WV TAP
WEST VIRGINIA TESTING ASSESSMENT PROJECT

Sampling Plan Design

Sampling Plan Design

Jeffrey S. Rosen[1], Andrew J. Whelton[2], Jennifer L. Clancy[1] and Timothy A. Bartrand[1]

1 - Corona Environmental Consulting
2 - University of South Alabama

0.0 Executive Summary

Following the spill of approximately 10,000 gallons of crude 4-methylcyclohexanemethanol (MCHM[1]) into the Elk River on January 9th, 2014, there have been persistent calls from the area residents for in-home sampling to establish the concentrations of the constituent chemicals in people's residences. This paper explores the properties of a sampling program that can answer the main questions being asked by residents and government officials. Statistical sampling design principles are applied to estimate the amount of certainty that can be established based on different sampling strategies. The evaluation of sampling strategies is based on a pilot sampling program that was implemented by the West Virginia Testing Assessment Project (WV TAP).

Designing a sampling program requires a clear understanding of the questions that the sponsors want to answer. In this case there are many questions that are being posed. Since no single design is optimized for all questions a subset of the questions that have been posed to WV TAP were selected and sampling designs to answer these questions were explored. The key questions addressed are:

1. What is the concentration of MCHM in people's residences?
2. Is the average concentration observed in homes below a level of concern?
3. What proportion of the homes has MCHM concentrations below a level of concern?

Sampling plans were evaluated that would allow testing whether or not measured concentrations were statistically different from critical values established by the Centers for Disease Control and Prevention (CDC), the State of West Virginia and WV TAP levels of concern. Standard statistical methods were used to estimate the likely confidence that would be observed for estimates of percentages of the residences in the affected area for which concentrations are above and below these critical values. Results from the 10 home sampling program demonstrated that more than one month after the spill that there were still detectable concentrations of MCHM in people's homes. The concentrations ranged from below detection levels of 0.5 parts per billion (ppb) up to 6.1 ppb. The standard deviation of the measurements ranged from a low of about 0.1 ppb to a high value of 1.5 ppb. These statistical properties were used to determine the number of residences that should be sampled and the number of samples that should be taken within each home.

A list of possible critical values to be considered is presented in Table ES -1. These are values that have been considered as critical health effect levels or levels of concern throughout the MCHM spill event.

[1] Crude MCHM is a mixture of pure 4-methylcyclohexanemethanol (referred to as MCHM in this report) and other organic compounds (Eastman Chemical Company, 2011). According to the Safety Data Sheet for crude MCHM (Eastman Chemical Company, 2011), pure MCHM makes up 68 - 89% of crude MCHM by weight. In this report crude MCHM denotes to the mixture spilled into the Elk River and MCHM denotes pure 4-MCHM

The values in the columns labelled as Differences in Measurements are the highest average value (4.9 ppb) and the highest individual measurement (6.1 ppb) recorded in the 10 home sampling program. In each of these columns is the difference between the critical value and the highest mean (4.9 ppb) and the highest individual measurement (6.1 ppb). One other difference, the difference between the current lowest detection level for MCHM and the Odor Threshold value which is the lowest concentration at which consumer panelists could consistently detect odors from crude MCHM, is evaluated.

Table ES-1 – List of critical values that might be evaluated relative to measured concentrations with the highest standard deviation observed in the 10 home sampling program (1.5ppb). The bottom row is a summary of the number of samples that would be needed to detect the difference between the Odor Threshold Concentration (0.55 ppb) and the lowest method detection limit (MDL) reported at the time this report was released (0.38 ppb). For this one row the standard deviation used is the lowest one observed (0.10 ppb).

		Difference from Measurements (ppb)			
Basis of Concern	**Level of Concern**	**4.9**	**6.1**	**0.38**	**Number of Samples Required**
CDC Screening Level	1 ppm	995.1	993.9		1
CDC Pregnancy Screening level[2]	50 ppb	45.1	43.9		1
WV TAP Health Effects Safe Level	120 ppb	115.1	113.9		1
Odor Recognition Concentration	7.4 ppb	2.5	1.3		17
Odor Objection Concentration	9.5 ppb	4.6	3.4		3
Odor Threshold Concentration	0.55			0.17	5

It is expected that once the GAC is completely replaced at the West Virginia American Water (WVAW) plant, the concentrations in the water delivered to the affected areas will be consistently lower than the concentrations measured in the 10 home sampling program. Therefore these sample sizes should be more than adequate to detect differences where they actually exist. The question that is asked will dictate the sampling design. If the main concern is that the water is safe for residents to use for all intended uses by all members of the community then all that is required is to evaluate samples relative to the WV TAP health effects safe level of 120 ppb which only requires a single sample given that the expected concentrations will be in the single digit ppb or lower. To ensure that each of the regions

[2] Personal communication from Dr. V. Kapil, CDC, to the WV TAP team, 26 March 2014

sampled in the early days of the MCHM event can be declared safe and to develop some confidence on the part of the citizens of the affected area it is advised that 30 homes in each of the 24 regions be tested. In order to verify that the variability is properly characterized by the 10 home sampling program it would be best to take two samples per home so there is a measure of the within home variability. Since the only chemical detected in the sampled homes was MCHM, it is the only chemical that needs to be analyzed.

This recommended sampling program would result in a total of 720 residences being sampled. This number of sampled homes would allow a good estimate of the percent homes in the affected area that are below any critical value of interest. For example, these data could be used to estimate the percentage of homes for which the MCHM concentration is below any of the critical values listed in Table ES-1. The confidence interval about any estimate of percent homes for the entire affected area would be in the range of ± 3% or better.

1.0 Introduction

This document describes a pilot sampling effort and follow-on analyses that were conducted to support development of a large-scale sampling plan and to answer questions related to the response to the Elk River MCHM spill that occurred in January of 2014. Those questions include how many samples should be taken within a single residence to determine whether MCHM concentration is above a level of concern, where the samples should be collected in a residence, and how many samples should be collected to determine the proportion of houses in an area of interest that must be sampled to establish the proportion of houses with MCHM concentration above a level of concern. The answers that statistics provide to these questions must be weighed against practical considerations such as budget, logistical constraints and public perceptions.

The Elk River Chemical Spill: Kanawha Valley Water Treatment Plant, Distribution System and Customers

On January 9, 2014 the State of West Virginia discovered that a major spill of “crude MCHM” was occurring from a chemical storage tank into the Elk River. This liquid industrial product contained 4-methylcyclohexanemethanol (MCHM) along with several other compounds in lesser quantity. The spill occurred at a site approximately 1.5 miles upstream of the Kanawha Valley water treatment plant (KVWTP) which is operated by West Virginia American Water (WVAW). The initial determination that the water was contaminated was based on complaints by residents of a licorice odor in the air. The licorice odor would become the *de facto* indicator for the people of the Kanawha Valley that there was something wrong with their drinking water.

The Kanawha Valley WTP supplies water to approximately 300,000 residents through a complex system that has the following characteristics:

- A span of nine counties,
- 1,900 miles of water mains,
- 100 water storage tanks,
- 179 separate pressure zones.

The area affected by the spill includes:

- Single family residences,
- Multiple residence buildings,
- Schools,
- All types of businesses,
    - Manufacturing
    - Office buildings
    - Restaurants
- Public buildings,
- Hospitals, and
- Other medical care facilities.

Premise plumbing is the plumbing that is under the control of the building owner after it leaves the water main at the service meter and enters the building. The water provider does not have jurisdiction of the plumbing at this point. Premise plumbing systems differ widely between types of buildings and even between buildings of the same type due to materials, design, and operation. Differences include the plumbing system components (water heaters, washing machines, water treatment devices), plumbing system materials (pipe type, valve type and materials, seal materials, connector materials), service connection within the distribution system, and operation (e.g., presence of pipe segments in the premise that are seldom used). Premise plumbing system complexity and diversity results in a wide variety of water retention times in within buildings connected to the system.

