Environ Monit Assess (2009) 150:91–100
DOI 10.1007/s10661-008-0685-8

Survey design for lakes and reservoirs in the United States to assess contaminants in fish tissue

Anthony R. Olsen · Blaine D. Snyder · Leanne L. Stahl · Jennifer L. Pitt

Received: 29 January 2008 / Accepted: 14 February 2008 / Published online: 4 December 2008

Abstract The National Lake Fish Tissue Study (NLFTS) was the first survey of fish contamination in lakes and reservoirs in the 48 conterminous states based on a probability survey design. This study included the largest set (268) of persistent, bioaccumulative, and toxic (PBT) chemicals ever studied in predator and bottom-dwelling fish species. The U.S. Environmental Protection Agency (USEPA) implemented the study in cooperation with states, tribal nations, and other federal agencies, with field collection occurring at 500 lakes and reservoirs over a four-year period (2000–2003). The sampled lakes and reservoirs were selected using a spatially balanced unequal probability survey design from 270,761 lake objects in USEPA's River Reach File Version 3 (RF3). The survey design selected 900 lake objects, with a reserve sample of 900, equally distributed across six lake area categories. A total of 1,001 lake objects were evaluated to identify 500 lake objects that met the study's definition of a lake and could be accessed for sampling. Based on the 1,001 evaluated lakes, it was estimated that a target population of 147,343 (±7% with 95% confidence) lakes and reservoirs met the NLFTS definition of a lake. Of the estimated 147,343 target lakes, 47% were estimated not to be sampleable either due to landowner access denial (35%) or due to physical barriers (12%). It was estimated that a sampled population of 78,664 (±12% with 95% confidence) lakes met the NLFTS lake definition, had either predator or bottom-dwelling fish present, and could be sampled.

Keywords Fish tissue · Contaminants · Lakes · Reservoirs · Probability survey design · PBTs

A. R. Olsen (✉)
U.S. Environmental Protection Agency,
ORD/National Health and Environmental Effects Laboratory, Western Ecology Division,
200 S.W. 35th Street, Corvallis, OR 97333-4902, USA
e-mail: olsen.tony@epa.gov

B. D. Snyder · J. L. Pitt
Tetra Tech, Inc., Center for Ecological Sciences,
400 Red Brook Boulevard, Suite 200, Owings Mills, Baltimore, MD 21117, USA

L. L. Stahl
U.S. Environmental Protection Agency,
OW/Office of Science and Technology,
1200 Pennsylvania Avenue, NW (MC 4305),
Washington, DC 20460, USA

Introduction

In 1998, the U.S. Environmental Protection Agency's (EPA's) Office of Science and Technology (OST) within the Office of Water (OW) held a workshop to initiate a national study of contamination in fish tissue for lakes and reservoirs in the 48 conterminous states. Workshop participants were asked to discuss alternative survey designs, potential analytes of concern and their analytical methods, field sampling methods, target fish species, data management, data quality objectives, and other operational aspects necessary to implement the study. This National Study of Chemical Residues in Lake Fish Tissue (or National Lake Fish Tissue Study, NLFTS) was initiated in response to EPA's *Multimedia Strategy for Priority Persistent, Bioaccumulative, and Toxic (PBT) Pollutants* (USEPA 1998). Prior to development of the NLFTS, a lack of data prevented EPA from reporting on the national extent of fish tissue contamination in lakes and reservoirs. Information was available for individual lakes, but it could not be used collectively to report on all lakes and reservoirs. The NLFTS addressed this lack of information.

Large-scale studies of lakes had been conducted previously, but they did not include evaluation of fish tissue contaminants. In the 1970s, more than 800 lakes were sampled across the United States for the National Eutrophication Study (USEPA 1978). Lakes were sampled to characterize water quality, focusing on nutrients and eutrophication status. The study targeted lakes in all 48 conterminous states that had sewage treatment facilities on or near the lake and were of interest to state environmental agencies. Since the lakes were targeted and did not comprise a probability sample of all lakes (or any definable class of lakes), no generalization to U.S. lakes was possible.

