JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

A Survey Sampling Approach for Pesticide Monitoring of Community Water Systems Using Groundwater as a Drinking Water Source

Roy W. Whitmore[†] and Wenlin Chen*[,‡]

[†]RTI International, P.O. Box 12194, Research Triangle Park, North Carolina 27709-2194, United States

[‡]Syngenta Crop Protection, LLC, P.O. Box 18300, Greensboro, North Carolina 27419-8300, United States

ABSTRACT: The ability to infer human exposure to substances from drinking water using monitoring data helps determine and/or refine potential risks associated with drinking water consumption. We describe a survey sampling approach and its application to an atrazine groundwater monitoring study to adequately characterize upper exposure centiles and associated confidence intervals with predetermined precision. Study design and data analysis included sampling frame definition, sample stratification, sample size determination, allocation to strata, analysis weights, and weighted population estimates. Sampling frame encompassed 15 840 groundwater community water systems (CWS) in 21 states throughout the U. S. Median, and 95th percentile atrazine concentrations were 0.0022 and 0.024 ppb, respectively, for all CWS. Statistical estimates agreed with historical monitoring results, suggesting that the study design was adequate and robust. This methodology makes no assumptions regarding the occurrence distribution (e.g., lognormality); thus analyses based on the design-induced distribution provide the most robust basis for making inferences from the sample to target population.

KEYWORDS: *water quality, sampling design, drinking water exposure, atrazine*

■ INTRODUCTION

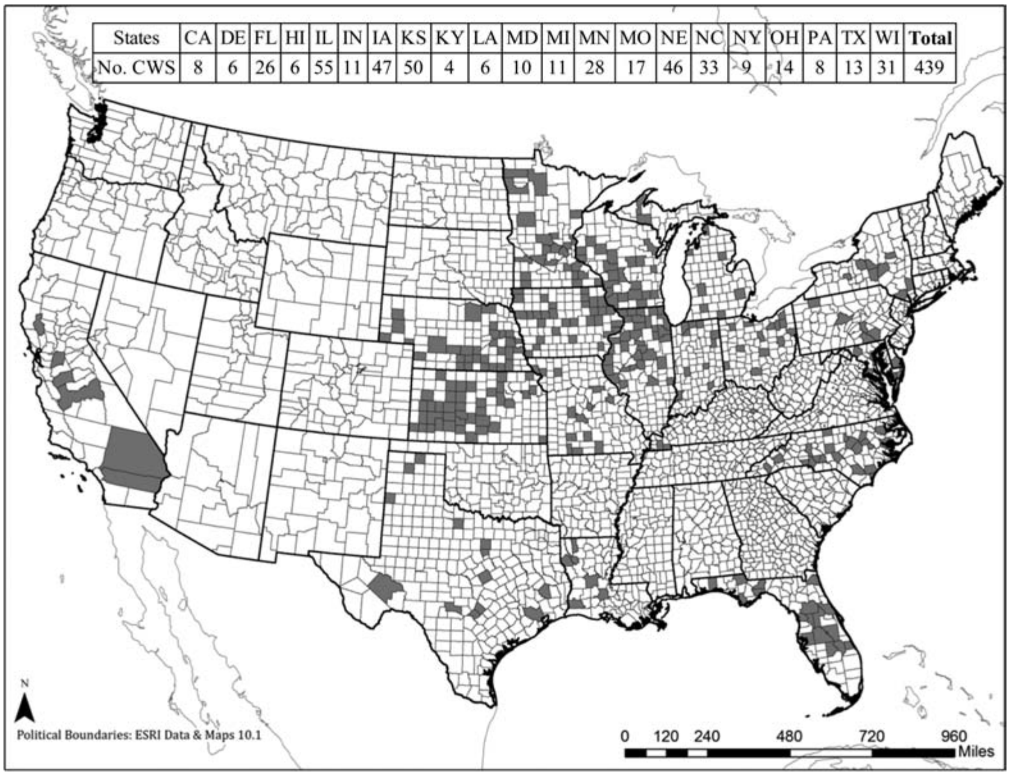
Regulatory assessments for potential drinking water exposure to crop protection chemicals are primarily conducted using computer models with a set of combined worst-case input assumptions to generate high-end (protective) estimates of exposure.[1,2] The process follows a tiered scheme of increasing complexity, beginning with a simple and lower tier model, and followed by various higher tiers with inputs reflecting agronomic, soil, and hydroclimatological conditions of vulnerable geographical locations.[2,3] Pursuit of refined and realistic exposure estimates using monitoring data to reduce model estimate uncertainty in pesticide regulation has been a continuous endeavor.[4−9] Recently, the Organization for Economic Co-operation and Development (OECD) published the guidance document for exposure assessment using environmental monitoring,[10] calling for consistency between monitoring data-based and model-based exposure estimation.

Conceptually, field monitoring reflects "real-world" conditions and should provide the most accurate exposure information for risk assessment. Yet, monitoring data are often less used in the regulatory assessment process due to questions about how to place in context and interpret measured exposure especially when there is lack of a scientifically sound monitoring design associated with data generation. A robust monitoring design requires a well-defined statistical population and a sampling plan that supports population inference requirements. It involves the use of a probability sampling method so that measured data can be used to provide design-based inference to address specific exposure questions for the target population of monitoring such as community water systems (CWS) that deliver drinking water derived entirely from groundwater sources.

Water quality at CWS is regulated by the U.S. Environmental Agency (EPA) under the Safe Drinking Water Act (SDWA). There are 40 000+ groundwater CWS in the United States, accounting for 78% of all CWS. Each CWS supplies drinking water to at least 25 people or 15 service connections.[11] Occurrence of potential pesticide residues in groundwater used by these CWS is expected to vary geographically due to pesticide use patterns (application timing, rate, and method), agronomic practices, and combinations of local weather, soil, and other hydrological characteristics. Monitoring to determine exposure on each CWS nationally is operationally challenging and costly. A robust sampling approach for water quality monitoring design is needed to accurately estimate exposure in an economically feasible manner. In this paper, we propose and demonstrate the utility of a survey sampling design with a case application to a groundwater CWS monitoring study for atrazine (6-chloro-*N*-ethyl-*N*′-(1-methylethyl)-1,3,5-triazine-2,4-diamine) across 21 states. Atrazine has been widely used over the last 50 years for weed control in corn, sorghum, and sugar cane production.

Our specific objectives were to present: (1) a stratified random sampling plan to support robust estimates of the distribution of atrazine concentrations across the population of CWS using groundwater as their source water; (2) a statistical method to characterize the measured exposure distribution and compare with historical SDWA data; and (3) recommendations of factors needed in statistical survey planning and design for monitoring raw and finished water samples collected from CWS.

Received: September 16, 2013
Revised: November 8, 2013
Accepted: November 12, 2013
Published: November 12, 2013



States	CA	DE	FL	HI	IL	IN	IA	KS	KY	LA	MD	MI	MN	MO	NE	NC	NY	OH	PA	TX	WI	Total
No. CWS	8	6	26	6	55	11	47	50	4	6	10	11	28	17	46	33	9	14	8	13	31	439

Figure 1. County level locations in the contiguous United States and number of community water systems (CWS) participated in the atrazine groundwater monitoring case study.

■ MATERIALS AND METHODS

Planning a monitoring survey begins with a clear definition of the target population to be sampled and the survey's objectives. Once defined, sample stratification, sample size and allocation, and analysis weights can be developed based on the monitoring objectives and prior knowledge of the target population (i.e., groundwater CWS), such as the geography of pesticide use and existing relevant monitoring data or relevant exposure modeling. In the following sections, an atrazine groundwater CWS monitoring program conducted in 2000 was used as a case study to illustrate design steps and statistical characterization of results.

