



MEMORANDUM

To: Dan Axelrad, EPA
From: Jonathan Cohen, ICF
Date: 5 November, 2010
Re: Statistical methods for testing for trends and year-to-year changes in air quality measures.

Introduction and Summary

America's Children and the Environment (ACE) brings together, in one place, quantitative information from a variety of sources to show trends in levels of environmental contaminants in air, water, food, and soil; concentrations of contaminants measured in the bodies of mothers and children; and childhood diseases that may be influenced by environmental factors. The ACE results have been published in two printed reports and also on a website <http://www.epa.gov/envirohealth/children/> that is updated annually. The 2003 edition of ACE and the current ACE website includes indicators for air quality, using air quality data from the EPA Air Quality System (AQS) and population data from the US Census Bureau. For the forthcoming third edition of ACE (ACE3), EPA plans to develop additional or revised indicators showing trends in levels of environmental contaminants. The current presentation of indicators in ACE includes providing time series data, but does not include evaluation of whether the trends and year-to-year differences are statistically significant. A goal for ACE3 is to add this type of statistical evaluation to the presentation of the indicators. The goal of the statistical methods presented in this memorandum is to be able to determine which of the observed trends and year-to-year differences are large enough to be highlighted in discussions about the findings rather than being attributed to random variation. This memorandum focuses on those indicators that are based on trends in annual summary statistics that are not derived from survey data. Two other memoranda present statistical methods used for the analyses of health-related indicators that are derived from complex survey data, such as from the National Health and Nutrition Examination Survey (NHANES) and the National Health Interview Survey (NHIS).

This memorandum presents methods we intend to apply to statistical analysis of annual summary statistics, such as the air quality indicators, to address these two questions:

1. Is there a trend in the indicator value over time?
2. Is there a statistically significant change in the indicator value for a given year compared with the previous year?

The term "indicator value" refers to the annual summary statistic of interest. For air quality indicator E1, this is the percentage of children living in counties that exceed air quality standards. For air quality indicator E2, this is the percentage of children's days with good, moderate, or unhealthy air quality.

A variety of statistical methods were used to develop these indicators of children's environmental health. The detailed methods used depend upon the question of interest and the nature of the available databases. The ACE website and published reports tabulate and graph trends for most of the indicators but do not currently evaluate whether the trends and year-to-year differences are statistically significant.

This memorandum describes various statistical methods for testing for trends and year-to-year changes in the ACE air quality indicators E1 and E2, and in other ACE indicators defined as annual summary statistics. These indicators are calculated from monitoring databases. Unlike data from sample surveys, the random mechanism that generated the data is not known. This influences the types of statistical methods available. The simplest and recommended approach treats these data as a series of annual values generated from some underlying simple statistical model developed by the analyst. Therefore the same statistical approaches can be applied to any of these measures. To illustrate the methods, they were applied to air quality indicator E1, more specifically to the annual proportions of children living in counties experiencing annual average PM_{2.5} levels exceeding the national air quality standard of 15 µg/m³. Since the annual statistics of interest are proportions p that are by definition bounded between zero and one, we recommend transforming the proportions to have values between minus and plus infinity by using the logit transformation $\text{logit}(p^*) = \log \{p^*/(1 - p^*)\}$ where $p^* = 0.05 + 0.9p$. The initial transformation from p to p^* rescales the proportion to the interval from 0.05 to 0.95 and thus avoids undefined values for proportions exactly equal to zero or one, and makes the results less sensitive to very low or very high proportions. The logit transformation then transforms the rescaled proportion to be between minus and plus infinity. To test for trends, we recommend regressing the transformed proportion against the year and testing whether the slope coefficient is statistically significant, i.e., testing whether the slope is zero. This approach looks for consistent year-to-year changes. To test the year-to-year change over the last two years $y - 1$ and y , we recommend a logistic regression approach where the logit-transformed proportion is assumed to have a linear trend up to year $y - 2$, and the values for years $y - 1$ and y are unrestricted. The statistical test is whether the expected difference between years $y - 1$ and y is zero, and the test is carried out by comparing the observed change with the variability of the transformed proportions about the trend line.

For these ACE indicators, a major limitation for the statistical analyses of trends and year-to-year changes is the small sample sizes, since the number of available annual statistics is generally less than 15. Each year of data provides only a single data point for the analysis, even though that single data point is calculated from a very large data base (e.g., for air quality data, the annual statistics are based on hourly or daily monitored concentrations at all monitors nationwide and all county children's populations). It is difficult to use the underlying database to estimate the variability of these annual statistics. We considered alternative approaches based on fitting detailed statistical models to the annual statistics using the underlying raw data, but do not recommend these methods because the analyses would be quite complex and resource-intensive and would need to be redeveloped for each annual update. Therefore our recommended approach uses only the short series of annual statistics to estimate variability and test for trends or year-to-year changes. With such a small sample size, these techniques will have limited power to find significant trends (or year-to-year changes).

