DESIGNING A SPATIALLY BALANCED, RANDOMIZED SITE SELECTION PROCESS FOR REGIONAL STREAM SURVEYS: THE EMAP MID-ATLANTIC PILOT STUDY

ALAN T. HERLIHY[1], DAVID P. LARSEN[2], STEVEN G. PAULSEN[2]
N. SCOTT URQUHART[3] and BARBARA J. ROSENBAUM[4]

[1]*Department of Fisheries & Wildlife, Oregon State University, c/o USEPA, 200 SW 35th Street, Corvallis, OR 97333 USA;* [2]*U.S. Environmental Protection Agency, NHEERL-WED, 200 SW 35th Street, Corvallis, OR 97333 USA;* [3]*Department of Statistics, Oregon State University, Corvallis, OR 97331 USA;* [4]*OAO Corp., c/o U.S. EPA, 200 SW 35th Street, Corvallis, OR 97333 USA*

Abstract. In 1993, the U.S. Environmental Protection Agency (EPA), as part of the Environmental Monitoring and Assessment Program (EMAP), initiated a sample survey of streams in the mid-Atlantic. A major objective of the survey was to quantify ecological condition in wadeable streams across the region. To accomplish this goal, we selected 615 stream sites using a randomized sampling design with some restrictions. The design utilized the digitized stream network taken from 1:100,000-scale USGS topographic maps as the sample frame. Using a GIS, first- through third-order (wadeable) stream segments in the sample frame were randomly laid out in a line and sampled at fixed intervals after a random start. We used a variable probability approach so that roughly equal numbers of first-, second-, and third-order stream sites would appear in the sample. The sample design allows inference from the sample data to the status of the entire 230,400 km of wadeable stream length in the mid-Atlantic study area. Of this mapped stream length, 10% was not in the target population because no stream channel existed (4%), the stream channel was dry (5%), or the stream was not wadeable (1%). We were unable to collect field data from another 10% of the mapped stream length due to lack of access (mostly landowner denials). Thus, the field data we collected at 509 sites allows inference to the ecological condition for 184,600 km of the mapped stream length in the region.

1. Introduction

Sample surveys have been used with great effectiveness in a variety of fields to describe the characteristics of populations that are too large to census efficiently. Best known are surveys associated with political polls (e.g., election polls). Although sample surveys have been used for many years in a variety of arenas, their use and acceptance for characterizing water resources has been limited. The approach is gaining attention, however. One major impetus was the implementation of the stream and lake surveys conducted by the EPA to characterize the regional status in stream and lake acid-base chemistry relative to acidic deposition (Kaufmann et al. 1991, Landers et al. 1988). With the creation of EMAP (Environmental Monitoring and Assessment Program), the EPA extended these concepts. EMAP is designed to answer questions about the current status, changes, and trends in indicators of ecological condition of the nation's ecosystems (Messer et al. 1991).

Environmental Monitoring and Assessment **63:** 95–113, 2000.
© 2000 *Kluwer Academic Publishers. Printed in the Netherlands.*

It has been estimated that there are at least 44,000 publicly owned lakes and 5.6 million km of rivers in the conterminous United States (USEPA 1994). Describing a national picture of the status of lakes and streams via a census of the resources clearly is not feasible. Indeed, it is not practical to census lakes and streams at even a regional or state level, particularly if regional or statewide trends in condition are of interest. Paulsen et al. (1998) gave examples from four aquatic case studies in which there were large discrepancies in regional assessments of condition between probability/randomized survey results and hand-picked site survey results. Thus, there is clear benefit to adopting sample survey techniques for assessing condition

Larsen et al. (1994) described a method for selecting a probability sample of lakes for regional surveys. Here we describe a process for selecting a probability sample of stream sites. This survey design is transferable to any stream monitoring project where the stream resource is too large to census, making a sample survey necessary. We first describe the important design criteria for such a survey. We then explain the mechanics of the method and follow with an illustration of how we selected a probability sample of stream sites in the mid-Atlantic region of the United States for an EMAP stream pilot survey. We conclude with an example of frequency distributions from data collected at those sites to illustrate the presentation of results of surveys of this nature.

2. Design Criteria

Forming the basis of our design choice are criteria that recognize: 1) the nature of stream systems, and 2) the goals for a national or regional scale monitoring program.

2.1 STREAMS AS A CONTINUOUS RESOURCE

Streams form a continuous network embedded in the watersheds they drain. That streams reflect the ecoregion within which they exist, the condition of the watersheds they drain and that the condition of the watersheds and that the character of a stream changes in a continuous fashion as water moves downstream are fundamental tenets of limnology (Vannote et al. 1980, Hynes 1975). A survey design failing to recognize or accommodate this fundamental nature of stream networks would be inadequate. However, the extent to which the condition of a particular point on a stream reflects the aggregate of the watershed it drains depends on the attribute of interest. Whereas the chemical quality of the water sampled at a point might integrate the entire watershed's influence, the species complex and stage of development of the riparian corridor at the same site might be much more locally determined. A sample design should be able to incorporate the varying impacts of

the range of local and distant influences on the variety of possible attributes ascribed to a sample location.

Although streams are continuous resources, they are often organized into discrete units (reaches) by confluences or abrupt shifts in geomorphic types (e.g., U.S. EPA's River Reach File, Horn and Grayman 1993). One thus has the choice of considering the stream resource as a discrete (finite) or continuous population. A substantial body of literature forms the statistical foundation for finite population sampling, which would be applicable if we treated the continuous stream resource as a discrete population (Cassel et al. 1977, Cochran 1977, Kish 1965). Recent advances extend this foundation to sampling continuous resources; consequently, stream networks can be treated as either discrete or continuous populations (Stevens 1997). This distinction is important, because many researchers tend to treat streams as finite populations by decomposing them into reach segments for site selection, but then treat them as continuous to summarize results.

Designing a survey from a discrete perspective leads to statistical issues, if the results are to be expressed from a continuous perspective. For a given sampling effort, the precision of the estimate of condition relative to length is lower than it would be if the design were based on a continuous perspective. The technical arguments supporting this contention are based on the need to convert inclusion probabilities derived from the discrete units into expansion factors appropriate for calculating length properties. The end result is that unnecessary variation in these occurs, leading to lower precision in estimates if length is the resource property to be characterized.