Monitoring Goal and Data for Sampling Frame. For the atrazine case monitoring study, the goal was to collect sufficient finished drinking water samples so that the 95th percentile of the concentration distribution could be statistically characterized for the population of groundwater CWS in the major atrazine use regions. Major use geography was determined by county level use data over four years (1995−1998) from Doane's market research survey, and comprised counties in 21 states or 90% of the total atrazine used in the U.S. (Figure 1): California, Delaware, Florida, Hawaii, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maryland, Michigan, Minnesota, Missouri, Nebraska, North Carolina, New York, Ohio, Pennsylvania, Texas, and Wisconsin.

A comprehensive database was compiled from the state CWS SDWA-compliance monitoring programs (1993−1998) to establish the atrazine groundwater sampling frame. The SDWA requires CWS to monitor for certain constituents including atrazine in finished drinking water samples (i.e., after water plant treatment). Water samples were typically taken quarterly in each year. The compiled database is referred to as Population Linked Exposure (PLEX) which links atrazine measurements, sampling date, source water types (surface water, groundwater, mixed, etc.) and other auxiliary information to the number of persons served at each CWS. On the basis of these data, 15 840 CWS satisfied the following criteria: CWS uses groundwater as sole drinking water source; groundwater source is not influenced by surface water; groundwater is not purchased from another source; and the CWS serves at least 25 persons as defined by the U.S. EPA.[11] These 15 840 CWS therefore became the sampling frame for the survey of CWS using groundwater as their sole source of drinking water.

Sample Stratification. An effective sampling design divides the target population into distinct groups called strata so that heterogeneity between strata can be the greatest while group members within each stratum remain relatively homogeneous. Such stratification improves sample representativeness and precision of population quantities of interest to the extent that outcomes (e.g., atrazine concentrations) are more homogeneous within strata. For the 15 840 groundwater CWS, two subpopulations (or domains) were first stratified based on historically measured atrazine concentrations in finished water: (1) nondetect domain (15 381 CWS) where atrazine was never detected (1993−1998); and (2) detect domain (459 CWS) where at least one sample with a detectable atrazine concentration (1993−1998) was measured.

For clarity, we referred to detect and nondetect strata as "domains", while substrata within these domains are referred to as "strata" (Figure 2). Stratification variables were investigated within each domain by evaluating the effectiveness of the resulting alternative stratification schemes. For this purpose, the 1993−1997 data were used to form strata and compared with percentages of atrazine measurements detected in 1998 in these strata. The evaluation was based on 256 CWS with at least one detectable atrazine concentration measurement (1993−1998) and had data (either detect or nondetect) for both 1998 and at least 1 year in 1993−1997. Within this context, the following metrics were examined to determine their utility for stratifying the sample:

(1) Number of persons served by each CWS.

(2) Average atrazine use intensity (1995−1997). This parameter was computed by dividing the estimated total mass of atrazine applied in a county in a given year by the total county area and averaging these rates over years for which estimated product use data were available. Total county area was used because the impact on the county's groundwater was expected to be related to the total amount applied

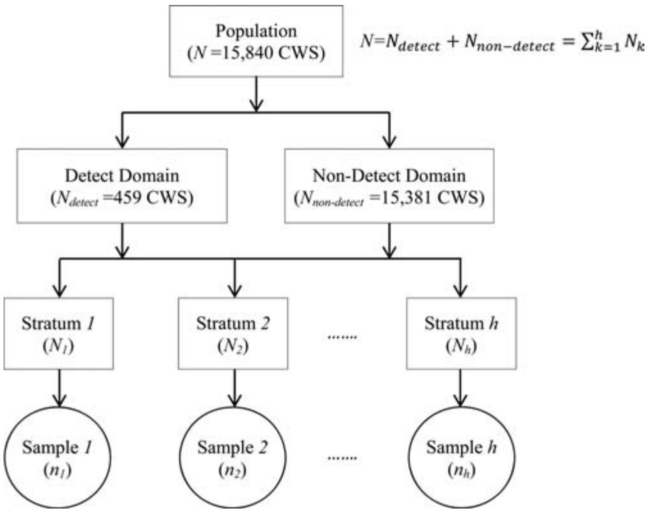


Figure 2. Stratification for the atrazine monitoring case study of community water systems (CWS) using groundwater as their sole drinking water source.

and the total groundwater volume, recharge and flow characteristics, rather than the specific crop acreage to which the atrazine was applied.

(3) Average quarterly maxima of atrazine concentrations (1993–1997). For each CWS in the SDWA database, the maximum concentration of all measurements in each quarter was determined. Then, the quarterly maxima were averaged. Using the maximum for each quarter guarded against averaging downward the data for a quarter in which there was one relatively high value followed by several small, or nondetect.

(4) Median of the quarterly maxima of atrazine concentrations (1993–1997).

(5) Maximum atrazine concentrations detected (1993–1997).

There are other factors such as climate (e.g., precipitation), soil, and hydrogeological characteristics which may influence groundwater quality on a local and/or regional scale and can be potentially used as stratification variables for monitoring design. These factors however were not considered in the current case study design because of the availability of a large amount of historical SDWA groundwater monitoring data (1993–1998) which would reflect the long-term impact of the environmental conditions.

Limit of quantification (LOQ) for the state SDWA monitoring programs used to determine atrazine levels in CWS water samples varied from 0.1 to 3 ppb (μg L^{-1}) (1993–1998). The varying LOQ could impact statistical estimates of lower quantiles but was expected to have little effect on estimates of higher quantiles of the atrazine concentration distribution (e.g., the 95th percentile). In the case study of atrazine monitoring (LOQ = 0.05 ppb as described later), statistical estimates were made with two approaches, one using data of all instrument-measurable concentrations and the other using 1/2 LOQ substitution for nondetection. There are other statistical approaches to deal with nondetections,[12] but these were not attempted as the focus of this study was on the higher centile distribution estimates.

Sample Size and Allocation. Sample size for monitoring depends on the targeted percentile of the population distribution, desired confidence level, and estimate precision. Sample size recommendations have been provided based on three methods: tolerance interval, relative standard error (RSE), and skewness criterion.[13] Minimum effective sample sizes resulting from these approaches are provided in Table 1, and details can be found in Mosquin et al.[13] A brief summary is provided below.

The tolerance interval approach finds the smallest sample size, n, such that the probability will be at least $100(1 - \alpha)$% that a simple random sample of size n contains the 100qth centile, Q_q, of a continuous distribution.[14] For large q values (close to 1, such as the 95th percentile), the required minimum sample size can be estimated as

Table 1. Smallest Effective Sample Size n Required to Estimate Selected Percentiles in the Population Distribution

Tolerance Method				
	estimated percentile $(100q)^a$			
confidence level $100(1 - \alpha)$	90.0%	95.0%	99.0%	99.9%
80.0%	16	32	161	1,609
90.0%	22	45	230	2,302
95.0%	29	59	299	2,995
99.0%	44	90	459	4,603
99.9%	66	135	688	6,905
Relative Standard Error (RSE) Method				
	proportion of true population $(p)^b$			
upper bound RSE (R)	10.0%	5.0%	1.0%	0.1%
50%	37	77	397	3,997
40%	57	119	619	6,244
30%	101	212	1,101	11,101
20%	226	476	2,476	24,976
10%	901	1,901	9,901	99,901
Method Based on Binomial Distribution Skewness^c				
	estimated percentiles $100(1 - p)$			
skewness (γ)	90.0%	95.0%	99.0%	99.9%
<1/2	29	69	389	3,989

^{a}Tolerance method estimates the smallest effective sample size n for which the probability of containing the 100qth centile, Q_q, has confidence level $100(1 - \alpha)$. ^{b}The relative standard error (RSE) method estimates the smallest effective sample size n such that the RSE of the estimate of a population proportion p will be no more than a prespecified upper bound R. The qth percentile estimate, Q_q, equals $100(1 - p)$. ^{c}The method based on the skewness of the binomial distribution does not explicitly specify the confidence intervals. As per Boos and Hughes-Oliver[15] and Brown et al.,[16] two-sided confidence intervals are expected to have reasonable coverage for small skewness of the binomial distribution (i.e., $\gamma \leq 0.5$).

$$n \approx \log(\alpha)/\log(q) \tag{1}$$

The RSE method assumes a prespecified upper bound, R. For simple random sampling with RSE $\leq R$, the required minimum sample size is

$$n = \frac{1 - p}{R^2 p} \tag{2}$$

Generally as a rule of thumb, RSE $\leq$ 10% corresponds to accurate estimates; RSE $\approx$ 30% corresponds to moderately precise estimates; and RSE $\geq$ 50% corresponds to estimates with minimal precision.

