

Measurement error in estimated levels of disinfection by-products in tap water: implications for exposure assessment

E Symanski, D A Savitz, P C Singer

Occup Environ Med 2004;**61**:65–72

Aims: To assess spatial fluctuations, temporal variability, and errors due to sampling and analysis in levels of disinfection by-products in routine monitoring tap water samples and in water samples collected in households within the same distribution system for an exposure assessment study.

Methods: Mixed effects models were applied to quantify seasonal effects and the degree to which trihalomethane (THM) levels vary among households or locations relative to variation over time within seasons for any given location. In a separate analysis, the proportion of total variation due to measurement error arising from sampling and analysis was also quantified.

Results: THM levels were higher in the summer relative to other seasons. Differences in the relative magnitude of the intra- and inter-household components of variation were observed between the two sets of THM measurements, with a greater proportion of the variation due to differences within seasons for the routine monitoring data and a greater proportion of the variation due to differences across locations for the exposure assessment study data. Such differences likely arose due to differences in the strategies used to select sites for sampling and in the time periods over which the data were collected. With the exception of bromodichloromethane, measurement errors due to sampling and analysis contributed a small proportion of the total variation in THM levels.

Conclusions: The utility of routine monitoring data in assigning exposure in epidemiological studies is limited because such data may not represent the magnitude of spatial variability in levels of disinfection by-products across the distribution system. Measurement error contributes a relatively small proportion to the total variation in THM levels, which suggests that gathering a greater number of samples over time with fewer replicates collected at each sampling location is more efficient and would likely yield improved estimates of household exposure.

See end of article for authors' affiliations

Correspondence to:
Dr E Symanski, University
of Texas School of Public
Health, 1200 Herman
Pressler Blvd, Houston,
Texas 77030, USA;
esyanski@
sph.uth.tmc.edu

Accepted 11 January 2003

Chlorine, used by municipal water treatment facilities to disinfect water, reacts with naturally occurring organic matter to produce a host of compounds known as disinfection by-products.¹ In addition to chloroform, brominated species such as bromodichloromethane, dibromochloromethane, and bromoform may also be formed if bromide is present in the source water. Together, these volatile compounds comprise the trihalomethanes (THMs). Non-volatile disinfection by-products are also produced, which include haloacetic acids and haloacetonitriles. The extent of formation of specific compounds depends on treatment parameters (for example, the type and amount of disinfectant used), water parameters (for example, pH and temperature) and the amount and type of organic matter in the source water.¹ In general, however, THMs are the most abundant class of compounds present in treated drinking water, followed by haloacetic acids.²

Because water temperatures, the organic load in the water, and chlorine dosages are typically higher during the summer than in other seasons, THM levels tend to be highest during the warmest months of the year.^{3–6} In addition to systematic changes that occur seasonally, the hydraulic characteristics of the distribution system and water treatment parameters such as adjustments to chlorine dosages contribute to variations in THM concentrations from day to day, as well as diurnally within a 24 hour period.⁷ Additionally, the travel time to various locations within the distribution system (referred to as the residence time or water age) influences concentrations of disinfection by-products at sites served by the water utility,

with levels generally increasing the longer the water ages in the system. Since residence time is a function of the size and geometry of the pipe network, water demand, and distance from the treatment facility, higher THM levels may be more likely for households farther away from the water plant, for those areas served by smaller pipes or by more complex routing schemes, and where demands for water are low.^{7–8}

Epidemiological studies have evaluated associations between exposure to disinfection by-products and cancer,⁹ as well as adverse reproductive outcomes.^{10–11} Such studies have relied on different methods of assessing exposure, which presents difficulties in making comparisons between investigations and in generalising results.^{12–13} Studies investigating cancer outcomes necessarily dictate the application of retrospective methods to evaluate exposures and, therefore, require the use of surrogate (and likely error prone) measures of exposure because of the long historical periods of time for which requisite water quality information is not available. In contrast, exposure assessment in studies of reproductive outcomes should be relatively simpler because of the focus on more recent exposures for which water quality data are more readily available. Yet, despite the shorter interval over which the potential for exposure exists, these investigations have also been hindered by assessment strategies that likely introduced error in classifying exposure because of their reliance on sparse data and the failure to account for spatial and temporal fluctuations in THM levels across the distribution system over the time periods (weeks or months) relevant to studying pregnancy. Finally, another

Main messages

- Quantifying sources of variability in THM levels allows inferences to be made regarding the relative influences on exposure of factors related to the distribution system, macro- and micro-environmental conditions, treatment characteristics, and errors in sampling and analysis.
- Information about the relative magnitude of important components of variation facilitates the design of efficient sampling strategies for assessing exposure to disinfection by-products, particularly those associated with short term exposure events such as potential adverse reproductive and developmental health effects.
- The utility of routine municipal monitoring data in assigning exposure in epidemiological studies might be limited because such data may not represent the magnitude of spatial variability in levels of disinfection by-products across the distribution system.

complicating factor in investigations of increased health risks associated with exposures to disinfection by-products is proper identification of the putative agent, which toxicological studies suggest may be the halogenated acetic acids¹⁰ rather than the more commonly measured and more abundant THMs.

