Information Collection Request for the 2020 Drinking Water Infrastructure Needs Survey and Assessment (DWINSA)

PART B OF THE SUPPORTING STATEMENT (FOR STATISTICAL SURVEYS

Prepared for:

U.S. Environmental Protection Agency

Office of Ground Water and Drinking Water

Drinking Water Protection Division

Table of Contents

[2020 STATE DWINSA 5](#_Toc44571487)

[B.1 SURVEY OBJECTIVES, KEY VARIABLES AND OTHER PRELIMINARIES 5](#_Toc44571488)

[B.1.a Survey Objectives 5](#_Toc44571489)

[B.1.b Key Variables 6](#_Toc44571490)

[B.1.c Statistical Approach 6](#_Toc44571491)

[B.1.d Feasibility 8](#_Toc44571492)

[B.2 SURVEY DESIGN 10](#_Toc44571493)

[B.2.a Target Population and Coverage 11](#_Toc44571494)

[B.2.b Sample Design 11](#_Toc44571495)

[B.2.c Precision Requirements 19](#_Toc44571496)

[B.2.d Data Collection Instrument Design 22](#_Toc44571497)

[B.3 PRE-TESTS AND PILOT TEST 23](#_Toc44571498)

[B.3.a Pre-tests 23](#_Toc44571499)

[B.3.b Pilot Test 23](#_Toc44571500)

[B.4 COLLECTION METHODS AND FOLLOW-UP 24](#_Toc44571501)

[B.4.a Collection Method 24](#_Toc44571502)

[B.4.b Survey Response and Follow-up 24](#_Toc44571503)

[B.5 ANALYZING AND REPORTING SURVEY RESULTS 24](#_Toc44571504)

[B.5.a Data Preparation 24](#_Toc44571505)

[B.5.b Analysis 25](#_Toc44571506)

[B.5.c Reporting Results 25](#_Toc44571507)

[2020 NATIVE AMERICAN DWINSA 26](#_Toc44571508)

[Introduction 26](#_Toc44571509)

[B.1 SURVEY OBJECTIVES, KEY VARIABLES AND OTHER PRELIMINARIES 26](#_Toc44571510)

[B.1.a Survey Objectives 26](#_Toc44571511)

[B.1.b Key Variables 26](#_Toc44571512)

[B.1.c Statistical Approach 26](#_Toc44571513)

[B.1.d Feasibility 27](#_Toc44571514)

[B.2 SURVEY DESIGN 27](#_Toc44571515)

[B.2.a Target Population and Coverage 27](#_Toc44571516)

[B.2.b Sample Design 27](#_Toc44571517)

[B.2.c Precision Requirements 28](#_Toc44571518)

[B.2.d Data Collection Instrument Design 29](#_Toc44571519)

[B.3 PRE-TESTS AND PILOT TEST 29](#_Toc44571520)

[B.3.a Pre-tests 29](#_Toc44571521)

[B.3.b Pilot Test 30](#_Toc44571522)

[B.4 COLLECTION METHODS AND FOLLOW-UP 30](#_Toc44571523)

[B.4.a Collection Method 30](#_Toc44571524)

[B.4.b Survey Response and Follow-up 30](#_Toc44571525)

[B.5 ANALYZING AND REPORTING SURVEY RESULTS 30](#_Toc44571526)

[B.5.a Data Preparation 30](#_Toc44571527)

[B.5.b Analysis 30](#_Toc44571528)

[B.5.c Reporting Results 31](#_Toc44571529)

List of Appendices

Appendix A – Public Notice Required Prior to ICR Submission to OMB (See Separate Document)

Appendix B – Data Collection Instrument and Lists of Codes (See Separate Document)

Appendix C – Comments and Response to Comments Received on the 2020 DWINSA Peer Review (See Separate Document)

Appendix D – Comments and Response to Comments Received on the First Federal Register (See Separate Document)

|  |
| --- |
| INTRODUCTION TO PART B  The Environmental Protection Agency (EPA) will conduct the following type of statistical survey for the 2020 State Drinking Water Infrastructure Needs Survey and Assessment (DWINSA). EPA will undertake an assessment of community water systems (CWSs) serving populations of more than 3,300. EPA will use the same methodology for collecting data for CWSs serving more than 3,300 persons as was used in the 2015 DWINSA. EPA will use this same approach to collect data for not-for-profit noncommunity water systems (NPNCWSs) serving more than 10,000 persons. EPA is proposing the same approach used in 2007 to collect data from CWSs serving 3,300 or fewer persons and the same approach used in 1999 to collect data from NPNCWSs serving 10,000 or fewer persons. EPA will send site visitors to collect data from CWSs serving 3,300 or fewer persons and to NPNCWSs serving 10,000 or fewer persons.  For the 2020 Native American DWINSA, EPA proposes a national sample for American Indian systems and a separate sample for Alaska Native Village water systems. EPA also proposes EPA Regions and the Navajo Nation collect the information for these systems. |

# 2020 STATE DWINSA

## B.1 SURVEY OBJECTIVES, KEY VARIABLES AND OTHER PRELIMINARIES

### B.1.a Survey Objectives

The primary objective of the 2020 State DWINSA is to collect information from water systems on the infrastructure they need to continue to provide safe drinking water to consumers. These data are used to produce a national estimate as well as state-specific estimates of water systems’ 20-year need. In addition, EPA will collect information through new questions in the data collection instrument. As previously mentioned in Section A, LSL questions are mandated by AWIA Section 2015(e)(2), the I&S questions will provide iron and steel construction material information to aid EPA in management of the AIS requirements under the SDWA section 1452(a)(4), and the OpW questions will provide important information on current and anticipated drinking water treatment and distribution system operator staffing concerns. EPA has established policies to ensure that the overarching goals of the survey are met:

* Estimate the total national 20-year need.
* Estimate the total 20-year need for medium and large CWSs for each fully participating state.
* Collect data on lead service lines nation-wide.
* Collect data on operator workforce concerns.
* Estimate the 20-year demand for iron or steel represented by DWINSA projects.
* Provide complete and accurate data to Congress.
* Provide a tool to fairly distribute DWSRF capitalization funds to states.
* Maintain the credibility of the DWINSA findings.

EPA proposes to collect information on the cost of water systems’ infrastructure needs. If cost data are not available from systems, EPA proposes to collect information that will enable the Agency to model costs. In the data collection instrument, the respondent will identify needs on a project-by-project basis and list the “type(s) of need” that the project will meet. The “types of need” include raw water source, transmission, source water treatment, storage, distribution, pumping stations, and other needs.

EPA will use the information from the DWINSA to estimate capital investment needs of drinking water systems. The information will be used to allot DWSRF monies among states.

For the 2020 DWINSA, EPA is proposing to use a modified panel approach to select survey respondents. The modified panel approach will involve dropping a random selection of 25 percent of the systems serving 3,301 to 100,000 persons that participated in the 2015 DWINSA and then drawing a random sample to replace those systems in the survey for the 2020 DWINSA. This will be done for each state and by strata. By primarily using information from the 2015 DWINSA, this approach would reduce the amount of time needed to prepare and review the responses from systems resurveyed in 2020. This approach was used for the 2015 State DWINSA.

For the new systems selected for the 2020 State DWINSA, EPA will use the same methodology as used in previous DWINSAs. The sampling design is discussed in detail below. As previously discussed in Part A of this ICR, EPA will administer new questions in addition to the 2020 State DWINSA: the LSL, OpW, and I&S questions. There are no separate precision targets for these questions, so EPA will administer these questions to the same systems sampled for the 2020 State DWINSA. The LSL, I&S, and OpW questions will be administered to all systems sampled for the 2020 State DWINSA in fully participating states. Medium CWSs in partial participation states will not be sent a survey asking for information about capital improvement projects, but they will still be sampled to respond to the LSL and OpW questions. A previously discussed in Part A, this ICR assumes that all states will have their medium systems participate in the LSL and OpW questions.

### B.1.b Key Variables

Several key variables are available from the Safe Drinking Water Information System (SDWIS). To ensure accuracy, the 2020 DWINSA will verify these data by asking respondents to confirm existing information (pre-populated on the data collection instrument) or correct it. These variables include population served, total design capacity, number of service connections, primary source of supply, ownership type (private or public) and whether the system purchases water from or sells water to another water system.

Information on capital needs will be collected from respondents on a project-by-project basis. For each project, respondents will be asked to provide the following types of information: type of need; reason for need; documentation of need and cost (if necessary); if the project is a new project or to replace, rehabilitate or expand existing infrastructure; if the project is needed now to protect public health or if it is needed over the next 20 years to continue to provide safe drinking water; the federal regulation or state requirement (if the project is to meet a current regulation or state requirement); design capacity of source, storage, and treatment projects; length and diameter of pipe projects; diameter for projects such as water meters; cost of the project (if available); and date of the cost estimate. For most of these variables, respondents will choose the appropriate “documentation,” “type of need,” “reason for need,” or “regulation or requirement” from EPA’s Lists of Codes.

The principal variable of interest is total projected capital needed for each water system in the 2020 DWINSA for the time period of January 1, 2020, through December 31, 2039. The total capital need for all systems in each state (to be derived from the statistical sample of systems) is the key variable that decision-makers at EPA use to allocate funds to states based on need.

