RTI/6922/371-04F

Testing Equivalence of Microfiltration and Continuous-Backwash-Upflow, Dual-Sand Filtration Technologies

Submitted to
U.S. Environmental Protection Agency
Office of Research and Development
Quality Assurance Division

Prepared by
Roy W. Whitmore
Work Assignment Leader
Research Triangle Institute

U.S. EPA Contract 68-D4-0091
Work Assignment 97-11
RTI Task 91U-6922-371

May 13, 1998

Executive Summary

In May of 1997, the New York City Department of Environmental Protection (NYC DEP) issued new rules and regulations governing activities in NYC's watershed. Contained in the City's regulations is the requirement that all wastewater treatment plants (WWTPs) discharging in either the Catskill, Delaware, or Croton watersheds be upgraded to incorporate advanced wastewater treatment. This upgrade, by virtue of the City's regulations, must include microfiltration (assumed to be the best available technology) or an approved equivalent. Moreover, the U.S. EPA Region 2 Filtration Avoidance Determination issued in May of 1997, also requires all WWTPs discharging into NYC's Catskill and Delaware watersheds, utilize by May 1, 2002, either microfiltration or an approved equivalent.

Accordingly, NYC DEP developed and implemented protocols for a study that compares the pathogen removal capabilities of microfiltration and continuous-backwash-upflow, dual-sand (CBUDS) filtration by operating a pilot facility of each system, side by side, with the same influent and testing methods. The test units are located at the Stamford WWTP, immediately downstream of the secondary clarifier. This plant treats mostly domestic wastewater. The microfiltration unit is Memtec America Corporation's Memcor 6M10C unit. The CBUDS test unit consists of two Parkson Corporation Dynasand filters operated in series. Both the CBUDS and the microfilter test units were operated using the manufactures' standard operating procedures.

During each run of the pilot facility, the mean concentration in the spike solution tank, from which the spiking material was pumped into the influent stream, was approximately 1×10^8 units per 100 liters for both Giardia cysts and Cryptosporidium oocysts. Approximately 42 liters (11 gallons) of this solution was delivered to each effluent stream. Hence, the total numbers of Giardia cysts and Cryptosporidium oocysts delivered to each unit was approximately 4.2×10^7.

After system stabilization (based on dye flow tests), three filters were used to monitor the effluent concentration from each system for 80 minutes. The effluent monitoring filters were all absolute pore size filters (Nuclepore # 641505 filters) designed to capture *all* Giardia cysts and Cryptosporidium oocysts.

The first run of the test facility was on Monday, June 2, 1997. Additional runs were conducted every other Monday through October 20, and the final run was on Monday, October 27. This report reviews the sample size recommendations made at the conclusion of the first five runs, and presents analysis of the results from all 12 runs.

The decision regarding equivalency was formulated in terms of the following null and alternative hypotheses regarding the effluent concentrations:

H_0: $O_m \geq O_s$ or $\Delta_O = O_m - O_s \geq 0$ (same or lower effluent concentration with sand)
versus
H_A: $O_m < O_s$ or $\Delta_O = O_m - O_s < 0$ (higher effluent concentration with sand)

where O_m is the mean effluent concentration with microfiltration and O_s is the mean with CBUDS filtration. That is, the null hypothesis is that the mean effluent concentrations are equivalent, and the alternative hypothesis is that the mean effluent concentration is less with microfiltration.

The stakeholders recommended conducting the hypothesis test at the 10 percent level of significance with sufficient power to be 95 percent certain of rejecting the hypothesis of equal effluent means for both Giardia cysts and Cryptosporidium oocysts if the microfiltration mean were 1.00 Giardia cyst or Cryptosporidium oocyst per filter and the CBUDS mean were 3.00 Giardia cysts or Cryptosporidium oocysts per filter (approximately 0.5 difference in log-removals). Determining the sample size (number of runs) necessary to achieve this level of power required an assumption regarding the variability in the Giardia cyst and Cryptosporidium oocyst concentrations in the effluent. The Poisson probability distribution was used as the basis for the assumption regarding variability because the Poisson probability distribution is the distribution of the number of events that occur during a particular time period when the events occur at random and independently in time with a fixed probability of an event in a short time interval.

The Poisson assumption of equal means and variances (100 percent coefficient of variation) appeared to be consistent with the results of the first five runs. Therefore, we recommended that a minimum of 12 runs be performed to reach a conclusion regarding the equivalency of microfiltration and CBUDS filtration for removal of Giardia cysts and Cryptosporidium oocysts. NYC DEP implemented 12 runs before winter weather forced the pilot test to be terminated.

The manufacturer of the microfiltration unit, Memtec, determined that one of the six filtration membranes was leaking slightly during the initial runs, and the leaking membrane was replaced after the third run. Therefore, the microfiltration results from the first three runs, which did appear to be anomalous, were excluded from the statistical analyses.

The mean effluent concentrations of Giardia cysts and Cryptosporidium oocysts are summarized in Table 6. The microfiltration means for runs 4 through 12 are 0.009 per 100 liters for both Giardia cysts and Cryptosporidium oocysts. The CBUDS mean effluent concentrations are 0.020 Giardia cysts per 100 liters and 0.034 Cryptosporidium oocysts per 100 liters. In both cases, the result of the hypothesis test, using the Pittman Test, a non-parametric permutation test, is that the null hypothesis of no difference in overall means cannot be rejected at the 10 percent level of significance.

The only remaining question is whether or not the 9 runs for microfiltration and 12 runs for CBUDS filtration were sufficient to achieve the desired 95 percent power for detecting the difference between a mean of one Giardia cyst or Cryptosporidium oocyst per filter for microfiltration and three per filter for CBUDS filtration. Based on the observed variability in the data, it appears that this level of power was achieved for the Giardia hypothesis test. The results for the Cryptosporidium hypothesis test were not quite as definitive (i.e., there was greater variability between runs for Cryptosporidium), but it appears that we did achieve approximately 94 percent power for this hypothesis test, also.

Notation

Log	=	Base 10 logarithm
I_m	=	Influent concentration for microfiltration
I_s	=	Influent concentration for CBUDS filtration
O_m	=	Effluent concentration for microfiltration
O_s	=	Effluent concentration for CBUDS filtration
R_t	=	Log-removal for treatment t = m or s
	=	$\text{Log}(I_t) - \text{Log}(O_t)$
	=	$\text{Log}\left(\frac{I_t}{O_t}\right)$
Δ	=	Difference in log-removals
	=	$R_m - R_s$
	=	$[\text{Log}(I_m) - \text{Log}(O_m)] - [\text{Log}(I_s) - \text{Log}(O_s)]$
	=	$\text{Log}\left[\frac{I_m/O_m}{I_s/O_s}\right]$

Definitions

CBUDS: continuous-backwash-upflow, dual-sand.

cyst: specialized cell produced either in response to environmental stress or as a natural part of the life cycle of a protozoan. Giardia cysts are round to oval (8 to 18 µm long by 5 to 15 µm wide).

oocyst: encrusted zoospore that results from completion of the protozoan life cycle. Cryptosporidium oocysts are ovoid to spherical (3 to 7 µm in diameter).

(oo)cyst: Giardia cyst or Cryptosporidium oocyst.

stakeholders: New York City Department of Environmental Protection, New York State Department of Health, and U.S. Environmental Protection Agency Region 2

Table of Contents

List of Tables

List of Figures

Testing Equivalence of Microfiltration and Continuous-Backwash-Upflow, Dual-Sand Filtration

1. Background

Preventing the discharge of pathogens from wastewater treatment plants (WWTPs) is important for protection of public health. This is especially true for the parasitic protozoans Giardia and Cryptosporidium. When present in potable water, these organisms have been identified as significant contributors to waterborne outbreaks of gastrointestinal disease. These organisms have also shown themselves to be resistant to conventional disinfection methods. Hence, for the past several years, the New York City Department of Environmental Protection (NYC DEP) has been evaluating (with technical assistance from the U.S. EPA and the NYS Department of Health) the adequacy of various technologies for removing these organisms.

