Information Collection Request for the Clean Watersheds Needs Survey (CWNS) (Renewal)
EPA ICR No. 0318.14, OMB Control No. 2040-0050
June 2024

PART B OF THE SUPPORTING STATEMENT (FOR STATISTICAL SURVEYS)

Prepared for:

U.S. Environmental Protection Agency
Office of Wastewater Management
Water Infrastructure Division

Table of Contents

INTRODUCTION TO PART B

The Environmental Protection Agency (EPA) will conduct the following type of statistical survey for the 18th Clean Watersheds Needs Survey (CWNS). EPA will undertake an assessment of POTWs with a design capacity less than 1 million gallons per day (MGD), referred to as small POTWs. EPA will continue perform a census of POTWs with a design capacity greater than or equal to 1 MGD and all other infrastructure types (stormwater, decentralized wastewater treatment, and nonpoint source control). EPA will use the same data collection instrument as was used in the 2022 CWNS and use virtual site visits to collect data from small POTWs.

18th CWNS

B.1   SURVEY OBJECTIVES, KEY VARIABLES AND OTHER PRELIMINARIES

B.1.a   Survey Objectives

The primary objective of the 18th CWNS is to collect technical and needs data on the existing and planned publicly owned wastewater conveyance and treatment facilities, combined sewer overflow correction, stormwater management, decentralized wastewater treatment, and nonpoint source pollution control in the United States (including territories and excluding Tribes). Technical data refers to refers to data collected for each submission not related to infrastructure needs (e.g., wastewater or stormwater flow, population served). "Needs data" refers to information about the capital improvement needs, including cost and documentation. These data are used to produce a national estimate as well as state-specific estimates of the 20-year need to meet the water quality goals of the Clean Water Act. The EPA has established policies to ensure that these overarching goals of the survey are met:

- Assess the capital improvement needs for all projects that are eligible for Clean Water State Revolving Fund (CWSRF) assistance
- Analyze the treatment level of the nation's POTWs
- Report accurate data to Congress

The EPA proposes to collect technical and needs data for publicly owned treatment works (POTWs) and needs data for other infrastructure types (stormwater, decentralized wastewater treatment, and nonpoint source control). The EPA proposes using the cost estimation tools (CETs) developed for the 2022 CWNS to model costs for certain project types where a need is identified but cost data is unavailable.

The EPA proposes to perform a census of all stormwater, decentralized wastewater treatment, and nonpoint source control facilities as well as POTWs with a treatment plant design capacity greater than or equal to 1 million gallons per day (MGD). EPA is proposing a census for these POTWs, referred to as "non-small POTWs," because they represent the majority of the nation's and each state's POTW need, because they have substantial capital improvement needs, and because they are likely to have acceptable forms of documentation (e.g., capital improvement plans). EPA will perform a census of infrastructure types other than wastewater because there is no nationwide database or inventory of these facilities; therefore, there is no clearly defined universe from which to draw a statistical sample.

For the 18th CWNS, the EPA is proposing to conduct a statistical survey of small POTWs. For the purposes of the CWNS, "small POTWs" are defined as POTWs that include a treatment plant with a design flow of less than 1 MGD. The sampling design is discussed in detail below. The rationale for conducting a statistical survey of these facilities is to improve the quality of data collected and reduce the burden on states and municipalities and is based on the following:

- These small POTWs represent a substantial number of facilities in each state (and therefore a substantial reporting burden) but do not represent the majority of the wastewater need in most states.
- There is a well-defined database from which to draw the random sample.
- Many communities with POTWs of this size lack the resources and capacity to thoroughly assess their capital improvement needs.
- The EPA has data regarding project-specific costs that can be relied upon to model the costs for projects in this size range where need is identified but costs are not available.

B.1.b Key Variables

The key variables for conducting the random sample survey are facility type(s), physical location, NPDES permit information, population served, design flow, discharge type(s), and effluent treatment level. They are available from the 2022 CWNS dataset. To ensure accuracy, the 18th CWNS will verify these data by asking states to confirm or correct the 2022 data (pre-populated on the data collection instrument).

Information on capital needs will be collected from respondents at the project level. A project is a capital investment in an asset or program that addresses a water quality problem or public health problem related to water quality. For each facility, respondents will be asked to describe the changes that the projects will have on the facility, including new construction, existing (e.g., rehabilitation, replacement), or abandonment. For each submission, respondents will document the water quality problem (or public health problem related to water quality) that the projects will address by choosing one or more compliance-related needs from a dropdown list. For each project, respondents will be asked to provide documentation of the cost or inputs for EPA's CETs if no cost documentation exists.