In this memorandum we illustrate the methods using annual statistics for years 1999 to 2008, as in the most recent PM_{2.5} data analyzed for ACE indicator E1. The same approaches apply with minor modifications for other start and end years.

Air Quality Indicator E1: PM2.5 Annual Standard

One of the summary statistics in Indicator E1 is the proportion of children living in counties experiencing annual average PM2.5 levels exceeding the national air quality standard of 15 $\mu\text{g}/\text{m}^3$. A county is assumed to exceed the standard in a given year if the annual average PM2.5 exceeds 15 $\mu\text{g}/\text{m}^3$ at one or more monitors in that county; in cases with two or more monitors at the same site location, only the lowest monitor number (POC) is used. A county is assumed not to exceed the standard in a given year if the annual average PM2.5 is at most 15 $\mu\text{g}/\text{m}^3$ at all of the monitors in the county, or if there are no monitors with valid data in that county for that year. The summary statistic is calculated by summing the county children's populations over all counties exceeding the standard and then dividing by the total US children's population. The data from 1999 to 2008 are tabulated in Table 1 and plotted as the black dots in Figures 1 and 2. Figure 1 shows the proportions plotted against the year. Figure 2 shows the logit-transformed proportions plotted against the year.

Table 1. Proportion of U.S. children living in counties experiencing annual average PM2.5 levels exceeding the national air quality standard of 15 $\mu\text{g}/\text{m}^3$. Proportions and logit-transformed proportions.

Year	Proportion children affected	Logit-Transformed Proportion*
1999	0.242	-1.008
2000	0.296	-0.770
2001	0.247	-0.982
2002	0.210	-1.161
2003	0.191	-1.253
2004	0.165	-1.395
2005	0.244	-0.995
2006	0.125	-1.641
2007	0.162	-1.413
2008	0.073	-2.033

* If p is a proportion between 0 and 1, then the logit-transformed proportion is defined as $\text{logit}(p^*) = \log \{p^*/(1 - p^*)\}$ where $p^* = 0.05 + 0.9p$. "Log" denotes the natural logarithm, base e .

Annual Statistics

For each ACE indicator for which the methods in this memorandum will be applied, there is an associated annual statistic. For PM2.5 in indicator E1, the annual statistic is the annual proportion of children living in counties experiencing annual average PM2.5 levels exceeding the national air quality standard of 15 $\mu\text{g}/\text{m}^3$. This annual statistic should be carefully distinguished from the annual average concentrations themselves which are not of direct interest for ACE (although the annual statistic is, of course, a function of the county monitor annual averages and the county populations). The regression methods to be applied assume that the annual statistic has been transformed so that its distribution is approximately normal. For indicators where the annual statistic is unbounded, no transformation may be needed and so the transformed and untransformed annual statistics are the same. For indicators such as E1 where the annual statistic is a proportion bounded between 0 and 1, we recommend transforming the annual statistic using a logit transformation (and a rescaling). The proportion p is first rescaled to a value p^* in the interval from 0.05 to 0.95. This avoids problems with calculating the logarithms of zero and reduces the sensitivity of the results to very small, but positive values. Then a logit transformation is applied to p^* giving $\text{logit}(p^*) = \log \{p^*/(1 - p^*)\}$. "Log" denotes the natural logarithm, base e . The logit transformation was chosen to make the proportion data more approximately normal. The logistic regression model for the proportions

assumes that there are equal annual changes in the logit, and thus approximately equal annual percentage changes in the corresponding odds ($p/(1-p)$).

Statistical Test for Trend

The annual statistics are assumed to have been generated from the statistical regression model:

$$\text{Transformed Annual Statistic (Year)} = \text{Intercept} + \text{Trend} \times (\text{Year} - 1999) + \text{Error},$$

for Year = 1999, 2000, ... 2008.

The transformed annual statistic for a proportion p is the logit-transformed proportion defined above. The error terms (observed statistic minus expected statistic) are independent and normally distributed with a mean of zero and an unknown variance. The Intercept and Trend are estimated using simple linear regression. If the estimated value of Trend is statistically significantly different from zero at the five percent level, then a statistically significant trend has been found. The trend is the estimated change in the transformed annual statistic between one year and the next year. Using the logit transform, 100 times the trend is approximately the annual percentage change in the odds ($p/(1-p)$).

