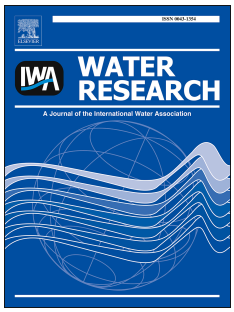
Accepted Manuscript



Effects of Holding Time and Measurement Error on Culturing Legionella in Environmental Water Samples

W. Dana Flanders , Kimberly H. Kirkland , Brian G. Shelton

PII: S0043-1354(14)00378-9

DOI: 10.1016/j.watres.2014.05.025

Reference: WR 10679

To appear in: *Water Research*

Received Date: 2 October 2013

Revised Date: 9 May 2014

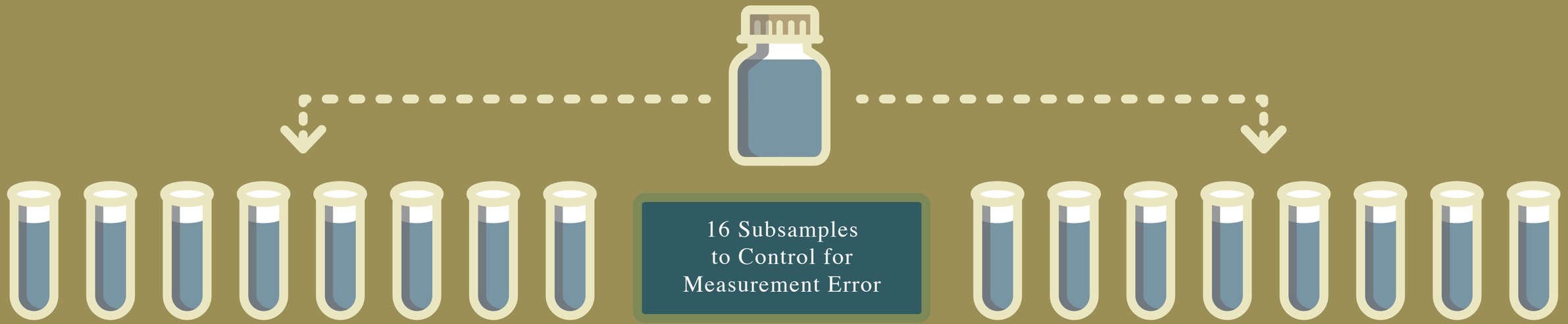
Accepted Date: 19 May 2014

Please cite this article as: Dana Flanders, W., Kirkland, K.H., Shelton, B.G., Effects of Holding Time and Measurement Error on Culturing Legionella in Environmental Water Samples, *Water Research* (2014), doi: 10.1016/j.watres.2014.05.025.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Effect of a Holding/Shipping Time on *Legionella* Culture Results?

159 Water Samples



16 Subsamples to Control for Measurement Error

Process 8 Subsamples Immediately

Hold 8 Subsamples for 24 hrs Before Processing

NO Meaningful Effect

ACCEPTED MANUSCRIPT

Highlights

- We evaluate the effect of holding/shipping time on *Legionella* culture results.
- We account for measurement error by replicating immediately-processed & held/ samples.
- Holding had a small effect on results relative to inherent measurement error (ME).
- After accounting for ME, shipped samples had very high sensitivity and specificity.
- Current practice of shipping samples (overnight express) does not invalidate results.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Effects of Holding Time and Measurement Error on Culturing *Legionella* in Environmental Water Samples

W. Dana Flanders, [1#]

Kimberly H. Kirkland, [2]

Brian G. Shelton, [2]

Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia 30322 (wflande@.emory.edu)[1] and PathCon Laboratories, Norcross, Georgia 30092[2]

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

ABSTRACT

Outbreaks of Legionnaires' disease require environmental testing of water samples from potentially implicated building water systems to identify the source of exposure. A previous study reports a large impact on *Legionella* sample results due to shipping and delays in sample processing. Specifically, this same study, without accounting for measurement error, reports more than half of shipped samples tested had *Legionella* levels that arbitrarily changed up or down by one or more logs, and the authors attribute this result to shipping time. Accordingly, we conducted a study to determine the effects of sample holding/shipping time on *Legionella* sample results while taking into account measurement error, which has previously not been addressed. We analyzed 159 samples, each split into 16 aliquots, of which one-half (8) were processed promptly after collection. The remaining half (8) were processed the following day to assess impact of holding/shipping time. A total of 2544 samples were analyzed including replicates. After accounting for inherent measurement error, we found that the effect of holding time on observed *Legionella* counts was small and should have no practical impact on interpretation of results. Holding samples increased the root mean squared error by only about 3 to 8%. Notably, for only one of 159 samples, did the average of the 8 replicate counts change by 1 log. Thus, our findings do

ACCEPTED MANUSCRIPT

not support the hypothesis of frequent, significant ($\geq$ = 1 $\log_{10}$ unit) *Legionella* colony count changes due to holding.

Keywords: *Legionella* monitoring, Sample holding time, Shipping effects, *Legionella* culture, Measurement error, Sensitivity

1. Introduction

Legionnaires' disease accounts for about 1-5% of community-acquired pneumonia with perhaps 8,000 to 18,000 cases occurring annually in the United States, and reported cases continue to increase each year following a substantial increase in 2003 (1, 2, 3). It is indicated that legionellosis is greatly underdiagnosed and underreported and the number of cases is likely greater than reported (4, 5). The disease has a fatality rate of about 5-30% and is higher among the immunocompromised (5, 6). Disease is caused by *Legionella* bacteria, usually *L. pneumophila* serogroup 1, although many species and serogroups of *Legionella* can cause disease (5). *Legionella* is an important waterborne bacterium that poses a significant health risk to people exposed to the organism in aerosolized water droplets from contaminated water systems (7). Water sampling for *Legionella* bacteria is an essential component of investigations of Legionnaires' disease outbreaks and sampling is useful in identifying potentially contaminated sources

with *Legionella* isolates sometimes used to identify the source of the implicated etiologic strain. In addition, water sampling for *Legionella* is sometimes utilized to assess the efficacy of maintenance programs and disinfection procedures where sample quantitation is particularly important. *Legionella* bacteria are widely found in low levels in natural bodies of water (8) and, at times, in potable and non-potable building water systems (7). Identification often involves cultures of the bacteria in samples of water to which people are exposed.

