Statistical Methodology: The Air Quality Data

Data Sources

The data on air quality throughout the United States were obtained from the U.S. Environmental Protection Agency's Air Quality System (AQS), formerly called Aerometric Information Retrieval System (AIRS) database. The American Lung Association contracted with Dr. Allen S. Lefohn, A.S.L. & Associates, Helena, Montana, to characterize the hourly averaged ozone concentration information and the 24-hour averaged $PM_{2.5}$ concentration information for the three-year period for 2013-2015 for each monitoring site.

Design values for the annual $PM_{2.5}$ concentrations by county for the period 2013-2015 came from data posted on July 29, 2016, at the U.S. Environmental Protection Agency's website at https://www.epa.gov/sites/production/files/2016-07/pm25_designvalues_20132015_final_07_29_16.xlsx.

Ozone Data Analysis

The 2013, 2014 and 2015 AQS hourly ozone data were used to calculate the daily 8-hour maximum concentration for each ozone-monitoring site. The hourly averaged ozone data were downloaded on August 11, 2016, following the close of the authorized period for quality review and assurance certification of data. Only the hourly average ozone concentrations derived from FRM and FEM monitors were used in the analysis. The data were considered for a three-year period for the same reason that the EPA uses three years of data to determine compliance with the ozone standard: to prevent a situation in any single year, where anomalies of weather or other factors create air pollution levels, which inaccurately reflect the normal conditions. The highest 8-hour daily maximum concentration in each county for 2013, 2014 and 2015, based on the EPA-defined ozone season, was identified.

The current national ambient air quality standard for ozone is 70 parts per billion (ppb) measured over eight hours. The EPA's Air Quality Index reflects the 70 ppb standard. A.S.L. & Associates prepared a table by county that summarized, for each of the three years, the number of days the ozone level was within the ranges identified by the EPA based on the EPA Air Quality Index:

8-hour Ozone Concentration	Air Quality Index Levels
0 – 54 ppb	Good (Green)
55 – 70 ppb	Moderate (Yellow)
71 – 85 ppb	Unhealthy for Sensitive Groups (Orange)
86 – 105 ppb	Unhealthy (Red)
106 – 200 ppb	Very Unhealthy (Purple)
>201 ppb	Hazardous (Maroon)

The goal of this report was to identify the number of days that 8-hour daily maximum concentrations in each county occurred within the defined ranges. This approach provided an indication of the level of pollution for all monitored days, not just those days that fell under the requirements for attaining the national ambient air quality standards. Therefore, no data capture criteria were applied to eliminate monitoring sites or to require a number of valid days for the ozone season.

The daily maximum 8-hour average concentration for a given day is derived from the highest of the 17 consecutive 8-hour averages beginning with the 8-hour period from 7:00 a.m. to 3:00 p.m. and ending with the 8-hour period from 11:00 p.m. to 7:00 a.m. the following day. This follows the process EPA uses for the current ozone standard adopted in 2015, but differs from the form used under the previous 0.075 ppm 8-hour

average ozone standard that was established in 2008. All valid days of data within the ozone season were used in the analysis. However, for computing an 8-hour average, at least 75 percent of the hourly concentrations (i.e., 6-8 hours) had to be available for the 8-hour period. In addition, an 8-hour daily maximum average was identified if valid 8-hour averages were available for at least 75 percent of possible hours in the day (i.e., at least 13 of the possible 17 8-hour averages). Because the EPA includes days with inadequate data (i.e., not 75 percent complete) if the standard value is exceeded, our data capture methodology also included the site's 8-hour value if at least one valid 8-hr period were available and it was 71 ppb or higher.

As instructed by the Lung Association, A.S.L. & Associates included the exceptional and natural events that were identified in the database and identified for the Lung Association the dates and monitoring sites that experienced such events. Some data have been flagged by the state or local air pollution control agency to indicate that they had raised issues with EPA about those data. For each day across all sites within a specific county, the highest daily maximum 8-hour average ozone concentration was recorded and then the results were summarized by county for the number of days the ozone levels were within the ranges identified above.

Following receipt of the above information, the American Lung Association identified the number of days each county, with at least one ozone monitor, experienced air quality designated as orange (Unhealthy for Sensitive Groups), red (Unhealthy) or purple (Very Unhealthy).

