negative errors), the ETV Technical Panel decided to incorporate the method into the draft ETV Protocol. As the ETV Protocol was being reviewed and finalized, the Required Method was additionally validated in three locations: Baltimore, MD; Sequim, WA; and Boothbay Harbor, ME (Steinberg et al. 2011). The validation exercise was conducted by NRLKW researchers, who also developed the method (NRLKW will not benefit financially if the Required Method is widely used). Using this information, the Technical Panel revised this section of the Protocol, and the method was incorporated into the final draft (U.S. EPA 2010).

Among all four study sites, the results showed low false negatives on the basis of cell motility and the markers' presence (moving but non-fluorescent cells, 0-2%). It is possible, however, that non-moving, non-fluorescent cells represented false negatives. The false positive errors (heat-killed and fluorescent) were low in Florida and Maryland (5% and 3%, respectively), but high at the Washington and Maine sites (36% and 19%, with faint, but visible fluorescence from heat-killed cells, typically from heterotrophic or mixotrophic dinoflagellates). Steinberg et al. (2011) recommended that the method should be validated at each site where it was used, and at sites with high false positives, it was recommended that carefully selected fluorescence thresholds would be necessary.

Upon subsequent examination, the rate of false negatives in the study by Steinberg et al. (2011) has been questioned. While moving but non-fluorescing cells were considered false negatives, it is possible that non-moving, non-fluorescing cells (~10-25% of cells, depending on the site) could be living but non-fluorescing (i.e., false negatives). Initial trials at NRLKW with cultured organisms (some of which were non-motile) showed that cultures (i.e., cells in conditions under optimum light and nutrient conditions) did fluoresce green (i.e., were living) but did not fluoresce after heat killing (i.e., were dead). Nonetheless, the point is valid. As stated in the ETV Protocol, methods should be validated at facilities to ensure their appropriateness. Non-moving, non-fluorescing cells could be interrogated to determine if they were, indeed, dead or if they represented false negatives (for example, using the uptake of radiolabeled macromolecules).

Unpublished data by MacIntyre and Cullen (described in Miller 2015a and Miller et al. 2015e) showed 50% of 24 species of phytoplankton cultures yielded false negatives after heat killing, and 7 of 24 species of phytoplankton cultures were classified accurately (live/dead) using FDA and CMFDA markers. Notably, these laboratory experiments used flow cytometry to count living and dead cells, which is in contrast to the Required Method, which employs microscopy. This distinction is important because the microscopy allows an analyst to score an unstained but moving cell as living; a flow cytometer cannot measure movement; thus, it potentially

overestimates false negatives. On the other hand, data from a flow cytometer can be analyzed using a specific threshold (and the set point of the threshold is critical to the measurement), and thus, samples are quantified in an objective manner. These issues aside, the Required Method has been used successfully by laboratories conducting BWMS verification testing (largely using ambient communities, not cultured organisms, another difference with the MacIntyre and Cullen paper); that is, there has been good agreement with the Required Method and other methods (e.g., chl *a* fluorescence), when used. Nonetheless, this work illustrates the need for validation studies on ambient communities at each test facility and for additional work to identify ways to reduce the potential for false negatives when using the Required Method to quantify the number of live organisms. To determine the rate of false negatives for non-motile, non-fluorescing cells, a biochemical assay would be in order, perhaps using cellular uptake of radiolabelled substrates.

3.7 Statistical Review of the Alternative Method

Given the importance of statistics in the Alternative Method, it is prudent to consider the method itself and underlying assumptions of the MPN analysis. The potential biases were addressed above (Section 3.3 Potential Biases and Drawbacks of the Alternative Method). Below, the statistical grounds for the MPN Method are presented, followed by passages about the method's assumptions and calculations of its CIs.

3.7.1 Statistical Justification for the MPN Method

The MPN method is an inverse approach for estimating concentrations of living organisms; that is, observations (here, the number of MPN tubes that show positive growth after a 14-d incubation) are used to hindcast the initial conditions (here, the number of viable organisms in the original sample). First described a century ago (McCrary 1915), the method is a common microbiological assay for calculating concentrations based on observations of population growth in replicate serial dilution cultures (APPENDIX 1—MPN Statistical Theory). The MPN method has been particularly valuable for estimating concentrations of monocultures (single species or strains) of single-celled microorganisms; indeed, it was originally designed for use in this manner with bacteria. In the context of ballast water, then, the method has been proposed for use with ambient aquatic organisms, which comprise a mixed assemblage of species rather than a monoculture. In a given water body, the number of species varies geographically and seasonally, but it is on the order of dozens, if not hundreds, of species. For example, Marshal et al. (2005) identified 1400 phytoplankton species in the Chesapeake Bay, its subestuaries, and tidal tributaries.

The optimal MPN tube (usually containing a volume of 5 or 10 mL) will, as a result of serial dilutions from the original sample, contain either 0 or 1 organism, but detecting a single microorganism in the sample using traditional microscopy would not be feasible. However, since many microorganisms, such as bacteria, fungi, and algae, can reproduce via asexual cell division, and under the appropriate conditions for growth, a single cell can reproduce to concentrations that are easily detectable. For example, a bacterial colony that forms on the surface of a nutrient agar plate represents the reproduction by a single microorganism, a colony-forming unit (cfu), resulting in a large number of cells (i.e., a colony, which is visible to the

naked eye). Similarly, a suspension containing a single organism, given ideal conditions (e.g., nutrients, substrate, temperature, and light), will undergo exponential population growth. After sufficient time has occurred to undergo exponential growth and to populate a sample, bulk metrics can be used to indicate the presence of that colonizing organism. Changes in common bulk metrics, including chl *a* fluorescence for algae and turbidity for bacteria, between initial measurements and measurements after a period of culturing, denote the presence of at least one organism in the initial diluted sample capable of undergoing reproduction. Importantly, the MPN assay was developed for monocultures of bacteria (*Bacillus coli*, McCrady 1915). This foundation complicates the use of the MPN method for use with ambient, mixed assemblages of phytoplankton, because applying the method to such diverse communities may violate the assumptions of the method (as discussed in the next section).

3.7.1.1 Key Assumptions of the MPN Method

The following assumptions, which were written in reference to bacteria but are also applicable for phytoplankton, are “necessary to support the MPN method” (U.S. FDA 2010; the assumptions are also presented below in Table 9, prepared for this Review):

- “The bacteria are distributed randomly within the sample.
- The bacteria are separate, not clustered together, and they do not repel each other.
- Every tube (or plate, etc.) whose inoculum contains even one viable organism will produce detectable population growth or change.
- The individual tubes of the sample are independent.”

Table 9. Summary of the key assumptions of the Most Probable Number (MPN) method.

Assumption	Relevance	Examples of Violations
All cells must be capable of growth	Every viable organism must be capable of reproducing	Viable cells that cannot be induced to reproduce in an MPN culture*
Random distribution	Organisms must be randomly distributed in the sample, so that the probability that an organism occurs in a subsample is a function of subsample volume and its concentration	Chain- or colony-forming species
Separate cells	Organisms must be separate, un-clustered, and not repelling each other	Chain- or colony-forming species
Independence	Individual tubes are independent of each other	Samples that are not independent

*These cells do not have to be cultured in perpetuity, but they must have the capacity to show population growth over the 14-d incubation period.