To identify *Legionella* in water samples, the U.S. Centers for Disease Control and Prevention (CDC) and the European Health Protection Agency recommend culture analysis. Culture analysis, however, has inherent variability – as do any other quantitative microbiological culture methods (9). For example, if culture analysis is performed on a particular water sample and repeated immediately on the same sample, the first concentration will likely not be identical to the second one, reflecting inherent measurement error. In part, because of this measurement error, proficiency testing of laboratories that perform *Legionella* analyses is conducted by the CDC Environmental *Legionella* Isolation Techniques Evaluation (ELITE) program in the U.S. and by the Centre for Infections Food and Environmental Proficiency Testing Unit (FEPTU) in Europe. The inherent measurement error in culture analysis is indicated by results from the CDC Elite proficiency testing

program (10); they report a between-laboratory standard deviation of 0.62 logs for the reported *Legionella* counts (log transformed), similar intra-laboratory variability, and an even greater deviation of reported counts from what was considered the true value.

In a recent publication, McCoy et al. (11) note that error in estimated counts from *Legionella* culture analysis could arise due to a delay in plating the cultures, such as would occur if a sample was shipped overnight from the collection site to the laboratory. They report that culture analyses they initiated immediately yielded different results than did analyses that were delayed by holding samples for 6 or more hours at room temperature before plating. Notably, they report that *Legionella* counts on 52% of their cultures plated immediately differ by one order of magnitude or more from counts obtained from a repeat culture of the same sample, apart from the 6 plus hour delay. They attribute the differences to the holding times. The authors report no systematic pattern of differences: they report that culture results processed after holding can be either substantially higher or substantially lower than immediately processed culture results, with no apparent systematic trend in either direction. If holding time does adversely impact sample results, their findings have significant implications for water sample collection and analysis for *Legionella* during outbreak investigations and risk assessments.

Measurement error is an unavoidable component of microbiological sampling, particularly when analyzing small-volume samples using culture media, such as testing for bacteria in water (12, 13). It can be introduced during a number of analytical steps, including unaccounted for variation in sample volume analyzed, pipetting, spread plating, selective procedures such as acid or heat treatment, and incubation conditions (9). Measurement error can also be due to variability in water sample characteristics including concentration of the organism in the sample, concentrations of competing organisms in the sample, amount of debris, and the non-uniform distribution of organisms in the sample. Despite its importance, we identified only two peer-reviewed, published studies reporting within-sample measurement error results for *Legionella* culture (10, 13), a third publication referring to one-order of magnitude "precision" without indicating how the estimate was derived (11), plus websites, such as those that had reported results from European proficiency testing (10).

A potentially important limitation of the study by McCoy et al. is that they did not account for the variability that is inherent in the microbiological culturing of *Legionella* samples ("measurement error"). Although they refer to 1-log "accuracy", the methods described for evaluation of the effect of holding time do

not account for measurement error, for example by replication or analytic correction. Instead, the authors attributed any difference between the immediately processed culture result and the corresponding result for the same sample obtained after a delay entirely to the holding time. However, if the inherent measurement error is important, it could account for most of the difference between the culture result obtained from the immediately processed sample and the result obtained from the sample processed after holding. On the other hand, if the measurement error is relatively small, it would not account for the differences between these culture results. Thus, it is important to account for inherent measurement error in evaluating the importance of any impact of holding time on *Legionella* culture results.

The primary goal of our study is to estimate the impact of holding time on culture results, *after* accounting for the random within-sample measurement error that affects culture analyses. In particular, we estimate the average change in culture results and the proportion of samples in which the *Legionella* count changes by at least an order of magnitude after a one-day delay. Secondarily, we assess the within-sample measurement error in culture results processed by direct culture, both with and without delays. To estimate and to account for inherent measurement error, we based analyses on replicate cultures – both for samples

ACCEPTED MANUSCRIPT

136 plated immediately in the field, and for samples processed in the laboratory after
137 holding/shipping for one day.
138
139
140 2. Methods and Materials
141
142 2.1. Water sample collection procedures and plating schedule
143 *Group A samples*
144 Ninety 125-ml samples were collected from six different hotel buildings using
145 sterile polypropylene containers containing sodium thiosulfate, a chlorine and
146 other oxidizing biocide neutralizer. The samples represent many types of water
147 systems (predominantly potable, but also non-potable water). Samples were
148 collected from hotel water systems and included showers, sinks, spa tubs, hot water
149 storage tanks and return systems, and two cooling towers. These samples were
150 collected in October 2012 in Nevada and California and shipped from these
151 locations.
152
153 Each of the original samples was split into16 subsamples (8 replicates to be
154 analyzed promptly in the field (Time=0) and 8 replicates to be analyzed after
155 shipping (Time=1)). Each of the 16 subsamples was labeled with a unique code

number to blind laboratory analysts to the time of sample processing and identity of the original sample. All eight of the Time=0 subsamples were promptly plated in the field (within a maximum of 2 hours of collection) and incubated at 35°C. The next day, the inoculated media plates for the 8 Time=0 subsamples were shipped via priority overnight service in insulated boxes to the laboratory. Upon receipt at the laboratory, the Time=0 plates were incubated under recommended conditions of 35°C with 3% CO_2 for the remainder of the analysis. The remaining 8 subsamples (Time=1) were shipped on the day of collection via priority overnight service in insulated boxes to the laboratory for receipt the following day. These samples were plated at the laboratory on the day of receipt and incubated at 35°C with 3% CO_2. All analytical procedures performed on the Time=0 samples (plated promptly in the field) and Time=1 samples (plated after shipping) were the same, except for the differences in *timing* of the plating, shipping, and incubation as described above.