The Elk River Chemical Spill Response

Once the State and water utility confirmed that the spill had occurred, rapid decisions were required regarding whether or not to shut down the raw water intake. Following the confirmation of the spill a decision was made to not shutdown the intake. Crude MCHM contaminated river water entered the treatment plant, underwent some treatment and was discharged from the plant into the distribution system with finished water. The Kanawha Valley WTP is a conventional treatment plant using GAC filtration with a relatively short empty bed contact time and a four-year regeneration cycle. The plant uses chlorine for both primary and secondary disinfection. Potassium permanganate is also applied at the facility for preoxidation, manganese oxidation, zebra mussel control and reduction of disinfection byproducts.

Crude MCHM is an industrial product used to remove coal dust from mined materials. It is used regularly in the coal mining and distribution industry. Because crude MCHM was never expected to be in drinking water sources, little was known about its health effects, treatability or odor characteristics before the Elk River spill. The West Virginia Department of Health and Human Resources (WVDHHR) requested help from the CDC in determining what concentrations of crude MCHM compounds were safe for human exposure. In the area affected, humans could be exposed to chemicals in water through oral ingestion, inhalation because of volatilization from contaminated water and through dermal exposure. After deliberations, the CDC established a screening level of 1 part per million in water (ppm) as a concentration below which water was safe for ingestion.

Within a short period after the spill was known, the WVDHHR and partner agencies conducted water sampling for MCHM:

- In the raw and finished water of the WTP,
- At multiple locations in the distribution system, and
- In public buildings including schools.

Sampling continued in these facilities for a number of weeks following the spill. Some independent sampling was done in personal residences but the sampling was not conducted by State, Federal, or WVAW officials, nor was it coordinated. Figure 1 shows the concentrations of MCHM in finished water, hydrants, and other locations such as schools and public buildings from samples coordinated by the WVDHHR from January 10 to February 5. The black symbols show MCHM concentration (in ppb which is equivalent to µg/Liter) for samples in which MCHM was detected and the grey symbols show the reported detection limit for samples in which no MCHM was detected. No systematic validation or interlaboratory comparison of analytical methods results was undertaken. It is important to note that many of these analyses were carried out using a variety of method detection levels that started out at 1000 ppb then dropped in incremental steps to a MDL at about 0.5 ppb for the work done on this project. Figure 1 indicates that for two weeks MCHM concentrations were frequently greater than 100 ppb, followed by decreasing MCHM concentration and an increasing proportion of non-detect observations.

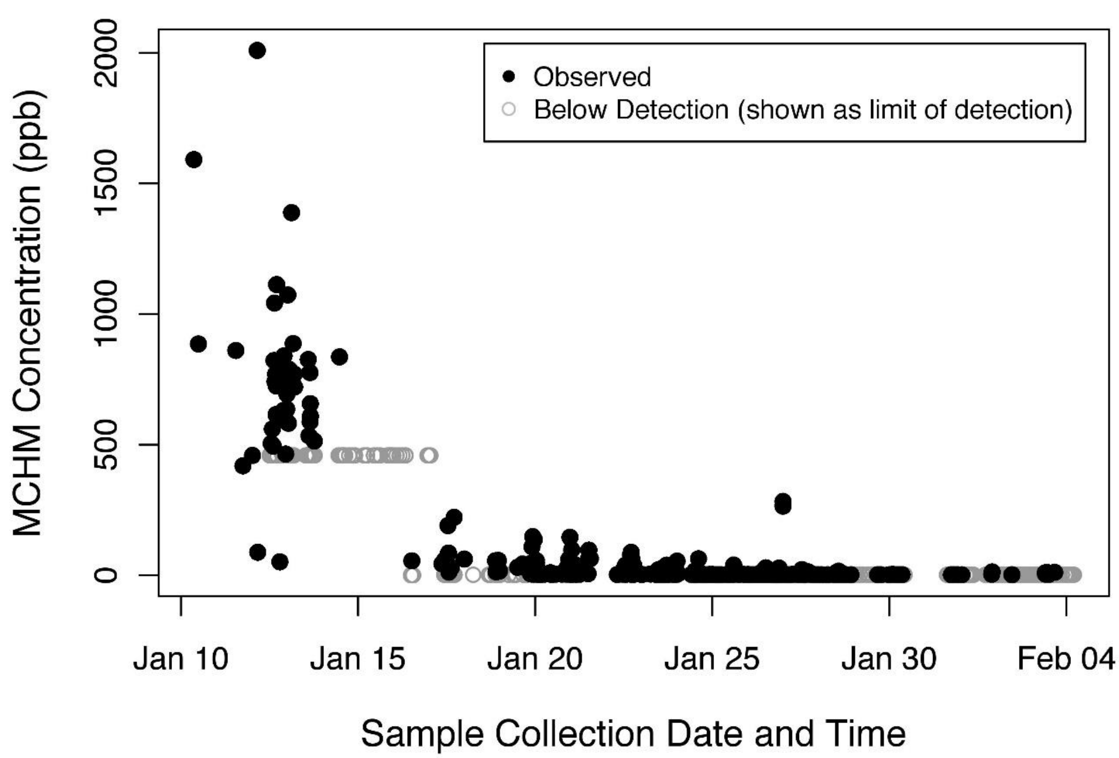

Figure 1. Time history of MCHM concentrations from samples drawn from the distribution system, school taps and faucets, and other public facilities. Note units on the y axis are equivalent to ppb.

After two weeks, most of the samples collected from the distribution system and in public buildings had MCHM concentrations below the detection level established by the laboratories (approximately 10 ppb). However, residents still reported smelling the licorice odor (and other odors described as medicinal, and sweet chemical) in their tap water, despite extensive flushing of the WVAW distribution system. The presence of sweet-licorice odors in their tap water along with analytical results that indicated that there were no detectable chemicals in their water resulted in distrust among the affected residents directed at WVAW, public health officials, the WV DEP and others. Many residents wondered if the concentrations in their homes were greater than the concentrations being measured in the distribution system and the public facilities. No in-home tap water testing had been carried out by the State at that time. Many residents called for an extensive residential sampling program to improve the understanding of their exposures to the contaminated water.

When the WVDHHR initiated the West Virginia Testing Assessment Project (WV TAP) one of the independent team's first tasks was to develop an in-home sampling plan that would address public concerns. The goal of this effort was to design a statistically defensible in-home water sampling plan for the residents affected by the Elk River spill. Specific objectives were to (1) describe important components of tap water testing, (2) quantify the variability in MCHM concentration within and among homes and its effect on in-home water sampling plan design, and (3) design a large-scale in-home sampling plan. To collect data for these purposes, the WV TAP team conducted a pilot sampling effort (called the 10 home study) in which extensive sampling of 10 homes was conducted. Among other objectives, the 10 home study was designed to develop an understanding of the variability of all components of the crude MCHM which might be measured within homes.

Objectives of this Study

The experience described above demonstrates the need for systematic, statistically designed sampling plans that produce actionable scientific data that can help restore customer confidence in their drinking water. To meet this need, the objectives of this study were to:

- Summarize available data,
- Estimate the number of samples that should be collected from an individual home to assess whether MCHM concentration in the home exceeds a level of concern, and
- Estimate the number of locations that must be sampled within a region of interest to establish the proportion of buildings with MCHM concentrations above a level of interest.

2.0 Factors that Determine In-Home Tap Water Sampling Design and Findings from the Ten-Home Study

The reason for an extensive sampling plan was to develop a good understanding of the concentrations of both the spilled industrial product, crude MCHM, its major and minor components and compounds which might be formed during water treatment or distribution or in the plumbing of affected homes. In an ideal world where funding, time and logistics are irrelevant, water samples would be collected at every residence in the affected area, at every tap in the building, and be analyzed for all components of crude MCHM along with testing that water for every possible breakdown compound. Table 1 is a summary of the customers affected by the crude MCHM spill. Multiplication by 3 gives an estimate of the total population affected (93,866 * 3 = 281,598 people).

