ORIGINAL ARTICLE

Identifying public water facilities with low spatial variability of disinfection by-products for epidemiological investigations

A F Hinckley, A M Bachand, J R Nuckols, J S Reif

See end of article for authors' affiliations
....

Correspondence to:
Dr A Hinckley, Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, CO 80523-1681, USA; ahinckley@cdc.gov

Accepted 1 February 2005
....

Background and Aims: Epidemiological studies of disinfection by-products (DBPs) and reproductive outcomes have been hampered by misclassification of exposure. In most epidemiological studies conducted to date, all persons living within the boundaries of a water distribution system have been assigned a common exposure value based on facility-wide averages of trihalomethane (THM) concentrations. Since THMs do not develop uniformly throughout a distribution system, assignment of facility-wide averages may be inappropriate. One approach to mitigate this potential for misclassification is to select communities for epidemiological investigations that are served by distribution systems with consistently low spatial variability of THMs.

Methods and Results: A feasibility study was conducted to develop methods for community selection using the Information Collection Rule (ICR) database, assembled by the US Environmental Protection Agency. The ICR database contains quarterly DBP concentrations collected between 1997 and 1998 from the distribution systems of 198 public water facilities with minimum service populations of 100 000 persons. Facilities with low spatial variation of THMs were identified using two methods; 33 facilities were found with low spatial variability based on one or both methods. Because brominated THMs may be important predictors of risk for adverse reproductive outcomes, sites were categorised into three exposure profiles according to proportion of brominated THM species and average TTHM concentration. The correlation between THMs and haloacetic acids (HAAs) in these facilities was evaluated to see whether selection by total trihalomethanes (TTHMs) corresponds to low spatial variability for HAAs. TTHMs were only moderately correlated with HAAs ($r = 0.623$).

Conclusions: Results provide a simple method for a priori selection of sites with low spatial variability from state or national public water facility datasets as a means to reduce exposure misclassification in epidemiological studies of DBPs.

A growing body of epidemiological evidence has been published suggesting that exposure of pregnant women to disinfection byproducts (DBPs) is associated with an increased risk of some adverse reproductive outcomes, as reported in reviews by Nieuwenhuijsen and colleagues[1] and Bove and colleagues.[2] However, there is general agreement that limitations in exposure assessment have resulted in substantial misclassification of exposure in these studies, making the true risk difficult to determine.[1–4] For example, in most epidemiological studies conducted to date, all persons living within the boundaries of a water distribution system have been assigned a common exposure value based on facility-wide averages of trihalomethane (THM) concentrations for a point in time of interest. Use of public water facility monitoring data can result in exposure misclassification when persons living at different locations within a distribution system are exposed to significantly different concentrations of THMs but are assigned a common exposure level.

Local conditions at the water treatment facility such as the quantity of organic matter in source water, chlorine dose, pH, temperature, and bromide ion concentration can vary across the distribution system, leading to variability in the formation of DBPs. In particular, prolonged chlorine contact times result in higher, non-uniform, THM concentrations across a distribution system.[5] In two previous studies, residences situated farthest hydraulically from the source of treatment or at dead-end points in the distribution system had THM levels 30–150% higher than those measured at the treatment plant. Increased THM production was primarily attributable to higher chlorine residual levels at these distant locations.[6 7] In contrast, haloacetic acids (HAAs) are more likely to degrade with increasing residence time in the distribution system, due to biological degradation, although they can continue to form in the presence of chlorine and organic precursors.[5 8]

The US Environmental Protection Agency (USEPA) collected data under the Information Collection Rule (ICR) to assemble a water facility database system to supply information on parameters related to pathogen occurrence and DBP formation.[9] The data were collected in quarterly sampling events from July 1997 to December 1998 at public water distribution systems across the USA that served at least 100 000 people and included measurements of THMs and HAAs. The ICR database lends itself to the study of intra-system spatial variability of THMs since it represents a

....

