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 EPA to identify and select a sample of PWSs that is representative of PWSs nationwide. The selected sample of PWSs will conduct monitoring of contaminants identified by the UCMR program. The representativeness of this sample of systems is critical to the UCMR program because the drinking water contaminant occurrence data collected by the PWSs will be used to: estimate national occurrence and exposure, establish a baseline for health effects and economic analyses, and provide information for regulatory determinations.

1(b) Key Variables

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

1(c) Statistical Approach

Section 1445(a)(2) of SDWA (as amended in 1996) requires that the UCMR program include only a representative sample of systems serving 10,000 or fewer people. In addition to satisfying statutory requirements, selection of a sample of systems 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 EPA to collect high quality data about contaminant occurrence.

1(d) Feasibility

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 have given EPA confidence that equivalent or better can be achieved during UCMR 2.
- The statistical approach to this data collection requires only a fraction of small systems to conduct monitoring, resulting in much smaller cost and burden at the national level than would be incurred if all systems monitored. Small systems that are selected for UCMR 2 monitoring will incur only a few hours of labor burden. EPA will pay for all laboratory fees and shipping costs related to small system testing.
- The survey results will be completed in time to inform the corresponding cycle of CCL regulatory determinations.

2   SURVEY DESIGN

2(a)   Target Population and Coverage

Public water systems are the target population for UCMR monitoring. Transient non-community water systems and those that purchase all of their water from another system are not subject to UCMR 2. All other PWSs that serve more than 10,000 people will be subject to at least the Assessment Monitoring component of UCMR 2 monitoring. Eligible small PWSs (serving 10,000 or fewer people) will only be required to conduct UCMR 2 monitoring if they are part of the statistical selection, as described below. Small PWSs will only be selected to monitor for either Assessment Monitoring or the Screening Survey, not both.

2(b)   Sample Design

2(b)(i) Sampling Frame

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

2(b)(ii) Sample Size

UCMR 2 monitoring will include: Assessment Monitoring conducted by all PWSs serving more than 10,000 people ("large" PWSs), and 800 representative PWSs serving 10,000 or fewer people ("small" PWSs); and Screening Survey conducted by all PWSs serving more than 100,000 people, and 800 representative PWSs serving 100,000 or fewer people.

2(b)(iii) Stratification Variables

In developing the representative sample, EPA considers factors such as population served, water source, and geographic location. The sample PWSs will be stratified by population served (system size), allocating samples proportionately to each State by system 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 systems will account for different system sizes, sources of water supply, and geographic location (e.g., States). The sample will be stratified water source type (i.e., ground or surface water) and by system size category (i.e., serves 25 to 500 people, 501 to 3,300 people, etc.). This

stratification allows EPA to account for different exposure risks of contaminant occurrence that may be related to the vulnerability differences between surface and ground water sources and differing management and financial capacity that can vary across system sizes.

With contaminant exposure assessment as a primary goal, systems will be selected in proportion to the population served, as was generally done under UCMR 1. 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, EPA will include at least two systems from each State. This additional PWS selection requirement will provide 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 water systems across the EPA Regions are grouped into a single category (equivalent to a "State") for the representative sample.

2(b)(v) Multi-Stage Sampling

Because PWS status often changes over time, EPA will also select "alternate" systems that fit the size/source water strata of the originally selected system. Through an interactive review process with the States, systems that no longer meet eligibility criteria (for example, if they are in a different size category than when originally selected, have become inactive, or purchase all of their water) will be replaced by an alternate system 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 systems in which a contaminant occurs) and exposure (the fraction of people exposed to a contaminant). For estimates of exposure fractions, EPA will allow 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%, EPA must 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) Nonsampling error

For those PWSs required to conduct UCMR monitoring, response is a requirement. As with any regulation, some non-compliance can be expected. However, high compliance levels (>95%) during UCMR 1 (attributable to extensive outreach and compliance assistance) give EPA

confidence that the same or better can be achieved during UCMR 2. EPA plans to continue outreach and compliance assistance efforts, as needed.

2(d)   Questionnaire Design

No questionnaires will be used for the UCMR. Analytical results for contaminant occurrence will be electronically reported directly by the laboratories to EPA's electronic reporting system.