In the 1980s, USEPA conducted two studies of lakes as part of the national acid precipitation assessment program. The Eastern Lake Survey (ELS) and the Western Lake Survey (WLS) were conducted to quantify the chemical status of lakes in the United States in regions potentially susceptible to the effects of acidic deposition (Landers et al. 1988; Blick et al. 1987). The probability survey design involved three levels of stratification: three geographic regions, subregions within the regions, and three alkalinity classes. Subregions identified areas within each region that were relatively homogeneous and were used as strata so that distinct geographic portions were represented. In all, 33 strata were used with a sample size of at least 50 in each. U.S. Geological Survey (USGS) topographic maps with a scale of 1:250,000 were used to identify all lake objects. Selected lake objects were then evaluated to determine if they were a lake of interest (e.g., only lakes >4 ha could reliably be determined from the maps).

In the early 1990s, the Environmental Monitoring and Assessment Program (EMAP) in EPA's Office of Research and Development (ORD) conducted a regional pilot study of lakes in the northeast (Larsen and Christie 1993; Larsen et al. 1994; Yeardley et al. 1998). The assessment focused on indicators of lake condition and associated stressors. Lakes were selected using a probability survey design, which applied spatial units of 40-km^2 hexagons and identified all lakes greater than 1 hectare within those hexagons on USGS topographic maps with a scale of 1:100,000. Once the lakes were identified, a probability sample was selected that was spatially balanced.

Based on the prior experience of EMAP and others with national lakes studies, EPA determined that probabilistic (random) surveys would be an appropriate approach to address the objectives of the NLFTS. The NLFTS was a precursor to the EPA's national aquatic resource assessments, which are designed to provide statistically defensible data to characterize the national condition of lakes, streams, rivers, and coastal waters (USEPA 2006, 2007a).

Lakes and reservoirs became the focus for this freshwater fish contamination study because they are environments where contamination may be more likely to accumulate than in flowing waters. They occur in a variety of landscapes (e.g., urban, agricultural, and wilderness), and they can receive contaminants from several sources, including direct discharges into the water, air deposition, and agricultural or urban runoff. Monitoring fish contamination in lakes and reservoirs is a critical activity for protecting human health because lakes and reservoirs provide important sports fisheries and other recreational opportunities. Lake ecosystems also provide critical habitat for aquatic species, and they support wildlife populations that depend on aquatic species for food. The 2007 update of the National Listing of Fish Advisories (USEPA 2007b) provided support for the value of monitoring fish contamination as one way to protect human health when it reported that 38% of the Nation's total lake acres were under fish consumption advisories at the end of 2006.

This paper describes the NLFTS probability survey design, including its implementation and statistical estimation procedures, and presents estimates on the extent of lakes and reservoirs in the 48 conterminous United States. Additional information about the study is available in USEPA (1999, 2000, 2005a, b) and Stahl et al. (2008).

Methods

Study design

The objective of the NLFTS was to estimate the national distribution of the mean levels (i.e., composite average concentrations or "lake means") of selected persistent, bioaccumulative, and toxic chemical (PBT) residues in fish tissue from lakes and reservoirs of the conterminous United States. The study design generated data to develop national estimates of the median concentrations of PBT chemicals in lake fish, to estimate the percentage of lakes and reservoirs with fish tissue concentrations above a specified threshold related to human health, and to define a national baseline for tracking changes in concentrations of PBT chemicals in freshwater fish as a result of pollution control activities. A probability-based survey design formed the basis for the study, and it required a definition of the target population of lakes, creation of a sample frame of lakes, and determination of an appropriate survey design. Implementation of the survey design required an evaluation of the selected lakes to determine if they were in the target population and could be sampled, field and laboratory operations to collect data, adjustment of weights associated with the survey design, and finally, estimation of the extent of the target population and distribution of PBTs in fish tissue. Field and laboratory procedures are described and results for PBTs are presented in Stahl et al. (2008).

Probability sampling provides the basis for estimating resource extent and condition, for characterizing trends in extent or condition, and for representing spatial pattern, all with known certainty. A probability sampling design has some inherent characteristics that distinguish it from other sampling designs. First, the population being sampled is explicitly described. Second, every element in the population has the opportunity to be sampled with known probability. Third, the selection process includes an explicit random element. Since the specific purpose of the NLFTS was to describe the condition of resources on a national basis, a probability-based design was an essential component of the study.

Survey design

The target population for this study was all lakes and reservoirs within the conterminous United States that met minimum size requirements and contained a permanent fish population, excluding the Laurentian Great Lakes and the Great Salt Lake. Use of the term "lakes" refers collectively to lakes and reservoirs. A lake was defined as a permanent body of water of at least one hectare (2.47 acres) in surface area with a minimum of 1,000 m^2 of open (non-vegetated) water and a depth of at least one meter. Since the lakes also needed to have a permanent fish population, lakes subject to annual fish winterkill or recently stocked with fingerlings were rejected during the lake evaluation process. Stocked lakes with adult fish introduced at least three years prior to sampling were accepted as having a permanent fish population since fish residence time would be sufficient to allow accumulation of contaminants.