Table 1. Affected customers.

Affected Customers	Number (Estimates)
Total	93,866
Residential	86,866
Commercial	5,435
Industrial	58
Public	557

Sampling all of the residential customers one time would require sampling 86,866 separate residences, at a cost of more than 10 million dollars (depending on the details of sampling) and would require more than a year for completion of sampling and laboratory analyses. While collecting and analyzing tap water from all homes is not feasible, sampling a proportion of the residences is feasible and, if done properly, can effectively answer many of the questions. Sampling would result in estimates of the concentrations and the variability that is characteristic of the exposure. Ideally, this sampling should have been initiated as soon as possible following the spill.

In order to design an effective in-home tap water sampling plan it is important to understand the questions that the program will be designed to answer. In the case of the West Virginia spill, the following questions might be asked of a sampling program that would be implemented many months after the spill. Some questions relate to the extent of contaminants still remaining in the overall system (including premises) and other questions relate to the conditions in a particular home. For example:

1. What is the concentration of MCHM in people's residences?
2. Is the average concentration observed in homes below a level of concern?
3. What proportion of the homes has MCHM concentrations below a level of concern?
4. Are there significant differences between concentrations in hot and cold water?
5. Are there significant differences in the concentrations between different locations in the residences?

Along with consideration of the questions that are asked, the level of precision required from the resulting data and the level of concern (i.e., a concentration of MCHM) compared to those measurements must also be considered. Question 1 (above) is a characterization of the chemical concentration in people's residences on a specific day. It implies a level of certainty and this level of certainty depends on the number of samples collected, the number of homes sampled and the variability observed in the data, which also is impacted by the quality of the analytical methods.

Questions 2 and 3 are comparisons of concentrations observed in a sampling program to a screening level established by authorities. In order to compare the concentrations the screening level must be known. A number of concern levels have been articulated over the course of the response to the Elk River chemical spill. Those levels are summarized below.

A. The CDC screening level of 1 ppm (1000 ppb; the original level of concern).
B. A few days after the CDC's issued its original guidance of 1 ppm they issued a second level of concern for pregnant women at 50 ppb which at the time was the detection level of the participating laboratories. (http://www.dhsem.wv.gov/WVTAP/test-results/Documents/POSTED_WV%20TAP%20TOX%20APP%20M%20%28CDC%29.pdf).
C. The WV TAP health review expert panel indicated a 120 ppb screening level below which the experts were willing to say that the water was safe for all people for all intended uses, as long as this was for an exposure less than 28 days.
D. The results of the crude MCHM Odor Threshold study (conducted with a consumer panel as part of the WV TAP independent analysis) indicated that the odor threshold concentration is 0.55 ppb, the odor recognition concentration is 7.4 ppb and the odor objection concentration is in the range of 7.7 – 9.5 ppb.

Question 2 is a comparison of the observed concentrations of chemicals to a level of concern. For example, if the WV TAP safe level of 120 ppb is the level for evaluation then the number of samples (n) that would be required to conclude with 95% confidence (α) that the measured values are below the safe level (120 ppb) is dictated by how close the observed value $(\overline{X})$ is to the level of concern, C (in this example the level of concern = the safe level = 120 ppb). The difference, δ, to be detected is the difference between the level of concern and the observed values. Further, the greater the variability in the measurements, σ, the larger the sample sizes that are necessary to detect a given difference. For data that are normally distributed or that can be modeled based on a normal distribution, the relationship between δ, n and σ is given by Equation 1.

$$n = \frac{(Z_\alpha + Z_\beta)^2\ \sigma^2}{\delta^2} \qquad (1)$$

In equation 1,

n is the number of samples per sample unit

$(z_\alpha + z_\beta)^2$ is a factor related to the level of significance and the power to detect real differences (α=0.90, β = 0.80)

σ is the standard deviation of the sampling unit and

δ is the difference that we wish to detect.

Similarly, the width of a confidence interval can be defined (Equation 2) for normally distributed data or data that can be modeled as normally distributed. When the variability of sampled data is not known and is based on fairly small sample sizes, the confidence interval is based on a student's t distribution instead of a normal distribution. The student's t distribution has longer tails than a normal distribution and accounts for uncertainty related to sample size.

$$\overline{x}\ \pm t_{(n-1,\alpha)}\, s/\sqrt{n} \qquad (2)$$

Where

$\bar{x}$ is the average concentration

$t_{(n-1,\alpha)}$ is a student t value with n-1 degrees of freedom for a particular confidence interval in this case (α = 0.9)

s is the standard deviation of the sample and

n is the number of samples included in the calculation of the average and the standard deviation.

To address question 3 above, it is necessary to determine the number of samples required to attain a particular level of confidence in the percentages estimated through a sampling program. If a number of buildings are sampled and the proportion of those buildings that are above any established critical level turns out to be some value $\hat{p}$, confidence levels around the proportion of buildings exceeding the level may be calculated using Equation 3.

$$\hat{p} \pm 1.96 * \sqrt{\hat{p}\hat{q}/_{n}} \tag{3}$$

Where

$\hat{p}$ is the proportion in the sample above the critical level,
$\hat{q}$ is the proportion in the sample below the critical level, and
n the number of sampling units sampled.

3.0 Sampling Strategy to Assess Individual Homes

Review and Analysis of Data from the 10 Home Study Pertinent to Sample Design

As part of the WV TAP project, a focused in-home tap water sampling effort was conducted for 10 homes in the affected area. This effort was conducted to establish the variability of chemical concentrations within each home, among other objectives. That variability includes differences between concentrations for samples taken from the same tap and variation in concentration among samples collected at different taps within a residence. Households were visited in eight of the nine counties (Boone, Cabell, Clay, Kanawha, Lincoln, Logan, Putnam, and Roane) from February 11, 2014 to February 18, 2014. Detailed results of the 10 home study are presented in separate reports describing tap water quality (http://www.dhsem.wv.gov/WVTAP/test-results/Documents/POSTED%2010%20Home%20Study%20Chemical%20Analysis%20Report_FINAL.pdf) and the interview conducted with the residents of those residences (http://www.dhsem.wv.gov/WVTAP/test-results/Documents/POSTED%2010%20Home%20Study%20Interview%20Report_FINAL.pdf).

In the 10 home study, water samples were collected in both a kitchen and a bathroom for both cold water and hot water samples. Nine samples were collected for each category of sample (i.e., Kitchen Cold, Kitchen Hot, Bathroom Cold, and Bathroom Hot). Three samples from each category were sent to each of two different laboratories (Eurofins and ALS). Three samples were held as backups in case there were samples lost in shipment. Table 2 summarizes the samples collected for each home.

Table 2. Samples collected and analyzed for each home in the 10home study.

			Samples		
Location	Tap	Location Code	Total Analyzed	Analyzed by Eurofins	Analyzed by ALS
Kitchen	Cold	1	6	3	3
Kitchen	Hot	2	6	3	3
Bathroom	Cold	3	6	3	3
Bathroom	Hot	4	6	3	3

The 10 home sampling plan was designed to evaluate whether there were differences between the locations in the homes and whether there were differences between the concentrations of chemicals in hot and cold water.

Table 3 is a summary of the total number of samples analyzed in all 10 homes and the number of detections for MCHM and PPH. The only chemical that was expected to be found and that was observed was MCHM (no PPH was detected). The only detections of MCHM were in analyses done by the Eurofins laboratory. The differences in the detections is due primarily to the differences in the detection levels, method reporting level (MRL) and method detection level (MDL) that the laboratories were able to attain along with the reliability of those detection levels (Table 4). Since the only results with positive detections were for MCHM samples analyzed by Eurofins, further analyses were performed on these results and additional sampling, if undertaken, should only be for MCHM.

Table 3. Summary of the total number of samples analyzed for the 10 home study.

