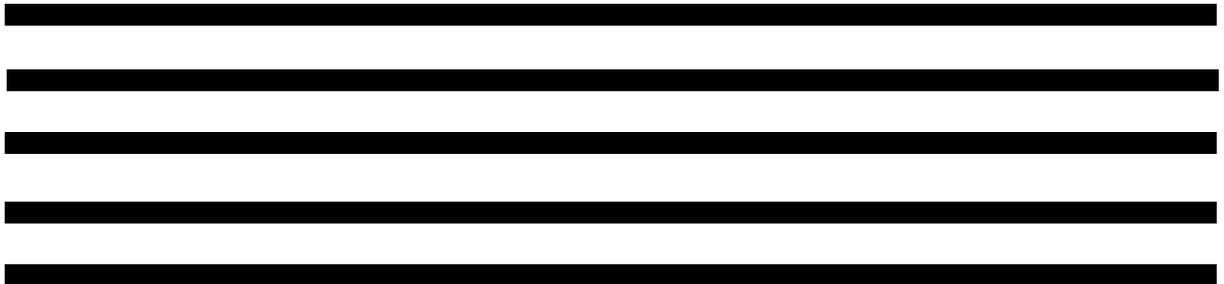




CRITERIA FOR EVALUATION OF PROPOSED PROTOZOAN DETECTION METHODS



CRITERIA FOR EVALUATION OF PROPOSED PROTOZOAN DETECTION METHODS

**Alfred P. Dufour¹
Maryann Feige²
H.D. Alan Lindquist¹, Editor
Michael Messner²
Stig Regli²
Crystal Rodgers²
Frank W. Schaefer¹, III
Steven Schaub²
James Sinclair²
Larry J. Wymer¹**

March 1999

¹Office of Research and Development

²Office of Water

CRITERIA FOR EVALUATION OF PROPOSED PROTOZOAN DETECTION METHODS

Introduction:

Currently, the only EPA approved method for detection and quantitation of protozoan cysts and oocysts in source and drinking water, is the “ICR Protozoan Method for Detecting *Giardia* Cysts and *Cryptosporidium* Oocysts in Water by a Fluorescent Antibody Procedure (ICR Microbial Laboratory Manual, Section VII, EPA/600/R-95/178, April 1996). Experience with the ICR protozoan method has shown that it usually underestimates the levels and occurrence of *Giardia sp.* and *Cryptosporidium sp.*, the two protozoan parasites that it was designed to detect. False positives and false negatives have been found during testing in a number of laboratories. The ICR method has also been criticized for being complex and difficult to perform, yielding highly variable results, demonstrating differential response in a variety of water matrices, and being costly (Clancy, et al. 1997, Schaefer, 1997). To address these issues, alternate methods have been proposed and preliminarily tested. Not all of the alternatives are complete methods. Some are method components meant to substitute for the sampling, processing, staining or detection steps used in the ICR method.

The result of the development of methods and method components has been a profusion of procedures that have not been tested in a consistent manner. Many of these methods have not been validated in a multi-laboratory, round robin format. A mechanism must be developed to evaluate these proposed methods and method components. The recent development of draft EPA method 1622 (SEP 98 DRAFT - Method 1622: *Cryptosporidium* in Water by Filtration/IMS/FA (270 KB) -(EPA 821-R-98-010) may provide a useful guideline for evaluating proposed methods. Method 1622 is a performance based method that includes multiple options for sampling and processing that may be selected.

The problem in evaluating methods is that great variances exist in the testing of the proposed methods. There are significant variations in the water matrices, the concentration of protozoan cysts, or oocysts, and the method for enumeration of these protozoan spikes, the quality of the spike material, and even the descriptions of the methods that have been provided by developers. As a result, there is little comparability in the data that have been developed for each proposed technique or method.

To evaluate the prospective utility of proposed methods and techniques it is necessary to establish a logical paradigm for categorization and preliminary evaluation to determine if further testing or investigation is warranted. Within the framework proposed, method replacements for the ICR method may be expeditiously screened. This framework is designed to be robust enough to allow comparison of widely divergent methods. This paradigm allows fair evaluation of a number of prospective proposals, allowing those with a demonstrated likelihood of being better than the ICR method to be tested further, while providing a set of necessary benchmarks for method and technique developers to accomplish before consideration for further testing. This framework will include a table that may be filled in by evaluators listing important characteristics of proposed methods. Knowing this framework will allow method developers to focus their energies on collecting the relevant data, and will guide them in the future research on method improvements.