For the PM2.5 data, the fitted trend line for the proportion affected is shown as the red dots in Figures 1 and 2. The slope of the straight line in Figure 2 is the value of Trend, -0.1010 . A 95% confidence interval for the trend is the interval from -0.1581 to -0.0440 . The p-value is 0.0035, which is less than 0.05, so the trend is statistically significant.

The results for the Trend test are tabulated in Table 2.

Table 2. Logistic regression trend test for the proportion of U.S. children living in counties experiencing annual average PM2.5 levels exceeding the national air quality standard of $15 \mu\text{g}/\text{m}^3$ in years 1999 to 2008.

Variable	N	Trend	95% Confidence Interval for Trend: Lower Bound	95% Confidence Interval for Trend: Upper Bound	P-value for Trend
Proportion of children living in counties experiencing annual PM2.5 $\geq 15 \mu\text{g}/\text{m}^3$	10	-0.1010	-0.1581	-0.0440	0.0035

Statistical Tests for Year-to-Year Change

The annual statistics are assumed to have been generated from the statistical model:

$$\text{Transformed Annual Statistic (Year)} = \text{Intercept} + \text{Trend2} \times (\text{Year} - 1999) + \text{Error},$$

for Year = 1999, 2000, ... 2006,

$$\text{Transformed Annual Statistic (2007)} = \text{Intercept} + \text{Lasttwo} + \text{Error}, \text{ and}$$

$$\text{Transformed Annual Statistic (2008)} = \text{Intercept} + \text{Lasttwo} + \text{Last} + \text{Error}.$$

The error terms (observed statistic minus expected statistic) are independent and normally distributed with a mean of zero and an unknown variance. The Intercept is assumed to have the same value for all three equations. The value of Lasttwo is assumed to have the same value for the last two equations.

The unknown regression parameters are Intercept, Trend2, Lasttwo, and Last. Under this model, there is a trend from 1999 to 2006, but the mean values in 2007 and 2008 are arbitrary (since there are the two parameters Lasttwo and Last). It can be shown that the values of Intercept, Trend2, and the error variance are all estimated from the first equation, which is a standard linear regression trend equation. Using the second equation, the estimated value of Lasttwo equals the transformed annual statistic for 2007 minus the estimated Intercept (from the first equation). Finally, because Lasttwo appears in both the second and third equations, the third equation implies that the estimated value of Last equals the difference between the transformed annual statistics for 2007 and 2008. If Last equals zero, then the means for 2007 and 2008 are equal, so there is no mean change in the transformed annual statistic from 2007 to 2008.

The parameters are estimated using linear regression of the transformed annual statistics. If the estimated value of Last is statistically significantly different from zero at the five percent level, then a statistically significant change from 2007 to 2008 has been found.

The variance is estimated by the mean square error, MSE, which is the sum of the squared differences between the observed and predicted values, divided by 6, the degrees of freedom:

$$\text{MSE} = \Sigma \{ \text{Transformed Statistic (Year)} - \text{Predicted Value (Year)} \}^2 / \{ \text{Number of years} - 4 \}.$$

Because the observed and predicted values are equal for 2007 and 2008, this is mathematically the same as the mean square error around the trend line from 1999 to 2006. Thus the variability around the trend line is used to estimate the variance of the annual statistics in 2007 and 2008, which in turn is used to test whether the change is statistically significant.

For the PM2.5 data, the fitted model for the proportion affected is shown as the red dots in Figure 3. Figure 3 shows the predicted and observed logit-transformed proportions plotted against the year using the statistical model described at the beginning of this section. The slope of the line from 1999 to 2006 is the value of Trend2, -0.0821 . The estimated value of Last is the observed change in the transformed statistic from 2007 to 2008, -0.6200 . The p-value for Last is 0.0737, which is greater than 0.05, so the change is not statistically significant. Also note that the estimated variance is 0.0411, the mean square error. Readers may be surprised from Figure 3 that the apparently large change from 2007 to 2008 was not statistically significant. However, the estimated standard deviation of each annual statistic is the square root of the mean square error, 0.2027, so that the estimated standard deviation of the change from 2007 to 2008 is 0.2867 (0.2027 multiplied by the square root of 2). Thus the observed change is only 2.2 standard errors away from zero ($0.6200/0.2867 = 2.2$). Because the initial trend line was fitted to only 10 years of data, and there are four unknown parameters, the available degrees of freedom is $10 - 4 = 6$, and 2.2 is not a statistically significant difference (at the five percent level) for a T distribution with only 6 degrees of freedom.