Despite widespread recognition of temporal and spatial variation in levels of disinfection by-products in distribution systems, investigations have not been conducted to quantify the degree to which these contaminants vary over time and across households served by the same water utility. Unlike the variability in occupational exposures that have been the focus of numerous investigations,^{14, 15} similar studies focusing on variation in contaminant levels in environmental samples have been limited^{16, 17} and no studies, to our knowledge, have examined disinfection by-products in tap water. Such questions are of critical importance when exposures to disinfection by-products are estimated on an individual (or household) basis because random variation in exposure over time produces error in the exposure assessment and tends to attenuate measures of effect.¹⁸

Given our lack of understanding of variability in levels of disinfection by-products, the primary objective of the current investigation was to evaluate sources of variation in THM levels in drinking water. In particular, we were interested in quantifying seasonal effects and the degree to which THM levels vary among households or locations in the same municipality (inter-household source of variation) relative to the variation in any one household (or location) over time (intra-household variation). Given that it is common practice to collect replicate tap water samples at each location during sampling surveys for quality control purposes, a secondary objective was to partition the intra-household source of variation into a component due to sampling and laboratory assay errors and a component due to the myriad factors that affect changes in THM levels from month to month within the same season.

MATERIALS AND METHODS

Compilation of a database of levels of disinfection by-products in tap water samples

One of the primary aims of a previous investigation conducted at the University of North Carolina (UNC), which was entitled The Chlorination By-product Exposure Assessment Study, was to develop and validate engineering

Policy implications

- Changes in municipal monitoring programmes in the United States with regard to the selection of sites for THM evaluation in the distribution system and in the frequency of sampling over the course of a year would enhance the usefulness of the data that are routinely collected.

models that predict individual household THM levels in tap water based on treatment parameters and hydraulic characteristics of the distribution system.^{19, 20} As part of this investigation, 17 households and other sites within the Orange Water and Sewer Authority (OWASA) service district in Orange County, North Carolina were identified for evaluation of THM levels in tap water. The households were selected to provide relatively uniform spatial coverage of the distribution system.

Water samples for this intensive survey were collected in July, August, September, and October of 1994 and in February, March, and April of 1995. During the seven sampling periods, samples were collected from each location on the same day (25/7/94, 3/8/94, 28/9/94, 26/10/94, 27/2/95, 31/3/95, 27/4/95); however, the time of day that each sample was collected was not recorded. Six locations were omitted during two of the sampling rounds (3/8/94 and 28/9/94) for logistic reasons. The water utility used free chlorine as both the primary and secondary disinfectant in their distribution system at the time these surveys were conducted. Chloroform comprises in excess of 80% of the total THM concentrations in this system.

Levels of chloroform, bromodichloromethane, dibromochloromethane, bromoform, and total THMs in tap water were evaluated using standard methods of sampling and analysis¹⁹⁻²¹ and are briefly described here. After allowing the water to run for five minutes, samples were collected into 40 ml glass vials containing a few crystals of ammonium sulphate to quench any free chlorine residual and prevent further formation of THMs. The vials were then sealed headspace-free with Teflon faced septa, placed in an ice chest, and refrigerated on return to the laboratory. THM concentrations were analysed using gas chromatography with electron capture detection after hexane extraction, with a method detection limit of 1 µg/L.²¹ For quality control, tap water samples were collected and analysed in triplicate. In the four surveys that were conducted in 1994, the concentration data of individual THM constituents and of total THMs in the triplicate samples were available;¹⁹ in the three surveys conducted in 1995, only the average of the measurements of the triplicate samples collected at each location were reported.²⁰

In addition to the monthly survey of 17 sites in the water utility's service district, quarterly routine compliance monitoring records were obtained for THM levels at five locations for 12 different sampling periods during the interval from February 1991 to November 1993.¹⁹ Sampling was carried out by the water utility in accordance with EPA procedures. Following standard practice by water utilities, all locations were sampled on the same day. The water samples were analysed for THMs by a commercial state approved laboratory. Routine monitoring data were available for chloroform, bromodichloromethane, dibromochloromethane, bromoform, and total THMs.

The routine monitoring data for 1994 and 1995 were also obtained as part of the current investigation to allow for more direct comparisons with the tap water measurements made

in the exposure assessment study. Nonetheless, the two data sets differ in that the exposure assessment study data were collected for four consecutive months in one year (in summer and autumn) and in three consecutive months in the second year (in winter and spring), whereas the routine monitoring data were gathered quarterly (approximately every three months) each year. Moreover, the 17 sites selected for sampling in the exposure assessment study were spatially distributed across the water system. In contrast, the routine monitoring data are from fewer sites, three of which represent average system residence times, the fourth the finished water at the point of entry to the distribution system, and the fifth a remote location with an expected high residence time.

The source of the data (water utility versus exposure assessment study), the dates and locations of sampling, a unique identifier for each of the triplicate samples collected at the same location (1994 exposure assessment study only), the specific contaminant measured, and the concentration values ($\mu\text{g/l}$) were input into the spreadsheet MS Excel (Microsoft Corp., Redmond, WA) and then converted into SAS (SAS Institute, Inc., Cary, NC) databases.

Assessment of variability in levels of disinfection by-products in tap water

Prior to our assessment of variability in levels of disinfection by-products in tap water, preliminary analyses were conducted to guide the selection of the most appropriate model. All of the bromoform measurements collected by the water utility were below detectable levels ($<1 \mu\text{g/l}$) and nearly 68% of the bromoform measurements were below detection limits in the exposure assessment study. Thus, subsequent analyses that focused on the individual constituents of the THMs excluded evaluation of the bromoform data. Also, higher than expected concentrations of brominated trihalomethanes during the October 1994 survey had been previously noted by Kachur,¹⁹ who speculated that they likely arose as a result of contamination of a chemical used in the water treatment process with bromide since it is known that the source water contains relatively low levels of naturally occurring bromide. Consequently, these data were excluded when the individual THM constituents were analysed to minimise effects due to the possibly anomalous levels of bromodichloromethane and dibromochloromethane detected during the last sampling round of 1994.