The technical criteria for evaluation of performance of each method is based on testing in a defined water matrix with defined oöcyst populations, at a specified spiking rate, based on a standardized method of spike enumeration. All of these criteria are reasonably obtainable. Methods selected by this framework should be further tested under a yet to be developed protocol using performance evaluation criteria, and a variety of water matrices. Description of this type of testing is beyond the scope of this protocol.

Although *C. parvum* and *G. lamblia* are specifically addressed in the criteria for evaluating methods proposed here, it should be possible to use these criteria to evaluate methods proposed for detection of other protozoa. Adaptation of this method for other protozoa will have to be thoughtful. Any adaptation to include further species should include species specific parameters for preparation of seed for spiking experiments, and may require modification in other criteria as well.

It is also not possible to include in any single document, or set of criteria, all of the nuances of the application of those criteria. Evaluation of methods is ultimately an exercise in decision making, which involves factors other than a strictly technical evaluation of the methods. These criteria are intended only to help ensure that each method has an opportunity for a scientifically sound evaluation of its respective technical merit.

Finally, these criteria are designed to evaluate entire methods. Individuals wishing to have specific components of a method evaluated by the criteria outlined in this document must include those components in the context of a complete method. Any method component, however promising, does not exist independently of the entire method in which it is used. This does not preclude re-assortment of method components at a future date to develop a better method.

Technical criteria:

Statistical performance of method:

A variety of statistical measures are required to evaluate each method. These measures must be reported on the basis of having run the entire method from sampling to the detection and identification. In this way, the statistical measures will have validity and be useful for comparison. Spiked sample sets must include blanks to determine if false positive results occur.

The only acceptable method for running spiked samples is to obtain a sufficient quantity of reagent water for a typical application of the method, and to seed this quantity of water at a variety of concentrations with cysts and oöcysts from known sources maintained in defined conditions. These must be counted by hemocytometer count (EPA 821-R-98-010, September 1998), or by another method demonstrated to give results of equal or greater precision and accuracy to hemocytometer counting.

All parameters in the statistical performance of the method must be based on at least 10 spiked samples and one blank sample prepared in reagent grade water being run through the entire method. Spike material must be obtained from a documented source of laboratory animal derived cysts and oöcysts, purified by at least three levels of purification with the final method being cesium chloride purification (Arrowood and Donaldson, 1996) or equivalent. Cysts of *G. lamblia* should be 2 weeks old or less and should have at least 90% phase bright (highly birefringent) organisms, and oöcysts of *C. parvum* must be less than three months old and have

at least 80% phase bright organisms. The cysts and oöcysts should not be preserved, but should be stored in a medium for decontamination, for example water containing antibiotics. Care should be taken to avoid influencing the viability or infectivity of test organisms during purification and storage.

Spike material must be counted using a bright line hemocytometer using appropriate technique (EPA 821-R-98-010, September 1998). A suitable aliquot is then added to the sample of reagent grade water. Reagent grade water has a resistivity of 18 megaohms or greater, and is substantially free from bacterial contamination by filtration through a 0.2 micron pore sized filter. Sterility of this water must be checked at regular intervals by culture. The final spike dose is to be 10 cysts and oöcysts per liter, and the minimum sample volume should be 10 liters, although larger volumes may be tested if the method design is intended for use with larger volume applications. Spiked water samples must be used within 12 hours of enumeration of the spike material. After spiking a carboy with a seed of organisms, an inoculum of equal volume to the volume of oöcyst stock spiked into the water sample must be enumerated by a method other than the initial method of counting. This may include staining the oöcysts in this inoculum and counting stained cysts and oöcysts, or some other method of enumeration. This is to ensure the accuracy of the spike dose.

The hemocytometer counts must be recorded in a signed and dated laboratory notebook, available for inspection at the time a method is presented. Failure to present this data shall result in a score of 0 for all statistical parameters. In addition, for every 10 samples run through the complete method, at least one blank must be run. Blanks may be used for statistical significance only if at least ten blanks are run. Although blanks may be run at a higher rate than one in eleven, care must be taken to intersperse blanks and positive samples to attempt to blind the analysis. All samples and the blank should be prepared by an investigator other than the individual analyzing the data, in an attempt to operate the test in a blind fashion. In the event that the spike for any sample is prepared by an individual who is the analyst for that sample, the fact that the test was not run in a blind fashion must be reported.

