INTERNATIONAL

MEMORANDUM

To: Dan Axelrad, EPA

From: Jonathan Cohen, ICF

Date: 3 October, 2011

Re: Selected statistical methods for testing for trends and comparing years or demographic groups in other ACE health-based indicators.

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. For the forthcoming third edition of ACE (ACE3), EPA has developed indicators for the prevalence of children's emergency room visits and hospitalizations for respiratory diseases, the prevalence of childhood obesity and cancer, and the prevalence of adverse birth outcomes such as low birthweight or preterm births. The indicators for emergency room visits and hospitalizations for respiratory diseases use data from the National Hospital Ambulatory Medical Care Survey (NHAMCS) and the National Hospital Discharge Survey (NHDS), respectively. The indicator for obese children comes from the National Health and Nutrition Examination Survey (NHANES). The indicator for childhood cancer comes from the Surveillance, Epidemiology, and End Results (SEER) national cancer data compilation. The indicator for adverse birth outcomes comes from the National Vital Statistics System (NVSS). NHAMCS, NHDS, and NHANES are all complex surveys conducted on a continuous basis by the National Center for Health Statistics in the Centers for Disease Control and Prevention, which also compiles the birth certificate data for the NVSS. SEER data are compiled by the National Cancer Institute. These surveys and data compilations provide high-quality nationally representative data over a long period, with detailed demographic information in addition to the detailed health data. Most of the data are publicly available, although some data elements are masked to protect confidentiality. The current presentation of indicators in ACE includes providing time series data and values stratified by race/ethnicity and family income, but does not include evaluation of whether the trends and 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 enable researchers and other readers of the ACE3 report to determine which of the observed trends and differences are large enough to be highlighted in discussions about the findings rather than being attributed to random variation. Ideally this determination should include cases where the difference is above a certain threshold that is large enough to be of concern, but for most indicators there is no good consensus about which differences are numerically large enough to be of concern, thus we use statistical significance as a guideline. Very large differences would usually be noted together with information about whether or not the difference was statistically significant. Very small, but statistically significant, differences would usually not be highlighted.

This memorandum presents methods that can be applied to statistical analysis of these selected health-based indicators to address these three 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 (or other time period) compared with the previous year?
3. Is there a statistically significant difference in the indicator value between different demographic groups?

The term “indicator value” refers to the rate of emergency room visits or hospitalizations for the NHAMCS and NHDS respiratory disease indicators within a defined population, or to the prevalence of a disease or condition within a defined population.

This memorandum describes the various statistical methods for testing for trends, analyzing year-to-year changes, and comparing demographic groups in the values of the health-based indicators listed above. A companion memorandum describes similar statistical analysis methods applicable for percentiles of body burdens measured in the NHANES and for disease prevalence measured in the National Health Interview Survey (NHIS). Another companion memorandum describes statistical trend analyses for annual indicators of children’s air quality.

For the indicators from NHAMCS and NHDS, weighted linear regression is used, so that the rate of visits is regressed against the covariates, inversely weighting each rate by its estimated variance. For the childhood obesity indicator from NHANES and for the NVSS birth outcomes indicator, the logit of the prevalence (i.e., the logarithm of the odds of having the disease or condition) is regressed against the covariates. For the cancer indicator from SEER, the numbers of cases (incidence or mortality) are assumed to have a Poisson distribution and the logarithm of the rate is regressed against the covariates. In all of these cases, the unadjusted analysis is compared with an analysis that adjusts for possible confounding effects by including other demographic variables in the regression. We also analyze the trends and year-to-year changes for different demographic groups, to evaluate whether the trend is different for different demographic groups.

The methods for testing for trends, year-to-year changes, and demographic group differences are illustrated for NHAMCS and NHDS using the 1996 to 2008 trends and 2005-2008 demographic group differences in emergency room visits data for asthma and other respiratory causes, and for NHANES using the 1976 to 2008 trends and 2005-2008 demographic group differences in the prevalence of overweight children. Example analyses of the SEER cancer data and NVSS adverse birth outcomes data are not provided here, since the methods used are similar to the methods described in detail for the NHAMCS and NHANES data.

National Hospital Ambulatory Medical Care Survey (NHAMCS)

The National Hospital Ambulatory Medical Care Survey (NHAMCS) is conducted by the National Center for Health Statistics, a division of the Centers for Disease Control and Prevention. The complex multi-stage survey is designed to collect data on ambulatory care services in hospital emergency and outpatient departments; these analyses only used the emergency department visits. Sampled hospitals are noninstitutional general and short-stay hospitals located in all states and Washington DC, but exclude federal, military, and Veteran’s Administration hospitals. Data from sampled visits are obtained on the demographic characteristics, expected source(s) of payments, patients’ complaints, physician’s diagnoses, diagnostic and screening services, procedures, types of health care professionals seen, and causes of injury.

