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, 1999-2016.

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, 1999-2016.

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; PM_{10}, 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. Monitoring data are submitted by state and local environmental agencies to the national EPA Air Quality System 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. The supplementary tables for Indicator E1 provide 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 (Table E1a) or family income (Table E1b), for the year 2016. Indicator E2 is the total number of children ages 0 to 17 years living in counties with a given number of occurrences of

[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.

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, NO_2, ozone, SO_2, $PM_{2.5}$ (24-hour standard only) and PM_{10}. 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 air quality 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. Although this file has county exceedance data for other criteria pollutants, those data were not used. 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

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

obtained from the URL: http://epa.gov/airtrends/values_previous.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 2013 to 2015. Year 2013from 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 to 2016. Years 2014 to 2016 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.

- 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.

  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 2010-2016, we obtained this information from the bridged race Vintage 2016 postcensal population file:

National Center for Health Statistics. Vintage 2016 postcensal estimates of the resident population of the United States (April 1, 2010, July 1, 2010-July 1, 2016), 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 June 26, 2017, following release by the U.S. Census Bureau of the unbridged Vintage 2016 postcensal estimates by 5-year age group on June 22, 2017.

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/.

[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.

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
PM_{10}	24-hour[e]	150 µg/m^3	155 µg/m^3
$PM_{2.5}$	24-hour[f] Annual[g]	35 µg/m^3 12 µg/m^3	35.5 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/m^3 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 PM_{10} 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/m^3 to 35 µg/m^3 in 2006 (71 FR 61144, October 17, 2006).

[g]The level of the annual standard for $PM_{2.5}$ was revised from 15 µg/m^3 to 12 µg/m^3 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 2016 were obtained from the EPA website https://aqs.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 Includced" (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 Inclucded"). 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:

- 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
- NO_2 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
- PM_{10} 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
- SO_2 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 2016 were obtained directly from EPA in October and November 2017 (data for the years 1999 to 2010 were previously obtained 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

[v]David Mintz, EPA OAQPS, 919- 541-5224. Mintz.David@epamail.epa.gov

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 µg/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 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 years 2014 to 2016, 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.

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 µg/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.

[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.

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 the indicator E2, we used the same Annual Summary files to obtain the numbers of days that the ozone or $PM_{2.5}$ 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 $PM_{2.5}$ 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 $PM_{2.5}$ 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 1999-2016.

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 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 2016, we used the bridged race Vintage 2016 postcensal population file:

> National Center for Health Statistics. Vintage 2016 postcensal estimates of the resident population of the United States (April 1, 2010, July 1, 2010-July 1, 2016), 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 June 26, 2017, following release by the U.S. Census Bureau of the unbridged Vintage 2016 postcensal estimates by 5-year age group on June 22, 2017.

The bridged race 2010 to 2016 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, and July 1, 2016 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, and pop2016. 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 2016 were obtained by summing the variables pop2010, pop2011, pop2012, pop2013, pop2014, pop2015, and pop2016 over all ages <= 17, all values of "racesex," and all values of "hisp." The county children's populations for each year 2010-2016 and each race/ethnicity group were obtained by summing the variable popXXXX over all ages <= 17 for the applicable values of "racesex" and "hisp."

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

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.,

Children Affected (year Y) = Σ [Pop (county C, year Y) $\times$ Exceed (county C, year Y)]

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

Children (year Y) = Σ 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.

Percentage Children Affected = [Children Affected (year Y) / 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.

ExceedRange (county C, year Y) = 1, if county C has K exceedances of the NAAQS in year Y and $a \leq K \leq b$,
= 0, otherwise

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.,

Children Affected (year Y) = Σ [Pop (county C, year Y) × ExceedRange (county C, year Y)]

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

Children (year Y) = Σ Pop (county C, year Y)

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

The supplementary tables 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 (Table E1a) or family income (Table E1b), for the year 2016 only. The calculation of Tables E1a and E1b is exactly the same as for 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 2016.

For the supplementary table E1a, children's populations stratified by race/ethnicity for 2016 were estimated for each county by using the bridged-race Vintage 2016 post-censal population file for July 1, 2016 at the url http://www.cdc.gov/nchs/nvss/bridged_race.htm to obtain race/ethnicity distributions. The race/ethnicity groups were summed as follows:

- White non-Hispanic: racesex = 1 and 2, hisp = 1, age <= 17
- Black non-Hispanic: racesex = 3 and 4, hisp = 1, age <= 17
- AIAN non-Hispanic: racesex = 5 and 6, hisp = 1, age <= 17
- API non-Hispanic: racesex = 7 and 8, hisp = 1, age <= 17
- Hispanic: racesex = 1, 2, … 8, hisp = 2, age <= 17
- All: racesex = 1, 2, … 8, hisp = 1 and 2, age <= 17

For the supplementary table E1b, children's populations stratified by income for 2016 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 2016 post-censal population file for July 1, 2016. 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/2016/demo/saipe/2016-state-and-county.html

as the Excel file est16ALL.xls (downloaded in December, 2017). 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 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]

[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.

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 2016.

