Technical Support Document: Revisions to the Air Emissions Reporting Requirements (Proposed Rule)

Office of Air Quality Planning and Standards

7/10/2023

Table of Contents

1 Overview

In support of the Air Emissions Reporting Requirements (AERR) proposed revisions, EPA performed several analyses to inform various technical aspects of the proposal. This document provides the additional technical detail for interested parties to better understand and obtain the underlying data used for the analyses.

Section 2 of this document provides an overview of the emissions data and risk data that EPA used, including descriptions of EPA's additional efforts on the emissions inventory data for identifying major sources and other analyses needed for this proposed rule. Section 3 documents the approaches that EPA used to establish the proposed emissions reporting thresholds for hazardous air pollutants (HAP). Section 4 provides a summary and reference materials that further illustrate EPA's approach to determine those industries to include for non-major sources in the proposed rule. Section 5 describes the approach EPA used to estimate the total number of facilities and micro-businesses that could be affected by the rule, which were needed for creating the Regulatory Impact Assessment (RIA) and draft Information Collection Request (ICR) for the proposed rule. The analysis in Section 5 relies heavily on approaches developed during the Small Business Advocacy Review (SBAR) Panel for this proposed rule. Finally, Section 6 provides EPA's analysis of those facilities for which EPA expects emissions reports would be collected under this proposed rule and their proximity to certain population demographics relative to the average national demographics.

This document references several Excel® workbooks with additional information about the data and steps described here. Some workbooks include a "Readme" worksheet, which describes the worksheets within the workbooks to help reviewers find data of interest. In addition, column headers within each worksheet may include comments (denoted by a purple triangle in the upper right corner of the spreadsheet cell). These comments provide explanations of the purpose for the column, and they can be viewed by using positioning the cursor over the cell.

2 Emissions and risk data

The analyses used to develop some aspects of the proposed AERR revisions relied on emissions data from the 2017 National Emissions Inventory (NEI), which has its own extensive Technical Support Document (TSD) available on EPA's website.[1] In addition, EPA relied on results from the 2017 AirToxScreen effort, which used the 2017 NEI as well as air quality modeling and risk information to

---

[1] US EPA, 2017 National Emissions Inventory Technical Support Document, February 2021, EPA-454/R-21-001, https://www.epa.gov/air-emissions-inventories/2017-national-emissions-inventory-nei-technical-support-document-tsd, also available in the docket for this proposal, EPA-HQ-OAR-2004-0489.

estimate risk at locations nationwide.[2] The 2017 AirToxScreen results also include an extensive TSD.[3] The EPA used the 2017 NEI because it was the latest triennial inventory available at the time the work was performed to draft the proposed revisions and the only inventory year (at the time) that included the new AirToxScreen product. Although the analysis was based on the 2017 data, EPA's analytical approach was designed so that conclusions from it are not limited to 2017, but rather are useful for drawing the conclusions needed for informing regulatory choices for the AERR in future years.

The 2017 NEI data are organized by data categories: point, nonpoint, onroad mobile, nonroad mobile, and events (fires). The analyses for the AERR proposal primarily focused on the point source data from the 2017 NEI. Point sources are those facilities from which state, local, and tribal (SLT) agencies collect emissions data and then report that data to EPA under the current AERR. Only certain point sources are required based on their emissions of criteria pollutants and precursors (CAPs), though states report many more facilities than they are required to report. The EPA uses these point source emissions from SLTs along with data from the Toxics Release Inventory (TRI) and other sources to compile the NEI. One analysis assessed SLT submissions of nonpoint data, as described in Section 2.3.

2.1   Identification of major sources within the 2017 NEI point source data

While the NEI database has the capability to document whether each facility is a major source or not, there is no requirement under the current AERR that states report the major source status of a facility. Additionally, certain facilities may be major sources only for HAP and do not meet the major source CAP emissions reporting thresholds. As a result, the NEI provides an incomplete list of Clean Air Act (CAA) major point sources. The major/non-major distinction is significant under this proposal development effort because the proposed AERR revisions for HAP reporting are different for major and non-major sources. Thus, to be able to estimate cost impacts of the proposed AERR revisions, EPA estimated which sources in the 2017 NEI were major sources.

To do this, EPA started with major source flags for facilities from three data sources: (1) ICIS Air, (2) a list of HAP Major sources from rulemakings, and (3) the major source labels from the point sources, stored in the "facility inventory" part of the Emissions Inventory System (EIS). Table 1 provides the list of datasets and fields EPA used to label facilities as CAP Major, HAP Major, and both HAP/CAP Major facilities. These datasets are provided as attachments to this TSD in the docket for the AERR proposal (see docket number EPA-HQ-OAR-2004-0489). The EPA combined these data files with an R markdown script and then used Microsoft® Access® to further manipulate the information and assign it to the 2017 NEI list of facilities.

---

[2] US EPA, 2017 AirToxScreen website, https://www.epa.gov/AirToxScreen/2017-airtoxscreen-assessment-results.

[3] US EPA, Technical Support Document: EPA's Air Toxics Screen Assessment – 2017 AirToxScreen TSD, March 2022, https://www.epa.gov/AirToxScreen/2017-airtoxscreen-technical-support-document, also available in the docket for the proposal EPA-HQ-OAR-2004-0489.

Table 1: Datasets used to determine facility major source status.

Data Source	Workbook and Worksheet Names	Data fields and usage

ICIS Air[a]	"ICIS Air poll_classif_eis_ids 21Oct2021.xlsx" Worksheet "Final ICIS-based Major list"	**PGM_SYS_ID,** usage:
		  This is the unique EIS facility identifier
		AIR_POLLUTANT_CLASS_CODE, usage:
		  CAP Major when field = "CAP MAJ"
		  HAP Major when field = "HAP MAJ"
		  HAP/CAP Major when field = "CAP/HAP MAJ"

Rulemaking data	"AERR Proposal TSD Supporting data.xlsx", Worksheet: "OAQPS Rule Data"	**EIS_ID_1**, usage: This is the unique EIS facility identifier All facilities were labeled as HAP Major based on emissions included in the worksheet, which may have been revised beyond data available in the 2017 NEI.

EIS	"AERR Proposal TSD Supporting data.xlsx", Worksheet: "EIS Categories"	**EIS_FACILTY_SITE_ID**, usage:
		  This is the unique EIS facility identifier
		FAC_CATEGORY_CD, usage:
		  CAP Major when field = "CAP"
		  HAP Major when field = "HAP"
		  HAP/CAP Major when field = "HAPCAP"

2017 NEI	"AERR Proposal TSD Supporting data.xlsx", Worksheet: "Major source results"	This file was created by using the 2017 NEI facility emissions totals, summing emissions for each pollutant and across all HAP and making the following assignments: - CAP Major for CAP pollutants when total emissions $\geq$ 100 tons/year[b] (but not HAP Major) - HAP Major for HAP total $\geq$ 25 tons/year or for an individual HAP $\geq$ 10 tons/year (but not CAP Major) - HAP/CAP Major when both CAP Major and HAP Major for a single facility

[a] The "SQL" worksheet (tab) in this workbook provides the SQL query used to pull the data from ISIS-Air.

[b] The CAP Major threshold is based on potential-to-emit rather than actual emissions; however, if a facility has actual emissions above the potential-to-emit threshold, then it can be reasonably flagged as also exceeding the potential-to-emit level that would be higher than the actual emissions level. In addition, lower thresholds for certain pollutants apply to facilities in nonattainment areas, but this aspect was not considered in this analysis.

The R-Markdown script used the "CAP Major", "HAP Major", and "Both Major" columns in the "Final ICIS-based Major list" worksheet by flagging facilities by EIS IDs from that worksheet when a "Y" appears in those three columns. The script also used the data provided in the "OAQPS Rule Data" tab by marking as a HAP Major source any facility based on the EIS_ID in this file that had a total of 25 tons or more emissions from the "Best_Total_HAP_Estimate_TPY" field or 10 tons or more emissions from the

"Best_Highest_Single_HAP_Estimate_TPY" field. For example, the facility with EIS ID 12682311 on row 11 of this worksheet was labeled as HAP Major by the script because the Best Highest Single HAP emissions from the rule data was 11.98 tons, which exceeded the 10 ton major source threshold. Similarly, the facility with EIS ID 863411 on row 20 of the worksheet was labeled as HAP Major because the Best Total HAP Estimate was 27.75 tons, which exceeds the 25 ton major source threshold for all pollutants. The script also used the categories from the EIS as shown in the "EIS Categories" worksheet by marking any records with "CAP" listed in the FAC_CATEGORY_CD field as CAP Major, any with "HAP as HAP Major, and any with "HAPCAP" as HAP/CAP Major.

EPA used the resulting labels from the R-Markdown script and the 2017 NEI to set final major source labels to facilities via Microsoft Access queries. The queries identified CAP Major, HAP Major, or HAP/CAP Major based on the 2017 NEI emissions levels using the major source definitions (*i.e.*, 25 tons of any HAP or 10 tons of a single HAP). The queries then combined these emissions-based designations with the R-Markdown results to set a final determination of major source status. When "CAP Major" was the only label across any of these datasets for a facility, then EPA labeled the 2017 NEI as CAP Major. Similarly, when "HAP Major" was the only label across any of these datasets, then EPA labeled the 2017 NEI facility as HAP Major. Finally, if a combination of labels appeared for "CAP Major", "HAP Major" or "HAP/CAP" Major across any dataset, then EPA labeled the 2017 NEI facility as HAP/CAP Major.

The "Major source results" tab provides resulting list of major source assignments based on each dataset individually as well as the overall determination. The "Emis-based Facility Category" field provides the results based on 2017 NEI emissions. The "ICIS Facility Category", "RTR_Major_Status" and "EIS Facility Category" shows the results from the R-Markdown script for the ICIS, OAQPS Rule data, and EIS HAP categories respectively. The "Final Facility Category" field provides the final determination that combines the results across all fields. EPA used these assignments for subsequent analyses for this proposal whenever the major source categorization or the major/non-major status was needed.

Historically in the AERR, EPA has estimated about 13,400 title V sources nationally (for CAP reporting). The analysis described above identifies only about 11,700 of them. Some may be missing from the NEI, but most are probably included and may be mislabeled as non-major sources. Title V experts within EPA also cite 14,300 as the number of title V sources, but because there is no single compilation across state and federal permitting programs, uncertainties exist about how many sources should be treated as major sources for purposes of this rule development process.

2.2 Analysis to discern impact of partial use of TRI data

As part of outreach for the Small Business Advocacy Review (SBAR) panel, the American Composites Manufacturers Association (ACMA) commented that in development its draft proposed emissions reporting thresholds, the EPA did not include all the data from the TRI in the NEI.

To assess the impact of EPA's incomplete use of TRI for augmenting NEI, EPA examined how much styrene data was not included from TRI for the North American Industry Classification System (NAICS) codes for industries associated with the ACMA. The EPA chose styrene as a proxy for illustration in this

document because it is a common pollutant with significant mass from many facilities for those NAICS codes.

Table 2: Comparison of 2017 NEI and 2017 TRI styrene emissions by NAICS associated with composites manufacturing.

	Facility counts		Emissions (tons/yr)	
NAICS code and description	**TRI**	**NEI[a]**	**TRI**	**NEI**
326122 - Plastics Pipe and Pipe Fitting Manufacturing	20	24 (50)	280	142
326130 - Laminated Plastics Plate, Sheet (except Packaging), and Shape Manufacturing	39	22 (46)	423	260
326191 - Plastics Plumbing Fixture Manufacturing	79	95 (100)	2,298	2,017
326199 - All Other Plastics Product Manufacturing	263	205 (431)	4,249	2,953
327991 - Cut Stone and Stone Product Manufacturing	6	11 (34)	49	61
336214 - Travel Trailer and Camper Manufacturing	14	14 (31)	247	315
336612 - Boat Building	109	144 (180)	3,716	3,605
Total	**530**	**515 (872)**	**11,263**	**9,953**

[a] The first number is the number with styrene emissions, and the number in parenthesis is the number with any pollutant.

The results of this comparison show that the total number of facilities for these NAICS between TRI and NEI is similar (530 vs. 515). For some NAICS, TRI has more facilities while for other NAICS, NEI has more. Overall, for these NAICS, the styrene average emissions per facility reported to 2017 TRI is 21.25 tons/year compared to the 18.16 tons/year for the 2017 NEI. The proposed threshold for styrene reporting for non-major sources is 10 tons/year (i.e., the default for major sources) due to the lack of dose-response information for styrene, so the level of the NEI emissions was not critical to setting the styrene threshold. However, this example shows that the NEI is not necessarily significantly less complete than TRI. Had EPA needed to rely on styrene levels in the NEI to develop a proposed emissions reporting threshold for styrene through risk results, a difference of 15 facilities out of 530 would not have been likely to impact those results. Further, when states report HAP data to NEI, that information is available by release point rather than as a facility total, which can be very impactful to risk estimates and any proposed emissions thresholds derived from them. Finally, since the threshold approach adjusts emissions to levels that would cause 1/million risk, the threshold approach depends not so much on the value of emissions, but rather on the locations of those facilities and their source release parameters from the NEI.

2.3   Evaluation of state submissions for 2017 nonpoint emissions

Section IV.A.J of the preamble to the proposed AERR revisions indicates that "For the 2017 NEI triennial inventory, EPA identified about 53,000 instances where state emissions data submissions were inconsistent with EPA's expectations and were, therefore, removed from the inventory."