The ACE analyses focused on visits to emergency rooms by children ages 17 and under for respiratory diseases. Emergency room data were selected by using the ED files only. The age variable was used to select visits by children ages 17 and under. The respiratory disease categories were selected based on the first physician's diagnosis code (DIAG1) using the International Classification of Diseases (ICD-9), first three characters:

- Asthma and all other respiratory causes: codes 460-466, 480-488, 490-496
- All respiratory causes other than asthma: codes 460-466, 480-488, 490-492, 494-496
- Upper respiratory: codes 460-466
- Pneumonia or influenza: codes 480-488
- Other lower respiratory: codes 490-492, 494-496
- Asthma: code 493

The NHAMCS uses a complex multi-stage, stratified, clustered sampling design. The statistical analyses used the patient visit survey weights (PATWT) to re-adjust the sample of visits to represent the total national population of emergency room visits in each calendar year. The rates of emergency room visits were calculated by dividing the number of visits by the associated population of noninstitutionalized civilian children ages 17 or under. The uncertainties of the rates were calculated using SUDAAN® (Research Triangle Institute, Research Triangle Park, NC 27709) statistical survey software. SUDAAN was used to calculate the estimated percentages and the standard errors of the estimated percentages, treating the population as known without uncertainty. The standard error is the estimated standard deviation of the percentage, and this depends upon the survey design. For this purpose, the public release version of NHAMCS includes the following variables:

- Masked Stratum (CSTRATM)
- Masked Primary Sampling Unit (CPSUM)

These variables are "Masked" so that the sample design represented by these variables is an approximation to the true sample design, which was not made publicly available in order to protect confidentiality

National Hospital Discharge Survey (NHDS)

The National Hospital Discharge Survey (NHDS) is conducted by the National Center for Health Statistics, a division of the Centers for Disease Control and Prevention. The complex multi-stage survey is designed to collect data on inpatients discharged from non-federal short-stay hospitals. Sampled hospitals are short-stay general or children's general hospitals located in all states and Washington DC, with an average length of stay of fewer than 30 days and six or more beds staffed for patients' use. Federal, military, and Veteran's Administration hospitals are excluded, as are hospital units of institutions. Data from sampled visits are obtained on the demographic characteristics and physician's diagnoses.

These analyses focused on hospital discharges by children ages 17 and under for respiratory diseases. The age variable was used to select visits by children ages 17 and under. The respiratory disease categories were selected based on the first physician's diagnosis code (DIAG1) using the International Classification of Diseases (ICD-9), first three characters:

- Asthma and all other respiratory causes: codes 460-466, 480-488, 490-496
- All respiratory causes other than asthma: codes 460-466, 480-488, 490-492, 494-496
- Upper respiratory: codes 460-466
- Pneumonia or influenza: codes 480-488
- Other lower respiratory: codes 490-492, 494-496
- Asthma: code 493

The NHDS uses a complex multi-stage, stratified, clustered sampling design. The statistical analyses used the survey analysis weights to re-adjust the sample of discharges to represent the total national population of hospital discharges in each calendar year. The rates of hospitalizations were calculated by dividing the number of hospital discharges by the associated population of noninstitutionalized civilian children ages 17 or under. The uncertainties of the rates were calculated using equations provided in the documentation that estimate the relative standard error of the number of discharges as $\sqrt{(a + b / \text{Total Discharges})} \times 100\%$, where a and b have different tabulated values for each year, treating the population as known without uncertainty.

National Health and Nutrition Examination Surveys (NHANES)

Since the 1960s, the National Center for Health Statistics, a division of the Centers for Disease Control and Prevention, has conducted the National Health and Nutrition Examination Surveys (NHANES), a series of U.S. national surveys of the health and nutrition status of the non-institutionalized civilian population. Interviewers obtain information on personal and demographic characteristics, including age, household income, and race and ethnicity by self-reporting or as reported by an informant. A health examination includes numerous body measurements. Samples of blood, serum, and urine are collected in Mobile Examination Centers and analyzed for numerous chemical contaminants. The NHANES use a complex multi-stage, stratified, clustered sampling design. Certain demographic groups were deliberately over-sampled, including Mexican-Americans and Blacks. The publicly released data includes survey weights to adjust for the over-sampling, non-response, and non-coverage.

The indicator for obese children uses measurements of the body mass index, The body mass index is the weight in kilograms divided by the square of the height in meters. Obese is defined as having a body mass index that is above the 95^{th} percentile for the age and sex, based on the 2000 CDC growth charts. For these analyses, the child's age was defined using the age in months at the time of the NHANES examination. The statistical analyses use the demographic file survey weights to re-adjust the examination data to represent the national population. SUDAAN statistical software for analyzing survey data was used together with SAS® (Cary, NC) software in order to calculate variances and p-values accounting for correlations within clusters.

Surveillance, Epidemiology, and End Results (SEER)

Since 1973, the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI) has been collecting and publishing cancer incidence data from population-based cancer registries that currently cover a total of 17 geographical areas in the United States and one quarter of the population. Since the coverage area has expanded over time, the trend analyses for the cancer indicator were based on the SEER 13 registries that cover 13.8% of the U.S. population for the years 1992 and later. Data include the type of cancer, age at diagnosis, year of diagnosis, sex, race, and ethnicity. Cancer incidence rates for malignant cancers in children ages 0 to 19 years were calculated using the SEER*Stat software that gives total counts of cancer cases in the SEER registries and populations from census data. Cancer mortality rates for malignant cancers in children ages 0 to 19 years were calculated using the SEER*Stat software that gives counts of cancer mortality cases in the SEER registries derived from the National Vital Statistics System (administered by the National Center for Health Statistics) and populations from census data.