Pollutant	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)	18	-0.027	-0.040	-0.014	< 0.001
PM_{10} (24-hour)	18	-0.019	-0.036	-0.002	0.031
$PM_{2.5}$ (24-hour)	18	-0.106	-0.129	-0.082	< 0.001
$PM_{2.5}$ (annual)	18	-0.149	-0.180	-0.119	< 0.001
Carbon monoxide	18	-0.034	-0.051	-0.018	< 0.001
Lead	18	-0.003	-0.029	0.023	0.809
Sulfur dioxide (one-hour)	18	-0.106	-0.115	-0.097	< 0.001
Nitrogen dioxide (one-hour)	18	-0.081	-0.099	-0.063	< 0.001
Any standard	18	-0.045	-0.058	-0.032	< 0.001

Methods

Indicator

E3. Percentage of days with good, moderate, or unhealthy air quality for children ages 0 to 17 years, 1999-2015.

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 tables for indicator E3 provide the percentages of days for children ages 0 to 17 years with good, moderate, or unhealthy air quality, stratified by race/ethnicity (Table E3a) or income (Table E3b), for the year 2015.

Overview of Data Files

The following files are needed to calculate this indicator:

- Daily monitoring and AQI data. This file contains the state and county FIPS codes, site id, POC, date, parameter code, daily maximum, event type, and the corresponding AQI for each parameter code.
- 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.

  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 2010-2015, we obtained this information from the bridged race Vintage 2015 postcensal population file:

> National Center for Health Statistics. Vintage 2015 postcensal estimates of the resident population of the United States (April 1, 2010, July 1, 2010-July 1, 2015), 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 June 24, 2015, following release by the U.S. Census Bureau of the unbridged Vintage 2015 postcensal estimates by 5-year age group on June 23, 2015.

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

Air Quality Data

Air Quality Index (AQI) and daily maximum monitoring data from the EPA Air Quality System (AQS) for the years 1999 to 2015 and for 8-hour ozone, 24-hour PM_{10}, 24-hour $PM_{2.5}$, 8-hour CO, 1-hour SO_2, and 1-hour NO_2 were downloaded in March 2017 from the daily summary files at the EPA Air Data website:

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

AQI monitoring data for the years 1999-2015 and for 1-hour ozone and 24-hour SO_2 were obtained directly from OAQPS staff.[viii] These daily summary files give the air quality index for each monitor, day, and pollutant standard for which the AQI is defined. The Event Type variable for a criteria air pollutant flags any days deemed to have exceptional events (e.g., forest fires). For these analyses, any exceptional event days were included because we excluded data with the Event Type "Excluded." For $PM_{2.5}$, if a monitoring site had AQI values for multiple POCs on the same date, then the AQI for the lowest POC was used. For the 8-hour ozone standard, all the AQI values were recomputed using the daily maximum 8-hour average ozone values, as described below. The AQI for a given county on a given day is obtained by determining the maximum of all available AQI values for that day across all monitors in the county and across all criteria air pollutants in the index. The following air quality index categories are assigned based

[viii] David Mintz, EPA OAQPS, 919- 541-5224. Mintz.David@epamail.epa.gov

on the breakpoints shown in Table 3: Good, Moderate, and Unhealthy (includes Unhealthy for Sensitive Groups, Unhealthy, Very Unhealthy, and Hazardous).

Table 3 shows the calculation of the air quality index for each day and criteria air pollutant, based on the measured concentrations on that day. The overall AQI category for a given monitor and day is defined as the highest AQI category among all the pollutants measured on that day, if any. The overall AQI category for a given county and day is defined as the highest AQI category among all the monitors in the county with air quality measurements on that day, if any.

Table 3. Breakpoints for the AQI

This Breakpoint...							...equals this AQI	…and this category
O_3 (ppm) 8-hour Code 44201	O_3 (ppm) 1-hour[a] Code 44201	PM_{10} ($\mu g/m^3$) 24-hour Code 81102	$PM_{2.5}$ ($\mu g/m^3$) 24-hour Code 88101	CO (ppm) 8-hour Code 42101	SO_2 (ppb) 1-hour Code 42401	NO_2 (ppb) 1-hour Code 42602	AQI	
0.000 – 0.054	-	0 – 54	0.0 – 12.0	0.0 – 4.4	0 – 35	0 – 53	0 – 50	Good
0.055 – 0.070	-	55 – 154	12.1 – 35.4	4.5 – 9.4	36 – 75	54 – 100	51 – 100	Moderate
0.071 – 0.085	0.125 – 0.164	155 – 254	35.5 – 55.4	9.5 – 12.4	76 – 185	101 – 360	101 – 150	Unhealthy for Sensitive Groups
0.086 – 0.105	0.165 – 0.204	255 – 354	55.5 – 150.4	12.5 – 15.4	186 – 304[c]	361 – 649	151 – 200	Unhealthy
0.106 – 0.200	0.205 – 0.404	355 – 424	150.5 – 250.4	15.5 – 30.4	305 – 604[c]	650 – 1249	201 – 300	Very Unhealthy
[b]	0.405 – 0.504	425 – 504	250.5 – 350.4	30.5 – 40.4	605 – 804[c]	1250 – 1659	301 – 400	Hazardous
[b]	0.505 – 0.604	505 – 604	350.5 – 500.4	40.5 – 50.4	805 – 1004[c]	1650 – 2049	401 – 500	Hazardous

Adapted from "Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI)," EPA-454/B-16-002, May 2016. Available at http://www.airnow.gov/index.cfm?action=pubs.index.