The skewness criterion method is based on the skewness parameter γ of the binomial distribution:

$$\gamma = \frac{1 - 2p}{\sqrt{np(1 - p)}} \tag{3}$$

It has been shown that the two-sided confidence intervals obtained by the Woodruff method[13] are expected to have reasonable coverage provided that the binomial distribution associated with the centile does not have excessive skew (i.e., $\gamma \leq 0.5$) under simple random sampling.[15,16] Given the criterion $\gamma \leq 0.5$, effective sample sizes n can be estimated for a range of percentiles (Table 1).

The monitoring goal of the case study was to make inferences about the 95th percentile of the atrazine concentration distribution with moderate precision (RSE $\approx$ 30%) from finished drinking water samples in each of the two sampling domains. Thus, the design selected a target sample size of 212 observations for each domain based in Table 1 and allocated proportionately to the strata within each domain based on the equation:

$$n_h = \frac{N_{d,t}}{N_d} N_h \tag{4}$$

where n_h is the number of samples for stratum h of domain d (d = detect or nondetect); N_d is the total CWS number of domain d; $N_{d,t}$ is the target sample size for domain d; and N_h is the total CWS number in stratum h of the detect or nondetect domain. Note that either n_h or N_h were not indexed per domains in eq 4 as we assigned distinct stratum numbers for each of CWS domains in the computer analysis process.

To ensure participation of at least 212 CWS in the monitoring for each of the two domains, a larger target sample size of 314 CWS was suggested for initial planning, assuming that at least 75% of the CWS selected for the study would be eligible and agree to provide a water sample, and at least 90% of those would be successfully analyzed. The number of CWS for each stratum within each domain was allocated proportionally and selected randomly so that the total number of participants in each domain was close to the target sample size of 314 or at least 212 CWS. A comprehensive questionnaire was prepared to identify CWS drinking water well characteristics including well location, construction year, depth, capacity, casing, aquifer type (confined or unconfined), influenced by surface water, distance from crop field, past atrazine detection, etc. Each of the randomly selected CWS was sent the questionnaire to determine CWS eligibility for the study using the criteria described above. If a questionnaire response was received from a CWS and it was subsequently determined as qualified, the CWS was then asked to provide a finished water sample for laboratory analysis.

Statistical Analysis Weights. Given the same target number of samples, different sampling rates for each of the two CWS domains resulted from the difference in domain size. For example, of the total 15 840 CWS, nondetect CWS were predominant with the domain size consisting of 15 381 systems. Because of this, statistical analysis weights were needed to produce design-unbiased estimates of overall population parameters (e.g., numbers and proportions of wells with detectable concentrations of the target analyte). Because of differential rates of CWS eligibility and nonresponse (i.e., CWS who chose not to participate in monitoring), statistical analysis weights needed to be adjusted to reduce the potential for nonresponse bias. Initial sampling weights were developed as a part of sample design activities. After data collection, sampling weights were adjusted to compensate (at least partially) for the potential bias resulting from survey nonresponse. Procedures for weighting and weight adjustments are described below.

Sampling weights, or initial statistical analysis weights, w_1, are reciprocals of the probabilities of selection for CWS i to be sampled:

$$w_1(h, i) = N_h/n_h \tag{5}$$

where N_h and n_h are defined in eq 4.

There are two types of nonresponse of a CWS to monitoring: (1) a CWS may be sampled but its eligibility cannot be verified, and (2) a verified eligible CWS may not provide water samples for analysis for various reasons. For the first type, a weight adjustment factor is needed to compensate for the unknown eligibility status of the CWS to reduce bias. This is essentially two weight adjustments for nonresponse to determine the eligibility of a CWS. Since each sampling stratum had a sufficient number of respondents to be used as a weighting class (except for several Kansas CWS with false positives in PLEX as discussed later), two weight adjustments were made based on sampling strata instead of using separate weighting classes. Hence, for the first weight adjustment, a factor wf_1 to compensate for nonresponse to the determination of eligibility was computed as

$$wf_1(h, i) = \frac{\sum_{i=1}^{n_h} w_1(h, i)}{\sum_{i=1}^{n_h} I_k(h, i) w_1(h, i)} \tag{6}$$

where

$$I_k(h, i) = \begin{cases} 1, \text{ if the } i\text{th CWS had known eligibility} \\ 0, \text{ and otherwise} \end{cases} \tag{7}$$

The adjusted weight w_2 combining the initial weight w_1 and adjustment factor wf_1 for the ith CWS in stratum h was then computed as

$$w_2(h, i) = w_1(h, i) wf_1(h, i) I_E(h, i) \tag{8}$$

where

$$I_E(h, i) = \begin{cases} 1, \text{ if the } i\text{th CWS was eligible} \\ 0, \text{ otherwise} \end{cases} \tag{9}$$

From eq 8, the sum of the adjusted weights, w_2, estimates the number of $I_E(h,i)$ eligible CWS in the weighting class (i.e., stratum) because it is the estimated proportion of eligible CWS in the weighting class multiplied by the total number of CWS on the stratum sampling frame, i.e.,

$$\sum_{i \in c}^{n_h} w_2(h, i) = \frac{\sum_{i=1}^{n_h} I_E(h, i) w_1(h, i)}{\sum_{i=1}^{n_h} I_k(h, i) w_1(h, i)} \sum_{i=1}^{n_h} w_1(h, i) \tag{10}$$

For the second nonresponse type, a similar weight adjustment factor, wf_2, was computed to compensate for no water sampling of a selected known eligible CWS:

$$wf_2(h) = \frac{\sum_{i=1}^{n_h} w_2(h, i)}{\sum_{i=1}^{n_h} I_R(h, i) w_2(h, i)} \tag{11}$$

where

$$I_R(h, i) = \begin{cases} 1, \text{ if the } i\text{th CWS was a participant} \\ 0, \text{ otherwise} \end{cases} \tag{12}$$

The final statistical analysis weight, w_3, for the i^{th} CWS in stratum h was then computed as

$$w_3(h, i) = w_2(h, i) wf_2(h) I_R(h, i) \tag{13}$$

For each stratum (h), and hence, the population as a whole, the sum of the final weights, w_3, is the same as the sum of the unadjusted weights, w_2, even though the nonrespondents have a nonzero value of w_2 and a zero value for the final weight, w_3. In this way, the nonresponse adjustment reduces bias to the extent the non-respondents are similar to respondents from the same stratum, i.e.,

$$\sum_{i=1}^{n_h} w_3(h, i) = \sum_{i=1}^{n_h} w_2(h, i) \tag{14}$$

Data Analysis. Proper analysis of data collected for members of a probability sample requires that all observations be weighted inversely to their probabilities of selection as described above. Sampling weights enable design-unbiased estimation of linear population parameters, such as population totals. A common example requiring weighted data analysis is estimation of a population proportion (or percentage), such as the percent measurable (i.e., estimated proportion of population units having detectable concentrations). To estimate a proportion P_x, the general form of the estimate is

$$\widehat{P}_x = \frac{\sum w_3(h, i) X_i}{\sum w_3(h, i)} \tag{15}$$

where summations are over all sample units, $w_3(h,i)$ denotes the analysis weight associated with CWS i in stratum h, and X_i is an indicator variable with a value of 1 (or 100) if CWS i has the characteristic of interest (e.g., has a detectable level) and with a value of 0 otherwise. The numerator is an estimate of the total number of CWS units in the population having the characteristic; the denominator is an estimate of the total number of units in the population.