This number reflects the number of county/process/pollutant combinations that EPA excluded from the 2017 NEI based on EPA quality assurance, despite having been submitted by SLTs. As described in the 2017 NEI TSD, EPA excluded these data using a process calling "tagging out," which prevents the data from being included in the final inventory. More information on this process is available in sections 2.2.6 and 2.2.7 of the 2017 NEI TSD, with information specific to nonpoint sources on page 4-3 of that document.

The report from the EIS that EPA used to estimate the 53,000 number based on tagging of state data for the 2017 NEI is available from the worksheet "Nonpoint tags example" within the "AERR Proposal TSD Supporting data.xlsx" workbook.

3 Approaches for HAP emissions reporting thresholds

The approach for defining risk-based emission thresholds for the proposed AERR relies on air quality modeling conducted for the 2017 AirToxScreen assessment. This section provides additional information on the risk-based modeling approach used for the proposed AERR that uses the AirToxScreen modeling results in a different way from the AirToxScreen assessment. EPA used this modeling approach to develop the proposed HAP emissions reporting thresholds in Table 1B to Appendix A of 40 CFR 51 Subpart A. This section also provides alternative approaches that EPA used to develop proposed HAP reporting thresholds for mercury, polycyclic organic matter (POM), and pollutants without a cancer unit risk estimate (URE) or a non-cancer reference concentration (RfC).

3.1 Risk-based thresholds for non-major sources

This section describes the approach taken to develop the proposed reporting thresholds for non-major sources and provides references to additional data tables that provide the data that EPA used to develop the proposed thresholds. First, EPA modeled air quality pollutant concentrations around facilities (major and non-major emissions sources) based upon the 2017 National Emissions Inventory (NEI) as part of the AirToxScreen assessment.

EPA used the model outputs from the 2017 AirToxScreen assessment to estimate ambient air concentrations of air toxics based upon EPA's American Meteorological Society/Environmental Protection Agency Regulatory Model (AERMOD). AERMOD is a steady-state plume model that incorporates air dispersion based on planetary boundary layer turbulence structure and scaling concepts.[4] AERMOD is EPA's preferred near-field modeling system of emissions for distances up to 50

---

[4] Cimorelli, A.J., Perry, S.G., Venkatram, A., Weil, J.C., Paine, R.J., Wilson, R.B., Lee, R.F., Peters, W.D. and Brode, R.W. 2005. AERMOD: A Dispersion Model for Industrial Source Applications. Part I: General Model Formulation and Boundary Layer Characterization. *Journal of Applied Meteorology,* 44: 682–693.

kilometers (km).[5] AERMOD (version 19191) was used for the modeling run with receptor elevations and hill heights being modeled with AERMAP (version 18081).

The 2017 AirToxScreen modeling that EPA used for the AERR proposal analysis was based upon (12 km x 12 km) gridded meteorological data for the contiguous United States (CONUS) and the three non-CONUS domains from the Weather Research and Forecasting Model (WRF), Advanced Research WRF (ARW) core[6] . The CONUS WRF modeling was based on the version 3.8 of WRF and the non-CONUS WRF modeling was based on version 3.9.9. The WRF Model is a state-of-the-science mesoscale numerical weather prediction system developed for both operational forecasting and atmospheric research applications.[7] For further information on the above modeling platforms and details on the modeling conducted for the 2017 AirToxScreen, refer to the TSD (reference provided above).

The 2017 AirToxScreen AERMOD runs utilized four types of receptors: 1) populated census block centroids (2010 Census data), 2) non-populated census block centroids (2010 Census data) in Alaska, 3) monitor locations, and 4) a network of gridded receptors within each CMAQ grid cell. An AERMOD run for a given facility could contain all four types of receptors or a combination of the four. Within the contiguous United States (CONUS), the gridded receptors' resolution varied between one and four kilometers. For CMAQ, grid cells located in Core Base Statistical Areas with populations of 1 million people or more (highly populated areas), the resolution was 1 km (shown in Figure 3-1), otherwise it was 4 km. For the non-CONUS areas, Alaska, Hawaii, and Puerto Rico and U.S. Virgin Islands, the gridded receptors' resolution was fixed at 3 km for Alaska and 1 km for Hawaii and Puerto Rico/U.S. Virgin Islands.

---

[5] EPA. 2017. Revisions to the Guideline on Air Quality Models: Enhancements to the AERMOD Dispersion Modeling System and Incorporation of Approaches to Address Ozone and Fine Particulate Matter. 82 Federal Register 10 (17 January 2017), pp. 5182-5231.

[6] Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Duda, M.G., Huang, X., Wang, W. and Powers, J.G. 2008. A Description of the Advanced Research WRF Version 3. Available online at http://www2.mmm.ucar.edu/wrf/users/pub-doc.html. Last updated 5 December 2014. Last accessed 16 December 2015.

[7] National Center for Atmospheric Research: (http://mmm.ucar.edu/models/wrf).

Figure 3-1: Example of receptor grid for modeling point sources within CONUS for highly populated areas.
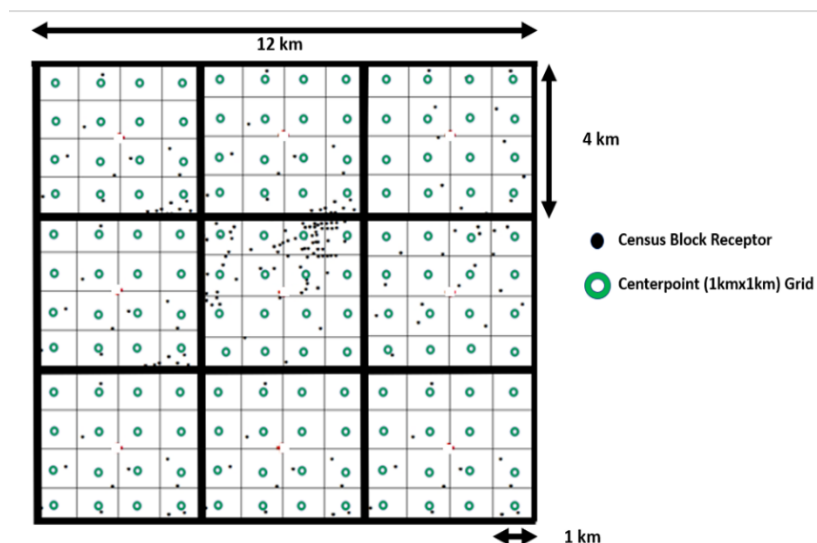
EPA post-processed the model outputs to utilize only air concentrations from the receptors that are no closer than 100 meters from each emission point within the facility. This 100-meter approach avoids overly high concentrations that can occur within the fence lines of facilities. As shown in Figure 3-2, about 65 percent of the distances between emission release points and the location of maximum total facility concentration was between 100 and 500 meters (and the remainder were farther away). The data also indicates that within the distance of 100 to 500 meters approximately 70 to 80 percent of the maximum modeled concentrations were at populated receptor locations at the census-block centroid.

The maximum concentrations no closer than 100 meters for those pollutants for which EPA has a URE or RfC are available in the "2017 Max Conc part 1" and "2017 Max Conc part 2" worksheets of the Microsoft Excel® workbook "AERR Proposal TSD Supporting data supplement.xlsx" available as an attachment to this TSD in the docket for the AERR proposal. The data are split into two tabs because the full dataset has more than 1.6 million records and is therefore too large to be provided in a single Excel worksheet (which supports about 1 million records per worksheet).

To keep supplementary Excel file smaller, the data included by facility has the facility identifiers but not other facility-level attributes. Reviewers of the spreadsheets can use the "2017 NEI facilities" worksheet of the "AERR Proposal TSD Supporting data.xlsx" workbook to see additional information about the facilities, including names, industry codes, addresses, and locations.

Figure 3-2: Receptor distance at maximum concentration location for modeled NEI facilities.
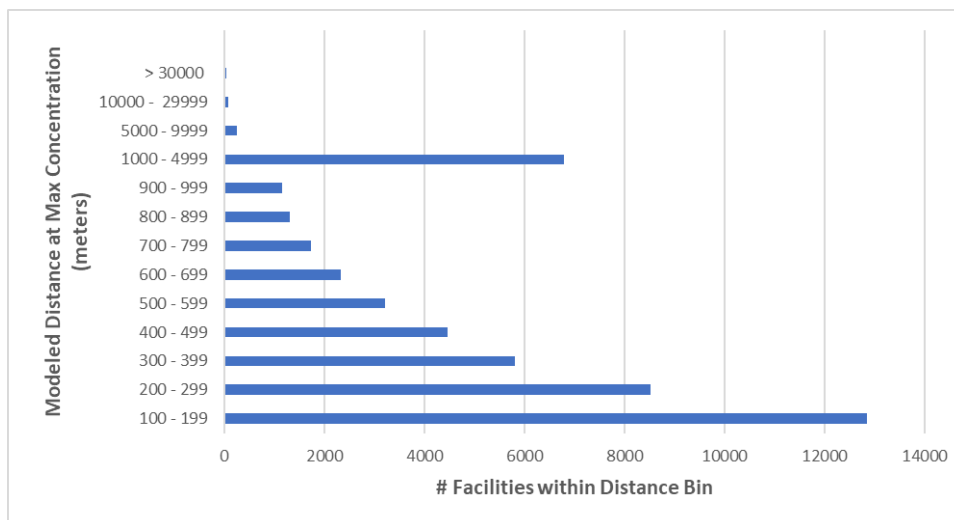
The EPA then used these concentrations to compute cancer risk estimates with pollutant-specific UREs and non-cancer impacts (*e.g.*, respiratory, neurological) based upon the RfC for the most sensitive organ system. Generally, the EPA used the same UREs and RfCs to calculate cancer risk and non-cancer hazard index (HI) EPA would use for regulatory actions[8] under Section 112(f) of the Clean Air Act (CAA) residual risk program. The URE and RfC values that EPA used are included in the "URE and RfC" worksheet (tab) of the workbook "AERR Proposal TSD Supporting data.xlsx."

For modeling purposes, EPA groups pollutants belonging to pollutant groups with the same URE or HI are grouped, since the results for these pollutants would be the same for a given emissions release point. To do this, EPA uses a single modeled pollutant to represent the pollutants in the group. EPA mapped the component pollutants emitted from each of the facility/release points in the 2017 NEI to the associated modeled pollutant. EPA grouped such pollutants by modeled pollutants using a crosswalk available in the "SMOKE-poll cross-walk" worksheet of the workbook "AERR Proposal TSD Supporting data.xlsx." The modeled pollutants are listed in the "smoke_name" column of the worksheet (column A). For example, the modeled pollutant named "PAH_176E2" represents Dibenzo[a,h]Pyrene (CAS 189640), Dibenzo[a,i]Pyrene (CAS 189559) and Dibenzo[a,l]Pyrene (CAS 191300). EPA used the sum of the component pollutants to calculate the emissions of the modeled pollutant, which through AERMOD modeling determines the locations of maximum facility concentrations for each grouped pollutant.

To guard against including release points and pollutants that contribute very minor impact to the overall facility risk, EPA excluded from subsequent analysis steps any release point/pollutant that contributed to less than 20 percent of the cancer risk or HI for that pollutant at the facility. Dropping the lower 20

---

[8] At the time of proposal, an IRIS health benchmark for cobalt is under development. For the purposes of this rule, EPA will rely on the available CalEPA health benchmark for insoluble cobalt to identify an emission threshold for data collection.

percent removes the smaller release points at a facility from the data, which guards against too low of a threshold given that the data used are incomplete and may be skewed by inconsistent voluntary reporting of lower emissions levels. Those release point/pollutant combinations that EPA retained for the threshold calculations are provided in the "Adj Emissions Results" worksheet of the workbook "AERR Proposal TSD Supporting data.xlsx." The percent contributions of each release point to cancer risk and HI are included in the table via the "FractionPolRisk" and "FractionPolHI" data fields, respectively. Whether or not the release point was the maximum risk or HI release point for the facility are also included in the table via the "MaxRiskRelPt?" and "MaxHIRelPt?" data fields, respectively.

Using the pollutant-specific cancer risk and HI estimates, EPA calculated the level of emissions ("adjusted emissions") that would be needed to cause one in a million risk and/or a 0.5 HI for each release point and HAP at all facilities in the 2017 data. This calculation is possible because the cancer risk and HI results from the modeling performed can be scaled linearly based on emissions. Rather than rely on a single facility or selected facilities, the approach provides for a distribution of emissions reporting thresholds to be explored so that EPA can ensure that emissions reporting thresholds are both robustly based on available data and not overly low causing undue burden. This approach allows for the large variety of stacks and fugitive releases with varied parameters to contribute to the information with which EPA could develop emissions reporting thresholds. The adjusted emissions values for each of the retained release points are also available in the same table/worksheet described in the previous paragraph. The adjusted emissions values are provided via the data fields "2017EmisRiskEq1 (tons)" field for cancer risk and the "2017EmisHIEqPt5 (tons)" field for HI.

For pollutants that are part of pollutant groups (e.g., the PAH_176E2 example above), EPA calculated adjusted emissions for the pollutant group rather than for the individual pollutants. Like ungrouped pollutants, EPA made this calculation for all release points that are no closer than 100 meters to the location of maximum concentration of the pollutant group for the facility. As a result, all pollutants associated with the pollutant group have the same value for the proposed emissions reporting threshold.