We examined temporal effects by visually inspecting graphs of the mean levels of total THMs during each sampling period (see fig 1). Consistent with results reported previously,³⁻⁶ these plots indicated seasonal trends in the data with noticeable peaks during the summer months. Finally, histograms of the measurements for total THMs suggested that the data were approximately lognormally distributed. Since chloroform is the predominant species among the THMs in the distribution system served by the water utility, a frequency distribution of those data produced almost identical plots as those for total THMs, whereas histograms of measurements of bromodichloromethane and dibromochloromethane appeared somewhat less right-skewed as concentrations of these species tended to be relatively small. Because the untransformed data were skewed to the right, we used the natural logarithm of the THM concentrations in our modelling of variability (see below) to produce distributions that were approximately normal.

Evaluation of seasonal effects and intra- and inter-household variation in levels of disinfection by-products

To account for seasonal effects while incorporating intra- and inter-household sources of variation in levels of

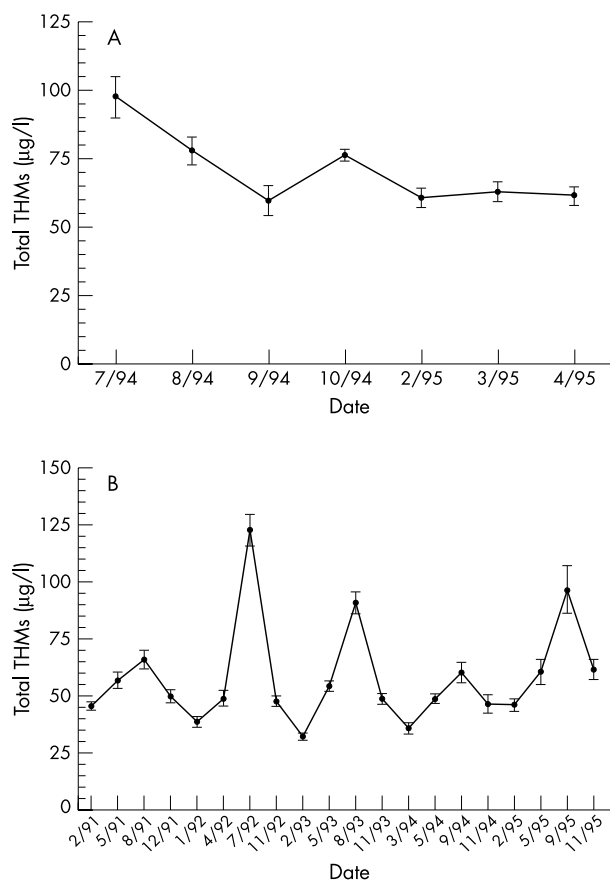


Figure 1 Temporal variations of the mean levels (± 1 SE) of total THMs in tap water samples ($\mu\text{g/l}$) from residences and other locations in Orange County, North Carolina; (A) exposure assessment study data; (B) routine compliance monitoring data.

contaminants in drinking water, a mixed effects linear model was applied to total THMs measured in tap water samples collected at 17 households surveyed during 1994 and 1995 and to routine monitoring data (total THMs; chloroform; bromodichloromethane; dibromochloromethane) collected at five locations by the water utility from 1991 through 1995. Separate analyses were also conducted on the subset of the monitoring samples collected during the same years as the exposure assessment study (1994–95) to enhance comparisons between the two sets of data. Seasonal differences were evaluated as a fixed effect with data stratified on the basis of when the samples were collected. The timing of sampling was broken down categorically into four periods (quarters): “winter” (January to March), “spring” (April to June), “summer” (July to September), and “autumn” (October to December). Two components of random variation were modelled: (1) a component of variation due to location or household; and (2) a component of variation due to differences across sampling periods within each season for each location as categorised above. These two components of variation comprise the inter- and intra-household sources of variation in disinfection by-product levels, respectively. Due to differences in the sampling strategies that were used to collect the data, the intra-household source of variation reflects year-to-year differences in THM levels within seasonal periods for the routine monitoring data and month-to-month differences within a seasonal period of the same year for the exposure assessment study data.

Briefly, the mixed effects model is specified as follows:

$$Y_{ijk} = \ln(X_{ijk}) = \mu_Y + \alpha_i + \beta_{ij} + \epsilon_{ijk} \quad (1)$$

for:

$i = 1, \dots, a$ seasonal effects

$j = 1, 2, \dots, b$ households

$k = 1, \dots, n_{ij}$ measurements from the j -th household in the i -th season,

and where:

X_{ijk} = the THM concentration for the j -th household on the k -th day in the i -th season

Y_{ijk} = the natural logarithm of the THM concentration

μ_Y = the overall mean of Y_{ijk}

α_i = the fixed effect due to the i -th season

β_{ij} = the random effect due to the j -th household in the i -th season

ϵ_{ijk} = the random error due to the k -th measurement from the j -th household in the i -th season.

