

## **– PART B OF THE SUPPORTING STATEMENT –**

### **1 SURVEY OBJECTIVES, KEY VARIABLES, AND OTHER PRELIMINARIES**

#### **1(a) Survey Objectives**

The primary objective of the statistical methods applied in this information collection is for the EPA to identify and select a sample of small PWSs that is representative of small PWSs nationwide. The selected sample of PWSs monitors for contaminants identified by the UCMR rule. The representativeness of this sample of PWSs is critical to the UCMR program because the drinking water contaminant occurrence data collected by the PWSs is used to: estimate national occurrence and exposure; establish a baseline for health effects and economic analyses; and provide information for regulatory determinations and, as appropriate, regulatory development.

#### **1(b) Key Variables**

Key variables associated with selecting a nationally representative sample of small PWSs include: PWS size, source water type and geographical location.

#### **1(c) Statistical Approach**

Section 1445(a)(2) of SDWA (as amended in 1996 and 2018) requires that the UCMR program include only a representative sample of PWSs serving 3,300 or fewer people. In addition to satisfying statutory requirements, selection of a sample of PWSs for participation in UCMR allows for significant national costs savings, as compared to monitoring by all PWSs. To estimate national occurrence and exposure, the primary UCMR program objective, the representative sample of PWSs must allow the EPA to collect high quality data about contaminant occurrence.

#### **1(d) Feasibility**

The EPA anticipates that the survey (the statistical sample) objectives are achievable given the existing time and resource constraints.

- High PWS response/participation rates (>95%) during UCMR 1, UCMR 2, UCMR 3, and UCMR 4 give the EPA confidence that equivalent or better participation rates can be achieved during UCMR 5.
- The statistical approach to this data collection requires only a fraction of small PWSs to conduct monitoring, resulting in much smaller cost and burden at the national level than

would be incurred if all small PWSs monitored. Small PWSs that are selected for UCMR 5 monitoring incur only a few hours of labor burden. The EPA pays for all laboratory fees and shipping costs related to small PWS testing.

- The survey results will support future CCLs and regulatory determinations.

## **2 SURVEY DESIGN**

### **2(a) Target Population and Coverage**

The SDWA, as amended by America's Water Infrastructure Act of 2018 (AWIA) (Public Law 115-270), specifies that, subject to the availability of appropriations for such purpose and appropriate laboratory capacity, the EPA's UCMR program must require all systems serving between 3,300 and 10,000 persons to monitor, and ensure that only a nationally representative sample of systems serving fewer than 3,300 persons are required to monitor. Therefore, PWSs are the target population for the UCMR 5 monitoring. The proposed UCMR 5 monitoring covers Assessment Monitoring, and does not include Screening Survey or Pre-Screen Testing.

### **2(b) Sample Design**

#### *2(b)(i) Sampling Frame*

The EPA developed the sample frame for the statistical selection of UCMR PWSs, including the system PWSID Code, name, source water category, and population-served data for each UCMR-eligible PWS. Initial data were pulled from the EPA's Safe Drinking Water Information System (SDWIS/Fed) inventory database and were adjusted to account for known anomalies in population and inventory reporting (for example, how wholesalers report their population data).

#### *2(b)(ii) Sample Size*

UCMR 5 will include Assessment Monitoring conducted by all PWSs serving 3,300 or more people.

#### *2(b)(iii) Stratification Variables*

In developing the representative sample, the EPA considers factors such as population served, water source and geographic location. The sample of PWSs is stratified by population served (PWS size), allocating samples proportionately to each state by PWS size and then by water source type. Other provisions, presented below, ensure broad geographic coverage.

### *2(b)(iv) Sampling Method*

To satisfy the specifications of SDWA section 1445(a)(2)(A), the representative sample of PWSs accounts for different PWS sizes, sources of water supply and geographic location (e.g., states). The sample is expected to be stratified by water source type (i.e., ground water or surface water and ground water under the direct influence of surface water) and by PWS size category (i.e., serves 25 to 500 people, 501 to 3,300 people, etc.). This stratification allows the EPA to account for different exposure risks of contaminant occurrence that could be related to the vulnerability differences between surface water and ground water sources and differing technical, management, and/or financial capacity that can vary across PWS sizes.

With contaminant exposure assessment as a primary goal, PWSs are expected to be selected in proportion to the population served. This population-weighted allocation leads to statistically valid estimates of national exposure. To ensure the sample provides equity across states for involvement in the UCMR, the EPA expects to include at least two PWSs from each state. This additional PWS selection requirement provides allocation across all the states and territories to account for differences in spatial vulnerability and contaminant occurrence, and to ensure equity in participation. Small tribal PWSs across the EPA Regions are expected to be grouped into a single category (equivalent to a “state”) for the representative sample.

### *2(b)(v) Multi-Stage Sampling*

Because PWS status can change over time, the EPA also expects to select “alternate” PWSs that fit the size/source water strata of the originally selected PWS. Through an interactive review process with the states, PWSs that no longer meet eligibility criteria (for example, if they are in a different size category than when originally selected, have become inactive, or do not have a retail customer base) will be replaced by an alternate PWS that meets the stratification criteria.2(c) Precision Requirements

### *2(c)(i) Precision Targets*

The representative sample of PWSs must be selected so that the data collected yield accurate and precise estimates of national contaminant occurrence (the fraction of PWSs at which a contaminant occurs) and exposure (the fraction of people exposed to a contaminant). For estimates of exposure fractions, the EPA specified a margin of error of  $\pm 1\%$  with 99% confidence, when the estimated exposure fraction is 1%. That is, if the estimated exposure fraction is 1%, the EPA will be able to state with 99% confidence that the true exposure fraction is between 0% and 2%. Because there are uncertainties and sources of variation in this and other such sampling programs, statistical sampling theory used to derive levels of accuracy and precision may not account for all of these sources of variation. Hence, the high confidence level, low allowable error, and consequent large sample size should help ensure adequate data to meet the objectives of the UCMR program.