In contrast to earlier lake surveys that used topographic maps to identify lakes, the sample frame used to generate the list of lakes was River Reach File Version 3 (RF3) (Horn and Grayman 1993). When the lakes were selected in early 1999, RF3 provided the best available GIS coverage for lakes and reservoirs in the United States. One important exception in the coverage was newly constructed large reservoirs. A list of all lakes and reservoirs greater than 5,000 ha in RF3 was provided to federal, state, and tribal study partners who were asked to verify the lakes and add any lakes greater than 5,000 ha that were not on the list. The resulting information was used to update RF3 before initiating lake selection. Point coverage was also created prior to lake selection.

Fig. 1 Sample frame from River Reach File Version 3 (RF3)

Figure 1 shows the geographic distribution of lakes in the sample frame. Past experience with RF3 had shown that it includes lake objects that are not lakes (e.g., glaciers or sandpits) and that more errors occur with smaller lakes than larger lakes. This information was considered while developing the survey design.

The study used a spatially balanced, unequal probability of selection survey design. Lakes were selected using the procedures described by Stevens and Olsen (1999) and Stevens and Olsen (2004). For the unequal probability of selection, lakes were divided into six size categories based on surface area of the lake expressed in hectares (ha): ≥1–5 ha, >5–10 ha, >10–50 ha, >50–500 ha, >500–5,000 ha, and >5,000 ha. Table 1 lists the number of lakes available in RF3 by size category. Note that lakes in the smallest category (≥1–5 ha) accounted for more than 60% of the lakes available to be drawn, whereas lakes in the two largest categories (>500–5,000 ha and >5,000 ha) represented less than 1% of the lakes in RF3. The probability of selection for a lake was chosen so that an approximately equal number of lakes would be selected in each lake size category. The inclusion probability was adjusted for the three smaller size categories to account for the larger number of lakes expected that would not meet the study definition of a lake. These adjustments were as follows: increase by 40% for ≥1–5 ha, increase by 30% for >5–10 ha, and increase by 20% for >10–50 ha. No adjustment was made for the remaining size categories (>50–500 ha, >500–5,000 ha, and >5,000 ha). These adjustments were based on limited information from the EMAP northeastern lake survey (Larsen and Christie 1993; Larsen et al. 1994). The impact of an incorrect adjustment would be that the number of lakes actually sampled by size category would be unequal.

Table 1 Number of lakes by lake area in sample frame (RF3)

Lake area (ha)	Number of lakes	Frequency (%)	Cumulative number of lakes	Cumulative frequency (%)
≥ 1 – 5	172,747	63.8	172,747	63.8
> 5 – 10	44,996	16.6	217,743	80.4
> 10 – 50	40,016	14.8	257,759	95.2
> 50 – 500	11,228	4.1	268,987	99.3
> 500 – 5,000	1,500	0.6	270,387	99.9
> 5,000	274	0.1	270,761	100.0

Table 2 Number of lakes selected for potential sampling

Lake area (ha)	Year 1		Year 2		Year 3		Year 4		All Years	
	Base	Reserve	Base	Reserve	Base	Reserve	Base	Reserve	Base	Reserve
$\geq 1-5$	39	47	41	48	47	48	47	49	174	192
$> 5-10$	44	45	40	52	47	40	46	42	177	179
$> 10-50$	32	36	47	39	46	42	25	41	150	158
$> 50-500$	34	36	37	26	29	40	34	22	134	124
$> 500-5,000$	36	38	30	29	31	30	41	37	138	134
$> 5,000$	40	23	30	31	25	25	32	34	127	113
Total	225	225	225	225	225	225	225	225	900	900

It is important to recognize that an unequal probability survey design was used and not a stratified survey design. Stratifying the design would have required lakes to be replaced within each size category. Although that may have been desirable, it would have complicated field operations. The major impact of not stratifying is on variance estimation. The local neighborhood variance estimator (see below) is applied to all lakes in the neighborhood regardless of lake size category. Smaller variance estimates (compared to a stratified variance estimator) may occur if less variation in fish tissue concentrations occurs locally in space compared to variation across lake size categories.