In making these adjusted emissions calculations, EPA excluded NEI emissions based on "HAP augmentation," whereby EPA estimates emissions of HAP based on emission factor ratios of HAP to a related CAP. These are excluded because they are among the least robust data in the NEI; therefore, EPA used only HAP data that were reported by states, that were reported to TRI, or that were included from rule development data.

The EPA evaluated the distributions of adjusted emissions data by using histograms, considering both the raw data and log-transformed data. The raw data are extremely skewed with a few high values and long tails, so EPA log-transformed the data. While a handful of the log-transformed histograms approximated a normal distribution, most of the distributions had a significant high value bias or low value bias. Because most histograms of the log data did not appear normally distributed, EPA has chosen not to use an approach that would rely on standard deviation from the median of adjusted emissions.

Example histograms from this analysis are available in the Excel workbook "AERR Proposal TSD Supporting data.xlsx" in the "Example Histograms" worksheet.

The EPA also evaluated using the median values of the adjusted emissions distributions to set an emissions reporting threshold, but these values were often several orders of magnitude higher than values estimated to cause significant risks based on the 2017 Air Toxics Data Update. For example, a median approach would have resulted in a threshold for hexavalent chromium of 0.0013 tons, whereas smaller levels of chromium than this included estimated cancer risks in the 2017 results from 47/million to 1/million for about 450 release points. One release point with hexavalent chromium emissions of 2.6E-5 tons/year contributed to a 2/million risk. Another example is benzidine for which the median-based threshold would have been 0.026, but the 2017 risk results at emissions as low as 0.00013 tons caused estimated cancer risks of 5/million. In reviewing the range of values from the adjusted emissions distributions, EPA determined that the 10th percentile of the adjusted emissions levels provided a reasonable reporting threshold. Percentiles below that level too often approached the minimum emissions levels causing risk in the 2017 Air Toxics Update, and percentiles above that level may not be rigorous enough to ensure that EPA collects sufficient data to be protective of human health. The median and 10th percentile adjusted emissions values, the latter of which were used to set the proposed emissions reporting thresholds, are available in the "Threshold results" tab of the "AERR Proposal TSD Supporting data.xlsx" workbook.

3.2   Mercury

It is important to ensure complete mercury reporting from sources because, in addition to using mercury data for risk analysis, EPA reports trends in total national mercury emissions based on international agreements such as the Minamata Convention on Mercury and the Convention on Long-Range Transboundary Air Pollution. The risk-based approach was insufficient for mercury compounds because they have multi-pathway (air, water, soil) effects that were not captured by the analysis described above. As described in the preamble for this proposed action, the EPA has proposed a reporting threshold of 0.0026 tons, which was set such that it would capture 95 percent of the mass of mercury nationally, based on the mercury data in the 2017 NEI.

The EPA used 2017 NEI facility total emissions of mercury to perform this calculation. These data are available in the "Mercury Threshold" worksheet of the "AERR Proposal TSD Supporting data.xlsx" workbook, included as an attachment to this TSD in the proposal docket.

3.3   POM

The risk-based HAP reporting threshold approach described in Section 3.1 only works when EPA has pollutant emissions data available in the 2017 NEI, but EPA was lacking data for some POM substances. POM include organic compounds with more than one benzene ring, which also have a boiling point greater than or equal to 100 degrees Celsius. POM represent a broad group of chemicals, including a subclass of POM pollutants called polycyclic aromatic hydrocarbons (PAHs). The AERR proposal does not provide an exhaustive list of POM HAP compounds but provides thresholds for general POM groups to suffice when a specific POM threshold is not provided.

Approximately 43 POM substances have a URE value, while EPA modeled only nine POM substances and thus developed risk-based thresholds only for these nine POM substances. To develop reporting thresholds for other HAP for which EPA does not have emissions data, EPA developed proposed reporting thresholds using a modified risk-based approach. To implement the approach, EPA started with the proposed AERR reporting thresholds for the nine POM substances shown in Table 3 that *were* included in the risk-based analysis ("modeled POM") and mapped these pollutants to "other POM" for which EPA does *not* have emissions data. To do this, EPA found the modeled POM with the closest URE to the other POM. Pollutants with the same URE as other pollutants received the same emissions threshold in the AERR proposal.

Table 3: AERR Proposed Emissions reporting thresholds for POM/PAH Substances based upon modeled NEI data and URE values

POM HAP	SMOKE name	CAS Number	Number 2017 NEI Facilities	Proposed AERR Threshold	URE $1/(\mu g/m^3)$
7,12-Dimethylbenz[a]Anthracene	PAH_114E1	57976	1,779	4.9E-05	0.1136
3-Methylchlolanthrene	PAH_101E2	56495	1,505	4.7E-04	0.01008
Dibenzo[a,h]Pyrene	PAH_176E2	189640	6	0.0011	0.0096
7H-Dibenzo[c,g]carbazole	PAH_176E3	194592	4,694	0.0028	9.6E-04
Benz[a]Anthracene	PAH_176E4	56553	8,539	0.028	9.6E-05
Benzo[k]Fluoranthene	PAH_176E5	207089	8,814	0.31	9.6E-06
Acenaphthene	PAH_880E5	83329	16,258	0.027	4.8E-05
Coal Tar	PAH_192E3	8007452	29	0.0035	9.9E-04

For POM substances without emissions data being reported to the NEI and that have a URE, a proposed AERR emissions reporting threshold was determined based upon those POM pollutants that were modeled with the closest URE. The higher the URE, the higher the cancer risk, resulting in a lower proposed AERR emissions reporting threshold. This approach was taken for 6-Nitrochrysene and 2-Aminoanthraquinone.

For 6-Nitrochrysene, which has a URE of 0.0096 $1/(\mu g/m^3)$, the closest URE for a POM with reported emissions is Dibenzo(a,h)Pyrene with the same URE. As a result, the proposed AERR emissions reporting threshold of 0.0011 tons/year for 6-Nitrochrysene is the same as for Dibenzo(a,h)Pyrene.

For 2-Aminoanthraquinone, which has a URE of 1.5E-5 $1/(\mu g/m^3)$, the closest URE of the modeled POM is for benzo(k)fluoroanthene. As a result, the proposed AERR emissions reporting threshold for 2-Aminoanthraquinone has been calculated by linearly scaling the 0.31 tons/year proposed emissions reporting threshold from benzo(k)fluoroanthene. The calculation for the resulting threshold of 0.20 tons/year is shown below, which includes a rounding to two significant figures that has been taken for all proposed emissions reporting thresholds:

Threshold for 2-Aminoanthraquinone = (0.31 tons/year) $\times$ (9.6E-6/1.5E-5)

= 0.198 tons/year or 0.20 tons/year

The PAH group "Polycyclic aromatic compounds (includes 25 specific compounds)" with pollutant code "N590" has a URE and is mapped to the modeled pollutant Acenaphthene, so the proposed emissions reporting threshold for N590 is 0.027 tons/year. Rather than have different thresholds for the various grouped POM and PAHs, EPA is proposing to set the emissions reporting threshold for "PAH/POM – Unspecified" (CAS) and "PAH, total" (CAS) to the 0.027 tons/year level. Likewise, for specific PAHs without UREs, we defaulted those thresholds to the 0.027 tons/year.

Finally, there are many other specific POM pollutants outside of the PAH group for which no URE is available. These pollutants would receive the default threshold as described in Section 3.4 below.

3.4 Default emissions reporting threshold approach for certain pollutants

The EPA considered how to set a default emissions reporting threshold for all remaining pollutants without an URE or RfC.[9] Without risk data to use to inform such an approach, EPA has proposed to use the major source threshold of 10 tons/year for a single pollutant.

Another default approach was necessary for pollutants *with* UREs or an RfC but that have not been included in the 2017 NEI. Since EPA did not have data for these pollutants, the modeling and risk results used to set thresholds for other pollutants were not available. Without modeling and risk results to use to define a threshold, the EPA is also proposing the 10 tons/year threshold based on the major source threshold for a single pollutant.

3.5 Uncertainties

The following list provides the known uncertainties from the analyses described in Section 3.

- The analysis leading to the proposed emissions reporting thresholds does not include facilities that are not included in the NEI; consequently, we know that we are missing many facilities. This statement is supported by other analyses described in Section 5 below. Despite not having all sources, EPA has a wide span of emissions data values and types of facilities because many states report far more sources than they are required to report. As a result, the analysis approach for using percentile values to set emissions reporting thresholds is valid, even if in some cases, more data could be available if data from more sources were available. Given that

---

[9] For assessments of HAP, EPA generally use UREs from EPA's Integrated Risk Information System (IRIS). For carcinogenic pollutants without IRIS values, we look to other reputable sources of cancer dose-response values, often using California EPA (CalEPA) UREs, where available. In cases where new, scientifically credible dose-response values have been developed in a manner consistent with EPA guidelines and have undergone a peer review process like that used by EPA, we may use such dose-response values in place of, or in addition to, other values, if appropriate.

EPA has enough data to sufficiently represent all industries covered by the proposal, it is reasonable that EPA is proceeding with the data that it has available.

- Most of the HAP emissions data in the NEI are based on emission factors that may be out of date because that is the only data available to states (and sources that report to states). Outdated emission factors could negatively impact the quality of the proposed emissions reporting thresholds. However, such data is still the best available data that is practical to use and is consistent with the 2017 Air Toxics Data Update information and other HAP-related efforts by EPA and states.

4 Analysis and selection of proposed NAICS for non-major sources

For non-major sources, EPA started with a presumed list of the "industrial" NAICS that start with 2 or 3, because the sectors that EPA currently regulates for HAP are often associated with those NAICS. To consider what additional NAICS to add, EPA analyzed the available 2017 NEI HAP emissions data to assess the point sources contribution for each pollutant by NAICS code. EPA used all data sources from the 2017 NEI, including HAP augmentation for this analysis. The EPA applied a threshold of 1 percent contribution by NAICS grouped to the first 4 digits of the NAICS code for each pollutant. The 1 percent threshold was set as a conservative approach to identify NAICS-pollutant combinations for consideration and further review for possible inclusion in the proposal. The initial list of 4-digit NAICS including HAP pollutant contributions to the total pollutant based on the 2017 NEI is available in the worksheet "Initial NAICS4-Poll list" of the workbook "NAICS analysis for non-major.xlsx," which is available as an attachment to the TSD in the docket for the AERR proposal.

As an initial screen, EPA considered which 4-digit NAICS should be excluded for various factors, which are listed in the "Initial Screen" column of the "Initial NAICS4-Poll list. Some NAICS4 groups were excluded because they are agricultural activities (1111, 1113, and 1151), choosing instead to limit proposed reporting to major sources for these NAICS. EPA also excluded the NAICS4 groups for Specialized Freight Trucking (4842) and Services to Buildings and Dwellings (5617) because company activities for these industrials are dispersed widely given the nature of the work, and as such, do not lend themselves to emissions collection for an individual facility. Finally, EPA excluded the Support Activities for Rail Transportation (4882) because rail emissions are covered in the proposed AERR through a different approach than collection of emissions data from these companies.

To review the remaining NAICS4 groups, EPA mapped the selected NAICS4 to the full NAICS and included several other data fields. The results of this work are provided in the "NAICS adds, keeps, drops" worksheet of the "NAICS analysis for non-major.xlsx" workbook. As part of this analysis, EPA included the number of facilities for each NAICS in the 2017 NEI and information about the number of small businesses and total number of businesses based on the analysis approach described in Section 5 below. For each NAICS under consideration, the worksheet provides the number of facilities in the 2017 NEI, the number of 2017 NEI facilities with HAP, and the estimated number of small and not-small businesses based on data from the U.S. Economic Census and the Small Business Administration (SBA) definition of

14

small businesses.[10] The "Decision for Proposal" column of the "NAICS adds, keeps, drops" worksheet provides the rational for dropping or adding NAICS.

The EPA further excluded NAICS using its judgement to ensure a manageable number of facilities for which EPA can handle collection and that are focused, as explained in the worksheet for each NAICS, on those sectors making a significant contribution to individual pollutants balanced with the number and/or concentrated nature of the activities for the NAICS. The EPA excluded NAICS based on a variety of rationales, including:

- The NAICS is for a company that has its emissions at numerous locations rather than a single facility (*e.g.*, construction and contractors (236xxx and 237xxx) and certain waste collection industries (5621xx);
- The NAICS is not currently widely reported as point sources by states because there are many small disperse sources such as retail establishments or for other reasons (*e.g.*, retail bakeries (311811), building supply stores (4441xx), gasoline stations (4471xx), offices of physicians (6211xx), and certain laundry facilities (8123xx except for industrial launderers);
- The NAICS is in an agricultural production sector more likely to contribute many small sources that would better be captured as nonpoint emissions (e.g., agricultural product merchant wholesalers (4245xx);
- The NAICS has a small contribution to total emissions of any pollutant and there is a high proportion of small businesses and/or a large number of businesses in the industry that would make collection of such data too burdensome relative to the impact on the data (*e.g.*, various merchant wholesalers (4233xx and 4239xx), plastics and related wholesalers (4246xx), fuel dealers and other direct selling establishments (4543xx), testing labs (541380), and veterinary services (541940)); and
- The NAICS is in a service sector that is not expected to include significant pollutant emissions (*e.g.*, marketing research and public opinion polling (541910) and translation and interpretation services (541930).