Group B samples

In addition, 69 samples were collected from building water systems within close proximity to our laboratory which is located near Atlanta, Georgia. These samples were collected from one hospital and from multiple buildings at a large industrial complex and types of sources included sinks, showers, hot water tanks and four

cooling towers. These samples were collected in July and August 2012 (26 samples) and in March 2014 (43 samples). All of these samples were immediately transported to the laboratory where each, original sample was split into16 subsamples (8 replicates were plated and incubated promptly (Time=0) and the other 8 replicates were held overnight at room temperature (21-23^{o}C) prior to analysis the next day (Time=1)). Because of the close proximity to our laboratory, the plating of the Time=0 subsamples was performed within 2 hours of collection and incubated immediately without any need for interrupting incubation for shipping samples to the laboratory. The remaining 8 subsamples (Time=1) were held overnight at room temperature (21-23^{o}C) (to simulate delays due to shipping) and plating was initiated within 22 to 26 hours of collection using identical methods to the Time=0 samples. The results from these 69 samples (Group B) were similar to the 90 samples (Group A) collected from other sites, so we reported results from all samples combined in Section 3.1 (159 samples, 1272 replicates at Time=0 and 1272 replicates at Time=1, n=2544). In section 3.2, we also present results for Group B samples only. Importantly, there was no difference in sample preparation and culture analysis used for samples processed immediately in the field (Time=0) and after shipping to the laboratory (Time=1) in both the original 69 samples and the larger sample size of 90 samples.

ACCEPTED MANUSCRIPT

196 2.2. Culture analysis for *Legionella*

197 All water samples (both from Time=0 and Time=1) were analyzed using methods
198 described below which include minor modifications to the published CDC method
199 (14). Direct plating as well as acid treatment of the samples (1:1 and 1:2 ratios)
200 was conducted in the analysis. It should be noted that all concentration steps were
201 omitted from the analysis (Time=0 and Time=1), as filtration is not practical to
202 perform in the field, outside of the laboratory. For this study, 0.1 ml of the water
203 sample was spread plated onto two media: buffered charcoal yeast extract (BCYE)
204 agar and modified GPVC (glycine, polymyxin B, vancomycin - without
205 cycloheximide). A total of three BCYE agar plates and three modified GPVC
206 plates were inoculated for each sample and incubated at 35^{o}C with 3% CO_2. After
207 4 days of incubation, all media were examined initially for the presence of
208 bacterial colonies having characteristics of *Legionella bacteria*. Incubation of all
209 culture plates continued for a minimum of 7 days and a maximum of 9 days with
210 all final visual examinations for presence of *Legionella* colonies occurring no
211 earlier than Day 7. *Legionella* colony counts were recorded as colony-forming
212 units per milliliter of sample (CFU/ml). Final concentrations for each sample were
213 calculated using the sample treatment that resulted in the best recovery of
214 *Legionella* bacteria. Both types of media as well as the direct plate and acid
215 treated portions of the sample were evaluated to determine which resulted in the

greatest recovery of *Legionella* colonies. The limit of detection (LOD) for this culture method is 10 CFU/ml. Suspect colonies were identified to genus level based on microscopic examination of colony characteristics and demonstrating the requirement of L-cysteine Some isolates (those detected from local samples) were further identified to the species and serogroup level by serologic methods using monovalent and polyvalent direct fluorescent antibody reagents and/or slide agglutination tests (15, 16).

2.3. Data analysis

We calculated descriptive statistics, including the proportion of culture results in which *Legionella* was detected, mean, median and geometric mean counts, and standard deviations by experimental group referred to as the "Time=0" and "Time=1" groups. To reduce the possible impact of a few high values, most analyses are based on logarithmic transformation (base 10). Before taking logarithms, we replaced values less than the limit of detection (LOD = 10 CFU/ml which is reported by Lucas et al. to be approximately the LOD (10)), with the LOD divided by 10; with this substitution the difference on the log scale between a count at the LOD and a value less than the LOD is treated as a 1 log difference.

We used a number of measures to characterize the effect of holding time on *Legionella* counts. One measure of the impact of holding time is the overall average difference between the counts at Time=0 and those at Time=1; thus, we compare the means, medians and geometric means at Time=0 with those at Time=1. A second measure of the effect of holding time is the absolute difference between the mean count of the 8 subsamples at Time=0 (on the log scale) and the corresponding mean of the 8 subsamples from the same sample at Time=1. We refer to this measure, when averaged over all 159 samples, as the mean absolute difference (MAD).

To assess within-sample measurement error, we calculated the within-sample standard deviation at Time=0 and Time=1. We also calculated the root mean squared error for the Time=0 and for the Time=1 subsamples (see Appendix A for the equation used for the estimate and for an explanation of why it is unbiased, if the assumption that the *mean* of the 8 replicates at Time=0 is unbiased).

We also evaluated how a binary analytic approach might change by accounting for within-sample measurement error. Therefore, we present results of "sensitivity" and "specificity" analyses with counts dichotomized at the LOD (10 CFU/ml). To account for (most of) the within-sample measurement error, we based classification

255 on the median of the 8 Time=0 subsamples. For these analyses, a “true positive”
256 was operationally defined as a sample in which the median of the 8 subsamples at
257 Time=0 was greater than the LOD; all other samples were operationally defined as
258 “true negative”. Using the true positive samples, sensitivity was then calculated as
259 the proportion of subsamples at Time=1 that were above the LOD; using the true
260 negative samples, specificity was calculated as the proportion of subsamples at
261 Time=1 that were below the LOD. In sensitivity analyses and for completeness,
262 we also analyzed these data using mixed, random effects linear models (methods
263 and results in Appendix B).
264
265 We conducted statistical analyses using all samples (n=159 samples, 1272
266 replicates at Time=0 and 1272 at Time=1) and then repeated analyses, restricting to
267 those samples (n=82) for which 1 or more of the 16 subsamples was at or above
268 the LOD (see Appendix C). We also performed separate analyses for the 69 Group
269 B samples (552 subsamples at Time=0 and 552 subsamples at Time=1) that were
270 collected near our laboratory. These Time=0 subsamples were processed
271 immediately and analyzed without interruption (see Section 3.2 and Table 3). In
272 sensitivity analyses, we replaced values below the LOD with the LOD divided by
273 the square root of 2 (rather than 10) and re-estimated the root mean squared error
274 and repeated analyses based on random effects models. We also repeated analyses

with no transformation, or using random rather than fixed effects for sample, conducted analyses using a variance components model with restricted maximum likelihood, and maximum likelihood and type I sum of squares methods - all sensitivity analyses led to similar conclusions.