The critical assumption of the MPN method is that each viable cell is capable of reproducing, so its original presence in the diluted subsample is detectable through population growth and an increase in the measured parameter over the course of the MPN incubation. The ability of an organism to reproduce must be independent of other organisms, so individuals that require conjugation for reproduction and population growth (Coats and Heinbokel 1982) or organisms that only reproduce in the presence of a symbiont (e.g., *Dinophysis sp.*, a phytoplankter that ingests and harbors algae; Myung et al. 2006) may violate the assumption of independent reproduction. Also, inhibitory factors, such as toxicity due to the presence of trace metals (e.g., Paytan et al. 2009) and viruses (e.g., Fuhrman 1999) in the cultures, may negatively affect the population growth rates of marine algae.

If every cell may not be capable of being cultured in MPN tubes, and in turn, may not be detected, then the mathematical underpinnings of the MPN calculations are not applicable. In fact, the Proposal includes data showing not all species of phytoplankton are capable of being cultured—at least under the conditions that were used—at all times (see Section 3.5.3.2 Method Specificity). This point is not addressed in the application other than to indicate it will present a small bias, approaching 0% when the historical record is considered (but that percentage is greater when the ability to grow in each test is considered). The ability of every cell to be cultured, however, is a fundamental tenant of the MPN Method, and not meeting that criterion may invalidate the use of the Alternative Method. Thus, this critical shortfall needs to be resolved. Indeed, the ETV Technical Panel has formed a task group to address this issue, and the data presented in the Proposal have been the basis for many of the task group's meetings to assess the MPN Method. The assumption that tubes are independent can be met if initial samples are independently collected (i.e., not pseudoreplicates, that is, collected from separate samples). As far as the assumptions about clumping, they can be ameliorated by mixing, given that the TF can verify that mixing is sufficient to disperse colonial organisms at that location and that the specified procedure for mixing does not lead to mortality.

Regarding the assumptions of randomness and individual cells, these are partly addressed in the discussion of biases in Table 3. The issue of positive or negative interactions within species within MPN tubes, however, is not addressed. Positive interactions (in which one species spurs the growth of one or more other species) would result in an overestimation of the number of viable cells (the MPN) in the original sample, and while that is an error, it is conservative. The error of greater concern would be an underestimation from negative interactions among species. It would be impracticable to test all of the potential interactions at a given location. Regardless, assuming the tubes in an MPN matrix with the highest dilution would contain only one cell—not interacting with other cells—the problem is largely ameliorated. Negative interactions among species could occur in the tubes at lower dilutions (and containing multiple species), but presumably, some cells would survive. Even if none survived the incubation period, the most dilute tubes would show growth (assuming they contained culturable cells), so the density of cells determined by the MPN assay would be >0 .

3.7.1.2 Confidence Intervals in the Autotroph Method

A description of the equation used to calculate MPN is found in APPENDIX 1—MPN Statistical Theory. The calculations of uncertainty for the approximation of the true density of organisms (δ) are derived elsewhere (Hurley and Roscoe 1983). When δ is defined, that is, when δ is not equal to 0 or infinity, the standard error (SE, which is calculated to show the reliability of the mean) is a function of δ , v (subsample volume) and n (number of subsamples), which is summated for the number of dilutions (m):

$$\text{Eq. 1} \quad SE_{\log\delta} = \left(\delta^2 \sum_{i=1}^m \frac{v_i^2 n_i}{e^{(v_i\delta)} - 1} \right)^{-1/2}$$

The 95% CIs are calculated below using the δ and the SE (a CI of 95% is commonly used in statistical analyses). The CI indicates the probability (in this case, 95% probability) that the *actual* δ , if known, would fall within the range of values between the upper and lower CI. The upper and lower CI (CI_{Upper} and CI_{Lower} , respectively) are derived using the following equations:

$$\text{Eq. 2} \quad CI_{Upper} = e^{(\log(\delta) + 1.96 \cdot SE_{\log\delta})}$$

$$\text{Eq. 3} \quad CI_{Lower} = e^{(\log(\delta) - 1.96 \cdot SE_{\log\delta})}$$

Typically, the upper and lower CIs are reported with MPN estimates. As an aside, other metrics of uncertainty have also been described and may also be reported. For example, the rarity index identifies unexpected outcomes—such as more dilute samples with higher growth than less dilute samples (Jarvis et al. 2010). Standard methods for MPN analyses, including methods published by the EPA (U.S. EPA 1978), the U.S. Food and Drug Administration (U.S. FDA 2010), and other scientific authorities (APHA et al. 1999), include tables that list the corresponding 95% CIs. Inclusion of these values when reporting the outcome of an MPN analysis, while not stated explicitly by the EPA or FDA, is an appropriate practice for assuring data quality (e.g., U.S. EPA 1984). Likewise, the calculators evaluated in the submitted documents all report CIs (Trojan Marinex 2015b), even though the Proposal does not advocate reporting the CIs.

As for all SE calculations, larger sample sizes (n) generally result in smaller SE. To demonstrate this, a single dilution MPN was simulated for this Review by assuming a sample volume of 0.1 mL and various numbers of tubes (Table 10). In each case, 20% of the sample tubes were positive for growth, and in all cases, the MPN (cells mL⁻¹) in the original undiluted sample was set at 2.23 viable organisms mL⁻¹. The precision of the estimate of a 5-tube MPN was limited: the range of CI spanned two orders of magnitude (from 0.31-15.9 organisms mL⁻¹). Higher numbers of sample tubes yield narrower CI ranges. The Proposal stipulates an MPN matrix should consist of 3 dilutions x 5 tubes per dilution. From Table 10, it is recommended that 10

tubes are used per dilution to ensure that an MPN value generated from an MPN table does not have a large CI, e.g., a value of 2.23 with 5 tubes had an upper CI of 15.9 (over the discharge standard) whereas a value of 2.23 with 10 tubes had an upper CI of 8.95 (below the discharge standard) (Table 10).

Table 10. Results of a theoretical matrix of MPN tubes with standard error (*SE*) and 95% confidence intervals (*CI*). In all cases, 20% of the total number of tubes (*n*) show positive growth (*p*); calculations were performed in the most simple case where there was only one sample volume (*v* = 0.1 mL).

Number of tubes (<i>n</i>)	Number of tubes with growth (<i>p</i>)	MPN (cells mL ⁻¹)	<i>SE</i>	Lower <i>CI</i> (cells mL ⁻¹)	Upper <i>CI</i> (cells mL ⁻¹)
5	1	2.23	1.00	0.31	15.9
10	2	2.23	0.71	0.56	8.95
20	4	2.23	0.50	0.84	5.96
50	10	2.23	0.37	1.20	4.16
100	20	2.23	0.22	1.44	3.46

MPN = Most probable number.

Notably, the Alternative Method does not account for the CIs generated by MPN tables. This aspect is troubling, as the CI can be relatively large (as noted above), and excluding the CIs can potentially result in a BWMS being considered to meet the discharge standard on the basis of the MPN, whereas the BWMS may not meet the discharge standard if the upper CI was taken into consideration (as is done in the statistical example in the ETV Protocol [U.S. EPA 2010, Table 12 therein]). The lack of the CI is not explained in the Method, other than by a statement that they are not used. The CI should be explained in the Alternative Method with a note that it should be reported with all results; for that matter, all data using the Required Method should also include the CI around the reported value.

4 EQUIVALENCY OF THE PROPOSED METHOD WITH THE REQUIRED METHOD

This Review’s technical opinion of the equivalency of the Alternative Method with the Required Method is provided in this section first by considering whether they measure the same endpoint. Next, the equivalency of the measurements made by the Alternative Method to the Required Method is considered. This task is accomplished by reviewing the uncertainties inherent in both the Required and Alternative Methods, chronicling (for completeness) relevant portions of the validation of the Required Method, evaluating the equivalence in which organisms are quantified by the Alternative Method (considering the methods’ precision, etc.), and reviewing the validation of the Alternative Method.