A similar expression is used to estimate target population mean, where Y_i denotes a measured quantity for CWS i (e.g., a concentration of atrazine):

$$\bar{Y} = \frac{\sum w_3(h, i)Y_i}{\sum w_3(h, i)} \quad (16)$$

The numerator estimates the total of the Y variable that would have been obtained if all CWS of the target population had been observed and measured; the denominator estimates total size of the target population.

We also wanted to estimate proportions and means for individual domains of the target CWS population, for example, percent measurable and average concentration for the detect domain. For a domain d, such a proportion is estimated as

$$\widehat{P}_x(d) = \frac{\sum w_3(h, i)I_d X_i}{\sum w_3(h, i)I_d} \quad (17)$$

where $I_d = 1$ if unit i is in the domain d, and $I_d = 0$ otherwise. Analogously, the domain means are estimated as

$$\bar{Y}(d) = \frac{\sum w_3(h, i)I_d Y_i}{\sum w_3(h, i)I_d} \quad (18)$$

Note that if the I_d values are identically 1 in eqs 17 and 18, then the domain of interest is the entire target population.

Population precision and domain parameter estimates (e.g., proportions, means, and percentiles) are expressed in terms of their variance or standard error. Estimating sampling variances and standard errors for statistics calculated from probability sampling data is based on the randomization distribution induced by the sampling design (i.e., they should account for all sampling design features, such as stratified random sampling). Such an approach is robust because it makes no assumptions regarding the distribution of occurrence (e.g., lognormality) of the survey items. Hence, analyses based on the design-induced distribution provide the most robust basis for making inferences from the sample to the target population.

The classic approach to estimating standard errors for nonlinear statistics, such as means and proportions, from complex probability sampling designs is a first-order Taylor Series linearization method. Alternative variance estimation techniques for such designs include jackknifing and balanced repeated replication. We employed the first-order Taylor Series method as implemented in the software Survey Data Analysis (SUDAAN) to estimate standard errors.[17] The SUDAAN DESCRIPT procedure was used to estimate proportions, means, geometric means, and percentiles. Approximate 95% confidence intervals for parameter estimates were calculated. For percentiles, confidence intervals were obtained directly by DESCRIPT. For means, intervals were obtained as [estimated mean] ± 2 [its estimated standard error] based on the assumption that estimates are approximately normally distributed.

Case Study. After the monitoring design (sample stratification and allocation) was completed and the CWS eligibility questionnaire and participation agreement were received, sampling began in May 2000 and completed in October 2000. The historical SDWA monitoring data (1993−1998) suggested that, where present, atrazine residues were not highly temporally variable. Thus potential temporal concentration fluctuations in finished drinking water at the tap of each groundwater CWS was not expected to vary significantly. The majority of CWS collected one single finished water sample in May, June, July, August, September, or October 2000. Eighty-nine CWS collected two finished water samples from different groundwater wells used by the same CWS.

Sample collection, shipment, storage, and disposition were recorded with a chain-of-custody form to track the process from the time of sampling to the end of sample analysis and disposal. Water samples were collected in two 1-L amber glass bottles as replicate A and B. Samples were shipped from the CWS to the Syngenta analytical laboratory with frozen blue ice packs in specially designed insulated containers to ensure samples were maintained at ∼4 °C during the Federal Express overnight or second-day shipment. Upon receipt, all water samples were kept in a refrigerator maintained at ∼4 °C prior to chemical analysis. The gas chromatography/mass selective detection (GC/MSD) method developed by Yokley and Cheung[18] was used to determine atrazine concentrations (LOQ = 0.05 ppb).

RESULTS AND DISCUSSION

Sample Stratification. Potential stratification variables were evaluated to determine which would be most predictive of atrazine detections within the PLEX data set. This evaluation was based on a subset of the 459 CWS with at least one atrazine detection (1993−1998); the subset had 256 CWS with an atrazine measurement in 1998 and at least one in prior years. Evaluation results for the five potential stratification variables are provided in Table 2.

Table 2. Potential Stratification Variables for 256 CWS with at Least One Detectable Atrazine Concentration (1993−1998) and a Measurement (Detect or Nondetect) in Both 1998 and Prior Years (1993−1997)

	distribution of CWS with 1998 data		distribution of CWS with 1998 detections		
potential strata[a]	*n*	%	*n*	%	percent with detectable atrazine in 1998
persons served at CWS					
[Q3, ∞)[b]	64	25.0	34	23.9	53.1
[Q2, Q3)	64	25.0	38	26.8	59.4
[Q1, Q2)	64	25.0	34	23.9	53.1
[0, Q1)	64	25.0	36	25.4	56.3
average atrazine use intensity (95−97)					
[Q3, ∞)	64	25.0	28	19.7	43.8
[Q2, Q3)	64	25.0	35	24.6	54.6
[Q1, Q2)	64	25.0	38	26.8	59.3
[0, Q1)	64	25.0	41	28.9	64.1
average of quarterly maxima (93−97)					
[Q3, ∞)	64	25.0	46	32.4	71.9
[Q2, Q3)	64	25.0	33	23.2	51.6
[0, Q2)	128	50.0	63	44.4	49.2
median of quarterly maxima (93−97)					
[Q3, ∞)	64	25.0	51	35.9	79.7
[0, Q3)	192	75.0	91	64.1	47.4
maximum concentration (93−97)					
[Q3, ∞)	65	25.4	37	26.1	56.9
[Q2, Q3)	63	24.6	35	24.6	55.6
[0, Q2)	128	50.0	70	49.3	54.7
detectable (93−97)					
yes	219	85.5	105	73.9	47.9
no	37	14.5	37	26.1	100.0

[a]Q1 = first quartile; Q2 = second quartile or median; Q3 = third quartile. [b][x1, x2) = the interval: x1 ≤ x < x2.

The rate of occurrence of detectable atrazine concentrations in 1998 was not strongly related to quartiles of the number of persons served by CWS, varying from 53.1% to 59.4%. As expected, this result indicates a low likelihood of predicting atrazine detection based on the number of persons served by CWS. However, the number of people served by CWS is an important parameter for the population-based regulatory aggregate dietary assessments that include drinking water as a route of exposure.