National Vital Statistics System (NVSS) Natality Data

The National Vital Statistics System (NVSS) is maintained by the National Center for Health Statistics, a division of the Centers for Disease Control and Prevention. The NVSS compiles national registration certificate data for births, deaths, marriages, divorces, and fetal deaths provided by various jurisdictions, including states. The ACE indicators use NVSS birth data to determine preterm and term low birth weight births. The rate of adverse birth outcomes is defined as the total number of births with adverse outcomes (e.g., preterm births) divided by the total number of births with known gestational age. For these statistical analyses we treated all births as a large random sample, so that each birth independently has a certain probability of having an adverse birth outcome.

The preterm indicator uses the gestation period variable GESTAT3 (GESTREC3 from 2003 to 2007) coded as 1 for Under 37 weeks, 2 for 37 weeks and over, and 3 for Not stated. This variable is a recode of the variable DGESTAT (COMBGEST from 2003 to 2007) that ranges from 17 to 47 and provides the number of completed weeks of gestation of the mother. A preterm birth is defined as a birth with GESTAT3 = 1 and births with GESTAT3 = 3 were excluded from the indicator calculation.. The term low birth weight indicator uses the same gestation period variable GESTAT3 (GESTREC3 from 2003 to 2007) and the birth weight variable BIRWT4 (BWTR4 for 2003 to 2007) coded as 1 for birth weights of 1499 grams or less, 2 for 1500 to 2499 grams, 3 for 2500 grams or more, and 4 for Unknown or not stated gives the birth weight in grams. A term low birth weight birth has a gestation period of 37 or more completed weeks and a birth weight of less than 2,500 grams. Thus a term low birth weight birth is defined byGESTAT3 = 2 and BIRWT4 = 1 or 2.. Birth certificates with an unknown gestation period (GESTAT3 = 3) and/or an unknown birth weight (BIRWT4 = 4) were excluded from the calculation of the term low birth weight indicator.

In the following sections we will describe the methods used in more detail. We first describe the NHAMCS and NHDS statistical analyses of trends, year-to-year changes, and demographic group differences using the example of the rates of emergency room visits for asthma and other respiratory diseases. We then give the corresponding statistical methods and results for the NHANES obese and overweight children statistical analyses using the example of the percentages of obese children. Finally, we will describe the methods used to analyze the SEER cancer data and the NVSS adverse birth outcomes data. The analyses were performed using SAS and SUDAAN statistical programming languages.[1]

NHAMCS and NHDS Indicators-–Trend Estimation and Year-to-Year Changes

Trend Estimation

We will illustrate the statistical approach using the 1996 to 2008 data for the annual rate of emergency room visits for asthma and all other respiratory causes by children ages 0 to 17 years. The same methods can be applied for analyzing trends over a shorter or longer period. For NHAMCS and NHDS, in general, an initial sampling frame of hospitals was chosen that gets updated annually (NHAMCS) or every few years (NHDS), and several hospitals are surveyed in

[1] SAS Version 9.2 and SAS-callable SUDAAN Version 10.0.1 were used for these statistical analyses. The software code is available by request.

multiple years. However the patient emergency room visits or hospital discharges were randomly selected each year. For these analyses we make the approximation that the rates are statistically independent from year to year.

The rates of emergency room visits for asthma and all other respiratory diseases data for all demographic groups combined, stratified by year, are shown in Table 1. Table 1 also includes the standard errors of the estimated rates.

Table 1. Rate of emergency room visits for asthma and all other respiratory causes by children ages 0 to 17 years, by year, for 1996-2008.

Year	**Emergency room visits for asthma and all other respiratory causes per 10,000 children.**	**Standard Error**
1996	636.4	40.4
1997	631.5	49.4
1998	654.7	54.7
1999	619.9	47.0
2000	622.7	55.6
2001	624.0	51.2
2002	721.1	83.6
2003	740.2	51.0
2004	528.8	42.7
2005	639.8	57.1
2006	584.3	41.5
2007	625.1	76.1
2008	619.1	50.5

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

Rate (year) = Intercept + Trend $\times$ Year + Error.

In this equation, "Year" is assumed to be a numerical variable rather than a categorical variable. The error terms (observed statistic minus expected statistic) are assumed to be independent and normally distributed with a mean of zero and a known variance, v(y),that depends upon the year. The analysis makes the approximation that the data from different years are independent; the likely impact of this approximation is that the statistical significance of the trend will be slightly overstated. The normality holds approximately because of the central limit theorem. The variance v(y) is the variance of the rate for year y, calculated as the square of the standard error. For this method, we ignore the fact that v(y) is estimated from survey data and so v(y), itself, is uncertain.[2]

The Intercept and Trend are estimated using weighted linear regression. The "weight" for year y is the reciprocal of the variance v(y).[3] To implement this calculation in SAS, the SAS procedure

[2] One reviewer of a companion statistical methodology memorandum proposed using a bootstrap approach to account for the uncertainty of v(y). Although it is possible to adapt the bootstrap approach to survey data, it appears to be too computer-intensive for routine application to ACE indicators, which are frequently updated to incorporate new data.