It is assumed that the α_i sum to zero and thus have a population variance defined as

$$\sum_{i=1}^a \alpha_i^2 / (a - 1)$$

where $a = 4$ seasons in this application. Under the model, it is also assumed that β_{ij} and ϵ_{ijk} are mutually independent and normally distributed with zero means and variances σ_B^2 and σ_W^2 , respectively. Thus, σ_B^2 and σ_W^2 represent the inter- and intra-household variance components. We assumed a common covariance for all pairs of measurements collected at the same household (which is referred to as a compound symmetry covariance structure). It directly follows from the model that $E(Y_{ijk}) = \mu_Y + \alpha_i$ for all i, j, k ; $Var(Y_{ijk}) = \sigma_B^2 + \sigma_W^2$ for all i, j, k ; and $Cov(Y_{ijk}, Y_{ijk'}) = \sigma_B^2$ for $k \neq k'$ and for all i, j .

Quantification of measurement error related to sampling and analysis and intra- and inter-household variation in levels of disinfection by-products

Given the availability of the raw data containing the values of the THM concentrations in the triplicate samples collected and analysed during 1994, we were able to partition the variation in the measurements within a household from one sampling period to another (expressed by the term ϵ_{ijk} in model 1) into two separate components: (1) a component associated with errors in sampling and analysis; and (2) a component associated with factors that affect changes in THM levels from one month to the next. As before, "household" was modelled as a random effect. A fixed effect for season was omitted in the model because of the exclusion of the October 1994 data, which left only data collected during the summer (July, August, and September) to be analysed. Briefly, the two way random effects model can be specified as follows:

$$Y_{jkl} = \ln(X_{jkl}) = \mu_Y + \beta_j + \epsilon_{jk} + \eta_{jkl} \quad (2)$$

for:

$j = 1, 2, \dots, b$ households

$k = 1, \dots, n_j$ measurements from the j -th household

$l = 1, 2, 3$ triplicate samples for the k -th measurement from the j -th household

and where:

X_{jkl} = the THM concentration of 1-th replicate for the k -th measurement from the j -th household

Y_{jkl} = the natural logarithm of the THM concentration

μ_Y = the overall mean of Y_{jkl}

β_j = the random effect due to the j -th household

ϵ_{jk} = the random effect due to the k -th measurement from the j -th household

η_{jkl} = the random error due to the 1-th replicate for the k -th measurement from the j -th household.

It is assumed that β_j , ϵ_{jk} , and η_{jkl} are mutually independent and normally distributed with zero means and variances σ_B^2 , σ_W^2 , and σ_e^2 , respectively. As in model 1, a compound symmetry covariance structure was assumed. It follows that $E(Y_{jkl}) = \mu_Y$ for all j, k, l ; $Var(Y_{jkl}) = \sigma_B^2 + \sigma_W^2 + \sigma_e^2$ for all j, k, l ; $Cov(Y_{jkl}, Y_{jkl'}) = \sigma_B^2 + \sigma_W^2$ for $l \neq l'$ and for all j, k and $Cov(Y_{jkl}, Y_{jk'l'}) = \sigma_B^2$ for $l \neq l', k \neq k'$, and for all j .

Under models 1 and 2, restricted maximum likelihood estimates of the variance components were obtained using PROC MIXED available with SAS software (SAS Institute, Cary, NC).

RESULTS

Table 1 summarises the concentrations of disinfection by-products in tap water samples collected by the water utility and in the exposure assessment study. As expected, chloroform comprised approximately 80% of the total THM concentrations in this system. While the periods of data collection do not directly coincide, it appears that average levels of total THMs were lower for the five locations sampled by the water utility (mean 56.8 $\mu\text{g/l}$; 3/94–11/95) compared to the 17 locations sampled in the exposure assessment study (mean 71.1 $\mu\text{g/l}$; 7/94–4/95). In evaluating the results of the correlation analysis to investigate relations between levels of total THMs and individual THM constituents (see table 2), total THM levels were highly correlated with chloroform and bromochlorodichloromethane, but not with dibromochloromethane. The individual THM species correlated well with each other with the exception of chloroform and dibromochloromethane.

Evaluation of seasonal effects and intra- and inter-household variation in levels of disinfection by-products

We used mixed effects models to evaluate sources of variation in levels of disinfection by-products and obtain estimates of the fixed effect due to season and the intra- and inter-household variance components. As shown in table 3, our results indicate that seasonal effects were important predictors of THM levels. Relative to summertime levels of THMs, concentrations were considerably lower in the autumn and spring with the lowest levels detected during the winter months for the routine monitoring data. Seasonal differences were also detected in the exposure assessment study data with total THM levels decreasing during the winter and spring periods. To further examine the influence of seasonal effects, a random effects model that excluded the fixed effect for season was fitted to each data set. In making comparisons between the random effects and mixed effects models, accounting for seasonal influences decreased the estimated total variance in total THM levels for the exposure assessment study data from 0.080 to 0.068 (15%). A much larger decrease in the total variance between the two models was observed for the routine monitoring data (1991–95), which fell from 0.123 to 0.048 (61%). We also computed the proportion of the intra-household variation explained by season

$$\left[i.e., \left(\frac{\hat{\sigma}_{W_1}^2 - \hat{\sigma}_{W_2}^2}{\hat{\sigma}_{W_1}^2} \right) \right]$$

where $\hat{\sigma}_{W_1}^2$ and $\hat{\sigma}_{W_2}^2$ are the intra-household variance components obtained from the one way random effects and mixed effects models, respectively.²² The intra-household