Table 3. Potential Strata Based on Population Served and Historical Atrazine Concentration Data or Average Atrazine Use Intensity for 256 CWS with at Least One Atrazine Detection from 1993 to 1998 and a Measurement (Detect or Nondetect) in Both 1998 and Prior Years (1993−1997)

stratum	number of persons served	median of quarterly concentration maxima or average atrazine use intensity	distribution of CWS with 1998 data		distribution of CWS with 1998 detections		% detectable atrazine in 1998
			n	%	*n*	%	
		Strata Based on Population Served and Historical Atrazine Concentration Data					
1	first (highest) quartile	top 25% in median of quarterly conc. maxima	16	6.3	12	8.5	75.0
2	first (highest) quartile	remaining 75% in median of quarterly conc. maxima	48	18.8	22	15.5	45.8
3	second quartile	top 25% in median of quarterly conc. maxima	16	6.3	14	9.9	87.5
4	second quartile	remaining 75% in median of quarterly conc. maxima	48	18.8	24	16.9	50.0
5	third quartile	top 25% in median of quarterly conc. maxima	16	6.3	13	9.2	81.3
6	third quartile	remaining 75% in median of quarterly conc. maxima	48	18.8	21	14.8	43.8
7	fourth (lowest) quartile	top 25% in median of quarterly conc. maxima	16	6.3	11	7.7	68.8
8	fourth (lowest) quartile	remaining 75% in median of quarterly conc. maxima	48	18.8	25	17.6	52.1
		Strata Based on Population Served and Average Atrazine Use Intensity					
1	first (highest) quartile	top 25% in average atrazine use intensity	16	6.3	7	4.9	43.8
2	first (highest) quartile	remaining 75% in average atrazine use intensity	48	18.8	27	19.0	56.3
3	second quartile	top 25% in average atrazine use intensity	16	6.3	9	6.3	56.3
4	second quartile	remaining 75% in average atrazine use intensity	48	18.8	29	20.4	60.4
5	third quartile	top 25% in average atrazine use intensity	16	6.3	7	4.9	43.8
6	third quartile	remaining 75% in average atrazine use intensity	48	18.8	27	19.0	56.3
7	fourth (lowest) quartile	top 25% in average atrazine use intensity	15	5.9	5	3.5	33.3
8	fourth (lowest) quartile	remaining 75% in average atrazine use intensity	49	19.1	31	21.8	63.3

Table 4. Strata and Target Sample Allocation for the Detect Domain (459 CWS with at Least One Atrazine Detection) and the Nondetect Domain (15 381 CWS with No Atrazine Detection) from 1993 through 1998 in the PLEX Database

stratum	number of persons served	median of quarterly concentration maxima or average atrazine use intensity	total CWS on sampling frame (N_h)	proportionate allocation	
				$N_{d,t}$ = 212	$N_{d,t}$ = 314
		Detect Domain			
11	first (highest) quartile	top 25% in median of quarterly conc. maxima	28	13	19
12	first (highest) quartile	remaining 75% in median of quarterly conc. maxima	86	40	59
13	second quartile	top 25% in median of quarterly conc. maxima	28	13	19
14	second quartile	remaining 75% in median of quarterly conc. maxima	86	40	59
15	third quartile	top 25% in median of quarterly conc. maxima	28	13	19
16	third quartile	remaining 75% in median of quarterly conc. maxima	86	40	59
17	fourth (lowest) quartile	top 25% in median of quarterly conc. maxima	28	12	19
18	fourth (lowest) quartile	remaining 75% in median of quarterly conc. maxima	89	41	61
		Nondetect Domain			
21	first (highest) quartile	top 25% in average atrazine use intensity	943	13	19
22	first (highest) quartile	remaining 75% in average atrazine use intensity	2,834	40	58
23	second quartile	top 25% in average atrazine use intensity	957	13	19
24	second quartile	remaining 75% in average atrazine use intensity	2,887	40	59
25	third quartile	top 25% in average atrazine use intensity	949	13	19
26	third quartile	remaining 75% in average atrazine use intensity	2,851	40	59
27	fourth (lowest) quartile	top 25% in average atrazine use intensity	957	13	19
28	fourth (lowest) quartile	remaining 75% in average atrazine use intensity	2,876	40	59
29	false positives in Kansas		127	2	3

Quartiles of average atrazine use intensity derived from 1995 to 1997 county level data appeared inversely related to the rate of occurrence of atrazine detections, varying from 43.8% for the highest quartile of usage to 64.1% for the lowest (Table 2). This may seem counterintuitive, but local and regional factors including well depth, treatment, soil and groundwater travel time, and hydrogeological characteristics can all complicate the relationship between atrazine use and detection in groundwater. Local groundwater recharge potential may not relate to finished groundwater atrazine residues especially where the CWS acquires water from confined aquifers.

Of all CWS in the highest quartile of the average quarterly concentration maxima (1993−1997), 71.9% had detectable atrazine levels in 1998. Conversely, ∼50% of the CWS in the other three quartiles had detectable concentrations in 1998. The first and second quartiles were combined for this analysis because many CWS had nondetectable quarterly maximum.

Table 5. Comparison of Target and Actual Number of Participants by Sampling Strata of the Atrazine Groundwater Monitoring Case Study

stratum	population served	median of quarterly concentration maxima or average atrazine use intensity	target number of respondents	actual number of respondents	percentage of target achieved
		Detect Domain			
11	first (highest) quartile	top 25% in median of quarterly conc. maxima	13	13	100.0
12	first (highest) quartile	remaining 75% in median of quarterly conc. maxima	40	39	97.5
13	second quartile	top 25% in median of quarterly conc. maxima	13	13	100.0
14	second quartile	remaining 75% in median of quarterly conc. maxima	40	39	97.5
15	third quartile	top 25% in median of quarterly conc. maxima	13	14	107.7
16	third quartile	remaining 75% in median of quarterly conc. maxima	40	38	95.0
17	fourth (lowest) quartile	top 25% in median of quarterly conc. maxima	12	12	100.0
18	fourth (lowest) quartile	remaining 75% in median of quarterly conc. maxima	41	36	87.8
		Subtotal for detect domain	212	204	96.2
		Nondetect Domain			
21	first (highest) quartile	top 25% in average atrazine use intensity	13	15	115.4
22	first (highest) quartile	remaining 75% in average atrazine use intensity	40	46	115.0
23	second quartile	top 25% in average atrazine use intensity	13	18	138.5
24	second quartile	remaining 75% in average atrazine use intensity	40	41	102.5
25	third quartile	top 25% in average atrazine use intensity	13	13	100.0
26	third quartile	remaining 75% in average atrazine use intensity	40	37	92.5
27	fourth (lowest) quartile	top 25% in average atrazine use intensity	13	16	123.1
28	fourth (lowest) quartile	remaining 75% in average atrazine use intensity	40	47	117.5
29		false positives in Kansas	2	2	100.0
		Subtotal for nondetect domain	214	235	109.8
		Total	426	439	103.1

Using the median of quarterly maxima (1993−1997), 79.7% of the CWS in the top quartile (25%) had a detectable concentration in 1998. For the remaining quartiles (75%), 47.4% of the CWS had a detectable concentration. Hence, the median of quarterly maxima seemed to be more predictive than the average for atrazine detections.

CWS separation by the maximum concentration (1993−1997) in terms of percent detectable atrazine measurements in 1998 varied only from 54.7% to 56.9% based on quartiles of this metric (Table 2). Again, the lowest two quartiles were combined due to the number of nondetectable samples in the group. Overall detection rate was 85.5% for the 256 CWS sampled in 1993−97; in 1998, the detection rate reduced to 55.5% (i.e., 142 out of 256 CWS with detectable atrazine concentrations in 1998).

Only two of the five variables examined in detail were feasible to stratify the nondetect domain of 15 381 CWS since there were no atrazine detects to use − number of persons served and average atrazine use intensity. Average atrazine use intensity could potentially be related to future atrazine detections (Table 2). Hence, both variables were used to stratify the nondetect domain.

The median of the quarterly atrazine maxima (1993−1998) was used as a useful stratification variable for the detect domian (Table 2). Although the number of persons served by CWS has low predictability for atrazine detection, it was selected as a stratification variable because of its potential use in aggregate dietary risk assessment for expressing exposure probability in terms of proportions of population-served. Using quartiles of the number of persons served as the first stratification variable, we investigated the utility of using either the median of the quarterly atrazine maxima or atrazine use intensity as a secondary stratification variable (Table 3). Using persons served as the second stratification variable after stratification by atrazine use intensity or quarterly maxima was also evaluated but was less effective (results not shown). Median of quarterly maxima was more effective, resulting in large interstratum heterogeneity in terms of percent wells with detections in 1998 (Table 3).