4.1 Ability of the Alternative Method to Measure the Same Proximate Aspect or Function of the BWMS

The Alternative Method and the Required Method must evaluate the same function of the BWMS. In this instance, the proximate issue is the effectiveness of the BWMS in reducing the concentration of “living” organisms in treated water to less than the criteria established in the USCG’s ballast water discharge standard. However, as discussed above, conceptually, non-viable organisms *do*, in the authors’ opinion, meet the intent of the USCG’s ballast water discharge rule, although not the letter of the regulation. Notably, the Alternative Method does not exclusively measure viability, since it combines a viability measurement (the Autotroph [MPN] Method) with a live-dead measurement (Heterotroph Method).

Of course, the fundamental tenet in equating non-viable organisms to dead organisms is that organisms scored as “non-viable” truly do not have the means to reproduce following treatment. That point has not been demonstrated unequivocally in the Proposal. The Alternative Method presumes that a sufficiently high dose of UV is provided by the treatment system that it results in permanent effective damage to reproductive ability in all organisms in all ecosystems. While UV radiation can damage cellular membranes and cytoplasmic proteins (Schwartz 1998), its primary mode of sterilization is through damage to DNA. Specifically, UV radiation causes dimerization (coupling) between pyrimidine bases, which in turn, interferes with DNA transcription and replication (Goodsell 2001; Oguma et al. 2002).

We know, however, that organisms with sublethal damage may undergo repair (Sinha and Häder 2002). Because the repair mechanisms exist, it is possible that a UV-damaged organism may be able to resume its basic cellular functions, including DNA replication and reproduction. For example, in light-dependent repair, photo-activated enzymes cleave the UV-created double bond between the pyrimidine bases, whereas in light-independent repair, UV-damaged portions of the DNA strand are excised (Goosen and Moolenaar 2008, Lesser et al. 1994). Regardless of the mechanism, if repair occurs within the MPN incubation period, a damaged organism may reproduce to reach a population size above the limit of detection, but after the conclusion of the 14-d incubation period. In this instance, the Autotroph Method would not count organisms capable of DNA repair as living organisms. This capability would confound the use of the Alternative Method (specifically, the Autotroph Method, which employs an MPN method), as it is predicated on the assumption that treated cells will be *permanently* non-reproductive, posing no risk of invasion, and thus equivalent to dead cells. The Required Method, on the other hand, does not distinguish among organisms that are (1) capable of growth and detection, (2) organisms that are potentially capable of growth and detection after repair, and (3) organisms that are not capable of growth and detection.

While a 14-d grow-out period (as stipulated in the Proposal) addresses self-repair at a given location, organisms in one geographic location may require a higher dose to achieve non-repairable damage than organisms in another location. Exploring this idea, it is important to note that planktonic organisms are not exposed to the germicidal wavelengths of UV-C (i.e., UV that kills organisms, wavelength peak = 254 nm) in the environment, as UV-C is attenuated in the atmosphere. However, UV-B (290-320 nm) does penetrate into the surface waters. In tropical

waters, the daily maximum fluence (energy per unit area) of UV-B at the surface waters is 4 W m⁻² (equivalent to 0.4 mW cm⁻²), however, approximately 90% of this UV-B radiation is attenuated in the top 5 m of the water column—even in water with low turbidity (Dunne and Brown 1996).

The realized incident dose of an organism in the surface water would depend upon its exposure time. Empirical studies have estimated the repair rate from examinations of phytoplankton growth along a series of wavelengths at various light intensities, estimating the ability of a population to repair UV damage relative to the amount of UV damage (Xing et al. 2015). The studies reveal that repair rates are species-specific (Smith and Cullen 1995), that organisms acclimated to high-intensity light—such as in shallow waters—may be more likely to undergo repair (Neale et al. 1998), and that other factors (e.g., nutrient concentrations) may influence the repair rate (Heraud et al. 2005).

The specific response of phytoplankton to germicidal UV—because this process does not occur in the environment and because, until recently, sterilization of non-pathogenic phytoplankton was not an industrial process—is not widely investigated. It is possible to extrapolate from the results of studies on environmental organisms that are acclimated to small fluences of UV radiation. That said, this Review hypothesizes that the potential for repair and regrowth following organisms' exposure to germicidal doses of UV light is minor considering that cellular adaptations, such as interfering compounds (the “UV sunscreens”) produced to attenuate UV and to mitigate damage are not optimized for UV-C (e.g., Gao and Garcia-Pichel 2011). Organisms that *could* undergo repair within the incubation period would be included within the concentration estimate, as they would have ample time to reproduce to achieve detectable concentrations. Nonetheless, it is advisable that the potential effect of self-repair on method parameters such as accuracy and precision is assessed prior to the method being approved. Particularly, empirical estimates of the maximum length of time required for repair would be useful in assessing whether the 14-d incubation period is sufficient for repair and reproduction. Note that such an experiment could be done in a straightforward manner at a land-based TF for the species that occur in one location, but making such a determination for shipboard testing, in which species from potentially all ports in the world could be in the ballast water, would be impracticable. The same issues would apply to any approach used in shipboard testing, including the Required Method.

An experiment should be conducted so the portion of the Alternative Method that measures viability (the Autotroph [MPN] Method) is compared to the Required Method using samples that are treated with a BWMS that results in immediate mortality without the possibility of repair, such as electrochlorination. Here, organisms would be treated with a method intended to kill organisms (e.g., electrochlorination) or treated with a method to render them unable to reproduce (here, irradiated with UV at doses bracketing those used in BWMS), and then all organisms would be evaluated using Required (FDA/CMFDA) and Autotroph (MPN) Methods, which would, ideally, yield the same results (no living cells and no viable cells, respectively) for organisms treated with electrochlorination and UV, respectively. Additionally, the UV samples should be re-evaluated at a later time to determine if photorepair occurred. For example, samples would be held in nutrient-rich, conditions with appropriate light for an additional incubation

period (for a total incubation of 28 d) and re-evaluated to determine if any tube previously scored as negative would show growth. While such an experiment could not capture all of the variability inherent with geography or temporal changes, results showing no growth would be a good indicator that the UV treatment rendered cells permanently non-viable.

4.2 Comparison of Uncertainties

Measurement uncertainty is inherent with any approach for quantifying living organisms. The uncertainty, however, can be estimated following standard approaches (ISO 2008). As such, all factors or variables used to calculate the concentration of living organisms ≥ 10 and $< 50 \mu\text{m}$ are considered below. Sources of uncertainty are classified as *random uncertainty*, where measurement values are normally distributed around the actual value, or as *systematic uncertainty*, where measurements diverge from the actual measurement in a direction (e.g., organisms resistant to fluorescence labeling lead to a systematic undercounting of living organisms).

Both of the approaches described in the Required Method and the Alternative (Heterotrophic) Method use manual microscopy to directly count organisms ≥ 10 and $< 50 \mu\text{m}$. The measurement variables associated with the calculation of the population concentration (P) of living organisms via direct microscope counts can be generalized in the following equation (with variables defined in Table 11):

Eq. 4
$$P = \frac{I \cdot C \cdot D}{A \cdot S}$$

Table 11. Variables used to calculate concentrations of living organisms using microscope counting. Typical values are from the Naval Research Laboratory, Key West (unpubl.).