Treating streams as a discrete population in survey design carries some additional disadvantages. One is that most field sampling is performed at points in a stream network, especially sampling for water column and sediment conditions and generally for biological attributes. If a stream network is divided into discrete reaches, one is faced with the need to argue that the point samples are representative of the specified reach. This introduces some unnecessary assumptions that can be avoided by treating the stream resource as continuous. Point locations can be selected at which samples or measurements (within an associated support area) can be taken, and from the measurements at the set of selected points, inferences can be drawn about the continuous stream resource.

Our choice of representing streams as a continuous or finite population must also be shaped by the assessment questions that various agencies ask. The U.S. EPA, in its National Water Quality Inventory, has phrased its questions in terms of the length of stream resource in various conditions (e.g., how many miles of stream are impaired relative to aquatic life use). Taking all these issues into account, we chose to adopt a continuous perspective, recognizing that streams are continuous and reflecting the desire to characterize responses in terms of the proportion or length of the target stream network in one condition or another, rather than in terms of the number or percent of reaches or segments.

2.2 RANDOMIZATION IN SITE SELECTION

One of the foundations of a statistically defensible and scientifically sound sample survey is that randomization must be incorporated into the site selection process (Cochran 1977, Kish 1965). Randomization is necessary to avoid the biases, even unrecognized ones, inherent in selecting sample sites by other methods and to assure that the sample is representative of the targeted resource. An important factor related to randomization is that each element in the target population has a known, non-zero probability of being included in the sample, which allows the uncertainty in the population inferences arising from the selection process to be calculated.

2.3 SPATIAL BALANCE IN SITE SELECTION

As are most ecological resources, stream networks are spatially distributed across the landscape. Our objective was to achieve a sample that reflects the density and distribution of the stream resource while promoting spatial balance in the selected sample sites. A simple random sample of a stream network can produce a pattern that appears aggregated; although unintended, some parts of the network might be sampled with greater density than others (e.g., all sample sites could fall in one small corner of the study area by random chance). Incorporating a method that assures spatial balance, as well as randomization, can produce a relatively even distribution of sample points. This spatial balance represents the nature of a distributed resource such as stream networks better than would a strictly random sample.

2.4 BALANCING SAMPLE SIZE AMONG SUB-POPULATIONS

Incorporating the ability to allocate different sample sizes to different regional or stream type sub-populations would be desirable because of the varying interests among user groups. Although statements about the condition of the entire stream network are desirable, it may also be important to make separate statements about the different classes or types of streams or to make comparisons among various regional sub-populations. This criterion is especially germane for stream networks, because small headwater streams are widely distributed and make up most of the stream length in a region; a simple random sample will allocate most of the sampling effort to these small, but extensive, streams. A second concern is that it may be important to describe the condition of any of a number of different stream classes, or to make separate statements about certain areas, such as a political or physiographic region, or a drainage basin of special interest. A sampling design that permits differential allocation of sampling effort among various potential sub-populations is desirable.

This set comprises the primary criteria we sought to satisfy in designing stream surveys. In the next section we first describe the general features of the stream site selection process, then discuss some of the mechanics of the selection.

3. Selecting Stream Sites

3.1 EMAP'S BASIC DESIGN

Early in the design process we decided to link the stream sample selection to the primary design structure originally envisioned for EMAP (Overton et al. 1991, Stehman and Overton 1994, Stevens 1994). This structure consists of a randomly placed triangular grid of points (and their hexagonal tessellations) draped over the continental United States, fit within a global framework. The original design envisioned an array of approximately 12,500 points, with their corresponding 635 km^2 hexagons as the basic structure. The design allows for a finer or coarser grid, depending on desired sample size and spatial coverage, yet retains the basic triangular arrangement of the points and the surrounding hexagonal areas. The original design also incorporated a smaller 40 km^2 hexagonal search area associated with each grid point, on or within which the actual field sampling sites would be located and the measurements would be made on the variety of EMAP resource types. The use of such co-located sites across resource types would facilitate synthesis and integration of results.

Important reasons for constructing the search area included allowing for a two-stage sampling process and for the selection of sampling locations that were not exactly coincident with a grid point (e.g., it is unlikely that a grid point would fall exactly on a stream or lake). The first-stage sample is an area sample cut out by the 40 km^2 hexagonal search area. For a second-stage sample, relatively inexpensive measurements could be used to refine the location and types of measurements to be made. For streams, the first-stage sample consisted of the stream length determined by the search areas. The second-stage sampling process involves selecting a subset of the first-stage sample to be used as the sample survey sites for making regional inferences (e.g., sites where field measurements will be made). Data from maps, aerial or satellite imagery, or any other source that can be obtained for the entire stream length in the first-stage sample can be used to drive the site selection process during the second-stage sampling.

The original triangular grid design incorporated a four-year rotating panel sampling design. During year 1, a systematic subset of the grid would be sampled, during year 2, a second subset, and so on, until completion of a four-year cycle. Then in year 5, the year 1 sites would be revisited; likewise in year 6, the year 2 sites would be revisited, and so on, until a second four-year cycle was completed. The four-year cycle would continue as part of the basic monitoring program to allow for trend detection. This rotating pattern has been shown by Urquhart et al.

(1998) to be an effective approach for detecting regional trends as well as describing regional status. Slight modifications to the sampling plan could incorporate yearly repeat visits, as well as future substitution of sites, if sites would "wear out" with repeated sampling.

3.2 THE TARGET POPULATION AND FRAME

Two primary features of a sample survey are: 1) to explicitly identify the target population about which the descriptions will be made, and 2) to identify a frame or frame materials used to represent the target population and from which the sample sites will be drawn (Cochran 1977). We chose a digital frame that had sufficient resolution for small wadeable streams and was relevant to other assessment or monitoring projects. The digital versions of the USGS 1:100,000-scale hydrographic maps, as modified by the U.S. EPA in the creation of the River Reach File (Horn and Grayman 1993), have been chosen by the U.S. EPA and USGS as the basis to represent the aggregate of the US stream resource and was our choice for a sample frame. We used Version 3 of the River Reach File, which is known by the acronym RF3. These digital coverages contain traces of stream resources covering the entire conterminous United States, from the largest rivers to a substantial portion of the smallest tributaries. These coverages represent over 2 million km of flowing waters. The use of geographic information systems (GIS) allows us to manipulate this database in a variety of ways to identify various potential regional target populations, for example, by ecoregion, state, hydrographic unit, or other spatial unit. Furthermore, RF3 contains some attribute information, such as network connectivity, and a separation of large and small rivers that allows some refinement of the spatial target populations. We added Strahler stream order as an additional attribute.

