

Methods

Indicator

E1. Percentage of children ages 0 to 17 years living in counties with pollutant concentrations above the levels of the current air quality standards, 1999-2021.

E2. Percentage of children ages 0 to 17 years living in counties with 8-hour ozone and 24-hour PM_{2.5} concentrations above the levels of air quality standards, by frequency of occurrence, 1999-2021.

Summary

EPA's Office of Air Quality Planning and Standards (OAQPS) has set primary (health-based) National Ambient Air Quality Standards (NAAQS) for six common air pollutants, often referred to as criteria air pollutants (or simply criteria pollutants). Each NAAQS has an averaging period (e.g., one-hour, eight-hour, 24-hour, three-month, annual) and a level.ⁱ For Indicator E1 we analyzed the following NAAQS: carbon monoxide, eight-hour; ozone, eight-hour; PM₁₀, 24-hour; PM_{2.5}, annual; PM_{2.5}, 24-hour; sulfur dioxide, one-hour; nitrogen dioxide, one-hour; lead, three-month. For indicator E2 we analyzed the annual frequency with which pollutant concentrations were above the current levels of the eight-hour ozone and 24-hour PM_{2.5} standards. EPA, state, local, and tribal air pollution control agencies collect and report data from thousands of monitors to the Air Quality System (AQS) database. For each NAAQS standard, the criteria pollutant concentrations are averaged over the associated averaging period. For each NAAQS, monitor, and year, we used air quality summary statistics to determine whether there was a concentration above the current level of the standard. For each NAAQS, county, and year, the county exceeds the level of an air quality standard if there was a concentration greater than the level of the standard at any of the county's monitors. In addition to presenting data for each of the criteria pollutants separately, Indicator E1 presents the percentage of children living in counties in which concentrations were above the level of a NAAQS for any criteria air pollutant (i.e., exceedance of standard levels for one or more criteria air pollutants).

Indicator E1 is the total number of children ages 0 to 17 years living in counties with pollutant concentrations above the levels of air quality standards at any time during the year, divided by the total number of children ages 0 to 17 years in the United States. Indicator E1a provides the percentages of children ages 0 to 17 years living in counties with pollutant concentrations above the levels of air quality standards, stratified by race/ethnicity for the year 2021. Indicator E1b provides the percentages of children ages 0 to 17 years living in counties with pollutant concentrations above the levels of air quality standards, stratified by family income, for the year 2021. Indicator E2 is the total number of children ages 0 to 17 years living in counties with a

ⁱ Each NAAQS also has a statistical form: for example the annual 4th highest daily maximum eight-hour average value for ozone, averaged over three years, and has various data completeness requirements. These aspects of the NAAQS are not used in these indicators.

Environments and Contaminants: Criteria Air Pollutants

given number of occurrences of concentrations above the levels of air quality standards for short-term exposure to ozone or PM_{2.5} during the year, divided by the total number of children ages 0 to 17 years in the United States.

Overview of Data Files

The following files are needed to calculate these indicators:

- Annual Summary files. There is one file for each year. For these analyses we extracted data for CO, NO₂, ozone, SO₂, PM_{2.5} (24-hour standard only) and PM₁₀. These comma-separated files contain the pollutant parameter name (PARAMETER NAME), pollutant standard (POLLUTANT STANDARD), unit of measure (UNITS OF MEASURE), code (PARAMETER CODE), Event Type (EVENT TYPE), and sample duration (SAMPLE DURATION), the state and county FIPS codes (STATE CODE, COUNTY CODE), site number (SITE NUM), Pollutant Occurrence Code (POC), year, maximum value (IST MAX VALUE), arithmetic mean (ARITHMETIC MEAN), and the number of exceedances of the primary standard (PRIMARY VIOLATION COUNT). These files were obtained from the website:
https://aqs.epa.gov/aqsweb/airdata/download_files.html
- County PM_{2.5} annual mean exceedance summary data. There is one file for each year. This file contains the state and county names, state and county FIPS codes, and the variable Pm25wtdamviol that indicates whether the PM_{2.5} annual mean NAAQS was exceeded; i.e., whether the weighted annual mean concentration was above the level of the NAAQS at any monitor in the county with sufficiently complete data. These files were obtained as SAS datasets directly from EPA.ⁱⁱ
- Lead maximum rolling three-month averages data for 1999 to 2006. This file contains the site ID, year, and the maximum rolling three-month average lead concentration for each year from 1999 to 2006. This file was obtained as an Excel file directly from EPA.ⁱⁱⁱ
- Lead maximum rolling three-month averages data for 2007 to 2012. Year 2007 from Pb_DesignValues_20072009_Final.xls. Year 2008 from Pb_DesignValues_20082010_FinalRevised.xlsx. Year 2009 from Pb_DesignValues_20092011_FINAL_07_26_12.xlsx. Year 2010 from Pb_DesignValues_20102012_FINAL_07_22_13.xlsx. Year 2011 from Pb_DesignValues_20112013_FINAL_08_13_14.xlsx. Year 2012 from Pb_DesignValues_20122014_FINAL_08_03_15.xlsx. These files contain the site ID, the rolling three-month average lead concentration for each month and year from 2007 to 2012, and the annual maximum rolling three-month average lead concentration for each year from 2011 to 2014. The files were originally obtained from the URL: http://epa.gov/airtrends/values_previous.html. The same

ⁱⁱ David Mintz, EPA OAQPS, 919- 541-5224. Mintz.David@epamail.epa.gov

ⁱⁱⁱ Halil Cakir, EPA OAQPS, 919-541-2416. Cakir.Halil@epa.gov.

Environments and Contaminants: Criteria Air Pollutants

information can currently be found at the URL: <http://www.epa.gov/air-trends/air-quality-design-values>.

- Lead rolling three-month averages data for 2013. Year 2013 from Pb_DesignValues_20132015_FINAL_08_29_16.xlsx. This file contains the site ID and the annual maximum rolling three-month average lead concentration for each year from 2013 to 2015. The files were originally obtained by clicking the link for “Lead Detailed Information” from the URL: <http://epa.gov/airtrends/values.html>. The same information can currently be found at the URL: <http://www.epa.gov/air-trends/air-quality-design-values>.
- Lead rolling three-month averages data for 2014. Year 2014 from Pb_DesignValues_20142016_final_29june17.xlsx. This file contains the site ID and the annual maximum rolling three-month average lead concentration for each year from 2014 to 2016. The file was found at the Lead Design Values, 2016 link at the URL: <http://www.epa.gov/air-trends/air-quality-design-values>.
- Lead rolling three-month averages data for 2015. Year 2015 from pb_designvalues_20152017_final_07_24_18.xlsx. This file contains the site ID and the annual maximum rolling three-month average lead concentration for each year from 2015 to 2017. The file was found at the Lead Design Values, 2017 link at the URL: <http://www.epa.gov/air-trends/air-quality-design-values>
- Lead rolling three-month averages data for 2016 to 2018. Year 2016 from pb_designvalues_20162018_final_06_28_19.xlsx. This file contains the site ID and the annual maximum rolling three-month average lead concentration for each year from 2016 to 2018. The file was found at the Lead Design Values, 2018 link at the URL: <http://www.epa.gov/air-trends/air-quality-design-values>
- Lead rolling three-month averages data for 2017 to 2019. Years 2017 to 2019 from pb_designvalues_2017_2019_final_05_05_20.xlsx. This file contains the site ID and the annual maximum rolling three-month average lead concentration for each year from 2017 to 2019. The file was found at the Lead Design Values, 2019 link at the URL: <http://www.epa.gov/air-trends/air-quality-design-values>
- Lead rolling three-month averages data for 2019 to 2021. Years 2019 to 2021 from pb_designvalues_2019_2021_final_05_04_22.xlsx. This file contains the site ID and the annual maximum rolling three-month average lead concentration for each year from 2019 to 2021. The file was found at the Lead Design Values, 2021 link at the URL: <http://www.epa.gov/air-trends/air-quality-design-values>
- Census data. This file contains the state and county FIPS codes, year, and children’s population. For 1999, we obtained this information from the U.S. Census Bureau files:

Environments and Contaminants: Criteria Air Pollutants

Estimates of the Population of Counties by Age and Sex: 1990-1999, August 30, 2000. The file headers were “(C0-99-9) Population Estimates for Counties by Age and Sex: Annual Time Series July 1, 1990 to July 1, 1999.”