3. Results

3.1. Results for all samples (Group A and Group B)

As shown in Table 1, the geometric mean *Legionella* count for the 1272 subsamples processed immediately was 3.43 (arithmetic mean40.0) and for those processed after holding was 3.61 (arithmetic mean47.6). The count was about 0.02 logs (4 %) or 7.5 CFU/ml (19%) higher, on average, after holding. Approximately 31% of the 1272 subsamples had a *Legionella* count of 10 CFU/ml or greater, both at Time=0 and Time=1.

The average of the 159 within-sample *absolute* differences between the mean of the 8 replicates at Time=0, and the mean of the 8 replicates of the same sample at Time=1 was 0.121 logs (Table 2). In other words, after accounting for (most of) the within-sample measurement error by averaging the 8 replicates, the count changed by only 0.121 logs, on average. The maximum absolute difference between these means was 1.06 logs and only a single value of the 159 absolute

differences changed by 1 or more logs, after accounting for within-sample error. The average of the 159 within-sample standard deviations, an indicator of within-sample measurement error, was 0.202 logs at Time=0 and it was only slightly greater at Time=1 (0.208 logs).

The estimated root mean squared error at Time=0 is 0.337 logs using the Time=0 sample-specific mean concentration as the true value. The estimated root mean squared error at Time=1 is 0.370 logs, again using the Time=0 sample-specific mean concentration as the true value. Thus, we estimate that holding time increases the root mean squared error by about 9.8%, again assuming that the subsamples processed immediately are unbiased.

Fifty-two samples were operationally defined as "true positive" when we dichotomized samples using the median of the 8 subsamples processed at Time=0 to *partially* account for within-sample random measurement error. With the Time=0 median as the "gold standard" for each sample, the sensitivity of the cultures obtained at Time=1 was 81.7 % and the specificity was 91.6%. However, when we restricted the positive samples to those for which the median of the 8 Time=0 results was greater than twice the limit of detection (> 20 CFU/ml), the sensitivity of the individual Time=1 subsamples was 92.7% (i.e., without

accounting for measurement error at Time=1). The median of the 8 subsamples at Time=1 exceeded the LOD for these 49 of these 52 true positives (sensitivity would be 94.2%, if based on the median of the Time=1 subsamples) and the median of the 8 subsamples at Time=1 for 106 of the 107 "true negatives" were less than the LOD (specificity would be >99 %, if based on the median of the Time=1 subsamples). We repeated the analysis without accounting for measurement error by randomly selecting 1 of the 8 Time=0 replicates, treating it as the gold standard and comparing it with one of the randomly chosen Time=1 replicates. To increase stability, we repeated this process 50 times. Without accounting for within-sample measurement error at all, our estimates of differences were lower (sensitivity = 80.1 %, specificity = 90.9 %).

3.2. Results for only Group B samples

We also examined the 69 samples for which the Time=0 samples were processed at the laboratory within 2 hours of collection and the Time=1 samples were held until the following day (to simulate shipping) prior to processing (Table 3). The average of the count in these 552 subsamples when processed immediately was 18.7 CFU/ml (geometric mean3.40) and the mean was 18.7 CFU/ml (geometric mean3.30) for samples processed after holding. On the log scale, the counts increased, on average, by 0.03 logs from Time=0 to Time=1.

335

336 The average of the 69 within-sample *absolute* differences between the mean of the
337 8 replicates at Time=0 and the mean of the 8 replicates of the same sample at
338 Time=1 was 0.125 logs. In other words, after accounting for most of the random
339 measurement error, the absolute difference in counts was 0.125 logs, on average.
340 The maximum absolute difference between these means was 0.875 logs and no
341 value of the 69 absolute differences exceeded 1 or more logs, after accounting for
342 within-sample error. The average of the 69 within-sample standard deviations, an
343 indicator of within-sample measurement error, was 0.229 logs at Time=0 and
344 0.215 logs at Time=1.

345

346 The root mean squared error at Time=0 was 0.360 using the Time=0 sample-
347 specific concentration as the truth. The root mean squared error at Time=1 was
348 0.388 using the Time=0 sample-specific mean concentration as the truth. Thus, we
349 estimated that the root mean squared error increased by 7.8% after holding – if we
350 assume that the subsamples processed immediately have no bias. We found similar
351 results, in sensitivity analyses using mixed random effects linear models (see
352 Appendix B).

353

354

ACCEPTED MANUSCRIPT

4. Discussion

The results of our study suggest several important conclusions concerning *Legionella* culture analysis. First, we found that *Legionella* levels were about 0.02 to 0.05 logs higher, on average, and that the root-mean squared error was less than 10% higher after holding for 1 day. These changes associated with holding time are relatively small compared to the within-sample measurement error. Second, when accounting for measurement error, we found that the absolute difference between the mean Time=0 and Time=1 results was small or modest in nearly every sample, and for only one of 159 samples (less than 1%) changed by 1 log after holding. Thus, a delay in processing such as that associated with the common procedure of overnight shipping of water samples appears to allow for reliable enumeration of *Legionella* bacteria. Third, we found that within-sample measurement error (without using concentration steps to supplement the method, i.e., direct plating only) was about 0.3-0.5 logs. This was non-negligible, but likely consistent with values reported from the European proficiency testing (10). Therefore, there is inherent measurement error within *Legionella* culture analysis, even in subsamples processed identically and without delay, which cannot be disregarded.

When we did not account for within-sample measurement error by using only one of the replicates, the sensitivity and specificity of the held/shipped samples were relatively lower – if we treat the immediately plated samples as the "gold standard". This lower sensitivity and specificity were due primarily to measurement error and not to holding time because once we accounted for measurement error in both the Time=0 and Time=1 for direct plate (unfiltered) samples, the estimated sensitivity and specificity increased (estimated 100% and 97.7%, respectively, when based on the median of replicated subsamples). Thus, if we had ignored within-sample measurement error we might have had very different findings. The sensitivity and specificity reported here would be even higher if concentration steps typically used as part of our laboratory procedure for in-house laboratory analysis, were applied in this study to both Time=0 and Time=1 samples.