In May of 1997, the NYC DEP issued new rules and regulations governing activities in NYC's watershed. Contained in the City's regulations is the requirement that all WWTPs discharging in either the Catskill, Delaware, or Croton watersheds be upgraded to incorporate advanced wastewater treatment (AWT). This upgrade, by virtue of the City's regulations, must include microfiltration (assumed to be the best available technology) or an approved equivalent. Moreover, the U.S. EPA Region 2 Filtration Avoidance Determination (FAD) issued in May of 1997, also requires all WWTPs discharging into NYC's Catskill and Delaware watersheds, utilize by May 1, 2002, either microfiltration or an approved equivalent.

NYC DEP is interested in establishing criteria by which to evaluate equivalency of the various technologies for removing Giardia and Cryptosporidium. NYC DEP contends that continuous-backwash-upflow, dual-sand (CBUDS) filtration is as efficient as microfiltration for removing the pathogens Giardia and Cryptosporidium. In addition, they have maintained that CBUDSF is also less expensive, simpler to operate, and more reliable.

Through 1996, NYC DEP had conducted three series of pathogen tests on these treatment technologies to determine equivalency. After reviewing the results of these studies, the U.S. EPA determined that more information was needed to determine equivalency. Data were needed in which the two systems were run under similar conditions, at the same time, with identical measurement methods. Accordingly, NYC DEP developed and implemented protocols for a study that compares the pathogen removal capabilities of microfiltration and CBUDSF by operating a pilot facility of each system, side by side, with the same influent and testing methods.

The test units are located at the Stamford WWTP. This plant treats mostly domestic wastewater. The test unit is located immediately downstream of the WWTP's secondary clarifier. The microfiltration unit is Memtec America Corporation's Memcor 6M10C unit. The CBUDSF test unit consists of two Parkson Corporation Dynasand filters operated in series. Figure 1 provides a schematic of the test facility. Both the CBUDSF and the microfilter test units were

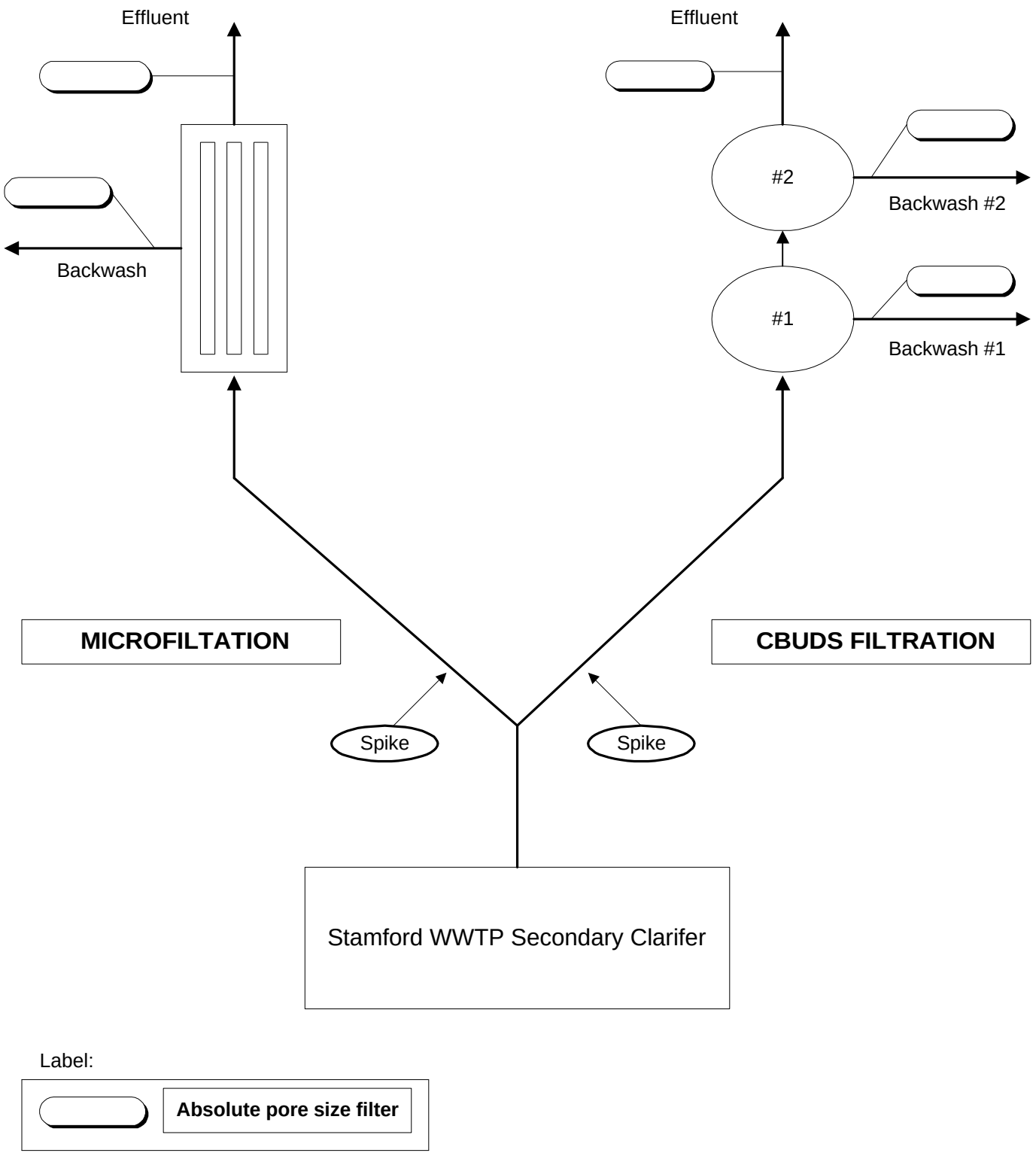

Figure 1. Schematic of Test Facility

operated using the manufactures' standard operating procedures (SOPs). For the microfiltration unit, different SOPs are applicable depending upon whether the microfilter is run manually or automated. Because the microfiltration units, if installed as part of the WWTP upgrade, will be automated systems, the microfiltration unit was operated using the SOPs applicable for NYC's Grahamsville WWTP (an automated system).

During each run, a concentration of approximately 10^8 formalin-treated Giardia cysts per 100 liters and 10^8 formalin-treated Cryptosporidium oocysts per 100 liters were uniformly injected into the influent line for each treatment system over a period of approximately 100 minutes.[1] Except for the first run, two aliquots (approximately 0.75 ml each) were taken at systematic time intervals from each spiking vat to verify the spiking concentration, and three filters were used to monitor the effluent concentration from each system for 80 minutes after system stabilization (based on dye flow tests). The monitoring filters were all absolute pore size filters (Nuclepore # 641505 filters) designed to capture *all* Giardia cysts and Cryptosporidium oocysts.

The first run of the test facility was on Monday, June 2, 1997. Additional runs were conducted every other Monday through October 20, and the final run was on Monday, October 27. The current report reviews the sample size recommendations made at the conclusion of the first five runs, and presents analysis of the results from all 12 runs.

2. Initial Formulation of the Hypothesis Test

The U.S. EPA requirement for drinking water filtration plants designed to remove Giardia cysts is a log-removal of three (e.g., from 10^8 cysts in the influent to 10^5 cysts in the effluent). Hence, the criteria initially agreed upon by representatives of the NYC DEP, NY State, and the U.S. EPA for determining equivalency was a hypothesis test that would be sufficiently powerful to be 95 percent confident of detecting a difference if the true difference in log-removals was 0.5 or more.

In this case, the question of equivalence of microfiltration and CBUDS filtration could be applied independently for Giardia cysts and Cryptosporidium oocysts using the following hypothesis testing framework:

H_o: $R_s \geq R_m$ or $\Delta_R = R_s - R_m \geq 0$ (same or more removal with sand)

versus

H_A: $R_s < R_m$ or $\Delta_R = R_s - R_m < 0$ (less removal with sand).