The broad coverage of stream resources, available electronically, was a prime consideration in the selection of RF3 as the frame. A digitized sample frame was highly desirable so that it could be combined with the EMAP hexagons electronically in a GIS. This combination allows sample selection for surveys at numerous scales, from national to very local, such as within a particular watershed, state, or other desired geographic area. In addition, it permits development of routine protocols for site selection, so that, once the objectives of a particular survey have been developed, sub-populations identified, and sample sizes chosen, relatively little effort is required to select the actual sample site location (i.e., latitude and longitude of the locations comprising the probability sample).

Our experience in the mid-Atlantic has been that large streams and rivers, and the headward extent of the main channels are well represented in RF3. However, despite the fact that the total length of first-order streams accounts for about 60–70% of the total stream length in most areas we have investigated, we've noticed that many first-order side tributaries are absent when the coverages are compared with stream networks delineated on the USGS 1:24,000-scale topographic

maps. We have also discovered some inconsistencies in stream density among adjoining map panels. However, because RF3 is electronic, in those areas where exclusion of stream resources is problematic, the electronic files can be augmented by digitizing the missing information from the necessary source information.

3.3 THE SITE SELECTION PROCESS

Key elements of the stream site selection process are:

- a continuous sampling perspective as the foundation for the site selection process (i.e., selection of points on a stream network rather than discrete segments);
- randomization at various stages of the site selection process;
- two-stage site selection, based on a point grid with an associated search area;
- opportunity to vary sampling density by features in the sampling frame, and
- a digital hydrographic layer as the frame from which the sample is drawn.

The mechanics of sample selection occur in two stages, as outlined in the following subsections. However, before performing the actual sample selection, it is important to explicitly define the target population, identify some possible sub-populations one might wish to describe and compare, and finally, determine the sample size.

3.3.1 *The First-Stage Sample*

The first-stage sample consists of the intersection of the 40 km^2 EMAP hexagon (referred to as a 40-hex from now on) grid search area with RF3 (Figure 1). A GIS coverage is prepared for both in the target region of interest. Their intersection creates an area sample consisting of all the RF3 stream traces identified by the 40-hex search area. The aggregate length of the stream trace in this set of 40-hexes is the first-stage sample. This sample can be characterized in a variety of ways to refine the second-stage sample. As mentioned earlier, some information about the stream traces is available in RF3; initially, we calculated the relative proportion of stream length as intermittent and perennial (as coded in the frame), and the relative length in each of the Strahler stream order classes. More complex classifications can be envisioned, such as proportions in various land types or ecoregions, or riparian condition. Characterization of the first-stage sample, however, can be resource intensive because of the requirement that the entire first-stage sample be evaluated. How this information would be used for refining the second-stage sample should be evaluated carefully before embarking on the stage 1 characterization.

3.3.2 *The Second-Stage Sample*

The second-stage sample is selected conceptually by straightening out and lining up all the first-stage stream traces in a random order along a long line and then systemati-

cally locating sampling points on the line at equal intervals, after a random start. The distance between points is determined by dividing the length of the line by the desired sample size. This task is accomplished mechanically by linking all the first-stage segments electronically into one long trace. The traces are linked first within each 40-hex, and then across all 40-hexes. At this stage, two principal criteria guide the selection process, an even spatial spread and randomization in the site selection. To achieve an even spatial spread to the sites in the study area, the 40-hexes are grouped into spatial clusters, such that each cluster contains approximately the same stream length. The clustering assures that samples are spread out evenly across the region of interest. A similar process was used in the selection of lakes for the lake pilot survey (Larsen et. al. 1994).

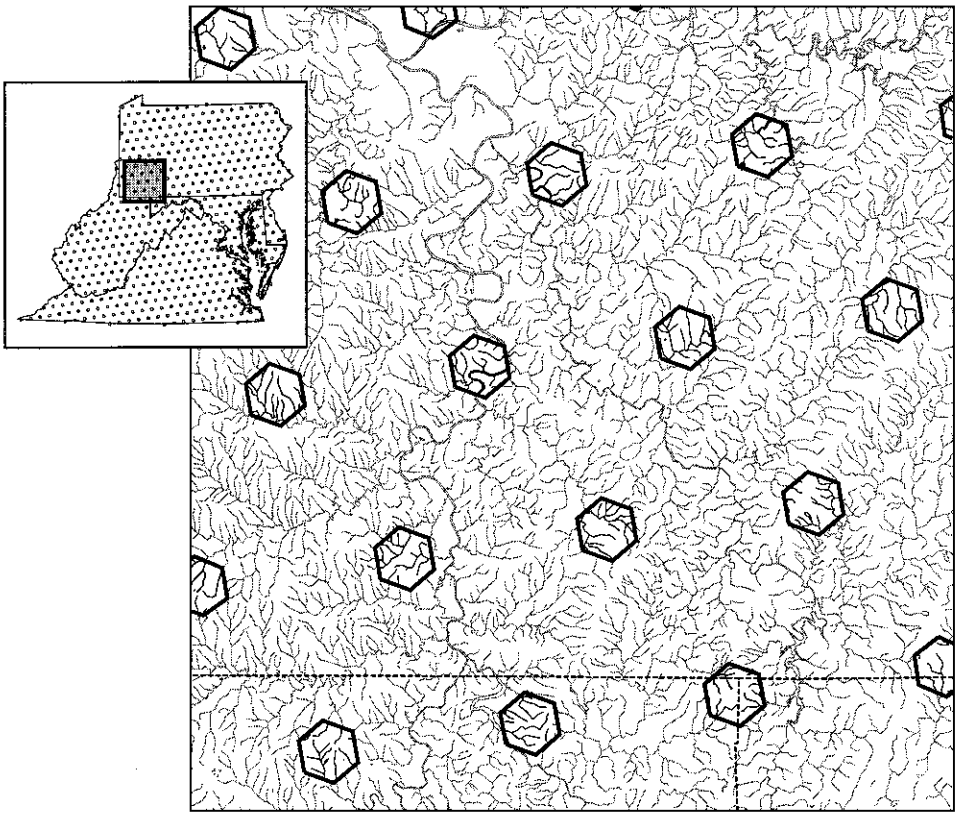

Figure 1. *The stage 1 selection process in the mid-Atlantic. The small inset map shows the distribution of the 40-hex grid across the five mid-Atlantic states in the study region. The larger map is an enlargement showing the 40-hexes and RF3 stream traces. Stream traces that fall within the hexagon search area are "clipped out" and used in the stage 2 selection process for picking sample sites.*