This procedure is not designed for the specific testing of method components. If method components are to be tested, they should be included in a complete method. An example of this would be to include a new processing step into Method 1622, using one of the recommended filtration devices, the new method component for processing, and one of the Method 1622 analysis protocols.

Specific Criteria

A set of criteria defining significant technical parameters and characteristics to be used as evaluation measures for proposed methods was developed to determine if further testing of proposed methods should be undertaken. These parameters were separated into criteria with each criterion being assigned a number of response levels designed to be inclusive of the possible set of responses. It is understood that it may not be possible to account for every contingency when designing a comprehensive set of responses to a variable, and that good scientific judgement should be used when assigning appropriate scores to methods that fall outside the parameters of the response set for any criterion.

Requirements for Data Generation to Measure the Statistical Performance

A variety of statistical measures are required to evaluate each method. They must be reported on the basis of having conducted the entire method from sampling to detection and identification to have validity or usefulness for comparison. All parameters in the statistical performance of the method must be based on at least 10 spiked samples and one blank sample prepared in accordance with the following guidelines.

All spiked samples and blanks are to be prepared using reagent grade water with a resistivity of 18 milliohms or greater, and is substantially free from bacterial contamination by filtration through a 0.2 micron pore sized filter. Sterility of this water must be checked at regular intervals by culture. The sole exception to the use of distilled water is the case wherein the absence of particulates is an interference to the correct operation of the proposed method. In this case, chemically defined particulate may be added at a defined rate to allow the method to work.

Spike material must be obtained from a traceable source of laboratory animal derived cysts and oöcysts. A traceable source of oöcysts should be reasonably available to all qualified researchers, should maintain documentation of the species of animal from which the parasites were initially isolated, the species of laboratory animals within which the parasites have been passed, and the methods used in propagation of the parasite. Parasite spike material should be purified to a nearly homogeneous population of monodispersed parasites with final processing method of cesium chloride purification (Arrowood, and Donaldson, 1996) or equivalent. Cysts of *G. lamblia* must be less than 3 weeks old and oöcysts of *C. parvum* must be less than six months old. While these organisms should not be preserved, they should be stored in a medium designed to retard the growth of bacteria, for example, reagent grade water containing antibiotics.

All protozoan cysts or oöcysts must be enumerated by hemocytometer count, or by another method demonstrated to give results of equal or greater precision and accuracy than hemocytometer counting. Enumeration by hemocytometer refers to material being counted using a bright line hemocytometer designed for phase optics. After the hemocytometer chamber is filled with an appropriate volume and the hemocytometer placed on the microscope stage, it must be allowed to settle for 2 minutes, then protozoa are counted using phase contrast optics at no less than 200 x total magnification. Estimates are made by counting the four, 1 mm² corners of the hemocytometer grid etched into the platform. Cysts or oöcysts touching either the top or bottom, and either the right or left line are included, those touching the opposite line are excluded. At least 50, and not more than 120, cysts and oöcysts must be counted per hemocytometer platform. If cysts and oöcysts are not monodispersed, then the cyst or oöcyst preparation must be treated to provide for monodispersion (i.e. by addition of 0.01% (v/v) Tween 20 and vortexing for 2 minutes) and recounted. The average of 6 hemocytometer platforms counted in this way is taken and used to calculate the dilution required to prepare a concentration of 10 cysts and/or oöcysts per 1 liter. A suitable portion is then added to the sample of reagent grade water. The final spike dose is calculated to be 10 cysts and/or oöcysts per liter, and the minimum sample volume should be 10 liters, although larger volumes may be tested if the method design is intended for use with larger volume applications. Spiked water samples must be used within 12 hours of enumeration of the spike material.

Well documented raw data should be available for inspection at the time a method is presented. If this type of data is not available, the evaluation of statistical parameters will not be

possible.

In addition, for every 10 samples analyzed by the complete method, at least one blank must be included. Blanks may be used for statistical significance only if at least ten blanks are used. Care must be taken to intersperse blanks and positive samples. If possible, all samples and the blank should be prepared by an investigator other than the individual analyzing the data, in an attempt to operate the test in a blind fashion. In the event that the spike for any sample is prepared by an individual who is the analyst for that sample, the fact that the test was not conducted in a blind fashion must be reported.

This procedure is not designed for the specific testing of method components. If method components are to be tested, they should be included in a complete method. An example of this would be to include a new processing step into Method 1622, using one of the recommended filtration devices, the new method component for processing, and one of the Method 1622 analysis protocols.