In two cases, EPA added NAICS because of its analysis of the results. First, EPA added some air transportation industries (481211, 481212, and 481219) because the related NAICS (starting with 4881) was also included and the available NEI data showed that both sets of NAICS are used to identify the facilities that are operating just outside of airports in support of airport activities. While these NAICS are also used to identify airports, the proposed approach for airport emissions is different, but would not include activities at facilities co-located with airports that are performing support activities for the airports. Second, EPA added specialty hospitals (622310) because of the known use of ethylene oxide (EtO) sterilizers associated with other hospitals included based on EtO emissions in the 2017 NEI, but a lack of any data in NEI for these specialty hospitals. Including this NAICS in the AERR requirements would

---

[10] U.S. Small Business Administration, table of size standards, https://www.sba.gov/document/support-table-size-standards, Excel® file "Table of Size Standards Effective July 14 2022.xlsx".

ensure that any significant emissions would be captured based on the proposed risk-based emissions reporting thresholds.

5   Estimated number of facilities

EPA has estimated the number of facilities, small businesses, and micro-businesses that would potentially be affected by the proposed AERR revisions. These efforts have included estimates of small businesses using both the small business size standards from the SBA and a threshold of 100 employees to approximate the CAA definition of small businesses. EPA used the NAICS codes identified in Section 4 along with data from the 2017 County Business Patterns and Economic Census and the small business thresholds provided by the SBA. The original data sources were:

- 2017 NEI
- Small business size standards from the SBA (see footnote 10)
- Total US count of businesses by NAICS and number of employee size groups[11]
- Total US count of businesses by NAICS and receipts size groups[12]

The EPA took the following steps and made the following assumptions to estimate the number of facilities, small businesses, and micro-businesses for each NAICS. All Excel references in this section are to the workbook "Facility Count analysis for ICR Draft.xlsx," provided as an attachment to this TSD in the docket for the AERR proposal. That workbook is referenced in the remainder of this section as "the workbook." EPA has removed the formulas from the workbook that implement the approaches described in this section (*e.g.*, looking up values in tables, determining percentiles from a list of values) and has included the resulting values instead. This was done because the workbook runs very slowly with the formulas included. The workbook provided will more readily facilitate stakeholder review of the data that EPA used in developing proposed emissions reporting thresholds, but the original Excel file with formulas is also provided in the as an attachment to this TSD in the proposal docket via the file "Facility Count analysis for ICR Draft with formulas.xlsx."

5.1   Prepare the 2017 NEI data for processing

From the 2017 NEI, EPA selected all facilities except those with NAICS starting with 9 (because the economic census data does not include public sector facilities). EPA also does not include airports because airports would be captured via a different approach in the proposed rule and there are more than 30,000 airports in the NEI that would have cluttered the data tables. We additionally excluded any pollutants for which there are fewer than three facilities with that pollutant because a sample size of three facilities for a pollutant/NAICS combination provides a lower confidence that the pollutant is

---

[11] US Census Bureau, 2017 SUSB Annual Data Tables by Establishment Industry, https://www.census.gov/data/tables/2017/econ/susb/2017-susb-annual.html, May 2021, Excel file "us_state_naics_detailedsizes_2017.xlsx".

[12] Same web page as previous footnote, May 2021, Excel file "us_6digitnaics_rcptsize_2017.xlsx".

routinely associated with a given NAICS. The results of this 2017 NEI extraction are available in the "Raw data" worksheet of the workbook.

EPA also adjusted any 5-digit NAICS to the best available 6-digit NAICS, since the Economic Census data are available by 6-digit NAICS. EPA used the workbook to make these adjustments using the information provided in the "NAICS comparison" worksheet. The NAICS revised" column of the "Raw data" worksheet provides the resulting adjusted NAICS.

5.2   Prepare economic census data for processing

EPA Used the enterprise[13] size groups from the 2017 economic census to set the next number of employees or receipts for each size group in the Economic Census data. For example:
For employees, enterprise size = "06: <20 employees" would be 20

- For employees, enterprise size = "17: 300-399 employees" would be 400
- For receipts, enterprise size = "05: 1,000-2,499" would be 2,500.

This resulting number of employees to use for assessing whether a size group should be included in any count of small businesses is provided by the "category minimum" field of the "5 Calc by NAICS per size group" worksheet. This "category minimum" is the only field used to determine whether an enterprise size group should be associated with the SBA small business definition. If the "category minimum" is less than or equal to the size standard, then the enterprise size group is labeled as meeting the SBA small business definition.

Then, EPA mapped the small business size standard to the nearest enterprise for each enterprise size group to assign whether each enterprise size group from the economic census data would be counted as a "small business" or not. The "small business threshold" column of the "5 Calc by NAICS per size group" worksheet provides the SBA small business threshold. The "Small business flag" column of the same worksheet provides the status as "small" or "not small" based on the SBA small business threshold.

As an example, for the NAICS 211120 (Crude Petroleum Extraction), the following results come from the "5 Calc by NAICS per size group" worksheet. First, the SBA size standard for this NAICS is 1,250 employees, and the nearest enterprise size group to the SBA size threshold is 1,500 employees. To estimate the number of small firms, EPA counts all firms as small for size groups less than 1,500 employees (for SBA small) or less than 100 employees (CAA small).

---

[13] As defined by the Census Bureau: "An enterprise (or "company") is a business organization consisting of one or more domestic establishments that were specified under common ownership or control. The enterprise and the establishment are the same for single-establishment firms. Each multi-establishment company forms one enterprise - the enterprise employment and annual payroll are summed from the associated establishments". This definition is found at https://www.census.gov/programs-surveys/susb/about/glossary.html."

Table 4: Example assignments of small status to the Economic Census data, for NAICS = 211120.

Enterprise Size	Firms	Establishments	SBA Small	CAA Small
02: <5 employees	3,195	3,196	Small	Y
03: 5-9 employees	611	619	Small	Y
04:10-14 employees	208	216	Small	Y
05: 15-19 employees	98	106	Small	Y
07: 20-24 employees	58	62	Small	Y
08: 25-29 employees	53	60	Small	Y
09: 30-34 employees	37	43	Small	Y
10: 35-39 employees	34	39	Small	Y
11: 40-49 employees	39	53	Small	Y
12: 50-74 employees	49	73	Small	Y
13: 75-99 employees	23	36	Small	Y
14: 100-149 employees	25	43	Small	
15: 150-199 employees	17	27	Small	
16: 200-299 employees	28	58	Small	
17: 300-399 employees	4	23	Small	
18: 400-499 employees	9	25	Small	
20: 500-749 employees	17	63	Small	
21: 750-999 employees	8	36	Small	
22: 1,000-1,499 employees	11	128	Small	
23: 1,500-1,999 employees	7	44	Not small	
24: 2,000-2,499 employees	7	27	Not small	
25: 2,500-4,999 employees	15	114	Not small	
26: 5,000+ employees	17	242	Not small	
Total small establishments:			4,906	4,503
Total not small establishments:			427	830

EPA additionally considered, and ultimately is proposing, to use the Clean Air Act small business threshold as defined by CAA section 507(c). To approximate this, EPA assumed that any business tallied in the economic census with fewer than 100 employees would be considered "small." In the "5 Calc by NAICS per size group" worksheet, EPA set enterprise size groups as "CAA small" if the "category maximum" was less than 100 or, in the case of NAICS associated with a receipt threshold for the SBA definition, if the total employment of the enterprise group divided by the number of firms in that enterprise size group was less than 100. The resulting CAA small business designation for each enterprise size group is provided in the "CAA small business guess?" column.

Using the number of firms for each (from the "Firms" column) along with the small or not-small designations in the "Small business flag" and "CAA small business guess?" columns, the "5 Calc by NAICS per size group" worksheet allows EPA to add up the number of small businesses and total number of businesses. Similarly, the "Establishments" column, which represents individual facilities associated with

the firms, can be used to estimate the number of facilities associated with each NAICS. Such summation is accomplished using the "6 Results Pivot" worksheet.

The EPA assumes in this approach that the enterprise size thresholds are sufficient for comparing to the SBA small business thresholds. Since the "enterprise" as defined in the business census data are (by definition) larger than the establishments, some of the employees and receipts of an enterprise may not be a part of the establishments associated with a particular NAICS. However, this assumption is reasonable because most of the small businesses are not part of larger enterprises.

In addition, when a small business threshold maps exactly to an enterprise size threshold, a slight underestimate could occur, but this is preferable than a large overestimate. For example, for NAICS 212321 (Construction Sand and Gravel Mining), the small business threshold is 500 employees, and the Enterprise size range is "19: <500 employees". In this case, the small business threshold is mapped to this enterprise size, which does not include any firms with exactly 500 employees associated with the next higher enterprise size group of "20: 500-749 employees." This approach avoids including all firms in the 500-749 employee range of this example as being 500 employees or fewer.

5.3 Methodology for estimating number of small businesses potentially affected by HAP thresholds

The EPA has proposed HAP emissions reporting for major sources and for non-major sources that meet proposed emissions reporting thresholds. As just described above, the 2017 Economic Census data provides counts of firms based on employees or receipts and by using small business definitions from SBA and a proxy of fewer than 100 employees, the EPA can estimate the total number of small businesses for each NAICS for the two small business criteria. The challenge is how to connect the number of small businesses to the emissions thresholds.

To address this challenge, EPA developed an approach to estimate emissions/employee and emissions/receipts for each NAICS code, using the available 2017 NEI data. The EPA then used these data to estimate the number of employees or the receipts that would trigger reporting based on the HAP reporting thresholds. By mapping these numbers to the Economic Census data and associated small business thresholds, the EPA created reasonable estimates of the number of small businesses that could potentially be affected by the proposed AERR revised requirements. The results of this approach are available in the "Final results for ICR& preamble" worksheet of the workbook.

5.3.1 Step 1: Map Economic Census data on number of establishments by employees and by receipts to NEI facilities

To determine an emissions/employee or emissions/receipts value, EPA estimated the number of employees or the receipts for each facility in the 2017 NEI, because this information is not included in the NEI. The approach assumes that the facilities with larger emissions in the NEI have larger numbers of employees or larger receipts, within groups of facilities at the 4-digit NAICS level (NAICS4). EPA used the NAICS4 groups because all NEI facilities have at least NAICS4 codes, which was the minimum detail of NAICS required for reporting to the NEI for the 2017 NEI.

This step involves two parts. First, EPA ranked the NEI facilities based on emissions and second, EPA mapped the NEI facilities to the establishment counts in the Economic Census data by assuming the largest NEI facilities should map to the largest average number of employees/establishment and receipts/establishment.

To determine the largest NEI facilities for this purpose, the EPA used the sum of: 1) carbon monoxide (CO) as a reasonable surrogate for combustion activities, 2) volatile organic compounds (VOC) as a reasonable surrogate for solvent and other evaporative activities (after adjustment, see below), and 3) primary particulate matter with diameter less than or equal to 2.5 microns (PM25-PRI) as a reasonable surrogate for dust-generating activities. These types of activities were targeted because they capture those activities that influence a facility's size. To account for certain VOC controls, the EPA summation described above adjusted VOC. To do this, EPA assumed that VOC emissions from solvent processes from point sources are on average controlled by 85% and used this assumption to calculate adjusted VOC and VOC HAP values.

$$\text{Adjusted VOC} = \text{NEI VOC} / (1-0.85)$$

$$\text{Adjusted VOC HAP} = \text{NEI VOC HAP} / (1-0.85)$$

To account for facilities with missing or under-reported VOC, EPA compared facility total adjusted VOC HAP to facility total adjusted VOC emissions and used the greater of the two. Similarly, EPA compared facility total PM HAP to the PM25-PRI and used the greater of the two. Finally, if no CO data were present at a facility, EPA included emissions from the acid gases Hydrochloric Acid and Hydrogen Fluoride instead of the CO. This approach allowed EPA to rank all facilities in the NEI for the relevant NAICS4 groups.

With the first part of this step completed, EPA next mapped the largest facilities to the largest average number of employees. The worksheet "1 Facility Assign Census data" of the workbook accomplishes this step. This part included computing average employees/establishment and average receipts/establishment for each Enterprise size provided in the 2017 Economic Census. In the first case, EPA divided the total number of employees by the total number of establishments for each of the employee size groups and rounded down. Rounding is needed because fractions of employees cannot exist and rounding down is conservative for this analysis (i.e., a lower employee number will increase the emissions/employee values). In the second case, EPA divided the total receipts by the total number of establishments for each of the receipts size groups. For each NAICS4 group, EPA then sorted the enterprise size groups by descending employees/establishment (see the "SBA_Employment_data" worksheet) and separately by descending receipts/establishment (see the "SBA_Receipt_data" worksheet). The sorting was needed to facilitate the next part of the analysis.

Table 5 below provides an example of mapping between the NEI facilities and the 2017 Economic Census employee data for NAICS4 2111 (Oil and Gas Extraction), which is a group of NAICS that uses an SBA small business threshold based on number of employees. In Table 5, the enterprise sizes are ranked based on employees/establishment as they are in the "SBA_Employment_data" worksheet. The largest

employees/establishment is 131 followed by 107 then 81, etc. Table 6 shows that the largest 25 NEI facilities are assigned 131 employees, while the next largest 65 facilities are assigned 107 employees and so forth. Table 6 is a summary of the information provided in the "1 Facility Assign SBAdata" worksheet.