Table 1 Trihalomethane concentrations in tap water (µg/l) from residences and other locations in Orange County, North Carolina from two data sources

	b*	n*	N*	Mean	Median	Min	Max	SD
Exposure assessment study (7/94–10/94)								
Chloroform	17	3–4	56	62.3	56.5	30.2	141	23.1
Bromodichloromethane	17	3–4	56	14.3	13.4	9.41	19.7	3.32
Dibromochloromethane	17	3–4	56	3.08	2.26	1.48	5.86	1.46
Total trihalomethanes	17	3–4	56	79.7	75.0	41.7	161	24.0
Exposure assessment study (7/94–4/95)†								
Total trihalomethanes	17	6–7	107	71.1	67.5	41.7	161	21.7
Routine monitoring data (2/91–11/95)								
Chloroform	5	20	100	46.9	41.1	22.4	128	21.0
Bromodichloromethane	5	20	100	9.38	8.95	4.00	19.7	3.07
Dibromochloromethane	5	20	100	1.48	1.20	<1	3.70	0.734
Total trihalomethanes	5	20	100	57.7	52.2	27.4	142	23.3
Routine monitoring data (3/94–11/95)								
Chloroform	5	8	40	44.4	40.7	23.4	98.3	18.5
Bromodichloromethane	5	8	40	10.4	10.2	5.60	16.1	2.34
Dibromochloromethane	5	8	40	1.95	1.80	<1	3.70	0.697
Total trihalomethanes	5	8	40	56.8	52.3	30.6	114	20.3

*b = no. of sampled locations (or households); n = no. of measurements per location; N = total number of measurements.

†In the three surveys conducted in 1995, measurements were only reported for total trihalomethane levels.

variance component was reduced by 44% for the exposure assessment study data (with a drop from 0.048 to 0.027), and by 61% for the routine monitoring data (with a drop from 0.115 to 0.045). The analyses were repeated for the subset of monitoring data collected during 1994–95 and similar reductions were observed.

In evaluating the relative contributions of the sources of variation in total THM levels, inconsistent results were obtained. Controlling for seasonal effects, more of the variation in the exposure assessment study data was attributable to differences in average levels of total THMs across residences ($\hat{\sigma}_B^2 = 0.041$) compared to fluctuations over time ($\hat{\sigma}_W^2 = 0.027$). Viewed in terms of the proportion of the variation explained by each component of variance, almost two thirds of the total variance ($\hat{\sigma}_Y^2 = 0.068$) was due to differences across residences (61%) and a little more than one third of the variance was due to intra-seasonal differences over time (39%). In contrast, the source of variation across locations contributed much less to the total variability in total THM levels in the routine monitoring data collected during the same period (1994–95) (22%) when compared to the source of variation within seasons (78%). A similar breakdown in the proportion of the variation explained by each component of variance was obtained for the monitoring data collected over the period 1991–95 for total THM levels and individual THM species with the

exception of dibromochloromethane, which exhibited far less variation between locations (2%) than variation over time (98%).

Quantification of measurement error related to sampling and analysis, and intra- and inter-household variation in levels of disinfection by-products

As shown in table 4 for total THMs, the estimates of the components of variation (and percentage of total variation) due to differences between households, differences over time, and differences in concentration levels among the triplicate samples arising from errors in sampling and analysis were 0.077 (58%), 0.041 (31%), and 0.013 (10%), respectively. Similar results were observed for chloroform, with somewhat less variation in average levels exhibited among households (52%) and somewhat more variation observed across sampling periods (39%). Whereas a substantial proportion of the variation in bromodichloromethane levels was due to differences between households (82%), the household source of variation (37%) was far less influential for dibromochloromethane.

DISCUSSION

In this study, we were able to examine seasonal effects on levels of disinfection by-products along with the degree to

Table 2 Estimated Spearman rank correlation coefficients between levels of total trihalomethanes and individual THM constituents

	Chloroform	Bromodichloromethane	Dibromochloromethane	Total trihalomethanes
Exposure assessment data: 7/94–9/94 (n = 39)				
Chloroform	–	0.833*	0.226	0.996*
Bromochlorodichloromethane	0.833*	–	0.672*	0.856*
Dibromochloromethane	0.226	0.672*	–	0.270
Total trihalomethanes	0.996*	0.856*	0.270	–
Routine monitoring data: 2/91–11/95 (n = 100)				
Chloroform	–	0.707*	0.237*	0.982*
Bromochlorodichloromethane	0.707*	–	0.762*	0.810*
Dibromochloromethane	0.237*	0.762*	–	0.376*
Total trihalomethanes	0.982*	0.810*	0.376*	–
Routine monitoring data: 3/94–11/95 (n = 40)				
Chloroform	–	0.754*	0.063	0.981*
Bromochlorodichloromethane	0.754*	–	0.465*	0.834*
Dibromochloromethane	0.063	0.465*	–	0.182
Total trihalomethanes	0.981*	0.834*	0.183	–

*p<0.05.

Table 3 Seasonal effects and inter- and intra-household variation in (natural log transformed) levels of disinfection by-products in tap water ($\mu\text{g}/\text{l}$) from two different data sources

	Seasonal effect† [$\hat{\alpha}_i$ (SE)]			Inter-household variation ($\hat{\sigma}_B^2$)	% of total variation‡	Intra-household variation ($\hat{\sigma}_W^2$)	% of total variation‡
	Autumn	Winter	Spring				
Exposure assessment study (1994–95) (n = 107)							
Total trihalomethanes	-0.057 (0.048)	-0.285* (0.039)	-0.287* (0.048)	0.041	61	0.027	39
Routine monitoring data (1994–95) (n = 40)							
Total trihalomethanes	-0.349* (0.095)	-0.632* (0.095)	-0.332* (0.095)	0.013	22	0.045	78
Routine monitoring data (1991–95) (n = 100)							
Total trihalomethanes	-0.511* (0.053)	-0.765* (0.053)	-0.451* (0.053)	0.012	26	0.036	74
Chloroform	-0.557* (0.060)	-0.806* (0.060)	-0.495* (0.060)	0.011	20	0.045	80
Bromodichloromethane	-0.305* (0.060)	-0.635* (0.060)	-0.262* (0.060)	0.016	26	0.045	74
Dibromochloromethane	-0.081 (0.126)	-0.319* (0.126)	-0.050 (0.126)	0.003	2	0.197	98

†Summer: referent category.