			Samples			Detections Eurofins		Detections ALS	
Location	Tap	Location Code	Total Analyzed	Analyzed by Eurofins	Analyzed by ALS	MCHM	PPH	MCHM	PPH
Kitchen	Cold	1	60	30	30	27	0	0	0
Kitchen	Hot	2	60	30	30	27	0	0	0
Bathroom	Cold	3	60	30	30	28	0	0	0
Bathroom	Hot	4	60	30	30	28	0	0	0

Table 4. Summary of the method detection and reporting limits for the two laboratories involved in the analyses.

	Eurofins		ALS	
Analyte	Method Detection Level (MDL)	Method Reporting Level (MRL)	Method Detection Level (MDL)	Method Reporting Level (MRL)
MCHM	0.5	0.94	2.7	5.0
PPH	0.5	0.94	3.6	5.0

Of the 120 analyses of MCHM performed by Eurofins, 10 were below the laboratory's MDLs. The non-detect samples provide useful information (i.e., that the concentration of MCHM is lower than in samples in which it was detected) and this information should be included in the assessment of MCHM concentrations and variability and in development of a sampling plan. In order for these analyses to be included in statistical analysis and design procedures, values need to be assigned to the below detection level (BDL) values. Many approaches may be used to assign values to these BDL responses (Helsel 2005). All approaches, however, have limitations.

Several simple, but standard approaches to characterizing non-detect observations were compared to determine how sensitive results are to the choice of approach and which approach is appropriate. Four approaches were evaluated for these data. First, BDL values were omitted from analyses. Second, BDL values were assigned the value 0. Third, the BDL values were assigned half the detection limit. Fourth, BDL values were assigned the detection limit. With these assignments summary statistics for the entire data set were generated. Table 5 shows the impact of the four approaches (all of which introduce biases) on overall mean and standard deviation for the full data set and Table 6 shows the impact of the four approaches on mean and standard deviation for each house. The results of this analysis suggest that the approach for handling non-detect data does not have a very large effect on the summary statistics. Only three of the 10 homes tested had BDL observations for MCHM as measured by Eurofins. Houses 4, 9 and 10 had 4, 1 and 5 measurements BDL, respectively.

Table 5. Summary statistics over all locations for four approaches to account for a below detection observations.

Statistic	Result	BDL = Zero	BDL = Half MDL	BDL = MDL
Minimum	0.49	0.00	0.24	0.47
Mean	1.55	1.42	1.44	1.46
Standard Deviation	1.168	1.197	1.175	1.157
Maximum	6.10	6.10	6.10	6.10

Table 6. Demonstration of the impact of approaches for assigning values to below detection level (BDL) measurements of MCHM in the 10 home pilot testing study. Cells for which the BDL count value is not 0 denote homes for which there were BDL observations.

House Number	N Rows	BDL count	drop BDL		BDL 1/2 MDL		BDL = MDL	
			Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation
1	12	0	1.408	0.193	1.408	0.193	1.408	0.193
2	12	0	1.283	0.225	1.283	0.225	1.283	0.225
3	12	0	1.383	0.535	1.383	0.535	1.383	0.535
4	12	4	0.878	0.186	0.665	0.347	0.745	0.246
5	12	0	1.001	0.136	1.001	0.136	1.001	0.136
6	12	0	1.633	0.227	1.633	0.227	1.633	0.227
7	12	0	1.298	0.468	1.298	0.468	1.298	0.468
8	12	0	4.408	1.428	4.408	1.428	4.408	1.428
9	12	1	0.877	0.178	0.824	0.251	0.844	0.205
10	12	5	0.649	0.176	0.478	0.248	0.577	0.157

Since the method used to quantify the BDL values does not have a big effect on the means and standard deviations, any of the approaches described can be used. To make the mean and standard deviations conservative (i.e., biased toward overestimation) BDL values were assigned to the MDL in further analyses. Summary results with BDL replaced by the full detection limit are presented in Table 7 and Figure 2. These results are used to estimate the number of samples required to accurately characterize the concentration of MCHM in homes in the affected area.

Table 7. Summary statistics for MCHM concentration by home in ppb. All non-detect concentrations replaced with the method detection level.

House Number	Minimum	25th %ile	Average	Median	Standard Deviation	75th %ile	Maximum
1	1.20	1.20	1.41	1.40	0.193	1.60	1.70
2	1.00	1.13	1.28	1.20	0.225	1.53	1.70
3	0.79	0.91	1.38	1.30	0.535	1.80	2.40
4	0.48	0.48	0.75	0.82	0.246	0.97	1.10
5	0.82	0.88	1.00	0.98	0.136	1.10	1.20
6	1.30	1.50	1.63	1.60	0.227	1.85	2.00
7	0.71	0.79	1.30	1.25	0.468	1.70	1.90
8	2.20	2.93	4.41	4.45	1.428	5.88	6.10
9	0.48	0.67	0.84	0.87	0.205	0.99	1.20
10	0.47	0.48	0.58	0.54	0.157	0.56	0.93

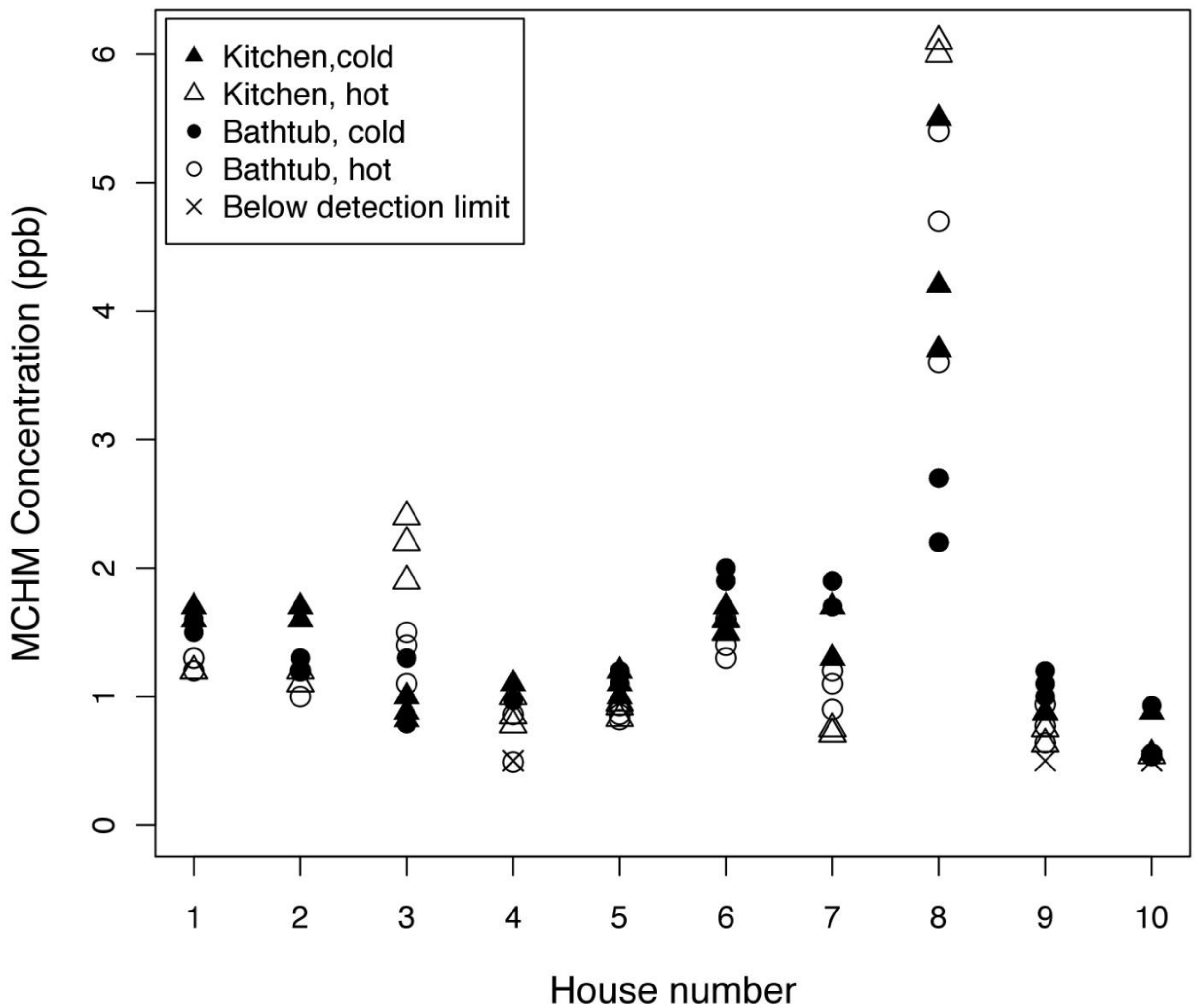

Figure 2. MCHM concentration (ppb) by house. Non-detect observations are shown as the detection limit.