Randomization in the second-stage sample is accomplished at several levels in the selection process. The digital database stores stream traces as stream fragments, identified by stream order. Within each 40-hex, these fragments are randomly ordered, with the exception that fragments from the same continuous stream with the same order are joined together before randomizing; the resultant unit is randomly ordered. This first step creates a single electronic stream trace for each 40-hex, consisting of a random arrangement of fragments. Then 40-hexes within clusters are arranged in random order, and finally, the clusters are arranged in random order. The resultant single, long stream trace is used as the base for choosing the systematic locations of points, after a random start. Because the stream traces are georeferenced, the geographic coordinates of the sample points can be retrieved and plotted on an ordinary map and ultimately located in the field.

The general procedure just outlined is relatively flexible and can be modified in a variety of ways while retaining the basic principles. One common modification we use is to allocate variable sampling probability to different stream sizes. For example, using the above procedure, a simple random sample of streams will select sites in proportion to the abundance of their type in the target population. For streams, this means that 60–70% of the selected sample sites would be in first-order (headwater) segments. The different stream orders, however, can be selected with variable probability to adjust the allocation of sites among the order classes. This selection is accomplished by electronically stretching the stream traces for which differential sampling is desired. For example, suppose we wished to allocate approximately 50% of our samples to first-order streams and 50% to second- and third-order streams, and on average, first-order streams accounted for 75% of the target stream length. Even allocation could be achieved by tripling the length of the second- and third-order streams, yielding an apparent second- and third-order stream length equivalent to the first-order length. Applying the selection rules described above would allocate half the sampling effort to the first-order streams and half the effort to the second- and third-order streams. In the analysis stage, different weights are then applied to first- versus second- and third-order streams in making regional inferences. The variable probability scheme can be applied to any stream classification system (e.g., ecoregion, state) that is identified for every stream trace in the first-stage sample.

A second type of modification we often use is to intensify the sample selection for some regional sub-populations by increasing the number of grid points and corresponding search areas; the size of the search hexagon can be modified as well. For example, one might want to characterize a particular watershed or ecoregion separately from the overall regional population within which the sub-population occurs. In the EMAP mid-Atlantic case study described in the next section, we selected more stream sites in areas known to be sensitive to acidic precipitation, to meet requirements for monitoring effects associated with atmospheric sulfur and nitrogen emissions reductions directed by the Clean Air Act.

The outcome of the two-stage site selection process is a set of field sample points (with their mapped locations and latitude/longitude) that constitutes the stream survey sample. In the field, sample sites were identified to within 100 m, using topographic map orienteering and global positioning system (GPS) devices. Measurements at these sample locations form the basis for making inferences about the target stream population. At the field sample points, some field measurements can be made, for example, withdrawing a water sample for chemical analyses. However, point samples make little sense for some attributes, such as measurements of fish assemblages at a single point. In these cases, a *support area*, or sample reach, is established around the selected point, within which measurements or collections are made, representing the sample point. Also, since each point on a stream can be thought of as the outlet of the drainage basin upstream of that point, each sample point can be characterized by its watershed as well. For example, watershed attributes associated with sample points can include land use/land cover information (e.g., percent of the watershed above the sample point that is forested or clearcut), population or road densities, presence of point source discharges, or other watershed attributes of interest. These watershed measures become attributes of the selected sample point, and population inferences can be made relative to these watershed attributes in a manner analogous to making inferences based on site measurements.

Key to making population inferences from the statistical samples is using the sample site expansion factors (also known as weights). As noted earlier, the selection process allows any location represented on the frame to be included in the sample with known probability. The selected sample size and the allocation of samples to various target sub-populations in relation to the total population controls the chance that any particular location will be chosen. In a simple random sample consisting of 50 sites along 1,000 km of stream length, each site has an inclusion probability of 50/1,000, or .05, resulting in a sample weight of 20 km (the sample weight is the inverse of the inclusion probability). In other words, each site "represents" 20 km of stream length. If 30% of these sites were characterized as degraded, we would infer that 300 km of stream length would be in degraded condition. This approach for attributing "length" to a sample point is rigorous and consistent. It removes much of the subjectivity that historically has gone into these decisions by states in preparing their water quality reports to U.S. EPA. Differentially varying the sample size in relation to potential target sub-populations, or in relation to stream order, produces different inclusion probabilities or weights for the respective sample units. These inclusion probabilities are attributes of the sample sites, just as are the field measurements. The sample weight for each site must be used in order to make valid population inferences from the sample data.

4. EMAP Mid-Atlantic Stream Sample: A Case Study

To illustrate the sample design, we describe in detail the site selection process for the base EMAP sample for the first year of the mid-Atlantic pilot study (1993). The initial step in any sample survey requires a variety of decisions about target populations and sample sizes. The survey often is designed to meet several objectives; thus, the general design framework should be flexible to accommodate this diversity. The design issues tend to revolve around the selection of the target areas and the stream sub-populations for which estimates and comparisons are to be made. The case study used for illustration is no exception, and the following description reflects the kinds of decisions required before the sample sites can be selected.

EMAP identified the mid-Atlantic United States as the region within which to conduct a pilot stream sample survey. The area of interest covers the states of Pennsylvania, Maryland, Delaware, Virginia, and West Virginia, and the Catskill Mountain area of southeastern New York. The survey was to focus on wadeable streams across the region.

At the time we designed the EMAP pilot study, we were also considering conducting a national stream survey whose goal would be to sample approximately 800 stream sites, allocated in such a manner that separate statements could be made about each of the 10 Standard Federal Regions in the conterminous United States. For the mid-Atlantic region (corresponding to Region 3), we assigned 100 sites, representing a 25% increase over an even allocation of the 800 sites to the 10 Regions, to accommodate non-targets, non-permission sites, etc. The 100 sites spread across Region 3 comprised the base EMAP sample. The decisions we made prior to selecting the sample included:

- defining the target population as first-, second-, and third-order streams (what we considered to be wadeable streams) in the region, using RF3 as the sample frame, and the basic EMAP design as the sampling framework.
- allocating sample size in the ratio of 2:1:1 for first:second:third order streams. The goal was to sample approximately equal number of sites on first-, second-, and third-order streams. Our initial guess was that approximately 50% of the first-order streams would be dry or otherwise non-target. Thus, we over-selected first-order streams to compensate for the non-target types of stream traces.