### *2(c)(ii) Non-sampling error*

For those PWSs selected to conduct UCMR monitoring, response is a requirement. As with any regulation, some non-compliance can be expected. However, high compliance levels (>95%) during prior UCMR monitoring (attributable to extensive outreach and compliance assistance) give the EPA confidence that the same or better compliance levels can be achieved during UCMR 5. The EPA plans to continue outreach and compliance assistance efforts as needed.

### **2(d) Questionnaire Design**

No questionnaires are anticipated for UCMR 5. Analytical results for contaminant occurrence are expected to be reported directly by the laboratories to the EPA's electronic data reporting system.

## **3 PRETESTS AND PILOT TESTS**

This is the fifth cycle for UCMR monitoring, and the statistical methods have been very successful over the first four cycles. For UCMR 5, the EPA expects to apply the same basic statistical methods that were used to select the UCMR 1, 2, 3, and 4 national representative samples of small PWSs. Historically, more than 99% of small PWSs completed their required monitoring, and more than 95% of large PWSs completed their required monitoring and reporting.

## **4 COLLECTION METHODS AND FOLLOW-UP**

### **4(a) Collection Methods**

Large PWSs are required to submit their data (posted to the EPA's electronic data reporting system by their laboratories) to the EPA. Small PWSs work directly with an EPA-appointed UCMR Sampling Coordinator, and monitoring data for the small PWSs are submitted directly to the EPA's electronic data reporting system by the laboratories conducting the analyses.

### **4(b) Survey Response and Follow-up**

High compliance levels (>95%) during prior UCMRs give the EPA confidence that equivalent or better levels can be achieved during UCMR 5. The EPA continues outreach and compliance assistance efforts as needed. Each small PWS works with a UCMR Sampling Coordinator, and has minimal reporting requirements and one-on-one compliance assistance.

## **5 ANALYZING AND REPORTING SURVEY RESULTS**

### **5(a) Data Preparation**

After laboratories post UCMR 5 monitoring results and required data elements to the EPA's electronic data reporting system, the data is reviewed by the EPA's electronic data reporting system, the EPA, and states before placing the data in the NCOD for public access.

Data problems may occur, but the EPA takes the following efforts to reduce problems and increase the dependability and quality of the occurrence data. The EPA's electronic data reporting system and the EPA QA/QC assessments screen for the use of inappropriate measurement units and other improper data. Additional automated QC functions are already in place to identify possible data quality issues such as duplicate data submissions and incomplete data. All Assessment Monitoring samples are collected by knowledgeable PWS staff and analytical results are generated by laboratories that are approved for UCMR 5 drinking water analysis. Electronic data submission also avoids potential re-keying errors. As part of the data QA/QC procedures, all edits or changes made to the data are documented.

### **5(b) Analysis**

For UCMR 1, 2, 3 and 4, the EPA developed a two-stage analytical approach for the evaluation of the national occurrence of contaminants. The EPA will use the same two-tier approach to analyze the data for UCMR 5.

The first stage of analysis, Stage 1, provides a straightforward evaluation of occurrence for simple and conservative assessments of contaminant occurrence. The Stage 1 analysis of the UCMR data consists of non-parametric, unweighted counts and simple descriptive statistics of analytical results for each of the contaminants. These occurrence analyses are conducted at the sample level, PWS level and population-served level. For each contaminant, occurrence measures include the number and percent of samples with analytical detections and the minimum, median, maximum, and 99th percentile values of those detections. PWS-level occurrence measures include the number and percent of PWSs with one or more analytical detections and the number and percent of PWSs with two or more analytical detections of a given contaminant. Population-served occurrence measures include: the number and percent of population served by PWSs with one or more analytical detections, and the number and percent of population served by PWSs with two or more analytical detections of a given contaminant. Similar measures may also be conducted for each EPTDS for each PWS. Since these contaminant- and PWS- level occurrence measures are based on raw occurrence data (that have not been adjusted for population-weighting and sampling), they are less accurate representations of national occurrence than occurrence measures based on adjusted occurrence data.

Based on the findings of the Stage 1 analysis, the EPA can select contaminant(s) for which more detailed and sophisticated statistical evaluations – the Stage 2 analysis – may be warranted as a next step to generate national probability estimates of contaminant occurrence and exposure.

Specifically, the modeling and estimation of PWS mean contaminant concentrations may be desired. The Stage 2 analysis uses a Bayesian-based hierarchical model to estimate the percent (and number) of PWSs with a mean contaminant concentration above any specified concentration threshold. The Bayesian-based Hierarchical Model also provides quantified error of estimation, and enables estimates of mean contaminant concentrations below the minimum reporting level. This statistical model was used to generate the contaminant occurrence estimates for 60 regulated contaminants for the first Six-Year Review of National Primary Drinking Water Regulations, an approach that underwent peer review.

### **5(c) Reporting Results**

After final review and formatting, the data collected through this ICR are expected to be made available to the public through the NCOD, as was done with the data collected for UCMR 1, 2, 3 and 4. The analytical results from UCMR 5 monitoring will support regulatory determinations and, as appropriate, regulation development. For contaminants with significant occurrence and health effects, the EPA expects to use the results: to support an exposure assessment; to establish the baseline for health effects and economic analyses; to analyze contaminant co-occurrence; and to evaluate treatment technology, including contaminant source management. Further, the results may suggest that the occurrence of certain contaminants may be significant enough to initiate research on health effects and treatment technology. Finally, the data may guide future source water protection efforts.