The sample size for the study was 500 lakes to be sampled over 4 years. To insure that sufficient lakes were available to meet this target, 900 lakes were initially selected for the study. This was based on prior information from the northeastern lake survey regarding the number of lakes that were expected not to meet the target population definition and that were expected not to be accessible due to landowner denial or physical barriers (Peterson et al. 1995, 1998). An additional reserve list of 900 lakes was also selected. Table 2 provides a summary of the number of lakes selected by size category and year for base and reserve lakes. The lakes selected for each sampling year comprise an annual statistical subset that provides a nationally representative sample. For example, if all lakes designated for sampling in year 1 were sampled in the same year, then a national estimate of fish tissue contamination could be made based on those lakes. This provision was incorporated in the design for three reasons: to distribute field operations over four years to make it feasible to complete the study within field crew constraints, to allow differences across years to be examined, and to provide the option of sampling less than the full complement of 500 lakes, if necessary. An attempt was made to sample each annual panel in the year specified, although field operational considerations precluded this from being the case. Most lakes in each annual panel were sampled in the designated year, but not all. Consequently, annual estimates cannot be derived unless an assumption is made that lakes from a panel not sampled in the designated year can be combined with those sampled in the designated year.

Lake evaluation and field sampling

A critical element in the implementation of the survey design was the determination of the status of each lake in the sample. This means that each lake was checked to determine if it met the target population definition of a lake for the study. Where possible, the determination was made without a site visit; however, field reconnaissance was necessary for some lake evaluations. A record of the lake evaluation was maintained and was used to estimate the number of lakes in the target population. Two other situations occurred that resulted in a lake not being sampled. First, some lakes were on private land and required landowner permission to access. All landowner refusals were documented and recorded. Secondly, some lakes were physically inaccessible. When logistical or safety constraints made a lake inaccessible, the reason was recorded.

Field sampling for the NLFTS began in 2000 and continued through 2003. Fish sampling consisted of collecting two fish composites per lake (one composite of predators and another of bottom dwellers). Each composite consisted of five adult fish of the same species and of similar size (so that the smallest individual within the composite was no less than 75% of the total length of the largest individual). Replicate samples were collected from about 10% of the lakes to estimate sampling variability. Most fish composites were collected during the summer and fall of each sampling year. Stahl et al. (2008) provides additional details on fish sampling, contaminants measured, and analytical results.

Calculating the sample weights

A critical activity for analyzing data from a study with an unequal probability survey design is deriving the sample weights. For this study, lake size (surface area) was used to assign a probability of selection (or inclusion probability) to each lake. Statistical analyses required the use of the weights derived from the unequal probability of selection (Thompson 1992).

The statistical weight for each lake is the inverse of its probability of selection or inclusion probability. Initial sample weights were calculated for each lake based on the unequal probabilities of selection. The inclusion probabilities for the smallest to the largest lake size categories were as follows: 0.001065142, 0.003822562, 0.003898441, 0.011756323, 0.088000000, and 0.452554745. These probabilities resulted in expected total sample sizes of 184, 172, 156, 132, 132, and 124 lakes, respectively. They also resulted in initial design weights of 938.84, 261.60, 256.51, 85.06, 11.36, and 2.21, respectively. Note that the sample weights are expressed as numbers of lakes. The final sample weights were derived from adjusted inclusion probabilities based on the total number of lakes identified for potential sampling at the conclusion of the lake evaluation process. This weight adjustment was made individually for each lake size category. The adjusted weight is the initial weight multiplied by the ratio of the number of lakes in the sample frame divided by the product of the number of lakes evaluated and their initial weight. For example, if the number of lakes evaluated in the >5000 ha size category was 139, then the adjustment would be:

$$2.21 \times \frac{274}{(139 \times 2.21)} = 1.97$$

Population estimation

The data necessary for estimating the number of target lakes are the evaluation status results compiled for all lakes statistically drawn for potential field sampling. Diaz-Ramos et al. (1996) describe the statistical procedure used to estimate the total (target) population from the unequal probability sample of lakes. An associated variance estimate, called a local neighborhood variance estimate, is described by Stevens and Olsen (2003).

The target population of lakes for this study included lakes on both public and private lands. All target lakes met the study definition of a lake, but some lakes could not be sampled. Landowner denial to access lakes on private property was a major factor in preventing field teams from sampling privately owned target lakes. Another important factor was physical barriers that made some target lakes inaccessible (e.g., reservoir draw downs). The number of inaccessible target lakes was used to develop an estimate of the sampled population of lakes (i.e., the number of target lakes that could be sampled).