The method of data collection has been designed to minimize burden on respondents while ensuring that information is collected in a consistent manner. Collecting information on a project-by-project basis, for example, will be particularly helpful in reducing burden since most respondents develop Capital Improvement Plans (CIPs) on a project-by-project basis.

Information on type of need will be used to disaggregate total capital needs for EPA’s Report to Congress. Information on the reason for need will be used to verify the public health benefit of the need. Information on the date of the cost estimate will be used to provide a consistent basis for cost estimates across systems. Information on a regulation or requirement will be used to determine the reported project costs related to federal regulations or state requirements.

If a system cannot provide cost estimates, additional data are necessary so that the Agency can impute costs. Each of these variables is described in greater detail later in this document.

### B.1.c Statistical Approach

The 2020 DWINSA is being designed to achieve a desired level of precision for state-level estimates of total capital needs for medium and large CWSs. It also is being designed to estimate the total capital needs of small systems and NPNCWSs for the nation as a whole. EPA proposes to use an approach that includes a census of large CWSs and a survey of a statistical sample of medium-sized CWSs to estimate total capital needs. This statistical approach minimizes burden while achieving the desired level of precision. For NPNCWSs, EPA proposes a census of NPNCWSs serving more than 10,000 persons and a sample of NPNCWSs serving 10,000 or fewer persons. Each approach is described in more detail below.

##### Medium and Large CWSs

The 2020 DWINSA design divides CWSs serving populations of more than 3,300 into two groups: CWSs serving populations of more than 100,000 and systems serving populations of 3,301 to 100,000. EPA proposes to sample with certainty systems serving more than 100,000 persons. These systems have the largest capital needs and they have the staff to respond efficiently to the 2020 DWINSA. EPA proposes to use a random sample of systems serving 3,301 to 100,000 persons. This methodology can reduce burden and still achieve the DWINSA data quality objectives.

To further reduce burden, EPA proposes using a modified panel approach for the 2020 DWINSA. Rather than select a completely new sample of systems in 2020, EPA will reassess the needs of most of the systems that participated in the 2015 DWINSA. EPA will replace 25 percent of the sample of systems serving 3,301 to 100,000 persons. By state and stratum, EPA will randomly select 25 percent of the sample to drop and will then randomly select replacement systems from the sampling frame. By primarily using information from the 2015 DWINSA, this approach would reduce the amount of time needed for systems to prepare and states to review the responses from systems resurveyed in 2020. This approach will maintain EPA’s sampling targets for each stratum and ensure that EPA continues to meet its precision targets for each state. Additionally, by using information from the 2015 review, the modified panel approach would reduce the amount of time necessary for EPA’s contractor to review each system’s response. By replacing 25 percent of the sample, EPA will reduce a potential source of bias introduced by the panel. (When a completely new sample was selected for each assessment, the sampling error was a random component that changed from survey to survey. With the panel approach, this error becomes systematic.) By refreshing 25 percent of the sample, the approach alleviates this potential source of bias and helps to ensure that the 2020 sample represents the need as it exists in 2020. This approach was used for the 2015 State DWINSA.

To meet the state-level precision targets, EPA will use the same strata as in the 2015 DWINSA. As previously mentioned, EPA will adjust the sample size to accommodate changes in the sample frame. These changes may address new systems, systems that are no longer active and systems that have “migrated” between strata (become smaller or larger, or changed source). If EPA determines that there have been substantial changes in the size of the sample frame since the 2015 DWINSA, EPA will adjust the sample size as needed to ensure that the precision targets are met for each state. As in the 2015 DWINSA, EPA will first determine the total sample size for each state to meet the target level of precision. EPA will then allocate the sample to strata in order to maximize the efficiency of the design.

##### Small CWSs

The objective of the 2020 DWINSA is to develop state-level estimates of total capital needs for CWSs. For large and medium systems, as explained above, this objective is achieved by selecting samples that are allocated across various strata in the population of systems to achieve an overall precision level for each state. Several barriers prevent us from developing state-level estimates for systems serving populations 3,300 or fewer:

* First, a mail survey is not an effective approach to collection of data from these small CWSs. State experience with mail surveys for small CWSs suggests that total non-response and item non-response would be very high with a mail survey. Also, states believe that the absence of knowledgeable respondents at small CWSs limits the general reliability of the responses. Therefore, the best way to gather information from small CWSs is through site visits made by EPA contractors. This will minimize total non-response, eliminate item non-response, and significantly improve the reliability of data collected.
* Second, if EPA assumes that all data collected from small CWSs will require site visits, then the number of such visits is constrained by the budget allocated for the 2020 DWINSA. EPA’s current budget provides for approximately 606 site visits. Including small CWSs in the state-level design proposed for the medium and large systems, however, would require several thousand site visits. Thus, the statistical design for medium and large systems cannot be applied to small CWSs.
* Given this dilemma, EPA will adopt a different approach for small CWSs, one that focuses on national-level estimates. This approach is consistent with the approach used for the 2007 DWINSA. The direct sample estimates of total capital needs at the national level will be used to infer the total capital needs for small CWSs in each state. The workgroup for the 2007 DWINSA preferred this approach, and no problems were identified in the 2007 DWINSA when this approach was implemented.

##### NPNCWSs

The 2020 DWINSA design divides NPNCWSs into two groups: NPNCWSs serving populations of more than 10,000 persons and systems serving populations of 10,000 or fewer persons. EPA proposes to sample with certainty systems serving more than 10,000 persons (approximately 13 systems). These systems have the staff to respond efficiently to the 2020 DWINSA. For NPNCWSs serving 10,000 or fewer persons, EPA proposes to use a random sample. This methodology can reduce burden and still achieve the DWINSA data quality objectives.

Similar to small CWSs, the same barriers prevent EPA from developing state-level estimates for NPNCWSs. Therefore, EPA proposes to adopt national-level estimates for NPNCWSs and to use sample estimates of total capital needs at the national level to infer the total capital needs for NPNCWSs serving 10,000 or fewer persons in each state. This approach was applied successfully to small CWSs and NPNCWSs in the 1999 State DWINSA and to the small CWSs in the 2007 State DWINSA, and EPA believes it will have similar success for NPNCWSs in the 2020 State DWINSA.

##### Medium CWSs in Partial Participation States

The 2020 DWINSA will not be administered to medium CWSs in partial participation states, but a sample of these systems will receive the LSL and OpW questions. Unlike medium systems in states that are fully participating in the survey, EPA will develop national-level estimates for these systems. EPA will sample 362 medium systems in partial participation states to receive the LSL and OpW questions.

EPA is designing and conducting the 2020 DWINSA with the assistance of a contractor:

|  |  |
| --- | --- |
| **Contractor**  The Cadmus Group LLC  100 5th Avenue, Suite 100  Waltham, MA 02451  (617) 673-7000 | **Contractor Roles**   * Oversight of data collection instrument design. * Oversight of statistical sample design. * Training. * Data collection instrument package distribution: logistics. * Technical support for respondents and states. * Review of all survey responses for conformance to survey policies and project allowability criteria. * Model development. * Data processing. * Statistical sample design. |

### B.1.d Feasibility

The 2020 DWINSA data collection instrument has been designed with the capabilities of the typical respondent in mind. To fully assess feasibility, EPA undertook the following steps. EPA convened a workgroup (see Section A.5.b) to comment on the proposed data collection and its feasibility. The data collection instrument to be used for the 2020 DWINSA is generally the same form as used for the past three DWINSAs. For the 2007 DWINSA, EPA conducted a pre-test in which EPA’s contractor met with individual CWS operators and discussed the proposed survey. System operators were asked to comment on all proposed data elements and the feasibility of collecting information by a mail survey. The Agency recognizes that most systems serving fewer than 50,000 persons and some that serve 50,000 or more may not have cost data or documentation of costs for some projects. In those cases, the 2020 DWINSA data collection instrument requests other readily available information that EPA can use to model costs. EPA will emphasize to respondents that they are not expected to develop cost estimates for the purposes of the 2020 DWINSA. In addition, EPA (or states) will provide systems with technical assistance for completing the data collection instrument.

EPA has developed cost models for most of the infrastructure needs included in the 2020 DWINSA based on the size and capacity of a project. These cost models were originally developed during the 1995 DWINSA, have been updated during subsequent assessments, including the 2015 DWINSA, and will be used again for the 2020 DWINSA. New cost models may be developed for weaker cost models, influential cost models, and new technology.

Unlike the medium and large systems, the 2020 DWINSA will not be self-administered by small CWSs or NPNCWSs serving 10,000 or fewer persons; rather, EPA contractors, accompanied by state personnel if state personnel choose to participate in this portion of the 2020 DWINSA, will visit the small CWSs and NPNCWSs serving 10,000 or fewer persons. Prior to the visit, the contractors will have access to all state records on the system (e.g., the results of recent sanitary surveys and inspections). The contractors will spend approximately 3.34 hours with the small CWS owner or operator and approximately 1.75 hours with the small NPNCWS owner or operator, requesting information that will be helpful in estimating system infrastructure needs and conducting a physical inspection of the system to confirm information provided by the owner or operator.

The EPA contractor will focus attention on the capital needs associated with treatment of source water, transmission, storage, and distribution. Capital needs associated with treatment will be modeled using methods similar to those currently used by EPA in the development of economic analyses. (In these analyses, data on occurrence of contaminants and cost estimates for treatment of source water to remove contaminants yield the cost of compliance with regulations that require the removal of contaminants from finished water.)