As described earlier, the site selection process occurred in two stages. The first-stage sample consisted of the intersection of search 40-hexes with the RF3 coverage, retaining only the first- through third-order streams identified in RF3 that were captured by the 40-hexes (Figure 1). The total stage 1 sample corresponds to a $^1/_{16}$ area sample evenly spread across the mid-Atlantic region (any site

has a $^1/_{16}$ chance of being included). In 1993, we followed the 4-year EMAP rotating hexagon design and used only $^1/_4$ of the 40-hexes in stage 1 sampling, so that the probability of any stream trace being included in the 1993 stage 1 sample was $^1/_{64}$ or .015625. For stage 2 sampling, stream fragments were randomized and linked, as described in section 3.3.2. To get the desired sample allocation among order classes, we used variable probabilities by stretching the sample line for second- and third-order stream traces twofold and fourfold, respectively.

This series of steps produced a single electronic stream trace, reflecting the desired spatial spread, randomization, and multipliers. As a final series of steps, we divided the length of the stretched sample line by the target sample size of 100 for the base EMAP sample to create an interval length (Table I). For the 1993 base EMAP sample, the interval length was 50.9024 km. Beginning at the origin of the trace, the first site was located randomly within the first interval; then each successive site was located by moving down the trace a distance equal to the sampling interval (50.9024 km in this case). The final result of the series of selection steps was a sample consisting of 100 sites spread across the region (Figure 2). Associated

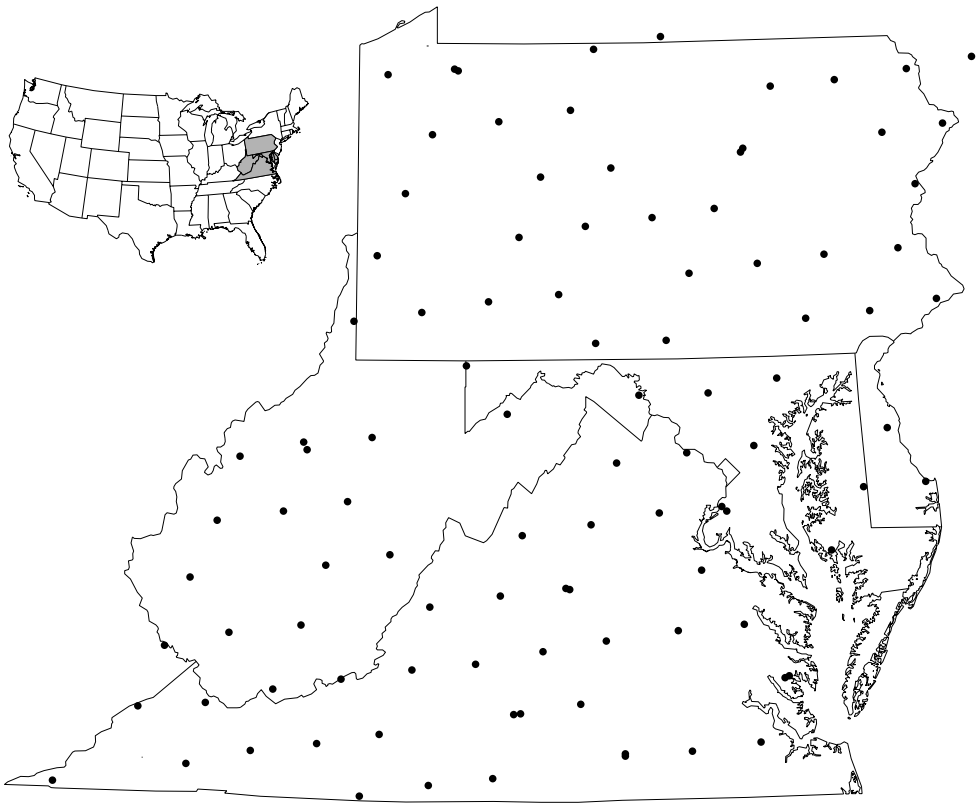

Figure 2. *Location of the 100 probability sample sites in the base 1993 EMAP survey in the mid-Atlantic.*

with each site is its respective weight or expansion factor (Table I). As variable probability was used for each order class, the different orders have different weights. Details of the weight calculation are given in Table I.

Table I

Stream length and design information used to calculate the sample expansion factors (weights) in the 1993 base EMAP probability survey; sample size was 100 sites.

Order	Stage 1 Length (km)	Stage 2 Var. Prob. Multiplier	Multiplied Line Length	Stage 2[a] Inclusion Probability	Expansion[b] Factor (km)
1	2544.670	1	2544.670	0.019645	3257.754
2	609.391	2	1218.782	0.039291	1628.878
3	331.697	4	1326.788	0.078582	814.438
ALL	3485.758		5090.240		

[a]Stage 2 inclusion probability = (sample size * stage 2 multiplier) / total multiplied line length.

[b]Expansion factor (or weight) = 1/(stage 1 inclusion probability * stage 2 inclusion probability). The stage 1 inclusion probability for the 1993 EMAP base survey was $^1/_{64}$ or 0.015625 (stream traces were clipped out of $^1/_{64}$ of the land area in the region).

4.1 INTENSIFYING THE SAMPLE DENSITY

In addition to the base EMAP sample, we selected a more spatially dense set of sample sites for regional acidic deposition stream monitoring and specific EPA regional efforts in the study area. One option would have been to identify a shorter sampling interval on the original stage 2 sample stream line for those sections of the stream line within the regions of interest. However, this would have concentrated the sampling in the vicinity of the base grid point and caused multiple sample sites within each 40-hex, contradicting our desire to spread the samples out. Instead, we identified a set of additional hexagons affiliated with each base 40-hex by increasing the density of the grid sevenfold; this step produced an additional 6 grid points (Figure 3). Around each of the larger, original 40 km^2 hexes, we used 6 smaller hexagons (13 km^2) to extract the stream traces for the stage 1 intensified sample (Figure 3).