Abbreviations: ANOVA, analysis of variance; AVG1 and AVG2, average residence time of water in the distribution system; BIF, bromine incorporation factor; $CHBrCl_2$, bromodichloromethane; $CHBr_3$, bromoform; $CHCl_3$, chloroform; $CHBr_2Cl$, dibromochloromethane; DBP, disinfection by-product; DSE, distribution system equivalent; HAA, haloacetic acid; ICR, Information Collection Rule; HAA5, sum of five haloacetic acids (monobromoacetic, monochloroacetic, dibromoacetic, dichloroacetic and trichloroacetic acids); MAX, maximum residence time of water in the distribution system; THM, trihalomethane; TTHMs, total trihalomethanes; USEPA, United States Environmental Protection Agency

Main messages

- Epidemiological studies of disinfection by-products (DBPs) and reproductive outcomes have been hampered by misclassification of exposure, in part due to spatial variability of these contaminants within water distribution systems.
- One approach to mitigate exposure misclassification is to select communities for epidemiological investigations that are served by distribution systems with consistently low spatial variability of DBPs.
- We investigated two approaches to identify low spatial variability: the first based on a statistical analysis of variability (ANOVA) and the second based on concentration cut-points for total trihalomethanes (TTHM) derived from prior epidemiological studies of birth outcomes. Both methods appear to be useful for community selection.

Policy implications

- Application of these techniques to identify communities with low spatial variability will reduce exposure misclassification and make the results of studies of DBPs and reproductive outcomes more useful for regulatory purposes and policy formulation.
- Performing epidemiological investigations of DBPs and adverse reproductive outcomes in communities served by water facilities with characteristically low spatial variability of THM contaminants will minimise the bias due to spatial variability and will help to elucidate the health effects these contaminants.

national database with samples collected under a uniform protocol.

The relative time and cost efficiency make it probable that public water facility monitoring data will continue to be used in future studies, despite the potential for exposure misclassification due to spatial variability and other sources of error. The purpose of this study was to explore methods that can be used to reduce exposure misclassification in epidemiological studies of THMs through water facility/community site selection. We evaluated two methods for selecting water facility distribution systems with spatially consistent THM levels from the USEPA ICR database. Since recent toxicological[10–12] and epidemiological[13–15] evidence suggests that exposure to brominated THMs may be important predictors of risk for adverse reproductive outcomes, we also evaluated the usefulness of these methods in selecting sites with a high proportion of brominated compounds.

METHODS

Under the guidelines of the ICR, water samples were collected from the distribution systems of 198 water facilities. Samples were collected at the following locations: (1) two locations of the "average" residence time of water in the distribution system (AVG1 and AVG2); (2) the point farthest along in the distribution system or at a dead-end representing the maximum residence time of water in the distribution system (MAX); and (3) at a location with a known retention time within the distribution system (termed the distribution system equivalent, DSE). Samples for each quarterly "sampling event" were collected during a single 24 hour window.[9]

We limited our study to facilities in the ICR database that had collected data for total THMs (TTHMs), the sum of individual THM species (chloroform ($CHCl_3$), bromodichloromethane ($CHBrCl_2$), dibromochloromethane ($CHBr_2Cl$), and bromoform ($CHBr_3$)), over four consecutive quarters (all seasons) and at four different sampling points in the distribution system from November 1997 to October 1998. Eighty four of the 198 facilities in the ICR database met this requirement. We extracted data for total (TTHMs) and individual THM species, and the sum of five haloacetic acids (HAA5) (monobromoacetic, monochloroacetic, dibromoacetic, dichloroacetic, and trichloroacetic acids) for each of these facilities.

We selected facilities exhibiting low spatial variability in each of the four quarters using two methods: (1) a method based on two way analysis of variance (ANOVA); and (2) a method based on concentration cut-points for TTHM derived from prior epidemiological studies of birth outcomes. We compared these methods with respect to factors that limited their validity and number of sites found eligible.

Two way ANOVA

We used a two way ANOVA procedure to compare TTHM concentrations between the four ICR sampling points in the presence of an extraneous variable, quarter, which reflected season. This allowed us to evaluate spatial variability between sampling points independently from temporal variability between quarters. Facilities with low intra-system spatial variability were defined as those where spatial variability did not depend on season according to Tukey's test for non-additivity, and where there was no evidence of a statistically significant difference ($p \geqslant 0.05$) between concentrations of TTHM by sampling point.