Reliance on site visits to small CWSs was strongly recommended by the 2007 DWINSA EPA workgroup to avoid problems that have faced every state survey of small CWS infrastructure needs:

* *Total non-response.* Since many systems have not clearly identified responsible parties, and since responsible parties often are unwilling to respond to data collection instruments, it is difficult to use a mail survey to obtain the necessary information. Working with participating state regulatory agencies and representatives of small CWSs should minimize non-response problems.
* *Item non-response.* System owners and operators often are not knowledgeable about the capital needs of their systems. Unlike larger systems, who may maintain CIPs, small CWSs lack information to answer questions. Since the EPA contractor engineers will conduct site visits to gather data, item non-response should be eliminated.
* *Reliability.* State drinking water regulators are suspicious of information provided directly from owners or operators of small CWSs. Unlike larger systems, small CWSs usually do not have professional, certified operators. Instead, one is likely to meet mobile home park owners, volunteers from homeowners’ associations, and others who are not water supply professionals. State drinking water administrators clearly prefer the judgments of the EPA contractor engineers, accompanied by their own staff, for reliable information on capital needs.

NPNCWSs serving 10,000 or fewer persons typically face these same challenges, and EPA proposes to treat them the same as small CWSs in the 2020 State DWINSA to minimize the burden on these systems and maximize survey response.

Finally, employing site visitors will substantially reduce the burden on small CWSs and NPNCWSs serving 10,000 or fewer persons. Total burden on the systems, on average, will be about 3.59 hours for small CWSs and 2 hours for NPNCWSs serving 10,000 or fewer persons. Instead of completing a data collection instrument, the system owner or operator can answer questions asked by the visiting engineer. The approach was discussed with knowledgeable state drinking water regulators as well as representatives of small CWSs and NPNCWSs serving 10,000 or fewer persons, and all parties agreed that it was the best approach to achieve the desired results of the 2020 DWINSA.

The time frame for the 2020 DWINSA is acceptable to the users of data within the Office of Ground Water and Drinking Water (OGWDW) and sufficient to complete a report to Congress by its anticipated due date in 2022. The schedule also is acceptable to other users of the data.

## B.2 SURVEY DESIGN

This section contains a detailed description of the statistical survey design and modified panel approach including a description of the sampling frame, sample identification, precision requirements and data collection instrument.

##### Medium and Large CWSs

The sample design for the 2020 DWINSA is stratified random sampling within each state. In cases where the state is not participating in the data collection for systems serving 3,301 to 100,000 persons, EPA will not provide state-specific results, as the data collection for these states does not meet the DWINSA data quality objectives. EPA will include an overall national result for the systems serving 3,301 to 100,000 persons, using the average need by strata of the systems in states that are participating in the full 2020 DWINSA. For states that are fully participating in data collection for systems serving 3,301-100,000 persons, the 2020 DWINSA will use a modified panel approach for sampling these systems within each state. This approach is described in more detail in Section B.2.b.

Stratification increases the precision of estimates compared with a simple random sample of the target population of systems. In stratified samples, the target population is divided into non-overlapping groups, known as strata, from which separate samples are drawn. The goal of stratified sampling is to choose sample sizes within each stratum in a manner designed to obtain maximum precision in the overall estimate for the population. Stratification variables for this study include: population size (populations of: 3,301 to 10,000; 10,001 to 25,000; 25,001 to 50,000; 50,001 to 100,000 and more than 100,000) and primary sources of supply (surface and ground). Systems serving more than 100,000 persons are selected with certainty. For the 2020 DWINSA, the survey will rely on a modified panel approach in which 75 percent of the 2015 DWINSA respondents serving populations of 3,301 to 100,000 were resampled, and 25 percent were put back into the frame and a new 25 percent were drawn. Sampling for the 2020 DWINSA will follow this same approach.

EPA’s precision target for the 2020 DWINSA is to be 95 percent confident that the true need for each state lies within an interval of plus or minus 10 percent of the estimated need. These precision targets are identical to the targets for the 2015 DWINSA. The 2015 sample, modified as described above, will meet the assessment’s precision target. The sample sizes will be adjusted to account for changes in the inventory of systems, if necessary, to ensure the 2020 sample meets the precision targets.

##### Small CWSs

The 2020 DWINSA design for small CWSs, like that for medium and large systems, is stratified random sampling. The stratification variables for small CWSs are the same as those for other systems: size of population served and primary source of supply.

Unlike the medium and large systems, the design for small CWSs is driven by significant budgetary constraints: EPA cannot afford to complete more than approximately 606 site visits. EPA’s objective in sampling is to achieve the maximum level of precision on a national basis without exceeding that budgetary constraint. Precision targets will be discussed in Section B.2.c, below.

##### NPNCWSs

NPNCWSs serving more than 10,000 persons are selected with certainty. As previously discussed in Part A, larger NPNCWSs are complex systems (such as airports) relative to smaller NPNCWSs (such as mobile home parks). Due to the significant difference in complexity, EPA proposes to survey larger NPNCWSs serving more than 10,000 persons with certainty. The 2020 DWINSA design for small and medium NPNCWSs serving 10,000 or fewer persons is random sampling. Similar to small CWSs, the design for NPNCWSs serving 10,000 or fewer persons is driven by significant budgetary constraints: EPA cannot afford to complete more than approximately 100 site visits. EPA’s objective in sampling is to achieve the maximum level of precision on a national basis without exceeding that budgetary constraint. Precision targets will be discussed in Section B.2.c, below.

##### Medium CWSs in Partial Participation States

For systems in partial participation states, a stratified random sample of approximately 352 CWSs will be conducted in each state to determine which systems will receive the LSL and OpW questions. The sample design for these systems will be identical to that used for the medium CWSs in fully participating states.

### B.2.a Target Population and Coverage

The target population for the 2020 DWINSA is the number of CWSs and NPNCWSs in the nation. A CWS is a public water system (PWS) that serves at least 15 service connections used by year-round residents or regularly serves at least 25 year-round residents (40 CFR 141.2). An NPNCWS is a non-profit water system that does not regularly supply water to the same population year-round. The 2020 DWINSA is designed to produce estimates of the capital need CWSs and NPNCWSs for each participating state. In partial participation states, EPA will be able to provide state-specific results for systems serving 100,000 or more persons. EPA will include an overall national result for the systems serving 3,301 to 100,000 persons using the average need by strata of the systems in participating states and the total number of systems by strata in the partial participation state. The 2020 DWINSA is designed to produce estimates of the capital need of CWSs serving 3,300 or fewer persons and NPNCWSs for the nation as a whole.

### B.2.b Sample Design

This section describes the sample design. It includes a description of the sampling frame, target sample size, stratification variables and sampling method. The sampling design employed is a stratified random sample of CWSs. NPNCWSs serving 10,000 or fewer persons are not stratified for the sample. The strata employed in the design are discussed in Section B.2.b.iii. Neyman allocation is used to efficiently allocate the sample of water systems among the strata.

#### B.2.b.i Sampling Frame

The sampling frame is developed from SDWIS. SDWIS is a centralized database for information on PWSs, including their compliance with monitoring requirements, maximum contaminant levels (MCLs) and other requirements of the SDWA. The following information will be extracted from SDWIS for the statistical survey and verified by participating states:

* Name of system.
* Type of system (CWS).
* Retail population served.
* Consecutive population served.
* Total population served.
* Primary source (surface water or ground water).
* PWS identification number (PWSID).
* Ownership type.

From these data, EPA will develop the frame from which EPA will calculate summary statistics (e.g., number of systems per state in pre-defined strata) for use in calculating sample size. For the modified panel approach, the 2020 sampling design will use the 2015 DWINSA sample and make targeted modifications to account for changes in the inventory of systems between 2015 and 2020. Systems that have closed since the 2015 DWINSA will be removed from the 2020 survey. The needs of merged systems will remain in the survey as the need of the combined system. New systems added to the inventory (i.e., NPNCWSs, systems serving 3,300 or fewer persons, systems that served 3,300 or fewer persons in 2015 that now serve more than 3,300 persons, or newly created systems) will be sampled to ensure the sample is representative of systems in 2020. New large systems (those serving more than 100,000 persons) will also be added to the census.

##### Justification for the Use of SDWIS

The following criteria are often used in assessing a proposed sampling frame:

* It fully covers the target population.
* It contains no duplication.
* It contains no foreign elements (i.e., elements that are not members of the population).
* It contains information for identifying and contacting the units selected in the sample.
* It contains other information that will improve the efficiency of the sample design.

The units of observation for this survey are CWSs and NPNCWSs, a subset of PWSs. SDWIS is the ideal choice for a sample frame because of its inclusive coverage of all units of observation for the 2020 DWINSA. In addition, SDWIS has two other advantages: it contains information that will facilitate contacting the respondents and it contains other information that is useful in stratifying the sample, thereby improving the efficiency of the sample design.

In previous surveys where SDWIS was used as a sample frame, there have been criticisms of its utility. Since 1989, EPA has conducted audits of the quality of SDWIS data. As a result, EPA is aware of the problems with SDWIS. The audits, however, show that errors in classification of systems by strata proposed for the 2020 DWINSA are rare. The audits show that systems are misclassified by population or source in less than one percent of all cases.