<http://www.census.gov/popest/data/counties/asrh/1990s/CO-99-09.html>.

These files give county populations by age and sex for 1990 to 1999. We summed these populations by year and county across all ages 0 to 17 years and both sexes.

For 2000-2009, we obtained this information from the bridged-race intercensal population files:

National Center for Health Statistics. Intercensal estimates of the resident population of the United States for July 1, 2000-July 1, 2009, by year, county, single-year of age (0, 1, 2, ..., 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative arrangement with the U.S. Census Bureau. Available from: http://www.cdc.gov/nchs/nvss/bridged_race.htm as of October 26, 2012, following release by the U.S. Census Bureau of the revised unbridged intercensal estimates by 5-year age group on October 9, 2012.

The children’s populations by year and county were obtained by summing across the ages 0 to 17 years inclusive.

For 2010-2020, we obtained this information from the bridged race Vintage 2020 postcensal population file:

National Center for Health Statistics. Vintage 2020 postcensal estimates of the resident population of the United States (April 1, 2010, July 1, 2010-July 1, 2020), by year, county, single-year of age (0, 1, 2, ..., 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative arrangement with the U.S. Census Bureau. Available from: http://www.cdc.gov/nchs/nvss/bridged_race.htm as of September 22, 2021, following release by the U.S. Census Bureau of the unbridged Vintage 2019 postcensal estimates by 5-year age group on June 17, 2021.

For 2021, because bridged-race Vintage postcensal population data are no longer provided by the National Center for Health Statistics, we obtained this information from the U.S. Census Bureau files:

“County Characteristics Resident Population Estimates,” July 1, 2021. The file header is “Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2020 to July 1, 2021.” The file description and data link can be found at <https://www.census.gov/data/datasets/time-series/demo/popest/2020s-counties-detail.html>. The data set is located at <https://www2.census.gov/programs-surveys/popest/datasets/2020-2021/counties/asrh/>.

Environments and Contaminants: Criteria Air Pollutants

The children’s populations by year, county, and race/ethnicity were obtained by summing across the ages 0 to 17 inclusive.

National Ambient Air Quality Standards (NAAQS)

Table 1 lists the NAAQS for the criteria pollutants used in indicator E1. The first two columns give the pollutant and the averaging time. The third column gives the corresponding level of the standard. The fourth column gives the value used in determining, for purposes of this indicator, whether a concentration is above the level of a standard, and incorporates a rounding convention. For example, the eight-hour average CO standard’s level of 9 ppm has a rounding convention of 1 ppm, so that the level of the standard is exceeded if any eight-hour average CO concentration rounded to the nearest 1 ppm exceeds 9 ppm; i.e., the unrounded eight-hour average has to equal or exceed the target value of 9.5 ppm. A county exceeds the level of the standard if one or more of the monitors in the county has a concentration greater than the level of the standard. Except for PM_{2.5}, co-located monitors are treated as separate monitors. For PM_{2.5}, if there are several co-located monitors, then only the monitor with the lowest Monitor Number (also known as the Pollutant Occurrence Code, abbreviated as POC) is used. We use PM_{2.5} data from the lowest POC (typically POC 1, the primary sampler) to represent an area.^{iv} For other pollutants, we use data from all monitor/POC combinations.

It should be noted that counties with concentrations above the levels of the standards are not necessarily non-attainment counties for the NAAQS. A non-attainment designation is generally based on three years of data, certain data completeness criteria, and, for short-term standards, requires multiple daily exceedances of the NAAQS. For example, non-attainment of the carbon monoxide eight-hour standard only uses eight-hour averages with at least six hourly values, and occurs when there are two or more eight-hour averages above the NAAQS level. Non-attainment of the eight-hour ozone standard occurs if the three-year average of the fourth highest daily maximum eight-hour average exceeds the level of the standard and the data completeness criteria are met. For a detailed description of the NAAQS attainment and non-attainment calculations, see the NAAQS website at <http://www.epa.gov/naaqs/>.

Table 1: Levels of the National Ambient Air Quality Standards (NAAQS) and the Values Used to Define Exceedances for Indicators E1 and E2*

Pollutant	Standard Averaging Time	Level of the Standard	Target value to define exceedance for indicator calculation**
Carbon monoxide	1-hour 8-hour ^a	35 ppm 9 ppm	Not considered 9.5 ppm
Nitrogen dioxide	Annual 1-hour ^b	0.053 ppm 100 ppb	Not considered 100.5 ppb
Ozone	8-hour ^c	0.070 ppm	0.071 ppm
Lead	Rolling 3-month average ^d	0.15 µg/m ³	0.155 µg/m ³

^{iv} A similar approach was used in the OAQPS report “Analyses of Particulate Matter (PM) Data for the PM NAAQS Review,” June 30, 2005, http://www.epa.gov/ttn/naaqs/standards/pm/data/schmidt_63005.pdf.

Environments and Contaminants: Criteria Air Pollutants

PM ₁₀	24-hour ^c	150 µg/m ³	155 µg/m ³
PM _{2.5}	24-hour ^f	35 µg/m ³	35.5 µg/m ³
	Annual ^g	12 µg/m ³	12.05 µg/m ³
Sulfur dioxide	1-hour ^h	75 ppb	75.5 ppb

* Indicators E1 and E2 are calculated with reference to the current averaging time and level of the air quality standard for all years.

** Standards not used for indicator E1 are shown as “Not considered.”

^a The carbon monoxide 8-hour standard was established in 1971 (See 59 FR 38906, August 1, 1994).

^b The nitrogen dioxide 1-hour standard was established in January 2010 (75 FR 6474, February 9, 2010).

^c The ozone 8-hour standard was adopted in July 1997 (62 FR 38856, July 18, 1997). The level of the ozone 8-hour standard was changed from 0.08 to 0.075 ppm in March 2008 (73 FR 16436, March 27, 2008). The level of the ozone 8-hour standard was changed again from 0.075 to 0.070 ppm in October 2015 (80 FR 95292, October 26, 2015). For the 2015 ozone standard level calculations, all hourly values are 8-hour average values and are reported to three decimal places, truncating digits beyond the third decimal place.

^d The level of the lead standard was changed from 1.5 to 0.15 µg/m³ in October 2008 (73 FR 66964, November 12, 2008). The revised standard uses the maximum rolling three-month average lead concentration.

^e The form, but not the level, of the 24-hour standard for PM₁₀ was revised in July 1997 (62 FR 38652, July 18, 1997).

^f The level of the 24-hour standard for PM_{2.5} was revised from 65 µg/m³ to 35 µg/m³ in 2006 (71 FR 61144, October 17, 2006).

^g The level of the annual standard for PM_{2.5} was revised from 15 µg/m³ to 12 µg/m³ in 2012 (78 FR 3086, January 15, 2013).