Short-Term Particle Pollution Data Analysis

A.S.L. & Associates identified the maximum daily 24-hour AQS $PM_{2.5}$ concentration for each county in 2013, 2014 and 2015 with monitoring information. The 24-hour $PM_{2.5}$ data were downloaded on August 10, 2016, following the close of the authorized period for quality review and assurance certification of data. In addition, hourly averaged $PM_{2.5}$ concentration data were characterized into 24-hour average $PM_{2.5}$ values by the EPA and provided to A.S.L. & Associates. Using these results, A.S.L. & Associates prepared a table by county that summarized, for each of the three years, the number of days the maximum of the daily $PM_{2.5}$ concentration was within the ranges identified by the EPA based on the EPA Air Quality Index, as adopted by the EPA on December 14, 2012:

24-hour $PM_{2.5}$ Concentration	Air Quality Index Levels
0.0 mg/m^3 to 12.0 mg/m^3	Good (Green)
12.1 mg/m^3 to 35.4 mg/m^3	Moderate (Yellow)
35.5 mg/m^3 to 55.4 mg/m^3	Unhealthy for Sensitive Groups (Orange)
55.5 mg/m^3 to 150.4 mg/m^3	Unhealthy (Red)
150.5 mg/m^3 to 250.4 mg/m^3	Very Unhealthy (Purple)
greater than or equal to 250.5 mg/m^3	Hazardous (Maroon)

All previous data collected for 24-hour average $PM_{2.5}$ were characterized using the AQI thresholds listed above.

The goal of this report was to identify the number of days that the maximum in each county of the daily $PM_{2.5}$ concentration occurred within the defined ranges. This approach provided an indication of the level of pollution for all monitored days, not just those days that fell under the requirements for attaining the national ambient air quality standards. Therefore, no data capture criteria were used to eliminate monitoring sites. Both 24-hour averaged PM data, as well as hourly averaged PM data averaged over 24 hours were used. Included in the analysis are data collected using only FRM

and FEM methods, which reported hourly and 24-hour averaged data. As instructed by the Lung Association, A.S.L. & Associates included the exceptional and natural events that were identified in the database and identified for the Lung Association the dates and monitoring sites that experienced such events. Some data have been flagged by the state or local air pollution control agency to indicate that they had raised issues with EPA about those data. For each day across all sites within a specific county, the highest daily maximum 24-h $PM_{2.5}$ concentration was recorded and then the results were summarized by county for the number of days the concentration levels were within the ranges identified on the chart on the preceding page.

Following receipt of the above information, the American Lung Association identified the number of days each county, with at least one $PM_{2.5}$ monitor, experienced air quality designated as orange (Unhealthy for Sensitive Groups), red (Unhealthy), purple (Very Unhealthy) or maroon (Hazardous).

Description of County Grading System

Ozone and Short-Term Particle Pollution (24-hour $PM_{2.5}$)

The grades for ozone and short-term particle pollution (24-hour $PM_{2.5}$) were based on a weighted average for each county. To determine the weighted average, the Lung Association followed these steps:

1. First, assigned weighting factors to each category of the Air Quality Index. The number of orange days experienced by each county received a factor of 1; red days, a factor of 1.5; purple days, a factor of 2; and maroon days, a factor of 2.5. This allowed days where the air pollution levels were higher to receive greater weight.
2. Next, multiplied the total number of days within each category by their assigned factor, and then summed all the categories to calculate a total.
3. Finally, divided the total by three to determine the weighted average, since the monitoring data were collected over a three-year period.

The weighted average determined each county's grades for ozone and 24-hour $PM_{2.5}$.

- All counties with a weighted average of zero (corresponding to no exceedances of the standard over the three-year period) were given a grade of "A."
- For ozone, an "F" grade was set to generally correlate with the number of unhealthy air days that would place a county in nonattainment for the ozone standard.
- For short-term particle pollution, fewer unhealthy air days are required for an F than for nonattainment under the $PM_{2.5}$ standard. The national air quality standard is set to allow 2 percent of the days during the three years to exceed 35 $\mu g/m^3$ (called a "98th percentile" form) before violating the standard. That would be roughly 21 unhealthy days in three years. The grading used in this report would allow only about 1 percent of the days to be over 35 $\mu g/m^3$ (called a "99th percentile" form) of the $PM_{2.5}$. The American Lung Association supports using the tighter limits in a 99th percentile form as a more appropriate standard that is intended to protect the public from short-term spikes in pollution.