In summary, three stratification variables evaluated above were used in the final design of the atrazine groundwater CWS case study. Resulting frame counts (N_h) were given in Table 4 for each stratum in the detect (459 CWS) and nondetect (15 381 CWS) domains, respectively. An additional stratum was added to the nondetect group for 127 CWS in Kansas when the state determined that these CWS previously reported as having a detection actually did not have an atrazine detection. These 127 CWS with false positives (1993−1998) were set aside as a separate stratum in the nondetect domain.

Sample Size Requirement. Sample size results are provided in Table 1 based on three methods: tolerance interval, RSE, and skewness criterion.[13] According to tolerance interval, the smallest sample size of 59 is required to achieve 95% confidence such that the sample contains the 95th centile value. This requirement is similar to the estimate by the skewness criterion method which requires 69 samples. However, the sample size requirement increases to 212 if the goal is to estimate the 95th percentile with a predetermined precision at 30% RSE. Sample size increases rapidly as the target percentile, precision, or the confidence level increases.

Table 6. Response Rates for Determining Eligibility, Water Sampling Participation, and the Overall Response Rates for the Atrazine Groundwater Monitoring Case Study

	eligibility determination		water sampling		overall response rate	
stratum (h)	unweighted response rate	weighted response rate	unweighted response rate	weighted response rate	unweighted	weighted
			Detect Domain			
11	100.0	100.0	92.9	92.9	92.9	92.9
12	100.0	100.0	86.7	86.7	86.7	86.7
13	100.0	100.0	92.9	92.9	92.9	92.9
14	100.0	100.0	79.6	79.6	79.6	79.6
15	100.0	100.0	100.0	100.0	100.0	100.0
16	98.0	98.0	84.4	84.4	82.7	82.7
17	100.0	100.0	80.0	80.0	80.0	80.0
18	94.6	94.6	80.0	80.0	75.7	75.7
Subtotal	98.5	98.6	84.7	84.8	83.4	83.6
			Nondetect Domain			
21	87.5	87.5	79.0	79.0	69.1	69.1
22	100.0	100.0	59.7	59.7	59.7	59.7
23	93.3	93.3	64.3	64.3	60.0	60.0
24	100.0	100.0	69.5	69.5	69.5	69.5
25	95.0	95.0	68.4	68.4	65.0	65.0
26	97.2	97.2	56.1	56.1	54.5	54.5
27	96.9	96.9	51.6	51.6	50.0	50.0
28	91.5	91.5	58.0	58.0	53.1	53.1
29	100.0	100.0	66.7	67.3	66.7	67.3
Subtotal	96.0	96.2	61.4	62.1	58.9	59.7
Total	97.0	96.3	70.4	62.7	68.2	60.4

On the basis of these considerations, a sample size was selected that was sufficiently large to yield ~212 eligible, participating CWS with water samples in each of the two domains. No inflation of the sample size was necessary to account for the survey design effect as a result of proportionally allocating the sample to the strata within each domain. Hence, the design should be at least as efficient as a simple random sample.

On the basis of the RSE method, the sample size of 212 should achieve moderate precision or better (RSE ≤ 30%) for 95th percentile concentrations (Table 1). From the tolerance interval method, this sample size should provide high confidence levels (>99%) and that the actual 95th percentile of the distribution will be observed for each domain. For the skewness criterion approach, the 95th percentile or above can be estimated from 212 measurements.

Sample Allocation. Sample allocation results are shown in Table 4 for eight strata in each of the detect and nondetect domains, respectively. To account for nonresponders, ineligible CWS and CWS not providing an analyzable water sample, a sample of 314 CWS were initially targeted from each domain to ensure that the final number of participants would be close to the goal of 212.

Actual number of participating CWS totaled 439 (204 from the detect domain; 235 from the nondetect domain), and these were distributed among 21 major atrazine use states (Figure 1). Actual sample numbers were compared to target sample sizes for each stratum in Table 5. The number of participants was close to the target sample size, achieving ≥95% of the target for all but two strata. Only one stratum in each of two CWS domains achieved lower than 95% of the target sample size; 87.8% in the detect domain and 92.5% in the nondetect domain. For the whole sample, 439 CWS participated in the study, exceeding the target of 426. The high yield rate for all strata (all >85% compared to the target sample size) ensures the sufficiency of collected data for strong statistical inferences regarding estimation of the 95th and other high percentiles of the distribution of concentrations. Strong inference requires that the eligibility and response status of each sample CWS be documented and that water samples be successfully obtained and analyzed for a high proportion of eligible CWS. Statistical analysis weights that compensate for nonresponse from eligible CWS should be used in the final analyses to reduce nonresponse bias as described below.

CWS Response Rates. The response rates by strata to determine CWS eligibility and water sampling participation are summarized in Table 6. A random sample of CWS was selected from each stratum within the PLEX database for determining the response rates. The number of CWS was selected to be large enough to produce the desired number of participating CWS in each stratum based on the two target sample sizes listed in Table 4. A total of 695 CWS were actually selected randomly, and 674 of these facilities were determined to be eligible based on criteria provided in the methods section.

Weighted and unweighted response rates for the eligibility determination are provided in Table 6. Unweighted response rates are proportions of responding vs sampled CWS. Weighted response rates utilize sampling weights to estimate the proportion of eligible CWS in the population. Weighted response rate was 98.6% for the detect domain and 96.2% for the nondetect domain. The weighted overall response rate was 96.3%. The weighted overall response rate is similar to the response rate for the nondetect domain because the detect domain was heavily oversampled.

Response rates for collection and analysis of water samples and questionnaires among eligible CWS are also provided in Table 6. Results were used to adjust the analysis weight due to the second type of nonresponse (i.e., likelihood of an eligible CWS not to supply a water sample). Water samples were obtained and analyzed for 439 of 624 sample CWS that were known to be eligible. The unweighted overall response rate for participation in the study sampling (among verified eligible CWS) was 70.4% (84.7% for the detect domain, and 61.4% for the nondetect domain). Accounting for differing sampling and response rates for eligibility determination, the weighted estimate of the population response rate for well water sampling is 62.7% for the overall population (84.8% for the detect domain, and 62.1% for the nondetect domain).

The overall study response rate is the product of the response rates at the two stages of nonresponse discussed above. Weighted and unweighted overall study response rates are shown in Table 6. The unweighted response rate is 68.2% for the overall population, 83.4% for detect domain, and 58.9% for the nondetect domain. Correspondingly, the weighted response rate is 60.4% for the overall population, 83.6% for CWS in the detect domain, and 59.7% for CWS in the nondetect domain.

Monitoring Results. A histogram of all measured atrazine concentrations was plotted in Figure 3. The overall distribution was skewed toward the nondetection or low end of the measured values. Estimates of the target population distributions and their 95% confidence intervals of atrazine concentrations using the analysis weight method described above are presented in Table 7. Two approaches were used to evaluate the impact of nondetect (ND) samples (i.e., <LOQ) on statistical estimates. One approach was to replace all ND records with 0.025 ppb, which is half of the LOQ (0.05 ppb). The second approach was to use all instrument readings as low as 0.01 ppb (instrument measurable limit, IML). Estimated mean atrazine concentration for the overall population was 0.030 ppb (95% CI: 0.028, 0.033) by the 1/2 LOQ substitution approach and 0.0097 ppb (95% CI: 0.0069, 0.013) by the IML approach. For the detect domain, the estimated mean, 90th and 95th percentiles using the 1/2 LOQ approach were not statistically different (confidence interval overlap) from the corresponding values derived from the IML approach. The 90th and 95th percentile estimates were the same (0.36 and 0.51 ppb, respectively) for the 1/2 LOQ and IML methods. As expected, the detect domain had a higher estimated mean and other percentile values than the nondetect domain.