[3] The statistical model for year y is of the form:

Rate (y) = Intercept + Trend $\times$ y + Error, where Error is normally distributed, mean zero, variance v(y). This model is mathematically equivalent to:

GENMOD was used to regress the annual statistic against the year. The WEIGHT variable was set to equal the reciprocal of v(y). The option NOSCALE for the MODEL statement was applied, because the variance for year y is given by v(y) itself and not some unknown multiple of v(y) that needs to be estimated.

The estimated value of Trend is the predicted annual change in the rate from one year to the following year. If the estimated value of Trend is statistically significantly different from zero at the 5 percent level, then a statistically significant trend has been found.

In this method, the survey design is taken into account, since the survey design is used to compute the annual rate and its estimated variance. The method assumes that the estimated rates are normally distributed (which holds, approximately), and ignores the uncertainty in the estimated variances. The method also assumes that the rates for different years are statistically independent, which is approximately valid because the visits were independently selected each year, although in some cases the visits were to the same hospital for different years.

Adjusted Test for Trend

An easy generalization of the above regression model is of the form:

Rate (for demographic group G in year y) =
Intercept + Trend × Year + g(demographic group) + Error

where the demographic group is a categorical explanatory variable, e.g., race/ethnicity or sex. In SAS, the race/ethnicity model can easily be fitted by defining a classification variable RACE that has different values for different race/ethnicity groups, and writing the model terms as "race year". Alternatively, dummy indicator variables can be created for each group, such as race1 = 1 if White non-Hispanic, race1= 0 otherwise, and the model terms are of the form "race1 race2 … racen year" Similarly for other demographic groupings.

Using this more general model, the p-value for Trend tests for a trend adjusted for the demographic variables. This would account for the possible confounding effects caused by changes in the demographics of the national populations between the years. For example, the results in Table 3, below, show that for non-Hispanics, Black children have higher rates of emergency room visits for asthma and other respiratory diseases than White children. Suppose that the Black children's population increased from one year to the next year, but the visit rates remained constant for both groups. In this case, the overall visit rate given by the value of Trend in the basic model would show an increase, but this increase would be due to the population demographic changes rather than changes in the disease visit rate due to other factors (e.g., air pollution). For this example, the value of Trend in the generalized model with an extra race variable would be zero. This shows that there would have been no trend in the emergency room visit rate had the race/ethnicity distributions remained the same from year to year. Note that this generalized model adjusts for changes in the demographics of the national population from year to year. The survey weights already adjust for changes in the demographics of the sample populations.

Rate (y)/√v(y) = Intercept/√v(y) + Trend × y /√v(y) + Error2, where Error2 is normally distributed, mean zero, variance 1.
If we define the weight w(y) = 1/v(y), then the sum of squares of Error2 is given by:
$\Sigma\{\text{Rate (y)}/\sqrt{v(y)} - \text{Intercept}/\sqrt{v(y)} - \text{Trend} \times y/\sqrt{v(y)}\}^2 =$
$\Sigma w(y) \times \{\text{Rate (y)} - \text{Intercept} - \text{Trend} \times y\}^2$.
By definition, weighted linear regression chooses the regression parameters to minimize this weighted sum of squares.

Although a wide variety of possible weighted regression models could be used to adjust for confounding effects, we restrict our modeling to a few basic models, which has practical advantages and also ensures consistency. We restrict the adjusted analysis to the following model formulation:

$$\text{Rate (year, age group, sex, race/ethnicity)} = \text{Intercept} + \text{Trend} \times \text{Year} + f(\text{age group}) + g(\text{sex}) + h(\text{race/ethnicity}) + \text{Error}$$

This model adjusts for the confounding effects of age, sex, and race/ethnicity. Since this model adds separate terms for each demographic factor, we do not consider the possibility of interactions between the demographic and year factors, so that, for example, the effect of age on emergency room visits for asthma and other respiratory diseases is assumed to be the same for each year, sex, and race/ethnicity. Moreover, we adjust for all three demographic factors, even if some or all of the terms f, g, and h are not statistically significant.

In this equation, "Year" is assumed to be a numerical variable rather than a categorical variable. The functions f, g, and h define the categorical variables for age group, sex, and race/ethnicity, represented by linear combinations of dummy variables. The error terms (observed statistic minus expected statistic) are assumed to be independent and normally distributed with a mean of zero and a known variance, v(y, a, s, r), that depends upon the year y, age group a, sex s, and race/ethnicity r. The variance v(y, a, s, r) is the variance of the visit rate for year y and the given demographic group combination, calculated as the square of the standard error. For this method, we ignore the fact that v(y, a , s, r) is estimated from survey data and so v(y, a, s, r), itself, is uncertain. The p-value for Trend tests for a trend adjusted for the demographic variables.