Because of the difference between enterprises and establishments, larger enterprise sizes have employee size ranges that reflect all their employees over the entire firm, which might span multiple NAICS. This leads to a feature in Table 5 that appears to be an inconsistency. For example, the first row shows that the establishments with the largest employee/establishment for this NAICS4 have 131 employees on average and are owned by enterprises with 400-499 employees. The apparent inconsistency is that 131 is not within the 400-499 range. This can be explained, however, because the other employees for the enterprise are associated with other NAICS4 (a single enterprise can have operations across many industries) and are not included in that row of data. Those other employees are also not relevant for the per-NAICS and per-establishment calculations done to match with NEI facilities.

Table 5: Selected Economic Census data for estimating number of employees per facility, NAICS4=2111

NAICS4	Enterprise Size	No. of Firms	No. of Establishments	Employment	Employees / Establishment (rounded down)
2111	18: 400-499 employees	10	26	3,431	131
2111	24: 2,000-2,499 employees	9	65	6,956	107
2111	21: 750-999 employees	11	49	4,005	81
2111	16: 200-299 employees	35	95	6,957	73
2111	25: 2,500-4,999 employees	16	233	14,893	63
2111	20: 500-749 employees	21	96	5,930	61
2111	15: 150-199 employees	26	47	2,879	61
2111	14: 100-149 employees	32	54	3,054	56
2111	22: 1,000-1,499 employees	12	144	7,834	54
2111	26: 5,000+ employees	22	401	20,862	52
2111	17: 300-399 employees	5	24	961	40
2111	13: 75-99 employees	34	57	2,215	38
2111	23: 1,500-1,999 employees	8	105	3,782	36
2111	12: 50-74 employees	55	91	2,646	29
2111	10: 35-39 employees	36	41	1,206	29
2111	11: 40-49 employees	49	68	1,936	28
2111	09: 30-34 employees	46	57	1,228	21
2111	08: 25-29 employees	59	72	1,521	21
2111	07: 20-24 employees	70	77	1,485	19
2111	05: 15-19 employees	120	131	1,941	14
2111	04:10-14 employees	249	260	2,855	10
2111	03: 5-9 employees	723	734	4,672	6
2111	02: <5 employees	3,473	3,478	5,736	1

Table 6: Mapping NEI facilities to average employee per establishment, NAICS4 = 2111

2017 NEI		**Enterprise Size Group (employees)**	**No. Establishments**	**Employment**	**Employee/ Establishment**
Largest 25 facilities	maps to →	400-499	25	3,008	131
Next largest 65 facilities	maps to →	2,000-2,499	65	6,956	107
Next largest 49 facilities	maps to →	750-999	49	4,005	81
…					

The example below provides a similar example of mapping between the NEI and the 2017 Economic Census *receipts* data for NAICS4 4931 (Warehousing and Storage), which is a group of NAICS with a small business threshold based on receipts. Table 7 provides the receipts per establishment for this NAICS4, sorted from high to low as they are in the "SBA_Receipts_data" worksheet. Table 8 shows that the largest 274 NEI facilities are assigned receipts of $4,596K, while the next largest 386 facilities are assigned receipts of $4,451K and so forth.

Table 7: Selected Economic Census data for estimating receipts per facility, NAICS4=4931

NAICS4	**Enterprise Size ($1000)**	**Firms**	**No. of Establishments**	**Receipts ($1000)**	**Receipts/ Establishment ($1000)**
4931	11: 20,000-24,999	148	274	$1,259,265	$4,596
4931	16: 50,000-74,999	202	386	$1,718,150	$4,451
4931	13: 30,000-34,999	117	197	$800,612	$4,064
4931	14: 35,000-39,999	105	184	$694,442	$3,774
4931	09: 10,000-14,999	296	463	$1,714,938	$3,704
4931	17: 75,000-99,999	150	523	$1,929,154	$3,689
4931	15: 40,000-49,999	113	283	$1,034,199	$3,654
4931	08: 7,500-9,999	214	299	$1,073,073	$3,589
4931	12: 25,000-29,999	101	180	$640,924	$3,561
4931	10: 15,000-19,999	200	364	$1,185,637	$3,257
4931	07: 5,000-7,499	386	528	$1,683,695	$3,189
4931	06: 2,500-4,999	652	763	$1,874,110	$2,456
4931	18: 100,000+	1,416	8,402	$20,034,826	$2,385
4931	05: 1,000-2,499	1,094	1,171	$1,634,882	$1,396
4931	04: 500-999	835	857	$576,617	$673
4931	03: 100-499	1,421	1,425	$370,827	$260
4931	02: <100	594	602	$27,106	$45

Table 8: Mapping NEI facilities to average receipts per establishment, NAICS4=4931

2017 NEI		Enterprise Size Group ($1000)	No. Establishments	Receipts ($1,000)	Receipts ($1000)/ Establishment
Largest 274 facilities	maps to →	20,000-24,999	274	$1,259,265	$4,596
Next largest 386 facilities	maps to →	50,000-74,999	386	$1,718,150	$4,451
Next largest 197 facilities	maps to →	30,000-34,999	197	$800,612	$4,064
…					

5.3.2 Step 2: Across all facilities by NAICS4 and HAP calculate median emissions per employee and per receipts

At this step of the analysis, each facility in the NEI has been assigned an assumed employee count and a total annual receipts value. These are listed in the "1 Facility Assign SBAdata" worksheet. Each facility also has its own level of emissions for whichever pollutants have been compiled for that facility from the 2017 NEI. The EPA dropped all facility/pollutant records with zero emissions from the analysis. Using this information, EPA computed emissions/employee and emissions/receipts for each facility and pollutant (see "2_Sort_Faciliies_by_Emis" worksheet). Next, the EPA grouped these results by NAICS4 and across all facilities and pollutants at the NAICS4 level, identified the median emissions/employee and emissions/receipts. Table 9 below shows 8 pollutants for NAICS4 2111 (Oil and Gas Extraction) in no significant order, out of a total 60 pollutants in the NEI for that NAICS4 group, and the table provides the number of facilities for that NAICS4 and each pollutant. The information in this table can be found in the "3_Thresholds_median" tab of the workbook for all NAICS4 groups and pollutants from the 2017 NEI.

Table 9: Example median emissions/employee and emissions/receipt by pollutant, NAICS4 = 2111

NAICS4	Pollutant	No. Facilities	Median emissions (lb)/employee	Median emissions (lb)/receipts ($1000)
2111	Glycol Ethers	6	4.49E-04	1.60E-05
2111	Formaldehyde	2,972	3.36E-02	2.65E-04
2111	Benzo[a]pyrene	288	1.94E-09	1.83E-11
2111	Dibenzo[a,h]anthracene	100	5.97E-10	8.87E-12
2111	Carbon Tetrachloride	404	5.98E-05	2.00E-07
2111	3-Methylcholanthrene	80	5.79E-10	6.37E-12
2111	Benz[a]anthracene	435	3.08E-08	3.64E-10
2111	7,12-Dimethylbenz[a]anthracene	82	5.09E-09	5.19E-11
…				

Step 3: Calculate the employees and receipts that would trigger reporting based on HAP thresholds for each NAICS4.

Next, for each full NAICS, the EPA identified the pollutants present (which could be fewer pollutants than for the NAICS4 group). For these pollutants, EPA used the formula below to combine the median emissions/employee or emissions/receipt with the HAP threshold to compute the employee number and receipts amounts that would trigger reporting by an establishment.

$$Trigger = \underset{P1 \rightarrow Pn}{\text{Min}} \ (\text{HAP threshold / median rate})$$

For the employee-based trigger, the number of employees is rounded up to the next whole number. The resulting trigger values are available in the "3_Thresholds_median" worksheet I the columns labeled "Employee trigger w/Median" and "$1000 Trigger w/Median." The minimum trigger in the formula above is identified over all pollutants P1 through *Pn* for the NAICS, which is shown in the "4_Helper_data" worksheet, accomplished via sorting those records based on the NAICS and trigger values.

Table 10 below shows the resulting lowest 10 median employee thresholds for NAICS 211120 (Crude Petroleum Extraction). The employee-based thresholds are shown because small businesses for this NAICS are defined by the number of employees (i.e., 1,250 employees). The number of facilities in this table are for the more detailed full NAICS (as compared to the NAICS4 group of the previous table). The trigger in this example is 3 employees, which has been computed based on the proposed formaldehyde threshold 0.083 tons (166 lbs) divided by the median emissions rate for formaldehyde from facilities within NAICS4 2111 of 0.0336 lbs/employee (from Table 9 above). This information in Table 10 can be found in the "4_Helper_data" tab of the workbook (columns J through Z).

Table 10: Lowest 10 median employee thresholds for crude petroleum extraction (NAICS = 211120)

NAICS	Pollutant	No. Facilities[a]	Median employee threshold[b]
211120	Formaldehyde	487	3 (Trigger)
211120	Benzene	544	33
211120	Acetaldehyde	404	61
211120	Acrolein	384	63
211120	Ethylene Dibromide	22	72
211120	1,3-Butadiene	292	73
211120	Naphthalene	292	223
211120	PAH, total	145	230
211120	Cadmium	76	373
211120	2-Methylnaphthalene	114	466
…			

[a] column W of the "4_Helper_data" tab
[b] column T of the "4_Helper_data" tab

Before moving to the next step, EPA additionally reviewed and selectively revised the trigger pollutant based on expert judgement in cases where the pollutant with the lowest trigger value was not representative of the NAICS. This is necessary because EPA is aware that the NEI data are incomplete for HAP, and using an uncommon pollutant to determine expected facilities that would report for an entire NAICS could unreasonably overestimate the number of facilities.

For example, the EPA skipped benzidine for NAICS 221112 for Fossil Fuel Electric Power Generation (see the "Skip Number List" worksheet of the workbook). In this case, benzidine was present for only 17 facilities at the full NAICS level out of more than 2000 facilities in the NEI, but had EPA used this pollutant to determine facility counts, it would have resulted in an employee threshold of 1 employee. This in turn would have resulted in *every* facility for NAICS 221112 being included in the estimated count of facilities that would need to report. The EPA determined that based on the available NEI data, it would be unreasonable to assume that all facilities within NAICS 221112 *could* have benzidine for the purposes of estimating the potential number of facilities affected by the proposal. Instead, EPA selected benzyl chloride, which was reported at 143 facilities in the 2017 NEI for NAICS 221112. The resulting estimated number of facilities that would report assumes that all facilities with this NAICS could emit benzyl chloride, but if all did not emit benzyl chloride, then the actual number that would need to report would be lower.

As a result of this approach, for this example, EPA did not include all facilities within NAICS 221112 in its count of potentially subject facilities. Rather, EPA estimated 1,623 facilities would report, based assumption that all facilities in this NAICS emit benzyl chloride (even though < 10% of facilities in the NEI for NAICS 221112 have this pollutant reported) and that the level of such emissions is directly correlated with the number of employees or receipts (depending on the NAICS code as explained in the examples provided in Table 5 and Table 7 above). This estimation approach is still extremely conservative, but not excessively so, and is necessary because the NEI data are incomplete. The number of facilities that would truly need to report under such a rule as EPA has proposed would depend on whether the facility is major source or not, and if not, the levels of emissions for each pollutant in comparison to the threshold.

The EPA skipped one or more pollutants for 30 NAICS, which EPA has provided in the "Skip number list" worksheet of the workbook. To use this information, for each NAICS/pollutant combination in the list, refer to the "4_Helper_data" worksheet for the same NAICS/pollutant combination. The Skip Number value is the number of rows to count down from the first pollutant in the sorted list of pollutants for each NAICS in the "4_Helper_data" worksheet. For example, for NAICS 221112, the highest-ranking pollutant is benzidine, but the Skip Number of 2 means that the employee number trigger of 24 employees was used from Benzyl Chloride rather than the trigger of 1 employee for Benzidine or 2 employees for N-Nitrodosimethylamine. These pollutants were skipped because the available data had emissions from just 17 and 4 facilities, respectively, for NAICS 221112. Given the low trigger number of employees resulting from this approach, using benzyl chloride with 132 facilities in the 2017 NEI reported with that pollutant was a more reasonable approach because more facilities emitted that pollutant.

5.3.4 Step 4: Classify each NAICS and SBA enterprise size group as (a) small business and (b) above trigger level

For each NAICS and enterprise size threshold, we have described in Section 5.2 how EPA assigned whether the firms in each group would be considered small firms for the purposes of this analysis (based on both the SBA definition and the CAA definition approximation). In Sections 5.3.1 through 5.3.3, we have described how we have determined the number of employees or receipts that would be needed to trigger reporting for each NAICS and pollutant. In the final step provided in this section, EPA describes how it determined the number of firms and establishments (by enterprise size groups) that EPA estimates would need to report emissions data to EPA under the AERR proposal. This step is implemented in the "5 Calc by NAICS per size group" tab of the workbook.

For NAICS that use employees to determine small business status, the EPA compared the maximum range of the employee size threshold to the trigger employees (see column "1/ Million AERR Report?"). If the size category maximum was greater than or equal to the trigger threshold, then EPA assumed that all firms and establishments within that size threshold would need to report to the NEI. A similar approach was taken for NAICS that use receipts to determine small business status, but both the size range and trigger were receipts rather than employees. This approach is conservative because all firms in the enterprise size group are included, even if some of the firms are smaller than the trigger value.