Scoring Criteria

Statistical Performance of Method

All of the technical parameters must be adhered to in testing to develop statistical parameters. Failure to adhere to these guidelines will result in a score of 0 for all technical statistical parameters.

Percent recovery:

The percent recovery is the percent of the initial spike dose recovered at the end of a method trial. In order for percent recovery to be valid, it must be noted as a method percent recovery in which the spike was added to a water sample, and the entire method performed, from initial concentration to final sample enumeration. A spiked sample that is used to test the efficacy of a method component may provide a component percent recovery, but not a method percent recovery.

Percent recovery:	Rank
Not reported or 0%	0
<25%:	1
25-≤50%	2
50-≤75%	3
75- 100%	4

Limit of detection:

The number of cysts and oöcysts that may be reliably detected per unit volume reported as # cysts and oöcysts per liter by species is the limit of detection.

Limit of Detection:	Rank
Not reported or >10,000/l	0
10,000->1,000:	1
1,000- >100	2
100- >10	3
<10	4

Precision:

Precision relates to repeatability, and is the percent coefficient of variation (standard deviation as a percent of mean) for the method when performed in its entirety using spiked samples.

Precision:	Rank
Not reported, or $\geq 100\%$	0
<100-75%	1
<75- 50%	2
<50- 25%	3
<25%	4

Lower 95% confidence limit:

This confidence limit represents the minimum number of cysts or oöcysts that will be detected in 95 of 100 trials if the method were to be performed using the conditions specified in this document. The limit is calculated from the sample mean using the central limit theorem and is a single tailed test derived from the percent recovery data. An example calculation of this would be: the average recovery - [1.645 · (the standard deviation of the recovery/the square root of the sample size)].

Lower 95% confidence limit	Rank
Not reported, or ≤ 0	0
> 0 - 25	1
>25 - 50	2
>50 - 75	3
>75	4

Specificity:

Specificity is the true negative rate. True negatives can only be determined by using a confirmatory test. All of the cysts and oöcysts in an inoculated sample of reagent grade water should in fact be cysts and oöcysts. Therefore specificity should be a measure of the ability to confirm the identity of cysts or oöcysts. For example, in the ICR method, it may not be possible to confirm by differential interference contrast (DIC) all objects that react with the fluorescent antibody and are therefore only presumed positive. This presumptive nature may confound analysis of the results of the method. It is the nature of certain test types that confirmation may be easier or more difficult. This difference may change or be magnified as the matrix becomes more complex. Many proposed methods will not have addressed the matrix effect or ability to confirm presumptive objects. When only known positive material is added, the specificity is the ratio of confirmed to presumptive objects as a percent.

Specificity:	Rank
Not reported, or 0%	0
<25%	1
25- <50%	2
50- <75%	3
75- 100%	4

False positive rate:

The false positive rate is the apparent detection of cysts and oöcysts in known blank samples. Since, in some samples, less than 100 percent of the sample volume may routinely be analyzed, this is generally a presence/absence criterion. This is reported as the number of positives detected in blank trials randomly interspersed with positive trials.

False positive rate:	Rank
Not reported, or blank samples not performed, or false positives \geq 20%	0
False positives present but fewer than 10 blank samples were analyzed	1
No false positives present, but fewer than 10 blank samples analyzed	2
More than 10 blanks analyzed, and percent of false positives $>$ 5%, and $<$ 20%	3
More than 10 blanks analyzed and percent of false positives \leq 5%	4

Collaborative testing:

Collaborative multi-laboratory testing is exemplified by the ASTM format (D 2777-96, 1996).

If possible all statistical measures should be derived from collaborative testing. In the earliest stages of testing, there should be at least confirmation of methods and results by one or more independent laboratories. If no independent laboratory testing has been carried out the score for this category would be zero, however, independent laboratory testing to achieve a 1 or 2 score need not be in the above described multi-laboratory collaborative format.