For analyses of emergency room visits by children ages 0 to 17 using NHAMCS, we use the following categories:

Age groups

- < 12 months
- 1 to < 2 years
- 2 to < 3 years
- 3 to < 6 years
- 6 to < 11 years
- 11 to < 16 years
- 16 to < 18 years

Sex groups

- Males
- Females

Race/ethnicity groups

- White non-Hispanic
- Black non-Hispanic
- Asian or Native Hawaiian or Pacific Islander (API) non-Hispanic
- American Indian or Alaska Native (AIAN) non-Hispanic
- Hispanic

- Other

For the NHAMCS data, the “Other” race/ethnicity category denotes children reporting multiple races and was not an available category for the years 1996 to 1998.

A statistically significant adjusted trend would show that the trend in the rate of emergency room visits for asthma and all other respiratory diseases is statistically significant after adjusting for possible confounding effects of age, gender, and race/ethnicity.

Year-to-Year Change

The second statistical issue is to test for year-to-year changes. Of interest is whether the change in rate of emergency room visits for asthma and all other respiratory diseases between the most current year and the prior year is statistically significant. For this issue, the statistical approach is to apply exactly the same set of weighted regression analyses as the tests for trend, but restrict the analyses to the data from the last two years only. If the trend from the last two years is statistically significant, then the change from year to year must also be statistically significant. Note that the value of the Trend parameter will then equal the estimated annual change in the visit rate (after adjusting for any other demographic variables in the weighted regression model).

For the unadjusted model there are only two rates in the year-to-year change dataset, for the the most recent year and the previous year. In this case the trend is the difference in rates and its standard error is the square root of the sum of their known variances. Thus the test is equivalent to a z-score test of the rate difference divided by its standard error. For the adjusted model, the much larger dataset consists of the rates and standard errors for each combination of year and demographic variables.

Analyses of NHDS Indicators

Exactly the same approach is used for the analyses of trends and year-to-year changes in children’s hospital discharges for asthma and all other respiratory diseases, with the following modifications: Since NHDS does not include ethnicity data, the race/ethnicity demographic variable is replaced by the race variable:

Race groups

- White
- Black
- Asian or Native Hawaiian or Pacific Islander (API)
- American Indian or Alaska Native (AIAN)
- Other

For these NHDS data the “Other” race category includes children of Other races,[4] children of multiple races (for 2000 or later), and children with a race that was not stated. Note that following NCHS recommendations, due to concerns about high uncertainty, detailed results are not presented in the ACE report for the American Indian/Alaskan Native, Asian and Pacific Islander, and Other categories. However all five categories were used to define the race groups for the statistical comparisons, since those comparisons take the uncertainty into account.

[4] Although the NHDS hospital discharge data includes Other races as a possible category, the corresponding census population data only provides estimates for the single race groups: White, Black,, Asian, AIAN, Hawaiian and Pacific Islander; and for multiple races.

Normality test of residuals for the rate and its logarithm

These trend models assume that the error terms are normally distributed with a mean of zero and a variance of v(y, a, s, r). To evaluate this assumption, the standardized Pearson residuals were calculated and tested for normality using the Shapiro-Wilk normality test. The standardized Pearson residual is defined as the observed rate minus the expected rate, divided by the standard error of this difference. For emergency room visits for asthma and all other respiratory causes, the p-value for the Shapiro-Wilk normality test of the residuals using the unadjusted trend test was 0.1832. Since this p-value is greater than 5%, the statistical test does not reject normality at the 5% level. The corresponding p-value for the adjusted trend test was < 0.0001, so that normality of the residuals was rejected at the 5% level. For hospitalizations for asthma and all other respiratory causes, the p-values for the Shapiro-Wilk normality test of the residuals were 0.5712 for the unadjusted trend test and < 0.0001 for the adjusted trend test. The rejection of normality for the adjusted trend tests suggests that the form of the statistical model might not be an adequate description of the data so that the associated trend tests may be a poor approximation. Normality tests for these and additional model formulations are presented in Table 2.

We considered trying to improve the model by replacing the dependent variable Rate by the natural logarithm of the rate. For this logarithmic model, the trend parameter is the estimated annual change in the logarithm of the rate instead of the annual change in the rate itself, and so has a different interpretation. To fit this model, the variance of the logarithm of the rate was approximated as $v(y, a, s, r)/rate^2$, using a standard Taylor series approximation (also known as the "delta" method). The p-values for the Shapiro-Wilk test of normality for residuals from the various alternative models are tabulated in Table 2, including the models for the adjusted test for demographic differences that are presented below. Table 2 shows that for hospitalization the residuals from the rate models consistently show more support for normality than the residuals from the models using the logarithm of the rate, but for emergency room visits the opposite tends to hold. For consistency of these two companion indicators, we choose to use the untransformed rate as the dependent variable, since the more meaningful trend of interest is a trend in the rates themselves and since this approach avoids the additional approximation needed to estimate the variance of the logarithm of the rate. Furthermore, we note that in most cases the p-values for the statistical trend tests using the rate and the logarithm of the rate were very close.

Table 2. Shapiro-Wilk statistical significance tests of normality for the residuals from various models of the rates of emergency room visits and hospitalizations for asthma and all other respiratory causes by children ages 0 to 17.