Variable	Description	Typical Values	Uncertainty Type
I	Individual (Ind.) count	30 – 300 Ind.	Random and Systematic
C	Concentrated sample volume	50 mL	Random
D	Dilution factor	1.015*	Random
A	Aliquot volume	1 mL	Random
S	Sample volume	5 L	Random

* Dilution factor (no units) accounts for the addition of the fluorescent labels (15 μL of the dissolved fluorochrome mixture is added to 985 μL of sample water).

In a forthcoming study (First et al. in prep.), the uncertainties associated with all of these variables were estimated. In empirical studies, the accuracy and precision of volumetric measurements were estimated for graduated cylinders, flasks and beakers, and volume-displacement pipettors. The highest CV observed, 2.4%, was for a pipettor used to dispense volumes of 10 μL ; in general, CVs were $< 1\%$. Therefore, sources of uncertainty in volume measurements are minor and random, and they apply to both methods. For simplicity, they will

be omitted in the comparison between the Required and Alternative Methods. The major uncertainty is aggregated under the individual counts of organisms, which includes both random error associated with the variability of detection and systematic error. The latter includes organism loss due to mortality during the sample processing, inefficient labeling or fluorescence detection, lack of detection of motility or chl *a*, or incorrectly including or excluding organisms based upon size. These factors are addressed below.

4.2.1 Loss and Mortality

Within samples, changes in organism concentrations occur continuously due to growth, predation, and senescence; some of these dynamic processes are affected by physical conditions within the sample container in a laboratory environment that will inevitably differ from the *in situ* conditions. In either the Required or Alternative Method, analysis should occur as soon after sample collection and processing to minimize these effects. The Required Method dictates that samples are processed immediately, with a sample holding time (prior to analysis) no longer than a hold time of six hours, with the caveat that each TF should validate its holding times. The Heterotroph Method may be completed up to 48 h after collection, which is based upon a set of validation experiments reported in the Proposal (Miller and Petri 2015). This guidance should be validated at each test facility using the Alternative Method, as experiments have shown a six-hour hold time to be appropriate at another location (U.S. EPA 2010).

Sample processing, particularly filtering, can also contribute to loss and mortality. Organisms >10 µm may be forced through the mesh netting as organisms are concentrated on sieves, a process used in both methods. An additional source of loss, however, can occur during physical disruptions (due to shear caused by pressure gradients near the mesh during filtering or mechanical damage during agitation to disassociate colonial species into individual cells) and rapid changes in water characteristics (e.g., temperature, salinity, dissolved oxygen). Complete filtration—i.e., allowing the entire filtrate to drain through a sieve, leaving organisms in a thin film of water on the mesh surface—would be expected to introduce variability into the analysis. Changes in water characteristics and resulting stress to organisms would be rapid in a thin film of water with a relatively high surface area exposed to the atmosphere. Thus, the suitability of using a filtration step in the Alternative Method (and the Required Method, for that matter) should be determined.

4.2.2 Inefficient Labeling

As noted in the paper describing the Required Method (Steinberg et al. 2011), certain organisms are weakly labeled with the combination of markers that are transformed by cellular enzymes. Using motility as a definitive indicator of living organisms, the proportion of living organisms not detected by their fluorescence signal (i.e., false negatives) and heat treated organisms exhibiting weak autofluorescence (i.e., false positives) varied among locations. Rates of false negatives were <2%, but false positive rates were 3%, 5%, 15% and 36% for samples from Baltimore, MD; Key West, FL; West Boothbay Harbor, ME; and Sequim, WA, respectively (Steinberg et al. 2011). Thus, the ETV Protocol notes that before the Required Method or any alternative method is used, “it is necessary that it undergo on-site validation by preparing, examining, and analyzing ambient samples that are killed (i.e., negative controls)” (EPA 2010).

Within the Autotroph (MPN) Method, the inability of all phytoplankters to grow in MPN assays may be considered to have a similar effect as inefficient—or insufficient—labelling. This fraction of the community must be accounted for. It may be possible to accommodate these organisms in MPN calculations of the concentration of viable organisms, but there is not yet agreement on the statistical adjustments for doing so.

4.2.3 Detection of Motility

Living heterotrophs are identified in the Alternative Method by motility and the absence of chl *a* fluorescence. Because not all living heterotrophs may be visibly moving at the time of viewing, this approach would potentially lead to the exclusion of organisms, and the rate of false negatives would depend upon the relative abundance of slow moving or motionless heterotrophs. The filter set prescribed by the Alternative Method may lead to the opposite error, one of false positives (Section 3.3.3 Drawbacks—Heterotrophs).

4.2.4 Detection of Chlorophyll *a*

In the Heterotroph Method, motile organisms are viewed using epifluorescence microscopy to detect chl *a* fluorescence, and organisms with a visible fluorescence signal are excluded from the tally, since they would be autotrophs that are quantified using the Autotroph (MPN) method. When chl *a* is present but not detected, the organism is classified as a heterotroph and counted as living. This situation represents a false positive and increases the total count of organisms, which yields a conservative estimate of organism concentrations for judging whether the sample meets or exceeds the discharge standard. False negatives can occur when fluorescence of chl *a* in algae contained within the food vacuoles of heterotrophic organisms leads to the false categorizing of such organisms as autotrophs, and erroneous exclusion of the heterotrophic (algae-consuming) organisms from the total count. The likelihood of this occurrence would vary among locations and depend upon the relative abundance of herbivores, which tend not to dominate this size class. This idea is not considered in the Proposal, nor is the threshold for chl *a* fluorescence. Further, it should be restated that the filter set described in the Alternative Heterotroph Method is not optimized to detect chl *a*.

4.2.5 Size Classification

Both the Alternative and Required Methods stipulate that organisms $<10\ \mu\text{m}$ and $>50\ \mu\text{m}$ are excluded from tally. In the Required Method, microbeads with dimensions near these size thresholds may be used as visual references during microscopic examinations. The error associated with incorrect sizing of organisms is inherent to both methods, but the exact approach for categorizing organisms by size that is used in the Heterotroph Method (whether comparisons to microbeads or measuring with an eyepiece micrometer) should be detailed in the Proposal.

4.2.6 Uncertainty of the Alternative Autotroph (MPN) method

Although the MPN method is complex in design, there are actually only three values that are needed for calculating the concentration of organisms: sample volume per tube, the total number of tubes, and the number of tubes exhibiting positive (algal, in this case) growth. Measurements of volumes, as discussed above, are generally accurate and precise if calibrated pipettes are used. Similar to the random error associated with other volumetric measurements, the uncertainty of the measurement of volume is excluded, as it is small relative to other sources of error. The total number of tubes is known, so there is no uncertainty associated with this value. However, the uncertainty of the number of positive tubes results from the ability to detect reproduction within the tubes, including the degree to which the fundamental MPN assumptions (that organisms are randomly distributed and capable of growth independently to reach detectable concentrations [Cochran 1950]) are violated. These explicit assumptions and the implicit assumptions that follow are addressed Section 3.7 Statistical Review of the Alternative Method). The Proposal listed all of these considerations and provided arguments to address the major considerations. Although grazing was addressed, the possibility of viral lysis was not considered. If MPN tubes show an unusual lack of growth, tubes should be examined for the presence of virus-like-particles, which can be done relatively easily, using epifluorescence microscopy (e.g., Hennes and Suttle 1995), rather than electron microscopy. Since viruses could be numerically more abundant than algae, it is possible that multiple viruses could suppress the growth of a population. Of greater concern, however, are the inability of some phytoplankton cells to grow and the lack of standardization among calibration techniques among different models of fluorometers used at various laboratories, as discussed above.