National estimates of fish tissue concentrations for each fish composite type (i.e., predator fillets and bottom-dweller whole bodies) for all the target chemicals were also calculated using procedures described by Diaz-Ramos et al. (1996). Tissue concentration data from laboratory analysis of the fish composite samples, along with the final sample weights associated with each lake, were used to derive these estimates. The tissue concentration distributions are reported as percentiles, including the 50th percentile or median concentration, for each target chemical and composite type. The estimated proportion (p_c) below a specific value for a concentration (C) is:

$$p_C = \frac{\sum_{i=1}^{n} w_i * x_i}{\sum_{i=1}^{n} w_i}$$

where: $x_i = 1$ if concentration for ith lake is below C and equals 0 otherwise, $w_i =$ the adjusted weight for ith lake, and n = total number of lakes sampled.

Variance estimates were derived using the local neighborhood variance estimator described by Stevens and Olsen (2003, 2004). Analyses were completed using the R statistical software (R Development Core Team 2007) and an R contributed library for probability survey population estimation (spsurvey) that is available online at http://www.epa.gov/nheerl/arm/analysispages/software.

The objectives for the NLFTS focused only on national estimates. Estimates can also be made for sub-populations, such as geographic regions or lake area size classes. Sub-population estimates will have wider confidence intervals since the sample sizes will be smaller. With a sample of 500 lakes nationally and with the sample being spatially balanced, it is unlikely that reliable estimates for all ten USEPA regions would be possible. All that is required to derive sub-population estimates is to define the sub-population as a particular subset of the target population, identify which lakes in the sample are in that population, and then complete the estimation as described above using only those lakes.

Results

A total of 1,001 lakes were evaluated to obtain a sample size of 500 lakes where composite fish samples could be collected. Since the initial list included only 900 lakes, 101 reserve lakes were evaluated in the final year of field sampling. During the four years of sampling, field teams collected fish from 443 lakes on the initial lake list and 57 lakes on the reserve lake list. Figure 2 shows the geographic distribution of all 1,001 lakes identified as non-NLFTS lakes, NLFTS lakes not sampled, and the 500 NLFTS sampled lakes. Note that the geographic distribution of the 1,001 lakes evaluated has a similar density of lakes as the sample frame, a property of the spatially balanced survey design. The sampled lakes also have similar density, although that is partially the result of non-NLFTS lakes and non-sampled NLFTS lakes being distributed across the 48 states.

Two sets of lakes, the target population and the sampled population, define how broadly the fish tissue concentration results apply to lakes and reservoirs in the United States. The target population consists of all lakes and reservoirs in the lower 48 states that met the study lake size requirements and contained a permanent fish population. Based on the evaluation of the 1,001 lakes, it is estimated

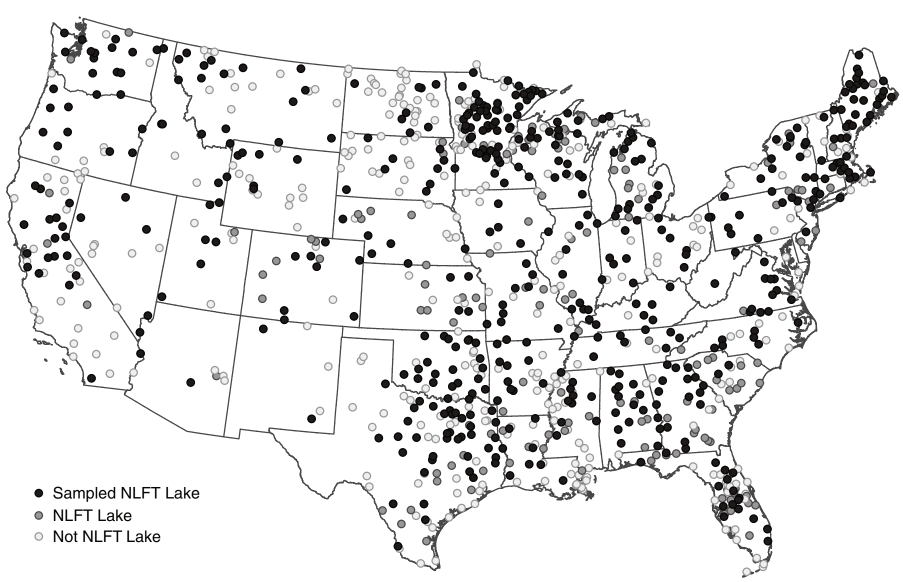

Fig. 2 Lakes evaluated and sampled for the NLFTS

Table 3 Estimated number of lakes by evaluation status, lake size category, and composite type

