

PART B OF THE SUPPORTING STATEMENT

## INFORMATION COLLECTION EFFORT FOR FACILITIES WITH COMBUSTION UNITS

### 1. Respondent Universe

The vacated Boiler NESHAP and the remanded CISWI standard were developed using a 1997 a database of combustion units. The database was built using data from the Industrial Combustion Coordinated Rulemaking “inventory” and additional data received from non-fossil facilities “survey”. These inventory and survey databases contained 3,894 solid fuel fired units, or units that fired a combination of solid fuels and other fuels/materials. Since the vacature of the boiler and process heater NESHAP and the CISWI definition rule, the Agency has been working with state agencies to re-establish the population of facilities with boilers or process heaters that were subject to the NESHAP. This data has been collected mostly on the facility level, and at this time the Agency has minimal unit level data to generate a revised estimate of the population of boilers and process heaters that have emission test data gaps. However, the new initial notification and permit data indicates that the total number of facilities having boilers and process heaters potentially subject to the vacated NESHAP is approximately 2,360, which is much less than the 15,900 facilities with affected units listed in the 1997 inventory and survey databases. Given that the new population of facilities is much less, the Agency assumes that the number of solid-fuel units within this population will also be much less than the 3,894 solid-fuel units contained in the 1997 data. Therefore, a sample size of 350 units, which was based on the much larger 1997 dataset, is unlikely to be needed, and the number of actual stack tests is expected to be much less, assuming that the number of potential CISWI units is also less than expected. The agency expects that the number of CISWI units responding will be those covered by the 2000 standard as well as those specifically exempted from that standard. Thus, the Agency expects that there will be no more than 1,036 facilities with projected CISWI units. Once the Agency reviews and analyzes the data received from the electronic questionnaire, it will be able to revise the number of stack tests needed in order to obtain a statistically valid sample based on the new respondent universe. The sample design described below is an example approach of how the Agency would select test site candidates. Although the example is based on solid-fuel units from the 1997 database, the Agency may identify data gaps in other combinations of fuel and combustor type. For example, the survey may not provide sufficient emission data for CO or organic HAP from gas-fired units, and the Agency may request that a sample of gas-fired units test for CO only.

### 2. Respondent Universe Stratification

Two variables that appear to have an effect on HAP emissions are fuel/material type and combustor design (for combustion-based HAP such as CO). The emission control device also has an impact on HAP emissions, however the process for establishing a MACT floor will take into account the top 12 percent of the best performing units, regardless of control device, and so we have not included this variable in our sample design. For the purposes of grouping the combustion units into categories, these two variables will be used so to select a representative sample of combustion units for testing. Commenters requested that the Agency add a third variable, combustion unit size (i.e., design capacity), to the sample design. The Agency notes that combustion HAP emissions are

a function of combustor design, not the size of the boiler. In the vacated boiler and process heater NESHAP, the Agency defined the small subcategory of boilers to be units that are less than 10 million Btu per hour or units that have a fire tube type combustor design. The Agency will collect data on the design capacity during the survey component of this ICR, but it will not use combustor size as a variable in its sample design.

In the example below, solid fuel types are categorized as they were in the model units for the vacated boiler NESHAP: coal (fired-alone), a blend of coal/wood/non-fossil liquid/non-fossil solids, a blend of gas/wood/other biomass/liquid fossil fuel, blend of non-fossil liquid/non-fossil solid, blend of gas/non-fossil liquid/non-fossil solid, wood (fired alone), and a blend of wood/other biomass/non-fossil liquid/non-fossil solid. Combustor design was also categorized as it was used in the model units for the boiler NESHAP: all (incorporating multiple combustor types), wall-fired/PC/Other, and Fluidized Bed/Stoker/Dutch Oven/Other.

In the 1997 population dataset, some of the fuel categories did not contain the full array of combustor designs. Based on the 1997 data, and the six categories of solid fuel types, there were 8 defined categories that each combustion unit could be assigned to. The number of units in each category is shown on the following table.

<b>1997 Boiler Population</b>			
<b>Material</b>	<b>Combustor Type</b>		
	<b>Fluidized Bed/ Stoker/ Dutch Oven/ Other</b>	<b>Wall-fired/ PC</b>	<b>All</b>
Coal	1,778	634	<b>2,412</b>
Coal/Wood/NFF Liquid/NFF Solid			<b>173</b>
Gas/Wood/Other Biomass/Liquid FF			<b>302</b>
NFF Liquid/NFF Solid/Gas			<b>122</b>
Wood	638	54	<b>692</b>
Wood/Other Biomass/NFF Liquid/NFF Solid			<b>193</b>
<b>Total</b>	<b>2,416</b>	<b>688</b>	<b>3,894</b>

Please note that the variation in combustor design and total number of units assigned to each fuel/material category, and even the types of fuel/material categories amongst the new data to be collected under the first component of this ICR, could be strikingly different than the 1997 boiler and process heater population. Therefore, the Agency may use different variables of fuel/material and combustor design when designing the revised sample size. Several commenters requested that a facility should be given the opportunity to opt out of a test if it stops burning a fuel/material selected as part of the sample design. The goal of this testing program is to fill data gaps. If a facility elects to permanently stop burning a fuel/material identified as having a data gap, the Agency would not require a test at this facility.