Cut-points

Epidemiological studies investigating the effects of TTHMs have suggested increased risks of low birth weight, small for gestational age, spontaneous abortion, and selected birth defects at specific estimated exposure levels of TTHMs.[13 14 16–19] In general, positive associations have been described at TTHM levels greater than 80 µg/l, whereas TTHM levels below 40 µg/l do not appear not to be associated with adverse effects. Facilities exhibiting consistently low intra-system spatial variability through four quarters were defined as those where, in a given quarter, all four sample values fell within one of three exposure groups (low, <40 µg/l; medium, 40–80 µg/l; or high, >80 µg/l).

Validation

To ensure that sites selected as having low intra-system spatial variability would also be selected in subsequent years, we performed validation analyses on results from both ANOVA and cut-point methods by substituting TTHM concentrations from the fall 1997 and summer 1998 seasons with fall measurements from 1998 and summer measurements from 1997. The additional data were available for 52 (62%) of the 84 sites that had data for six quarters. The validation was conducted by calculating the sensitivity and specificity of our spatial variability classification methods. We cross-tabulated the number of sites identified by classification as low or high spatial variability for each method according to the original and repeat analyses. The results from the original analyses were used as the reference or "gold standard" data.

Concentration profiles and brominated THMs

We classified those facilities in the USEPA ICR database selected as having consistently low spatial variability by either the ANOVA or cut-point method into one of three

Table 1 Total trihalomethane concentrations (µg/l) by quarter and sampling point for 84 sites reporting for four quarters; Information Collection Rule database 1997–1998

ICR time period	Sampling point	Mean	SD	Median	Range
Quarter 1 (Oct–Dec 97)	DSE	39.20	22.33	33.30	ND–108.9
	AVG1	41.24	23.44	35.95	ND–130.8
	AVG2	42.41	23.15	38.10	ND–104.4
	MAX	48.76	25.35	42.35	ND–129.0
Quarter 2 (Jan–Mar 98)	DSE	30.91	16.59	28.35	3.8–74.8
	AVG1	32.82	18.76	28.30	5.9–96.1
	AVG2	32.91	16.26	31.30	ND–80.5
	MAX	38.71	21.09	34.20	3.6–108.5
Quarter 3 (Apr–Jun 98)	DSE	45.17	27.37	39.15	1.1–136.0
	AVG1	47.56	26.59	39.80	7.2–133.1
	AVG2	48.10	28.34	42.05	3.2–136.0
	MAX	54.68	28.68	48.60	4.3–136.4
Quarter 4 (Jul–Sept 98)	DSE	56.19	35.31	47.35	ND–193.0
	AVG1	58.15	33.65	50.40	1.2–188.0
	AVG2	61.12	38.31	53.00	ND–222.0
	MAX	67.75	35.99	58.90	ND–204.7

exposure profiles: (1) high TTHM ($\geqslant$80 µg/l), high proportion of brominated THMs; (2) high TTHM ($\geqslant$80 µg/l), low proportion of brominated THMs; and (3) low overall TTHMs (<40 µg/l). A facility had to meet one of three criteria to be classified as having predominantly brominated THMs: (i) the mean concentration of brominated THM species for all sampling points exceeded 50 µg/l across all four quarters; (ii) the proportion of brominated species exceeded 50% at all sampling points during all four quarters; or (iii) the average bromine incorporation factor (BIF) across quarters was greater than or equal to 1.5. The BIF describes the molar contribution of all brominated species and is equal to:

$$\frac{0 \times [CHCl_3] + 1 \times [CHBrCl_3] + 2 \times [CHBr_2Cl] + 3 \times [CHBr_3]}{[CHCl_3] + [CHBrCl_3] + [CHBr_2Cl] + [CHBr_3]}$$

where in the numerator, the stoichiometric number of bromine moles in each THM species is multiplied by their respective concentration and the denominator is the concentration of TTHMs.[20] Our cut-point of 1.5 is the midpoint of the possible range in BIFs, 0 to 3.