		P-value for Shapiro Wilk test of normality (P-values > 0.05 fail to reject normality)	
Indicator	**Statistical Test**	**Dependent Variable = Rate**	**Dependent Variable = Log(Rate)**
Emergency Room Visits	Unadjusted test for trend	0.1832	0.1401
	Adjusted test for trend	< 0.0001	0.2363
	Adjusted test for year to year change	< 0.0001	0.1620
	Adjusted test for demographic differences	0.0025	0.6102

Hospitalizations	Unadjusted test for trend	0.5712	0.4918
	Adjusted test for trend	< 0.0001	< 0.0001
	Adjusted test for year to year change	0.0012	0.0002
	Adjusted test for demographic differences	0.2698	0.0425

Results for Trends and Year-to-Year Changes in Emergency Room Visits for Asthma and All Other Respiratory Causes

These weighted regression tests were applied to the emergency room visits data from NHAMCS. The trend test was applied to the trend over the period 1996 to 2008. The year-to-year change test was applied to the change from 2007 to 2008. Table 3 gives the statistical significance tests. From and To give the range of years analyzed. Against is the variable of interest, which is year for the trend and year-to-year change tests. The P-VALUES columns give the p-value of the statistical test, so that p-values at or below 0.05 are statistically significant at the 5 percent significance level. In this memorandum, results will be referred to as statistically significant if they are significant at the 5 percent level. “Unadjusted” gives the p-values for the trend and year-to-year change test without adjusting for potential demographic confounders. “Adjusted” gives the p-values for the trend and year-to-year change test after adjusting for potential demographic confounders, in this case for age, sex, and race/ethnicity. For example, from row 1, we can see that the trend is not statistically significant if not adjusted for demographic confounders, but the trend is statistically significant after adjusting for demographic confounders (unadjusted p-value = 0.381, and adjusted p-value < 0.001). The year-to-year change from 2007 to 2008 is not statistically significant.

Table 3. Statistical significance tests of trends and year-to-year changes in the rates of emergency room visits for asthma and all other respiratory causes by children ages 0 to 17.

				P-VALUES	
Variable	**From**	**To**	**Against**	**Unadjusted**	**Adjusted***
All asthma and other respiratory causes	1996	2008	year	0.381	< 0.001
All asthma and other respiratory causes	2007	2008	year	0.948	0.177

*For Against = "year," the p-values are adjusted for age, sex, and race/ethnicity.

NHAMCS and NHDS Indicators -–Compare Demographic Groups

The third statistical issue is to test for differences in prevalence among demographic groups. In this case the analysis is restricted to a group of years, and the comparison is between the overall rate of emergency room visits or hospitalizations for that period in different demographic groups, such as for different age groups. Weighted regression modeling again is proposed. The same demographic group categories as the trend analyses are used.

There is a wide variety of possible demographic comparisons that could be addressed using statistical tests. These statistical tests are intended to confirm or evaluate the patterns that we can see when comparing the rates for different race/ethnicities or races as well as for other demographic subgroups. The results of these statistical tests can then be used to decide which differences should be highlighted in our discussions of the indicators. The methods for NHAMCS are described below. The methods for NHDS are essentially the same, except that the six race/ethnicity groups for NHAMCS are replaced by five race groups for NHDS .

<u>Unadjusted Test for Demographic Group Differences</u>

Consider the case of testing for sex differences. The regression model can be written in the general form:

Rate (sex) = Intercept + g(sex) + Error

This is an unadjusted model, since the possible confounding effects of other demographic variables are not accounted for. The test for no demographic differences tests that the overall sex effect defined by the function g(sex) is zero, which implies that the prevalence for boys is the same as the prevalence for girls.

It is probably easier to understand the model if it is rewritten using dummy indicator variables, using an intercept together with a linear combination of the indicator I_b that equals 1 for boys and 0 for girls, and the indicator I_g that equals 1 for girls and 0 for boys:

Rate (sex) = Intercept + bI_b + gI_g + Error

The test for no demographic differences tests if the two slope parameters b and g are equal. The alternative hypothesis is that b and g are not equal. In the full rank parameterization, g is set to equal zero, so that the null hypothesis is that b equals zero.

In a similar manner we can apply tests for differences among age groups or among race/ethnicity groups. The general formulations are:

Rate (age group) = Intercept + f(age) + Error; test f(age) = 0

Rate (race/ethnicity) = Intercept + h(race/ethnicity) + Error;

test h(race/ethnicity) = 0

Adjusted Test for Demographic Group Differences

Just as for the trend test, it is possible that the overall differences could be confounded. For example, any difference (or lack of difference) between the rates of emergency room visits for boys and girls might be partly attributable to differences between boys and girls in their population distributions of race/ethnicity and/or age group, assuming that race/ethnicity and/or age group also affect emergency room visits for asthma and other respiratory diseases. To adjust for the confounding effects, the following regression model is proposed:

Rate (age group, sex, race/ethnicity) =
Intercept + f(age) + g(sex) + h(race/ethnicity) + Error

The same model is used for all three adjusted tests. As before, this model assumes that there are no interactions between the different demographic group effects.

Adjusted test for age: Test f(age) = 0
Adjusted test for sex: Test g(sex) = 0
Adjusted test for race/ethnicity: Test h(race/ethnicity) = 0

Race/Ethnicity and Income Paired Comparisons.