The subsequent stage 2 process used in picking the sample sites was identical to the one that produced the base 100 sites, except that we wanted 150 additional sites. As we wanted intensified sites only in certain areas of the region, we only clipped out stream traces for the stage 1 sample that were within the intensified study area. Both the base sample and the intensified sample can be used independently to make population inferences. It is desirable to combine them, however, to make more precise estimates due to the larger combined sample size. The combined inclusion probability is calculated as the sum of the base and intensified inclusion probabilities. Within the intensified areas, a new combined weight (equal to 1/combined inclusion probability) was necessary because both the base and the intensified sampling offered chances to select sites. Outside the intensified study

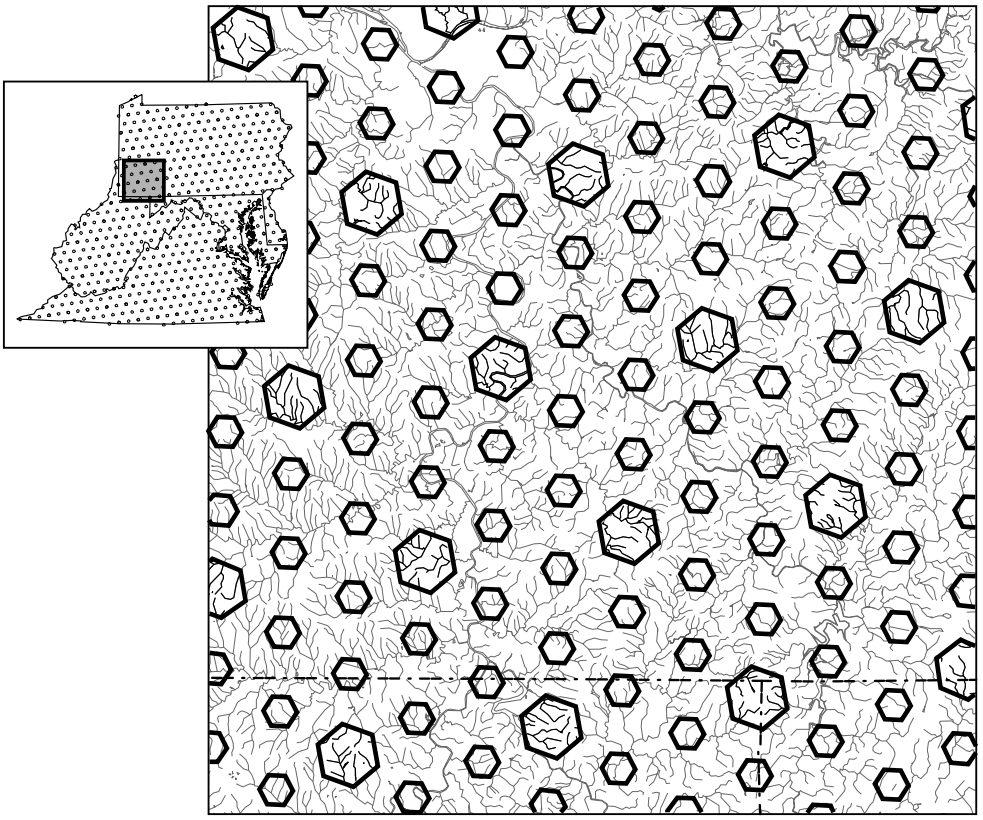

Figure 3. Illustration of the increased areal coverage in the stage 1 selection process using a finer scale hexagon grid density.

area, the combined weight was the same as the base survey weight because the intensified inclusion probability was zero.

4.2 EXAMPLE STREAM POPULATION INFERENCES

In addition to the 1993 EMAP pilot survey and the intensified surveys, we carried out similarly designed surveys in 1994 and 1995 in the mid-Atlantic. These survey results can be combined into one regional estimate by adjusting the inclusion probabilities. Note that all regional estimates presented here are weighted and thus represent an estimate of the condition of the length of stream resource in the study area, not the frequency distribution of the collected samples. In all, 615 sites were selected for field visitation during spring baseflow (May–June), 1993–1995. Important components of field sampling included conducting reconnaissance to de-

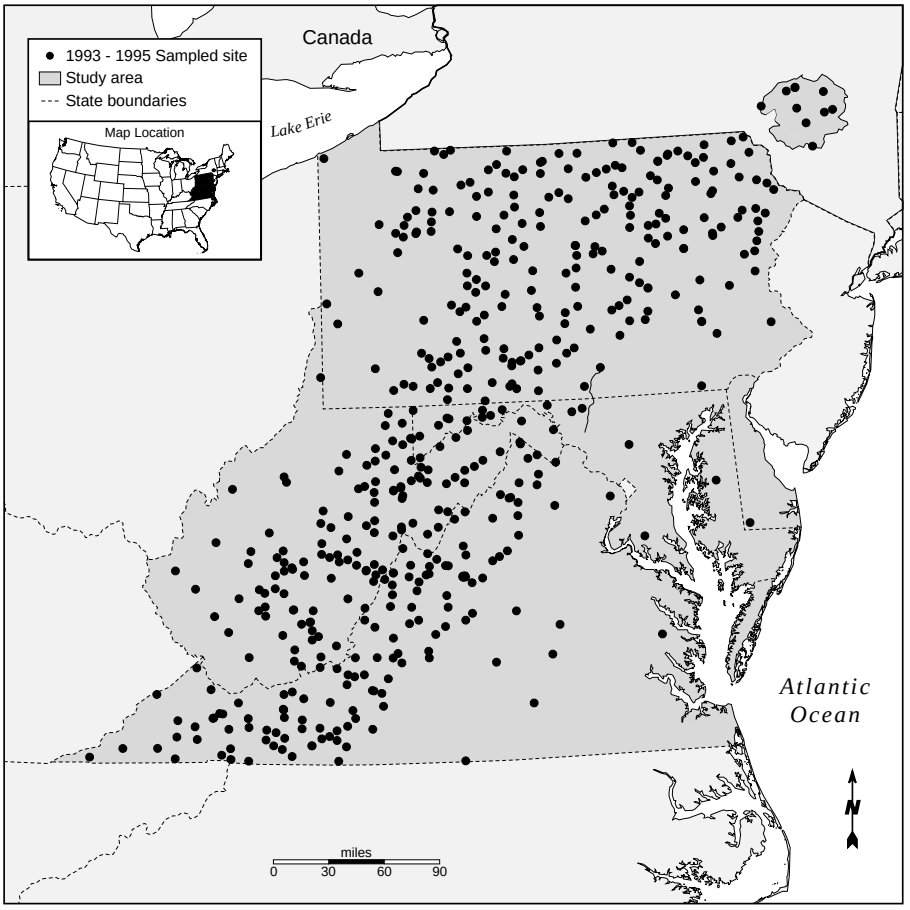

Figure 4. *Location of all 509 sites where field sample data was collected in the 1993–1995 mid-Atlantic EMAP probability surveys.*