The mean and standard deviation estimates (using replacement of BDL data with the full detection limit) for each house were used in Equation 2 to produce 95th percentile confidence interval estimates for the mean concentration of MCHM in each house and for each tap condition. Results are presented in Table 8. Data subsets were also assigned to a non-parametric group based on Wilcoxon Rank tests in an attempt to discern similarities between houses or tap conditions. Non-parametric groups with the same letter within each home are not different from one another. For example, for house number 1 the Kitchen Cold and the Bathtub Cold samples are not statistically different from one another (both listed as non-parametric group A1) while the Kitchen Hot and the Kitchen Cold (both listed as non-parametric group B1) are also not different from one another.

Table 8. Summary results for MCHM in ppb by home and location. Columns marked CI (Confidence Interval) Low and CI High are the boundaries of a 95% confidence interval calculated according to equation 2. The column marked non-parametric group shows which results within a home are significantly different from one another. The rows with the same letter and number are not different from one another.

House Number	Location	Sample Size	Average	Standard Deviation	CI Low	CI High	Non Parametric Group
1	Kitchen Cold	3	1.63	0.058	1.4899	1.7768	A1
	Kitchen Hot	3	1.20	0.000	1.2000	1.2000	B1
	Bathtub Cold	3	1.53	0.058	1.3899	1.6768	A1
	Bathtub Hot	3	1.27	0.058	1.1232	1.4101	B1
2	Kitchen Cold	3	1.63	0.058	1.4899	1.7768	A2
	Kitchen Hot	3	1.13	0.058	0.9899	1.2768	B2
	Bathtub Cold	3	1.23	0.058	1.0899	1.3768	B2
	Bathtub Hot	3	1.13	0.115	0.8465	1.4202	B2
3	Kitchen Cold	3	0.90	0.092	0.6723	1.1277	B3
	Kitchen Hot	3	2.17	0.252	1.5415	2.7918	A3
	Bathtub Cold	3	1.13	0.294	0.3985	1.8615	B3
	Bathtub Hot	3	1.33	0.208	0.8162	1.8504	B3
4	Kitchen Cold	3	0.69	0.358	-0.2025	1.5759	A4
	Kitchen Hot	3	0.88	0.112	0.5975	1.1559	A4
	Bathtub Cold	3	0.81	0.283	0.1039	1.5094	A4
	Bathtub Hot	3	0.61	0.217	0.0720	1.1480	A4
5	Kitchen Cold	3	1.10	0.100	0.8516	1.3484	A5
	Kitchen Hot	3	0.90	0.062	0.7449	1.0551	B5
	Bathtub Cold	3	1.13	0.058	0.9899	1.2768	A5
	Bathtub Hot	3	0.87	0.056	0.7317	1.0083	B5
6	Kitchen Cold	3	1.60	0.100	1.3516	1.8484	B6
	Kitchen Hot	3	1.53	0.058	1.3899	1.6768	B6
	Bathtub Cold	3	1.97	0.058	1.8232	2.1101	A6
	Bathtub Hot	3	1.43	0.153	1.0539	1.8128	B6
7	Kitchen Cold	3	1.57	0.231	0.9930	2.1404	A7
	Kitchen Hot	3	0.72	0.023	0.6660	0.7807	B7
	Bathtub Cold	3	1.83	0.115	1.5465	2.1202	A7
	Bathtub Hot	3	1.07	0.153	0.6872	1.4461	B7
8	Kitchen Cold	3	4.47	0.929	2.1585	6.7748	A8
	Kitchen Hot	3	6.07	0.058	5.9232	6.2101	A8
	Bathtub Cold	3	2.53	0.289	1.8162	3.2504	B8
	Bathtub Hot	3	4.57	0.907	2.3126	6.8207	A8
9	Kitchen Cold	3	0.87	0.006	0.8590	0.8877	A9
	Kitchen Hot	3	0.62	0.135	0.2840	0.9560	B9
	Bathtub Cold	3	1.10	0.100	0.8516	1.3484	A9
	Bathtub Hot	3	0.78	0.150	0.4096	1.1571	B9
10	Kitchen Cold	3	0.64	0.212	0.1142	1.1658	A10
	Kitchen Hot	3	0.50	0.035	0.4139	0.5861	A10
	Bathtub Cold	3	0.65	0.248	0.0310	1.2624	A10
	Bathtub Hot	3	0.52	0.044	0.4117	0.6283	A10

Initial inspection of Table 7, Table 8 and Figure 2 suggests the following.

1. The concentrations of MCHM observed in the 10 homes studied are all lower than all the levels of concern listed above, except the odor threshold concentration.
2. While most of the measured values are below 2 ppb, there are measurements above 2 ppb in home 3 and in home 8. Home 8 appears to be an outlier for this data set.
3. There is no clear pattern between concentrations in either the bathroom versus the kitchen or between hot and cold water.

These initial observations are confirmed below applying standard statistical tests. In order to determine which tests should be used and whether or not Equation 1 can be used, an evaluation of the normality of the data was performed both on untransformed and on log transformed data. The data sets are small but there consistently was no indication that the data are normally distributed or that a simple transformation like a log transformation would make the data normal. Therefore, any comparisons done were made using non parametric procedures. All comparisons were made using Wilcoxon Rank Test (also known as the Kruskal- Wallis Rank Sums) (calculated by JMP version 11). The differences between the locations within the homes are presented in Table 8. There are no consistent differences in these results. In fact for most of the analyses completed, the results do not vary in meaningful ways. However, the results do demonstrate that there are real differences between locations in some of the homes and therefore it would be useful to sample from more than one location in a home. The variability of the results within homes ranges from a low of 0.2 up to a high of 1.5. This range includes all the estimates of variability observed in the data, irrespective of how the data are grouped.

The power analysis equation presented in Equation 1 depends on an assumption of normality. While the data gathered from the 10 home sampling are not normally distributed, they are not widely divergent of normality. The function of the equation is to develop a relationship between the sample size, the variability in the data and the differences that the experimenters want to detect for a certain level of certainty and level of power. Since the data do not vary from normality in a meaningful way the power analyses will be applied and the results will be considered with additional variability and with some added samples to account for the failure to meet the assumption that the data are normally distributed.

Power Analysis to Estimate the Number of Samples per Home

To perform the power analysis four values need to be selected or estimated. First, a difference needs to be defined which will be the object of the sampling. In this case the difference can be between the highest concentrations (or mean concentrations) observed in the 10 home pilot study. The greatest MCHM concentration observed was 6.1 ppb and the greatest mean was 4.41 ppb, both observed at house number 8. The differences that an in-home tap water monitoring program might want to detect are the differences between these values and the critical values listed above and summarized in Table 9. For example, a sampling program might be designed to answer the question: “Is the average MCHM concentration in a specific home greater than the CDC pregnancy screening level?” In this case the number of samples could be chosen such that a difference of 6.1 ppb between the in-home concentration and the screening level (50.0 – 6.1 = 43.9 ppb) would be confidently detected.