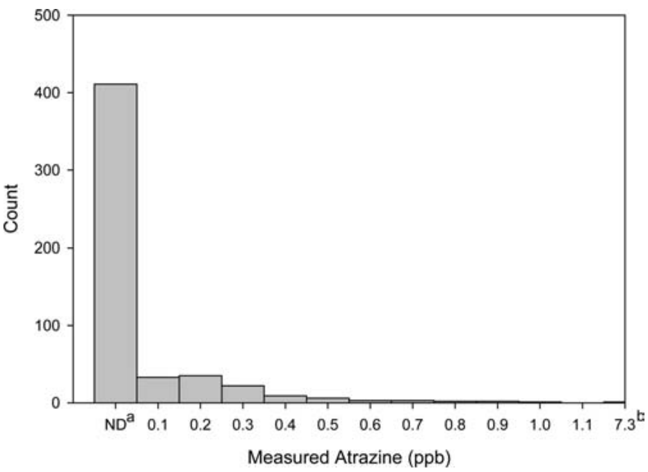


Figure 3. Measured atrazine concentrations (ppb) in finished groundwater samples from the atrazine monitoring case study of community water systems (CWS) using groundwater as their sole drinking water source. [a]ND = Nondetection (<0.05 ppb LOQ). [b]One detection of a nondrinking water well sample

Corresponding estimates for historical PLEX data using the maximum last quarter measurements between 1993 and 1998 were also provided in Table 7. The 50th, 90th, and 95th percentile values were not estimated for the nondetect domain and overall population (nor the 50th percentile of the detect domain) due to large number of ND values. Estimated mean for the nondetect domain and overall population was 0.00 and 0.0037 ppb, respectively, lower than the atrazine case study results with either the 1/2 LOQ or IML approach. For the detect domain of the PLEX data, estimated mean, 90th and 95th percentiles were 0.13, 0.28, and 0.44 ppb, respectively and were not statistically different (confidence internal overlap) from the corresponding estimates in the current case monitoring study.

In the PLEX database (1993−1998), the overall percent detectable rate for atrazine is 2.8%. The LOQs reported in the PLEX database for this time period ranged from 0.1 to 3 ppb. In the groundwater monitoring case study, 3.3% of the CWS water samples showed detectable levels (LOQ = 0.05 ppb), which was consistent with the PLEX data. However, percent detectable increased to 14% if the IML (0.01 ppb) was used for

Table 7. Statistical Estimates of Atrazine Concentration Distributions (95% Confidence Intervals) from the Groundwater Monitoring Case Study and Comparison with the Corresponding Results of the SDWA Maximum Last Quarter Data in PLEX (1993−1998)

	atrazine groundwater CWS Monitoring, 2000			
percentile	mean (ppb)	50th (ppb)	90th (ppb)	95th (ppb)
Using Half of the Laboratory Limit of Quantification (0.05 ppb) at 0.025 ppb				
detect domain	0.17 (0.091, 0.24)	NE[a]	0.36 (0.27, 0.46)	0.51 (0.40, 0.71)
nondetect domain	0.026 (0.025, 0.028)	NE	NE	NE
overall population	0.030 (0.028, 0.033)	NE	NE	NE
Using Lowest Analytical Instrument Cutoff at 0.01 ppb (μg/L)				
detect domain	0.16 (0.085, 0.23)	0.049 (0.034, 0.076)	0.36 (0.27, 0.46)	0.51 (0.39, 0.69)
nondetect domain	0.0053 (0.0035, 0.0072)	0.0021 (0.0020, 0.0023)	0.0089 (0.0057, 0.013)	0.015 (0.011, 0.043)
overall population	0.0097 (0.0069, 0.013)	0.0022 (0.0020, 0.0024)	0.0096 (0.0068, 0.015)	0.024 (0.013, 0.25)
1993−1998 PLEX (Maximum Last Qtr. Measurement)				
detect domain	0.13 (0.089, 0.18)	NE	0.28 (0.23, 0.37)	0.44 (0.30, 0.93)
nondetect domain	0.00 (0.00, 0.00)	NE	NE	NE
overall population	0.0037 (0.0025, 0.0049)	NE	NE	NE

[a]NE = Not able to estimate.

detect/nondetect determination. As expected, decreasing LOQs significantly increases the percent detectable rate.

The methodology underlying the design of the atrazine groundwater monitoring study is that of a probability-based sample survey, which is commonly used in social science research.[19] However, these same principles are applicable to other scientific disciplines for statistical inference regarding a population based on observing and measuring the members of a sample selected from that population.[20] An early application of these methods to environmental science research was the U.S. EPA "total exposure assessment methodology (TEAM") studies,[21] which sampled people residing in specific cities or counties to measure their exposures to volatile organic compounds. In the 1990s, the EPA decided to extend these studies to multiple sources and routes of exposure in EPA Region 5, six states in the Midwest area of the United States. The survey sampling issues that had to be considered when designing such studies were addressed by Callahan et al.[22] These principles were then applied to selection of individuals throughout EPA Region 5 for a field test for the National Human Exposure Assessment Survey.[23] These same principles have also been applied to surveys of drinking water wells[24] and industrial surface water impoundments.[25] These surveys are relatively straightforward applications of probability-based sampling in which the population is a countable set of "units" (people, wells, or industrial establishments). However, probability-based sampling methods also have been applied to environmental surveys of continuous spatial populations, such as the lakes, rivers, forests, and estuaries that comprise the EPA's Environmental Monitoring and Assessment Program (EMAP).[26,27] The probability-based survey sampling methods applied in the atrazine groundwater CWS study are statistically robust for making inferences regarding any well-defined population.[28]

In conclusion, the design and data analysis of the atrazine groundwater monitoring program by a survey sampling approach were described to characterize the distribution of concentrations in 15 840 groundwater CWS in 21 major atrazine use states. Methods for determining each design element were provided and discussed for sample stratification, sample size and allocation, analysis weight determination, and weighted population estimates. Three factors were found to best stratify samples for the atrazine case study: atrazine use, number of persons served by the CWS, and historical (1993–1998) atrazine detection rate. For the case study, only 3.3% of the finished groundwater samples contained detectable residues of atrazine (LOQ = 0.05 ppb). Means and percentiles were estimated for two scenarios of treating nondetection samples: (1) substituting nondetection with 1/2 LOQ and (2) including all instrument readings including values below LOQ as long as detectable by the analytical instrument. The 1/2 LOQ substitutions affected percentile estimates for the overall population of CWS and the nondetect domain. However, little effect was observed on higher centile estimates (>90th centile) of the detect domain. Including all instrument readings in the data analysis, median and 95th percentile estimates (95% confidence intervals) were 0.0022 ppb (0.0020, 0.0024) and 0.024 ppb (0.013, 0.25) for the overall population of CWS, respectively. For the detect domain, the median and 95th percentile estimates were 0.049 ppb (0.034, 0.076) and 0.51 ppb (0.39, 0.69), respectively. Corresponding estimates for the nondetect domain were lower than the overall population and the detect domain. Estimates of the mean, 90th and 95th percentile values for this study were in good agreement with corresponding results from the historic atrazine data of the SDWA monitoring programs, suggesting that the study design was adequate and effective to address previously stated study objectives. The robustness of this sampling approach should provide improved monitoring efficiency that can result in significant cost savings for similar regional and national monitoring programs. The elements of this sampling approach are applicable to the design of other environmental monitoring efforts aimed at probability-based exposure determination with statistical confidence and model validation.