To mitigate any potential problems with the sample frame, the 2020 DWINSA design anticipates substantial state involvement in the 2020 DWINSA process. For example, states will be checking the sample frame of systems that will be used to determine the final sample. In EPA’s experience, states often have in-house data systems with very accurate data. Even if these data are not transmitted to SDWIS, they are available and can be used by states to check the sample frame. Examples of state information that ensures an accurate frame is developed include the population served by consecutive systems that purchase sufficient water to affect the size of the infrastructure of the wholesale system, and knowledge of recent or pending source water changes to ground water or to surface water.

#### B.2.b.ii Sample Size

*Medium and Large CWSs*

Exhibit B-2-1 at the end of this subsection shows the preliminary sample sizes for the 2020 State DWINSA. For the modified panel approach, sample sizes for the 2020 State DWINSA will be the same as for the 2015 DWINSA except for changes to accommodate: (1) the addition of large systems serving more than 100,000 persons since the 2015 survey; (2) partial participation states from the 2015 survey that fully participate in the 2020 survey; and (3) changes in the inventory of medium systems, including systems changing size or source categories or the creation of new systems. As shown on Exhibit B-2-1, the sampling design will be implemented to achieve state-level precision targets for CWSs serving more than 3,300 persons. Precision targets are discussed in Section B.2.c.

The task of determining the sample size for each stratum requires two steps. The first step determines the sample size for each state that achieves the precision targets for that state. The second step allocates the sample among the relevant strata in the state. The strata are described in section B.2.b.iii.

The first step calculates the total number of systems required at the state level to meet the precision requirements. The sample size is given by:



|  |  |
| --- | --- |
| Where: | n0g = the sample size for state g (prior to the finite population correction)  Ngh = the total number of systems in the gth state in the hth stratum (taken from SDWIS)  sgh = the standard deviation of the variable of interest in the gth state for the hth stratum (estimated using data from the data from previous assessments)  H = the number of strata defined in the sample design for the gth state  Vg = the desired sampling variance for the total system (those serving more than 3,300 persons) capital needs estimate for state g. |

The desired error in the sample is expressed as a relative error. In the above equation, Vg = (d/Zα \* )2.  is an estimate of the total capital needs for a given state.  is computed for each state by calculating the mean total capital needs for stratum h (from the prior DWINSAs) and multiplying this mean by the actual number of systems in each stratum for that state (Ngh). Summing across strata provides an estimate of . d is the half-width of the desired confidence interval (0.10 for the Assessment). Zα is the value of a standard normal distribution for a confidence level of 1- α, (1.96 for the Assessment).

Because the number of water systems is known and finite, the following population correction is applied:



The second step allocates the total sample to each sampling stratum. EPA will randomly draw this number of samples from each of these strata. The Neyman allocation is used to determine the sample size for each stratum:[[1]](#footnote-2)



(Because systems serving populations more than 100,000 are to be sampled with certainty, H is the number of strata of systems serving 100,000 or fewer persons.)

In order to implement these sample size and sample allocation equations, EPA needs estimates for Vg, Ngh, sgh and mean total capital needs by stratum. Information on mean total capital needs by stratum and sgh were estimated using data from the prior DWINSAs.

Exhibit B-2-1 State Sample Sizes

| **State** | **Estimated Total Number of CWSs Serving More Than 3,300 Persons** | **Estimated Sample Size for CWSs Serving More Than 3,300 Persons\*** | **Estimated Sample Size for CWSs Serving More Than 3,300 Persons Receiving Full Survey\*\*** |
| --- | --- | --- | --- |
| Alaska \* | 17 | 17 | 2 |
| Alabama | 341 | 116 | 116 |
| Arkansas | 181 | 81 | 81 |
| American Samoa | 1 | 1 | 1 |
| Arizona | 134 | 34 | 34 |
| California | 689 | 174 | 174 |
| Colorado | 181 | 65 | 65 |
| Connecticut | 59 | 24 | 24 |
| District of Columbia | 1 | 1 | 1 |
| Delaware \* | 37 | 19 | 3 |
| Florida | 383 | 84 | 84 |
| Georgia | 255 | 54 | 54 |
| Guam | 3 | 3 | 3 |
| Hawaii \* | 33 | 22 | 2 |
| Iowa | 149 | 72 | 72 |
| Idaho \* | 52 | 24 | 2 |
| Illinois | 506 | 83 | 83 |
| Indiana | 216 | 77 | 77 |
| Kansas | 113 | 64 | 64 |
| Kentucky | 258 | 70 | 70 |
| Louisiana | 251 | 96 | 96 |
| Massachusetts | 258 | 56 | 56 |
| Maryland | 64 | 32 | 32 |
| Maine | 38 | 19 | 19 |
| Michigan | 308 | 63 | 63 |
| Minnesota | 184 | 116 | 116 |
| Missouri | 243 | 104 | 104 |
| Northern Mariana Islands | 2 | 2 | 2 |
| Mississippi | 209 | 80 | 80 |
| Montana \* | 32 | 16 | 1 |
| North Carolina | 300 | 70 | 70 |
| North Dakota \* | 50 | 43 | 1 |
| Nebraska \* | 48 | 21 | 2 |
| New Hampshire \* | 40 | 28 | 1 |
| New Jersey | 246 | 58 | 58 |
| New Mexico \* | 66 | 48 | 2 |
| Nevada | 35 | 17 | 17 |
| New York | 341 | 53 | 53 |
| Ohio | 325 | 71 | 71 |
| Oklahoma | 175 | 71 | 71 |
| Oregon | 120 | 37 | 37 |
| Pennsylvania | 353 | 54 | 54 |
| Puerto Rico | 100 | 42 | 42 |
| Rhode Island \* | 27 | 20 | 4 |
| South Carolina | 166 | 78 | 78 |
| South Dakota \* | 52 | 25 | 2 |
| Tennessee | 274 | 91 | 91 |
| Texas | 1162 | 115 | 115 |
| Utah | 127 | 42 | 42 |
| Virginia | 160 | 46 | 46 |
| Virgin Islands | 2 | 2 | 2 |
| Vermont \* | 35 | 20 | 0 |
| Washington | 223 | 56 | 56 |
| Wisconsin | 181 | 39 | 39 |
| West Virginia \* | 116 | 52 | 2 |
| Wyoming \* | 30 | 21 | 0 |
| **Total** | **9,952** | **2,889** | **2,537** |

\* Fourteen states are “partial participation” states and are expected not to participate in the statistical portion of the 2020 State DWINSA (i.e., collecting data for the full survey from systems serving 3,301 to 100,000 persons). However, these systems serving 3,301 to 100,000 persons in partial participation states will receive the LSL and OpW questions. The number in the “Estimated Sample Size for CWSs Serving More Than 3,300 Persons” column includes systems in partial participation states which will not participate in the statistical portion of the 2020 State DWINSA but that will receive the LSL and OpW questions. This table does not include NPNCWSs.

\*\* The number in the “Estimated Sample Size for CWSs Serving More Than 3,300 Persons Receiving Full Survey” column does not include systems that serve 3,301 to 100,000 persons in the 14 partial participation states. However, the number in this column does include systems that serve more than 100,000 persons because these large systems will participate in the census portion of the survey (i.e., collecting data from systems serving more than 100,000 persons). This table does not include NPNCWSs.

*Small CWSs*

The total small system sample is set at 606 by available resources. EPA will allocate the sample among eight strata to produce the most efficient estimate of small sample need, given this sample size. Section B.2.b.iii discusses the how the sample will be stratified. The sample for systems serving 3,300 or fewer persons is allocated among source water and population-served strata using a Neyman allocation.

*NPNCWSs*

There are 13 NPNCWSs serving more than 10,000 persons, and these systems will be sampled with certainty. The total sample of NPNCWSs serving 10,000 or fewer persons is set at 100 by available resources. EPA will select a random sample of 100 NPNCWSs serving 10,000 or fewer persons within the clusters or counties with small systems that are selected to be surveyed.

*Medium CWSs in Partial Participation States*

The total sample for medium CWSs in partial participation states is set at 352 by available resources. These systems are included in the counts in Exhibit B-2-1 represented in the column titled “Estimated Sample Size for CWSs Serving More Than 3,300 Persons.” EPA will deliver the LSL and OpW questions to a random sample of 352 medium CWSs in partial participation states.

#### B.2.b.iii Stratification Variables

The objective of stratification is to increase the efficiency of the sampling design (thereby reducing the number of systems to be sampled for a given level of precision). Stratified sampling may produce a gain in precision in the estimates of the characteristics of the target population as compared to simple random sampling. In stratified sampling, the target population (i.e., CWSs) is divided into non-overlapping strata that are internally homogeneous, in that the measurements vary little from one unit to another (i.e., the within-stratum variance is minimized). If the within-stratum variance is relatively small, then a precise estimate of the variable of interest can be obtained with relatively small samples. Each of the strata estimates can be combined to obtain a precise estimate for the overall target population. If the strata are constructed correctly, the target population estimate can be achieved with greater precision and with fewer samples than the estimate obtained from simple random sampling.