Haloacetic acids
The HAAs are a class of non-volatile by-products with demonstrated reproductive effects in animals at higher doses;[1] however, with few exceptions,[18 21] measurements of haloacetic acids have not been incorporated into epidemiological studies to date. In order to determine whether TTHM concentrations could predict HAA concentrations at a water facility, we evaluated the relation between concentrations of HAA5 and TTHM for the four sampling points using Spearman correlation coefficients for all sites identified by our methods as having consistently low spatial variability. We performed the same analyses to determine whether brominated THM concentrations could be used to predict brominated HAA (monobromoacetic and dibromoacetic acid) concentrations.

RESULTS
Descriptive statistics for TTHMs for the 84 US sites with complete reporting for four consecutive quarters are shown in table 1. Stratification by quarter revealed the expected seasonal fluctuation in TTHM concentrations with highest levels from July to September and lowest levels from January

to March. The means at the four sampling points in the distribution system in each quarter suggest that TTHM concentrations increased with water residence time in the distribution system, with the lowest values seen at the earliest (DSE) and the highest values seen at the latest (MAX) sample points.

Two way ANOVA
Twenty five sites were found to have low spatial variability based on the ANOVA. Eight of these sites, however, were excluded after Tukey's test for non-additivity suggested that spatial variability depended on season and was not consistently low throughout the year.

Cut-points
Based on the cut-point method, we found 20 sites that had low spatial variability in each season. Twelve of the sites had TTHMs in the same exposure group (low, <40 µg/l; medium, 40–80 µg/l; or high, >80 µg/l) in any given season. Of the 20 sites, only four sites were also selected by the ANOVA method.

Validation
Twelve of 17 facilities characterised as having consistently low spatial variability according to the ANOVA method had six quarters of data. When TTHM concentrations from the summer and fall quarters were replaced, 10 (sensitivity = (10/12) = 83%) were again found to exhibit low spatial variability using ANOVA. Of the sites originally excluded by the ANOVA method, 40 had data for six quarters. After replacing quarters, 28 (specificity = (28/40) = 70%) of these were found to have high spatial variability and 12 were considered to have low spatial variability.

We also reevaluated facilities selected by the cut-point method that had six quarters of data in the ICR database. In this case, 10 of the 12 facilities (sensitivity = (10/12) = 83%) retained the classification of having consistently low spatial variability. Of the two facilities that were excluded, one had consistent low spatial variability of TTHM concentrations for three quarters, but a single abnormally high value for the MAX sample site led to high spatial variability in the last quarter. The other site was excluded for having a range (26–55 µg/l) including the lower cut-point of TTHM concentrations across sample locations for the summer quarter. All sites (n = 40) originally excluded due to spatial variability by

Table 2 Total and individual THMs and sum of 5 HAAs (µg/l) for plants categorised into each of three profiles by the ANOVA and cut-points methods

Method	DBP	Profile 1: High TTHM*, high brominated (n = 1)		Profile 2: High TTHM, low brominated (n = 3)		Profile 3: Low TTHM, low brominated (n = 6)	
		Mean	SD	Mean	SD	Mean	SD
ANOVA (n = 10)	TTHM	115.7	–	82.3	1.21	30.2	8.5
	$CHCl_3$	64.5	–	60.5	14.3	20.5	9.1
	$CHBrCl_2$	28.5	–	16.1	8.5	6.7	2.5
	$CHBr_2Cl$	17.9	–	8.4	0.2	3.3	3.1
	$CHBr_3$	4.8	–	0.3	0.1	0.9†	0.5
	HAA5	104.0	–	46.0	18.9	21.5	17.7

Method	DBP	Profile 1: High TTHM*, high brominated (n = 0)		Profile 2: High TTHM, low brominated (n = 3)		Profile 3: Low TTHM, low brominated (n = 13)	
		Mean	SD	Mean	SD	Mean	SD
Cut-points (n = 16)	TTHM			82.1	2.3	21.2	8.6
	$CHCl_3$			62.1	8.3	14.6	8.8
	$CHBrCl_2$			15.9	4.3	4.4	2.6
	$CHBr_2Cl$			4.0	3.7	2.6†	2.4
	$CHBr_3$			0.2	0.2	0.5†	1.0
	HAA5			71.6	16.8	16.7	12.5

*High TTHM includes all values ⩾80 µg/l.
†Value does not include non-detect results.

the cut-point method were again excluded (specificity = (40/40) = 100%) in the validation analyses.