termine whether the locations were truly streams (sometimes map errors identify streams where no streams exist), and obtaining site access permission. The result of this screening process was a summary of population stream length represented by the sample frame, classified into different target status categories (Table II). Of the 230,400 km of first- through third-order streams represented in RF3 in the mid-Atlantic, we were able to collect stream ecology data from 509 sites (Figure 4) representing 80% of the length. The remainder of the length could not be sampled due to: 1) inability to access the site (10%; primarily land owner permission denials), 2) dry stream channels (5%), 3) non-wadeable streams (1%), or 4) no stream channel present (4%). Thus, our sample data can be used to draw inferences to 184,600 km of wadeable stream in the mid-Atlantic region (Table II).

From the field measurements, we can draw inferences to the full length of streams represented by this set of sites. To illustrate this process, we selected a

Table II
Estimated status of stream length in the RF3 sample frame in the mid-Atlantic by stream order inferred from 1993-1995 EMAP probability survey data.

Order	Length (km)	% Sampled	% Dry	% Non-wadeable	% No Channel	% No Access
1	172,000	79	6	0	4	11
2	33,200	81	1	6	1	11
3	25,000	85	0	4	7	4
ALL	230,400	80	5	1	4	10

chemical indicator (total phosphorus) for which measurements were made during all field visits (sampling protocols are summarized in Lazorchak et al. 1998). We also obtained information about the immediate watershed associated with each sample point. Examples of the relationships between watershed condition and stream chemistry are given in Herlihy et al. (1998). The database from which we make the population inferences consists of site identifying information (name, latitude, longitude), expansion factor (km), and an array of environmental data affiliated with the site. The intent is to describe the characteristics of the stream population, and the usual form is a frequency distribution of sample scores weighted by the appropriate expansion factors, expressed as a cumulative distribution function (Figure 5). For total phosphorus, 84% of the stream length had concentrations less than 50 μg/L. The 90% confidence bounds around that estimate were plus or minus 9 percentage points. In other words, we're 90% confident that between 75% and 93% of the stream length in the region had total phosphorus con-

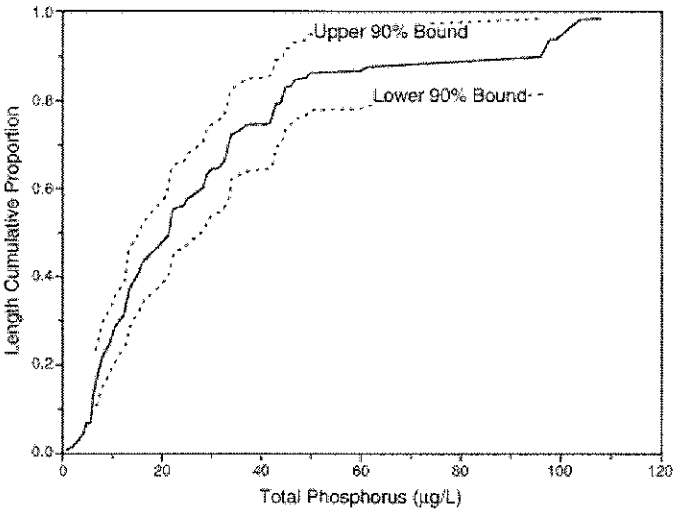

Figure 5. Cumulative frequency distribution of total phosphorus in mid-Atlantic wadeable streams. These frequency distributions can be used to estimate the proportion of the stream resource that does - n't achieve, or that exceeds, particular levels. The solid line is the population estimate and the dotted lines are the upper and lower 90% confidence bounds.

centrations less than 50 μg/L. Confidence bounds on the length estimates were calculated using the Horvitz-Thompson estimation procedure (Cochran 1977).

With adequate sample sizes, it is also possible to infer the condition of any sub-population in the study area. Sub-populations can be political (states, counties), physiographic (coastal plain, ecoregions), drainage basins (Ohio, Potomac), or any other definable group (e.g., streams over 2,000 feet elevation). Sub-populations can be defined both before and after survey design. For example, a sub-population of streams with pH < 6 could be created and reported on. To make sub-population estimates, only sites in the defined sub-population are used in the analysis. A general rule of thumb is that a sub-population sample size of 30–50 produces a reasonable estimate of a frequency distribution; more sites are better, of course, and as a general rule, the precision of the estimates doubles with a fourfold increase in sample size. With our mid-Atlantic data, for example, we could construct state-level estimates of condition (Figure 6). In this example, stream length in the state of Pennsylvania tended to have higher total phosphorus concentrations than West Virginia.

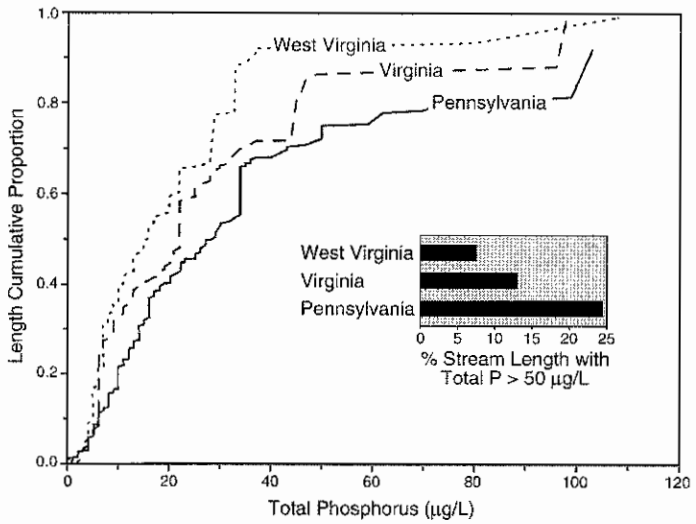

Figure 6. *Cumulative frequency distributions of total phosphorus in three mid-Atlantic states. Error bounds for each state are not shown but were in the range of ± 10–18 percentage points. The bar diagram inset shows another way this type of data can be displayed.*

The examples are meant as simple illustrations of the cumulative frequency distributions that form the basic data summaries for sample surveys. Usually, site visits and watershed compilations create a large matrix of site scores covering the various biotic, chemical, physical, riparian, and watershed characterizations forming the basis of extensive analyses of interrelationships; here we intend to describe

the site selection process, not report findings on all of these variables. However, gathering this type of information on a set of sites selected to represent specific target populations allows the investigator to make inferences about the resource as a whole. There is also a basis for believing inferences derived from these characteristics and associations are applicable across the population. The basis for making these population and sub-population inferences relies on the selection of a representative sample (i.e., an unbiased sample).