EPA’s drinking water programs have historically evaluated CWSs based on (1) the number of persons served and (2) the primary water source (ground water and surface water).[[2]](#footnote-3) Using total capital need information obtained from prior DWINSAs, EPA evaluated several classification schemes. This analysis showed that the stratification scheme used in prior assessments (10 strata based on size and source) would be appropriate for the 2020 DWINSA. For some states, EPA may combine the 10,001 to 25,000 and 25,001 to 50,000 size categories within each source category, resulting in 8 rather than 10 strata. EPA will combine these two size categories only if the sample using 8 strata is more efficient than the sample using 10 strata. The proposed strata for systems serving more than 3,300 persons are as follows:

| **Size of Population Served** | **Source** | **Sample Methodologies** |
| --- | --- | --- |
| 3,301 – 10,000 | Ground | Panel approach with 25 percent refresh using a random sample. |
| 3,301 – 10,000 | Surface |
| 10,001 – 25,000 | Ground | Panel approach with 25 percent refresh using a random sample. In some states the number of strata will be reduced based on analysis of optimal stratum boundaries. Specifically, in some states systems serving between 10,001 and 50,000 will be in one size group rather than two. |
| 10,001 – 25,000 | Surface |
| 25,001 – 50,000 | Ground |
| 25,001 – 50,000 | Surface |
| 50,001 – 100,000 | Ground | Panel approach with 25 percent refresh using a random sample |
| 50,001 – 100,000 | Surface |
| More than 100,000 | Ground | Sampled with certainty |
| More than 100,000 | Surface |

EPA’s sample design for small CWSs is also stratified based on the size of the population served and the source water of the system.

The proposed strata are as follows:

|  |  |
| --- | --- |
| **Water Source** | **Population Served** |
| Surface Water Systems | Fewer than 101 |
| 101 – 500 |
| 501 – 1,000 |
| 1,001 – 3,300 |
| Ground Water Systems | Fewer than 101 |
| 101 – 500 |
| 501 – 1,000 |
| 1,001 – 3,300 |

As noted previously, NPNCWSs serving 10,000 or fewer persons are not stratified for the sampling method.

#### B.2.b.iv Sampling Method

As indicated above, all CWSs serving populations of more than 100,000 and NPNCWSs serving populations of more than 10,000 will be sampled with certainty.

For CWSs serving 3,301 to 100,000 persons, all CWSs will be allocated to eight strata, based on the population served and primary water source. The sample size for each stratum in each state will be determined by the sampling strategy outlined above. As previously described, the modified panel approach that will be used for the 2020 sample will begin with the 2015 DWINSA sample and make targeted modifications to account for changes in the sample between 2015 and 2020. EPA will then “refresh” the sample by randomly replacing 25 percent of the 2015 sample of systems serving 3,301 to 100,000 persons with systems that were not included in the 2015 Assessment. To refresh the sample, 25 percent of systems serving 3,301 to 100,000 persons will be dropped from the survey and returned to the pool of systems that were not selected in 2015. The “refresh” will then randomly select systems from among the pool of systems that were not in the 2015 sample, the systems that were randomly dropped and new systems so that the number of systems in the sample reaches the same size as the 2015 sample.

The sampling method for the 25 percent refresh will be similar to the approach used in the 2015 survey. An equal probability random sample will be drawn from each stratum. Anticipating a level of non-response, EPA will over-sample the refresh systems to achieve the desired number of completed data collection instruments. Since the expected response rate for systems serving 3,301 to 100,000 persons is approximately 90 percent, EPA has increased the sample by approximately 10 percent. However, as discussed below, the DWINSA has consistently achieved a higher response rate than estimated. Therefore, EPA has included the full sample size estimate in the burden estimate of this ICR.

CWSs serving 3,301 to 100,000 persons in partial participation will be sampled for the LSL and OpW questions using the same strata as for CWSs serving 3,301 to 100,000 persons in fully participating states. However, there will be no refresh for these systems in partial participation states because these systems were not sampled in the 2015 State DWINSA.

All CWSs serving populations of 3,300 or fewer will be allocated to four strata, based on population served and primary source. The sample size for each stratum will be determined by the sampling strategy outlined above. The sampling method will be a two-stage probability proportional to size random sample within each stratum. Past response rates for these systems exceeded 90 percent. EPA will over sample to account for non-response and will draw a sample of 606.

NPNCWSs serving populations of 10,000 or fewer will be randomly sampled within the clusters or counties chosen to survey small systems. Past response rates for these systems exceeded 90 percent. EPA will over sample to account for non-response and will draw a sample of 100.

#### B.2.b.v Multi-Stage Sampling

To achieve the required precision, reduce the burden to small CWSs and NPNCWSs serving 10,000 or fewer persons, and to keep costs down, a two-stage cluster sample will be used for systems serving 3,300 or fewer persons and NPNCWSs serving 10,000 or fewer persons. The use of a two-stage sample design will result in slightly reduced precision for the stratum-level estimates.

*First-Stage Sample*

All small CWSs and NPNCWSs serving 10,000 or fewer persons will be assigned to a county (or county equivalent in jurisdictions that do not have counties). Data on all small CWSs and NPNCWSs serving 10,000 or fewer persons will be sorted by county so that EPA can determine the number of systems, by strata, in each county. If a particular county does not contain the required number of systems (a minimum of 6 systems), it is grouped with an adjacent county; the combined county group is referred to as a county-cluster or the primary sampling unit (PSU). The first-stage sample will be approximately 120 counties, selected with probability proportional to size, where size is a composite measure of the number of small systems in each county. This method ensures that counties with more CWSs serving 3,300 or fewer persons and NPNCWSs serving 10,000 or fewer persons have a greater probability of being selected.[[3]](#footnote-4)

States will be given a SDWIS list of small CWSs and NPNCWSs serving 10,000 or fewer persons in the county (or counties) selected in the first-stage sample for their jurisdictions. EPA will ask states to verify that (1) the systems on the list are active CWSs with populations of 3,300 or fewer persons and assigned to the appropriate county, and (2) the systems on the list are active NPNCWSs serving 10,000 or fewer persons. If the number of systems in a county is large (e.g., 100 or more), EPA will select a sub-sample of the systems in that county to reduce the burden on the state. This review by the states will produce a clean sample frame for the second-stage sample.

*Second-Stage Sample*

In the second stage, a stratified random sample of six CWSs is drawn from each of the PSUs selected in the first stage of sampling. An additional 100 NPNCWSs serving 10,000 or fewer persons will be randomly sampled from the PSUs.

### B.2.c Precision Requirements

#### B.2.c.i Precision Targets

The sampling design for the 2020 State DWINSA will be implemented at the state level. EPA’s goal is to be 95 percent confident that the margin of error, when estimating the total capital needs facing these systems in each state, will be plus or minus 10 percent of the total need for these systems. For example, if the total need for these systems in a state is estimated to be $2 billion, EPA will be 95 percent confident that the actual total need is between $1.8 billion and $2.2 billion.

The size of the sample of small CWSs and NPNCWSs serving 10,000 or fewer persons is driven by budget constraints, not precision targets. EPA estimates that the sample size of 606 for small systems and the sample size of 100 for NPNCWSs serving 10,000 or fewer persons will allow the Agency to estimate the national capital need of these systems with a 95 percent confidence interval equal to plus or minus 15 percent of the national small systems and small NPNCWS need. This precision level will be less than the level for estimates developed for medium and large CWS and NPNCWSs serving more than 10,000 persons, but it will not materially reduce the overall precision for total cost estimates at the state level. Costs for small CWS and NPNCWSs serving 10,000 or fewer persons are a small portion of total system costs in each state. Thus, the lack of precision for these systems will not significantly reduce the overall precision of the state-level estimates.

#### B.2.c.ii Nonsampling Error

EPA has developed an assessment approach that will employ several quality assurance techniques to maximize response rates, response accuracy, and processing accuracy to minimize non-sampling error.

Particular emphasis will be placed on maximizing response rates. Standard methods that have proved effective in other surveys involving states and water systems will be used, including the following:

* EPA and the states will coordinate in the production of a cover letter for the 2020 DWINSA. EPA’s opinion (shared by state drinking water administrators and trade associations) is that surveys on state letterhead will be better received than surveys on EPA letterhead. Therefore, states can use state-level cover letters signed by a senior state official instead of EPA letter.
* The states will place a telephone call to each participating system to ensure that they understand the survey process and their role.
* The data collection instrument design, content and format were reviewed by states that participated in the 1995, 1999, 2003, 2007, 2011, and 2015 DWINSAs.
* Questions being asked are those that owners or operators of systems should know. EPA does not ask questions that require monitoring, research, or calculations on the part of the respondent.
* The data collection instrument design is limited to a cover page of system information and characteristics and one project table, with three tables the system may use to record general information about the system’s infrastructure inventory. By limiting the information requested, EPA believes that the average American Indian and Alaska Native Village water system and CWS respondent in fully participating states can complete the data collection instrument in approximately 5.88 hours. Exhibit A-6-16 shows the breakdown of the total burden hours for CWSs by system size in states fully participating in the 2020 State and Native American DWINSAs.
* As previously described, a mail survey is not an effective approach to collection of data from small CWSs or NPNCWSs serving 10,000 or fewer persons. Site visits to these systems will minimize total non-response, eliminate item non-response, and significantly improve the reliability of data collected. Furthermore, site visits will minimize the burden on systems by having contractors complete the survey instrument, rather than the systems. EPA believes that the average small CWS respondent would spend approximately 3.59 hours to support the site visit, and the average small NPNCWS would spend approximately 2.00 hours. Exhibit A-6-17 shows the breakdown of total burden hours for NPNCWSs.
* Medium CWSs in partially participating states will receive a survey instrument that consists only of LSL and OpW questions. The burden on these systems will be minimal because they will not receive the full survey instrument. EPA believes that the average medium CWS in partially participating states would spend approximately 1.05 hours to respond to the LSL and OpW questions. Exhibit A-6-18 shows the breakdown of burden hours for medium CWSs in partial participation states.
* Respondents will be encouraged to call state personnel who will be trained to answer questions. In addition, EPA will provide technical assistance to states and water systems.
* The electronic format of the survey will make returning the data collection instrument convenient.