The unadjusted and adjusted tests for race/ethnicity comparisons described in the last two subsections are each a broad test for whether there are any differences in rates of emergency room visits for asthma and other respiratory diseases among race/ethnicity groups or whether all the race/ethnicity groups have the same rate. If significant race/ethnicity differences are found overall, then we would like to use the statistical analysis to evaluate statements that say that one race/ethnicity group has a higher rate than another group or that one race/ethnicity group has the highest rate among all the groups. If the differences are statistically significant, then these statements are supported. Otherwise, the observed differences are small enough that they are applicable to the sample but have not been shown to be applicable to the U.S. population. For example, we can make a statistical comparison between the rate for the two groups White non-Hispanic and Black non-Hispanic. These specific comparisons are easily made using appropriately defined contrasts applied to the regression model that tests whether the corresponding regression coefficients are equal.

Assume that the six race/ethnicity groups are numbered 1 to 6 and that I_j is the dummy variable that equals 1 for members of group j and equals 0 otherwise. Then the unadjusted regression model for race/ethnicity is of the form:

Rate (race/ethnicity) = Intercept + h(race/ethnicity) + Error

which is equivalent to the model:

$$\text{Rate (race/ethnicity)} = \text{Intercept} + h(1)I_1 + h(2)I_2 + h(3)I_3 + h(4)I_4 + h(5)I_5 + h(6)I_6 + \text{Error}$$

To compare groups j and k, the appropriate statistical test is of the null hypothesis h(j) = h(k), which tests the null hypothesis that the two groups have the same regression parameters, and hence the same predicted rate. This test is easily performed in the SAS procedure GENMOD using an effects contrast for this regression model. For example, to compare groups 1 and 3, the contrast statement is coded as:

```
contrast '1v3' race 1 0 -1;
```

A p-value is calculated for each pair of race/ethnicity groups. The same set of contrast comparisons are used for the adjusted model, with additional terms for age and sex.

Since there are 15 paired comparisons between the six race/ethnicity groups, there is a high probability that at least one paired comparisons will be found statistically significant even if there are no race/ethnicity differences in the emergency room visit rates. While it is possible to make adjustments to the p-values to account for this multiple comparisons issue, we choose not to make an adjustment for multiple comparisons, but users of the results should be aware of the increased probability of an erroneous statistical decision when multiple comparisons are made.

As mentioned above, we would like to use the results to evaluate statements about the race/ethnicity group with the highest rates. For example, the ACE analyses of the 2005-2008 NHAMCS data showed that Black non-Hispanic children had the highest rates of emergency room visits for asthma and other respiratory diseases. A statistical test for the specific alternative hypothesis that Black non-Hispanic children had the highest rate among the six race/ethnicity groups is not available. Instead we can examine the paired comparisons between the Black non-Hispanic group and each of the other five race/ethnicity groups; if we find that all the differences are statistically significant, then we can confirm that the Black non-Hispanic children had the highest rate among the six race/ethnicity groups since the differences are not attributable to random variation.

Results for Tests on Demographic Group Differences in Emergency Room Visits for Asthma and All Other Respiratory Causes

Table 4 shows the rates of emergency room visits for asthma and all other respiratory causes by children ages 0 to 17 years for the different race/ethnicity groups, for 2005-2008, together with their standard errors.

Table 4. Rate of emergency room visits for asthma and all other respiratory causes by children ages 0 to 17 years, by race/ethnicity, for 2005-2008.

Race/Ethnicity	Emergency room visits for asthma and other respiratory causes per 10,000 children	Standard Error
White non-Hispanic	486.6	32.9
Black non-Hispanic	1240.1	129.8
Asian or Native Hawaiian or Pacific Islander (API) non-Hispanic	371.4	71.2
American Indian or Alaska Native (AIAN) non-Hispanic	536.2	148.8
Hispanic	671.5	78.5
Other	52.9	23.0

Comparisons of emergency room visit rates for asthma and other respiratory causes between pairs of race/ethnicity groups are shown in Table 5. For the “Unadjusted” comparisons, the only explanatory variables are terms for each race/ethnicity group. For these unadjusted comparisons, the statistical tests compare the rates for each pair of race/ethnicity groups. For the “Adjusted for age, sex” comparisons, the explanatory variables are terms for each race/ethnicity group together with terms for each age group and sex. For these adjusted comparisons, the statistical test compares the pair of race/ethnicity groups after accounting for any differences in the age and sex distributions between the race/ethnicity groups.

Table 5. Statistical significance tests comparing the rates of emergency room visits for asthma and other respiratory causes by children ages 0 to 17, between pairs of race/ethnicity groups, for 2005-2008.