Grading System		
Grade	**Weighted Average**	**Approximate Number of Allowable Orange/Red/Purple/Maroon days**
A	0.0	None
B	0.3 to 0.9	1 to 2 orange days with no red
C	1.0 to 2.0	3 to 6 days over the standard: 3 to 5 orange with no more than 1 red OR 6 orange with no red
D	2.1 to 3.2	7 to 9 days over the standard: 7 total (including up to 2 red) to 9 orange with no red
F	3.3 or higher	9 days or more over the standard: 10 orange days or 9 total including at least 1 or more red, purple or maroon

Weighted averages allow comparisons to be drawn based on severity of air pollution. For example, if one county had nine orange days and no red days, it would earn a weighted average of 3.0 and a D grade. However, another county that had only eight orange days but also two red days, which signify days with more serious air pollution, would receive a F. That second county would have a weighted average of 3.7.

Note that this system differs significantly from the methodology the EPA uses to determine violations of both the ozone and the 24-hour $PM_{2.5}$ standards. The EPA determines whether a county violates the standard based on the fourth maximum daily 8-hour ozone reading each year averaged over three years. Multiple days of unhealthy air beyond the highest four in each year are not considered. By contrast, the system used in this report recognizes when a community's air quality repeatedly results in unhealthy air throughout the three years. Consequently, some counties will receive grades of "F" in this report, showing repeated instances of unhealthy air, while still meeting the EPA's 2015 ozone standard. The American Lung Association's position is that the evidence shows that the 2015 ozone standard, although stronger than the 2008 standard, still fails to adequately protect public health.

The Lung Association calculates the population at risk from these pollutants based on the population from the entire county where the monitor is located and the largest metropolitan area that contains that county. Not only do people from that county or metropolitan area circulate within the county and the metropolitan area, the air pollution circulates to that monitor through the county and metropolitan area.

Counties were ranked by weighted average. Metropolitan areas were ranked by the highest weighted average among the counties within a given Metropolitan Statistical Area as of 2015 as defined by the White House Office of Management and Budget (OMB).

Year-Round Particle Pollution (Annual $PM_{2.5}$)

Since no comparable Air Quality Index exists for year-round particle pollution (annual $PM_{2.5}$), the grading was based on the 2012 National Ambient Air Quality Standard for annual $PM_{2.5}$ of 12 µg/m^3. Counties that EPA listed as being at or below 12 µg/m^3 were given grades of "Pass." Counties EPA listed as being at or above 12.1 µg/m^3 were given grades of "Fail." Where insufficient data existed for EPA to determine a design value, those counties received a grade of "Incomplete."

EPA officially recognized that data collected in all Illinois counties, in some Maine counties and in most Tennessee counties were processed in certain laboratories where quality control issues meant that available data could not be considered for development of an official design value. For short-term and annual particle pollution, those counties received a grade of "Incomplete."

Design value is the calculated concentration of a pollutant based on the form of the national ambient air quality standard and is used by EPA to determine whether or not the air quality in a county meets the standard. Counties were ranked by design value. Metropolitan areas were ranked by the highest design value among the counties within a given Metropolitan Statistical Area as of 2015 as defined by the OMB.

The Lung Association received critical assistance from members of the National Association of Clean Air Agencies and the Association of Air Pollution Control Agencies. With their assistance, all state and local agencies were provided the opportunity to review and comment on the data in draft tabular form. The Lung Association reviewed all discrepancies with the agencies and, if needed, with Dr. Lefohn at A.S.L. & Associates. Questions about the annual PM design values were discussed with EPA; however, the Lung Association made final decisions to grade counties as "Incomplete" where EPA considered $PM_{2.5}$ data to have inadequate quality assurance. The American Lung Association wishes to express its continued appreciation to the state and local air directors for their willingness to assist in ensuring that the characterized data used in this report are correct.

Calculations of Populations at Risk

Presently county-specific measurements of the number of persons with chronic conditions are not generally available. In order to assess the magnitude of chronic conditions at the state and county levels, we have employed a synthetic estimation technique originally developed by the U.S. Census Bureau. This method uses age-specific national and state estimates of self-reported conditions to project disease prevalence to the county level. The exception to this is poverty, for which estimates are available at the county level.

Population Estimates

The Lung Association includes the total county population in discussions of populations at risk from exposure to pollution in each county. The Lung Association uses that conservative count based on several factors: the recognized limited number and locations of monitors in most counties and metropolitan areas; the movement of the population both in daily activities, including outdoor activities, such as exercise or work; and the transport of emission from sources into and across the county to reach the monitor.

The U.S. Census Bureau estimated data on the total population of each county in the United States for 2015. The Census Bureau also estimated the age-specific breakdown of the population and how many individuals were living in poverty by county. These estimates are the best information on population demographics available between decennial censuses.

Poverty estimates came from the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program. The program does not use direct counts or estimates from sample surveys, as these methods would not provide sufficient data for all counties. Instead, a model based on estimates of income or poverty from the Annual Social and Economic Supplement (ASEC) to the Current Population Survey (CPS) is used to develop estimates for all states and counties.