Standard methods to reduce other sources of non-sampling error also will be used:

* EPA expects complete coverage of the target population using SDWIS, supplemented by state review of all systems.
* Data will be 100 percent independently keyed and verified.
* The data collection instrument is pre-coded to improve accuracy by eliminating unnecessary processing steps.

Supplementing these standard methods, EPA proposes several unique steps to eliminate non-sampling error which have been developed in concert with organizations representing the states and water systems. These organizations believe that the 2020 DWINSA is important and that a high level of participation by all water systems is essential to its success. Because of the substantial commitment being made by states and water systems to the 2020 DWINSA, EPA believes that response rates will be higher than most surveys of similar respondents. To ensure success, states and organizations representing water systems are taking the following steps.

* ***Participation of the states***. Because the 2020 State DWINSA will be used to allocate DWSRF funds to states, each entity has a strong interest in achieving a high response rate. EPA believes that their participation will be a key factor in guaranteeing high response rates and low item non-response. Personnel who work with water systems every day are in a strong position to encourage systems to complete the 2020 State DWINSA form. States have committed to assisting EPA in achieving a high response rate by participating in follow-up activities. EPA will provide technical assistance to any system that has questions about the 2020 State DWINSA.
* ***Participation of Organizations Representing Water Systems***. EPA anticipates public support of organizations representing water systems. The prior assessments were supported by groups such as the American Water Works Association (AWWA), the National Association of Water Companies (NAWC), and the Association of Metropolitan Water Agencies (AMWA).

This support by the organizations representing the respondents for the 2020 State DWINSA can be helpful in many ways to minimize non-sampling errors. For example,

* + In past DWINSAs, national water associations sent letters to each system in their membership, stressing the importance of surveying drinking water infrastructure needs. These letters, along with the letter from the states, helped convince water systems to respond. EPA will seek similar support from these associations for the 2020 State DWINSA effort to encourage systems to complete the data collection instrument.
  + In the past DWINSAs, the largest association representing water systems serving populations greater than 3,300, AWWA, provided support through its national organization. To improve the response rate, AWWA enlisted the support of its state affiliates to conduct telephone follow-up calls to encourage response. AWWA assisted in past DWINSAs to help achieve high response rates. EPA will seek similar support from AWWA in support of the 2020 DWINSA.
* ***Communications Strategy***. EPA has developed a comprehensive communications strategy that will inform likely respondents of the need for their participation. This strategy includes articles in magazines, newsletters, and bulletins of all major organizations that represent (or communicate with) water systems. This includes publications of all of the organizations mentioned above, plus the state and local affiliates of these organizations. The strategy is designed to develop widespread peer-group support for participation in the 2020 State DWINSA.

### B.2.d Data Collection Instrument Design

Questions about system characteristics (name, population served, number of connections, and other customary business information) will be pre-populated on all data collection instruments. The respondent needs only to enter accurate information if any pre-populated information is not correct.

The 2020 State DWINSA is based on a matrix project table that requests a list of capital water system infrastructure projects that the system plans for the period 2020 through 2039. For each project listed, the water system is asked to provide:

* Type of need.
* Reason for need.
* Documentation of need.
* If the project is for new infrastructure or to replace, rehabilitate, or expand existing infrastructure.
* If the project is needed now to protect public health or if it is needed over the next 20 years to continue to provide safe drinking water.
* The federal regulation or state requirement if the project is needed to meet a current federal regulation or state requirement.
* Design capacity of source, storage, and treatment projects.
* Length and diameter of pipe projects.
* Diameter for projects such as meters.
* Cost of the project (if available).
* Date of the cost estimate (if necessary).
* Documentation of cost (if necessary).

For most of these variables, respondents will choose the appropriate “documentation,” “type of need,” “reason for need,” or “regulation or requirement” from EPA’s “Lists of Codes” (Appendix B). The data collection instrument has been designed to be concise, to avoid jargon, and to avoid ambiguous words or instructions. Terms and formats have been standardized to the extent possible. There is no intentional bias in the ordering of the items.

The data collection instrument will also include the LSL, I&S, and OpW questions, depending on the system being surveyed (see Exhibit A-1-1 for a breakdown of which systems will receive each of these categories of questions). These questions have been designed to be simple, clear, and concise, to avoid jargon, and to collect information that survey respondents are likely to provide. As previously described, the LSL information is mandated by AWIA Section 2015(e)(2) and will include questions that help EPA understand what is known about lead service lines at CWSs and NPNCWSs. The OpW questions will generate important information on current and anticipated drinking water treatment and distribution system operator staffing concerns. The I&S questions will provide construction material information on specific types of need to estimate the 20-year demand for iron and steel represented by DWINSA projects to aid EPA’s management of the AIS requirements under the SDWA section 1452(a)(4).

## B.3 PRE-TESTS AND PILOT TEST

### B.3.a Pre-tests

For the 2007 DWINSA the data collection instrument and some policies were modified substantially. EPA conducted two pre-tests of the data collection instrument for the 2007 DWINSA. These pre-tests were conducted by EPA’s contractor, The Cadmus Group LLC. The pre-tests gathered feedback on the effectiveness of the data collection instrument; highlighted imprecise, ambiguous, or redundant questions; and indicated where further inquiry was needed. A pre-test was held in both Maine (four participants) and Montana (three participants). These states were chosen because they were both partial participation states and therefore most of their systems did not participate in the 2007 DWINSA. Also, the contractor conducting the pre-tests has offices in both these states and by conducting the pre-test in these states was able to reduce costs. The names of the seven systems were provided to EPA by the 2007 DWINSA state contacts. Based on the comments received, EPA made modifications to the data collection instrument. Since EPA’s pre-tests of the 2007 DWINSA data collection instrument were so extensive, and because few changes have been made to the data collection instrument since the 2007 DWINSA, EPA believes that a pre-test is not needed for the 2020 State DWINSA.

The data collection instrument was modified for the 2011 DWINSA by the addition of questions and codes to gather information on projects with “green” and climate readiness attributes. Consequently, EPA conducted a limited peer review focused on these new questions. For the reasons identified above, EPA did not conduct a pre-test of the 2011 DWINSA data collection instrument. Based on the limited number of states that submitted projects with “green” or climate readiness attributes indicated in the 2011 effort, EPA concluded these attributes were likely underreported. For the 2015 DWINSA, the “green” and climate readiness questions were removed from the data collection instrument. Instead, EPA explored streamlined approaches that might enable the Agency, during the Survey review process, to identify and flag projects that are likely to have “green,” climate readiness or climate resilience attributes. The same approach will be used for the 2020 DWINSA.

The data collection instrument will be modified for the 2020 DWINSA by the addition of questions to gather information on lead service lines (LSLs), pipe and storage iron and steel (I&S) materials, and operator workforce (OpW) issues. These questions are minor additions to the overall survey and have undergone a peer review. The peer review comments and EPA’s responses are summarized in Appendix C. The questions do not change any procedural aspects of the DWINSA, so no pre-tests were needed. This is consistent with the approach EPA used for the 2011 DWINSA when the “green” and climate readiness questions were added.

### B.3.b Pilot Test

To eliminate unnecessary burden on states and water systems, it has been decided that no pilot test for the 2020 State DWINSA will be conducted. A pilot test was conducted for the 1995 DWINSA and consisted of 60 CWSs from New York and Texas.

Starting with the 2015 State DWINSA, EPA began delivering all data collection instruments electronically rather than mailing hard copies, as was done for the 1995, 1999, 2003, 2007 and 2011 DWINSAs. States used the electronic data collection instruments in the 2015 DWINSA and expressed no concerns about using the electronic format. EPA believes this approach has been well tested and has proven to be successful; therefore, it is not necessary to repeat this testing step. Similarly, site visits were conducted for small CWSs in the 2007 State DWINSA and for small CWSs and NPNCWSs in the 1999 DWINSA, and EPA believes that this approach has proven to be successful and will not conduct any additional testing for the site visits to small CWSs and NPNCWSs in the 2020 DWINSA.

## B.4 COLLECTION METHODS AND FOLLOW-UP

### B.4.a Collection Method

The proposed collection method is an electronic survey. The data collection instrument including the Lists of Codes will be sent to the states via e-mail. State drinking water agencies will provide the data collection instrument (with the project table prepopulated for systems that participated in 2015 and blank for those that did not) and other necessary documents to the systems in the sample. They will follow-up if the data collection instrument has not been returned in 30 days. For a complete description of the follow-up procedures proposed to increase the response rate, see section B.2.c.ii.