Collaborative testing	Rank
No independent laboratory testing undertaken	0
Independent laboratory confirmation underway	1
Independent laboratory confirmation undertaken successfully	2
Collaborative multi-laboratory (ASTM-like) testing begun, not yet finished	3
Collaborative, multi-laboratory testing completed	4

Nature of Data Generated:

This is a description of the data provided by the method. It must include the following:

Organism:

This is a taxonomic description of that which is detected by the method. The taxonomic level to which a target organism is reliably identified and the method of this determination should be described. Taxonomic status of *Giardia sp.* cysts and *Cryptosporidium sp.* oöcysts may be confirmed by examination under DIC microscopy. The use of 4,6-diamidino2-phenyl-indole dihydrochloride (DAPI) staining in Method 1622 is also intended as a confirmatory technique. All antibody based methods reported to date have shown some level of cross reactivity with other species. Because of the difficulty of confirming protozoan presumptive identification, and the lack of specificity of antibody based methods, any proposed complete method must include some mechanism of confirmation of taxonomic status, beyond antibody recognition. Modern molecular methods offer the possibility of highly specific identification techniques. In the case of methods incorporating these techniques, it is necessary to ensure that the identification encompasses all isolates of a species that are infective to humans. Any new, proposed method of taxonomic confirmation other than enhanced contrast microscopy should be validated against DIC microscopy in a multi-laboratory study.

Organism:	Rank
The method to identify an organism to genus is not confirmed by multi-laboratory study, or it does not include a method to confirm the identification of the organism, or the method does not detect all strains of the target species known to be infective to humans	0
Identification to genus of the organism(s) is by non-selective methods	1
Organisms of interest are identified to genus with a selective test (that selectively identifies target organisms of interest from background particles), and an independent confirmatory test	2
Two independent testing mechanisms, at least one is valid to species level identification, and at least one is a selective test	3
Organisms of interest identified, and confirmed by an independent test as a strain(s) significant to human health	4

Viability:

Viability should always be described in terms of evolutionary significance. In this sense, a viable organism has the ability to pass genetic material on to the next generation. A demonstration of the ability to excyst represents excystation ability. The ability to infect a cell is a measure of cellular infectivity. A test that demonstrates that an organism has the ability to infect and cause clinical disease has shown clinical infectivity. A test demonstrating the production of a larger number of viable cysts or oöcysts than was present in the infective dose, or a test that directly correlates with this is a viability test. The reference method for demonstrating viability is parasite proliferation via infecting susceptible animals resulting in cyst or oöcyst production, or death of the host due to parasite proliferation, or detection of infective forms that have proliferated in the appropriate tissue on necropsy. Surrogate viability indicators are tests that have been tested alongside the reference method and have been found to be correlated with the reference method. Most surrogate viability indicators, may be found to deviate from this correlation under certain specific conditions. If the method is not designed to give viability information, then this column should be left blank.

Viability:	Rank
Viability data indicated by a surrogate viability indicator that has not been validated against the reference standard (animal infectivity model)	0
Surrogate viability indicator tested against one defined set of conditions and poorly correlated to reference standard, or found to be poorly correlated to the reference standard under conditions expected in the type of sample analyzed	1
Surrogate viability indicator data validated against reference standard, under one defined set of conditions	2
Surrogate viability indicator data validated against reference standard using numerous conditions, correlating well with the reference standard in several of these conditions	3
Surrogate viability indicator data validated against reference standard, high correlation between the test and animal infectivity found in all of a broad range of conditions tested, or method includes the reference test	4

Method Description

The description of the method must be thorough enough to allow evaluation of the completeness, availability and practicality of the method.

General description:

The general description should be sufficient to allow a scientist or technician, unfamiliar with the method, but with relevant experience to perform the method and achieve results comparable to those reported by the developers. The format used is to be determined by the method developer, but should include: scope and application, summary of method, definitions, interferences, safety, equipment and supplies, reagents and standards, sample collection, preservation and storage, quality control, calibration and standardization procedure, data analysis and calculations, method performance, pollution prevention, waste management, references, tables, diagrams, flowcharts, and validation data. The description should serve as a complete, self contained instruction set for evaluating a source or drinking water sample, from sampling through analysis, interpretation and reporting of results. Failure to include any of these steps will result in a finding that the method description is incomplete.

General description:	Rank
Incomplete, not given	0
Well described but not complete	1
Apparently complete enough to allow independent testing	2
Complete, tested by one independent laboratory	3
Complete, tested by independent laboratories, and it is written in a format recognized by the U.S. Environmental Protection Agency, or by an international consensus methods validation organization (such as the American Society for Testing and Materials or AOAC International)	4

Sampling:

The section on sampling must include information as to the range of sample volumes within which the method will perform as described including sample holding times and preservation techniques, ability to filter sample at sampling location, or at the laboratory. Other special limitations on sampling should be described here.