We note that sensitivity and specificity can be somewhat artificial measures of data quality for *Legionella* culture counts if the results are reported quantitatively, as we and several others do. Furthermore, we and some others recommend a graded interpretation of and response to *Legionella* culture results, based on 4 or 5 levels or categories. Successively higher *Legionella* levels and increased potential for exposure to aerosols require greater need for response and action (18, 19, 20, 21,

394 22). Also, a count that, for example, erroneously falls into an action level range
395 that is higher than the true level for the sample would likely be close to the cut
396 point between the levels (since root mean squared error is not large).
397
398 Our results concerning the impact of holding time are not inconsistent with those
399 of Barbaree et al. (17), although they evaluated much longer holding times (30 and
400 150 days). As did we, they used replication. Their samples when held for 30 days
401 at 25^{o}C, had an overall decrease in counts – but despite the much longer delay,
402 they found, much like us, that the counts did not change in any sample by 1 log or
403 more. Our results are partly consistent with those of Boulanger and Edelstein (13),
404 although they addressed a different goal using a different study design: they
405 primarily addressed the recovery of *Legionella* from seeded tap water. However,
406 they report, as do we, substantial measurement error (which they characterized as
407 variability in the recovery rate). On the other hand, they report lower sensitivity
408 (18 – 30% for counts <50 CFU/ml) than did we (75% at Time=1, restricted to
409 samples with a count <50 CFU/ml based on treating Time=0 median as the truth),
410 although this might be accounted for by differences in culture methods and our use
411 of real-world samples and an operational gold standard, rather than seeded samples
412 with known concentrations. Furthermore, Boulanger and Edelstein report that

413 reduced recovery of *Legionella* is attributed to cast membrane filtration,
414 centrifugation, and acid treatment (13).
415
416 We found less than 1% of the *Legionella* counts changed by 1 log or more after
417 holding once we accounted for within-sample measurement error. A key reason
418 for our finding probably reflects replication to account for within-sample
419 measurement. This contention is supported by a computer experiment and by
420 theoretical calculations: if the delays had had no effect, one would have expected
421 to find, on average, approximately half the samples changing by 1 log or more
422 from Time=0 to Time=1. In our computer experiment, we simulated no effect of
423 holding, but included normally distributed measurement errors having a 1-log
424 standard deviation (for reference one group (11) refers to an "accuracy" of about 1
425 order of magnitude for real-world samples). In 100,000 simulated subsamples,
426 48% of samples changed by 1 log or more. These simulated percentages are much
427 higher than those we found– reflecting the importance of accounting for
428 measurement error. In another computer experiment, we also simulated an effect of
429 holding combined with the measurement error; in this second experiment more
430 than 70% of samples changed by 1 log or more (depending on the magnitude of
431 holding effect) –more than the 48% seen when there is no effect of holding time.
432 [The r-program we used to simulate measurement error and sample-to-sample

433 variability before splitting the samples is available on request.] This computer
434 experiment and theoretical calculation strongly suggest that results can be heavily
435 influenced by measurement error alone. If measurement error is ignored
436 differences can occur and give the improper impression that holding time is having
437 an effect.
438
439 There are some limitations to our study that should be noted. For some samples
440 (Group A) the Time=0 plates were shipped overnight thus interrupting the
441 incubation time, but for other samples (Group B) the Time=0 plates did not have
442 an interruption in incubation. Also, the Time=1 subsamples for Group A were
443 shipped, but the Time=1 subsamples for Group B were held overnight at room
444 temperature (21-23^{o}C) to simulate a delay in processing due to shipping. However,
445 the results from these two groups were very similar (see, e.g. Section 3.2 and Table
446 3).
447 Another possible limitation is that we only assessed a holding/shipping time of
448 approximately one day. It is possible that samples shipped by methods slower than
449 overnight delivery, or from more distant locations requiring longer shipping times,
450 could experience higher holding time effects than what we report. However, a
451 majority of our samples (Group A, Time=1) were actually shipped across the
452 country so they are representative of delays due to real-world overnight shipping

453 which we and others recommend. Also, using various statistical approaches, we
454 provide several measures of the amount of error introduced by holding time. In
455 reality, these estimates for Group A include error not only from holding time, but
456 also from the limitations introduced by performing sampling in the field rather than
457 under controlled laboratory conditions. For example, it could be anticipated that
458 shipping the field inoculated petri dishes in less than ideal incubation conditions
459 during the critical growth phase of the organisms may have an effect of slowing
460 growth and potentially lowering the resulting count. However, this limitation does
461 not apply to the 69 Group B samples (all processing and holding occurred in the
462 laboratory).
463
464 Because of practical limitations in the field portion of this study and for
465 consistency of the field and laboratory analyses, we did not include filter
466 concentration steps as a component of sample processing. Filtration otherwise
467 would be a normal component of our analytical procedure for samples processed at
468 our laboratory. Because we accounted for within-sample measurement error (by
469 replication), the added step of filtration should have had a relatively smaller effect
470 and is not required for our assessment of the impact of holding time, our primary
471 study goal. In particular, *Legionella* counts changed only slightly after
472 holding/shipping (about a 1-day delay) and in only 1 sample did the sample-

specific mean change by 1 log or more. Our secondary goal, assessing the magnitude of measurement error before and after holding, concerns primarily the direct culture (unfiltered) results. Our supplemental results (Appendix D) for within-sample measurement error in the 26 cultures processed with filtration (and also without) give some guidance for within-sample measurement error when filtration is also performed. It is likely, and consistent with our supplemental results, that the inherent measurement error we report would be similar or even lower, and sensitivity and specificity higher, for samples processed using concentration steps - especially so for samples with lower counts, closer to the detection limit.

5. Conclusions

- In our evaluation of the effect of holding/shipping time on *Legionella* culture results, we found that measurement error that is inherent in culture results was important.
- After fully accounting for measurement error, the sensitivity and specificity of held/shipped samples were both very high.
- Compared with the inherent measurement error in culture results, holding had only a small effect on results. In fact, holding increased the estimated root mean-squared error by less than 10%.