Concentration profiles and brominated THMs
Table 2 describes the facilities with low spatial variability by exposure profile. Of the 17 water facilities selected by the ANOVA method, one had high and predominantly brominated TTHMs, three facilities had high TTHM levels with a low proportion of brominated compounds, and six had low overall TTHM levels. Seven facilities selected by the ANOVA had medium TTHM concentrations (40 µg/l ⩽ mean TTHM < 80 µg/l) and were not classifiable into any profile.

Among the 20 facilities that met the criteria for low spatial variability using the cut-point method, there were no water facilities with high TTHM concentrations that were predominantly brominated (profile 1). Three facilities met the definition of low proportions of brominated THMs with high mean TTHM concentrations (profile 2). Thirteen facilities with low TTHM exposure fit into profile 3 with a mean concentration of 21.2 µg/l TTHMs; five of these were classified as predominantly brominated sites. Four facilities selected by the cut-point method had medium mean TTHM concentrations (40 µg/l ⩽ mean TTHM < 80 µg/l) and were not classifiable into any profile.

In table 3, the mean concentration of the three brominated species, proportion of brominated compounds, BIF, and exposure profile are presented for the nine sites (one site was chosen by both methods) characterised as having low spatial variability and a predominantly brominated THM mixture. The exposure profile for these nine sites was obtained from their categorisation in table 2. The summary statistics suggested that the most useful predictor of a highly brominated system was the proportion of brominated compounds; the other criteria appeared to be highly variable.

HAAs
A moderate correlation (r = 0.667, p < 0.001) was found between TTHM values and HAA values for sampling events from facilities characterised as having consistently low spatial variability. The correlation was similar between mean concentrations of the brominated THM species and mean concentrations of the brominated HAA species (r = 0.642, p < 0.001).

DISCUSSION
We investigated two approaches to identify low spatial variability of DBPs in water distribution systems: the first based on a statistical ANOVA, and the second based on concentration cut-points for TTHMs derived from prior epidemiological studies of birth outcomes. Both methods appear to be useful for community selection. Identification of communities with low spatial variability will reduce exposure misclassification and make the results of studies of DBPs and reproductive outcomes more useful for regulatory purposes and policy formation.

In most epidemiological studies conducted to date, exposure assessment has been based on water facility data without accounting for spatial variability within the distribution system.[14–17 19 22 23] A few investigators have attempted to account for intra-system spatial variability. Klotz and Pyrch and Dodds et al used residential tap water sampling to validate exposure for each subject.[18 24] However, performing comprehensive water sampling for multiple DBP species is time and cost intensive and is difficult to accomplish in large studies. In addition, misclassification can occur when the samples are taken after the birth has occurred or a participant has moved. Gallagher et al incorporated a hydraulic model of the distribution system into a geographic information system, and assigned exposures to individual census block groups.[25] While this method shows promise for prediction of DBP concentrations throughout the system, extensive validation of the model for each class of by-product is necessary. Recently, Waller et al used a weighting procedure to reduce the influence on risk estimates of individuals with less accurate exposure values.[26] Weightings were based on degree of spatial variability in individual distribution systems and were assigned to each subject by residential location. This method may be appropriate for adjustment of biased exposure estimates when systems with high spatial variability must be utilised; however, as recently reported in a paper by Wright and Bateson, the applicability of this method may depend on degree of spatial variability and magnitude of "true" risk estimated for a given birth outcome.[4]

We evaluated two methods for selection of populations served by water distribution systems with minimal spatial variability in the concentrations of TTHMs. Using the two way ANOVA method, we compared TTHM levels between four sampling points in the distribution system controlling

Table 3 Total trihalomethane concentration, mean concentration of brominated trihalomethanes*, proportion brominated species, bromine incorporation factor (BIF), and exposure profile number for nine predominantly brominated sites selected using ANOVA and cut-point methods