That is, the null hypothesis is that the treatment systems are equivalent, and the alternative is that the log-removal with the CBUDS filtration system is less than that with microfiltration.

[1] Live (oo)cysts would pose an unacceptable health risk both to the public and to the project staff.

When a decision is made based on this hypothesis test (or any other hypothesis test), there are two possible errors:

(1) Type I error—Reject H_o when it is true: Decide that $\Delta_R < 0$ when, in fact, $\Delta_R \geq 0$; alternatively, declare the systems not equivalent when they are, in fact, equivalent.

(2) Type II error—Fail to reject H_o when it is false: Decide that $\Delta_R \geq 0$ when, in fact, $\Delta_R < 0$; alternatively, declare the systems to be equivalent when they are, in fact, not equivalent.

Protecting against Type I error protects against requiring installation of microfiltration systems when they are not necessary. Hence, it is desirable for the probability of a Type I error, the significance level of the test (α) to be small. Alternatively, protecting against Type II error protects against installation of CBUDS filtration systems when they are not as effective as microfiltration. Hence, it is important that the probability of a Type II error (β) be small when the true difference, Δ_R, is the smallest difference in effectiveness that is considered to have potentially important public health consequences. Therefore, the EPA Region 2 objective is to require that the probability of a Type II error (β) be 0.05 or less (i.e., that the power of the test be 0.95 or more) when the true difference in log-removals is $\Delta_R = 0.5$ or more.

The pilot test will subject each system (microfiltration and CBUDS) to approximately the same massive loading concentration (10^8 per 100 liters) of Giardia cysts and Cryptosporidium oocysts. What does it mean for the difference in log-removal to be 0.5 when the influent concentrations of (oo)cysts will be essentially constant for the two treatment systems?

When the influent concentrations are the same for the two systems (i.e., $I_m = I_s$), the difference in log-removals depends only on the effluent concentrations. The difference in log-removals is then

$$\begin{aligned} \Delta_R &= \text{Log}\,(O_s / O_m) \\ &= \text{Log}\,(O_s) - \text{Log}\,(O_m)\,. \end{aligned}$$

Hence, when the influent concentrations are equal, a test of equivalent log-removals is actually a test of equal effluent concentrations.

The pilot test subjects each treatment system to a massive dose of protozoans (approximately 5×10^5 Giardia cysts and 5×10^5 Cryptosporidium oocysts per minute for 100 minutes). If the null hypothesis is true, each treatment system removes virtually all of the protozoans. Therefore, the variability among log-removal observations, or any other metric based on the difference between influent and effluent concentrations, will be dominated by the variability among the influent measurements. Therefore, a statistical test based directly on the log-removal observations, or any other difference between influent and effluent concentrations, will be a test mainly of differences between influent concentrations. Clearly, that is not the objective of the pilot plant study.

3. Revised Formulation of the Hypothesis Test

Therefore, the hypothesis test was reformulated in terms of effluent concentrations, as follows:

H_0: $O_m \geq O_s$ or $\Delta_O = O_m - O_s \geq 0$ (same or lower effluent concentration with sand)
versus
H_A: $O_m < O_s$ or $\Delta_O = O_m - O_s < 0$ (higher effluent concentration with sand).

That is, the revised null hypothesis is that the effluent concentrations are equivalent, and the alternative hypothesis is that the effluent concentration is higher with CBUDS filtration (assuming equal influent concentrations).

Another, more statistical reason for using this revised formulation of the hypothesis test is that the influent and effluent concentrations (for a system) are not likely to be correlated when the systems are challenged by nearly identical massive influent concentrations, and the systems remove virtually all of the protozoans. We can test for such correlations, but the physical conditions suggest that influent and effluent will be virtually uncorrelated. If they were correlated, then a test based directly on a measure of the differences, like log-removals, would be more appropriate.

Nevertheless, we used log-removal considerations to determine the difference in effluent concentrations that the hypothesis test should be able to detect. As noted above, when the influent concentrations are the same for the two systems (i.e., $I_m = I_s$), the difference in log-removals can be expressed as

$$\Delta_R = \text{Log}(O_s / O_m) = \text{Log}(O_s) - \text{Log}(O_m) .$$

Hence, when the influent concentrations are identical, the effluent concentration with CBUDS treatment, O_s, can be expressed as a function of the microfiltration effluent concentration, O_m, and the log-removal difference, Δ_R, as follows:

$$O_s = \text{Antilog}[\text{Log}(O_m) + \Delta_R] .$$

Hence, for a difference of $\Delta_R = 0.5$ log-removals, we can calculate the mean CBUDS effluent concentration for various assumed mean microfilter effluent concentrations. Using this correspondence, (oo)cyst concentrations that correspond to a difference of 0.5 log-removals are presented in Table 1 in terms of two concentration scales — (oo)cysts per 100 liters and (oo)cysts per 435 liters, the target volume corresponding to 80 minutes of monitoring for each effluent filter.

Table 1. Microfilter and CBUDS Effluent Concentrations Corresponding to 0.5 Difference in Log-Removals When the Influent Concentrations are Identical

Microfilter Concentration, O_m		CBUDS Concentration, O_s	
((oo)cysts / 100 l)	((oo)cysts / 435 l)	((oo)cysts / 100 l)	((oo)cysts / 435 l)
0.19	0.83	0.60	2.6
0.21	0.91	0.66	2.9
0.23	1.0	0.73	3.2
0.25	1.1	0.79	3.4
0.50	2.2	1.6	6.9
0.75	3.3	2.4	10
1.00	4.4	3.2	14

One can get a better understanding of the meaning of these differences in mean (oo)cyst concentrations by considering not only the differences in means but also the differences in the likely distributions of (oo)cyst counts when 435 liters of effluent are filtered. The Poisson probability distribution is the distribution of the number of events that occur during a particular time period when the events occur at random and independently in time with a fixed probability of an event in a short interval time. Therefore, the Poisson distribution would appear to be ideal for modeling the number of (oo)cysts captured on a filter under the conditions of the field experiment, given that each effluent filter is filtering approximately 435 liters (115 gallons) of water. The Poisson probability distributions corresponding to the mean (oo)cyst counts per 435 liters (per filter) shown in Table 1 are presented in Appendix A and serve to illustrate the differences in distributions of counts per filter that correspond to the differences in means shown in Table 1.

For example, Table 1 shows the following. If the mean number of (oo)cysts on the microfiltration effluent filter, O_m, is 1.0, then the mean number of (oo)cysts on the CBUDS effluent filter, O_s, corresponding to a difference of 0.5 log-removals is 3.2. Tables A.3 and A.6 show the corresponding differences in the distributions of the numbers of (oo)cysts captured by the effluent filters. For example, if the mean effluent filter count is 1.0, then the probability of observing 5 or more (oo)cysts on a filter is only 0.004 (assuming a Poisson distribution). However, if the mean effluent filter count is 3.2, then the probability of observing 5 or more (oo)cysts on a filter is 0.219 (1 - 0.781).

4. Sample Size Recommendations Based on the First Five Runs

Determining the appropriate sample size for the hypothesis test requires first determining what statistical procedures will be used to perform the hypothesis test. We can then determine the sample size needed for that test to achieve the desired power.

A test based on dependent samples (like a paired t-test) would be appropriate if the pairs of observations from the two treatment systems were correlated (across runs). That could occur if time-varying characteristics of the wastewater stream resulted in both systems yielding relatively low effluent concentrations of protozoans for some runs and relatively high concentrations for other runs. Although such an outcome would be possible, we did not expect that the effluent protozoan counts would be affected by ancillary characteristics of the influent stream. If the effluent counts are not correlated, tests based on independent samples are more powerful. Hence, sample size requirements were calculated assuming that the microfiltration and CBUDS filtration effluent means would not be correlated.