5. Summary

A probability-based survey is a useful tool for estimating population attributes when the population is too large to census. Thus, a probability survey is a good way to design stream surveys when the objective is to estimate condition over large spatial areas. We describe a two-stage sampling process for picking a spatially balanced, randomized set of stream monitoring sites for estimating regional condition. As part of the stream pilot activities for EPA EMAP, we conducted a survey of wadeable streams in the mid-Atlantic region of the United States. The design utilized the digitized stream network taken from 1:100,000-scale USGS topographic maps as the sample frame. In the first stage, first- through third-order (wadeable) stream traces that fell within the EMAP hexagon search area in the region were clipped for the next stage of analysis. In stage 2, the clipped traces were randomly laid out in a line and sampled at a fixed interval, after a random start, to pick the sample sites.

During 3 years of sampling in the mid-Atlantic, 615 stream sites were selected. The sample design allows inference from the sample data to the status of the entire 230,400 km of wadeable stream length in the study area. Field sample data were successfully collected from 509 sites representing 80%, or 184,600 km, of the mapped stream length. Ten percent of the length in the sample frame was non-target (no stream channel present, a dry channel, or not wadeable). In another 10% of the frame length, we were unable to visit the site due to lack of access (land owner permission denials or inaccessibility). Key to making these population inferences is the use of the sample site expansion factors (weights) derived from the inclusion probabilities set during the survey design.

Acknowledgments

The research described in this paper has been funded by the U.S. Environmental Protection Agency in part through a cooperative agreement with Oregon State University (#CR824682) and a contract with OAO Corporation (68-W5-0065). It has been subjected to the Agency's peer and administrative review and approved for publication. Mention of trade names or commercial products does not consti-

tute endorsement or recommendation for use. We thank all the EMAP team for sharing design thoughts and Ian Waite and Lou Seivard for helpful comments on an earlier draft of this manuscript.

References

Cassel, C.M., Sarndal, C.E. and Wretman, J.H.: 1977, *Foundations of Inference in Survey Sampling*, John Wiley & Sons, New York, 192 pp.

Cochran, W.G.: 1977, *Sampling Techniques*, 3rd ed., John Wiley & Sons, New York, 428 pp.

Herlihy, A.T., Stoddard, J.L. and Burch-Johnson, C.: 1998, 'The Relationship Between Stream Chemistry and Watershed Land-Cover Data in the Mid-Atlantic Region, U.S.', *Water Air Soil Pollut.* **105**, 377–386.

Horn, C.R. and Grayman, W.M.: 1993, 'Water-quality modeling with the EPA reach file system', *J. Water Resources Plan. Manage.* **119**, 262–274

Hynes, H.B.N.: 1975, 'Edgardo Baldi Memorial Lecture: The Stream and Its Valley', *Verh. Inernat. Verein. Limnol.* **19**, 1.

Kaufmann, P.R., Herlihy, A.T., Mitch, M.E., Messer, J.J. and Overton, W.S.: 1991, 'Chemical Characteristics of Streams in the Eastern United States: I. Synoptic Survey Design, Acid-Base Status and Regional Chemical Patterns', *Water Resour. Res.* **27**, 611–627.

Kish, L.: 1965, *Survey Sampling*, John Wiley & Sons, New York, 643 pp.

Larsen, D.P., Thornton, K.W., Urquhart, N.S. and Paulsen, S.G: 1994, 'The Role of Sample Surveys for Monitoring the Condition of the Nation's Lakes', *Environ. Monit. Assess.* **32**, 101–134.

Landers, D.H., Overton, W.S., Linthurst, R.A. and Brakke, D.F.: 1988, 'Eastern Lake Survey: Regional Estimates of Lake Chemistry', *Environ. Sci. Technol.* **22**, 128–135.

Lazorchak, J.L., Klemm, D.J. and Peck, D.V. (eds.): 1998, *Environmental Monitoring and Assessment Program – Surface Waters: Field Operations and Methods for Measuring the Ecological Condition of Wadeable Streams*, EPA/620/R-94/004F, U.S. EPA, Washington DC.

Messer, J.J., Linthurst, R.A. and Overton, W.S.: 1991, 'An EPA Program for Monitoring Ecological Status and Trends', *Environ. Monit. Assess.* **17**, 67–78.

Overton, W.S.,White, D. and Stevens, D.L., Jr.: 1991, *Design Report for EMAP, the Environmental Monitoring and Assessment Program*, EPA/600/3-91/053, U.S. EPA, Washington DC.

Paulsen, S.G., Hughes, R.M. and Larsen, D.P.: 1998, 'Critical Elements in Describing and Understanding Our Nation's Aquatic Resources', *J. Amer. Water Works Assoc.* **34**, 995–1005.

Stehman, S.V. and Overton, W.S.: 1994, 'Environmental Sampling and Monitoring', pp. 263–305 in G.P. Patil and C.R. Rao (eds.), *Handbook of Statistics 12*, Elsevier Science, Amsterdam.

Stevens, D.L., Jr.: 1994, 'Implementation of a National Environmental Monitoring Program', *J. Environ. Manage.* **42**, 1–29.

Stevens, D.L., Jr.: 1997, 'Variable Density Grid-Based Sampling Designs for Continuous Spatial Populations', *Environmetrics* **8**, 167–195.

Urquhart, N. S., Paulsen S.G. and Larsen, D.P.: 1998, 'Monitoring for Regional and Policy-Relevant Trends Over Time', *Ecol. Applic.* **8**, 246–257.

USEPA: 1994, *National Water Quality Inventory: 1992 Report to Congress*, EPA/841/R-94/001, Office of Water, Washington DC.

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R. and Cushing, C.E.: 1980, 'The River Continuum Concept', *Can. J. Fish. Aq. Sci.* **37**, 130–137.