3   PRETESTS AND PILOT TESTS

For UCMR 2, EPA will apply the same basic statistical methods that were used for the UCMR 1 national representative sample of small systems. Following sample adjustments made through communications with States, >99% of the final sample of small systems (and >95% of large systems) completed their required monitoring and reporting.

4 COLLECTION METHODS AND FOLLOW-UP

4(a) Collection Methods

Large PWSs (those serving more than 10,000 people) are required to submit their data through EPA's electronic data reporting system. Small PWSs (those serving 10,000 or fewer people) will be working directly with an EPA-appointed UCMR Sampling Coordinator, and monitoring data from the small PWSs will be submitted directly to EPA's electronic reporting system by the laboratories conducting the analyses.

4(b) Survey Response and Follow-up

High compliance levels (>95%) during UCMR 1 have given EPA confidence that equivalent or better levels can be achieved during UCMR 2. EPA plans to continue outreach and compliance assistance efforts, as needed. Each small system will be working with a UCMR Sampling Coordinator, and will have minimal reporting requirements and one-on-one compliance assistance.

"Lessons learned" during UCMR 1 helped refine several UCMR 2 requirements. Monitoring schedules in UCMR 2 will be specified for all large systems (whereas in UCMR 1, large systems could choose when to monitor across a multi-year monitoring period). In addition, reporting deadlines have been more clearly defined. In some cases during UCMR 1, EPA had to wait until the end of a three-year monitoring period to make a compliance determination and contact systems with appropriate follow-up. With the revisions under UCMR 2, EPA or participating States will have the ability to quickly contact systems regarding sample results that are not received when expected.

5 ANALYZING AND REPORTING SURVEY RESULTS

5(a) Data Preparation

After PWSs or their laboratories post their UCMR 2 monitoring results and required data elements to EPA's electronic reporting system, EPA allows a specified time for quality control review by the PWS, States, and the Agency before placing the data in NCOD for public access.

Data problems will likely exist, but the following efforts will be taken by EPA to reduce problems and increase the dependability and quality of the occurrence data. The UCMR electronic data reporting system and EPA QA/QC assessments will screen for the use of inappropriate measurement units and other improper data. In addition, EPA plans to have other automated QC functions in place to identify possible data quality issues such as duplicate data submissions, and data completeness. All samples will be collected by trained PWS staff and analytical results will be generated by laboratories that are approved for UCMR 2 drinking water analysis. Electronic data submission also avoids potential re-keying errors. Therefore, some assumptions are made regarding the general quality of the raw data that will be received. As part of the data QA/QC procedures, all edits or changes made to the data will be documented.

5(b) Analysis

For UCMR 1, EPA developed a two-stage analytical approach for the evaluation of the national occurrence of contaminants. UCMR 2 expects to use the same 2-tier approach to analyzing the data.

The first stage of analysis, Stage 1, provides a straight-forward 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, system level and population-served level. For each contaminant, occurrence measures include the number and percent of samples for each contaminant with analytical detections, and the minimum, median, maximum, and 99th percentile values of those detections. System-level occurrence measures include the number and percent of systems with one or more analytical detections, and the number and percent of systems with two or more analytical detections of a given contaminant. Population-served occurrence measures include: the number and percent of population-served by systems with one or more analytical detections, and the number and percent of population-served by systems with two or more analytical detections of a given contaminant. Similar measures may also be conducted for each entry point to the distribution system for each system. Since these contaminant and system 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, 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 system mean contaminant concentrations may be desired. The Stage 2 analysis uses a Bayesian-based hierarchical model to estimate the percent (and number) of systems 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 MRL. This statistical model was used to generate the contaminant occurrence estimates for 60 regulated contaminants for the first Six-Year Review of NPDWRs, for which it underwent a peer review.

5(c)   Reporting Results

After final review, formatting, and analysis of the data collected through this ICR, the data will be made available to the public through the NCOD, as is being done with the data collected and analyzed for UCMR 1. A full report, including detailed descriptions of the occurrence analysis approach; data characteristics (including data quality and limitations); analytical methodology; and analytical results, will be provided to the public and posted on the UCMR 2 Web site once complete. The analytical results and report will provide support for regulatory determinations for the drinking water CCL.