Notably, the Alternative Method does not require that a measurement of uncertainty (e.g., CI) be reported. This omission is particularly important because the MPN is an approximation, and the uncertainty of the estimation varies greatly with the number of tubes and sample volume (e.g., Table 10 in this Review). For this reason, the uncertainty range should be reported and taken into consideration in the resultant conclusion.

To determine equivalency, the accuracy and precision of the two methods should be assessed. As reviewed above, there is no means to determine the “true” concentration of organisms, and the methods measure different endpoints (living or viable organisms), although the experiment described above using two BWMS approaches does provide insight into the relationship between living and viable cells. Additionally, precision can be used to evaluate the methods. From the data presented in the Proposal, the CV (a measure of precision) is greater (indicating lower precision) in the Alternative Method than the Required Method (Figure 3).

4.2.7 Propagation of Uncertainty

When multiple measurements are used for calculations, the uncertainty should be propagated using standard formulas (Sokal and Rohlf 1995). However, it is not possible to calculate CIs when growth is observed in all tubes (e.g., the resulting value is “>1600 cells mL⁻¹”; See Section 3.5.3.1 Method Accuracy and Precision). In this case, the inequality should be reported. This situation would most likely be a concern when the concentrations of organisms in the uptake water are calculated, presuming the Alternative Method is used for uptake samples so that they can be directly compared to measurements of the discharge samples. The reason why it would be more likely to exceed the limit of detection for uptake samples is that the ambient concentrations of phytoplankton may vary over short time periods. While discharge samples are expected to contain <10 organisms mL⁻¹, concentrations in uptake samples would be harder to predict, and it would not be evident that the predicted concentration—the basis for the dilutions used for the MPN method—was incorrect until 14 d after collection. One way to avoid this situation is to perform an initial microscope count to estimate microalgal concentrations and to then adjust the MPN dilution series accordingly; however, the Alternative Method does not recommend such a step. In the experimental data provided in the Proposal, the values listed as greater or less than a certain values were considered categorical values. Effectively, this removes the inequality sign before the value (e.g., <1600 mL⁻¹ becomes 1600 mL⁻¹), which allows for averaging and other calculations (see Section 3.5.3.1 Method Accuracy and Precision).

4.3 Summary of Equivalency Determination

Below, the results of the equivalency determination are summarized. They are informed by the various types of uncertainty, as characterized immediately above, as well as by the validation experiments outlined in the Proposal. Additionally, the statistical nuances of the MPN method as it applies to ambient phytoplankton communities are considered.

4.3.1 Equivalency—Dead to Non-Viable Organisms

The first question addressed in this investigation of equivalency is the top-order issue of equivalence of “non-viable” organisms and “dead” organisms. While this Review finds them to be conceptually equivalent, the Proposal did not demonstrate that UV-treated organisms (using doses currently employed in BWMS) would remain non-viable. A validation experiment should be conducted to compare the results of the Autotroph (MPN) Method and the Required Method using samples treated with a BWMS (or simulated BWMS) and with a treatment that results in immediate mortality without the possibility of repair, such as electrochlorination. Samples

should be evaluated in the Autotroph Approach after the 14-d incubation period and then again after another 14-d period. This approach would provide an estimate of the frequency of cells capable of repair. If the estimates from both methods are equivalent after the 28-d period, then it would support the assertion that “non-viable” (UV-treated) organisms will not undergo repair.

4.3.2 Equivalency—Precision of the Required and Alternative Methods

While the validation for the Alternative Method involved more experiments than were undertaken for the Required Method, laboratory experiments showed the precision was greater in the Required Method than the Alternative Method: the CVs were 1-26% (average = 13%) for the Required Method and 20-135% (average = 50%) for the Autotroph (MPN) method (Table 8). In all but one case, the CV of the Required Method was lower than the CV of the Autotroph (MPN) Method. Thus, using a metric typically used to compare methods, precision, the Alternative Method was shown—using the data provided in the Proposal—to be less precise than the Required Method.

4.3.1 Equivalency—Agreement of the Required and Alternative Methods

Calculations of the FOA between the two Methods showed the Alternative Method at times yielded a count at least five times greater than the Required Method. The water samples were not treated with UV radiation, and in this case, the Required and Alternative Methods would be expected to have similar results (FOA of ~1). This result could be due to the inclusion of organisms <10 µm in the samples that were not removed by the filtration through the 10-µm filter.

4.3.2 Equivalency—Applicability and Use of the Most Probable Number Method

Regarding the MPN portion of the Alternative Method, statistical questions remain unresolved. Most importantly, the number of viable but unculturable organisms is not accounted for in the calculations of the number of viable cells. This omission will require data of the percentage of non-culturable taxa from a variety of locations and the assistance of statisticians to determine if it is appropriate to develop MPN tables to account for the percentage of non-growing organisms. In fact, at this point, it is uncertain if the percentage of non-growing organisms should be those that cannot be cultured in a single test (a relatively high number) or the percentage that cannot be cultured over a series of historical tests (a relatively low number).

This Review contends that the percentage of non-culturable organisms in a *single* test is relevant, as each MPN trial is of interest, not the sum of all trials; in any given test, the percentage of non-culturable organisms, whether due to BWMS treatment or non-culturability of the organism, should be known. Additionally, the percentage of non-growers in a single test is more relevant than an historical record, considering that a given assemblage of organisms’ capacity to reproduce under the defined set of culture conditions is not predictable. If the percentage of non-growers was well constrained (i.e., predictable with a narrow confidence range), then the final estimate of cell concentration could be increased based upon the known percentage of growers, if such a practice would not violate the mathematical principles of the MPN Method. However,

this approach would require an extensive validation effort, considering that physical factors—such as tidal cycles, day and night variations in solar radiation, and seasonal trends—are superimposed upon complex, dynamic physiological and life history characteristics of the plankton community. It is likely that converging on a predictable and stable estimate of non-growers would be challenging at land-based test facilities. Determining the appropriate percentage of non-culturable organisms for shipboard testing would be even more challenging. Again, the concerns with this method in shipboard testing apply to any approach, including the Required Method.

The MPN estimates are provided in this Proposal without measures of uncertainty (CIs or SEs), which is anomalous to other uses of the MPN method. In addition, at times, MPN estimates are undefined, e.g., results were reported in the Proposal as > or < a value, and in these instances, it is unclear how these results would be applied, such as when calculating the sample mean. In addition, the process for combining the uncertainty of the MPN estimate and the heterotrophic concentration is unclear; indeed, it is unclear how the uncertainty around the heterotrophic concentration is expressed (e.g., 1 SD, 95% CI, etc.). These omissions should be rectified.

5 COMPATIBILITY BETWEEN TYPE APPROVAL TESTING AND COMPLIANCE ASSESSMENT

The Alternative Method is discussed in the Proposal in the context of its use at land-based test facilities during TA testing, and it is possible to imagine the method being used during shipboard TA testing, if the water in the ports where ballast water was taken onboard was well characterized. Its applicability in compliance testing, however, is more difficult to envision. In fact, the only time compliance testing is mentioned in any of the documents comprising the Proposal is in the research paper by Cullen and MacIntyre (2015). Here, the authors indicate the Alternative Method could be useful in shipboard and compliance testing following careful experimentation. Below, this Review considers the applicability of the Alternative Method, specifically, the Autotroph (MPN) portion, in compliance testing.