ACCEPTED MANUSCRIPT

- Holding time, in particular for samples received at our laboratory within one hour of collection (Group B samples), appears to have minimal effect on quantitative results – in none of the Group B samples did the culture result change by 1 log or more.
- Our results suggest that delays in sample processing such as those due to shipping water samples via overnight services does not lead to invalid results and should not have a practical impact on interpretation of *Legionella* culture results.

Acknowledgements

The authors thank Daphne Cooper, Shawna Hawk, Krystal Miller, Mike Murphy, Grady Paxton, Margie Viskup, and MeiLin Whitehead for sample collection and analysis. David Jessup, PE, is acknowledged for his assistance with sample collection.

Role of the funding source

Funding for this research was underwritten by PathCon Laboratories.

Disclosure statement

At the time of the research, all authors had business affiliations with PathCon Laboratories.

ACCEPTED MANUSCRIPT

515 REFERENCES

516 **1. Marston BJ, Plouffe JF, File TM, Jr., Hackman BA, Salstrom SJ, Lipman**
517 **HB, Kolczak MS, Breiman RF.** 1997. Incidence of community-acquired
518 pneumonia requiring hospitalization – results of a population-based active
519 surveillance study in Ohio. Arch. Intern. Med. **157(15):**1709-1718.

520

521 **2. Centers for Disease Control and Prevention (CDC). Adams DA (Coord.).**
522 2012. Summary of notifiable diseases-United States, 2010. Morb. Mortal. Wkly.
523 Rep. **59(53):**1-105.

524

525 **3. Neil K and Berkelman R.** 2008. Increasing incidence of legionellosis in the
526 United States, 1990-2005: changing epidemiological trends. Clin. Infect. Dis.
527 **47:**591-599.

528

529 **4. Bohte R, van Furth R, van den Broek PJ.** 1995. Aetiology of community-
530 acquired pneumonia: a prospective study among adults requiring admission to
531 hospital. Thorax. **50(5):**543-547.

532

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

5. Marston BJ, Lipman HB, Breiman RF. 1994. Surveillance for Legionnaires' disease – risk factors for morbidity and mortality. Arch. Intern. Med. **154(21):**2417-2422.

6. Hubbard RB, Mathur RM, MacFarlane JT. 1993. Severe community-acquired legionella pneumonia: treatment, complications and outcome. Q. J. Med. **86(5):**327-332.

7. Fields BS, Benson RF, Besser RE. 2002. *Legionella* and Legionnaires' disease: 25 years of investigation. Clin. Microbiol. Rev. **15:**506-526.

8. Fliermans CB, Cherry WB, Orrison LH, Smith SJ, Tison DL, Pope DH. 1981. Ecological distribution of *Legionella pneumophila*. Appl. Environ. Microbiol. **41(1):**9-16.

9. Niemi RM and Niemel SI. 2001. Measurement uncertainty in microbiological cultivation methods. Accredt. Qual. Assur. **6(8):**372-375.

ACCEPTED MANUSCRIPT

10. Lucas CE, Taylor TH, Jr., Fields BS. 2011. Accuracy and precision of *Legionella* isolation by US laboratories in the ELITE program pilot study. Water Res. **45:**4428-4436.

11. McCoy WF, Downes EL, Leonidas LF, Cain MF, Sherman DL, Chen K, Devender S, Neville MJ. 2012. Inaccuracy in *Legionella* tests of building water systems due to sample holding time. Water Res. **46:**3497-3506.

12. American Public Health Association (APHA), American Water Works Association (AWWA), Water Environment Federation (WEF). 2005. ASTM 1030 Data Quality. Standard Methods for the Examination of Water and Wastewater, 21st ed. American Public Health Association, Washington, DC.

13. Boulanger CA and Edelstein PH. 1995. Precision and accuracy of recovery of *Legionella pneumophila* from seeded tap water by filtration and centrifugation. Appl. Environ. Microbiol. **61(5):**1805- 1809.

ACCEPTED MANUSCRIPT

14. Centers for Disease Control and Prevention (CDC). Fields BS (Ed.). 2005. Procedures for the Recovery of *Legionella* from the environment. US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, Atlanta, GA.

15. Benson RF and Fields BS. 1998. Classification of the genus *Legionella*. Semin. Respir. Infect. **13:**90-99.

16. Thacker WL, Plikaytis BB, Wilkinson HW. 1985. Identification of 22 *Legionella* species and 33 serogroups with the slide agglutination test. J. Clin. Microbiol. **21(5):**779-782.

17. Barbaree JM, Morrill WD, Fields BS, Martin WT. Detection and Recovery of *Legionella* in Water. Toxicity Assess. 1988; **3(5):**479-490.

18. Morris GK and Shelton BS. 1990. Technical Bulletin 1.5- *Legionella* bacteria in environmental samples: hazard analysis and suggested remedial actions. PathCon Laboratories, Norcross, GA.

ACCEPTED MANUSCRIPT

19. European Working Group for Legionella Infections (EWGLI), European Surveillance Scheme for Travel Associated Legionnaires' Disease (EWGLINET). 2005. European Guidelines for Control and Prevention of Travel Associated Legionnaires' Disease. European Parliament and the Council, 2119/98/EC.

20. Health and Safety Commission. 2000. Legionnaires' disease: the control of legionella bacteria in water systems. Approved code of practice and guidance (ACoP L8). HSE Books. Health and Safety Executive, Sudbury, UK.

21. Occupational Safety and Health Administration (OSHA). 1999. OSHA Technical Manual, 5th edition, Section III, Chapter 7: Legionnaires' disease. Rowman and Littlefield, Inc., Lanham, MD.

22. American Industrial Hygiene Association (AIHA). Hung, LL, Miller JD, Dillon HK (Eds.). 2005. Field Guide for the Determination of Biological Contaminants in Environmental Samples, 2nd edition, Chapter 6.2: *Legionella*. AIHA, Fairfax, VA.

ACCEPTED MANUSCRIPT

TABLE 1.Summary of *Legionella* culture results by time (all samples, n=2544)[a]

Time	Mean[b] Log10 (GM)	Median	Mean[c] (no transformation)	Percent ≥ 10 CFU/ml	SD[d]	Min	Max
0 (n=1272)	0.536 (3.43)	0	40.0	31.4%	0.845	0	3150
1 (n=1272)	0.557 (3.61)	0	47.6	31.4%	0.882	0	1980

[a] Before taking logarithms, all values < LOD replaced by LOD/10.