AUTHOR INFORMATION

Corresponding Author

*Telephone: +1.336.632.2015. Fax: +1.336.632.7846. E-mail: Wenlin.Chen@syngenta.com.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The atrazine groundwater monitoring program would never have been accomplished without contributions by many Syngenta scientists, chemists, and CWS sampling personnel. We would like to especially thank Max W. Cheung, Robert A. Yokley, and Richard H. Ross for their dedication to the completion of the monitoring project and Peter Hertl for suggestions and review of the overall statistical design and analysis. Part of the statistical analysis was completed by statisticians Andrew Clayton and Brian Sutton at RTI International. Suggestions by Clint Truman during the preparation of this manuscript were much appreciated.

REFERENCES

(1) *Guidance for Selecting Input Parameters in Modeling the Environmental Fate and Transport of Pesticides*, Version 2.1; Office of Pesticide Programs, U. S. Environmental Protection Agency: Washington, DC, 2009; URL (http://www.epa.gov/oppefed1/models/water/input_parameter_guidance.htm) (accessed June 19, 2013).

(2) *Assessing Potential for Movements of Active Substances and their Metabolites to Ground Waters in the EU*; Report of the FOCUS Ground Water Work Group, EC Document Reference Sanco/13144/2010 version 1; URL (http://focus.jrc.ec.europa.eu/gw/docs/FinalreportGW.pdf) (accessed September 4, 2013).

(3) *Development and Use of the Index Reservoir in Drinking Water Exposure Assessments*; Office of Pesticide Programs, U.S. Environmental Protection Agency: Washington, DC, 2010; URL (http://www.epa.gov/oppefed1/models/water/index_reservoir_dwa.html) (accessed June 19, 2013).

(4) Jackson, S.; Hendley, P.; Jones, R.; Poletika, N.; Russell, M. Comparison of Regulatory Method Estimated Drinking Water Exposure Concentrations with Monitoring Results from Surface Drinking Water Supplies. *J. Agric. Food Chem.* **2005**, *53*, 8840–8847.

(5) *A Framework for Estimating Pesticide Concentrations in Drinking Water for Aggregate Exposure Assessments*; International Life Sciences Institute: Washington, DC, 1999.

(6) Larson, S.; Gilliom, R. Regression Models for Estimating Herbicide Concentrations in U.S. Streams from Watershed Characteristics. *J. Am. Water Resourc. Assoc* **2001**, *37* (5), 1349–1367.

(7) Chen, W.; Hertl, P.; Chen, S.; Tierney, D. Pesticide Surface Water Mobility Index and its Relationship with Concentrations in Agricultural Drainage Watersheds. *Environ. Toxicol. Chem.* **2002**, *21*, 298–308.

(8) Chen, W. Simple methods for estimating exposure concentrations of pesticide resulting from non-point source applications in agricultural

drainage networks. In *Techniques in Aquatic Toxicology*; Ostrander, G. K., Ed.; CRC Press: New York, 2005; Vol. *2*, p 357.
(9) Stone, W.; Gilliom, R. Watershed Regressions for Pesticides (WARP) Models for Predicting Atrazine Concentrations in Corn Belt Streams. *J. Am. Water Resourc. Assoc.* **2012**, *48*, 970−986.
(10) *Guidance Document for Exposure Assessment Based on Environmental Monitoring*, Health and Safety Publications Series on Testing and Assessment No. 185, ENV/JM/MONO(2013)7; OECD (Organisation for Economic Co-operation and Development) Environment Directorate: Paris, France, April 25, 2013.
(11) *Factoids: Drinking Water and Ground Water Statistics for 2009*; EPA 816-K-09-004; Office of Water U.S. Environmental Protection Agency: Washington, D.C., 2009.
(12) Helsel, D. R.; Hirsch, R. M. *Statistical Methods in Water Resources Techniques of Water Resources Investigations*; U.S. Geological Survey: Reston, VA, 2002; Book 4, Chapter A3.
(13) Mosquin, P.; Whitmore, R.; Chen, W. Estimation of Upper Centile Concentrations Using Historical Atrazine Monitoring Data from Community Water Systems. *J. Environ. Qual.* **2012**, *41*, 834−844.
(14) Hahn, G.; Meeker, W. Q. *Statistical Intervals: A Guide for Practitioners*; Wiley & Sons: New York, 1991.
(15) Boos, D.; Hughes-Oliver, J. How large does n have to be for Z and t intervals? *Am. Stat.* **2000**, *54*, 121−128.
(16) Brown, L.; Cai, T.; Dasgupta, A. Interval estimation for a binomial proportion. *Stat. Sci.* **2001**, *16* (2), 101−133. Crawford, C. G. Sampling Strategies for Estimating Acute and Chronic Exposures of Pesticides in Streams. *J. Am. Water Resour. Assoc.* **2004**, *40* (2), 485−502.
(17) Shah, B.; Barnwell, B.; Bieler, G. *SUDAAN User's Manual*, Release 7.5; Research Triangle Institute (RTI): Research Triangle Park, NC, USA, 1977.
(18) Yokley, R.; Cheung, M. Analytical method for determination of atrazine and its dealkylated chlorotriazine metabolites in water using gas chromatography/mass selective detection. *J. Agric. Food Chem.* **2000**, *48*, 4500−4507.
(19) Kalton, G. *Introduction to Survey Sampling, Sage University Paper Series on Quantitative Applications in the Social Sciences 07-001*; Sage Publications: London, 1983.
(20) Levy, P. S.; Lemeshow, S. *Sampling of Populations*; Wiley: New York, 1999.
(21) Wallace, L. A. *The Total Exposure Assessment Methodology (TEAM) Study*; EPA/600/6−87/002a; U.S. EPA Office of Research and Development: Washington, DC, 1987.
(22) Callahan, M. A.; Clickner, R. P.; Whitmore, R. W.; Kalton, G.; Sexton, K. Overview of Important Design Issues for a National Human Exposure Assessment Survey. *J. Exposure Anal. Environ. Epidemiol.* **1995**, *5*, 257−282.
(23) Whitmore, R. W.; Byron, M. Z.; Clayton, C. A.; Thomas, K. W.; Zelon, H. S.; Pellizzari, E. D.; Lioy, P. J.; Quackenboss, J. J. Sampling Design, Response Rates, and Analysis Weights for the National Human Exposure Assessment Survey (NHEXAS) in EPA Region 5. *J. Exposure Anal. Environ. Epidemiol.* **1999**, *9*, 369−380.
(24) Holden, L. R.; Graham, J. A.; Whitmore, R. W.; Alexander, W. J.; Pratt, R. W.; Liddle, S. K.; Piper, L. L. Results of the National Alachlor Well Water Survey. *Environ. Sci. Technol.* **1992**, *26*, 935−943.
(25) Johnson, B.; Balserak, P.; Beaulieu, S.; Cuthbertson, B.; Stewart, R.; Truesdale, R.; Whitmore, R.; Young, J. *Sci. Total Environ.* **2003**, *307*, 1−22.
(26) Stevens, D. L. Variable Density Grid-Based Sampling Designs for Continuous Spatial Populations. *Environmetrics* **1997**, *8*, 167−195.
(27) Stevens, D. L.; Olsen, A. R. Spatially Balanced Sampling of Natural Resources. *J. Am. Stat. Assoc.* **2004**, *99*, 262−278.
(28) Valliant, R.; Dever, J. A.; Kreuter, F. *Practical Tools for Designing and Weighting Survey Samples*; Springer: New York, 2013.