Prevalence Estimates

Chronic Obstructive Pulmonary Disease, Cardiovascular Disease, Asthma and Diabetes. In 2015, the Behavioral Risk Factor Surveillance System (BRFSS) survey found that approximately 21.6 million (8.9 percent) of adults residing in the United States and 8.5 percent of children from 30 states and Washington, D.C., reported currently having asthma. Among adults in the Unites States in 2015, 15.5 million (6.3 percent) had ever been diagnosed with chronic obstructive pulmonary disease (COPD), 20.4 million (8.4

percent) had ever been diagnosed with cardiovascular disease and 25.6 million (10.4 percent) had ever been diagnosed with diabetes.

The prevalence estimate for pediatric asthma is calculated for those younger than 18 years. Local area prevalence of pediatric asthma is estimated by applying 2015 state prevalence rates, or if not available, the national rate from the BRFSS to pediatric county-level resident populations obtained from the U.S. Census Bureau website. Pediatric asthma data from the 2015 BRFSS were available for thirty states and Washington D.C., from the 2014 BRFSS for seven states, from the 2013 BRFSS for one state, from the 2012 BRFSS for two states, from the 2011 BRFSS for one state, and national data were used for the nine states[1] that had no data available. Data from earlier years were not used due to changes in the 2011 survey methodology.

The prevalence estimate for COPD, cardiovascular disease, adult asthma and diabetes is calculated for those aged 18-44 years, 45-64 years and 65 years and older. Local area prevalence for these diseases is estimated by applying age-specific state prevalence rates from the 2015 BRFSS to age-specific county-level resident populations obtained from the U.S. Census Bureau website. Cardiovascular disease included ever having been diagnosed with a heart attack, angina or coronary heart disease, or stroke.

Incidence Estimates

Lung Cancer. State- and gender-specific lung cancer incidence rates for 2013 were obtained from StateCancerProfiles.gov, a system that provides access to statistics from both the National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) program and the Center for Disease Control and Prevention's National Program of Cancer Registries.

Local area incidence of lung cancer is estimated by applying 2013 age-adjusted and sex-specific incidence rates to 2015 county populations obtained from the U.S. Census Bureau. Thereafter, the incidence estimates for each county within a state are summed to determine overall incidence. Estimates for Nevada are based on 2010 rates.

Limitations of Estimates. Since the statistics presented by the BRFSS and SAIPE are based on a sample, they will differ (due to random sampling variability) from figures that would be derived from a complete census or case registry of people in the U.S. with these diseases. The results are also subject to reporting, non-response and processing errors. These types of errors are kept to a minimum by methods built into the survey.

Additionally, a major limitation of the BRFSS is that the information collected represents self-reports of medically diagnosed conditions, which may underestimate disease prevalence since not all individuals with these conditions have been properly diagnosed. However, the BRFSS is the best available source for information on the magnitude of chronic disease at the state level. The conditions covered in the survey may vary considerably in the accuracy and completeness with which they are reported.

Local estimates of chronic diseases are scaled in direct proportion to the base population of the county and its age distribution. No adjustments are made for other factors that may affect local prevalence (e.g., local prevalence of cigarette smokers or occupational exposures) since the health surveys that obtain such data are rarely conducted on the county level. Because the estimates do not account for geographic differences in the prevalence of chronic and acute diseases, the sum of the estimates for each of the counties in the United States may not exactly reflect the national or state estimates derived from the BRFSS.

1 2014: Alabama, Kentucky, Maryland, North Carolina, Tennessee, Washington, West Virginia. 2013: Arizona. 2012: North Dakota and Wyoming. 2011: Iowa. National: Alaska, Arkansas, Colorado, Delaware, Florida, Idaho, South Carolina, South Dakota and Virginia.

References

Irwin, R. Guide to Local Area Populations. U.S. Bureau of the Census, Technical Paper Number 39 (1972).

Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System, 2015.

StateCancerProfile.gov, 2016. Cancer Incidence by State and Gender, 2013.

Population Estimates Branch, U.S. Census Bureau. Annual Estimates of the Resident Population by Selected Age Groups and Sex for Counties: April 1, 2010 to July 1, 2015.

Office of Management and Budget. Revised Delineations of Metropolitan Statistical Areas, Micropolitan Statistical Areas, and Combined Statistical Areas, and Guidance on Uses of the Delineations of These Areas. OMB Bulletin 15-01 July 15, 2015.

U.S. Census Bureau. Small Area Income and Poverty Estimates. State and County Data, 2015.