When any method is reviewed for its use in ballast water compliance testing, a number of factors must be considered. Paramount among them is the time to complete an analysis. The sampling time must also be brief, but because a sampling step is required prior to using any compliance method or tool, it will not be considered here. Since the Autotroph (MPN) Method entails a 14-d incubation period, it is uncertain how it could be used for compliance testing, which will, ideally, proceed quickly, preferably to provide an “answer” to the compliance officer while he or she is onboard a vessel.

Next (and in no particular order), compliance testing, preferably, will require minimal training and equipment. Again, it is unclear how the Autotroph (MPN) method would meet these criteria. Even if the culturing equipment were minimized, it still would require a skilled technician to properly prepare the media and evaluate the many MPN tubes.

Importantly, any reliable and robust compliance method must be able to accurately analyze ballast water taken up anywhere in the world. Thus, if the MPN method is used, the culturability

of the global community of photoautotrophs must somehow be assessed. Because not all photoautotrophs can be maintained in MPN tubes (e.g., Miller et al. 2015a), if the MPN method was to be applicable for compliance, the portion of the autotrophic community that does not grow would need to be known. Likewise, the percentage of organisms that undergo self-repair must be known, particularly, organisms that may be able to undergo repair but do not reproduce to detectable concentrations within the 14-d incubation period. These parameters would require enormous effort to determine accurately over broad geographic and spatial scales. Therefore, estimates of these parameters would need to be used, which would warrant an increase in the CI around the estimates of living organisms.

Given these hurdles, it is reasonable to assume that if the MPN method were used for BWMS TA testing, another method would be required for compliance testing. From the research completed and work underway sponsored by several Administrations, it appears that variable fluorescence of photoautotrophs in unfiltered (whole) water samples is the leading parameter being considered for use in compliance assessment (noting, of course, that other methods may be developed later). Regardless, once a method(s) for compliance testing is determined, the correspondence between the TA method (i.e., MPN) and the compliance method would need to be determined. This analysis is underway to compare fluorescence-based tools (to potentially be used in compliance testing) and the Required Method (Drake et al., 2015). Such an exercise would be warranted if the Autotroph (MPN) Method were to be used in TA testing. If the two methods did not yield the same “answer” when they were used in preliminary studies or in validation studies, it is unknown how the difference would be resolved.

6 SUMMARY

This Review provides a technical opinion regarding the Proposal submitted to USCG by Trojan Marinex to use an Alternative Method to quantify the number of living organisms $\geq 10 \mu\text{m}$ and $< 50 \mu\text{m}$. The Proposal’s practicability, applicability, clarity, equivalency with the currently Required Method, and its potential use in compliance testing were reviewed, and the findings are summarized below. Where additional information or revisions to the method are recommended, it is the opinion of the reviewers that the experimental work should not require an enormous effort. The major findings are:

6.1 Practicability and Applicability

The Proposal argued that the current method used to quantify living organisms is not practicable for UV-based BWMS intended to render organisms non-reproductive. It was discussed that it is infeasible to increase the UV dose currently used in BWMS so the Required Method would indicate cells were dead following UV treatment (thus obviating the need for the Alternative Method). Using a summary of laboratory data submitted (but as yet unpublished) to a peer-reviewed journal, the Proposal argued that in systems currently designed, a UV dose sufficient to damage cells’ enzyme systems (specifically, esterase systems)—the foundation of the Required Method—and reduce cell concentrations from $1000 \text{ cells mL}^{-1}$ to 10 cells mL^{-1} would necessitate a 10-fold increase in the UV dose. Some species of cultured algae could not be killed at the dose used in laboratory experiments, although the dose was not specified (it should be). The Proposal

goes on to state that a 10-fold increase in dose would incur a concomitant 10-fold increase in the footprint of the BWMS and its electrical power requirements, which would not be practicable to install aboard a ship. This argument, however, did not include the engineering or economic calculations to demonstrate this point. They should be provided.

6.2 Full Method Description

The method described in the Proposal was for the most part, clear and understandable. Field and laboratory experiments were summarized (and the raw data were provided) to support the assertions made in the Proposal. Further, the sequence of (the many) experiments was clear as the proposers gathered more information to refine the MPN method. Some of the documents provided were not referred to in the other documents, but this is a minor quibble, as they were nonetheless useful.

This Review identified several concerns in this portion of the evaluation; largely, they are a lack of data to validate steps in the method or a lack of standardization. The items and their potential outcomes are addressed above, so they are briefly listed below in decreasing order of importance.

- The threshold for determining growth in an MPN tube is quite low, and it will be dependent on the fluorometer's model and calibration. These items should be specified and standardized in the Proposal.
- The filter set described in the Heterotrophic Method is not optimized to detect chl *a*.
- Additional data are required to determine if filters should be removed from the MPN tubes.
- From calculations prepared for this Review, it appears the number of MPN tubes used in testing should be increased; this point should be investigated.
- Along those lines, MPN values may be inflated if organisms <10 µm are inadvertently added to the MPN tubes; this potential should be investigated.
- The CIs around the MPN results should be explained in the Alternative Method with a note that they should be reported with all results; for that matter, all data using the Required Method should also include CIs. Similarly, the means to combine the CIs for the Autotroph and Heterotroph Methods should be provided.
- The Alternative Method allows the TO to determine the approach in which samples are collected, and that should not be the case: the Alternative Method should stipulate how samples are collected.
- The means to categorize organisms by size (as part of the Heterotroph Method), whether by making comparisons to microbeads or measuring organisms with an eyepiece micrometer, should be stated explicitly in the Alternative Method.
- The number of samples analyzed for autotrophs and heterotrophs in tests of Initial and Ongoing Precision and Accuracy should be increased from one to at least three. Likewise, an upper limit of the FOA should be defined in the Proposal.

6.3 Justification of Equivalence

The Proposal did not provide adequate justification that the proposed Alternative Method is equivalent to the Required Method for three reasons:

- (1) The equivalence between dead and non-viable organisms was not demonstrated.
 - The equivalence between dead and non-viable organisms was not demonstrated and should be, e.g., in an experiment using phytoplankton that are treated to induce immediate mortality without the possibility of repair (such as electrochlorination) and treated by UV radiation, and comparing the outcomes using the Required and Autotroph (MPN) Methods over a period >14 d (to assess the capacity of cells to undergo repair). This longer assessment period is important for organisms that are capable of repair but do not reproduce to detectable levels within the span of the 14-d incubation identified in the Alternative Method. These results could be augmented or replaced with similar experiments already described in the peer-reviewed literature.
- (2) The equivalence between the Required and Alternative Methods was not demonstrated.
 - While the validation for the Alternative Method was more extensive than that undertaken for the Required Method, laboratory experiments showed the precision of the two methods was lower in the Required Method: the CVs were 1-26% (average = 13%) for the Required Method and 20-135% (average = 50%) for the Autotroph (MPN) method (Table 8). In all but one case, the CV of the Required Method was lower than the CV of the Autotroph (MPN) Method. Thus, using a metric typically used to compare methods, precision, the Alternative Method was shown—using the data provided in the Proposal—to be less precise than the Required Method. Further, empirical measurements of the FOA (using ambient communities) between two methods varied widely among different experiments. At times, the Alternative Method yielded a count at least five times greater than the Required Method. This result is surprising, since the water samples in these trials were not treated with UV radiation; in untreated samples, the Required and Alternative Methods would be expected to have similar results (FOA of ~1). This result could be due to the inclusion of organisms <10 µm in the samples that are not removed by the filtration through the 10-µm pre-filter prior to analysis with the Alternative Method.
 - In addition, in the Proposal, the Alternative Heterotrophic Method was not compared to the Required Method (it was compared to a microscopy-based method, but it was not the epifluorescence-microscopy method outlined in the Required Method). For a robust comparison, the Heterotroph and Required Methods should be compared.
- (3) Statistical questions remain unresolved.
 - Regarding the use of the Autotroph (MPN) Method, most critically, the percentage of non-culturable taxa is not accounted for in the calculations of the number of viable cells. This percentage should be known at each TF using the Alternative Method, and upon the advice of statisticians, somehow incorporated into the calculations of cell densities. It is unclear how this factor would be incorporated into tests for shipboard verification of BWMS.
 - Other statistical concerns are more easily addressed: the MPN estimates are provided without measures of uncertainty (e.g., CIs). Additionally, at times, MPN estimates are undefined, e.g., results are reported as > or < the