Method	Facility	Mean TTHM (µg/l)	Mean brominated THMs (µg/l)	% brominated species	BIF	Exposure profile
ANOVA	126	118.4	51.3	54.0	1.54	1
	567†	43.7	23.0	52.9	1.50	–
	163	32.2	19.0	60.1	1.51	3
	413	19.4	9.6	51.3	1.33	3
Cut-points	132†	51.8	36.6	62.7	1.50	–
	535	24.1	18.9	78.7	1.77	3
	415	19.9	9.8	51.0	1.34	3
	355	19.4	10.0	52.5	1.36	3
	413	19.4	9.6	51.3	1.33	3
	416	19.0	9.2	50.0	1.32	3
Summary‡ statistics	Mean	34.9	17.8	55.7	1.44	
	SD	32.2	13.6	9.1	0.15	
	Median	20.2	10.0	52.5	1.36	

*Bromodichloromethane, dibromochloromethane, bromoform.
†Site had medium average TTHM concentrations (40 µg/l ⩽ mean TTHM < 80 µg/l) and was not classifiable into any profile.
‡Site 413 was counted only once in the summary statistics.

for the influence of season and identified 17 sites with low spatial variability in four consecutive seasons. A limitation of this method was that an arbitrary alpha level for site inclusion was chosen ($\alpha = 0.05$). Further, the power of the F-test was potentially differential for each facility, depending on the relative difference between sampling point means. Lastly, applying an inherently statistical approach does not incorporate biological plausibility. To illustrate, consider a hypothetical scenario in which the four TTHM levels measured during one quarter in the distribution system of facility A are 150, 170, 180, and 200 (variance = 433). In the distribution system of facility B, levels 70, 70, 90, and 90 were measured (variance = 133). Although the variance in facility B is much lower, facility A is a better choice for a high exposure community since all four values far exceed 80 µg/l, indicating high exposure for all exposed individuals.

The method of site selection based on exposure cut-points identified 20 eligible facilities. The primary limitation of the cut-point method is reliance on previous epidemiological findings; cut-points may be subject to change as newer studies become available. In addition, several sites were not included because degree of spatial variability of TTHMs in a distribution system depended on season. This limitation also applied to sites chosen by ANOVA. These methods limit the number of sites available for inclusion in studies.

There was little overlap between the methods used to identify low spatial variability; only four water facilities were selected by both methods. Three of these sites exhibited low spatial variability by both methods of selection in the validation study, making them excellent candidates for an epidemiological investigation. The lack of concurrence in site identification was not surprising, given the differences between the statistical (ANOVA) and epidemiological approaches used. In the validation analysis, we found the sensitivity of each site selection method to be identical (83%). The ability to consistently exclude sites for not having low spatial variability (specificity), however, was dependent on the selection method. During validation, the ANOVA method only excluded 70% of the facilities originally judged ineligible, while the cut-point method excluded 100%. Several reasons for this disparity may exist, including the possibility that the ANOVA method did not adequately control for temporal variability in THM formation by using quarter of sampling as a marker for seasonal changes. To the extent that this data set allowed, the validation analyses suggested that

the cut-point method may be more helpful in identifying sites with consistent low spatial variability from available state or national databases.

The cut-point method can be used to select water facilities having low spatial variability for a variety of study objectives. Of the 20 facilities identified by this method, 12 were classified into the same exposure category in each of the four consecutive quarters and would be appropriate for inter-community studies in which disease rates in communities with higher TTHM were compared with those in communities with lower TTHM levels. The exposure group classification (high, medium, low) of the eight other facilities with consistently low spatial variability changed over time. These facilities may be most suitable for studies employing an intra-community approach in which the effects of an exceedance above a given threshold concentration are evaluated. This approach is well suited to the study of birth defects since the critical period of development is likely to be known and one can investigate the effects of high exposures during that window of gestation within a single community.[27]