If they are correlated, a test based on influent/effluent differences (equivalently, a randomized block analysis of variance) would be required and a larger sample size might be needed because of the smaller number of degrees of freedom available for estimating the residual variance.[2] For example, n runs would result in n pairs of observations and (n-1) degrees of freedom for a paired t-test. However, if the observations are uncorrelated, an independent-sample t-test with 2 (n-1) degrees of freedom for error can be used, resulting in a more powerful test for the difference between treatment means.

Nonparametric statistical hypothesis tests are more defensible when sample sizes are not large enough to validate the (parametric) assumptions underlying conventional, parametric hypothesis testing procedures. Initial sample size calculations indicated that a relatively small sample (e.g., 10 to 20 runs) would be sufficient for the present problem. Therefore, we proceeded with sample size calculations assuming that nonparametric hypothesis testing procedures would be required. The "randomization test for two independent samples," an application of Fisher's exact test, was considered to be a likely candidate.[3]

Sample size requirements for nonparametric hypothesis tests are usually determined based on the efficiency of the tests relative to comparable normal theory tests. Therefore, we derived sample size requirements by calculating the power of the two-sample t-test. We chose the smallest sample size that achieved the required power of 0.95 and inflated the calculated sample

---

[2] For general protocols for testing equivalence of treatment systems, we recommend that sample size requirements be calculated assuming a randomized block analysis of variance to protect against the possibility that the influent and effluent concentrations may be correlated.

[3] See Section 7.4 of Conover, W.J. (1971). Practical Nonparametric Statistics. Wiley, New York, NY, pp. 360-364.

size by 15 percent to account for using a nonparametric test. The power was calculated for a specified difference, δ, in mean filter counts (corresponding to a 0.5 difference in log-removals) as follows[4]:

$$\text{Power} = \text{Prob} \{ T(\nu, \Delta_O) > t_{1-\alpha}(\nu) \} ,$$

where T is a noncentral Student's t random variable, $t_p(\nu)$ is the p-th percentile of the (central) Student's t distribution with ν degrees of freedom,

$$\Delta_O = \frac{\delta}{\sqrt{\frac{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}{n}}} ,$$

$$\nu = 2 (n - 1) ,$$

and $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are estimates of the variances of the effluent filter counts from the two treatment systems.

Sample size requirements resulting from application of this formula, allowing for a 15 percent inflation of the required sample size to account for using nonparametric hypothesis testing procedures, are presented in Table 2. To illustrate the computations underlying Table 2, consider the entries on the first row. This row addresses the situation where the mean number of (oo)cysts per 100 liters is 0.19 for microfiltration, which corresponds to a mean of 0.83 (oo)cysts on each filter, which has filtered 435 liters of effluent. If, under the alternative hypothesis, the CBUDS filtration system has a log-removal that is 0.5 smaller, its mean number of (oo)cysts per 100 liters will be 0.60, which corresponds to a mean of 2.6 (oo)cysts on the filter after 435 liters of effluent have been filtered. Assuming that the number of (oo)cysts captured by a filter follows a Poisson distribution (for which the mean and variance are equal, or the CV is 100 percent), the estimated variances used to calculate the sample size requirements were: $\hat{\sigma}_1^2 = 0.83$ and $\hat{\sigma}_2^2 = 2.6$. Given a test conducted at the 10 percent level of significance and a minimum detectable difference of $\delta_1 = 2.6 - 0.83 = 1.77$ (for 0.5 difference in log-removal), the smallest sample size that achieved 95 percent power was 12 runs, allowing for 15 percent inflation in the sample size to account for the fact that we plan to use a nonparametric hypothesis testing procedure.

[4] See Section 4.2.2 of Desu, M.M. and Raghavaroo, D. (1990). Sample Size Methodology. Academic Press, Beston, MA, pp. 58-59.

Table 2. Sample Size Requirements

Microfilter		CBUDS		Prob. of Type I Error(α)	Prob. of Type II Error (β)	Number of Runs Required (Sample Size)
(Oo)cysts per 100 l	Mean count on filter	(Oo)cysts per 100 l	Mean count on filter			
0.19	0.83	0.60	2.6	0.05 0.10	0.05 0.05	15 12
0.21	0.91	0.66	2.9	0.05 0.10	0.05 0.05	14 11
0.23	1.0	0.69	3.0	0.05 0.10	0.05 0.05	14 12
0.23	1.0	0.73	3.2	0.05 0.10	0.05 0.05	13 10
0.25	1.1	0.79	3.4	0.05 0.10	0.05 0.05	12 10
0.50	2.2	1.6	6.9	0.05 0.10	0.05 0.05	7 6
0.75	3.3	2.4	10	0.05 0.10	0.05 0.05	6 5

Shaded row represents approximately 0.48 difference in log removals.
Type I Error: Reject equivalency when the systems actually are equivalent.
Type II Error: Accept equivalency when the systems actually are not equivalent.
Assumes: Mean and variance of the number of Giardia cysts and Cryptosporidium oocysts on a filter are equal, consistent with a Poisson process for filter loading.

The ability to conduct the proposed hypothesis test regarding the number of (oo)cysts captured on the effluent filters is highly dependent on the ability to accurately measure the number of (oo)cysts captured on each filter. Therefore, NYC DEP implemented the following stakeholder recommendations:

- that 100 percent of the pellet prepared from each effluent filter be characterized. NYC DEP was able to comply with this request because both pilot plant treatment systems performed so well that only a small amount of solid materials were being released. This eliminated one source of extraneous variation: sampling from the pellet and statistically extrapolating to the entire pellet.

- that multiple filters (at least two) be used to characterize the effluent from each system. NYC used three effluent filters on each system for each run after the first to reduce the variability of the effluent measurements and increase the power of the test.

The key assumption underlying the sample size recommendations was that the variance of the (oo)cyst counts on the effluent filters could be based on an assumed Poisson distribution, in which case the mean and variance of the counts would be equal. The sample means and variances calculated from the first five runs are summarized in Table 3. The results from the replicate filters were first averaged for each run, then the means and variances were calculated from these five values. This was necessary to obtain an unbiased estimate of the variance between independent observations of the process (needed for the sample size calculations) because replicate filters for the same treatment process and run are not independent observations.

Table 3. Observed Means and Variances for the Effluent Filter Counts for the First Five Runs

	Giardia Cysts		**Cryptosporidium Oocysts**	
	Microfiltration	**CBUDS**	**Microfiltration**	**CBUDS**
Mean	0.43	0.07	1.10	0.30
Variance	0.44	0.02	2.30	0.45
$s^2/\bar{x}$	1.02	0.29	2.09	1.50
Sample Size	5	5	5	5

Although five runs was a very small data base for testing the assumption of equal means and variances, the data in Table 3 suggested that this assumption was not unreasonable for

calculating sample size requirements, at least for Giardia cysts.[5] The observed means shown in Table 3 suggested that we should enter Table 2 to determine the sample size requirement using a mean microfiltration filter count, O_m, somewhere between 0.43 and 1.10 (oo)cysts per filter. However, the number of runs required went well beyond the approximately 10 runs that NYC believed could be performed, given the available resources and the winter weather conditions, if we used means and variances less than 1.00.

If the true mean filter count for either system is one or less, then the treatment system has performed well, given the massive protozoan load to which it has been subjected. Therefore, we recommended that the U.S. EPA require a test with sufficient power to be 95 percent certain of rejecting the hypothesis of equal effluent (oo)cyst means if the microfiltration mean were 1.00 (oo)cyst per filter and the CBUDS mean were 3.00 (oo)cysts (approximately 0.48 difference in log-removals). We see from Table 2, that this criterion translates to a requirement of 12 to 14 runs, assuming equal means and variances (per the Poisson distribution). Given that the stakeholders were willing to accept, *a priori*, a hypothesis test conducted at a 10 percent significance level,[6] we recommended that a minimum of 12 runs be performed to reach a conclusion regarding the equivalency of microfiltration and CBUDS filtration for removal of Giardia cysts and Cryptosporidium oocysts.