B.1.c Statistical Approach

The EPA proposes to select a random sample of POTWs with a design capacity less than 1 MGD in order to reduce burden and better capture these needs. Several barriers prevented the EPA from capturing the full 20-year needs for small POTWs in the 2022 CWNS. Results and feedback from the survey showed:

- Over 60 percent (14,457) of wastewater submissions in 2022 were for communities identified as small (those with populations of 10,000 or fewer). Of these, only 41 percent (5,976 submissions) reported any needs. As it is unlikely that any facility would have no capital improvement needs over the next 20 years, it can be reasonably assumed that communities with unreported needs reflect reporting difficulties rather than true lack of need.
    - Of the 11,398 submissions with POTW flow less than 1 MGD, 6,669 (or 59 percent) did not report needs.
- After the 2022 CWNS, states provided feedback about the survey and indicated that gathering data from small communities was particularly difficult. The EPA's small community form (available both online and hardcopy), developed to facilitate capturing the needs of small communities that do not have other acceptable forms of documentation, had a low response rate at only 6 percent.
- State coordinators noted that many small communities do not have acceptable documentation of projects and their costs readily available nor staff who could certify cost estimates through the small community form.
- Some states also noted that they focused their limited resources on collecting data for larger communities with readily available documentation and higher potential needs per project to the detriment of small community needs collection.

B.1.d Feasibility

A statistical sample of small POTWs is feasible due to the following reasons:

- EPA contractors will perform the bulk of the needs data collection and analysis for small POTWs. The POTW point of contact will only need to provide existing documentation or inputs for cost estimation tools during a virtual site visit.
- Collecting data on needs from a fraction of small POTWs will allow states to focus their efforts on larger POTWs and other infrastructure needs. State coordinators indicated in 2022 that the small community non-response was due to their apportionment of limited resources.

- EPA contractors will be familiar with the data collection instrument and CWNS policies, allowing them to complete data entry for small POTWs more efficiently and accurately than state coordinators.
- Total burden on the systems, on average, will be about 2 hours.

B.2 SURVEY DESIGN

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

B.2.a Target Population and Coverage

The target population for the 18th CWNS is small POTWs across the nation. As mentioned above, EPA will perform a census of non-small POTWs and other infrastructure types as was done in previous surveys. The EPA proposes to calculate small POTW needs by state by multiplying the average of the national needs of the sampled POTWs by the total number of small POTWs in the state.

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 random sample of small POTWs. The stratum employed in the design is discussed in Section B.2.b.iii.

B.2.b.i Sampling Frame

The sampling frame, or the portion of the POTW population from which samples will be drawn (i.e., all POTWs with a treatment plant and design flow less than 1 MGD), will be developed from the 2022 CWNS dataset. The following information will be extracted from the 2022 CWNS for the statistical survey and verified by participating states:

- Name of system
- CWNS ID
- Facility type(s)
- Physical location
- NPDES permit number
- Residential population served
- Total population served
- Design flow
- Discharge type(s)
- Effluent treatment level

The EPA will use the 2022 CWNS dataset to develop the frame and calculate the necessary sample size.

Justification for the Use of 2022 CWNS Data

The units of observation for this survey are small POTWs – a subset of facilities inventoried in the 2022 CWNS. The 2022 CWNS provides EPA's most up-to-date inventory of POTWs nationwide.

To mitigate any potential problems with the sample frame, the 18th CWNS design anticipates substantial state

involvement early in the process. States will confirm key variables for all POTWs prior to the start of data collection to ensure the accuracy of the sample frame of POTWs used to determine the final sample. In the EPA's experience, states often have in-house data systems with very accurate data that can be used by states to check the sample frame data.

B.2.b.ii   Sample Size

Equation 1 is a standard equation for the minimum number of samples needed to estimate a population mean within a given confidence interval and at a specified confidence level.

Equation 1. Minimum Sample Size

$$N \geq \left[\frac{z_{1-\alpha} S}{d} \right]^2$$

Where:

N = the minimum sample size to achieve the desired level of certainty and confidence
$Z_{1-\alpha}$ = the z-score corresponding to the two-tail, 1-α level of confidence
S = population standard deviation
d = half-width of the desired confidence interval in units of the population mean

Equation 1 is designed to be used on normally distributed data, as using a single variable for the confidence interval (d) assumes symmetrical variability around the population mean. However, 2022 CWNS needs data are log-normal. Application of Equation 1 directly to 2022 CWNS data therefore results in unrealistic results.