^h The final rule for the sulfur dioxide 1-hour standard was issued in June 2010 (75 FR 35220, June 22, 2010).

Air Quality Data

Annual Summary monitoring data from the EPA Air Quality System (AQS) Data Mart for the years 1999 to 2021 were obtained from the EPA website

https://aqz.epa.gov/aqsweb/airdata/download_files.html

The Annual Summary files include the year, state and county FIPS codes, site number, parameter (pollutant) code (PARAMETER CODE), pollutant standard (POLLUTANT STANDARD), and pollutant occurrence code (POC). The Annual Summary files also include the sample duration (SAMPLE DURATION), and the measurement units (UNITS OF MEASURE). The event type (EVENT TYPE) flags whether exceptional events are included in the annual summary statistics. For these analyses we included both the event types “No Events” and “Events Included” (sometimes coded as “Events Included,”sic) that denote cases where either there were no exceptional events for the year (“No Events”) or where there were exceptional events that year and the analysis included data for those events (“Events Included” or “Events Included”). Also included are the annual maximum value (1ST MAX VALUE) and the number of exceedances of the primary standard (PRIMARY VIOLATION COUNT).

The following combinations of the parameter code, pollutant standard, sample duration, measurement unit, maximum, and arithmetic mean were used to define exceedances of the standards used for indicator E1:

Environments and Contaminants: Criteria Air Pollutants

- CO 8-hour standard: PARAMETER CODE = 42101, POLLUTANT STANDARD = “CO 8-hour 1971”, SAMPLE DURATION = “8-HR RUN AVG END HOUR” or “8-HR RUN”, UNITS OF MEASURE = “Parts per million”, 1ST MAX VALUE ≥ 9.5
- NO₂ 1-hour standard: *Either* PARAMETER CODE = 42602, POLLUTANT STANDARD = “NO2 1-hour”, SAMPLE DURATION = “1 HOUR”, UNITS OF MEASURE = “Parts per million”, 1ST MAX VALUE ≥ 0.1005 *Or* PARAMETER CODE = 42602, POLLUTANT STANDARD = “NO2 1-hour”, SAMPLE DURATION = “1 HOUR”, UNITS OF MEASURE = “Parts per billion”, 1ST MAX VALUE ≥ 100.5
- Ozone 8-hour standard: PARAMETER CODE = 44201, POLLUTANT STANDARD = “Ozone 8-hour 2015”, SAMPLE DURATION = “8-HR RUN AVG BEGIN HOUR” or “8-HR RUN”, UNITS OF MEASURE = “Parts per million”, 1ST MAX VALUE ≥ 0.071
- PM₁₀ 24-hour standard: PARAMETER CODE = 81102, POLLUTANT STANDARD = “PM10 24-hour 2006”, SAMPLE DURATION = “24 HOUR”, “24 HOURS”, “24-HR BLK AVG” or “24-HR BLK”, UNITS OF MEASURE = “Micrograms/cubic meter (25 C)”, 1ST MAX VALUE ≥ 155
- SO₂ 1-hour standard: *Either* PARAMETER CODE = 42401, POLLUTANT STANDARD = “SO2 1-hour 2010”, SAMPLE DURATION = “1 HOUR”, UNITS OF MEASURE = “Parts per million”, 1ST MAX VALUE ≥ 0.0755 *Or* PARAMETER CODE = 42401, POLLUTANT STANDARD = “SO2 1-hour 2010”, SAMPLE DURATION = “1 HOUR”, UNITS OF MEASURE = “Parts per billion”, 1ST MAX VALUE ≥ 75.5
- PM_{2.5} 24-hour standard: PARAMETER CODE = 88101, POLLUTANT STANDARD = “PM25 24-hour 2012”, SAMPLE DURATION = “24 HOUR” or “24-HR BLK AVG”, UNITS OF MEASURE = “Micrograms/cubic meter (LC)”, 1ST MAX VALUE ≥ 35.5

The monitor exceeds the level of the standard if the corresponding set of conditions holds for the given year. A county exceeds the level of the standard in a given year if any of the monitors in that county exceed the level of the standard.

For the PM_{2.5} annual standard, summary data used for NAAQS attainment/non-attainment designations for the years 2011 to 2013 were obtained directly from EPA in October and November 2017 (data for the years 1999 to 2010 were previously obtained from EPA). The PM_{2.5} annual standard summary data for the year 2014 were obtained directly from EPA in November 2018. The PM_{2.5} annual standard summary data for the year 2015 were obtained directly from EPA in February 2020. The PM_{2.5} annual standard summary data for the years 2016 to 2021 were obtained directly from EPA.^v We extracted the state and county FIPS codes for all years. These were used together with the variable PM25wtdamviol, which has the value 1 if the county has a concentration above the level of the PM_{2.5} annual standard. In this case, a county exceeds the level of the standard if, at any monitor in the county, the NAAQS data completeness requirements are met and the weighted annual mean is above the target value of 12.05 $\mu\text{g}/\text{m}^3$. The weighted annual mean is the average of the four quarterly means. This approach often gives different results from the values for PM_{2.5} annual standard in the Annual Summary files, which

^vDavid Mintz, EPA OAQPS, 919- 541-5224. Mintz.David@epamail.epa.gov

Environments and Contaminants: Criteria Air Pollutants

do not apply data completeness rules and are based on the annual arithmetic mean of all the daily values.

For the lead standard, summary data with the annual maximum rolling three-month average lead concentrations for each monitor from 1999 to 2006 were obtained directly from EPA on October 14, 2009.^{vi} These values were calculated using the completeness criteria and averaging calculations detailed in the Federal Register, Vol. 73, No. 219. Thus, the maximum rolling three-month average is calculated by finding all 12 three-month average lead concentrations that begin in that calendar year, and finding the maximum of those 12 averages. For the year 2007, rolling three-month average lead concentrations were obtained from the Excel file Pb_DesignValues_20072009_Final.xls at the link for “Lead Detailed Information” for the end year 2009 from the URL: http://epa.gov/airtrends/values_previous.html. For the year 2008, rolling three-month average lead concentrations were obtained from the Excel file for the end year 2010 Pb_DesignValues_20082010_FinalRevised.xlsx at the same URL. For the year 2009, rolling three-month average lead concentrations were obtained from the Excel file for the end year 2011 Pb_DesignValues_20092011_Final.xlsx at the same URL. For the year 2010, rolling three-month average lead concentrations were obtained from the Excel file for the end year 2012 Pb_DesignValues_20102012_Final_07_22_13.xlsx at the same URL. For the year 2011, annual maximum rolling three-month average lead concentrations were obtained from the Excel file for the end year 2013 Pb_DesignValues_20112013_Final_03_13_14.xlsx at the same URL. For the year 2012, annual maximum rolling three-month average lead concentrations were obtained from the Excel file for the end year 2014 Pb_DesignValues_20122014_FINAL_08_03_15.xlsx at the URL: <http://epa.gov/airtrends/values.html>. For the year 2013, annual maximum rolling three-month average lead concentrations were obtained from the Excel file for the end year 2015 Pb_DesignValues_20132015_FINAL_08_29_16.xlsx at the URL: <http://epa.gov/airtrends/values.html>. For the year 2014, annual maximum rolling three-month average lead concentrations were obtained from the Excel file for the end year 2016 Pb_DesignValues_20142016_final_29june17.xlsx at the URL: <http://www.epa.gov/air-trends/air-quality-design-values>. For the year 2015, annual maximum rolling three-month average lead concentrations were obtained from the Excel file for the end year 2017 pb_designvalues_20152017_final_07_24_18.xlsx at the URL: <http://www.epa.gov/air-trends/air-quality-design-values>. For the year 2016, annual maximum rolling three-month average lead concentrations were obtained from the Excel file for the end year 2018 pb_designvalues_20162018_final_06_28_19.xlsx at the URL: <http://www.epa.gov/air-trends/air-quality-design-values>. For the years 2017 to 2019, annual maximum rolling three-month average lead concentrations were obtained from the Excel file for the end year 2019 pb_designvalues_2017_2019_final_05_05_20.xlsx at the URL: <http://www.epa.gov/air-trends/air-quality-design-values>. For the years 2019 to 2021, annual maximum rolling three-month average lead concentrations were obtained from the Excel file for the end year 2021 pb_designvalues_2019_2021_final_05_04_22.xlsx at the URL: <http://www.epa.gov/air-trends/air-quality-design-values>.