[a] Areas are required to report the AQI based on 8-hour ozone values. However, there are areas where an AQI based on 1-hour ozone values would be more protective. In these cases the index for both the 8-hour and the 1-hour ozone values may be calculated and the maximum AQI reported.
[b] 8-hour O_3 values do not define higher AQI values (≥ 301). AQI values of 301 or higher are calculated with 1-hour O_3 concentrations.
[c] 1-hour SO_2 values do not define higher AQI values (≥ 200). AQI values of 200 or higher are calculated with 24-hour SO_2 concentrations.

For the 8-hour ozone standard, many of the AQI values in the daily files downloaded in March 2017 were erroneously calculated using the previous ozone standard. The daily AQI values were

recomputed from the daily maximum values using the breakpoints given in Table 3. For each category in Table 3, let BP_{Lo} and BP_{Hi} denote the lower and upper breakpoints for 8-hour ozone given in the first column of Table 3, and let I_{Lo} and I_{Hi} denote the lower and upper AQI values in the AQI column. Let C denote the daily maximum 8-hour average ozone concentration truncated to 3 decimal places. For each C, we found the category so that BP_{Hi} is greater than or equal to C and BP_{Lo} is less than or equal to C. We used the following equation to compute the AQI for the 8-hour ozone standard:

$$AQI = \frac{I_{Hi} - I_{Lo}}{BP_{Hi} - BP_{Lo}}(C - BP_{Lo}) + I_{Lo}.$$

The AQI is rounded to the nearest integer.[ix]

For the calculations for indicator E3, the following overall AQI categories were combined into the category "Unhealthy": Unhealthy for Sensitive Groups, Unhealthy, Very Unhealthy, and Hazardous.

For each county and year, we counted 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.

Census Data

For the trend analyses we obtained children's populations by county for each year from 1999 to 2015.

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 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:

[ix] See Equation 1 in "Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI)," EPA-454/B-16-002, May 2016. Available at http://www.airnow.gov/index.cfm?action=pubs.index

> 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 2010-2015, we obtained this information from the bridged race Vintage 2015 postcensal population file:

> National Center for Health Statistics. Vintage 2015 postcensal estimates of the resident population of the United States (April 1, 2010, July 1, 2010-July 1, 2015), 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 June 24, 2015, following release by the U.S. Census Bureau of the unbridged Vintage 2015 postcensal estimates by 5-year age group on June 23, 2015.

The bridged race 2010 to 2015 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, and July 1, 2015 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, and pop2015. 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, 2011, 2012. 2013, 2014, and 2015 were obtained by summing the variables pop2010, pop2011, pop2012, pop2013, pop2014, and pop2015 over all ages <= 17, all values of "racesex," and all values of "hisp."

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.

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):

E3 = Percentage of children's days in category G

= [National children's G days (G, Y) / National children's days (Y)] × 100%

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 tables for Indicator E3 give the percentages of children's days with good, moderate, or unhealthy air quality stratified by race/ethnicity (Table E3a) or family income (Table E3b), for the year 2015 only. The calculation of Tables E3a and E3b is exactly the same as for 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 2015.

For the supplementary table E3a, children's populations stratified by race/ethnicity for 2015 were estimated for each county by using the bridged-race Vintage 2015 postcensal population file for July 1. 2015 at the URL http://www.cdc.gov/nchs/nvss/bridged_race.htm to obtain race/ethnicity distributions. The race/ethnicity groups were summed as follows:

- White non-Hispanic: racesex = 1 and 2, hisp = 1, age <= 17
- Black non-Hispanic: racesex = 3 and 4, hisp = 1, age <= 17
- AIAN non-Hispanic: racesex = 5 and 6, hisp = 1, age <= 17
- API non-Hispanic: racesex = 7 and 8, hisp = 1, age <= 17
- Hispanic: racesex = 1, 2, … 8, hisp = 2, age <= 17
- All: racesex = 1, 2, … 8, hisp = 1 and 2, age <= 17

For the supplementary table E3b, children's populations stratified by income for 2015 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 2015 postcensal population file for July 1, 2015. 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

http://www.census.gov/did/www/saipe/data/statecounty/data/2015.html

as the Excel file est15ALL.xls (downloaded in December, 2016). 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 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).[x]

[x] 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.

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 2015.

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	17	0.0388	0.0339	0.0437	< 0.001
Moderate	17	-0.0221	-0.0269	-0.0173	< 0.001
Unhealthy	17	-0.0444	-0.0524	-0.0363	< 0.001
No Monitoring Data	17	-0.0051	-0.0078	-0.0023	0.001