To provide a first approximation of minimum sample size, the EPA applied Equation 1 to log-transformed 2022 CWNS needs data using one, two, three and four strata. As input to Equation 1, the EPA assumed 95 percent confidence (two-tail z-score of 1.96) and a confidence interval around the mean of ±1 percent (Table 1Error: Reference source not found). As will be shown, this narrow confidence interval on a log-transformed basis translates to a larger confidence interval when data are untransformed. Results show that across the considered frames, minimum sample sizes range from 287 to 403, or 8 to 61 percent of frame populations.

To provide the resulting confidence interval in dollars (rather than log dollars), the log-transformed mean and confidence interval is "untransformed," or converted back into units of dollars using an exponential transformation (base e). Columns under the "Untransformed" heading in Table 1 show these results, where the exponential transformation of the log-transformed mean is equal to the geometric mean, and the confidence interval (±1 percent of the log-transformed mean) is expressed as a percent of that geometric mean.

When untransformed, the symmetrical confidence interval of ±1 percent becomes asymmetrical and larger, ranging from ±13 to 17 percent. These would be the resulting characteristics of each strata's sample. However, the geometric mean of a lognormal distribution is much less than its arithmetic mean, which is what a statistical survey seeks to reproduce. Thus, the confidence intervals in Table 1 should be interpreted as approximate and smaller than true arithmetic mean confidence intervals, but indicative of the approximate statistical power of sample sizes relative to frame characteristics.

Table 1. Sample Size Approximation

Number of Frames	Population	$n_{population}$	Population Mean	Log-transformed			Untransformed		
				Mean	SD	n_{sample}	Geometric Mean[a]	CI-[b]	CI+[b]
1	<1 MGD	4,725	$5,529,346	14.6	1.49	399	$2,135,526	-14%	16%

Number of Frames	Population	$n_{population}$	Population Mean	Log-transformed			Untransformed		
				Mean	SD	n_{sample}	Geometric Mean[a]	CI-[b]	CI+[b]
2	<=0.5 MGD	3,678	$4,432,083	14.4	1.46	394	$1,743,859	-13%	15%
	0.5-1 MGD	1,047	$9,383,913	15.3	1.36	306	$4,351,274	-14%	17%
3	0-0.33 MGD	3,163	$3,936,082	14.3	1.45	395	$1,573,229	-13%	15%
	0.33-0.67 MGD	1,026	$8,634,708	15.1	1.36	310	$3,734,359	-14%	16%
	0.67-1 MGD	536	$8,987,165	15.3	1.37	307	$4,446,996	-14%	17%
4	0-0.25 MGD	2,759	$3,602,355	14.2	1.45	403	$1,436,290	-13%	15%
	0.25-0.50 MGD	919	$6,923,073	15.0	1.31	294	$3,122,501	-14%	16%
	0.50-0.75 MGD	628	$9,559,210	15.2	1.36	307	$4,188,200	-14%	16%
	0.75-<1 MGD	419	$9,121,177	15.3	1.36	304	$4,607,658	-14%	17%

SD = standard deviation; CI = confidence interval around the geometric mean. a – geometric mean calculated as e^(log-transformed mean). b – CI calculated as e^(±0.01*log-transformed mean).

To obtain a more realistic estimate of the true confidence interval that would result from the range of sample sizes established in Table 1, the EPA performed Monte Carlo simulations to replicate hypothetical sample campaigns of the full population of small POTWs, treating them as one frame (n=4,725, arithmetic population mean=$5,529,346). Sample distributions of 200, 300, 400, and 500 were created by randomly sampling the actual frame needs. For each sample size, a unique sample set was generated 1000 times, with the means of each model iteration compiled into a summary distribution. To find the confidence interval at 95 percent confidence, the 2.5 and 97.5 percentile values of the summary distribution were identified (Table 2). These values bound the mean need that would result 95 percent of the time, given the stated sample size.

Table 2**. Results of 1000 hypothetical sample campaigns performed using Monte Carlo analysis.**

Sample Size	Mean	Confidence Interval (-)	Confidence Interval (+)
200	$5,822,663	-26%	30%
300	$5,722,000	-21%	25%
400	$5,732,888	-18%	22%
500	$5,697,188	-15%	20%

As shown above, precision improves with a greater sample size; therefore, EPA is proposing to use a sample size of 500 based on available resources and to provide a conservative estimate.