The approach of using the maximum of the enterprise size classifications will tend to overestimate the number of firms included. This is because the enterprise size maximums are greater than (or equal to) the establishment sizes on which the triggers are based. As a result, all firms in the enterprise size group are included, even if some of the firms are smaller than the trigger value. While most NAICS do not have a significant difference between enterprise size ranges and average employees/establishment for the smaller enterprise size ranges, some differences exist for the larger size ranges. This approach is therefore conservative (it likely overestimates) for the purposes of estimating the number of small businesses potentially subject to proposed reporting provisions.

To calculate the number of small businesses for each NAICS, EPA added up all the businesses (using the "Firms" column of the Economic Census data) for each NAICS across the enterprise size classifications that were both (1) labeled as small businesses and (2) identified as potentially needing to report to EPA given the preliminary emissions thresholds and this analysis approach. As a result, using the CAA definition of small business, EPA estimates about 35,000 small businesses (firms) and about 39,000 facilities (establishments) operated by small businesses could be subject to reporting at least one pollutant. Using the SBA definition of small businesses, the results are about 45,000 small businesses and about 57,000 facilities operated by small businesses.

Table 11 illustrates the results of this approach for NAICS 493110 (General Warehousing and Storage) and using firms and the SBA definition of small business, the data for which is available in the "5 Calc by NAICS per size group" tab of the workbook. In this example, the highlighted cells show the counts of firms that are within the enterprise size range that we estimated would be subject to reporting based on the receipts size trigger of $1.304M. Within those estimated for reporting, only those with receipts of

$30M or less would be considered small. The sum of the firms in the shaded area represents the number of small businesses estimated to need to report for NAICS 493110, which is 2,446.

- Small business definition: $41.5 million receipts
- Trigger pollutant: Benzene
- Trigger receipts: $1.305 million

Table 11: Example of classification of firms as reporting and small businesses, NAICS = 493110

NAICS	Enterprise Size (employees or $1000)	Firms	Category Maximum ($M)	Small Business Flag	1/Million AERR Report?	Est. AERR Firms
493110	02: <100	425	$0.1	Small		0
493110	03: 100-499	980	$0.5	Small		0
493110	04: 500-999	574	$1.0	Small		0
493110	05: 1,000-2,499	759	$2.5	Small	Y	759
493110	06: 2,500-4,999	461	$5.0	Small	Y	461
493110	07: 5,000-7,499	268	$7.5	Small	Y	268
493110	08: 7,500-9,999	148	$10.0	Small	Y	148
493110	09: 10,000-14,999	221	$15.0	Small	Y	221
493110	10: 15,000-19,999	152	$20.0	Small	Y	152
493110	11: 20,000-24,999	103	$25.0	Small	Y	103
493110	12: 25,000-29,999	86	$30.0	Small	Y	86
493110	13: 30,000-34,999	86	$35.0	Small	Y	86
493110	14: 35,000-39,999	82	$40.0	Small	Y	82
493110	15: 40,000-49,999	80	$50.0	Small	Y	80
493110	16: 50,000-74,999	149	$75.0	Not small	Y	149
493110	17: 75,000-99,999	120	$100.0	Not small	Y	120
493110	18: 100,000+	1133	$100.0	Not small	Y	1,133

5.4 Estimating the number of facilities and micro-businesses expected to report

In addition to allowing for estimation of small firms that may need to report under the proposed AERR, the approaches described above also allow for estimating the total number of facilities that would be reported. The draft ICR includes estimated hours for reporting on a per-facility basis, and thus to use the information, an estimated number of facilities is needed.

To estimate the number of facilities irrespective of whether they are small business or not, EPA used the "6 Results Pivot" worksheet of the workbook to summarize the number establishments from the Economic Census data. Using this approach, EPA created the resulting count of establishments and stored these in the "Final results for ICR& preamble" worksheet available in the workbook. The number of establishments identified with this approach is 125,368 with 39,203 small and 86,165 not small, based on the CAA small definition approach described above. In creating these totals, EPA excluded several NAICS from the counts. First, EPA excluded hydroelectric, solar, and wind energy NAICS from the counts

because the approach for the NAICS4 group these sectors appear in is dominated by emissions from fossil-based electric generation, and there is no evidence that the same trigger pollutants or levels should apply for these NAICS. The EPA also excluded small business from the Automotive Body, Paint, and Interior Repair and Maintenance industry (NAICS 811121) as per the results of the SBAR Panel report.[14]

Because the RIA and draft ICR consider additional costs for micro facilities to contract out emissions reporting responsibilities, EPA added a flag to the "5 Calc by NAICS per size group" worksheet called "Micro? Def<20 employees or <$3M receipts)." This field includes a "Y" for any Enterprise Sizes from the Economic Census data that have fewer the 20 employees or < $3M in receipts. Since there is no standard definition of micro facilities, EPA used this definition as being somewhat conservative as compared to a definition used by the SBA of less than 5 employees (15 U.S.C. § 6901).[15] Adding this field as a column via the "6 Results Pivot" tab provides a summary of the micro facilities that EPA estimates would need to report based on the proposed thresholds. The results of this summary are also included in the "Final results for ICR& Preamble" tab and show 19,024 facilities defined as micro facilities using this definition.

The analysis above cannot easily account for major versus non-major sources since it includes all facilities for the NAICS considered. The Economic Census does not provide individual facilities with their major/non-major status, thus separating out major sources is not possible. In addition, the number of facilities identified in the above analysis does not include all NAICS and certain major sources are known to have NAICS that are not included in the proposed non-major NAICS list. Since the proposed AERR would require all major sources to report, it is important to include all major sources in the ICR facility counts while avoiding any double counting of those major sources included in the analysis of these NAICS. To do this, EPA analyzed the 2017 NEI augmented with the major source assignments as described above in Section 2.1. These calculations are available in the "Facility counts worksheet" of the workbook "AERR Revision Burden Estimates v9 For Docket.xlsx" included as an attachment to the ICR Supporting Statement in the docket for the AERR proposal. These calculations show that of the 125,358 facilities identified by the analysis above, 9,533 of these are also major sources. So, for the final ICR count of facilities, EPA used 115,835 non-major sources along with the 13,420 major sources (as described in the ICR Supporting Statement available in the docket for this rule).

---

[14] U.S. EPA, "Panel Report of the Small Business Advocacy Review Panel on EPA's Planned Proposed Rule Revisions to the Air Emissions Reporting Requirements," January 3, 2023, also available in the docket for the proposal EPA-HQ-OAR-2004-0489.

[15] Title 15 of the U.S. Code, Chapter 95, §6901(10) defines "microenterprise" as "a sole proprietorship, partnership, or corporation that—(A) has fewer than 5 employees; and (B) generally lacks access to conventional loans, equity, or other banking services.

5.5   Uncertainties

There are numerous assumptions in this approach that are necessary given various limitations with the available data. These lead to uncertainties, but those cannot be measured or readily quantified. Thus, to try to make these estimates conservative, EPA has chosen assumptions that would tend to overestimate the resulting number of small businesses.

The key uncertainties based on this approach are:

- The use of 2017 Economic Census data, which is the most current available, but nevertheless is six years old. Final 2022 Economic Census data of use for this analysis will not be available until at least March 2024, according to Economic Census announcements (https://www.census.gov/topics/business-economy/library/flyers/economic-census-2022.html, accessed on August 25, 2022).
- Unknown actual number of employees or receipts per NEI facility – rather number of employees and receipts are assigned to NEI facilities by assuming emissions are correlated with the number of employees and receipts.
- The use of 2017 emissions data (to be consistent with the use of the 2017 Economic Census data) may include more or fewer facilities for each NAICS or higher or lower emissions than if we used more current data. The 2020 NEI emissions data were not completed when this analysis was performed.
- Missing pollutants from the NEI at existing facilities and missing facilities because HAP are collected voluntarily and because of limitations in available source test and emission factor data used to estimate emissions.
- Facilities in the NEI that use 4-digit and 5-digit NAICS, when a full 6-digit NAICS would be more precise definition of the industry. The lack of specific NAICS leads to use of NAICS4 for mapping facilities with employee counts and receipts.
- Uncertain emissions levels in the NEI because of current voluntary reporting requirements, limitations in the available source test data and use of low-quality emission factor data.

The key assumptions that EPA has made, and their associated directional effect on the analysis results are summarized in the list below:

Assumption	Expected Impact
• Use 2017 NEI data (rather than more current data since it was not available when this work was performed).	• Could be fewer or more businesses (and small businesses) now, depending on NAICS, given changes to the economy since 2017. • Would underestimate facilities that would need to report dioxins/furans, if those pollutants would have been a trigger pollutant (the 2017 NEI does not include dioxins/furans).

Assumption	Expected Impact
• NEI facility emissions are correlated to number of employees or receipts. • Use of average number of employees per establishment from the 2017 Economic Census data is sufficient to assign employees to NEI facilities.	• Introduces random error into emissions/activity rates. This concern is somewhat offset for NAICS with large sample sizes (i.e., large number of facilities on which trigger thresholds are based) and use of median rate since that is the most common rate over all facilities.
• NEI emissions are sufficiently complete for facility ranking.	• If missing CO, VOC, and/or PM2.5 emissions and HAP are incomplete or underestimated, then a larger facility may be ranked lower and therefore be assigned a lower number of employees or receipts amount. This could cause a larger amount of a trigger HAP to be associated with a smaller number of employees, which would tend to overestimate the number of facilities included in the estimated facility counts. While overestimated emissions are possible, it is EPA's experience that overestimates are more readily identified and corrected because higher emissions result in higher state fees for owners/operators.
• Assumed 85% control on NEI point source solvent VOC and VOC HAP.	• This assumption is dependent on the actual VOC or HAP controls (if any) at each facility and could cause emissions rates to be too high or too low. It is more likely that solvent VOC emissions are uncontrolled in the NEI, because the solvent VOC emissions are consistent with solvent reformulations to meet current EPA Clean Air Act standards rather than reliance on "end-of-pipe" emission controls. Thus, this assumption would tend to overestimate emission rates and reduce the employee thresholds and lead to an overestimate of small businesses expected to be subject to reporting.
• Pollutant selected is sufficiently representative and that all facilities would emit this pollutant.	• Tends to overestimate number of small businesses. For purposes of determining whether a facility would report to EPA under a final rule (not an estimated count), the facility emissions would need to be compared to any finalized AERR emissions reporting thresholds.
• Use of median emissions/employee and per receipts based on NEI.	• Median may not be representative depending on the NAICS, which could cause triggers to vary up or down depending on the circumstances. However, median is the best choice available because it represents the most frequently occuring value across the distribution.

Assumption	Expected Impact
• Assume all facilities in SBA size group have the minimum employees or minimum receipts in the range when calculating emissions/employee and emissions/receipts.	• Would lead to lower trigger thresholds for employees and receipts, thus overestimating the number of small businesses potentially subject to reporting.
• Use the enterprise size range maximum to compare to the trigger thresholds.	• Would overestimate the number of businesses because the enterprise size is being used instead of the establishment size. By definition, enterprise size is greater than or equal to the establishment size.
• Do not include 9xxxx series NAICS because government activities are not in Census data.	• Does not estimate small (governmental) entities that run municipal facilities that could be affected.

6 Environmental justice analysis

This section provides summary results and describes the approach used to evaluate the different socio-economic demographic groups within the population living near the known facilities (from the 2017 NEI) that would be subject to proposed revisions to the AERR. Because EPA does not have data on many of the facilities that EPA expects would need to report under this proposal, those facilities are not included in this analysis.

This work, which is provided for informational purposes, is referenced by Section IV.A.10 of the preamble to the proposed rule. The demographic analysis is for those known facilities identified by EPA as potentially being subject to the proposed revisions to the existing AERR rule under 40 CFR part 51, subpart A, analyzed by major and non-major sources of CAP and HAP. Because the EJ analysis was done before all thresholds were finalized for the proposal, the list of facilities used for this analysis is not the final list of known facilities potentially subject. Rather, the analysis is representative of the final list of facilities to provide the EJ analysis. The current analyses evaluate census blocks surrounding these facilities with census-based demographic data and present the demographic composition of the populations located within 5-km proximity to these facilities. The following demographic groups were included in these proximity analyses:

- Total population;
- White;
- Minority;
- African American (or Black);
- Native Americans;
- Other races and multiracial;
- Hispanic or Latino;
- Children 17 years of age and under;
- Adults 18 to 64 years of age;
- Adults 65 years of age and over;

- Adults 25 years of age and older without a high school diploma;
- People living below the poverty level; and
- Linguistically isolated people.

The current analysis identified all census blocks with centroids[16] located within specified radii of the latitude/longitude location of each facility, and then linked each block with census-based demographic data. In this analysis, if the centroid of a census block is located within the specified radius, the entire population of that census block is counted as within the radius. In addition to facility-specific demographics, EPA also computed the demographic composition of the populations within the specified radius collectively for all facilities potentially subject to the proposed AERR (*e.g.*, inventory -wide). The inventory-wide computation considers neighboring facilities with overlapping study areas and ensures populations in common are counted only once in this demographic analysis. Finally, this analysis compares these inventory-wide demographics at the specified radius (*i.e.*, 5 km) to the demographic composition of the nationwide population.

The 5-km distance was established based on an analysis of distance to populations most affected by facility emissions as described below.

The distance from the facility centroid to the modeled cancer maximum individual risk (MIR) location was used as a measure of the minimum distance to the most affected populations. The MIR location being the location of the census block centroid or user receptor that was modeled in HEM as having the highest cancer risk for that facility. It should be noted that highly affected populations would be located in areas around the MIR, not just at the MIR. So, including more area beyond the MIR would ensure that the most affected populations are captured.