^{vi} Mark Schmidt, EPA OAQPS, 919-541-2416. Schmidt.Mark@epamail.epa.gov. Current contact: Halil Cakir, EPA OAQPS, 919-541-2416. Cakir.Halil@epa.gov.

Environments and Contaminants: Criteria Air Pollutants

The annual maximum rolling three-month average lead concentrations for 2007 to 2010 were calculated from the monthly values. The monitor has a concentration above the level of the standard for the given year if the annual maximum rolling three-month average lead concentration is at least $0.155 \mu\text{g}/\text{m}^3$. A county is considered to have a concentration above the level of the standard in a given year if any of the monitors in that county exceed the level of the standard.

For indicator E1, we also evaluated the counties exceeding Any Standard; i.e., counties with pollutant concentrations above the levels of one or more of the eight NAAQS standards used for indicator E1.

For indicator E2, we used the same Annual Summary files to obtain the numbers of days that the ozone or $\text{PM}_{2.5}$ standards were exceeded. For ozone we used the observations for PARAMETER CODE = 44201, POLLUTANT STANDARD = “Ozone 8-hour 2015”, SAMPLE DURATION = “8-HR RUN AVG BEGIN HOUR” or “8-HR RUN”, UNITS OF MEASURE = “Parts per million” and the variable PRIMARY VIOLATION COUNT, giving the number of days where the maximum 8-hour average concentration exceeded the NAAQS level. For $\text{PM}_{2.5}$ we used the observations for PARAMETER CODE = 88101, POLLUTANT STANDARD = “PM25 24-hour 2012”, SAMPLE DURATION = “24 HOUR” or “24-HR BLK AVG”, UNITS OF MEASURE = “Micrograms/cubic meter (LC)” and the variable PRIMARY VIOLATION COUNT, giving the number of days where the 24-hour average concentration exceeded the NAAQS level. For $\text{PM}_{2.5}$ we only used the data from the lowest numbered POC at each monitoring site. The number of exceedances for a county is calculated as the maximum number of days with concentrations above the levels of the standards across all monitors in that county.

Census Data

For the trend analyses we obtained children’s populations by county for each year from 1999-2021.

For 1999, the source was U.S. Census Bureau files, Estimates of the Population of Counties by Age and Sex: 1990-1999, August 30, 2000. The file headers were “(C0-99-9) Population Estimates for Counties by Age and Sex: Annual Time Series July 1, 1990 to July 1, 1999.” <http://www.census.gov/popest/data/counties/asrh/1990s/CO-99-09.html>. These files give county populations by age and sex for 1990 to 1999. We summed these populations by year and county across all ages 0 to 17 and both sexes.

For 2000-2009, we used the bridged-race intercensal population files obtained from the CDC website:

National Center for Health Statistics. Intercensal estimates of the resident population of the United States for July 1, 2000-July 1, 2009, by year, county, single-year of age (0, 1, 2, ..., 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative arrangement with the U.S. Census Bureau. Available from: http://www.cdc.gov/nchs/nvss/bridged_race.htm as of October 26, 2012, following release by

Environments and Contaminants: Criteria Air Pollutants

the U.S. Census Bureau of the revised unbridged intercensal estimates by 5-year age group on October 9, 2012.

The bridged-race intercensal population files contain estimates of the resident population of the United States as of July 1, 2000; July 1, 2001; July 1, 2002; July 1, 2003; July 1, 2004; July 1, 2005; July 1, 2006; July 1, 2007; July 1, 2008; and July 1, 2009 by county, single-year of age (0, 1, 2,..., 85 years and over), bridged-race category (White, Black or African American, American Indian or Alaska Native, Asian or Pacific Islander), Hispanic origin (not Hispanic or Latino, Hispanic or Latino), and sex. There is one SAS file for the years 2000-2004 and another SAS file for the years 2005-2009. Files are available in SAS dataset and text formats; we used the SAS dataset format for these analyses.

We extracted the following variables: state, county, age, racesex, hisp, and pop2000 to pop2009. The racesex variable is a single coded value for each combination of race and sex, e.g., racesex = 1 denotes White males. The value popXXXX gives the population as of July 1 of the calendar year XXXX for a given state, county, age, racesex combination, and ethnicity. The county children's populations for each year 2000-2009 were obtained by summing the variable popXXXX over all ages <= 17, all values of "racesex," and all values of "hisp." The county children's populations for each year 2000-2009 and each race/ethnicity group were obtained by summing the variable popXXXX over all ages <= 17 for the applicable values of "racesex" and "hisp."

For 2010 to 2020, we used the bridged race Vintage 2020 postcensal population file:

National Center for Health Statistics. Vintage 2020 postcensal estimates of the resident population of the United States (April 1, 2010, July 1, 2010-July 1, 2020), by year, county, single-year of age (0, 1, 2, ..., 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative arrangement with the U.S. Census Bureau. Available from: http://www.cdc.gov/nchs/nvss/bridged_race.htm as of September 22, 2021, following release by the U.S. Census Bureau of the unbridged Vintage 2020 postcensal estimates by 5-year age group on June 17, 2021.

The bridged race 2010 to 2020 population files contain estimates of the resident population of the United States as of July 1, 2010, July 1, 2011, July 1, 2012, July 1, 2012, July 1, 2013, July 1, 2014, July 1, 2015, July 1, 2016, July 1, 2017, July 1, 2018, July 1, 2019, and July 1, 2020 by county, single-year of age (0, 1, 2,..., 85 years and over), bridged-race category (White, Black or African American, American Indian or Alaska Native, Asian or Pacific Islander), Hispanic origin (not Hispanic or Latino, Hispanic or Latino), and sex. Files are available in SAS dataset and text formats; we used the SAS dataset format for these analyses.

We extracted the following variables: state, county, age, racesex, hisp, pop2010, pop2011, pop2012, pop2013, pop2014, pop2015, pop2016, pop2017, pop2018, pop 2019, and pop2020. The racesex variable is a single coded value for each combination of race and sex, e.g., racesex = 1 denotes White males. The value popXXXX gives the population as of July 1, XXXX for a given state, county, age, racesex combination, and ethnicity. The county children's populations for the years 2010 to 2020 were obtained by summing the variables pop2010, pop2011, pop2012,

Environments and Contaminants: Criteria Air Pollutants

pop2013, pop2014, pop2015, pop2016, pop2017, pop2018, pop2019, and pop2020 over all ages <= 17, all values of “racesex,” and all values of “hisp.” The county children’s populations for each year 2010-2020 and each race/ethnicity group were obtained by summing the variable popXXXX over all ages <= 17 for the applicable values of “racesex” and “hisp.” For 2021, the source was U.S. Census Bureau files:

The file header is “CC-EST2021-ALLDATA-[ST-FIPS]: Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2020 to July 1, 2021.” The file is released in June 2022. <https://www.census.gov/data/datasets/time-series/demo/popest/2020s-counties-detail.html>.