Table 9. **– List of critical values that might be evaluated relative to measured concentrations with the highest standard deviation observed in the 10 home sampling program (1.5ppb). The bottom row is a summary of the number of samples that would be needed to detect the difference between the Odor Threshold Concentration (0.55 ppb) and the lowest method detection limit (MDL) reported at the time this report was released (0.38 ppb). For this one row the standard deviation used is the lowest one observed (0.10 ppb).**

		Difference from Measurements (ppb)		
Basis of concern	Level of Concern	4.9	6.1	0.38
CDC Screening Level	1 ppm	995.1	993.9	
CDC Pregnancy Screening level	50 ppb	45.1	43.9	
WV TAP Health Effects Safe Level	120 ppb	115.1	113.9	
Odor Recognition Concentration	7.4 ppb	2.5	1.3	
Odor Objection Concentration	9.5 ppb	4.6	3.4	
Odor Threshold Concentration	0.55			0.17

The second parameter needed for the power analysis is a range of values for the expected standard deviation (σ). Regardless of how the data are grouped and analyzed, the variability in MCHM concentrations (as expressed by the standard deviation) ranges from a low of about 0.1 ppb to a high value of 1.5 ppb. This standard deviation is characteristic of the range of values that were detected in the 10 home sampling study. It is likely that any sampling done after the release of this plan will be in this range or lower. Therefore, the power analysis is conducted over this entire range of variabilities.

The third parameter required is the confidence level desired (α). In this case a confidence of 90% has been selected. In many scientific studies the confidence is set at 0.95. Given the many uncertainties that will be inherent in any sampling plan done for the vast area affected by MCHM, the authors feel that a 95% confidence level would not be a reasonable expectation and therefore a 90% confidence level is used in all calculations in this report.

The final parameter required in the power analysis is the desired probability of finding a difference when there is a real difference. This value is the power of the test (β). For the analyses in this report the value of β is set at 0.80, which means that in about 20% of the tests there may be a real difference that will not be detected. This selection for the power of the analyses adds additional conservatism to the analyses and means that when a difference is detected it will likely be real.

If the question that is being addressed is whether the concentrations observed at a home are different than any of the screening levels, then applying Equation 1 using the range of variability observed in the 10 home sampling program results in the family of curves shown in Figure 3. Referring to Table 9 and Figure 3, the only values of interest that would require more than a single sample are the differences between the maximum measurements, maximum mean, odor recognition and odor objection levels. At the highest estimate of variability observed, to be able to detect differences of 1.3 ppb (the difference between the highest observed value and the odor objection level) requires 17 samples per home. If the monitoring program focuses on the question: "Is the concentration in a home below the odor recognition level?", three samples would be required based on the highest variability (1.5) and the highest measurement (6.1 ppb) observed in the 10 home sampling. Detecting differences between any of the values or means observed and the safe level established by the WV TAP program (120 ppb) would require only one sample. However, estimating the within-home variability requires a minimum of two samples in each home. Even for evaluation of whether the mean MCHM concentration in a given home is different from 120 ppb, at least two samples per home are required. In all cases the hypothesis that would be tested statistically is that the measured concentration is greater than the established critical value.

The persistence of odors being experienced by residents months after the spill and after multiple flushes of the system raises another possible question. Specifically, are the concentrations of MCHM in residences below the odor threshold concentration? Since Eurofins is able to detect to an MDL of 0.38 ppb and the odor threshold concentration as established by the consumer panel is 0.55 ppb, there is a comparison that can be made between an average concentration measured in homes. The difference that is being estimated is 0.55 - 0.38 ppb = 0 .17 ppb. Referring to Figure 3 (and the underlying equation) even at very low estimates of the standard deviation (σ = 0.1) 12 samples per home would allow a comparison between concentrations observed and the odor threshold value. However, testing of values below the MRL has additional challenges since there are additional uncertainties that cannot be easily quantified for values between the MRL and MDL.

In all cases the hypothesis that would be tested statistically is that the measured concentration is greater than the critical value of interest. Since the actual value can be either higher of lower than the critical value being test all tests would be based on two tailed tests.

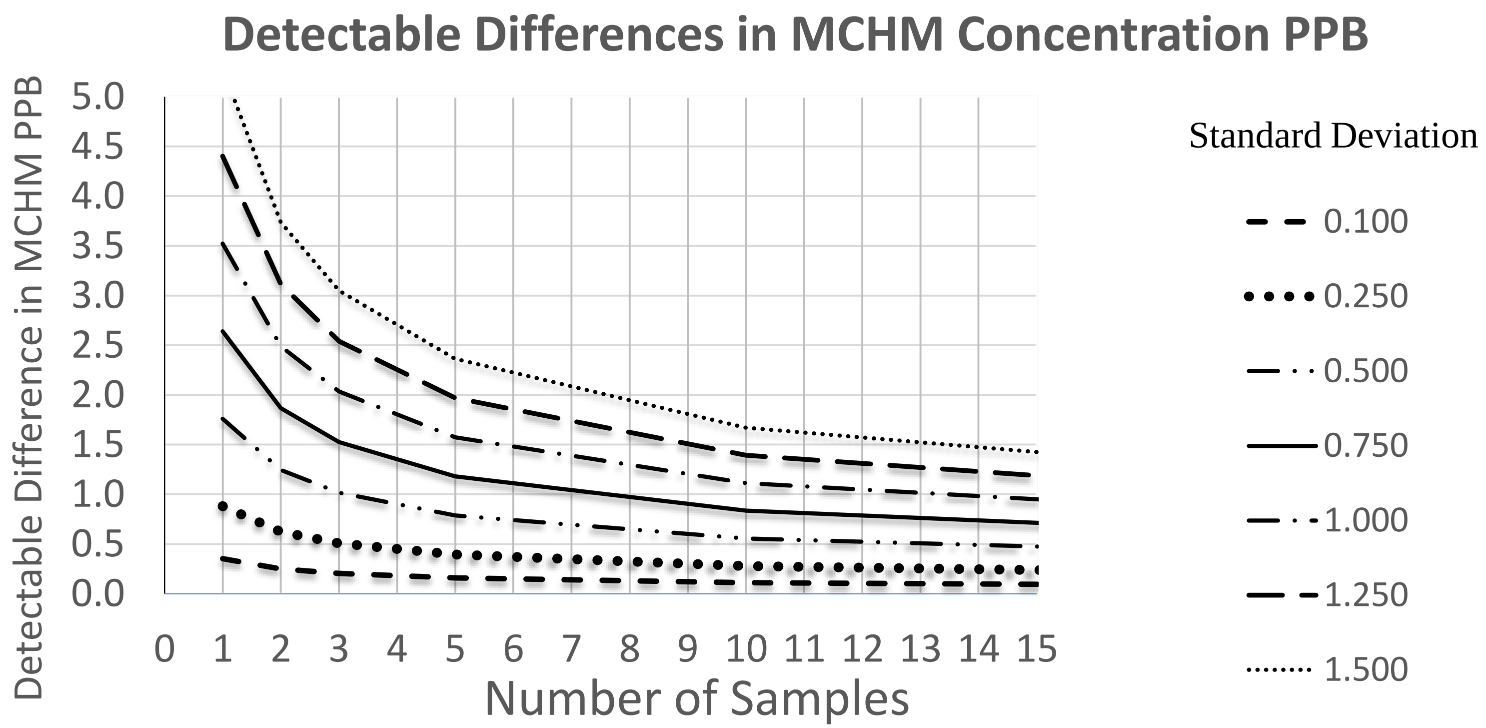

Figure 3. Results of the power analysis assuming a two-sided confidence interval with α (significance) = 0.1 and β (power) = 0.80.