The proposed collection method for small systems and NPNCWSs is to visit each small system and NPNCWSs in the sample. An EPA contractor, accompanied by state personnel that choose to participate, will interview the owner or operator and fill in the data collection instrument for all costs except treatment costs. (Costs of treatment will be modeled, using methods similar to those used by the OGWDW for regulatory impact analyses for new regulations.)

### B.4.b Survey Response and Follow-up

The target response rate (defined as the ratio of responses to eligible respondents) for the 2020 State DWINSA is 90 percent. EPA realizes that this is an ambitious target, but EPA believes that there are special circumstances that warrant such a target. Also, overall response rates of 94, 97, 96, 93, 97, and 99.7 percent were achieved in the 1995, 1999, 2003, 2007, 2011, and 2015 surveys, respectively. In the first six surveys, EPA conducted the following activities to achieve that high response rate.

* ***Seek Support from the Respondent Population***. This is a national survey of infrastructure needs for drinking water systems. EPA will work to bring to the attention of water systems, as well as all national organizations representing these systems, the importance of the DWINSA results. As with the previous six surveys, all national organizations will be contacted by EPA to seek their endorsement of the DWINSA and to communicate to their members the importance of a high response rate to their members. As discussed in Section B.2.c, in past surveys, organizations have provided access to their newsletters and magazines to publicize and endorse participation in the DWINSA. For the 2020 State Survey, EPA will seek similar efforts by these organizations.
* ***Follow-up by States and Respondent Peer Groups***. Since a majority of states have indicated their willingness to participate in follow-up activities, EPA has requested that state personnel, most of whom are personally familiar with the respondents, conduct follow-up procedures including the use of reminder letters and telephone calls to systems that have not responded with the needed information or documentation. If the follow-up fails after three attempts (one reminder letter plus two telephone follow-up calls), EPA is planning to shift to a second approach of peer-group follow-up by members of a trade association, such as AWWA.
* ***Recruitment by States and Respondent Peer Groups of Small Systems and NPNCWSs.*** In participating states, scheduling of site visits will be conducted by state personnel, most of whom are personally familiar with the respondents. If state personnel cannot schedule a visit with a system in the sample, EPA will turn to respondent peer groups.

## B.5 ANALYZING AND REPORTING SURVEY RESULTS

### B.5.a Data Preparation

State personnel will check all cost data and documentation for CWSs serving more than 3,300 persons and for NPNCWSs serving more than 10,000 persons to ensure that it is consistent with state and national standards. States will then send the completed and reviewed data collection instruments to EPA for a second round of review by EPA contractor staff. For CWSs serving 3,300 or fewer persons and for NPNCWSs serving 10,000 or fewer persons, EPA contractor staff will complete the data collection instrument through site visits, and a separate team of EPA contractor staff will conduct a round of review.

Once data have been checked, the contractor will key and verify the data. Senior data entry staff will be used for the verification process to improve quality control. Editing will include automated logic and range checks and checks for missing data. Missing cost data will be modeled, using other information provided by the respondents on the data collection instrument. When modeling is insufficient, missing data will be imputed using standard methods such as cell means and regression. The sample of water systems will be weighted so that stratum estimates can be summed to prepare state-level estimates for medium and large CWSs in the 2020 State DWINSA. EPA will report estimates on a national level for small CWSs and NPNCWSs.

### B.5.b Analysis

EPA will prepare a report that tabulates the results of the 2020 State DWINSA and explains the precision of the estimates of total capital needs. Examples of statistics that will be produced include:

* Total capital needs by state and by types of need for medium and large CWSs.
* Total capital needs for the nation by types of need for small CWSs and NPNCWSs.
* Total capital needs by domains within the total population, e.g., systems serving populations greater than 100,000.
* Total capital needs by system type.
* Number of lead service lines by state for medium and large CWSs and for the nation for small CWSs and NPNCWSs.
* Operator workforce information by system type and size.
* Information about iron and steel (I&S) materials represented by the projects reported in the 20-year infrastructure needs for the nation.
* Standard errors calculated for key statistics.

The analysis will be similar to that of previous DWINSAs.

### B.5.c Reporting Results

The 2020 State DWINSA results will be made available to EPA and the public through:

* A printed report that is submitted to Congress on drinking water infrastructure needs. This report will be made available to all participants in the 2020 State DWINSA and the public through EPA’s Safe Drinking Water website.
* Desktop computer access to state data on the DWINSA website without modeled project costs (each state can access only its own data).
* Desktop computer access to the entire data system (EPA only).

A report providing the cost models used to develop costs for the 2020 State DWINSA will be made available to EPA and the public through EPA’s Safe Drinking Water website.

# 2020 NATIVE AMERICAN DWINSA

## Introduction

In the following paragraphs, we present information on the survey of American Indian and Alaska Native Village water systems. The approach for the 2020 Native American DWINSA is largely the same as that for the 2020 State DWINSA, as described in Part B above, with minor differences. The discussion below repeats some of the sections of Part B where the approach is different from the approach being used for the 2020 State DWINSA as previously described.

## B.1 SURVEY OBJECTIVES, KEY VARIABLES AND OTHER PRELIMINARIES

### B.1.a Survey Objectives

The primary objective of the 2020 Native American DWINSA is the same as the objective described for the 2020 State DWINSA in Section B.1.a with the exception that the 2020 Native American DWINSA will not include the I&S questions because the AIS requirements of the SDWA do not apply to these systems.

EPA will use the information from the Native American DWINSA to estimate capital investment requirements of drinking water systems. The information will be used as part of an allotment formula for the DWSRF Tribal Set-Aside (TSA) Program.

For the 2020 Native American DWINSA, EPA is proposing to use a modified panel approach to select survey respondents, as described for the 2020 State DWINSA. The 2020 Native American DWINSA will draw from 2011 Native American DWINSA rather than the 2015 State DWINSA.

For the new systems selected for the 2020 Native American DWINSA, EPA will use the same methodology as used in previous DWINSAs. The sampling design is discussed in detail below. As previously described in Part A of this ICR, EPA will administer new questions in addition to the 2020 Native American DWINSA: the LSL and OpW questions. There are no separate precision targets for these questions, so EPA will administer these questions to the same systems sampled for the 2020 Native American DWINSA.

### B.1.b Key Variables

The key variables available for the 2020 Native American DWINSA are the same as those described in Section B.1.b for the 2020 State DWINSA with the exception that the survey instrument will be prepopulated with data from the 2011 Native American DWINSA for systems surveyed in the 2020 DWINSA.

### B.1.c Statistical Approach

The 2020 Native American DWINSA is designed to estimate the total capital needs of American Indian systems for the nation as a whole and for Alaska Native Village systems. EPA proposes a survey of a statistical sample to estimate total capital needs. This statistical approach minimizes burden while achieving the desired level of precision.

A mailed survey is not an effective approach to the collection of data from these water systems. EPA believes that the absence of knowledgeable respondents at these systems limits the general reliability of the responses. The best way to gather information from these systems is through direct contact by EPA Regions or the Navajo Nation. Thus, EPA Regions and the Navajo Nation will be responsible for collecting survey data, with assistance from the systems.

### B.1.d Feasibility

This section is the same for the 2020 Native American DWINSA as described in Section B.1.d for the 2020 State DWINSA, with exceptions noted below.

As noted in Section B.1.c above, the 2020 Native American DWINSA will not be self-administered by American Indian and Alaska Native Village water systems; rather, EPA regional offices and the Navajo Nation will complete the surveys with assistance from the water systems. EPA estimates that for each system, the survey will require approximately 3.13 hours (Exhibit A-6-16) from water system staff, 4.75 hours (Exhibit A-6-19) for the Navajo Nation, and 7.43 hours for the EPA regional office to complete. This approach substantially reduces the burden on Native American water systems and increases the accuracy of survey responses and the response rate for these systems.

## B.2 SURVEY DESIGN

The design for the 2020 Native American DWINSA, like that for the 2020 State DWINSA, is stratified random sampling. American Indian and Alaska Native Village water systems serving more than 3,300 persons will be sampled with certainty. The stratification variables for the small systems are the same as those for medium and large CWSs sampled in the 2020 State DWINSA: size of population served and primary source of supply. However, unlike the 2020 State DWINSA, the 2020 Native American DWINSA will select two separate samples: 1) American Indian systems in the continental U.S., and 2) Alaska Native Village systems. Stratification variables for both samples include population size (populations of: 25 – 500; 501 – 1,000; and 1,001 – 3,300), and primary sources of supply (surface and ground). Systems serving more than 3,300 persons are selected with certainty.

### B.2.a Target Population and Coverage

The target population is CWSs and NPNCWSs that have been designated as Native American. A CWS is a public water system that serves at least 15 service connections used by year-round residents or regularly serves at least 25 year-round residents. A NCWS is a “public water system that is not a community water system and that regularly serves at least 25 of the same persons over 6 months per year” (non-transient noncommunity water system) or is a public water system that is not a community water system and “does not regularly serve at least 25 of the same persons over six months per year” (transient noncommunity water system). (40 CFR 141.2)

### B.2.b Sample Design

The sample design for the 2020 Native American DWINSA is predominantly the same as that for the 2020 State DWINSA described in Section B.2.b. Differences between the 2020 State DWINSA and the 2020 Native American DWINSA are described in the relevant sections below.