Given that 12 runs was still a small sample size, we recommend that a variant of Fisher's exact test be used to test the hypothesis that the mean effluent filter counts are equal for the two treatment systems. As discussed earlier, the randomization test for independent samples is the recommended hypothesis testing procedure. This test can be implemented using the Pitman Test for two independent samples in the *StatXact 3* statistical software package for exact nonparametric inference.

The Pitman Test does not accommodate replicate observations (multiple filters per run). Therefore, we recommended that an analysis variable be computed by combining the results for the three effluent filters. Since the three effluent filters monitor the same effluent stream during the same period of time, the recommended metric is the total number of (oo)cysts detected on the three filters divided by the total volume of effluent that passed through the filters, which can be interpreted as the average (oo)cyst concentration measured by the three effluent filters. This average concentration for each run and treatment system is the metric we recommend for use in the Pitman Test to test the null hypothesis that the effluent concentrations are identical for the two treatment systems.

[5] The variances in Table 3 are between runs, but the replicate filters within runs are generally consistent. Hence, the variability between independent filters from different runs would only slightly exceed the variability between runs.

[6] 10 percent chance of rejecting equivalency when the systems actually are equivalent.

5. Results from All 12 Runs

Aliquots (approximately 0.75 ml each) were taken at systematic time intervals from each spike solution tank, the tank from which the spiking material was pumped into the influent stream, to verify the spiking concentrations. The results are presented in Table 4. We see that the overall average concentration in the spike solution tank was measured to be approximately 1×10^8 units per 100 liters for both Giardia cysts and Cryptosporidium oocysts. Given that approximately 42 liters (11 gallons) were delivered from the spike solution tank to the influent stream for each treatment unit, the total numbers of Giardia cysts and Cryptosporidium oocysts delivered to each unit was approximately 4.2×10^7. Given that effluent monitoring occurred for 80 of the 100 minutes during which the spiking load was delivered, approximately 3.3×10^7 Giardia cysts and Cryptosporidium oocysts were delivered to each unit during the 80 minutes during which the effluent concentrations were measured.

The effluent volumes filtered and numbers of Giardia cysts and Cryptosporidium oocysts detected on the individual effluent filters are presented in Table 5. Mean effluent concentrations and volumes monitored are presented in Table 6. The concentrations are all near zero. Hence, each unit achieved average removals greater than 7 logs.[7]

The Giardia analysis for one of the three filters from the microfiltration Run # 7 on August 25 resulted in an aberrant observation, 26 cysts. No Giardia cysts were found on the other two filters on the same microfiltration unit. Moreover, no Cryptosporidium oocysts were found on the same filter that had the 26 Giardia cysts. Although laboratory and field notes revealed no explanation for this anomaly, we believe that it is the result of some cross contamination. Therefore, it is not included in the statistical analyses.

Another data validity issue resulted from that fact that one of the six filtration membranes of the microfiltration unit appeared to be leaking slightly during the first two runs. This membrane was replaced by the manufacturer, Memtec, after the third run. Since we did not want to compare sand filtration to a leaking microfiltration unit, the data from the first three runs of the

[7] These calculations are all based on the measured influent and effluent concentrations. In fact, the recovery efficiency of the (oo)cyst measurement technique is estimated to be about 55 to 65 percent, certainly between 50 and 100 percent. Correcting for recovery would have a greater absolute effect on the influent concentrations because they are much higher concentrations. Therefore, the estimated average log removal would be increased. However, the magnitude of increase would be much less than an order of magnitude, hence, the average removal would still be on the order of 7 logs.

Table 4. Spiking Concentration Measurements

Run No.	Run Date	Filter No.	Giardia Cysts (per 100 l)		Crypto. Oocysts (per 100 l)	
			Microfiltration	CBUDS	Microfiltration	CBUDS
1	06/02/97	1	1.22×10^8	1.35×10^8	2.26×10^8	1.68×10^8
		2	1.48×10^8	1.46×10^8	2.15×10^8	2.07×10^8
		3	1.50×10^8	1.39×10^8	2.27×10^8	2.67×10^8
		Mean	1.40×10^8	1.40×10^8	2.22×10^8	2.14×10^8
2	06/16/97	1	0.42×10^8	0.43×10^8	1.21×10^8	1.01×10^8
		2	0.49×10^8	0.39×10^8	1.10×10^8	0.74×10^8
		Mean	0.45×10^8	0.41×10^8	1.15×10^8	0.88×10^8
3	06/30/97	1	0.92×10^8	1.08×10^8	0.85×10^8	0.68×10^8
		2	0.77×10^8	1.23×10^8	0.70×10^8	0.72×10^8
		Mean	0.84×10^8	1.15×10^8	0.77×10^8	0.70×10^8
4	07/14/97	1	0.90×10^8	0.61×10^8	1.47×10^8	1.70×10^8
		2	0.76×10^8	0.69×10^8	1.65×10^8	1.58×10^8
		Mean	0.83×10^8	0.65×10^8	1.56×10^8	1.64×10^8
5	07/28/97	1	1.38×10^8	1.43×10^8	1.13×10^8	0.55×10^8
		2	1.49×10^8	0.77×10^8	1.41×10^8	0.74×10^8
		Mean	1.44×10^8	1.10×10^8	1.27×10^8	0.64×10^8
6	08/12/97	1	0.79×10^8	0.89×10^8	0.78×10^8	1.10×10^8
		2	1.08×10^8	1.26×10^8	1.09×10^8	1.17×10^8
		Mean	0.94×10^8	1.07×10^8	0.93×10^8	1.13×10^8
7	08/25/97	1	1.40×10^8	1.33×10^8	1.50×10^8	1.21×10^8
		2	1.50×10^8	1.10×10^8	1.10×10^8	1.74×10^8
		Mean	1.45×10^8	1.22×10^8	1.30×10^8	1.48×10^8
8	09/08/97	1	1.13×10^8	0.70×10^8	0.99×10^8	0.60×10^8
		2	0.92×10^8	0.80×10^8	0.72×10^8	0.71×10^8
		Mean	1.03×10^8	0.75×10^8	0.85×10^8	0.65×10^8
9	09/22/97	1	1.02×10^8	0.92×10^8	0.40×10^8	0.65×10^8
		2	0.67×10^8	0.95×10^8	0.40×10^8	0.41×10^8
		Mean	0.85×10^8	0.94×10^8	0.40×10^8	0.53×10^8
10	10/06/97	1	0.74×10^8	0.87×10^8	0.68×10^8	0.51×10^8
		2	0.65×10^8	1.03×10^8	0.38×10^8	0.50×10^8
		Mean	0.69×10^8	0.95×10^8	0.53×10^8	0.51×10^8
11	10/20/97	1	0.79×10^8	0.63×10^8	1.05×10^8	1.62×10^8
		2	0.82×10^8	0.70×10^8	0.97×10^8	2.04×10^8
		Mean	0.81×10^8	0.67×10^8	1.01×10^8	1.83×10^8
12	10/27/97	1	1.72×10^8	1.65×10^8	0.87×10^8	0.67×10^8
		2	0.95×10^8	1.10×10^8	0.63×10^8	0.69×10^8
		Mean	1.33×10^8	1.37×10^8	0.75×10^8	0.68×10^8
Overall Mean[a]		All runs:	1.01×10^8	0.97×10^8	1.06×10^8	1.07×10^8
		Runs 4-12:	1.04×10^8		0.96×10^8	

[a] The overall mean was calculated by first averaging over filters within runs and then averaging over runs.