[b] Mean count after logarithmic transformation; geometric mean.

[c] Arithmetic mean.

[d] Standard deviation, after logarithmic transformation.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

TABLE 2.Summary of *Legionella* culture results based on mean of 8 replicates at each time (all samples, n=159; Group B, n=69)[a]

	Mean Absolute Difference[b]	Median Absolute Difference[c]	Max Absolute Difference	Proportion of Mean differences ≥ 1[b]
All (n=159)	0.121	0.011	1.06	0.006
Group B (n=69)	0.125	0.000	0.88	0

[a] Before taking logarithms, all values < LOD replaced by LOD/10.

[b] Difference between mean at time 1 and time 0: $|Mean_{s,1} - Mean_{s,0}|$.

[c] Difference between median at time 1 and time 0: $|Median_{s,1} - Median_{s,0}|$.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

TABLE 3.Summary of *Legionella* culture results by time (Group B samples only, n=1104)[a]

Time	Mean[b] Log10 (GM)	Median	Mean[c] (no transformation)	Percent ≥ 10 CFU/ml	SD[d]	Min	Max
0 (n=552)	0.53 (3.40)	1	18.7	34.4%	0.767	0	360
1 (n=552)	0.48 (3.03)	1	18.7	30.2%	0.767	0	330

[a] Before taking logarithms, all values < LOD replaced by LOD/sqrt(2).

[b] Mean count after logarithmic transformation; geometric mean.

[c] Arithmetic mean.

[d] Standard deviation, after logarithmic transformation.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

In this Appendix we justify our estimate of the root mean squared error (RMSE), when the goal is to estimate the true mean in each sample at time 0. We make a "worst case" assumption - that the true mean in each sample at time 0 is estimated without bias by the cultures processed immediately. In other words, we assume that with a very large number of repetitions (we used 8) the mean of the cultures processed at time 0 would be arbitrarily close to the true mean. If the assumption is incorrect and the time 1 mean is less biased than the time 0 mean, we would tend to underestimate the RMSE at time 0 and overestimate the RMSE at time 1.

With this worst case assumption, the mean squared error in the samples cultured at time 0 (denoted by MSE_0) is the average of the sample-specific variances for the samples processed at time 0. Thus, MSE_0 is consistently estimated by

1) $$MSE_0 = \frac{1}{7} \sum_{s=1}^{S} \sum_{k=1}^{8} (Y_{s,0,k} - y_{s,0})^2$$

so in expectation

2) $$E[MSE_0] = \frac{1}{S} \sum_{s=1}^{S} \frac{1}{7} \sum_{k=1}^{8} E[(Y_{s,0,k} - y_{s,0})^2] = \frac{1}{S} \sum_{s=1}^{S} \sigma_{s,0}^2$$

where: $Y_{s,t,k}$ is the cfu per ml, in sample *s*, at time *t*, repetition *k* for *s*= 1,..., S, *t* =0, 1 and *k*=1,...,8; is the $y_{s,0}$ is the observed mean of the 8 subsamples of sample *s* at time 0; and, $\sigma_{s,0}^2$ is the measurement error variance in sample s at time 0. We estimate the MSE in the time 1 samples (MSE_1) as:

3) $$MSE_1 = \frac{1}{8S} \sum_{s=1}^{S} \sum_{k=1}^{8} (Y_{s,1,k} - y_{s,0})^2 - MSE_0/8$$

The root mean squared error is estimated as the square roots of these quantities. We define the MSE in the sample at time 1 as:

4) $$MSE_1 = \frac{1}{8S} \sum_{s=1}^{S} \sum_{k=1}^{8} (Y_{s,1,k} - \mu_{s,0})^2$$

where $\mu_{s,1}$ and $\mu_{s,0}$ are the true means in sample *s* at time 1 and 0, respectively. The right hand side of Equation (4) is the overall mean squared error - the average over samples of the sample-specific mean squared errors.

We now show that the expected value of our estimate in Equation (3) equals MSE_1, as defined in Equation (4). By adding and subtracting the true means, we can rewrite Equation (3) as:

5) $$MSE_1 = \frac{1}{8S} \sum_{s=1}^{S} \sum_{k=1}^{8} (Y_{s,1,k} - \mu_{s,0} + \mu_{s,0} - y_{s,0})^2 - MSE_0/8$$

Taking Expectations, E[.] on both sides of Equation (5) and re-writing we obtain:

6) $$E[MSE_1] = \frac{1}{8S} E[\sum_{s=1}^{S} \sum_{k=1}^{8} \{ (Y_{s,1,k} - \mu_{s,0})^2 + (\mu_{s,0} - y_{s,0})^2 + 2(Y_{s,1,k} - \mu_{s,0})(\mu_{s,0} - y_{s,0}) \} - \frac{MSE_0}{8}]$$

$$= \frac{1}{8S} \sum_{s=1}^{S} \sum_{k=1}^{8} \{ E[(Y_{s,1,k} - \mu_{s,0})^2] + E[(\mu_{s,0} - y_{s,0})^2]$$

ACCEPTED MANUSCRIPT

$$+ 2E[(Y_{s,1,k} - \mu_{s,0})(\mu_{s,0} - y_{s,0})]\} - \frac{E[MSE_0]}{8}$$

$$= \frac{1}{8S} \sum_{s=1}^{S} \sum_{k=1}^{8} E[(Y_{s,1,k} - \mu_{s,0})^2] + \sigma_{s,0}^2/8 - \frac{E[MSE_0]}{8}$$

$$= \frac{1}{8S} \sum_{s=1}^{S} \sum_{k=1}^{8} E[(Y_{s,1,k} - \mu_{s,0})^2]$$

where we have used $E[(Y_{s,1,k} - \mu_{s,0})(\mu_{s,0} - y_{s,0})] = 0$; $\sum_{s=1}^{S} \sum_{k=1}^{8} E[(\mu_{s,0} - y_{s,0})^2] = \sum_{s=1}^{S} \sigma_{s,0}^2 = S \cdot E[MSE_0]$. The last line in Equation (6) is the same as definition of MSE_1 (right hand side of Equation 4), proving that the estimate of we use correctly estimates the mean squared error, averaged over samples, under our worst case assumption.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Appendix B – *Sensitivity analyses –mixed random effects linear model*