This file gives county populations by age groups, sex, unbridged race groups, and Hispanic Origin. We summed these populations by year and state across all counties in the state, all ages 17 years and under, and both sexes.

The unbridged race 2020 to 2021 population file contains estimates of the resident population of the United States as of July 1, 2020, and July 1, 2021. Unbridged race categories (American Indian and Alaska Native alone non-Hispanic, White alone non-Hispanic, Black or African American alone non-Hispanic, Asian alone non-Hispanic, Native Hawaiian and Other Pacific Islander alone non-Hispanic, Two or More Races non-Hispanic, and Hispanic) were used in the analyses of 2021.

We extracted the following variables: state, county, agegrp, and 7 race/ethnicity groups (“NHIA”, “NHWA”, “NHBA”, “NHAA”, “NHNA”, “NHTOM”, and “H”) for both males and females. The group “ALL” was the sum of all persons in those 7 race/ethnicity groups. For each state, county, and AGEGRP, we summed the male and female populations for each of the following list of race/ethnicity groups which use the unbridged race groupings: “All” = All persons, “AIANNH” = AIAN alone non-Hispanic (NH), “WNNH” = White alone NH, “BNH” = Black or African American alone NH, “ANH” Asian alone NH, “NHOPINH” = Native Hawaiian and Other Pacific Islander alone NH, “TwoNH” = two or more races NH, “H” = Hispanic. The populations of children ages 0 to 17 years were calculated by summing the populations for AGEGRP = 1, 2, and 3, and adding 3/5 of the population for AGEGRP = 4. That is because the AGEGRPs are “1” = ages 0 to 4, “2” = ages 5 to 9, “3” = ages 10 to 14, and “4” = ages 15 to 19, assuming the ages 15-19 are equally distributed.

Calculation of Indicator

Indicator E1 is calculated as follows. Define, for each NAAQS,

Pop (county C, year Y) = population of children in county C, year Y

Exceed (county C, year Y) = 1, if county C has a pollutant concentration greater than the level of the NAAQS in year Y
= 0, otherwise

Environments and Contaminants: Criteria Air Pollutants

These values are calculated as described above. Note that Exceed = 0 if the county has no air quality monitors or no air quality monitoring data in the given year. Counties outside the 50 states or Washington, DC were excluded.

The total number of children living in counties in which concentrations were above the levels of air quality standards is the sum of Pop over all counties where Exceed equals 1; i.e.,

$$\text{Children Affected (year Y)} = \sum [\text{Pop (county C, year Y)} \times \text{Exceed (county C, year Y)}]$$

The total number of children living in the United States is the sum of Pop over all counties. Thus,

$$\text{Children (year Y)} = \sum \text{Pop (county C, year Y)}$$

The percentage of children living in counties in which pollutant concentrations were above the levels of air quality standards equals the total number of children living in counties in which concentrations were above the levels of air quality standards divided by the total number of children living in the United States and multiplied by 100.

$$\text{Percentage Children Affected} = [\text{Children Affected (year Y)} / \text{Children (year Y)}] \times 100\%$$

Indicator E2 is calculated as follows. Define, for each NAAQS and each exceedance range a to b (e.g., 11 to 25, or 26 to infinity),

Pop (county C, year Y) = population of children in county C, year Y.

$$\begin{aligned} \text{ExceedRange (county C, year Y)} &= 1, \text{ if county C has K exceedances of the} \\ &\text{NAAQS in year Y and } a \leq K \leq b, \\ &= 0, \text{ otherwise} \end{aligned}$$

These values are calculated as described above. Note that ExceedRange = 0 if the county has no air quality monitors or no air quality monitoring data in the given year. Counties outside the 50 states or Washington, DC were excluded.

The total number of children living in counties in which there were between a and b exceedances of the levels of air quality standards is the sum of Pop over all counties where ExceedRange equals 1; i.e.,

$$\text{Children Affected (year Y)} = \sum [\text{Pop (county C, year Y)} \times \text{ExceedRange (county C, year Y)}]$$

The total number of children living in the United States is the sum of Pop over all counties. Thus,

$$\text{Children (year Y)} = \sum \text{Pop (county C, year Y)}$$

Environments and Contaminants: Criteria Air Pollutants

The percentage of children living in counties in which there were between a and b exceedances of the levels of air quality standards equals the total number of children living in counties in which there were between a and b exceedances of the levels of air quality standards divided by the total number of children living in the United States and multiplied by 100.

Percentage Children Affected = [Children Affected (year Y) / Children (year Y)] × 100%

Race/Income

Indicator E1a and E1b give the percentages of children living in counties in which pollutant concentrations were above the levels of air quality standards, stratified by race/ethnicity for the year 2021 (Indicator E1a) or family income for the year 2021 (Indicator E1b). The calculation of Indicators E1a and E1b is exactly the same as for the trend analysis, except that the populations of children by county are replaced by estimated populations of children by county for the given bridged or unbridged race/ethnicity or income groups.

For Indicator E1a, children's populations stratified by race/ethnicity for 2021 were estimated for each county by using the unbridged-race 2021 post-censal population file for July 1, 2021 at the url <https://www.census.gov/data/datasets/time-series/demo/popest/2020s-counties-detail.html> to obtain race/ethnicity distributions. The race/ethnicity groups were summed as follows:

- White alone NH: NHWA_MALE and NHWA_FEMALE, AGEGRP = 1, 2, and 3, and adding 3/5 of 4.
- Black or African American alone NH: NHBA_MALE and NHBA_FEMALE, AGEGRP = 1, 2, and 3, and adding 3/5 of 4.
- Asian alone NH: NHAA_MALE and NHAA_FEMALE, AGEGRP = 1, 2, and 3, and adding 3/5 of 4.
- Native Hawaiian and Other Pacific Islander alone NH: NHNA_MALE and NHNA_FEMALE, AGEGRP = 1, 2, and 3, and adding 3/5 of 4.
- American Indian and Alaska Native alone NH: NHIA_MALE and NHIA_FEMALE, AGEGRP = 1, 2, and 3, and adding 3/5 of 4.
- Two or more races NH: NHTOM_MALE and NHTOM_FEMALE, AGEGRP = 1, 2, and 3, and adding 3/5 of 4.
- Hispanic: H_MALE and H_FEMALE, AGEGRP = 1, 2, and 3, and adding 3/5 of 4.
- All: sum of populations above.

For Indicator E1b, children's populations stratified by income for 2021 were obtained using data from the Census Bureau's Small Area Poverty and Income Estimates (SAIPE) program combined with the populations from the bridged-race Vintage 2019 post-censal population file for July 1, 2020 and unbridged-race 2021 post-censal population file for July 1, 2021. The SAIPE population estimates were developed by the Census Bureau using the American Community Survey (ACS) and Census 2010 data, based on statistical models. The SAIPE data were obtained from the url

<https://www.census.gov/data/datasets/2021/demo/saipe/2021-state-and-county.html>

Environments and Contaminants: Criteria Air Pollutants

as the Excel file est21ALL.xls. The Excel file includes the state and county FIPS codes and the poverty percentage under age 18. For each county, the total children's population was calculated by summing populations across all children ages 0 to 17 years. The children's population below the poverty level was obtained by multiplying the total children's population by the poverty percentage under age 18. The children's population at or above the poverty level was obtained by multiplying the total children's population by 100 minus the poverty percentage under age 18:

$$\text{Population below poverty} = \text{Total population} \times \text{poverty percentage under age 18}/100,$$
$$\text{Population at or above poverty} = \text{Total population} \times (100 - \text{poverty percentage under age 18})/100.$$

Questions and Comments

Questions regarding these methods, and suggestions to improve the description of the methods, are welcome. Please use the "Contact Us" link at the bottom of any page in the America's Children and the Environment website.