 B.2.b.iii *Stratification Variables*

The EPA investigated the possibility of stratifying the small POTW universe to increase the efficiency and precision of the statistical sample. In order to justify using a stratified sample, needs within each strata must be relatively homogeneous and each strata must be relatively distinct from one another (i.e., the stratification variables would result in a statistically significant difference in total capital needs). The EPA evaluated 2022 CWNS data to determine whether any of the key variables could be used to stratify the small POTW universe and produce a more efficient sample or result in higher quality end data. Of the potentially relevant

variables collected in the 2022 CWNS, the EPA considered three that were both consistently collected and likely to have a material influence on total need. These variables are: future design capacity of the POTW, future effluent treatment level, and population served. Table 3 presents the 2022 CWNS data elements, quality, and completeness for each variable.

Table 3. Potential Stratification Variables

Data Element	2022 CWNS Data Element	Confidence in Data Quality	Completeness
Size	Treatment plant design flow	EPA reviewed for reasonableness (e.g., magnitude errors)	Reported for each treatment plant (either collected or carried over from 2012)
Size	Population receiving treatment	EPA reviewed for reasonableness, but some states indicated that they did not know where to obtain this data, and others did not update the starting values	Reported for each collection system (either collected or carried over from 2012) and calculated for each treatment plant
Type	Treatment plant effluent treatment level	EPA reviewed for reasonableness	Reported for each treatment plant (either collected or carried over from 2012)

The EPA eliminated population from consideration due to the inconsistent way in which it was reported by states and its lack of documentation. Additionally, treatment plant flow is considered a better indication of "size."

The EPA determined that, based on reviewing CWNS data, while differences in need by effluent treatment level exist, these differences are relatively minor relative to the variability within and across populations.

The EPA also evaluated the influence of future design capacity by dividing all needs data into two, three, or four frames. Results suggested that dividing the population into two or three frames may improve the estimate of needs for communities of different sizes, but dividing into four frames would not provide further improvements.

Based on the small difference in need between frames analyzed, as well as the high degree of variability in the data that necessitate similar minimum sample sizes for most frames, the EPA proposes to use one frame in the 18th CWNS. This allows the EPA to estimate need of the entire small POTW population with a similar level of accuracy as multiple frames but with a smaller sample.

B.2.b.iv Sampling Method

As indicated above, EPA will perform a census of POTWs with a design capacity greater than or equal to 1 MGD and all stormwater, nonpoint source, and decentralized wastewater treatment systems.

For POTWs with a design capacity less than 1 MGD, a random sample will be drawn from the total, ensuring that at least one small POTW is sampled from each state. Anticipating a level of non-response, the EPA will over-sample to ensure that the sample size is met.

B.2.b.v Multi-Stage Sampling

The EPA does not anticipate the need for multi-stage sampling.

B.2.c   Precision Requirements

B.2.c.i   Precision Targets

Based on 2022 CWNS data, with the proposed approach of selecting a random sample of 500 small POTWs from one unstratified frame, the EPA anticipates being able to collect data that provide an estimate of need with a 95 percent confidence interval plus or minus approximately 20-25 percent.

B.2.c.ii   Nonsampling Error

The EPA will 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 proven effective in other surveys involving states and municipalities will be used, including the following:

- The EPA and the states will coordinate in the preparation of an introductory email for the 18th CWNS that the state coordinator will send to the selected POTW. The EPA will reach out to the point of contact for the selected POTWs, setting up a virtual meeting to review the data collection instrument and collect the information required for the survey.

- The data collection instrument design, content, and format will be the Data Entry Portal and hardcopy small community form, which were developed for the 2022 CWNS with input from many state coordinators. Although the small community form had a low response rate, many states indicated having more success with one-on-one phone calls with the contacts to walk them through the form, which is the approach the EPA will take through virtual site visits.

- Data requested will be limited to information that the POTW point of contact can provide without requiring monitoring, research, or calculations on the part of the respondent.

- The EPA anticipates that virtual site visits to small POTWs will minimize non-response and improve the reliability of data collected. By having contractors complete the survey instrument through virtual site visits, rather than the POTW staff, burden on the POTWs and total non-response is expected to be minimized.