Methods: For completeness and as additional sensitivity analyses, we also analyzed our experimental data using a mixed, random effects linear model, with fixed effects for sample, a random effect for method within sample (either immediate or held), and a random error term. We used a logarithmic transformation (base 10) to improve normality and to decrease the impact of unusually high values. Although the distribution of counts even after logarithmic transformation was somewhat skewed when we studied all samples, they provided alternative, supplementary estimates of measurement error. We also use a Box-Cox approach; the inverse square root transformation yielded a slightly lower error sum of squares than other transformations, but even so use of this transformation yielded a similar pattern of results to use of the logarithmic transformation in that the within sample error (square root of the mean squared error) was substantially largerthan the average change after holding/shipping. Other models, such as including a random rather than fixed effect for sample also yielded similar patterns.

Results: The mixed random-effects linear models indicated a similar pattern. We found an average increase in counts from Time=0 to Time=1 of 0.02 logs. The estimated measurement error standard deviation was about 0.34 logs, and the additional error associated with holding time was small (0.10 logs) – both

consistent with our direct estimates. The pattern was similar with no transformation and with the inverse square root transformation.

Mixed random-effects linear models indicated a similar pattern when we evaluated the 82 positive samples. Here the distribution was more nearly bell-shaped after logarithmic transformation. We found an average increase from Time=0 to Time=1 of 0.04 logs. The estimated measurement error standard deviation was about 0.47 logs, and the additional error associated with holding time was small (variance = 0.02). The pattern was similar with no transformation, with the inverse square root transformation and with the logarithmic transformation.

ACCEPTED MANUSCRIPT

Appendix C- *Supplementary Results – results for positive samples only*

We also examined the 82 samples for which at least one culture in any of the 16 replicates for a sample was 10 CFU/ml or greater (Supplementary Table 2 below). The average of the count in these samples when processed immediately was 77.6 CFU/ml (geometric mean 10.9) and the mean was 92.3 CFU/ml (geometric mean 12.0) for samples processed after holding. On the log scale, the counts increased, on average, by 0.04 logs from Time=0 to Time=1.

The average of the 82 within-sample *absolute* differences between the mean of the 8 replicates at Time=0 and the mean of the 8 replicates of the same sample at Time=1 was 0.235 logs. In other words, after accounting for most of the random measurement error, the absolute difference in counts was 0.235 logs, on average. The maximum absolute difference between these means was 1.06 logs and only a single value of the 82 absolute differences exceeded 1 or more logs, after accounting for within-sample error. The average of the 82 within-sample standard deviations, an indicator of within-sample measurement error, was 0.391 logs at Time=0 and 0.404 logs at Time=1.

The root mean squared error at Time=0 was 0.506 using the Time=0, sample-specific concentration as the truth. The root mean squared error at Time=1 was

0.522 using the Time=0 sample-specific mean concentration as the truth. Thus, we estimated that the root mean squared error increased by 3.1% after holding – if we assume that the split samples processed immediately have no bias. We found similar results, in sensitivity analyses using mixed random effects linear models.

SUPPLEMENTARY TABLE 1: Summary of *Legionella* culture results by time (Positive samples only, n=1312)[a]

Time	Mean[b] Log10 (GM)	Median	Mean[c] (no transformation)	Percent ≥ 10 CFU/ml	SD[d]	Min	Max
0 (n=656)	1.04 (10.9)	15	77.6	60.8%	0.928	0	3150
1 (n=656)	1.08 (12.0)	15	92.3	61.0%	0.971	0	1980

[a] Before taking logarithms, all values < LOD replaced by LOD/sqrt(2).

[b] Mean count after logarithmic transformation; geometric mean.

[c] Arithmetic mean.

[d] Standard deviation, after logarithmic transformation.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Appendix D- *Supplementary Results – sample analysis included filter concentration*

For the 26 samples collected near the laboratory, the 16 split samples were 0.125 liters; we cultured each portion of each sample without filtration as described in method (only these results without filtration are presented in the main text). We also filtered the remaining portion of each split sample (about 0.100 liters), cultured 8 at Time=0 and 8 at Time=1 (as above). We now consider the culture results when filtration was used (total of 26x2x8=416 total culture results with filtration; Supplemental Table 1). We obtained similar results for these 26 samples, whether we used the results from processing with or without filtration. In particular, we found a small difference between the means at Time=0 and Time=1 (Supplementary Table 1) and the mean absolute differences were 0.095 with filtration and 0.096 without filtration. Furthermore, there was no sample that was positive at the LOD (10 CFU/ml) when filtration was used, but negative without (or, conversely) after accounting for within-sample measurement error. Estimates of overall measurement error, the root-mean squared error, were 0.33 and 0.28 at Time=0 and Time=1 respectively, both with and without filtration. The estimate of within-sample measurement error, characterized by the average within-sample standard deviation, was 0.19 at Time=0 and 0.15 at Time=1, both with and without filtration. However, when we used an LOD of 1 CFU/ml (attainable with

filtration), the mean absolute difference, root-mean squared errors and average within-sample standard deviations favored the filtered results (.e.g., average within-sample standard deviation was 0.19 for filtered and 0.23 for unfiltered results at Time=1).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

SUPPLEMENTARY TABLE 2.Summary of *Legionella* culture results by time using only results based on filtration (n=416) [a,b]

Time	Mean[c] Log10 (GM)	Median	Mean[d] (no transformation)	Percent ≥ 10 CFU/ml	SD[e]	Min	Max
0 (n=208)	0.325 (2.11)	0	9.43	22.1%	0.639	0	180
1 (n=208)	0.385 (2.43)	0	15.3	24.5%	0.709	0	270

[a]26 samples were processed both with and without filtration; for our main results, use only the results based on unfiltered samples – so that methods are consistent throughout;

[b] Before taking logarithms, all values < LOD replaced by LOD/10.

[c]Mean count after logarithmic transformation; geometric mean.

[d]Arithmetic mean.

[e]Standard deviation, after logarithmic transformation.

ACCEPTED MANUSCRIPT