Statistical Comparisons

Statistical analyses of the percentages of children ages 0 to 17 years living in counties with concentrations above the levels of air quality standards were used to determine whether the trends in the percentages for different years were statistically significant. Using a logistic regression model, the logarithm of the odds that a child lives in a county with concentrations above the levels of air quality standards is regressed against the calendar year. To avoid potential statistical issues associated with very low or very high percentages, each percentage was first rescaled to be between 5% and 95% before computing the odds. The error terms for this logistic regression were assumed to be approximately independent and normally distributed. For each criteria air pollutant, the slope of the regression line for the logarithm of the odds was computed, together with a 95% confidence interval and its p-value. The slope estimates the annual change in the logarithm of the odds. A p-value at or below 0.05 implies that the trend is statistically significant at the 5% significance level. No adjustment is made for multiple comparisons.

The results are presented in Table 2. For more details on these statistical analyses, see the memorandum by Cohen (2010).^{vii}

Table 2: Logistic regression trend test for the proportion of children ages 0 to 17 years living in counties in which pollutant concentrations were above the levels of air quality standards in years 1999 to 2019.

^{vii} Statistical methods for testing for trends and year-to-year changes to air quality measures. Memorandum from Jonathan Cohen, ICF, to Dan Axelrad, EPA, November 2010.

Environments and Contaminants: Criteria Air Pollutants

Pollution Type	N	Trend (annual change in log odds)	95% Confidence Interval for Trend: Lower Bound	95% Confidence Interval for Trend: Upper Bound	P-value for Trend
Ozone (eight-hour)	23	-0.037	-0.047	-0.026	< 0.001
PM₁₀ (24-hour)	23	0.001	-0.014	0.016	0.875
PM_{2.5} (24-hour)	23	-0.075	-0.099	-0.052	< 0.001
PM_{2.5} (annual)	23	-0.112	-0.136	-0.088	< 0.001
Carbon monoxide	23	-0.022	-0.033	-0.011	0.001
Lead	23	-0.016	-0.033	0.000	0.054
Sulfur dioxide (one-hour)	23	-0.095	-0.103	-0.088	< 0.001
Nitrogen dioxide (one-hour)	23	-0.090	-0.103	-0.078	< 0.001
Any standard	23	-0.048	-0.058	-0.039	< 0.001

Methods

Indicator

E3. Percentage of days with good, moderate, or unhealthy air quality for children ages 0 to 17 years, 1999-2021.

Summary

EPA's Air Quality Index (AQI) classifies air quality as good, moderate, or unhealthy on a daily basis, for each county in the United States. The AQI is based on monitoring data for six criteria air pollutants. For each county, year, and day, the air quality index category for that day was obtained from EPA OAQPS. For each county and year, the annual numbers of days categorized as good, moderate, or unhealthy were counted. Indicator E3 is the national weighted average of these annual numbers, where the weights are the numbers of children ages 0 to 17 years living in each county, obtained from the Census Bureau. This is mathematically equivalent to the percentage of children's days in each category. The supplementary indicators for Indicator E3 provide the normalized percentages of days for children ages 0 to 17 years with good, moderate, or unhealthy air quality, stratified by race/ethnicity (Indicator E3a) or income (Indicator E3b), for the year 2021.

Overview of Data Files

The following files are needed to calculate this indicator:

- Annual monitoring and AQI data. This file contains the state and county names, days with AQI, good days, moderate days, unhealthy days under multiple categories, max AQI, and corresponding AQI for each parameter.
- State and county FIPS data. This file contains FIPS code for each county in the U.S. We obtained this information from National Bureau of Economic Research: <https://www.nber.org/research/data/ssa-federal-information-processing-series-fips-state-and-county-crosswalk>. We also checked that the FIPS codes were consistent with codes provided by Federal Communications Commission: <https://transition.fcc.gov/oet/info/maps/census/fips/fips.txt>.
- Census data. This file contains the state and county FIPS codes, year, and children's population. For 1999, we obtained this information from the U.S. Census Bureau files:

Estimates of the Population of Counties by Age and Sex: 1990-1999, August 30, 2000. The file headers were "(C0-99-9) Population Estimates for Counties by Age and Sex: Annual Time Series July 1, 1990 to July 1, 1999."

<http://www.census.gov/popest/data/counties/asrh/1990s/CO-99-09.html>.

These files give county populations by age and sex for 1990 to 1999. We summed these populations by year and county across all ages 0 to 17 years and both sexes.

Environments and Contaminants: Criteria Air Pollutants

For 2000-2009, we obtained this information from the bridged-race intercensal population files:

National Center for Health Statistics. Intercensal estimates of the resident population of the United States for July 1, 2000-July 1, 2009, by year, county, single-year of age (0, 1, 2, ..., 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative arrangement with the U.S. Census Bureau. Available from: http://www.cdc.gov/nchs/nvss/bridged_race.htm as of October 26, 2012, following release by the U.S. Census Bureau of the revised unbridged intercensal estimates by 5-year age group on October 9, 2012.

The children's populations by year and county were obtained by summing across the ages 0 to 17 years inclusive.

For 2010-2020, we obtained this information from the bridged race Vintage 2020 postcensal population file:

National Center for Health Statistics. Vintage 2020 postcensal estimates of the resident population of the United States (April 1, 2010, July 1, 2010-July 1, 2020), by year, county, single-year of age (0, 1, 2, ..., 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative arrangement with the U.S. Census Bureau. Available from: http://www.cdc.gov/nchs/nvss/bridged_race.htm as of September 22, 2021, following release by the U.S. Census Bureau of the unbridged Vintage 2020 postcensal estimates by 5-year age group on June 17, 2021.

For 2021, because bridged-race Vintage postcensal population data are no longer provided by National Center for Health Statistics, we obtained this information from the U.S. Census Bureau files:

“County Characteristics Resident Population Estimates,” July 1, 2021. The file header is “Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2020 to July 1, 2021.” The file description and data link can be found at <https://www.census.gov/data/datasets/time-series/demo/popest/2020s-counties-detail.html>. The data set is located at <https://www2.census.gov/programs-surveys/popest/datasets/2020-2021/counties/asrh/>.

The children's populations by year, county, and race/ethnicity were obtained by summing across the ages 0 to 17 inclusive.

Air Quality Data

Annual county-level Air Quality Index (AQI) and maximum monitoring data from the EPA Air Quality System (AQS) for the years 1999 to 2021 were downloaded in June 2023 from the annual summary files by county at the EPA Air Data website:

Environments and Contaminants: Criteria Air Pollutants

https://aqs.epa.gov/aqsweb/airdata/download_files.html#AQI

These annual county-level summary files contain the state and county names, days with AQI, good days, moderate days, unhealthy days under multiple categories, max AQI, and corresponding AQI for each parameter. The following air quality index categories were assigned: Good, Moderate, Unhealthy (combined to include Unhealthy for Sensitive Groups, Unhealthy, Very Unhealthy, and Hazardous).