‡Estimated total variation: $\hat{\sigma}_Y^2 = \hat{\sigma}_B^2 + \hat{\sigma}_W^2$ * $p < 0.05$.

which contaminant levels vary among households served by the same distribution system relative to the magnitude of variability over time within seasonal periods. To our knowledge, this is the first study that provides quantitative estimates of both seasonal effects and of the inter- and intra-household sources of variation in THM levels. Such a methodological approach is beneficial in enhancing our understanding of the important sources of variation in levels of disinfection by-products in tap water and for assessing exposure when epidemiological studies are carried out.

In the mixed models that were applied, we confirmed seasonal differences in THM levels as reported previously in the literature.^{3–6} Yet, the magnitude of the effect was not consistent when comparisons were made between the two sets of data that were analysed. While the mean levels of total THMs in the water utility samples (1994–95 data) and exposure assessment study data collected during the summer months were similar (83.5 $\mu\text{g}/\text{l}$ and 80.24 $\mu\text{g}/\text{l}$, respectively), larger differences in THM levels were apparent for samples collected during spring (53.0 versus 75.78 $\mu\text{g}/\text{l}$), autumn (49.9 versus 60.2 $\mu\text{g}/\text{l}$), and winter (38.7 versus 60.3 $\mu\text{g}/\text{l}$). However, it is difficult to make direct comparisons because the seasonal effect was evaluated using seven months of data in the exposure assessment study and averaged over a full two year period in the monitoring data collected quarterly by the water utility. While differences in THM levels for any given season from year to year were not formally evaluated, seasonal levels in the routine monitoring data appeared to fluctuate over the five year period (data not shown). Likewise, annual total THM levels varied with values of 54.3, 64.3, 56.3, 47.6, and 66.0 $\mu\text{g}/\text{l}$ for 1991 to 1995. While these results contrast the findings reported in a five year study²³ conducted in the northwest of England in which yearly mean levels showed little variation, the earlier investigation relied on many more measurements of THMs in water samples collected in 288 water zones supplied by different distribution systems.

Partitioning the total variation in exposure into separate components²⁴ allows discrimination between inter- and intra-individual differences in exposure as well as evaluation of the magnitude of their relative effects. While such methods have been extensively applied in the occupational arena to both air¹⁴ and biological¹⁵ monitoring data, much less work has focused on investigations of similar variation in environmental contaminants irrespective of the type of samples collected.^{16–17, 25–26} In evaluating variability in levels of total THMs collected in the exposure assessment study of disinfection by-products in tap water, a substantial proportion of the variation (61%) was attributable to differences across residences as compared to variation over time (within

a season). While these results suggest a moderate degree of variation in THM levels from month to month, repeated measurements within each seasonal period were limited and based on only seven months of data. Thus, future studies should sample more frequently to ensure that the full range of exposures is reflected in the data that are collected.

Equivocal findings regarding the relative magnitude of the intra- and inter-household sources of variation could have arisen because differences in the objectives of each monitoring programme dictated differences in the strategies employed to select locations for sampling and in the time periods over which the data were collected. In contrast to the results obtained with the exposure assessment study data, the variation in THM levels among locations was small relative to that occurring between sampling periods in the data collected by the water utility. Similar results were observed when the routine monitoring data were restricted to the same years in which the exposure assessment study data were collected. Because different sampling strategies were used to collect the different sets of data, the intra-household source of variation reflects year to year differences in THM levels within seasonal periods for the routine monitoring data and month to month differences within seasons of the same year for the exposure assessment study data. Thus, the water utility samples would be expected to exhibit greater temporal variability within seasons, which they did. Diurnal variation may have contributed to the intra-household source of variation in levels of disinfection by-products,⁷ but it was not possible to evaluate the magnitude of this component of variation in the data that were collected.

With respect to spatial variability, the locations sampled for the exposure assessment study data set were chosen to capture differences across the distribution system. In contrast, only one of the water utility's sampling sites, in accordance with the US EPA's compliance monitoring requirements,^{27, 28} was selected at a remote location, and three of the other sites were selected to have average residence times and presumably similar THM levels. To the extent that our results reflect routine monitoring data, in general, choosing a small number of sites in this manner fails to represent the magnitude of spatial variability in levels of disinfection by-products across the distribution system and raises concerns regarding the utility of using such data in assigning exposure for epidemiological studies, particularly those associated with short term exposure events such as potential adverse reproductive and developmental health outcomes. However, in accordance with anticipated new regulatory requirements in the United States,²⁹ municipalities will soon be monitoring every two months at eight locations over a 12 month period as part of an initial distribution

Table 4 Sources of variation trihalomethane levels in tap water due to inter- and intra-household differences and measurement error

	Inter-household variation ($\hat{\sigma}_B^2$)	% of total variation*	Intra-household variation ($\hat{\sigma}_W^2$)	% of total variation*	Sampling and analytical error ($\hat{\sigma}_e^2$)	% of total variation*
Total trihalomethanes	0.077	58	0.041	31	0.013	10
Chloroform	0.090	52	0.067	39	0.017	10
Bromodichloromethane	0.026	82	0.0002	1	0.006	18
Dibromochloromethane	0.021	37	0.025	45	0.010	18

Data source: exposure assessment study, July–September 1994, n = 111 measurements.