The distance from the facility centroid to the facility MIR location was investigated for 22 source categories that were modelled in HEM during RTR development. These 22 source categories include a total of 1,612 facilities. Table 12 shows the number of facilities with the average, median, minimum, and maximum distances from the facility centroid to the MIR in kilometers.[17] The average distance for all categories investigated is about 2 km and the median is about 1 km. Minimum distances range from 0.01 km for the Miscellaneous Metal Parts Coating Operations source category to about 1 km for the

---

[16] A census block centroid is considered a central location of the block polygon it represents and contains the same census-based information as the block polygon (e.g., the same population). See https://www2.census.gov/geo/pdfs/reference/GARM/glosGARM.pdf.

[17] The distance in kilometers was calculated from the latitude and longitude coordinates of the facility centroid to the coordinates for the MIR using the great circle formula as follows:

=ACOS(COS(RADIANS(90-Lat1)) * COS(RADIANS(90-Lat2)) + SIN(RADIANS(90-Lat1)) * SIN(RADIANS(90-Lat2)) * COS(RADIANS(Long1-Long2))) * 6371

Cyanide Chemicals Production source category. Maximum distances range from about 1 km for Metal Furniture Coatings to over 40 km for the Lime Manufacturing source category.

Table 12: Summary of Distances from Facility Centroid to Facility MIR for 22 Source Categories

Source Category	Facilities with MIR	Distance from Facility Centroid to Facility MIR (Census Block Centroid or User Receptor) in km			
		Ave	Median	Min	Max
Taconite	8	3.16	2.07	0.73	8.99
Miscellaneous Metal Parts and Products Surface Coating (MMP)	305	0.47	0.36	0.01	3.03
Lime	34	3.90	1.14	0.09	40.03
Wet-Formed Fiberglass	7	0.69	0.71	0.40	1.01
Automobile and Light Duty Truck Surface Coatings (ALDT)	42	1.08	0.93	0.31	3.28
Coke Ovens	14	1.29	0.81	0.39	4.04
Metal Furniture Coatings	14	0.35	0.23	0.07	0.84
Boat Manufacturing	22	0.62	0.31	0.07	2.73
Engine Test Cells	57	1.23	0.84	0.08	5.46
Fabric Coating	28	0.36	0.27	0.09	1.50
Carbon Black	15	1.96	1.57	0.48	4.45
Ethylene Production	31	1.73	1.53	0.27	4.35
Flexible Polyurethane Foam Fabrication Operation[a]	0	NA	NA	NA	NA
Gas Turbines	253	2.04	1.05	0.12	28.64
Hydrochloric Acid (HCL) Production[a]	0	NA	NA	NA	NA
Cyanide Chemicals Production	11	2.98	2.50	0.88	6.86
Integrated Iron and Steel (IIS)	11	1.56	1.32	0.62	3.42
Iron and Steel Foundries	40	0.48	0.39	0.06	2.27
MATS	322	3.29	2.46	0.47	36.66
Metal Coil Coatings	40	0.70	0.57	0.07	2.04
Hazardous Organic National Emission Standards (HON)	216	1.60	1.25	0.14	8.11
Petroleum Refineries	142	1.65	1.07	0.20	16.48
TOTAL	**1613**	**1.74**	**1.01**	**0.01**	**40.03**

[a] Two source categories did not emit carcinogens, and therefore, had no cancer MIR data.

Table 13 shows the number and percent of facilities with their MIR located within a certain radius. Overall, at 2 km, 74% of the facility MIRs would be captured and at 3 km 85% would be captured. At 5 km, about 94% of the MIRs are captured. The 5-km radius captures a large portion of the MIR (94%) without being too large. In addition, the 5-km distance captures over 90% of the MIR locations for 14 of

the 18 source categories investigated with MIR data. The radius would have to be expanded another two-fold to 10 km to capture 98% of the MIRs.

Table 13: Number and Percent of Facilities with MIR within a Specified Radius

Source Category	Facilities with MIR	Number (Percent) of Facilities within Radius				
		2 km	3 km	4 km	5 km	10 km
Taconite	8	4 (50%)	5 (63%)	6 (75%)	6 (75%)	8 (100%)
MMP	305	302 (99%)	304 (99.7%)	305 (100%)	305 (100%)	305 (100%)
Lime	34	24 (71%)	29 (85%)	29 (85%)	29 (85%)	30 (88%)
Wet-Formed Fiberglass	7	7 (100%)	7 (100%)	7 (100%)	7 (100%)	7 (100%)
ALDT	42	40 (95%)	41 (98%)	42 (100%)	42 (100%)	42 (100%)
Coke Ovens	14	12 (86%)	13 (93%)	13 (93%)	14 (100%)	14 (100%)
Metal Furniture Coatings	14	14 (100%)	14 (100%)	14 (100%)	14 (100%)	14 (100%)
Boat Manufacturing	22	20 (91%)	22 (100%)	22 (100%)	22 (100%)	22 (100%)
Engine Test Cells	57	46 (81%)	53 (93%)	56 (98%)	56 (98%)	57 (100%)
Fabric Coating	28	28 (100%)	28 (100%)	28 (100%)	28 (100%)	28 (100%)
Carbon Black	15	9 (60%)	12 (80%)	13 (87%)	15 (100%)	15 (100%)
Ethylene Production	31	23 (74%)	27 (87%)	30 (97%)	31 (100%)	31 (100%)
Flexible Polyurethane Foam Fabrication Operation	0	-	-	-	-	-
Gas Turbines	253	184 (73%)	207 (82%)	226 (89%)	233 (92%)	247 (98%)
HCL Production	0	-	-	-	-	-
Cyanide Chemicals Produdction	11	4 (36%)	7 (64%)	8 (73%)	9 (82%)	11 (100%)
IIS	11	8 (73%)	10 (91%)	11 (100%)	11 (100%)	11 (100%)
Iron and Steel Foundries	40	39 (98%)	40 (100%)	40 (100%)	40 (100%)	40 (100%)
Mercury and Air Toxics Standards	322	124 (39%)	211 (66%)	253 (79%)	273 (85%)	311 (97%)
Metal Coil Coatings	40	38 (95%)	40 (100%)	40 (100%)	40 (100%)	40 (100%)
HON	216	157 (73%)	188 (87%)	205 (95%)	213 (99%)	216 (100%)
Petroleum Refineries	142	107 (75%)	123 (87%)	135 (95%)	138 (97%)	140 (99%)
TOTAL	**1,612**	**1190 (74%)**	**1,381 (86%)**	**1,483 (92%)**	**1,526 (95%)**	**1,589 (99%)**

The census data used in this analysis is described in Section 6.1. The algorithms used to compute the population of each demographic category surrounding the facility are presented in Section 6.2. The summary results of these analyses are presented in Section 6.3, and Section 6.4 provides a characterization of the uncertainties. Finally, Section 6.5 points to supplemental workbooks containing the detailed facility-specific results from the analyses described in this section.

6.1 Census data

The total population within a specified radius around each facility is the sum of the population for every census block within that radius, based on each block's population provided by the 2010 Decennial Census.[18] For the demographic analysis, statistics on total population, race, ethnicity, age, education level, poverty status and linguistic isolation are obtained from the Census' American Community Survey (ACS) 5-year averages for 2015-2019.[19] These data are provided at the block group level. A census block group contains about 28 blocks on average, or about 1,400 people.

Table 14 summarizes the census data used in the analysis, showing the source of each dataset and the level of geographic resolution.

Table 14: Summary of Census Data used for Different Demographic Groups

Type of population category	Source of data	Geographic resolution
Total population (sum of block counts within radius)	2010 Census SF1	Census block
Total population (sum of block group counts, used for demographic percentages)	ACS Table B03002 (e1)	Census block group
Race/ethnicity categories (percentages): White (non-hispanic): $e3/e1$ Minority (non-white + hispanic): $(e1-e3)/e1$ African American (non-hispanic): $e4/e1$ Native American (non-hispanic): $e5/e1$ Other & Mixed race (non-hispanic): $(e6+e7+e8+e9)/e1$ Hispanic (all races): $e12/e1$	ACS Table B03002, Hispanic or Latino Origin by Race (Tiger table X03):	Census block group
Age groups	ACS Table B01001, Sex by Age (Tiger table X01)	Census block group

---

[18] Block level population used by EPA will be updated based on the 2020 Decennial Census, once processed and quality-assured for these analyses.

[19] U.S. Census Bureau, 2020. Five-year American Community Survey – 2015-2019, United States: https://www2.census.gov/programs-surveys/acs/summary_file/2019/data/5_year_entire_sf/.

Type of population category	Source of data	Geographic resolution
Level of education – percentage of adults 25 years and older without a high school diploma	ACS Table B15002, Sex by Educational Attainment (Tiger table X15)	Census block group
Individuals living in households earning below the poverty level (percentage of individuals)	ACS Table C17002, Ratio of Income to Poverty Level (Tiger table X17): (e2+e3)/e1	Census block group
Individuals living in linguistically isolated households (percentage of households)	ACS Table C16002, Household Language by Household (Tiger table X16): (e4+e7+e10+e13)/e1	Census block group

The statistics for total minorities, age groups, educational attainment, poverty, and linguistic isolation are consistent with the demographic statistics used in EPA's EJSCREEN tool for Environmental Justice analysis.[20] We derive our demographic statistics from the ACS, which is the source of data for EJSCREEN's statistics. For the current analysis, however, we provide the impact on different racial and ethnic groups in more detail later in this document (see Table 15).

6.2   Calculation methods

EPA used the census block and census block group identification codes to link each block to the appropriate ACS block group demographic statistics. This allowed us to estimate the number of people in different demographic categories for each census block in a specified radius around each facility. As noted in Section 6.1, demographic data are available at the census block group level. To estimate more detailed block-level demographic percentages for the purposes of this analysis, the demographic characteristics of a given block group – that is, the percentage of people in different races/ethnicities, the percentage in different age groups, the percentage without a high school diploma, the percentage that are below the poverty level, and the percentage that are linguistically isolated – are presumed to also describe each census block located within that block group.

For comparison, the nationwide demographic percentages are computed from the Census' ACS 5-year averages for 2015-2019 ("2019 ACS"). The denominator for these percentages is the total nationwide population, which is likewise computed from the 2019 ACS and determined by summing the total population of all census block groups. We also provide the total population based on the 2010 Decennial Census for comparison, because the census block populations are based on 2010 Decennial Census data, as noted in Section 6.1.

Sections 6.2.1 through 6.2.4 describe calculation methods for racial, ethnic, age, education status, poverty status, and linguistic isolation demographic categories. Section 6.2.5 describes the gap-filling

---

[20] U.S. Environmental Protection Agency. 2020. EJSCREEN: EPA's Environmental Justice Screening and Mapping Tool (Version 2020), https://www.epa.gov/ejscreen.

approach used when block group statistics are not available for a given block, based on computing default averages for the missing demographic(s) at the tract or county level.

6.2.1 Race, ethnicity and age categories

Table B03002 (Hispanic or Latino origin by race) of the ACS data provides race/ethnicity statistics for each census block group nationwide. Table B01001 of the ASC data provides age statistics for the population by ranges (in years) for each census block group nationwide. For each census block in this analysis, the race/ethnicity (White, African American, Native American, Multiracial/Other, and Hispanic or Latino) and age range (0-17, 18-64 and ≥65 years) for that block is estimated based on the demographic information provided at the block group level, as follows:

$$N(s,b/bg) = N(t,b/bg) \times P(s,bg) / 100$$

where:

$N(s,b/bg)$ = number of people in racial/ethnic or age subgroup "s", in block "b" of block group "bg";

$N(t,b/bg)$ = total number of people in block "b" of block group "bg"; and

$P(s,bg)$ = percentage of people in racial/ethnic or age subgroup "s", in a block group "bg".

The number of people in each racial/ethnic and age category is calculated using the above equation, summed over all blocks that fall within the specified radius of each facility.

6.2.2 Level of education

Table B15002 (educational attainment) of the ACS data provides education attainment statistics for each census block group nationwide. For each census block in this analysis, the number of people 25-years and older without a high school diploma is estimated based on the demographic information provided at the block group level, as follows:

$$N(nhs,b/bg) = N(t,b/bg) \times P(nhs,bg) / 100$$

where:

$N(nhs,b/bg)$ = number of people 25-years and older without a high school diploma "nhs", in block "b" of block group "bg";

$N(t,b/bg)$ = number of people 25-years and older in block "b" of block group "bg"; and

$P(nhs,bg)$ = percentage of people 25-years and older without a high school diploma "nhs", in a block group "bg".

The number of people 25-years and older without a high school diploma is calculated using the above equation, summed over all blocks that fall within the specified radius of each facility.

6.2.3   Poverty level

Table C17002 (poverty) of the ACS data estimates the numbers of individuals within a census block group who live in households where the household income is below the poverty line, and below various multiples of the poverty line. For this analysis, we calculate the fraction of individuals living in households earning incomes below the poverty level. For each census block in this analysis, the block's household income level is estimated based on the demographic information provided at the block group level, as follows:

$$N(hi,b/bg) = N(t,b/bg) \times P(hi,bg) \, / \, 100$$

where "hi" indicates household income below the poverty level, and:

$N(hi,b/bg)$ =   number of people living in households "hi" below the poverty level, in block "b" of block group "bg"

$N(t,b/bg)$ =   total number of people in block "b" of block group "bg"

$P(hi,bg)$ =   percentage of people living in households "hi" below the poverty level, among the population for which poverty status is known, in block group "bg"

The number of people living in households earning below the poverty level is calculated using the above equation, summed over all blocks that fall within the specified radius of each facility.