For each county and year, we used the total number of days in the good, moderate, and unhealthy categories and also the number of days without data (by subtraction from the number of days in each year). Counties not in the 50 states or Washington, DC were excluded. For county/year combinations not in the AQI data, we assumed zero days in the good, moderate, and unhealthy categories and 365 (366 for leap years) days without data.

State and County FIPS data

FIPS codes for each county in the 50 states and D.C. were obtained from National Bureau of Economic Research:

<https://www.nber.org/research/data/ssa-federal-information-processing-series-fips-state-and-county-crosswalk>.

FIPS codes from the above source were spot-checked to ensure consistency with FIPS codes provided by Federal Communications Commission:

<https://transition.fcc.gov/oet/info/maps/census/fips/fips.txt>.

A SAS version of the FIPS code data was downloaded with only state name, county name, and corresponding FIPS code were retained for use. Counties that had duplicate FIPS codes (because two or more SSA codes were assigned) were identified and deduplicated. FIPS codes were merged into the annual AQI datasets based on state name and county name. Due to some naming differences between AQI data and FIPS records (e.g., DeKalb vs De Kalb), some counties in AQI data were not assigned with a valid FIPS code. Unmatched county names between the annual AQI datasets and the FIPS dataset were identified in each run by printing out counties with missing FIPS code. These were manually corrected by editing the county names in FIPS data to be consistent with the AQI datasets in the data processing steps.

Census Data

For the trend analyses we obtained children's populations by county for each year from 1999 to 2021.

For 1999, the source was U.S. Census Bureau files:

Environments and Contaminants: Criteria Air Pollutants

Estimates of the Population of Counties by Age and Sex: 1990-1999, August 30, 2000. The file headers are “(C0-99-9) Population Estimates for Counties by Age and Sex: Annual Time Series July 1, 1990 to July 1, 1999.”
<http://www.census.gov/popest/data/counties/asrh/1990s/CO-99-09.html>.

These files give county populations by age and sex for 1990 to 1999. We summed these populations by year and county across all ages 0 to 17 years and both sexes.

For 2000-2009, we used the bridged-race intercensal population files obtained from the CDC website:

National Center for Health Statistics. Intercensal estimates of the resident population of the United States for July 1, 2000-July 1, 2009, by year, county, single-year of age (0, 1, 2, ..., 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative arrangement with the U.S. Census Bureau. Available from:
http://www.cdc.gov/nchs/nvss/bridged_race.htm as of October 26, 2012, following release by the U.S. Census Bureau of the revised unbridged intercensal estimates by 5-year age group on October 9, 2012.

The bridged-race intercensal population files contain estimates of the resident population of the United States as of July 1, 2000; July 1, 2001; July 1, 2002; July 1, 2003; July 1, 2004; July 1, 2005; July 1, 2006; July 1, 2007; July 1, 2008; and July 1, 2009 by county, single-year of age (0, 1, 2, ..., 85 years and over), bridged-race category (White, Black or African American, American Indian or Alaska Native, Asian or Pacific Islander), Hispanic origin (not Hispanic or Latino, Hispanic or Latino), and sex. There is one SAS file for the years 2000-2004 and another SAS file for the years 2005-2009. Files are available in SAS dataset and text formats; we used the SAS dataset format for these analyses.

We extracted the following variables: state, county, age, racesex, hisp, and pop2000 to pop2009. The racesex variable is a single coded value for each combination of race and sex, e.g., racesex = 1 denotes White males. The value popXXXX gives the population as of July 1 of the calendar year XXXX for a given state, county, age, racesex combination, and ethnicity. The county children’s populations for each year 2000-2009 were obtained by summing the variable popXXXX over all ages ≤ 17 , all values of “racesex,” and all values of “hisp.”

For 2010-2020, we obtained this information from the bridged race Vintage 2020 postcensal population file:

National Center for Health Statistics. Vintage 2020 postcensal estimates of the resident population of the United States (April 1, 2010, July 1, 2010-July 1, 2020), by year, county, single-year of age (0, 1, 2, ..., 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative arrangement with the U.S. Census Bureau. Available from:
http://www.cdc.gov/nchs/nvss/bridged_race.htm as of September 22, 2021, following release

Environments and Contaminants: Criteria Air Pollutants

by the U.S. Census Bureau of the unbridged Vintage 2020 postcensal estimates by 5-year age group on June 17, 2021.

The bridged race 2010 to 2020 population files contain estimates of the resident population of the United States as of July 1, 2010, July 1, 2011, July 1, 2012, July 1, 2013, July 1, 2014, July 1, 2015, July 1, 2016, and July 1, 2017, July 2018, July 2019, and July 2020 by county, single-year of age (0, 1, 2,..., 85 years and over), bridged-race category (White, Black or African American, American Indian or Alaska Native, Asian or Pacific Islander), Hispanic origin (not Hispanic or Latino, Hispanic or Latino), and sex. Files are available in SAS dataset and text formats; we used the SAS dataset format for these analyses.

We extracted the following variables: state, county, age, racesex, hisp, pop2010, pop2011, pop2012, pop2013, pop2014, pop2015, pop2016, pop2017, pop2018, pop2019, and pop2020. The racesex variable is a single coded value for each combination of race and sex, e.g., racesex = 1 denotes White males. The value popXXXX gives the population as of July 1, XXXX for a given state, county, age, racesex combination, and ethnicity. The county children's populations for the years 2010 to 2020 were obtained by summing the variables pop2010, pop2011, pop2012, pop2013, pop2014, pop2015, pop2016, pop2017, pop2018, pop2019, and pop2020 over all ages <= 17, all values of "racesex," and all values of "hisp." The county children's populations for each year 2010-2020 and each race/ethnicity group were obtained by summing the variable popXXXX over all ages <= 17 for the applicable values of "racesex" and "hisp."

For 2021, the source was U.S. Census Bureau files:

The file header is "CC-EST2021-ALLDATA-[ST-FIPS]: Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2020 to July 1, 2021." The file is released in June 2022. <https://www.census.gov/data/datasets/time-series/demo/popest/2020s-counties-detail.html>.

This file gives county populations by age groups, sex, unbridged race groups, and Hispanic Origin. We summed these populations by year and state across all counties in the state, all ages 17 years and under, and both sexes.

The unbridged race 2020 to 2021 population file contains estimates of the resident population of the United States as of July 1, 2020, and July 1, 2021. Unbridged race categories (American Indian and Alaska Native alone non-Hispanic, White alone non-Hispanic, Black or African American alone non-Hispanic, Asian alone non-Hispanic, Native Hawaiian and Other Pacific Islander alone non-Hispanic, Two or More Races non-Hispanic, and Hispanic) were used in the analyses of 2021.

We extracted the following variables: state, county, agegrp, and 7 race/ethnicity groups ("NHIA", "NHWA", "NHBA", "NHAA", "NHNA", "NHTOM", and "H") for both males and females. The group "ALL" was the sum of all persons in those 7 race/ethnicity groups. For each state, county, and AGEGRP, we summed the male and female populations for each of the following list of race/ethnicity groups which use the unbridged race groupings: "All" = All

Environments and Contaminants: Criteria Air Pollutants

persons, “AIANNH” = AIAN alone non-Hispanic (NH) , “WNH” = White alone NH, “BNH” = Black or African American alone NH, “ANH” Asian alone NH, “NHOPINH” = Native Hawaiian and Other Pacific Islander alone NH, “TwoNH” = two or more races NH, “H” = Hispanic. The populations of children ages 0 to 17 years were calculated by summing the populations for AGEGRP = 1, 2, and 3, and adding 3/5 of the population for AGEGRP = 4. That is because the AGEGRPs are “1” = ages 0 to 4, “2” = ages 5 to 9, “3” = ages 10 to 14, and “4” = ages 15 to 19, assuming the ages 15-19 are equally distributed.