Number of Samples to Assess the Proportion of Buildings in a Region with Concentration Above a Level of Concern

Developing a sampling plan to determine the proportion (percent) of homes in a given region that are above a certain threshold requires a decision regarding how much uncertainty is acceptable (see Equation 3). Figure 4 shows the value that will be added to and subtracted from the percent determined in sampling based on the sample size. For example, if the percent homes in the entire affected area above a threshold of 10 ppb were to be reported as 10% with a sample size of 400 then the resulting confidence interval would be 10% ± 2.4 % (solid line for 400 samples intersects 10% from the x axis at about 2.4% from the y axis).

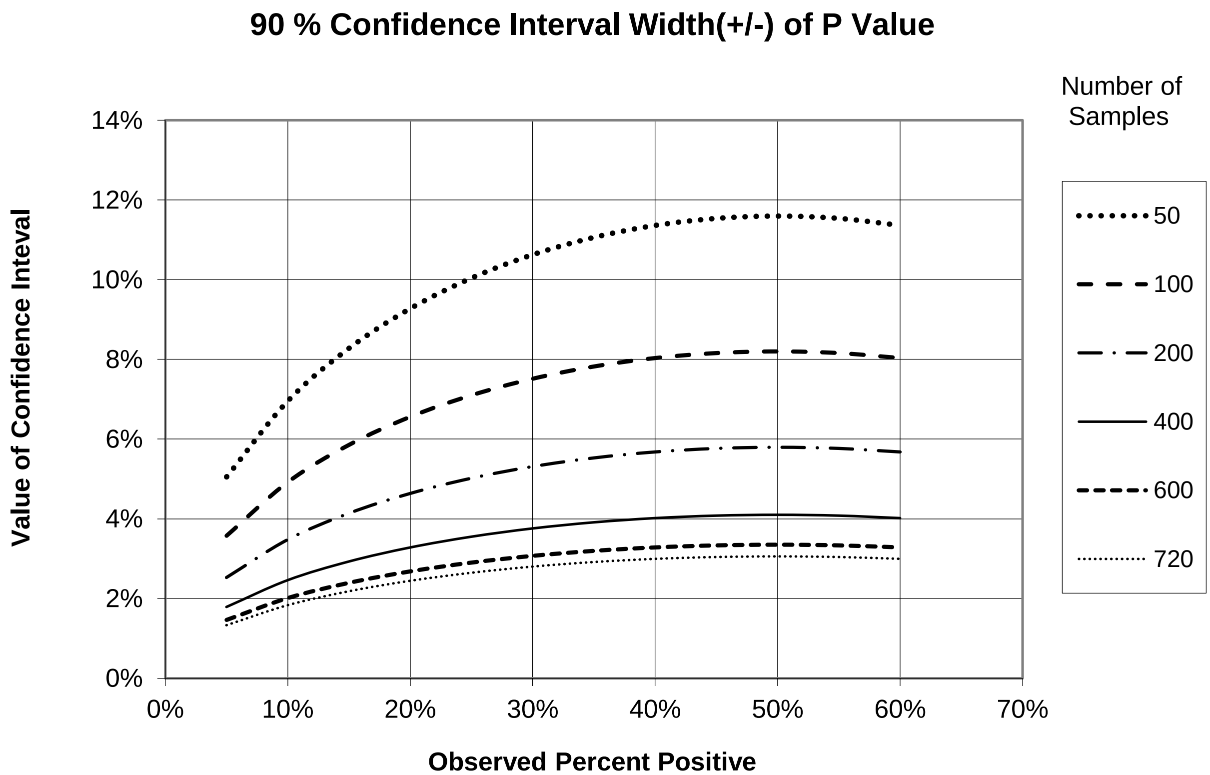

Figure 4. The width of the 90% confidence interval (value added to and subtracted from) the observed percent homes above a particular value. An illustrative example is provided in the text.

4.0 Large-Scale Sampling Plan

When the 10 home study was conducted, other sampling efforts conducted by the State had already indicated that the many MCHM concentrations at key locations within the distribution system were lower than the levels of concern discussed above. However, the 10 home study revealed that there were still significant concentrations of MCHM in all homes tested in eight counties and that there was some significant variability in both the concentrations and the variability between homes. The 10 home study was conducted in February 2014, approximately one month after the spill. This report is being completed in May 2014, four months after the spill. In the intervening months the concentrations in the WVAW distribution system will have continued to decrease due to flushing and turn-over of the water in the distribution system and in people's homes. Additionally, in March 2014 the WV TAP team suspected that there was low level of MCHM (< 1 ppb) leaching from within the WVAW treatment plant. WV TAP alerted WVAW of the possibility. WVAW immediately implemented a sampling plan that demonstrated that MCHM was desorbing off of the Granular Activated Carbon (GAC) filters. At the time this report was prepared, GAC was being removed from WVAW filters and was being replaced with fresh material. When the replacement of the GAC is completed, MCHM concentrations in the water distribution system and premise plumbing systems should decrease further as a remaining source of MCHM will have been removed. If the combination of flushing of the system and the removal of GAC as a persistent low level source causes continued decreases in MCHM concentration throughout the water system, then few samples, if any, will have concentrations near those observed in the 10 home study. It is possible that most, if not all, samples collected hereafter will be below the detection levels of 0.3 ppb, the MDL of the Eurofins laboratory.

However, it cannot be ruled out that there are reservoirs in the distribution system and in people's homes that could result in concentrations similar to those observed in house 8. If there is a desire to determine once and for all what MCHM levels are in affected homes and if, in fact, they are near zero, then an in-home tap water sampling program such as the one described in this report should be considered. This approach is an effective means for demonstrating that either the concentrations are well below the levels of concern or that there are persistent concentrations that need to be further addressed. Without the larger scale tap water sampling program, chemical levels in the affected area will remain unknown.

When the spill occurred, the distribution system was divided into 24 regions (see Figure 5) to expedite the water sampling and infrastructure flushing. It is proposed that these 24 regions also be used to organize and facilitate the proposed in-home tap water sampling approach; 30 homes per region should be sampled with three samples collected per home. Since the only contaminant that was detected in the 10 home study was MCHM, this is the only contaminant that is recommended for testing. This approach would address the questions outlined above and allow for robust answers to the questions. Based on the measurements made in the 10 home study including concentrations and variability the following conclusions can be reached regarding future sampling:

1. With a sample size of three samples per home, statistical power would be sufficient to determine if the concentrations observed in any one home could be safely considered to be below the upper value of the two estimates of the odor objection threshold concentration (9.5 ppb)
2. With a sample size of 13 samples per home, statistical power would be sufficient to determine if the concentrations observed in any one home could be safely considered to be below the odor recognition level (7.4 ppb).

3. If the goal of sampling is to determine if the concentrations measured in each home are below the Odor Threshold Concentration (0.55 ppb) then 5 samples per home would be required.
4. Sampling 30 homes per region will allow estimates of the average concentrations for each region with tight confidence interval that would allow for meaningful comparisons of the mean concentrations of all the regions.
5. If this hypothesis is rejected then at least one of the regions is different from the other regions. If this difference is positive and significant from a health or odor recognition perspective then more action may be required to continue the clean-up of the region(s) with higher concentrations.
6. A total of 720 homes would be sampled under this plan or 0.82% of the total number of residences affected. This sample size is statistically defensible and would allow for percentages of homes above or below any screening level to be calculated with very tight confidence levels even at very low percentages. The widths of the confidence interval for different percentage positive results at a sample size of 720 can be derived from Equation 3 and visualized on Figure 4. These estimates would be satisfactory for the results over the entire area affected but would not be useful for samples within any one of the 24 regions.