Estimated number of lakes by evaluation status	Lake area (ha)						Total
	≥1–5	>5–10	>10–50	>50–500	>500–5,000	>5,000	(95% confidence interval)
Lakes in RF3 (actual)	172,747	44,996	40,016	11,228	1,500	274	270,761
Non-target lakes							123,418 ± 8%
Saline lake	1,809	466	710	217	29	8	3,240 ± 66%
Not a lake	47,031	15,620	11,365	2,898	380	6	77,300 ± 12%
Lake <1 hectare	5,427		–	–	–	–	5,427 ± 63%
No fish in lake	17,184	3,031	1,657	72	–	2	21,947 ± 28%
No permanent fish	9,044	3,730	2,368	362	–	–	15,505 ± 30%
Target population of lakes							147,343 ± 7%
Inaccessible target lakes							
Land owner denied access	39,795	5,595	5,209	724	–	–	51,424 ± 16%
Lake physically inaccessible	9,948	4,430	2,605	362	10	–	17,355 ± 29%
Sampled population of lakes							
Predator sampled population	41,604	16,101	11,191	6,375	1,032	256	76,559 ± 12%
Bottom dweller sampled population	19,898	11,602	8,160	5,360	935	235	46,190 ± 17%
Total sampled population[a]	42,508	12,123	16,101	6,592	1,081	258	78,664 ± 12%

[a]The total sampled population of lakes includes lakes that yielded either predator or bottom-dweller species

that 147,343 (±7% with 95% confidence) lakes are in the target population for this study. The sampled population consists of all target lakes that were accessible for fish collection. Under ideal circumstances, the target and sampled populations should coincide. In this study, the sampled population is a subset of the target population. A number of target lakes were not accessible to field sampling teams because the lakes were located in remote wilderness areas or on private property where landowners denied permission to sample them (Table 3). The sampled population size is an estimated 78,664 (±12% with 95% confidence) lakes. The sampled population for all fish samples includes lakes where either predators or bottom dwellers or both are present in the lake. There is a different sampled population for each composite type based on differences in the occurrence of predators and bottom dwellers in the 500 sampled lakes. The sampled population for predators contains an estimated 76,559 (±12% with 95% confidence) lakes because most sampled lakes had predators present. The sampled population for bottom dwellers consists of an estimated 46,190 (±17% with 95% confidence) lakes since more lakes (e.g., high altitude lakes) did not have bottom dwellers. In both cases, these statements depend on the probability that predators or bottom dwellers were caught when they were actually present in the lake.

Of the 270,761 total number of lake objects in the sample frame, an estimated 123,418 (±8%) were non-target lakes (Table 3). Of these, approximately 63% did not meet the study definition of a lake for a reason other than not meeting size criteria, being saline, or not having a fish population. For example, these lakes may have been wetlands or commercial fish ponds, or they may have been too shallow or nonexistent. For 18% of the lakes,

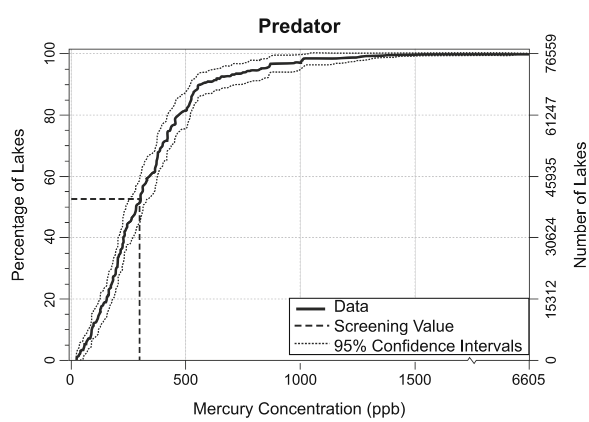

Fig. 3 Example cumulative distribution function estimate for mercury in predator fish (from Stahl et al. 2008)

fish were either known not to be present or were not caught. No permanent fish population (mainly due to winter fish kill) was the reason for 12.5% of the sites being listed as non-target lakes. Less than 5% of the lakes were non-target because they were saline or <1 ha in size.