Calculation of Indicator

Indicator E3 is calculated as follows. The percentage of days is calculated directly for three categories: Good, Moderate, or Unhealthy. The calculation is the same for each category; the following example is for the Good category. Define

Days (category G, county C, year Y) = number of days in AQI category G for county C, year Y

Pop (county C, year Y) = population of children in county C, year Y

These values are calculated as described above.

1. The county number of children’s days in category G is calculated as the number of G days in county C times the number of children in county C:

County children’s G days (G, C, Y) = Days (G, C, Y) × Pop (C, Y)

2. The national number of children’s days in category G is calculated by summing the county numbers over the counties:

National children’s G days (G, Y) = Σ County children’s G days in G (G, C, Y)

where this sum is taken across all counties in all 50 states plus Washington, DC.

3. The county number of children’s days (in all categories) is calculated as the number of days in year Y (365, or 366 for leap years) times the county number of children:

County children’s days (C, Y) = 365 (or 366) × Pop (C, Y)

4. The national number of children’s days is calculated by summing the county numbers over the counties:

National children’s days (Y) = Σ County children’s days (C, Y)

where this sum is taken across all counties in all 50 states plus Washington, DC.

Environments and Contaminants: Criteria Air Pollutants

5. The percentage of children's days in category G is calculated by dividing the national children's G days by the national children's days and multiplying by 100 (to make it a percentage):

$$\begin{aligned} E3 &= \text{Percentage of children's days in category G} \\ &= [\text{National children's G days (G, Y)} / \text{National children's days (Y)}] \times 100\% \end{aligned}$$

The numbers of unmonitored days are by definition the total number of good, moderate, and unhealthy days subtracted from the number of days in the year. To calculate the percentage of children's days that are unmonitored you may either use the same method applied to the category of unmonitored days, or you may subtract the total of the percentages for good, moderate, and unhealthy children's days from 100. Both approaches will give exactly the same answer.

Race/Income

The supplementary indicators for Indicator E3 give the percentages of children's days with good, moderate, or unhealthy air quality stratified by race/ethnicity (Indicator E3a) or family income (Indicator E3b), for the year 2021 only. The calculation of Indicators E3a and E3b is similar to the trend analysis except that the county children's populations are replaced by estimated county children's populations for the given race/ethnicity or income groups in 2021. To account for the stratified nature of air quality sampling, Indicators E3a and E3b present the fraction of each racial/income group in each air quality category normalized by the total population of that racial/income group for which data are available. Normalization is performed by dividing the fraction of the sampled population that falls into each air quality category for each racial/income group by the fraction of the population for which data were available in that group.

The normalization process is intended to provide a more balanced view of the data. For instance, the underrepresentation of rural counties may influence apparent trends if the air quality and racial characteristics of these counties are statistically distinct from those of the sampled counties. However, even the normalized data must be interpreted with caution in the absence of complete data representation.

For Indicator E3a, children's populations stratified by race/ethnicity for 2021 were estimated for each county by using the unbridged-race 2021 postcensal population file for July 1, 2021 at the url <https://www.census.gov/data/datasets/time-series/demo/popest/2020s-counties-detail.html> to obtain race/ethnicity distributions. The race/ethnicity groups were summed as follows:

- White alone NH: NHWA_MALE and NHWA_FEMALE, AGEGRP = 1, 2, and 3, and adding 3/5 of 4.
- Black or African American alone NH: NHBA_MALE and NHBA_FEMALE, AGEGRP = 1, 2, and 3, and adding 3/5 of 4.
- Asian alone NH: NHAA_MALE and NHAA_FEMALE, AGEGRP = 1, 2, and 3, and adding 3/5 of 4.

Environments and Contaminants: Criteria Air Pollutants

- Native Hawaiian and Other Pacific Islander alone NH: NHNA_MALE and NHNA_FEMALE, AGEGRP = 1, 2, and 3, and adding 3/5 of 4.
- American Indian and Alaska Native alone NH: NHIA_MALE and NHIA_FEMALE, AGEGRP = 1, 2, and 3, and adding 3/5 of 4.
- Two or more races NH: NHTOM_MALE and NHTOM_FEMALE, AGEGRP = 1, 2, and 3, and adding 3/5 of 4.
- Hispanic: H_MALE and H_FEMALE, AGEGRP = 1, 2, and 3, and adding 3/5 of 4.
- All: sum of populations above.

For Indicator E3b, children’s populations stratified by income for 2021 were obtained using data from the Census Bureau’s Small Area Poverty and Income Estimates (SAIPE) program combined with the populations from the bridged-race Vintage 2020 postcensal population file for July 1, 2020 and unbridged-race 2021 post-censal population file for July 1, 2021. The SAIPE population estimates were developed by the Census Bureau using the American Community Survey (ACS) and Census 2010 data, based on statistical models. The SAIPE data were obtained from the URL

<https://www.census.gov/data/datasets/2021/demo/saipe/2021-state-and-county.html>

as the Excel file est21ALL.xls. The Excel file includes the state and county FIPS codes and the poverty percentage under age 18. For each county, the total children’s population was calculated by summing populations across all children ages 0 to 17 years. The children’s population below the poverty level was obtained by multiplying the total children’s population by the poverty percentage under age 18. The children’s population at or above the poverty level was obtained by multiplying the total children’s population by 100 minus the poverty percentage under age 18:

Population below poverty = Total population × poverty percentage under age 18/100,

Population at or above poverty =
Total population × (100 – poverty percentage under age 18)/100.

Questions and Comments

Questions regarding these methods, and suggestions to improve the description of the methods, are welcome. Please use the “Contact Us” link at the bottom of any page in the America’s Children and the Environment website.

Statistical Comparisons

Statistical analyses of the percentages of days for children ages 0 to 17 years with good, moderate, or unhealthy air quality were used to determine whether the trends in the percentages for different years were statistically significant. Using a logistic regression model, the logarithm

Environments and Contaminants: Criteria Air Pollutants

of the odds that a children's day has good, moderate, or unhealthy air quality is regressed against the calendar year. To avoid potential statistical issues associated with very low or very high percentages, each percentage was first rescaled to be between 5% and 95% before computing the odds. The error terms for this logistic regression were assumed to be approximately independent and normally distributed. For each air quality index category, the slope of the regression line for the logarithm of the odds was computed, together with a 95% confidence interval and its p-value. The slope estimates the annual change in the logarithm of the odds. A p-value at or below 0.05 implies that the trend is statistically significant at the 5% significance level. No adjustment is made for multiple comparisons.

The results are presented in Table 4. For more details on these statistical analyses, see the memorandum by Cohen (2010).^{viii}

Table 4. Logistic regression trend test for the proportion of days for children ages 0 to 17 years with good, moderate, or unhealthy air quality in years 1999 to 2021.

Pollution Level	N	Trend (annual change in log odds)	95% Confidence Interval for Trend: Lower Bound	95% Confidence Interval for Trend: Upper Bound	P-value for Trend
Good	23	0.0312	0.0277	0.0348	< 0.001
Moderate	23	-0.0139	-0.0203	-0.0076	< 0.001
Unhealthy	23	-0.0284	-0.0350	-0.0219	< 0.001
No Monitoring Data	23	-0.0122	-0.0145	-0.0099	< 0.001

^{viii} Statistical methods for testing for trends and year-to-year changes to air quality measures. Memorandum from Jonathan Cohen, ICF, to Dan Axelrad, EPA, November 2010.