Sampling:	Rank
Not described, or described in general terms, or described with no test data presented, or well described but would not allow testing in suitable range of water types/volumes for practical use	0
Not completely described, some test data presented, full range of parameters unknown	1
Well described would allow water testing in an acceptable range of conditions	2
Tested by multiple laboratories, demonstrated to be effective with either source or drinking water	3
Tested in a collaborative multi-laboratory study demonstrating that sampling of both source and drinking water yields similar statistical recovery results	4

Managerial Criteria:

This element is intended to encompass the range of managerial concerns that may arise during the implementation of methods, either in a collaborative multi-laboratory effort, or during actual implementation of a method. These managerial considerations should be addressed within the methods description, however, since these considerations may exceed the requirements of the methods description format chosen by the developer, they may be included in a separate document. This information is to include a full description of: personnel required for method

performance including skill levels as well as training requirements for the personnel performing the method broken down by technique where appropriate, a list of equipment and supply providers and whether these providers meet industry quality assurance standards, and the open market availability for any critical or proprietary components.

Managerial Criteria:	Rank
If none of this information is given or it is not documented	0
If this information is partially complete, or any supporting documentation is only partially available	1
If all points are addressed, but some of the information is given in non-quantitative fashion, or includes general or anecdotal evidence	2
If all points are covered, but there is no evidence of independent confirmation of the information of some of the information	3
If all points are thoroughly addressed, and supported by independent confirmation, and all critical method components are readily available on the open market	4

Likely candidates for future research:

Certain methods or method components may not be available for immediate use, but might be likely candidates for future use. These may include methods wherein the current equipment/supply sources are limited in some way, but may become available in the future, or methods providing particular promise but still lacking critical elements that would allow a decision for further evaluation to be made. These candidates should be denoted by a non-numeric symbol. This denotation indicates that the method is unsuitable for use as an alternative to existing method/methods at this time.

A value of 0 in any of the reporting categories except viability would result in the method being sent to its originator for further information, revision, or further testing. All methods that have scored in each category are then compared for numerical score. The numerical score should have relevance to both the suitability of the method for use and for the state of preparedness of the method at the time of scoring. The higher the numerical score, the better the likelihood that the method is prepared to a sufficient state of readiness to be tested against the current reference method.

It may also be desirable to include a narrative section with a method rating, that addresses such issues as: market availability, comparative assessments with existing methods, overall cost, adaptability of the test for other organisms, etcetera, that were not addressed directly in the numerical ranking. A narrative section would provide an opportunity to specifically state any exceptional features of a proposed method, either positive or negative. Care should be taken to address both positive and negative aspects of each test being evaluated in any attached narrative.

Literature Cited:

- 61 *FR* 24354 (Information Collection Rule) - National Primary Drinking Water Regulations: Monitoring Requirements for Public Drinking Water Supplies; Final Rule.
- Arrowood, M.J., and Donaldson, K. 1996. Improved purification methods for calf-derived *Cryptosporidium parvum* oocysts using discontinuous sucrose and cesium chloride gradients. *J. Eukaryot. Microbiol* 43(5) p89S.
- Clancy, J.L., Hargy, T.M., and Schaub, S. 1997. Improved Sampling Methods for the Recovery of *Giardia* and *Cryptosporidium* from Source and Treated Water. In: 1997 International Symposium on Waterborne *Cryptosporidium* proceedings, March 2-5, 1997 Newport Beach California. Fricker, C.R., J.L. Clancy, and P.A. Rochelle, Eds. American Water Works, Denver CO. pp: 79-86.
- D 2777-96. 1996. Standard Practice for Determination of Precision and Bias of Applicable Test Methods of Committee D-19 on Water. Annual Book of ASTM Standards Vol. 11.01. ASTM West Conshohocken, PA.
- EPA 821-R-98-010 September 1998 DRAFT, Method 1622: *Cryptosporidium* in water by Filtration/IMS/FA. United States Environmental Protection Agency, Office of Water, Washington, DC 20460.
- Fout, G., Schaefer, III, F.W., Messer, J.W., Dahling, D.R., and Stetler, R.E. April 1996. ICR Microbial Laboratory Manual. EPA/600/R-95/178. United States Environmental Protection Agency, Office of Research and Development, Washington, DC 20460.
- Schaefer, III, F.W. 1997. Detection of Protozoan Parasites in Source and Finished Drinking Waters. In: Manual of Environmental Microbiology. Hurst, C.J., G.R. Knudsen, M.J. McInerney, L.D. Stetzenbach, and M.V. Walter. Eds. ASM Press. Washington D.C. pp. 153-167.