*Estimated total variation: $\hat{\sigma}_Y^2 = \hat{\sigma}_B^2 + \hat{\sigma}_W^2 + \hat{\sigma}_e^2$; percentages may not add up due to rounding.

system evaluation. These monitoring data may prove to be more informative in the future regarding variability in THM exposure in drinking water, particularly if coupled to additional sampling campaigns that would allow the full range of exposures (temporally and spatially) to be evaluated.

Partitioning of variability allows inferences to be made regarding the relative influences on exposure of factors related to the distribution system, macro- and micro-environmental conditions, treatment characteristics, and errors in sampling and analysis. Such information facilitates the design of efficient sampling strategies. For example, our results for THMs indicate that variation from month to month is more influential than errors associated with sampling and analysis (table 4). As such, in the face of limited resources, a sampling strategy that collected greater numbers of quality assured measurements over time with fewer replicates collected at each sampling location would be more efficient and would likely yield improved estimates of (household) exposure.

It is important to note, however, that the relative influence of important sources of variation depends on the interval over which inferences are to be drawn. On the one hand, for example, if average annual exposures were of primary interest as they are in relation to chronic exposure associated with carcinogenesis, then the variation in disinfection by-products would be more appropriately modelled without a seasonal component. On the other hand, for reproductive effects that likely manifest themselves during the first three months of pregnancy, modelling approaches that rely on traditional seasonal classifications may not provide enough resolution in assessing exposure and, instead, may have to be refined to reflect the actual periods in which the critical exposure is received.

The implications of exposure variability on epidemiological studies also depend on whether an individually or group based exposure assessment strategy is adopted. As noted earlier, random variation in exposure over time induces error and, under certain modelling assumptions, operates to attenuate measures of effect.¹⁸ The variability inherent in group based strategies in which, for example, the average THM level of a water distribution system over a specified period of time might be assigned to all individuals served by that system induces a different type of error (referred to as Berkson error), which tends to increase the imprecision of measures of effect and thereby decrease power rather than to bias results.³⁰ While it was beyond the scope of the current investigation to make comparisons to other distribution systems, a previous study²³ that relied on THM data routinely collected for compliance purposes by water companies in the United Kingdom from 1992 to 1996 reported that most of the variability was due to differences between water zones rather than within zones. It is important to note, however, that the sampling strategies employed by water municipalities in the UK and the USA are not the same in that THM concentrations are measured in samples from randomly selected taps in each water zone in the UK, whereas water samples are

collected from the same locations (chosen on the basis of their expected residence times) in a single distribution system in the USA. Such differences have implications for the structure of the multilevel models that the data support and the components of variability that the (different) models represent. Further research is needed to determine the pattern of variability in disinfection by-product levels between and within distribution systems, as well as to evaluate whether one source of variation appears to predominate for specific geographic locations or periods of calendar time.

In conclusion, more sophisticated strategies to evaluate exposures to chlorinated by-products in drinking water will depend on improved measures that account for variability in contaminant levels in tap water, as well as differences in water intake and personal water-use activities. Indeed, the methods used to examine intra- and inter-household variation in THM levels are applicable to examining intra- and inter-individual variability in water consumption and water related activities.³¹ While there has been some suggestion in the literature that variability in tap water THM levels may play a larger role than behavioural characteristics (for example, showering duration) in determining individual exposure,³² the relative influence of variation in water concentration and in patterns of water consumption and water use on individual exposure warrants further investigation.

ACKNOWLEDGEMENTS

The data used in the current investigation were collected for a study funded by the US Environmental Protection Agency (Cooperative Agreement CR 820076). The authors are grateful to Stephen Kachur and Daniel Hull for collecting and analysing the tap water samples, and to the Orange Water and Sewer Authority for providing us with their compliance monitoring data.

Authors' affiliations

E Symanski, Environmental Sciences, University of Texas School of Public Health, Houston, Texas, USA

D A Savitz, Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill, North Carolina, USA

P C Singer, Department of Environmental Sciences and Engineering, School of Public Health, University of North Carolina, Chapel Hill, North Carolina, USA

REFERENCES

- 1 **Singer PC**. Control of disinfection by-products in drinking water. *J Environ Eng* 1994;**120**:727–44.
- 2 **Krasner S**, McGuire M, Jacangelo J, *et al*. The occurrence of disinfection by-products in US drinking water. *J Am Water Works Assoc* 1989;**81**:41–53.
- 3 **Brett RW**, Calverley RA. A one-year survey of trihalomethane concentration changes within a distribution system. *J Am Water Works Assoc* 1979;**71**:515–20.
- 4 **Singer PC**, Barry JJ, Palen GM, *et al*. Trihalomethane formation in North Carolina drinking waters. *J Am Water Works Assoc* 1981;**73**:392–401.
- 5 **Arora H**, LeChevallier MW, Dixon KL. DBP occurrence survey. *J Am Water Works Assoc* 1997;**89**:60–8.
- 6 **Chen WJ**, Weisel CP. Halogenated DBP concentrations in a distribution system. *J Am Water Works Assoc* 1998;**90**:151–63.