#### B.2.b.i Sampling Frame

The sampling frame is developed for the 2020 Native American DWINSA using the same approach described in Section B.2.b.i for the 2020 State DWINSA with the exception that EPA Regions and the Navajo Nation will review the sample frame rather than states.

#### B.2.b.ii Sample Size

The procedures proposed for designing a sample size for the 2020 Native American DWINSA is the same as that proposed for 2020 State DWINSA. Equations 1, 2, and 3 still apply, except that a national sample size will be selected instead of state-by-state samples.

#### B.2.b.iii Stratification Variables

As with the design for the 2020 State DWINSA, the sample design for the 2020 Native American DWINSA is stratified on the basis of (1) size (number of persons served by the CWS or NPNCWS), and (2) primary source (ground water and surface water).

The proposed strata are as follows:

| **Size of Population Served** | **Source** | **Sample Methodologies** |
| --- | --- | --- |
| Fewer than 501 | Ground | Random sample |
| Fewer than 501 | Surface |
| 501 – 3,300 | Ground |
| 501 – 3,300 | Surface |
| More than 3,300 | Ground | Sampled with certainty |
| More than 3,300 | Surface |

#### B.2.b.iv Sampling Method

Similar to the 2020 State DWINSA, EPA will apply the panel approach to the 2020 Native American DWINSA based on participation in the 2011 Native American DWINSA. As indicated in Section B.2.b.iii above, all systems serving populations of more than 3,300 persons will be sampled with certainty.

For systems serving 3,300 or fewer persons, all systems will be allocated to six strata based on population served and primary source. The sample size for each stratum will be determined by the sampling strategy outlined above. The sampling method will be an equal probability random sample within each stratum. Anticipating a level of non-response, EPA will over-sample to achieve the desired number of completed data collection instruments. Since the expected response rate is 90 percent, EPA will draw a sample of 209 American Indian water systems and 95 Alaska Native Village water systems. However, the DWINSA has consistently achieved a higher response rate than estimated. Therefore, EPA has included the full sample size estimate in the burden estimate of this ICR.

### B.2.c Precision Requirements

#### B.2.c.i Precision Targets

The sampling design for the 2020 Native American DWINSA will be implemented at the national level for American Indian water systems and for the State of Alaska for Alaska Native Village water systems. EPA’s goal is to be 95 percent confident that the margin of error, when estimating the total capital needs facing these systems nationally (for American Indian water systems) and at the state (for Alaska Native Village water systems), will be plus or minus 10 percent of the total need for these systems.

#### B.2.c.ii Nonsampling Error

To minimize non-sampling error and maximize response rate and response accuracy, the 2020 Native American DWINSA data collection instrument will be completed by EPA Regions’ and the Navajo Nation personnel using information from the IHS SDS. In addition, EPA will employ several of the quality assurance techniques described in Section B.2.c.ii for the 2020 State DWINSA to minimize non-sampling error with the exception that EPA Regions and the Navajo Nation personnel will be performing the roles identified for states. These techniques include the following:

* EPA and the Navajo Nation will coordinate in the production of a cover letter for the 2020 Native American DWINSA.
* The data collection instrument design, content and format were reviewed by states that participated in the 1995, 1999, 2003, 2007, 2011, and 2015 DWINSAs.
* Questions being asked are those that owners or operators of systems should know, with assistance from the EPA Region’s and the Navajo Nation’s personnel. EPA does not ask questions that require monitoring, research, or calculations on the part of the respondent.
* The data collection instrument design is limited to a cover page of system information and characteristics and one project table, with three tables the system may use to record general information about the system’s infrastructure inventory. By limiting the information requested, EPA believes that the average American Indian and Alaska Native Village water system respondent require approximately 3.13 hours to assist EPA Regions and the Navajo Nation in the completion of the survey instrument. Exhibit A-6-16 shows the breakdown of the total burden hours for CWSs by system size in states fully participating in the 2020 State and Native American DWINSAs.
* The electronic format of the survey will make returning the data collection instrument convenient.

Standard methods to reduce other sources of non-sampling error also will be used:

* EPA expects complete coverage of the target population using SDWIS, supplemented by the EPA Region’s and the Navajo Nation’s review of all systems.
* Data will be 100 percent independently keyed and verified.
* The data collection instrument is pre-coded to improve accuracy by eliminating unnecessary processing steps.

### B.2.d Data Collection Instrument Design

The data collection instrument for the 2020 Native American DWINSA is the same as that for the 2020 State DWINSA with the exception that the 2020 Native American DWINSA will include the LSL and OpW questions but not the I&S questions.

## B.3 PRE-TESTS AND PILOT TEST

As previously indicated, the survey instrument used for the 2020 Native American DWINSA is the same as that for the 2020 State DWINSA and underwent the same pre-tests and pilot tests describe in section B.3 for the 2020 State DWINSA. Minor differences between the 2020 and 2011 Native American DWINSAs are discussed below.

### B.3.a Pre-tests

The data collection instrument for the 2020 Native American DWINSA is substantially similar to that used in the 2011 Native American DWINSA. However, as described in Section B.3.a for the 2020 State DWINSA, the “green” and climate readiness questions were removed from the data collection instrument that will be used for the 2020 DWINSA. Also, the data collection instrument will be modified for the 2020 Native American DWINSA by the addition of questions to gather information on lead service lines and certified operator workforce issues. These questions are minor additions to the overall survey, so no pre-tests were needed. The I&S and LSL questions were peer reviewed. The peer review comments and EPA’s response to comments are included in Appendix C.

### B.3.b Pilot Test

To eliminate unnecessary burden on the Navajo Nation and water systems, it has been decided that no pilot test for the 2020 Native American DWINSA will be conducted. As discussed in Section B.3.b for the 2020 State DWINSA, a pilot test was conducted for the 1995 DWINSA, and EPA used electronic data collection instruments for the 2015 State DWINSA. No concerns were expressed about using this electronic format. EPA believes this approach has been well tested and has proven to be successful; therefore, it is not necessary to repeat this testing step for the 2020 Native American DWISNA.

## B.4 COLLECTION METHODS AND FOLLOW-UP

### B.4.a Collection Method

The proposed collection method for the 2020 Native American DWINSA is for the EPA Region or the Navajo Nation to first preliminarily fill out the data collection instrument for each system in the sample based on information obtained from IHS and the water systems records. The Navajo Nation or the EPA Region will then contact each system and interview the respondent to identify possible additional projects and to concur on the final set of identified infrastructure investment needs. By having EPA Regions and the Navajo Nation conducting the survey in this manner for American Indian and Alaska Native Village water systems, the information collection burden on these water system respondents will be minimized.

### B.4.b Survey Response and Follow-up

The target response rate is the same for the 2020 Native American DWINSA as described in Section B.4.a for the 2020 State DWINSA, and EPA conducted the same activities to achieve the target 90 percent response rate (defined as the ratio of responses to eligible respondents) for the 2020 State DWINSA.

## B.5 ANALYZING AND REPORTING SURVEY RESULTS

### B.5.a Data Preparation

Like the state role for the 2020 State DWINSA, EPA regional staff and Navajo Nation will check all cost data and documentation to ensure that it is consistent with regional and national standards. EPA regional staff and the Navajo Nation will then send the completed and reviewed data collection instruments to EPA for a second round of review by EPA contractor staff.

Once data have been checked, the contractor will key and verify the data. Senior data entry staff will be used for the verification process to improve quality control. Editing will include automated logic and range checks and checks for missing data. Missing cost data will be modeled, using other information provided by the respondents on the data collection instrument. When modeling is insufficient, missing data will be imputed using standard methods such as cell means and regression. The sample of water systems will be weighted so that stratum estimates can be summed to prepare national-level estimates for the 2020 Native American DWINSA.

### B.5.b Analysis

EPA will prepare a single report that tabulates the results of the 2020 State and Native American DWINSAs and explains the precision of the estimates of total capital needs. Examples of statistics that will be produced for American Indian and Alaska Native Village water systems include:

* Total capital needs by types of need for the Native American DWINSA and for the American Indian and Alaska Native Village water systems.
* Total capital needs by domains within the total population, e.g., systems serving populations greater than 3,300.
* Standard errors calculated for key statistics.
* Number of lead service lines (LSLs).
* Operator workforce (OpW) information by American Indian and Alaska Native Villages.

The analysis will be similar to that of previous DWINSAs.

### B.5.c Reporting Results

The 2020 Native American DWINSA results will be made available to EPA and the public through the same means as described in Section B.5.c for the 2020 State DWINSA.

1. J. Neyman, “On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection,” *Journal of the Royal Statistical Society*, Vol. 97 (1934), pp. 558-606; as cited in William G. Cochran, *Sampling Techniques* (New York: John Wiley & Sons), 1977. [↑](#footnote-ref-2)
2. For the purposes of the 2020 DWINSA, purchased surface water systems are included with ground water systems. This design yields lower within-stratum variance and has been used since the 1999 DWINSA. [↑](#footnote-ref-3)
3. This method is based on Folsom, R.E., F.J Potter., and S.R. Williams, “Notes on a Composite Size Measure for Self-Weighting Samples in Multiple Domains,” *American Statistical Association 1987 Proceedings of the Section on Survey Research Methods*, August 1987, pp. 792-796. [↑](#footnote-ref-4)