Most previous epidemiological studies of reproductive outcomes have relied on TTHMs as the relevant exposure metric, while recent studies suggest that the composition of the mixture and the concentration of specific DBPs, especially brominated DBPs, may be of critical importance.[13–15 18 22] Approximately 60% of sites selected by ANOVA (10/17) and 80% of sites selected by the cut-point (16/20) method fell into one of the exposure profiles that was selected for future studies. The sites selected fell mainly into profiles of high TTHM concentrations with minimal proportions of brominated species (profile 2) and low overall TTHM concentrations (profile 3). Only one site, selected by ANOVA, had a high proportion of brominated THMs and TTHM levels that met the criteria for profile 1. The paucity of sites identified for this exposure profile is likely due to the relative rarity of the description, rather than an inherent limitation of the classification method. The analysis was designed to provide a comparison between tools available to identify a site with predominantly brominated DBPs. The summary statistics suggested that the most useful predictor of a highly brominated facility was the proportion of brominated compounds since the other criteria appeared to be highly variable.

Toxicological data[1] suggest that HAAs may be important contributors to risk for adverse reproductive outcomes, but with two exceptions,[18 21] exposure to HAAs has not been

incorporated into epidemiological studies. This is partly due to the deficiency of HAA data in the USA prior to 2002. Using data from the ICR for sites identified as having low spatial variability of TTHM concentrations, we found a moderate correlation with TTHM and HAA5. These inter-species correlations were generally consistent with the relation between TTHMs and the sum of five HAAs (r = 0.815) described by Villanueva and colleagues.[28] In a recent study, King *et al* found correlation coefficients of 0.74 and 0.52 between TTHM and HAA values for tap water samples taken in two regions of Canada.[29] These findings indicate that the concentrations of TTHMs are, at best, moderately associated with HAAs and suggest that researchers should not assume that selection of sites with low spatial variability in TTHMs will produce corresponding low spatial variability exposures to HAAs.

Improvements in exposure assessment for DBPs have been difficult to implement.[3] New methods of exposure assessment such as distribution system modelling, and use of biomarkers in blood, urine or exhaled breath have yet to be validated or applied in large epidemiological studies. Recent studies in the USA have relied on quarterly sampling data collected at multiple sampling sites[30] to meet regulatory requirements. Future studies will likely continue to use these data, due to their relative efficiency. The methods in this paper permit selection of sites with limited spatial variability from state or national DBP datasets to be used for future prospective and retrospective epidemiology studies. Ultimately, simple techniques such as these may improve both the quality of data and our understanding of the true risks associated with exposure to DBPs.

ACKNOWLEDGEMENTS

. . . .

Authors' affiliations
A F Hinckley, A M Bachand, J R Nuckols, J S Reif, Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, Colorado, USA