Table 5. Effluent Data

Run No.	Run Date	Filter No.	Volume Filtered		No. Giardia Cysts		No. Crypto. Oocysts	
			Micro	CBUDS	Micro	CBUDS	Micro	CBUDS
1	06/02/97	1	425	454	1	0	3	2
		2	434	447	2	0	2	1
2	06/16/97	1	427	428	0	0	0	0
		2	446	428	1	1	4	0
		3	424	430	1	0	5	0
3	06/30/97	1	428	424	0	0	0	0
		2	428	428	0	0	0	0
		3	428	428	0	0	0	0
4	07/14/97	1	435	436	0	0	0	0
		2	434	435	0	0	0	0
		3	435	434	0	0	0	0
5	07/28/97	1	425	422	0	0	0	0
		2	431	427	0	0	0	0
		3	422	425	0	0	0	0
6	08/12/97	1	417	427	0	1	0	0
		2	421	424	0	0	0	0
		3	410	424	1	0	0	0
7	08/25/97	1	427	433	0	0	0	1
		2	429	434	26[a]	0	0	0
		3	419	433	0	0	0	0
8	09/06/97	1	423	430	0	0	0	0
		2	436	441	0	0	1	0
		3	417	425	0	0	0	0
9	09/22/97	1	427	433	0	0	0	0
		2	429	434	0	0	0	0
		3	419	433	0	0	0	0
10	10/06/97	1	436	428	0	0	0	0
		2	440	433	0	0	0	0
		3	424	439	0	0	0	0
11	10/20/97	1	428	420	0	0	0	0
		2	435	441	0	0	0	0
		3	432	429	0	0	0	0
12	10/27/97	1	441	411	0	0	0	0
		2	445	411	0	1	0	0
		3	429	418	0	0	0	0
Overall Mean[b]		All runs:	428.8	430.5	0.21	0.08	0.49	0.15
		Runs 4-12:	428.4		0.04		0.04	

[a] This unusual result appears to be the result of cross contamination and, hence, is excluded from the statistical analyses.

[b] The overall mean was calculated by first averaging over filters within runs and then averaging over runs.

Table 6. Mean Effluent Volumes Monitored and Concentrations Measured

Run No.	Run Date	Volume (liters)		Giardia Cysts (per 100 l)		Cryptosporidium Oocysts (per 100 l)	
		Micro	CBUDS	Micro	CBUDS	Micro	CBUDS
1	06/02/97	429.5	450.5	0.349	ND	0.582	0.333
2	06/16/97	432.3	428.7	0.154	0.078	0.694	ND
3	06/30/97	428.0	426.7	ND[a]	ND	ND	ND
4	07/14/97	434.7	435.0	ND	ND	ND	ND
5	07/28/97	426.0	424.7	ND	ND	ND	ND
6	08/12/97	416.0	425.0	0.080	0.078	ND	ND
7	08/25/97	425.0	433.3	ND	ND	ND	0.077
8	09/06/97	425.3	432.0	ND	ND	0.078	ND
9	09/22/97	425.0	433.3	ND	ND	ND	ND
10	10/06/97	433.3	433.3	ND	ND	ND	ND
11	10/20/97	431.7	430.0	ND	ND	ND	ND
12	10/27/97	438.3	413.3	ND	0.081	ND	ND
Overall Mean[b]							
	All Runs:	428.8	430.5	0.049	0.020	0.113	0.034
	Runs 4-12:	428.4		0.009		0.009	

[a] ND = Non-detect (no (oo)cysts observed). The detection limit was approximately 0.08 (oo)cysts per 100 liters (1 (oo)cyst divided by the volume filtered). NDs were treated as zeros in all calculations.

[b] The overall mean was calculated by first averaging over filters within runs and then averaging over runs; non-detects (runs with no observed (oo)cysts) were treated as zeros.

microfiltration unit were not used in the statistical analyses.[8] However, since the hypothesis testing procedure recommended at the end of Section 4, the Pittman Test, treats the microfiltration and CBUDS runs as independent samples, we did retain the results of all 12 runs for the CBUDS unit in the statistical analyses.

Analysis of the results of the field test are organized around the following six issues:

1. Are the influent concentrations constant?
2. Are the influent and effluent concentrations correlated?
3. Are the microfiltration and CBUDS effluent concentrations correlated?
4. Are the monitored effluent volumes constant?
5. Are the treatment systems equivalent?
6. Was the sample size sufficient?

The first four issues have implications for what is the appropriate analysis for the hypothesis of equivalency of the two wastewater treatment systems. After exploring the implications of these issues, we discuss the hypothesis test outcome. Finally, as a quality assurance check we retrospectively investigate whether or not the number of runs appears to have been sufficient to achieve the desired power for the hypothesis test.

5.1 Are the influent concentrations constant?

As discussed in Section 2, if the influent concentrations are constant across all runs, then the original hypothesis test regarding log reductions in (oo)cyst concentrations is equivalent to the revised hypothesis test regarding effluent (oo)cyst concentrations only. Hence, we ran an analysis of variance to test two hypotheses — first, that the mean influent concentrations are the same for the microfiltration and CBUDS filtration systems and, second, that the mean concentrations are the same for all runs for each treatment system. The spiking material concentrations shown in Table 4 were treated as the influent concentrations for these analyses. The dilution of the spiking material into the 35 gallon per minute influent stream was not explicitly incorporated, without loss of generality. The results are presented in Tables 7 and 8.

[8] The test that detected leaking of the microfiltration membrane, a diffusion test, is not part of the standard operating procedures for units in field operation. Hence, one could argue that the first three runs should be included in the statistical analyses in spite of the fact that a membrane was replaced after the third run. If one were to include these first three runs, the microfiltration effluent means then would be consistently larger than the CBUDS filtration means, and the conclusion of the hypothesis test would not be altered. However, the desired 95 percent power would require additional runs for the Cryposporidium hypothesis test because of additional variability in the effluent results.

Table 7. Analysis of Variance for Giardia Influent Concentrations Excluding the First Three Microfiltration Runs

Source of Variation	df[a]	F Value	Prob > F
Treatment (micro vs sand)	1	1.08	0.3099
Runs (within treatment)	19	4.37	0.0006**
Filters (within treatment)	22		
Total	42		

[a] Degrees of freedom.

** Significantly different means at the 5% level of significance.

Table 8. Analysis of Variance for Cryptosporidium Influent Concentrations Excluding the First Three Microfiltration Runs

Source of Variation	df[a]	F Value	Prob > F
Treatment (micro vs sand)	1	2.50	0.1281
Runs (within treatment)	19	10.72	0.0001**
Filters (within treatment)	22		
Total	42		

[a] Degrees of freedom.

** Significantly different means at the 5% level of significance.

Review of Tables 4 and 7 shows that the influent concentrations of Giardia cysts were indeed massive (about 10^8 cysts per 100 liters); however, they were not constant. The analysis of variance in Table 7 shows that there was no significant difference between overall mean Giardia influent concentrations for the two treatment systems; however, there were significant differences in mean influent concentrations between runs.

Likewise, review of Tables 4 and 8 shows that the mean influent concentrations of Cryptosporidium oocysts exhibited essentially the same pattern of variability. The analysis of variance in Table 8 shows no significant difference between overall mean Cryptosporidium influent concentrations for the two treatment systems, but there are significant differences between runs.

The mean Giardia concentrations in the spike solution tank ranged from 0.41×10^8 to 1.45×10^8 cysts per 100 liters, and the mean Cryptosporidium influent concentrations ranged from 0.40×10^8 to 2.14×10^8 oocysts per 100 liters, when we exclude the first three microfiltration

runs. Hence, analysis of the effluent concentrations or (oo)cyst counts is not identical to an analysis of log removals. However, the mean spike concentrations are all close to 10^8 per 100 liters. Hence, interpretation in terms of log removals still seems reasonable.

5.2 Are the influent and effluent concentrations correlated?

Since the influent concentrations are not identical (could not be achieved practically), correlations between influent and effluent concentrations were possible. Significant correlations would suggest that the hypothesis test should be based on differences between influent and effluent concentrations, which would require much larger sample sizes because of the much greater variability in the observed influent/effluent differences, relative to the effluent counts or concentrations alone.