The results for the year-to-year change test are tabulated in Table 3.

Table 3. Logistic regression year-to-year change test for the 2007 to 2008 change in the proportion of U.S. children living in counties experiencing annual average PM2.5 levels exceeding the national air quality standard of $15 \mu\text{g}/\text{m}^3$ in years 1999 to 2008.

Variable	N	Year to Year Change	95% Confidence Interval for Change: Lower Bound	95% Confidence Interval for Change: Upper Bound	P-value for Change
Proportion of children living in counties experiencing annual PM2.5 $\geq 15 \mu\text{g}/\text{m}^3$	10	-0.6200	-1.3214	0.0813	0.0737

Figure 1. Statistical trend model for the proportion of children living in counties with PM2.5 annual mean exceedances. Linear vertical scale.

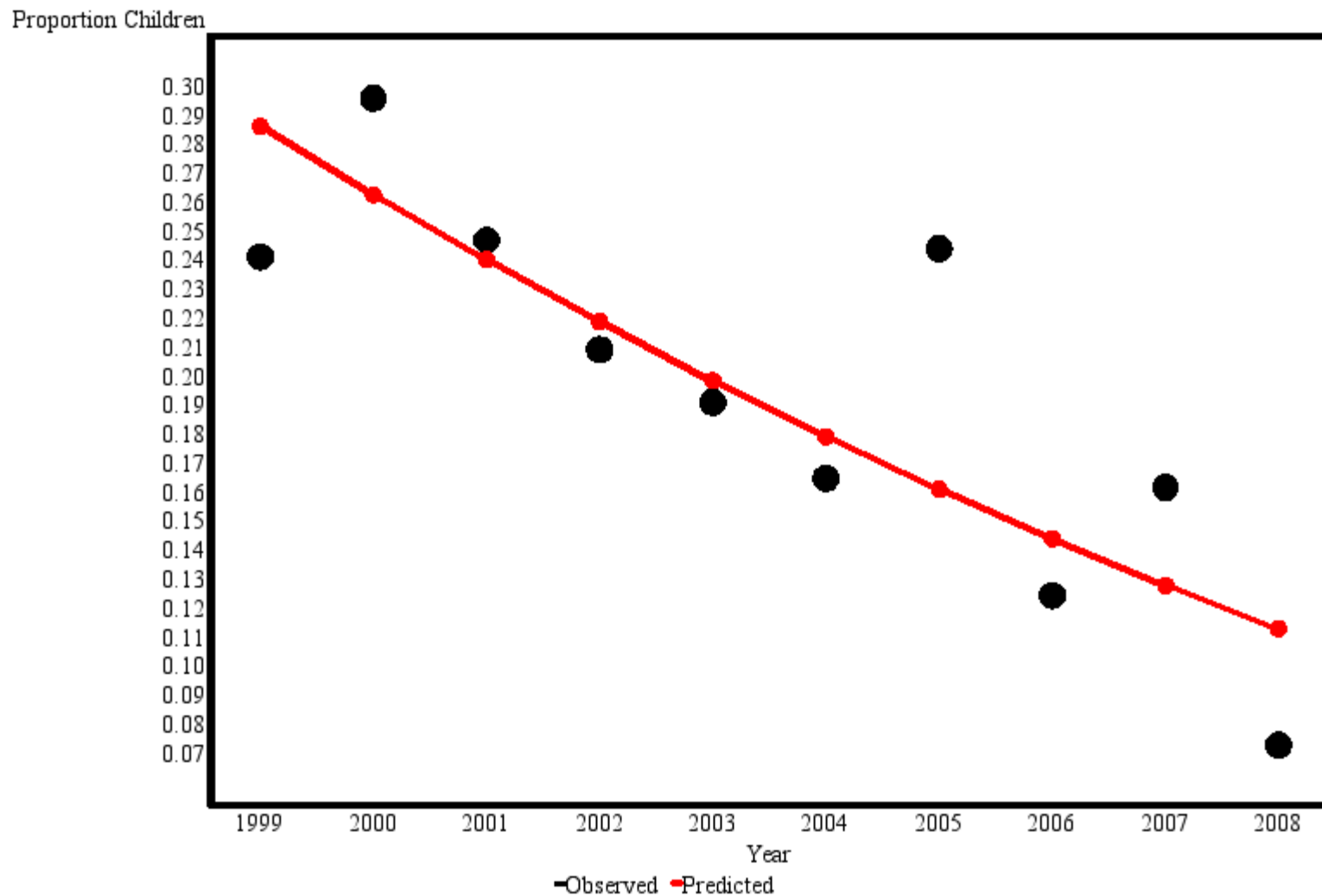


Figure 2. Statistical trend model for the proportion of children living in counties with PM2.5 annual mean exceedances. Logit-Transformed vertical scale.