detection limit, and in these instances, it is unclear how these results should be applied, such as when calculating the sample mean. Finally, the process for combining the uncertainty of the MPN estimate and the uncertainty in the heterotrophic concentration is unclear (i.e., how the CIs for both numbers are used—are they combined in a straightforward fashion, by adding them together, or are additional calculations required?). These omissions should be rectified.

6.4 The Alternative Method’s Use in Compliance Testing

It seems impracticable to use the Alternative Method (specifically, the MPN portion of it) for compliance testing. If the Alternative Method was used for BWMS TA testing and another method was used for compliance testing, it would be necessary to determine the correspondence between the two methods. At this point, it appears that variable fluorescence of photoautotrophs in unfiltered (whole) water samples may be the method used to determine ships’ compliance with the discharge standard. Developing the compliance assessment method includes a careful comparison of results obtained using the compliance method and the Required Method that is used for TA testing. If the Alternative Method was considered for use in Type Approval, it would be prudent to begin a similar study to determine how the Alternative Method and a variable fluorescence-based compliance method relate to one another. If the two methods did not yield the same “answer”, it is unclear how the difference would be resolved, or how compliance by ships using UV-based systems would be assessed.

ACKNOWLEDGEMENTS

This work was funded by the U.S. Coast Guard (USCG) Environmental Standards Division (CG-OES-3), MIPR HSCG23-14-X-MMS010, Task 5.7.1, although it does not represent the official policy of the USCG. We appreciate discussions with Gail Roderick (USCG Research and Development Center [RDC]) about this report. The reviews of this report by Collin Forbes, Edward Lemieux (Branch Head, Code 6130), and Barry Spargo (Acting Superintendent, Chemistry Division, NRL) improved it—thank you.

REFERENCES

- American Public Health Association (APHA), American Water Works Association, Water Environment Federation (1999) Standard methods for the examination of water and wastewater, 19th Ed., APHA, Washington, DC
- Coats DW, Heinbokel JF (1982) A study of reproduction and other life cycle phenomena in planktonic protists using an acridine orange fluorescence technique. *Mar Biol* 67:71–79
- Cochran WG (1950) Estimation of bacterial densities by means of the “most probable number”. *Biometrics* 6:105–116
- Cullen JJ, MacIntyre HL (2015) On the use of the serial dilution culture method to enumerate viable phytoplankton in natural communities of plankton subjected to ballast water treatment. *J Appl Phycol* DOI:10.1007/s10811-015-0601-x. Application documentation submitted to USCG
- DHI (2014) MPN Assay – Analyses of algal regrowth for performance evaluation of ballast water management systems. Application documentation submitted to USCG on 30-JAN-2015
- Drake LA, Riley SC, Robbins-Wamsley SH, Molina V, Tamburri MN, First MR (2015) Validation of tools used to assess compliance with ballast water discharge standards: Update of research activities. International Council for the Exploration of the Seas/Intergovernmental Oceanographic Commission/International Maritime Organization Working Group on Ballast and Other Ship Vectors. Bergen, Norway, March 16-18, 2015
- Drake LA, Wier TP, Grant JF, Parson EWJ, Lemieux EJ (2012) Intercomparison of U.S. ballast water test facilities - final report. Report No. CG-D-06-13, United States Coast Guard Research and Development Center, Groton, CT
- Dunne RP, Brown BE (1996) Penetration of solar UV-B radiation in shallow tropical waters and its potential biological effects on coral reefs; results from the central Indian Ocean and Andaman Sea. *Mar Ecol Prog Ser* 144:109–118
- Fuhrman JA (1999) Marine viruses and their biogeochemical and ecological effects. *Nature* 399:541–548
- Gao Q, Garcia-Pichel F (2011) Microbial ultraviolet sunscreens. *Nat Rev Microbiol* 9:791–802
- Goodsell DS (2001) The molecular perspective: Ultraviolet light and pyrimidine dimers. *Oncologist* 6:298–299
- Goosen N, Moolenaar GF (2008) Repair of UV damage in bacteria. *DNA Repair* 7:353–379
- Green JM (1996) A practical guide to analytical method validation. *Anal Chem* 68:305A–309A.
- Hennes KP, Suttle CA (1995) Direct counts of viruses in natural waters and laboratory cultures by epifluorescence microscopy. *Limnol Oceanogr* 40:1050–1055
- Heraud P, Roberts S, Shelly K, Beardall J (2005) Interactions between UV-B exposure and phosphorus nutrition. II. Effects on rates of damage and repair. *J Phycol* 41:1212–1218
- Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. *Ecol Monogr* 54:187–211
- Hurley MA, Roscoe ME (1983) Automated statistical analysis of microbial enumeration by dilution series. *J App Bacteriol* 55:159–164
- International Maritime Organization (2004) Convention BWM/CONF/36 ‘International convention for the control and management of ships’ ballast water and sediments, 2004’

- International Organization for Standardization (2005) ISO/IEC 17025:2005(E), General requirements for the competence of testing and calibration laboratories, second edition, Geneva
- International Organization for Standardization (2008) Evaluation of measurement data — Guide to the expression of uncertainty in measurement. Int Organ Stand Geneva ISBN 50:134
- Jarvis B, Wilrich C, Wilrich PT (2010) Reconsideration of the derivation of most probable numbers, their standard deviations, confidence bounds and rarity values. *J App Microbiol* 109:1660–1667
- Lemieux EJ, Grant JF, Wier TP, Robbins S, Riley S, Hyland W, Davis L, Donowick T, Stringham B, Kinee WJ, Brown B (2008) Pilot environmental technology verification (ETV) test report of the Severn Trent. NRL Letter Report 6130/6098, Washington, DC
- Lesser MP, Cullen JJ, Neale PJ (1994) Carbon uptake in a marine diatom during acute exposure to ultraviolet B radiation: Relative importance of damage and repair. *J Phycol* 30:183–192
- MacIntyre HL, Cullen JJ, Petri BM, Whitsitt TJ (2015) Toward best practices for assessing the effectiveness of ultraviolet radiation for treatment of phytoplankton in ballast water. Application documentation submitted to USCG
- Marshall HG, Burchardt L, Lacouture B (2005) A review of phytoplankton composition within Chesapeake Bay and its tidal estuaries. *J Plankton Res* 27: 1083–1102
- Maurer B, Welschmeyer N (2015a) Flow cytometric analysis of the relative abundance of heterotrophs and autotrophs in the regulated 10-50 µm size class. Application documentation submitted to USCG
- Maurer B, Welschmeyer N (2015b) Rationale for the use of Most Probable Number (MPN) technique in the evaluation of UV–based ballast water management systems. Application documentation submitted to USCG
- McCrary MH (1915) The numerical interpretation of fermentation-tube results. *J Infect Dis* 17:183–212
- Miller A (2015a) 1511138 - Request for approval of alternative method (Ref 46 CFR 162.060-10 (b) (1)) supplemental information. Application documentation submitted to USCG on 02-MAR-2015
- Miller A (2015b) 1511138 - Request for approval of alternative method (Ref 46 CFR 162.060-10 (b) (1)) updated documentation. Application documentation submitted to USCG on 06-MAY-2015
- Miller A (2015c) 1511138 - Request for approval of alternative method (Ref 46 CFR 162.060-10 (b) (1)) supplemental information. Application documentation submitted to USCG on 03-JUN-2015
- Miller A (2015d) 1511138 - Request for approval of alternative method (Ref 46 CFR 162.060-10 (b) (1)) supplemental information. Application documentation submitted to USCG on 30-JUN-2015
- Miller A (2015e) 1511138 - Request for approval of alternative method (Ref 46 CFR 162.060-10 (b) (1)) supplemental information. Application documentation submitted to USCG on 23-JUL-2015
- Miller A (2015f) 1511138 - Request for approval of alternative method (Ref 46 CFR 162.060-10 (b) (1)) supplemental information. Application documentation submitted to USCG on 26-JUL-2015