Hence, we tested for significant linear correlation between the pairs of influent and effluent mean concentrations for each treatment system and each type of protozoan, excluding the first three runs. For these tests, we treated non-detects (runs with no detected (oo)cysts) as zero concentrations, resulting in many ties among the output mean concentrations. The ties violate the assumptions underlying the hypothesis test for the Pearson correlation. Hence, we also ran nonparametric tests for correlations, the Kendall and Spearman tests. The results are presented in Table 9. The Kendall and Spearman tests are all non-significant at the 5 percent level. Therefore, we feel confident that basing the test for equivalency on the effluent concentrations or (oo)cyst counts alone is appropriate.

5.3 Are the microfiltration and CBUDS effluent concentrations correlated?

If the microfiltration and CBUDS effluent concentrations were correlated, paired tests based on the differences between the microfiltration and CBUDS effluent concentrations would be more appropriate. For a given number of runs, the paired tests, based on (n - 1) degrees of freedom, may be less powerful than independent sample tests, based on 2 (n - 1) degrees of freedom, depending on the strength of the correlation. Hence, additional runs might be necessary to achieve sufficient power.

Therefore, we tested for significant linear correlation between the pairs of microfiltration and CBUDS effluent mean concentrations, excluding the first three runs, for each type of protozoan. For these tests, we treated non-detects (runs with no detected (oo)cysts) as zero concentrations, resulting in many ties (see Table 6). Because the ties violate the assumptions underlying the hypothesis test for the Pearson correlation, we also ran Kendall and Spearman tests for correlation. The results are presented in Table 10. All tests are non-significant at the 5

Table 9. Test for Linear Correlation Between Influent and Effluent Concentrations[a]

Treatment	Sample Size	Protozoan	Measure of Correlation	Estimated Correlation	P-Value[b]
Microfiltration	9	Giardia	Pearson	-0.133	0.733
			Spearman	0.000	1.000
			Kendall	0.000	1.000
		Cryptosporidium	Pearson	-0.103	0.792
			Spearman	-0.137	0.725
			Kendall	-0.118	0.699
CBUDS	12	Giardia	Pearson	-0.033	0.919
			Spearman	0.064	0.843
			Kendall	0.045	0.856
		Cryptosporidium	Pearson	0.647	0.023
			Spearman	0.532	0.075
			Kendall	0.457	0.069

[a] Treating non-detects (effluent with no observed (oo)cysts) as zeros.
[b] Probability of a more discrepant result when the null hypothesis of zero correlation is true.

Table 10. Test for Linear Correlation Between Microfiltration and CBUDS Filtration Effluent Concentrations[a]

Protozoan	Sample Size	Measure of Correlation	Estimated Correlation	P-Value[b]
Giardia	9	Pearson	0.650	0.058
		Spearman	0.563	0.115
		Kendall	0.548	0.112
Cryptosporidium	9	Pearson	-0.125	0.749
		Spearman	-0.125	0.749
		Kendall	-0.125	0.724

[a] Treating non-detects (runs with no observed (oo)cysts) as zeros.
[b] Probability of a more discrepant result when the null hypothesis of zero correlation is true.

percent level.[9] Therefore, we feel confident that testing for equivalency by treating the microfiltration and CBUDS outcomes as independent sample outcomes is appropriate.

5.4 Are the monitored effluent volumes constant?

The staff implementing the pilot test attempted to keep the volume of effluent monitored constant across all runs. One benefit from monitoring a constant volume of effluent is a constant detection limit because the detection limit is one (oo)cyst (the minimum detection) divided by the volume of effluent monitored. Another benefit is that analysis of effluent concentrations is equivalent to analysis of (oo)cyst counts on the effluent filters because each concentration is the (oo)cyst count divided by the constant volume monitored.

Therefore, we conducted an analysis of variance to test two hypotheses — first that the mean volumes being monitored for the two treatment systems are the same and, second, that there is no difference in mean volumes between runs. The results of this analysis of variance are presented in Table 11.

Review of Tables 6 and 11 shows that there is no significant difference between the overall mean effluent volumes that have been monitored for the two treatment systems. However, there is a statistically significant difference in volumes between runs. Nevertheless, the mean

[9] Because there were so many ties, we also ran 2-way tables of detectable/non-detectable indicators for microfiltration versus CBUDS filtration for each protozoan, and tested for significant correlation using the Fisher's exact test. All correlations were non-significant at both the 5% and 10% significance levels.

volumes have ranged from 413.3 liters to 438.3 liters for all runs except for the CBUDS filters in the first run, for which the mean was 450.5 liters.

Because these differences are relatively small compared to the total volume monitored, we conclude that the difference in the volumes being filtered is not of practical importance, and the analysis of equivalency (the hypothesis test and the power calculations) can be based either on the effluent concentrations or on the numbers of (oo)cysts on the effluent filters.

Table 11. Analysis of Variance for Effluent Volumes Monitored Excluding the First Three Microfiltration Runs

Source of Variation	df[a]	F Value	Prob > F
Treatment (micro vs sand)	1	2.29	0.1381
Runs (within treatment)	19	5.60	0.0001**
Filters (within treatment)	41		
Total	61		

[a] Degrees of freedom.

** Significantly different means at the 5% level of significance.

5.5 Are the treatment systems equivalent?

As discussed in Section 3, the null hypothesis of interest is that the effluent concentrations are equivalent, and the alternative hypothesis is that the effluent concentration is higher with CBUDS filtration. That is, the null and alternative hypotheses are:

H_0: $O_m \geq O_s$ or $\Delta_O = O_m - O_s \geq 0$ (same or lower effluent concentration with sand)
versus
H_0: $O_m < O_s$ or $\Delta_O = O_m - O_s < 0$ (higher effluent concentration with sand).

As shown in Section 3, these hypotheses are equivalent to hypotheses regarding log reductions in (oo)cyst concentrations if the influent concentrations are equal. As discussed in Section 5.1, there is statistically significant variability in mean influent concentrations between runs. However, the ranges of mean influent concentrations are small relative to the overall mean concentrations, and interpretation of the results in terms of log removal still appears to be reasonable.

Because of the small sample sizes, 9 runs for microfiltration and 12 runs for CBUDS filtration, we implemented the Pittman Test based on the run mean concentrations, as recommended at the conclusion of Section 4. The results are presented in Table 12. We see that although the CBUDS mean effluent concentration is larger than the microfiltration mean concentration (excluding the first three microfiltration runs), this difference is not statistically significant. Therefore, there is not sufficient evidence to reject the null hypothesis of equivalence.

That is, the observed difference is not large enough to be particularly unusual if the true process means are equal, given the study sample sizes.[10]

Table 12. Hypothesis Test Results

Protozoan	Microfiltration		CBUDS		P-Value[a]
	Sample Size	Mean per 100 l	Sample Size	Mean per 100 l	
Giardia	9	0.009	12	0.020	0.3308
Cryptosporidium	9	0.009	12	0.034	0.4632

[a] Probability of a more discrepant result when the null hypothesis of equal means is true.

However, there is one remaining question that must be investigated. Are the sample sizes large enough to achieve the desired power, i.e., the desired probability of correctly rejecting the null hypothesis if it were false?

5.6 Are the sample sizes sufficient?

As discussed in Section 4, a recommendation was made by the stakeholders that the U.S. EPA require a test with sufficient power to be 95 percent certain of rejecting the hypothesis of equal effluent (oo)cyst means if the microfiltration mean were 1.00 (oo)cyst per filter and the CBUDS mean were 3.00 (oo)cysts per filter (i.e., 95 percent certain of not making a Type II Error when there is approximately 0.48 difference in log-removals). Given that the stakeholders were willing to accept, *a priori*, a hypothesis test conducted at a 10 percent significance level (for Type I Error), we recommended that a minimum of 12 runs be performed to reach a conclusion regarding the equivalency of microfiltration and CBUDS filtration for removal of Giardia cysts and Cryptosporidium oocysts. However, the primary assumption underlying this recommendation was that the variance of the filter counts was equal to the mean.