### 3. Sample Size

One of the most important factors in sample design is that of determining how large a sample is needed for the estimates obtained in those selected samples (or units) to be statistically reliable enough to meet the objectives of the study. In the determination of sample sizes for studies where virtual certainty (i.e., a high level of reliability) is

needed, a level of 95 percent confidence, and 5% margin of error is established to assure the objectives of the study will be met. A simple stratified sample design determines the sample size (n) by the following Microsoft Excel Function:

$$n = 0.25 * \text{NORMSINV}(0.95 + (1 - 0.95) / 2)^2 / ((0.05^2) + (0.25 * \text{NORMSINV}(0.95 + (1 - 0.95) / 2)^2) / N)$$

Setting N = 3,894, since there are 3,894 solid fuel units in the 1997 boiler and process heater population databases, a 95 percent confidence in the sample size obtained, and a 5 percent margin of error, the sample size n = 350.

Given the eight categories from which units to be sampled can be selected, the units to be sampled can be selected in a couple different ways: i) equally (or relatively so) among the eight categories, or ii) proportional allocation of units to be sampled to stratified population (units within each category). In proportional allocation, the sampling fraction ( $n_h/N_h$ ) is specified to be the same for each category. The number of units ( $n_h$ ) taken from each category is given by:

$$n_h = \frac{(N_h)(n)}{N}$$

where  $N_h$  is the number of units in each category, n is the total number of units to be sampled (i.e., 350), and N is the total number of units (i.e., 3,894). Since the results of the electronic questionnaire will replace the 1997 boiler and process heater population, and these new data may have a different distribution of units within each category, or perhaps even more specific fuel/material type or combustor design categories, the actual number to be obtained from each category cannot be determined at this time. Further, stack tests will only be obtained for those combinations of fuel/material type and combustor design where there are gaps in available emission data.

Some commenters suggested that the number of facilities required to test should be a function of the number of unit subcategories the Agency anticipates. The number of facilities required to test will be based upon the identified emission data gaps for certain fuel and combustor design combinations. The subcategories for the rulemaking cannot be identified prior to the survey collection and testing program.

#### 4. Respondent Sample Collection

A random selection process will be used to determine a statistically significant number of candidates within each category to participate in this testing program. The Agency will then submit this list of candidates to stakeholders for review and comment on the technical feasibility and least cost impact of testing units within each candidate pool. Several members of industry commented that not all facilities have accessible testing ports, some facilities may be small entities, some units may vent to common stacks, and other facilities may be shutdown due to

seasonal operations during the testing timeframe. Given these practical constraints, it is not possible to conduct a truly “random” sample, however the Agency will ensure that the test sites ultimately selected will represent a statistically significant pool of candidate test sites, and will not allow for industry self-selection of individual test sites. If possible, once a unit from a particular site has been selected, no other units at that site will be chosen for that particular category. In addition, the Agency will work to ensure that if a company has multiple plants, no more than two plants owned by the same entity will be selected for testing. This approach will provide us with more information from a larger number of facilities given that all operations are not the same due to differing environmental conditions (e.g., weather), equipment, and load (e.g., fuel feed rates). Each site will also have a different mix of fuel (even within the same fuel category), since most facilities will obtain solid fuels from multiple sources. Testing at multiple plants (sites) will provide additional information on the variability of emissions across the mix of fuels/materials.

Commenters also requested that the Agency exclude from any testing requirements units that were already tested to support compliance with the vacated NESHAP standard. The Agency recognizes that the results of any test conducted under the vacated NESHAP standard will be submitted as part of the survey component of this ICR. The Agency is only seeking new test data where there are data gaps. If a facility has already submitted test data for a defined set of pollutants under the questionnaire component, they would not be required to re-test for those pollutants under the stack test component of this ICR. However, this facility may be required to test for other HAP or HAP surrogates if the unit at the facility belongs to one of the fuel/material and combustor type groupings with an identified data gap or for section 129 pollutants if a data gap is identified.

#### 5. Response Rates

Since the information will be requested pursuant to the authority of section 114 of the Act, EPA anticipates that all respondents requested to submit information will do so.