6.2.4   Linguistic isolation

Linguistic Isolation is defined by in the ACS as "a household in which all members age 14 years and over speak a non-English language and also speak English less than "very well" (have difficulty with English)."[21] Table C16002 (Tiger table X16_language_spoken_at_home) of the ACS data provides the number of households in linguistic isolation in each block group. For each census block in this analysis, the number of people living in linguistic isolation is estimated based on the demographic information provided at the block group level, as follows:

$$N(li,b/bg) = N(t,b/bg) \times P(li,bg) \, / \, 100$$

where:

$N(li,b/bg)$ = number of people living in linguistic isolation "li", in block "b" of block group "bg";

---

[21] U.S. Census Bureau, 2020. American Community Survey and Puerto Rican Community Survey 2019 Subject Definitions. P. 49. https://www2.census.gov/programs-surveys/acs/tech_docs/subject_definitions/ 2019_ACSSubjectDefinitions.pdf.

N(t,b/bg) = total number of people in block "b" of block group "bg"; and

P(li,bg) =   percentage of linguistically isolated households "li", in block group "bg".

The number of people living in linguistic isolation is calculated using the above equation, summed over all blocks that fall within the specified radius of each facility.

6.2.5   Defaults

Block and block group designations used in the Census may be modified to accommodate population growth in some regions. As a result, certain census blocks which are based on the last Decennial Census, may not map to the block group designations used in the latest 5-year ACS. In addition, some statistics may not be reported in the ACS results for every block group. Race, ethnicity, and age statistics are generally reported for all block groups. However, poverty, linguistic isolation, and educational attainment statistics are not available for some block groups.

In these cases, we compute default estimates for the missing demographic statistics based on the average statistics for the tract in which the block is located. If no census tract-level data are available, EPA estimated demographic statistics based on the overall demography of the county in which the unmatched block is located. We performed this gap-filling exercise separately for each type of demographic data. That is, in the case where some categories of data are available (for instance, race, age and ethnicity) and others are not available (educational attainment, poverty, or linguistic isolation) we only computed defaults for the categories of data that are missing.

The EPA computed tract level defaults using weighted averages based on all of the other block groups in the tract for which data are available. The defaults are calculated as follows for race, ethnicity, and age subgroups:

$$P(s,T) = \{ \sum P(s,bg/T) \times N(t,bg) \} / \{ \sum N(t,bg) \}$$

where:

P(s,T) =   percentage of people in race, ethnicity, or age subgroup "s", in tract "T";

$\sum$ refers to the summation over all block groups in tract "T" for which data are available;

P(s,bg/T) = percentage of people in race, ethnicity, or age subgroup "s", in a block group "bg" of tract "T"; and

N(t,bg) =   total number of people in block group "bg".

The EPA calculated defaults for educational attainment, poverty, and linguistic isolation in a similar way, except that the population weighting term N was replaced by the population over age 25, the population for which poverty status is known, and the number of households, respectively. County level defaults were also calculated in a similar way, except that data were summed over the county instead of the census tract.

6.3 Results

The proximity results describe the demographics of the population surrounding these facilities subject to proposed revisions to the AERR rule. Table 15 presents the demographic composition of the population located within a proximity of 5 km to the 17,715 facilities as a whole, as well as within 5 km of subsets of these facilities based on emissions. For the purposes of this analysis, the demographics surrounding these 17,715 facilities are also examined according to their emissions as follows: non-major sources (6,096 facilities); criteria air pollutant (CAP) major sources (4,067 facilities); and hazardous air pollutant (HAP) or HAP-CAP major sources (7,552 facilities). For context, Table 15 also provides the nationwide percentages of these various demographic groups. The detailed facility-specific results underpinning these results are noted in the demographics report titled "Analysis of Demographic Factors for Populations Living Near Facilities Subject to the Proposed Revisions to the Air Emissions Reporting Requirements".

6.3.1   Non-major source facilities

For the 6,096 facilities subject to the proposed revisions to the AERR rule that are non-major sources that were identified for this analysis, the proximity results presented in Table 15 indicate that the population percentages for certain demographic groups within 5 km of these facility operations are greater than the corresponding nationwide percentages for those same demographics. The demographic percentage for populations residing within 5 km of facility operations is 8 percentage points greater than its corresponding nationwide percentage for the minority population (48% within 5 km of the facilities compared to 40% nationwide), 4 percentage points greater than its corresponding nationwide percentage for the Hispanic and Latino population (23% within 5 km of the facilities compared to 19% nationwide), 3 percentage points greater for the African American population (15% within 5 km of the facilities compared to 12% nationwide), 2 percentage points greater than its corresponding nationwide percentage for the other and multiracial population (10% within 5 km of the facilities compared to 8% nationwide), 2 percentage points greater than its corresponding nationwide percentage for the population living in linguistic isolation (7% within 5 km of the facilities compared to 5% nationwide), 2 percentage points greater than its corresponding nationwide percentage for the population over 25 years old without a high school diploma (14% within 5 km of the facilities compared to 12% nationwide), 2 percentage points greater than its corresponding nationwide percentage for people living below the poverty level (15% within 5 km of the facilities compared to 13% nationwide), and 2 percentage points greater than its corresponding nationwide percentage for the population aged 18 to 64 years old (64% within 5 km of the facilities compared to 62% nationwide). The remaining demographic groups (*i.e.*, Native Americans) within 5 km of non-major source facility operations are less than the corresponding nationwide percentages.

6.3.2   CAP major source facilities

For the 4,067 facilities identified as subject to the proposed revisions to the AERR rule that are also major sources of CAP, the proximity results presented in Table 15 indicate that the population percentages for certain demographic groups within 5 km of these facility operations are greater than the

corresponding nationwide percentages for those same demographics. The demographic percentage for populations residing within 5 km of facility operations is 10 percentage points greater than its corresponding nationwide percentage for the minority population (50% within 5 km of the facilities compared to 40% nationwide), 5 percentage points greater for the African American population (17% within 5 km of the facilities compared to 12% nationwide), 4 percentage points greater than its corresponding nationwide percentage for the Hispanic and Latino population (23% within 5 km of the facilities compared to 19% nationwide), 3 percentage points greater than its corresponding nationwide percentage for the population living in linguistic isolation (8% within 5 km of the facilities compared to 5% nationwide), 3 percentage points greater than its corresponding nationwide percentage for people living below the poverty level (16% within 5 km of the facilities compared to 13% nationwide), and 2 percentage points greater than its corresponding nationwide percentage for the population over 25 years old without a high school diploma (14% within 5 km of the facilities compared to 12% nationwide). The remaining demographic groups (*i.e.*, Native American and Other/Multiracial) within 5 km of CAP major source facility operations are less than, or within 1 percentage point greater than, the corresponding nationwide percentages.

6.3.3   HAP and HAP-CAP major source facilities

For the 7,552 facilities identified as subject to the proposed revisions to the AERR rule that are major sources of HAP and HAP-CAP, the proximity results presented in Table 15 indicate that the population percentages for certain demographic groups within 5 km of these facility operations are greater than the corresponding nationwide percentages for those same demographics. The demographic percentage for populations residing within 5 km of facility operations is 8 percentage points greater than its corresponding nationwide percentage for the minority population (48% within 5 km of the facilities compared to 40% nationwide), 4 percentage points greater for the African American population (16% within 5 km of the facilities compared to 12% nationwide), 3 percentage points greater than its corresponding nationwide percentage for the Hispanic and Latino population (22% within 5 km of the facilities compared to 19% nationwide), 3 percentage points greater than its corresponding nationwide percentage for people living below the poverty level (16% within 5 km of the facilities compared to 13% nationwide), 2 percentage points greater than its corresponding nationwide percentage for the population living in linguistic isolation (7% within 5 km of the facilities compared to 5% nationwide), and 2 percentage points greater than its corresponding nationwide percentage for the population over 25 years old without a high school diploma (14% within 5 km of the facilities compared to 12% nationwide). The remaining demographic groups (*i.e.*, Native American and Other/Multiracial) within 5 km of HAP and HAP-CAP major source facility operations are less than, or within 1 percentage point greater than, the corresponding nationwide percentages.

6.3.4   All facilities

Finally, for all facilities identified as subject to the proposed revisions to the AERR rule, the proximity results presented in Table 15 indicate that the population percentages for certain demographic groups within 5 km of the facility operations are greater than the corresponding nationwide percentages for

those same demographics. The demographic percentage for populations residing within 5 km of facility operations is 6 percentage points greater than its corresponding nationwide percentage for the minority population (46% within 5 km of the facilities compared to 40% nationwide), 3 percentage points greater for the African American population (15% within 5 km of the facilities compared to 12% nationwide), 2 percentage points greater than its corresponding nationwide percentage for the Hispanic and Latino population (21% within 5 km of the facilities compared to 19% nationwide), 2 percentage points greater than its corresponding nationwide percentage for the population living in linguistic isolation (7% within 5 km of the facilities compared to 5% nationwide), 2 percentage points greater than its corresponding nationwide percentage for the population over 25 years old without a high school diploma (14% within 5 km of the facilities compared to 12% nationwide), and 2 percentage points greater than its corresponding nationwide percentage for people living below the poverty level (15% within 5 km of the facilities compared to 13% nationwide). The remaining demographic groups (*i.e.*, Native American and Other/Multiracial) within 5 km of all facility operations are less than, or within 1 percentage point greater than, the corresponding nationwide percentages.

Table 15: Summary of Demographic Assessment for Facilities Subject to Proposed Revisions to the Air Emissions Reporting Rule: Proximity Statistics

Population Basis		Demographic Group[a]											
		Total	Minority[b]	African American	Native American	Other and Multiracial	Hispanic or Latino	Ages 0 to 17	Ages 18 to 64	Ages 65 and up	Over 25 Without a HS Diploma	Below Poverty Level	Linguistic Isolation[d]
Nationwide Demographics (2015-2019 ACS)		328,016,242	40%	12%	0.7%	8%	19%	22%	62%	16%	12%	13%	5%
Nationwide Block Counts (2010 Decennial Census)[e]		312,459,649											
	# Facilities	Population Surrounding the Facilities within 5 km[f]											
Non-Major Source Facilities	6,096	93,000,649	48%	15%	0.4%	10%	23%	22%	64%	14%	14%	15%	7%
CAP Major Source Facilities	4,067	69,683,592	50%	17%	0.4%	9%	23%	22%	63%	15%	14%	16%	8%
HAP/HAP-CAP Major Source Facilities	7,552	117,946,858	48%	16%	0.4%	9%	22%	22%	63%	15%	14%	16%	7%
All Facilities	17,715	171,011,126	46%	15%	0.4%	9%	21%	22%	63%	15%	14%	15%	7%

[a] The demographic percentages are based on the Census' 2015-2019 American Community Survey five-year averages, at the block group level, and include the 50 states, the District of Columbia, and Puerto Rico. Demographic percentages based on different averages may differ. The total population of each facility and of the entire run group are based on block level data from the 2010 Decennial Census. Populations by demographic group for each facility and for the run group are determined by multiplying each 2010 Decennial block population within the indicated radius by the ACS demographic percentages describing the block group containing each block, and then summing over the appropriate area (facility-specific or run group-wide).

[b] Minority population is the total population minus the white population.

[c] To avoid double counting, the "Hispanic or Latino" category is treated as a distinct demographic category for these analyses. A person is identified as one of five racial/ethnic categories above: White, African American, Native American, Other and Multiracial, or Hispanic/Latino. A person who identifies as Hispanic or Latino is counted as Hispanic/Latino for this analysis, regardless of what race this person may have also identified as in the Census.

[d] The linguistically isolated population is estimated at the block group level by taking the product of the block group population and the fraction of linguistically isolated households in the block group, assuming that the number of individuals per household is the same for linguistically isolated households as for the general population, and summed over all block groups.

[e] The nationwide 2010 Decennial Census population of 312,459,649 is the summation of all Census block populations within the 50 states, the District of Columbia, and Puerto Rico. Block level population used by EPA will be updated based on the 2020 Decennial Census, once processed and quality-assured for these analyses.

[f] The population tally and demographic analysis of the total population surrounding each group of facilities as a whole takes into account neighboring facilities with overlapping study areas and ensures populations in common are counted only once.

6.4   Uncertainty of results

Our analysis of the distribution of population across various demographic groups is subject to the typical uncertainties associated with census data (e.g., errors in filling out and transcribing census forms), which are generally thought to be small, as well as the additional uncertainties associated with the extrapolation of census block group data down to the census block level.

6.5   Additional supporting information for EJ analysis

Four workbooks contain the detailed facility-specific results underpinning the results presented in this section of the TSD (i.e., section 6). They are provided as attachments to this TSD in the docket with the names:

- AERR Non-Major Facility Demographic Results 2022-10-10.xlsx;
- AERR CAP Major Facility Demographic Results 2022-10-10.xlsx;
- AERR HAP and HAP-CAP Major Facility Demographic Results 2022-10-10.xlsx; and
- AERR All Facility Demographic Results 2022-10-10.xlsx.

The large datasets covering these groupings of the facilities make the facility-specific results more amenable to Excel workbooks, than a Microsoft Word® document. These separate workbooks also contain the summary results discussed in this section for each subset of facilities.