This assumption was postulated as a result of considering the process of filter loading to be approximated by a Poisson process, i.e., the event of an (oo)cyst being deposited on a filter occurs at random and independently in time (not in clusters) with a fixed probability of an event in a short time interval. The data from the first five runs did not appear to be inconsistent with this

[10]These hypothesis test results are based on the raw measurements without adjustment for recovery efficiency. Because of the nature of the nonparametric hypothesis testing procedure (a permutation test), the results of the test are unchanged if we make an adjustment for recovery efficiency so long as the same adjustment is applied for both treatment systems. Moreover, if one were to substitute an imputed concentration (e.g., the limit of detection) for the non-detects, the differences between the treatment systems would be even less significant because the detections are all near the limit of detection.

assumption, at least for Giardia cysts. Hence, the Poisson assumption was used as the basis for expected sample variability that was used to derive the recommendation of at least 12 runs in Section 4. To verify that 12 runs were sufficient, we need to verify that the data are consistent with the assumed population variability.

The data regarding the means and variances of the numbers of Giardia cysts and Cryptosporidium oocysts captured by the effluent filters are summarized in Table 13. The variances shown in this table are actually the variances between runs, while the parameter of interest is the variance between individual randomly selected filters from independent runs. Since the replicate measurements within runs are usually consistent (see Table 5), the variability between individual filters across independent runs would be only slightly greater than the variability between run means, shown in Table 13.

Table 13. Observed Means and Variances for the Effluent Filter Counts

	Giardia Cysts		Cryptosporidium Oocysts	
	Microfiltration	**CBUDS**	**Microfiltration**	**CBUDS**
Mean	0.037	0.083	0.037	0.153
Variance	0.012	0.023	0.012	0.189
$s^2/\bar{x}$	0.333	0.273	0.333	1.238
Sample Size	9	12	9	12

Table 13 suggests that the assumption that the variance of the number of Giardia cysts on a filter (approximately the number of cysts per 430 liters of effluent) is less than or equal to the mean is valid. We re-evaluated the power for detecting the difference between a mean of one cyst per filter for microfiltration and three cysts per filter for CBUDS filtration assuming 9 runs for microfiltration and 12 runs for CBUDS filtration, and we appear to have achieved at least 95 percent power (5 percent or less probability of a Type II Error) for this hypothesis test for Giardia removal.

However, Table 13 is not as definitive regarding the variability of the number of Cryptosporidium oocysts per filter. The observed variance ranges from 0.333 times the mean for microfiltration to 1.238 times the mean for CBUDS filtration. These results certainly are not inconsistent with the assumption that the process means and variances are equal. If, in fact, the overall means and variances are equal, the nine microfiltration runs and 12 CBUDS runs are sufficient to achieve the desired 95 percent power, as noted above regarding for the Giardia results. However, to be on the safe side we re-calculated the power assuming: (a) nine observations with equal overall mean and variance for microfiltration and (b) 12 observations with

the variance being 1.238 times the mean for CBUDS filtration. In this case, the power for detecting the difference between a mean of one oocyst per filter for microfiltration and three oocysts per filter for CBUDS filtration was reduced to 94 percent. Therefore, the power appears to be very close to the desired 95 percent for the Cryptosporidium hypothesis test, also.

Appendix A

Selected Poisson Probability Distributions

Table A.1 Poisson Distribution, Mean=0.83

Count	Probability	Cumulative Probability
0	0.436	0.436
1	0.362	0.798
2	0.150	0.948
3	0.042	0.990
4	0.009	0.998
≥5	0.002	1.000

Table A.2 Poisson Distribution, Mean=0.91

Count	Probability	Cumulative Probability
0	0.403	0.403
1	0.366	0.769
2	0.167	0.935
3	0.051	0.986
4	0.012	0.998
≥5	0.002	1.000

Table A.3 Poisson Distribution, Mean=1.0

Count	Probability	Cumulative Probability
0	0.368	0.368
1	0.368	0.736
2	0.184	0.920
3	0.061	0.981
4	0.015	0.996
≥5	0.004	1.000

Table A.4 Poisson Distribution, Mean=1.1

Count	Probability	Cumulative Probability
0	0.333	0.333
1	0.366	0.699
2	0.201	0.900
3	0.074	0.974
4	0.020	0.995
5	0.004	0.999
≥6	0.001	1.000

Table A.5 Poisson Distribution, Mean=2.2

Count	Probability	Cumulative Probability
0	0.111	0.111
1	0.244	0.355
2	0.268	0.623
3	0.197	0.819
4	0.108	0.928
5	0.048	0.975
6	0.017	0.993
7	0.005	0.998
≥8	0.002	1.000

Table A.6 Poisson Distribution, Mean=3.2

Count	Probability	Cumulative Probability
0	0.041	0.041
1	0.130	0.171
2	0.209	0.380
3	0.223	0.603
4	0.178	0.781
5	0.114	0.895
6	0.061	0.955
7	0.028	0.983
8	0.011	0.994
9	0.004	0.998
≥10	0.002	1.000

Table A.7 Poisson Distribution, Mean=3.3

Count	Probability	Cumulative Probability
0	0.037	0.041
1	0.122	0.171
2	0.201	0.359
3	0.221	0.580
4	0.182	0.781
5	0.120	0.895
6	0.066	0.955
7	0.031	0.980
8	0.013	0.993
9	0.005	0.998
≥10	0.002	1.000

Table A.8 Poisson Distribution, Mean=3.4

Count	Probability	Cumulative Probability
0	0.033	0.033
1	0.113	0.147
2	0.193	0.340
3	0.219	0.558
4	0.186	0.744
5	0.126	0.871
6	0.072	0.942
7	0.035	0.977
8	0.015	0.992
9	0.006	0.997
≥10	0.003	1.000

Table A.9 Poisson Distribution, Mean=4.4

Count	Probability	Cumulative Probability
0	0.012	0.012
1	0.054	0.066
2	0.119	0.185
3	0.174	0.359
4	0.192	0.551
5	0.169	0.720
6	0.124	0.844
7	0.078	0.921
8	0.043	0.964
9	0.021	0.985
10	0.009	0.994
11	0.004	0.998
≥12	0.002	1.000

Table A.10 Poisson Distribution, Mean=6.9

Count	Probability	Cumulative Probability
0	0.001	0.001
1	0.007	0.008
2	0.024	0.032
3	0.055	0.087
4	0.095	0.182
5	0.135	0.314
6	0.151	0.465
7	0.149	0.614
8	0.128	0.742
9	0.098	0.840
10	0.068	0.908
11	0.043	0.951
12	0.025	0.976
13	0.013	0.989
14	0.006	0.995
15	0.003	0.998
≥16	0.002	1.000

Table A.11 Poisson Distribution, Mean=10

Count	Probability	Cumulative Probability
≤1	0.001	0.001
2	0.002	0.003
3	0.008	0.010
4	0.019	0.029
5	0.038	0.067
6	0.063	0.130
7	0.090	0.220
8	0.113	0.333
9	0.125	0.458
10	0.125	0.583
11	0.114	0.697
12	0.095	0.792
13	0.073	0.864
14	0.052	0.917
15	0.035	0.951
16	0.022	0.973
17	0.013	0.986
18	0.007	0.993
19	0.004	0.997
≥20	0.003	1.000

Table A.12 Poisson Distribution, Mean=14

Count	Probability	Cumulative Probability
≤3	0.001	0.001
4	0.001	0.002
5	0.004	0.006
6	0.009	0.014
7	0.017	0.032
8	0.030	0.062
9	0.047	0.109
10	0.066	0.176
11	0.084	0.260
12	0.098	0.358
13	0.106	0.464
14	0.106	0.570
15	0.099	0.669
16	0.087	0.756
17	0.071	0.827
18	0.055	0.883
19	0.041	0.923
20	0.029	0.952
21	0.019	0.971
22	0.012	0.983
23	0.007	0.991
24	0.004	0.995
25	0.002	0.997
≥26	0.003	1.000