- Miller A, Norlin F, Petri B (2015a) MPN method development report experiment 4. Application documentation submitted to USCG on 20-JAN-2015
- Miller A, Norlin F, Petri B (2015b) MPN method development report experiment 5. Application documentation submitted to USCG on 20-JAN-2015
- Miller A, Petri B (2015) MPN method development experiment 6: Generating method performance data for the alternative method for analyzing 10-50 μm organisms in the ETV generic protocol for the verification of ballast water treatment technology. Application documentation submitted to USCG on 31-JAN-2015
- Myung GP, Kim S, Hyung SK, Myung G, Yi GK, Yih W (2006) First successful culture of the marine dinoflagellate *Dinophysis acuminata*. *Aquat Microb Ecol* 45:101–106
- Neale PJ, Cullen JJ, Davis RF (1998) Inhibition of marine photosynthesis by ultraviolet radiation: Variable sensitivity of phytoplankton in the Weddell-Scotia Confluence during the austral spring. *Limnol Oceanogr* 43:433–448
- Oguma K, Katayama H, and Ohgaki S (2002) Photoreactivation of *Escherichia coli* after low- or medium-pressure UV disinfection determined by an endonuclease sensitive site assay. *Appl Environ Microbiol* 68:6029–6035
- Paytan A, Mackey KRM, Chen Y, Lima ID, Doney SC, Mahowald N, Labiosa N, Post AF (2009) Toxicity of atmospheric aerosols on marine phytoplankton. *Proc Nat Acad Sci* 106:4601–4605
- Petri B (2015a) Evaluating the MPN dilution-culture method for the enumeration of viable phytoplankton cells. Application documentation submitted to USCG
- Petri B (2015b) MPN method development experiments 1 to 3 inter-lab comparison of the MPN dilution-culture method and fluorescein-based staining methods for the enumeration of viable or living phytoplankton cells. Application documentation submitted to USCG on 30-JAN-2015
- Schwarz T (1998) UV light affects cell membranes and cytoplasmic targets. *J Photochem Photobiol B* 44: 91–96
- Sinha RP and Häder DP (2002) UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci* 1:225–236
- Smith RC, Cullen JJ (1995) Effects of UV radiation on phytoplankton. *Rev Geophys* 33:1211–1223
- Sokal RR, Rohlf FJ (1995) *Biometry: the principles and practice of statistics in biological research*, 3rd ed. Freeman
- Steinberg MK, Lemieux EJ, and Drake LA (2011) Determining the viability of marine protists using a combination of vital, fluorescent stains. *Mar Biol* 158:1431–1437
- Trojan Marinex (2015a) An alternative method for determining the number of living organisms in the 10-50 μm size class for ballast water management system test samples. Application documentation submitted to USCG on 31-JAN-2015
- Trojan Marinex (2015b) An alternative method for determining the number of living organisms in the 10-50 μm size class for ballast water management system test samples (document updated using the "track changes" feature in Microsoft Word). Application documentation submitted to USCG on 05-MAY-2015
- U.S. Coast Guard (2012) Standards for living organisms in ships' ballast water discharged in U.S. waters final rule, *Fed Regist* 77:172541–17321

- U.S. Environmental Protection Agency (1978) Microbiological methods for monitoring the environment: Water and wastes. Report number EPA/600/8-78-017. Washington, DC
- U.S. Environmental Protection Agency (1984) Soil sampling quality assurance user's guide. Report number EPA/600/S4-84-043. Washington, DC
- U.S. Environmental Protection Agency (2010) Generic protocol for the verification of ballast water treatment technology, version 5.1. Report number EPA/600/R-10/146, United States Environmental Protection Agency Environmental Technology Verification Program, Washington, DC <http://www.uscg.mil/hq/cg5/cg522/cg5224/docs/600r10146.pdf>
- U.S. Environmental Protection Agency (2013) Vessel general permit for discharges incidental to the normal operation of vessels (VGP) authorization to discharge under the national pollutant discharge elimination system (NPDES). Washington, DC http://www.epa.gov/npdes/pubs/vgp_permit2013.pdf
- US Food and Drug Administration (2010) BAM Appendix 2: Most probable number from serial dilutions. Washington, DC
- Xing TK, Gao, Beardall J (2015) Response of growth and photosynthesis of *Emiliania huxleyi* to visible and UV irradiances under different light regimes. *Photochem Photobiol* 91:343–349

APPENDIX 1—MPN STATISTICAL THEORY

If organisms in a sample are uniformly distributed, the number of organisms present in a subsample is a function of the volume of the subsample and the initial concentration of organisms (in the sample). When multiple subsamples (n) of a known volume (v) are drawn from a sample, the probability that a subsample has at least one organism (i.e., a positive outcome, p) can be calculated based upon the following formula (McCrary 1915, but see Cochran 1950 for the derivation):

Eq. 5
$$nv = \frac{pv}{1 - e^{(-v\delta)}}$$

Where δ is the concentration of organisms. In MPN methods, δ is unknown. However, using observations of certain sets of positive outcomes, the equation can be used to solve for the unknown δ . The instances where δ cannot be calculated are when no positive outcomes are observed (i.e., $p = 0$) and all positive outcomes are observed ($p = n$). In these cases, δ is either undefined or incalculable.

Ideally, the volume chosen for the sample concentration will have an equal probability of containing 1 or 0 organisms. As there is no way to know this *a priori*, multiple volumes are usually incorporated in an MPN method, and the volume typically represent a serial dilution of the initial sample (e.g., 10^{-1} , 10^{-2} , 10^{-3} mL). As this method was commonly applied to estimate bacterial concentrations, which can range over several orders of magnitude, the broad range of dilutions and the large differences between any two volumes were justified. A single trial with multiple dilutions—each with different values for v , n , and p —requires that all multiple equations are solved simultaneously using an iterative approach to converge on a value for δ :

Eq. 6
$$\sum_{i=1}^m n_i v_i = \sum_{i=1}^m \frac{p_i v_i}{1 - e^{(-v_i \delta)}}$$

The equation is the summation of any number of dilutions (m). The iterative approach finds the value of δ that satisfies the equation, the most probable number.