Of the estimated 147,343 target lakes, 47% were estimated not to be sampleable either due to landowner access denial (35%) or physical barriers (12%). Most of the lakes where landowners denied access were small lakes (≥1–5 ha). Physically inaccessible lakes occurred in all but the largest lake area size category (>5,000 ha).

Figure 3 provides an example of an estimated cumulative distribution function (CDF) for mercury in predator fish composite samples from lakes (Stahl et al. 2008). The CDF (Fig. 3) showed that edible portions (fillets) of predators in 48.8% of the sampled population of lakes had tissue concentrations that exceeded EPA's 300 ppb fish tissue criterion for mercury, representing a total of 36,422 lakes nationwide. See Stahl et al. (2008) for other results.

Discussion

The NLFTS demonstrated that it is possible to conduct a national lake study based on a probability survey design (e.g., to investigate fish tissue contamination in lakes). Stahl et al. (2008) report PBT contaminants in fish tissue for predators and bottom dwellers collected during this survey. It is important to recognize that the target population of interest was lakes and that an "index" sample of fish within the lakes was used. That is, the study was not a probability sample of fish population. The latter is impossible to implement, especially when particular species are of interest. Collecting an index sample of fish from a probability sample of lakes is feasible and provides information on number of lakes that have fish tissue with contaminants present.

Implementation of a national study must address questions concerning the reliability of the sample frame, landowner access, and physical inaccessibility of lakes. A notable finding based on the evaluation of the lakes in RF3 is that RF3 includes many lake objects that are in fact not lakes under the definition of a lake used in this study. This statement will also apply to the new National Hydrography Dataset (NHD) since it is also based on the same 1:100,000 scale maps as RF3. Although probability surveys of lakes can still be based on RF3 or NHD as the sample frame, additional costs are incurred for office and field evaluation of the lake objects. For the NLFTS, 1,001 lakes had to be evaluated to identify 500 sampleable lakes. It is also important to note that, for 35% of the target lakes, landowners either denied access or did not respond to requests for access. If landowner access was attempted, it was assumed that the lake object was a lake. This may not be true in all cases. All lakes were physically accessible, although access to some lakes required significant cost in terms of field crew time and expense. It is important to minimize the loss of lakes due to physical inaccessibility, since these are more likely to be free of local human influences.

Without additional assumptions, inferences cannot be made regarding the levels of PBT contaminants in physically inaccessible lakes or those where landowners denied access. Since landowner denial lakes tended to be smaller, they may be more likely to have characteristics of smaller lakes that occur on private land. If such a class of lakes was defined and it is assumed that the lakes are missing at random, then estimates for that class could be applied to those lakes. Physically inaccessible lakes are unlikely to be missing at random and it may not be possible to define a class of similar lakes that could be sampled. Therefore, without additional assumptions and statistical modeling, it is unlikely that properties of physically inaccessible lakes can be inferred.

Spatial information from the NLFTS was used in the probability survey design for a subsequent national lake assessment by the EPA Office of Water (USEPA 2007c). Further refinement of a national sample frame for lakes in the 48 conterminous states would improve the ability to conduct national probability surveys of lakes.

Acknowledgements The information in this document was funded by the U.S. Environmental Protection Agency, and technical support was provided under EPA Contracts EP-C-04-030 and ET-W-06-046, MOBIS Contract GS-23F-9820, and Sample Control Center Contract 68-C-98-139. The manuscript has been subjected to review by the

Western Ecology Division of ORD's National Health and Environmental Effects Research Laboratory and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

References

Blick, D. J., Messer, J. J., Landers, D. H., & Overton, W. S. (1987). Statistical basis for the design and interpretation of the national surface water survey, phase I: Lakes and streams. *Lake and Reservoir Management*, *3*, 470–475.

Diaz-Ramos, S., Stevens, D. L., Jr., & Olsen, A. R. (1996). EMAP statistics methods manual. U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Corvallis, OR. EPA/620/R-96/002.

Horn, C. R., & Grayman, W. M. (1993). Water-quality modeling with EPA reach file system. *Journal of Water Resources Planning and Management*, *119*, 262–274.

Landers, D. H., Overton, W. S., Linthurst, R. A. & Brakke, D. F. (1988). Eastern lake survey: Regional estimates of lake chemistry. *Environmental Science & Technology*, *22*, 128–135.

Larsen, D. P., & Christie, S. J. (1993). EMAP-Surface waters 1991 pilot report. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory, Corvallis, Oregon. EPA-620-R-93-003.