- 7 **Singer PC.** Variability and assessment of disinfection by-product concentrations in water distribution systems. In: Craun GF, Hauchman FS, Robinson DE, eds. *Microbial pathogens and disinfection by-products in drinking water: health effects and management of risks*. Washington, DC: International Life Sciences Institute, 2001:211–13.
- 8 **Waller K,** Swan SH, Windham GC, *et al.* Influence of exposure assessment methods on risk estimates in an epidemiologic study of total trihalomethane exposures and spontaneous abortion. *J Exposure Anal Environ Epidemiol* 2001;**11**:522–31.
- 9 **Morris RD,** Audet A-M, Angelillo IF, *et al.* Chlorination, chlorination by-products and cancer: a meta analysis. *Am J Public Health* 1992;**82**:955–63.
- 10 **Nieuwenhuijsen MJ,** Toledano MB, Eaton NE, *et al.* Chlorination disinfection by-products in water and their association with adverse reproductive outcomes, a review. *Occup Environ Med* 2000;**57**:73–85.
- 11 **Bove F,** Shim Y, Zeitz P. Drinking water contaminants and adverse pregnancy outcomes: A Review. *Environ Health Perspect* 2002;**110**(suppl 1):61–74.
- 12 **Nieuwenhuijsen MJ,** Toledano MB, Elliott P. Uptake of chlorination disinfection by-products; a review and discussion of its implications for exposure assessment in epidemiological studies. *J Exposure Anal Environ Epidemiol* 2000;**10**:586–99.
- 13 **Arbuckle TE,** Hrudey SE, Krasner SW, *et al.* Assessing exposure in epidemiologic studies to disinfection by-products in drinking water: report from an international workshop. *Environ Health Perspect* 2002;**110**(suppl 1):53–60.
- 14 **Kromhout H,** Symanski E, Rappaport SM. A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. *Ann Occup Hyg* 1993;**37**:253–70.
- 15 **Symanski E,** Greeson NMH. Assessment of variability in biomonitoring data using a large database of biological measures of exposure. *Am Ind Hyg Assoc J* 2002;**64**:390–401.
- 16 **Brunekreef B,** Noy D, Clausen P. Variability of exposure measurements in environmental epidemiology. *Am J Epidemiol* 1987;**125**:892–8.
- 17 **Emond MJ,** Lanphear BP, Watts A, *et al.* Measurement error and its impact on the estimated relationship between dust lead and children's blood lead. *Environ Res* 1997;**72**:82–92.
- 18 **Armstrong BG.** Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ Med* 1998;**55**:651–6.
- 19 **Kachur SB.** An evaluation of methods for predicting total trihalomethane concentrations in exposure assessment studies. Master of Science technical report. Chapel Hill, NC: Department of Environmental Sciences and Engineering, University of North Carolina, 1994.
- 20 **Hull DT.** An evaluation of methods for predicting total trihalomethane concentration in finished drinking water. Master of Science technical report. Chapel Hill, NC: Department of Environmental Sciences and Engineering, University of North Carolina, 1995.
- 21 **American Public Health Association (APHA),** American Water Works Association (AWWA), Water Pollution Control Federation (WPCF), Clesceri LS, Greenberg AE, Trussell RR, eds. *Standard methods for the examination of water and wastewater*. Washington, DC: APHA, AWWA, and WPCF, 1989.
- 22 **Byrk AS,** Raudenbush SW. *Hierarchical linear models: applications and data analysis methods*. Newbury Park: Sage, 1992:64–5.
- 23 **Keegan T,** Whitaker H, Nieuwenhuijsen MJ, *et al.* Use of routinely collected data on trihalomethane in drinking water for epidemiological purposes. *Occup Environ Med* 2001;**58**:447–52.
- 24 **Searle SR,** Casella G, McCulloch CE. *Variance components*. New York: John Wiley & Sons, 1992.
- 25 **Gammon MD,** Wolff MS, Neugut AI, *et al.* Temporal variation in chlorinated hydrocarbons in healthy women. *Cancer Epidemiol Biomarkers Prevention* 1997;**6**:327–32.
- 26 **Macintosh DL,** Kabiru C, Scanlon KA, *et al.* Longitudinal investigation of exposure arsenic, cadmium, chromium, and lead via beverage consumption. *J Exposure Anal Environ Epidemiol* 2000;**10**:196–205.
- 27 **US Environmental Protection Agency (US EPA).** National interim primary drinking water regulations; control of trihalomethanes in drinking water. *Fed Reg* 44: 68624, 1979;(231).
- 28 **US Environmental Protection Agency (US EPA).** National primary drinking water regulations: disinfectants and disinfection byproducts; final rule. *Fed Reg* 1998;**63**(241):69390–9476.
- 29 **US Environmental Protection Agency (US EPA).** Stage 2. Microbial and disinfection by-products. Federal Advisory Committee agreement in principle. *Fed Reg* 2000;**65**(251):83015–83024.
- 30 **Armstrong BG.** The effects of measurement errors on relative risk regressions. *Am J Epidemiol* 1990;**132**:1176–84.
- 31 **Shimokura GH,** Savitz DA, Symanski E. Assessment of tap water use for estimating exposure to trihalomethanes. *Environ Health Perspect* 1998;**106**:55–9.
- 32 **Weisel CP,** Kim H, Haltmeier P, *et al.* Exposure estimates to disinfection by-products of chlorinated drinking water. *Environ Health Perspect* 1999;**107**:103–10.



Want to know more?

Data supplements

Limited space in printed journals means that interesting data and other material are often edited out of articles; however, limitless cyberspace means that we can include this information online. Look out for additional tables, references, illustrations.

www.occenvmed.com