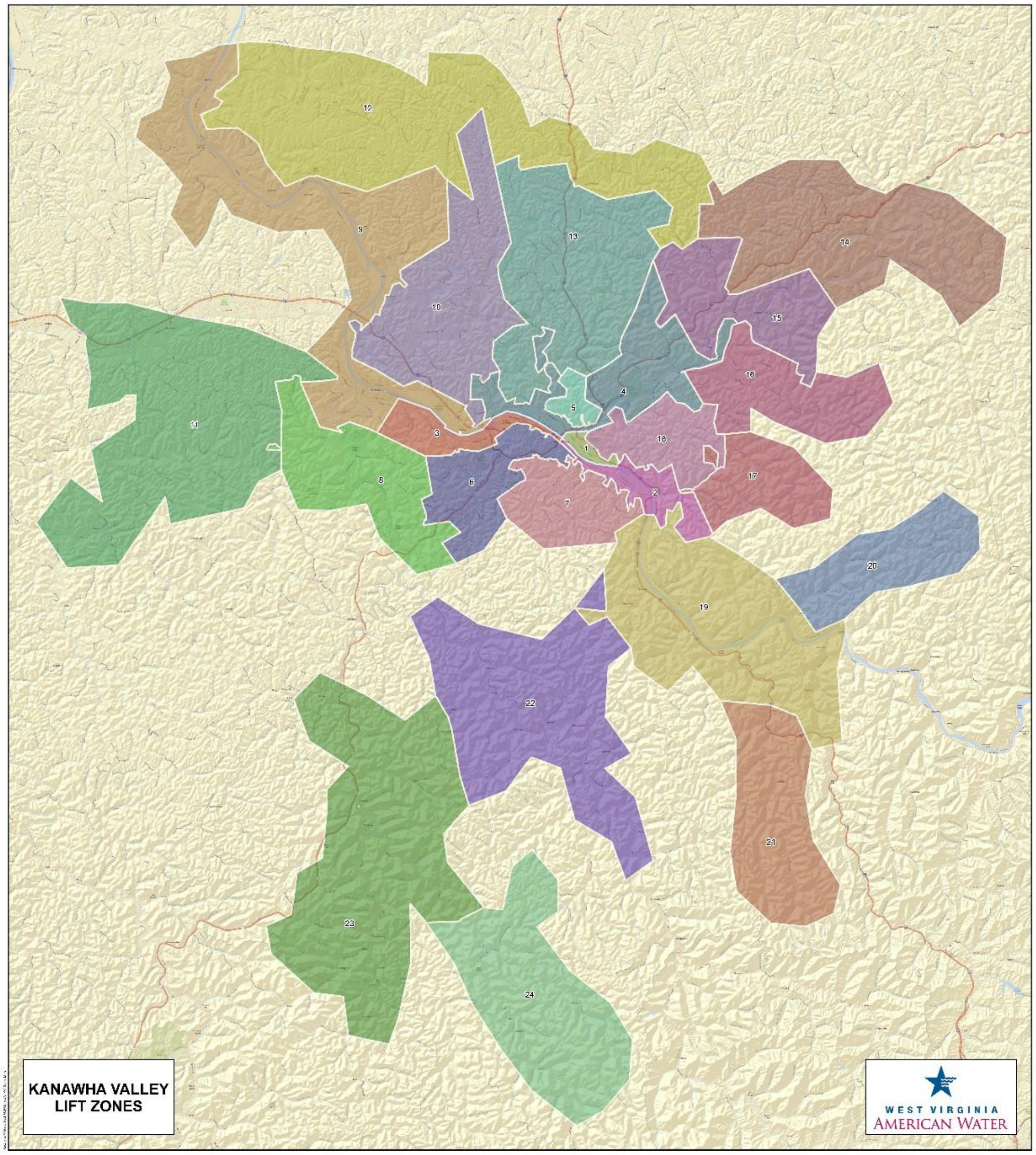

Figure 5. The 24 regions used to expedite sampling performed by the West Virginia National Guard.

Logistics

Implementation of this sampling plan will entail some complex logistics. First, a sampling plan of this type should be based on a random sample of homes. Since this sampling requires entry into private residences the plan cannot be completely randomized. A possible solution would be to develop a volunteer program to enable residents to volunteer to have their homes sampled. The volunteers would be grouped into the regions and random samples of 50 homes would be selected and prioritized. Fifty homes would be chosen to ensure that 30 homes within the set could be sampled. As with any sampling program, unforeseen difficulties are likely and an additional 20% of homes could be sampled to ensure a sufficient number of samples is collected and analyzed to allow answers to the questions with the certainty and power required. Second, the logistics of the 10 home sampling, which is a fraction of the effort that will be required for this project, demonstrated the difficulties in preparations, sample collection, sample shipment, sample tracking, data capture, data management, data analysis, results integration and reporting. A sampling project of this magnitude will require a significant logistical effort.

5.0 Recommendations

As noted in this and other reports from the WV TAP team, within a short time after the MCHM spill the MCHM concentrations to which people were exposed were, in general, far below the WV TAP health effects safe level and concentrations in the distribution system have likely decreased since then. Yet, as of the writing of this report, members of the public still report detecting odors in tap water that were not perceived before the spill. These persistent odors have contributed significantly to the continued distrust of authorities. This concern along with a general need to understand the science of MCHM fate and transport in distribution systems and premise plumbing may be sufficient motivation for a follow-on sampling effort.

Sampling that could improve our understanding of MCHM fate and transport in fully or partially-treated drinking water could include carefully designed:

- Sampling of the GAC filter effluent after replacement of the activated carbon,
- Sampling portions of the distribution system with different characteristics (e.g., different water ages), and
- Sampling homes with different premise plumbing configurations and materials.

Sampling that could improve confidence in WVAW and other authorities among consumers could entail a large-scale sampling effort aimed at establishing the proportion of buildings in different regions that have average MCHM concentrations above a level of concern. Two levels of concern that could be used are the WV TAP health effects safe level (120 ppb) and the odor objection threshold (9.5 ppb) established by consumer panel in the odor testing component of the WV TAP. The latter level is considerably more stringent than the former and would require significantly greater sample sizes. However, it may be more closely related to customer perception and a more appropriate target concentration if a major goal of the sampling program is to improve consumer confidence in the use of tap water that is currently being supplied to the affected area.

Sampling homes with different premise plumbing type would be very difficult because we do not know the plumbing type (materials and configuration) in each home are before sampling is conducted. Further, in many buildings, the premise plumbing types are mixed and inaccessible for inspection. While a sampling program with this kind of design would potentially be helpful in understanding differences in concentrations and odors, initial indications from the 10 home sampling program were inconclusive regarding the importance of plumbing materials and configuration and the logistical challenges of developing good statistical design would difficult to overcome.

Sampling to evaluate whether the concentrations measured in homes are below the Odor Threshold Concentrations (OTC) will be difficult to interpret because the OTC is so close to the current MDL. Reported values below the MRL (which is currently 0.8 ppb) have additional uncertainty which is difficult to quantify and therefore the power analysis is limited in it predictability.

Any sampling effort that is undertaken needs to carefully consider the logistics and quality control components of a defensible sampling program. This includes the details of:

- How the samples should be collected including the types of sample bottles and reagents included (like dechlorinating agent),
- Where and when the samples should be collected including if and how long the sample taps should be run before samples are collected,
- What requirements should be established for chain of custody tracking,
- What blanks and spikes should be included in the sampling program,
- Methods for sample shipments,
- Holding times for samples,
- Analytical methods,
- Required method and reporting limits,
- What surrogates should be used, and
- How the results will be reported, managed and analyzed.

The 10 home sampling did not address sampling and quantification of MCHM concentrations in apartment buildings or other larger structures. Sampling these structures would require a much more complicated sampling design. Neither resources nor time was available to attempt this more complicated sampling. In order to do this sampling, multiple units within the building would need to be sampled on multiple floors. Multiple locations in each unit would still be required. While this was not attempted as part of WV TAP it could be implemented but it would require extensive resources and planning.

6.0 References

Eastman Chemical Company, 2011. Safety Data Sheet; Crude MCHM. SDSUS / EN / 05, Version: 2.0, Revision date: 08/18/2011.

Helsel, Dennis R. 2005. Nondetects and Data Analysis: Statistics for Censored Environmental Data. Wiley and Sons, Inc. 250pp.