Logit Proportion Children

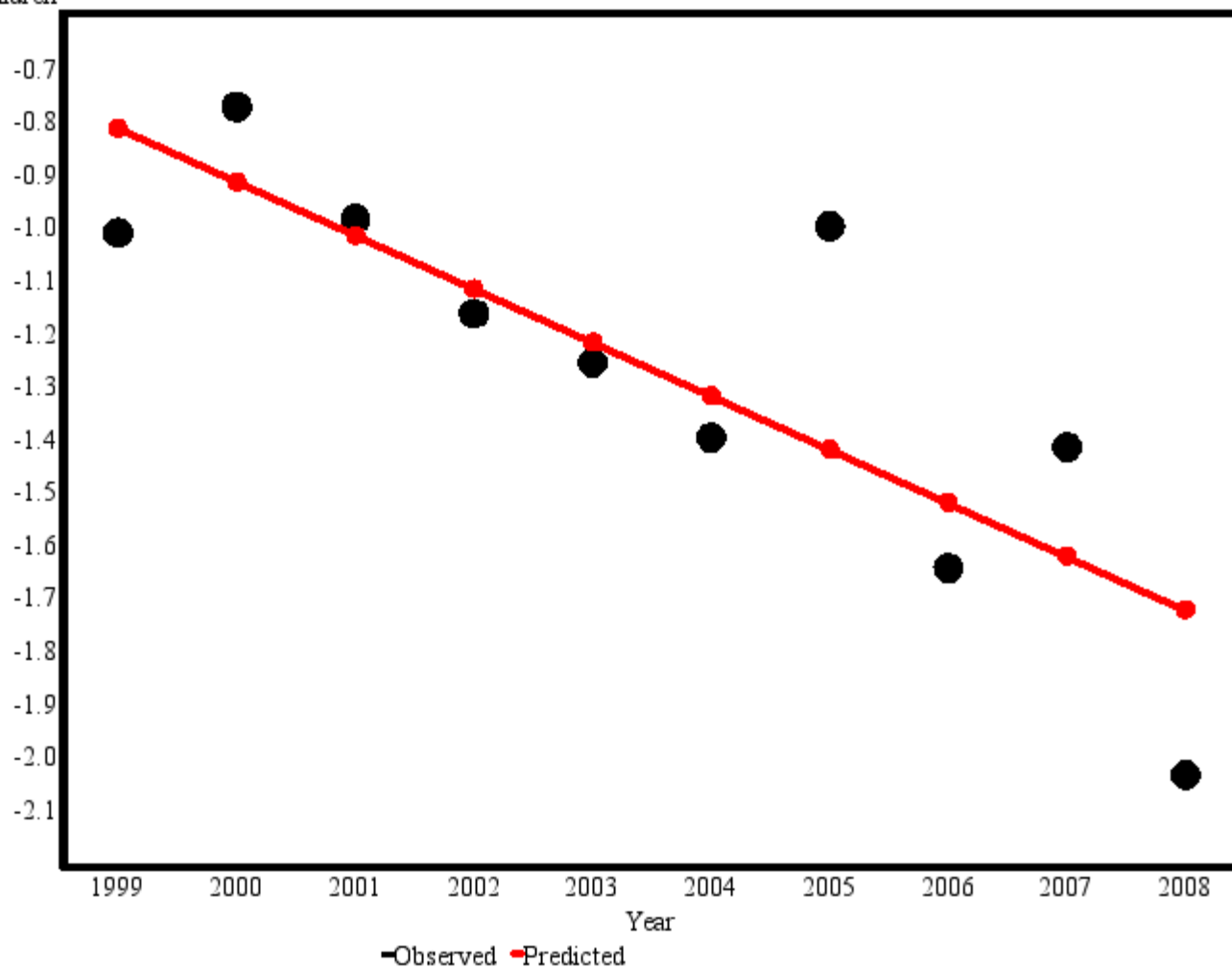


Figure 3. Statistical model for the proportion of children living in counties with PM2.5 annual mean exceedances. Logit-Transformed vertical scale. Model assumes a linear trend up to the last year minus 2 (i.e., 2006).

Logit Proportion Children