- The EPA expects complete coverage of the target population using past survey data, supplemented by state review of all POTWs.

- The EPA anticipates a target response rate (defined as the ratio of responses to proposed sample size) of 100 percent. If the EPA is unable to achieve the desired 500 responses with the initial sampling, they will replace refusals with a new random selection of small POTWs.

B.2.d   Data Collection Instrument Design

States will verify technical data for small POTWs based on 2022 CWNS data. The verified data will be pre-populated on the data collection instrument. The small POTW point of contact will be asked to confirm these data and provide information about the planned and/or needed changes at their facilities. They will also be asked to provide documentation of capital needs if this is available. If they have information about their needs, but not the costs of the projects, they will be instructed by the virtual site visitors to provide inputs for EPA's CETs (as was provided in the 2022 small community form). The EPA will populate information into the data entry portal based on information provided by the small POTW staff.

B.3   PRE-TESTS AND PILOT TESTS

B.3.a   Pre-tests

For the 2022 CWNS, EPA made substantial modifications to the data collection instruments (i.e., data entry portal and small community form) and policies. The EPA solicited feedback on these changes through a state coordinating committee and three subcommittees, consisting of CWNS state coordinators. State coordinators provided input on the data elements collected through the CWNS and made recommendations for data element requirements. They also provided feedback on the data entry portal through beta testing. State coordinator feedback during the data entry for the 2022 survey was primarily positive. The EPA anticipates making minor improvements to the data entry portal for the 18th survey based on this feedback. Because few changes to the data collection instrument are anticipated for 18th CWNS, the EPA believes that a pre-test is not needed.

B.3.b   Pilot Test

The data collection instruments (data entry portal and small community form) were used for the 2022 CWNS. To eliminate unnecessary burden, the EPA will consider the 2022 CWNS an adequate test and not conduct a pilot test for the 18th CWNS.

B.4   COLLECTION METHODS AND FOLLOW-UP

B.4.a   Collection Method

The proposed collection method is through the use of an electronic survey. The data collection instrument is EPA's data entry portal, which the EPA, EPA contractors, and state coordinators can access online. State coordinators will be provided with the data collection instrument (prepopulated with 2022 CWNS technical data) prior to data entry to confirm technical data.

The proposed collection process for small POTWs is to conduct a site visit (through a virtual meeting such as a webinar or video call) with the point of contact for each POTW in the sample. An EPA contractor, along with state personnel if they choose to participate, will interview the point of contact and complete the data collection instrument based on documentation provided or CET inputs provided.

B.4.b   Survey Response and Follow-up

The target response rate (defined as the ratio of responses to proposed sample size) for the 18th CWNS is 100 percent. The EPA will work with states to ensure POTW contact information is accurate, and states will send an "introductory email" priming POTW staff for EPA outreach and explaining the importance of the survey. The EPA will initially oversample with the expectation of receiving 90 percent participation. Refusal to participate will not be counted as non-response and the EPA will estimate their needs along with non-sampled small POTWs. If the EPA is unable to achieve 500 responses with the initial sampling, they will replace refusals with a new random selection of small POTWs.

B.5   ANALYZING AND REPORTING SURVEY RESULTS

B.5.a   Data Preparation

For POTWs with a design capacity less than 1 MGD, EPA contractor staff will enter data collected during virtual site visits into the data collection instrument, and a separate team of EPA contractor staff will review that data. Where a need is identified but cost data is unavailable, the EPA will model costs using other information provided by the respondents on the data collection instrument.

B.5.b   Analysis

Data collected using the statistical approach for small POTWs will be analyzed alongside other 18th CWNS data to produce statistics including:

- Total national capital needs by category for all infrastructure types.
- Total capital needs by state and category for all infrastructure types.
- Population served by treatment level nationally.
- Standard errors calculated for key statistics.

The analysis will be similar to that of previous CWNSs.

B.5.c   Reporting Results

Capital needs from the statistical sample with be reported alongside other (non-small POTWs and other infrastructure types) capital needs in the Report to Congress. Results will be made available to the EPA and the public through:

- A printed report that is submitted to Congress on clean water infrastructure needs. This report will be made available to all participants in the 18th CWNS and the public through EPA's CWNS website.
- Desktop computer access to state data on the CWNS DEP (each state can access only its own data).

A report providing the cost models used to develop costs for the 18th CWNS will be made available to the EPA and the public through EPA's CWNS website.