Larsen, D. P., Thornton, K. W., Urquhart, N. S., & Paulsen, S. G. (1994). The role of sample surveys for monitoring the condition of the Nation's lakes. *Environmental Monitoring and Assessment*, *32*, 101–134.

Peterson, S. A., Hughes, R. M., Larsen, D. P., Paulsen, S. G., & Omernik, J. M. (1995). Regional lake quality patterns: Their relationship to lake conservation and management decisions. *Lake and Reservoirs: Research and Management*, *1*, 163–168.

Peterson, S. A., Larsen, D. P., Paulsen, S. G., & Urquhart, N. S. (1998). Regional lake trophic patterns in the northeastern United States: Three approaches. *Environmental Management*, *22*, 789–801.

R Development Core Team (2007). The R Project for Statistical Computing. R Development Core Team, Vienna, Austria (website) http://www.R-project.org.

Stahl, L. L., Snyder, B. D., Olsen, A. R., & Pitt, J. L. (2008). Contaminants in fish tissue from U.S. lakes and reservoirs: A national probabilistic study. *Environmental Monitoring and Assessment*. doi:10.1007/s10661-008-0669-8.

Stevens, D. L., Jr., & Olsen, A. R. (1999). Spatially restricted surveys over time for aquatic resources. *Journal of Agricultural, Biological, and Environmental Statistics*, *4*, 415–428.

Stevens, D. L., Jr., & Olsen, A. R. (2003). Variance estimation for spatially balanced samples of environmental resources. *Environmetrics*, *14*, 593–610.

Stevens, D. L., Jr., & Olsen, A. R. (2004). Spatially-balanced sampling of natural resources. *Journal of American Statistical Association*, *99*, 262–278.

Thompson, S. (1992). *Sampling*. New York, NY: John Wiley and Sons.

U.S. Environmental Protection Agency (USEPA) (1978). *National eutrophication survey, working papers 474, 475, 476, 477*. Corvallis Environmental Research Laboratory and Las Vegas Environmental Monitoring and Support Laboratory, Washington, DC.

U.S. Environmental Protection Agency (USEPA) (1998). *A multimedia strategy for priority persistent, bioaccumulative, and toxic (PBT) Pollutants [Draft]*. U.S. EPA, Office of Water, Washington, DC. EPA-823-R-99-014.

U.S. Environmental Protection Agency (USEPA) (1999). *National study of chemical residues in lake fish tissue: Study design*. U.S. Environmental Protection Agency, Office of Science and Technology, Washington, DC.

U.S. Environmental Protection Agency (USEPA) (2000). *Field sampling plan for the National Study of Chemical Residues in Lake Fish Tissue*. U.S. EPA, Office of Water, Office of Science and Technology, Washington, DC. EPA-823-R-02-004.

U.S. Environmental Protection Agency (USEPA) (2005a). *Quality assurance report for the National Study of Chemical Residues in Lake Fish Tissue: Analytical data for years 1 through 4*. U.S. EPA, Office of Water, Washington, DC. EPA-823-R-05-005.

U.S. Environmental Protection Agency (USEPA) (2005b). *Quality assurance project plan for data analysis activities for the national study of chemical residues in lake fish tissue*. Washington, DC: U.S. Environmental Protection Agency, Office of Water.

U.S. Environmental Protection Agency (USEPA) (2006). *Wadeable stream assessment: A collaborative survey of the Nation's streams*. U.S. EPA, Office of Water, Washington, DC. EPA 841-B-06-002.

U.S. Environmental Protection Agency (USEPA) (2007a). *National estuary program coastal condition report*. U.S. EPA, Office of Water, Washington, DC. EPA-842-B-06-001.

U.S. Environmental Protection Agency (USEPA) (2007b). *2005/2006 National listing of fish advisories fact sheet*. U.S. Environmental Protection Agency, Office of Water, Washington, DC. EPA-823-F-07-003.

U.S. Environmental Protection Agency (USEPA) (2007c). *Survey of the Nation's lakes*. U.S. EPA (website). http://www.epa.gov/owow/lakes/lakessurvey/.

Yeardley, R. B., Jr, Lazorchak, J. M., Paulsen, S. G. (1998). Elemental fish tissue contamination in northeastern U.S. lakes: Evaluation of an approach to regional assessment. *Environmental Toxicology and Chemistry*, *17*, 1875–1884.