Competing interests: none declared

REFERENCES

1 **Nieuwenhuijsen MJ**, Toledano MB, Eaton NE, *et al.* Chlorination disinfection byproducts in water and their association with adverse reproductive outcomes: a review. *Occup Environ Med* 2000;**57**:73–85.
2 **Bove F**, Shim Y, Zeitz P. Drinking water contaminants and adverse pregnancy outcomes: a review. *Environ Health Perspect* 2002;**110**(suppl 1):61–74.
3 **Reif JS**, Hatch MC, Bracken M, *et al.* Reproductive and developmental effects of disinfection byproducts in drinking water. *Environ Health Perspect* 1996;**104**:1056–61.
4 **Wright JM**, Bateson TF. A sensitivity analysis of bias in relative risk estimates due to disinfection by-product exposure misclassification. *J Expo Anal Environ Epidemiol*, 2004 (published online 30 June 2004).
5 **Chen WJ**, Weisel CP. Concentration changes of halogenated disinfection by-products within a distribution system. *J Am Water Works Assoc* 1998;**90**:151–63.
6 **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.
7 **Rodriguez MJ**, Vinette Y, Sérodes JB, *et al.* Trihalomethanes in drinking water of greater Québec region (Canada): occurrence, variations and modeling. *Environ Monit Assess* 2003;**89**:69–93.
8 **Singer P**, Pyne RDG, Avs M, *et al.* Examining the impact of aquifer storage and recovery on DBPs. *J Am Water Works Assoc* 1993;**85**(11):85–94.
9 **US EPA**. ICR sampling manual, EPA 814-B-96-001. Washington, DC: United States Environmental Protection Agency, 1996.
10 **Narotsky MG**, Pegram RA, Kavlock RJ. Effect of dosing vehicle on the developmental toxicity of bromodichloromethane and carbon tetrachloride in rats. *Fundament Appl Toxicol* 1997;**40**:30–6.
11 **Bielmeier SR**, Best DS, Guidici DL, *et al.* Pregnancy loss in the rat caused by bromodichloromethane. *Toxicol Sci* 2001;**59**:309–15.
12 **Landi S**, Hanley NM, Warren SH, *et al.* Induction of genetic damage in human lymphocytes and mutations in Salmonella by trihalomethanes: role of red blood cells and GSTT1-1 polymorphism. *Mutagenesis* 1999;**14**:479–82.
13 **Waller K**, Swan SH, DeLorenze G, *et al.* Trihalomethanes in drinking water and spontaneous abortion. *Epidemiology* 1998;**9**:134–40.
14 **King W**, Dodds L, Allen A. Relation between stillbirth and specific chlorination by-products in public water supplies. *Environ Health Perspect* 2000;**108**:883–6.
15 **Dodds L**, King W. Relation between trihalomethane compounds and birth defects. *Occup Environ Med* 2001;**58**:233–7.
16 **Bove FJ**, Fulcomer MC, Klotz JB, *et al.* Public water contamination and birth outcomes. *Am J Epidemiol* 1995;**141**:850–62.
17 **Dodds L**, King W, Wolcott C, *et al.* Trihalomethanes in public water supplies and adverse birth outcomes. *Epidemiology* 1999;**10**:443–6.
18 **Klotz JB**, Pyrch LA. Neural tube defects and drinking water disinfection by-products. *Epidemiology* 1999;**10**:383–90.
19 **Wright JM**, Schwartz J, Dockery DW. Effect of trihalomethane exposure on fetal development. *Occup Environ Med* 2003;**60**:173–80.
20 **Miles AM**, Singer PC, Ashley DL, *et al.* Comparison of trihalomethanes in tap water and blood. *Environ Sci Technol* 2002;**36**:1692–8.
21 **Wright JM**, Schwartz J, Dockery DW. The effect of disinfection by-products and mutagenic activity on birth weight and gestational duration. *Environ Health Perspect* 2004;**112**:920–4.
22 **Kramer MD**, Lynch CF, Isacson P, *et al.* The association of waterborne chloroform with intrauterine growth retardation. *Epidemiology* 1992;**5**:407–13.
23 **Savitz DA**, Andrews KW, Pastore LM. Drinking water and pregnancy outcome in central North Carolina: source amount and trihalomethane levels. *Environ Health Perspect* 1995;**103**:592–6.
24 **Dodds L**, King WD, Allen AC, *et al.* Trihalomethanes in public water supplies and risk of stillbirth. *Epidemiology* 2004;**15**:179–86.
25 **Gallagher MD**, Nuckols JR, Stallones L, *et al.* Exposure to trihalomethanes and adverse pregnancy outcomes. *Epidemiology* 1998;**9**:484–9.
26 **Waller K**, Swan SH, Windham GC, *et al.* Influence of exposure assessment methods on risk estimates in an epidemiologic study of total trihalomethane exposure and spontaneous abortion. *J Expo Anal Environ Epidemiol* 2001;**11**:522–31.
27 **Hertz-Picciotto I**. Environmental epidemiology. In: Rothman KJ, Greenland S, eds. *Modern epidemiology*, 2nd edn. Philadelphia: Lippincott, Williams and Wilkins, 1998.
28 **Villanueva CM**, Kogevinas M, Grimalt JO. Haloacetic acids and trihalomethanes in finished drinking waters from heterogeneous sources. *Water Research* 2003;**37**:953–8.
29 **King WD**, Dodds L, Armson BA, *et al.* Exposure assessment in epidemiologic studies of adverse pregnancy outcomes and disinfection by-products. *J Expo Anal Environ Epidemiol* 2004;**14**:466–72.
30 **US EPA**. National primary drinking water regulations: disinfection by-products. *Federal Register* 1998;**63**(241):69390.