			P-VALUES	
Variable	**First race/ethnicity group**	**Second race/ethnicity group**	**Unadjusted**	**Adjusted for age, sex**
Asthma and all other respiratory causes	White non-Hispanic	Black non-Hispanic	< 0.001	< 0.001
Asthma and all other respiratory causes	White non-Hispanic	API non-Hispanic	0.745	0.032
Asthma and all other respiratory causes	White non-Hispanic	AIAN non-Hispanic	0.142	< 0.001
Asthma and all other respiratory causes	White non-Hispanic	Hispanic	0.030	0.084
Asthma and all other respiratory causes	White non-Hispanic	Other	< 0.001	< 0.001
Asthma and all other respiratory causes	Black non-Hispanic	API non-Hispanic	< 0.001	< 0.001
Asthma and all other respiratory causes	Black non-Hispanic	AIAN non-Hispanic	< 0.001	< 0.001
Asthma and all other respiratory causes	Black non-Hispanic	Hispanic	< 0.001	< 0.001
Asthma and all other respiratory causes	Black non-Hispanic	Other	< 0.001	< 0.001
Asthma and all other respiratory causes	API non-Hispanic	AIAN non-Hispanic	0.318	0.881
Asthma and all other respiratory causes	API non-Hispanic	Hispanic	0.421	0.007
Asthma and all other respiratory causes	API non-Hispanic	Other	0.001	< 0.001
Asthma and all other respiratory causes	AIAN non-Hispanic	Hispanic	0.005	< 0.001
Asthma and all other respiratory causes	AIAN non-Hispanic	Other	< 0.001	< 0.001
Asthma and all other respiratory causes	Hispanic	Other	< 0.001	< 0.001

For the overall unadjusted analyses, 11 of the 15 pairs of race/ethnicity groups had statistically significant differences at the 5 percent level. For the overall adjusted analyses, 13 of the 15 pairs of race/ethnicity groups had statistically significant differences at the 5 percent level. In particular, the adjusted analyses show that the differences between Black non-Hispanic, Hispanic, White non-Hispanic, and API non-Hispanic are all statistically significant after adjusting for age and sex, except for the comparison between White non-Hispanic and Hispanic. For the comparison between White non-Hispanic and API non-Hispanic children, the unadjusted p-value was not significant (at the 5 percent level) but the adjusted p-value was significant. This means that although the overall visit rates are similar for White non-Hispanic and API non-Hispanic children, the visit rates depend upon the age (and to a lesser extent on the sex) so that if you account for differences in the age distributions between White non-Hispanic and API non-Hispanic children, then the visit rates would be significantly different. On the other hand, for the comparison between White non-Hispanic and Hispanic children, the unadjusted p-value was significant but the adjusted p-value was not significant. In this case, a possible explanation is that the differences in the age distributions between White non-Hispanic and Hispanic children caused the visit rates to be significantly different, but if the differences in the age distributions are accounted for, then the visit rates would no longer be significantly different.

Additional comparisons are shown in Table 6. The Against = “age” unadjusted p-value compares the rates between all the age groups. The adjusted p-value includes adjustment terms for sex and race/ethnicity in the model. These results show statistically significant differences in the visit rates for different age groups.

Table 6. Statistical significance tests of age group differences in the rates of emergency room visits for asthma and other respiratory causes by children ages 0 to 17, for 2005-2008.

				P-VALUES	
Variable	**From**	**To**	**Against**	**Unadjusted**	**Adjusted***
Asthma and all other respiratory causes	2005	2008	age	< 0.001	< 0.001

*For Against = ”age,” the p-values are adjusted for sex and race/ethnicity.

NHANES Obese Children Indicator-–Trend Estimation and Year-to-Year Changes

Trend Estimation

We will illustrate the approach using the 1976 to 2008 NHANES data for the percentage of obese children ages 2 to 17 years. Obese children are defined as children who have a body mass index above the 95^{th} percentile for their age and sex from the CDC growth charts. The same methods can be applied for analyzing trends over a shorter or longer period. The statistical methods for analyzing the NHANES obesity data use a similar approach to the NHAMCS or NHDS weighted regression analyses. Each of the NHAMCS and NHDS weighted regression analyses of the emergency room visit or hospitalization rates is replaced by a logistic regression analysis of the percentage of obese children. For the trend analyses, the NHAMCS and NHDS year becomes the middle year of the NHANES period. This is necessary because the NHANES data were not collected continuously until 1999-2000, and the NHANES data are currently collected over a two-year period. Thus the reported NHANES trend results estimate annual changes rather than the changes between NHANES periods.

The analyses use the percentages of obese children ages 2 to 17 from the NHANES data for the following periods. The same methods would work as future periods of data are added.

- NHANES II 1976-1980: Middle year = 1978
- NHANES III Phase 1: October 1988 to October 1991. Middle year = 1990.3
- NHANES III Phase 2: September 1991 to October 1994. Middle year = 1993.3
- NHANES 1999-2000: Middle year = 1999.5
- NHANES 2001-2002: Middle year = 2001.5
- NHANES 2003-2004: Middle year = 2003.5
- NHANES 2005-2006: Middle year = 2005.5
- NHANES 2007-2008: Middle year = 2007.5

The NHANES survey design selects independent random samples of subjects for each survey period, treating NHANES III 1988-1994 as a single survey period. For NHANES II from 1976-1980, the strata were numbered from 1 to 21. For NHANES III, the strata over the entire six-year period were numbered from 1 to 49. For NHANES 1999-2000 and later, the strata were numbered from 1 to 14 in 1999-2000, 15 to 28 in 2001-2002, etc. To properly analyze the NHANES data, it is necessary to distinguish the strata in NHANES II and NHANES III from the same numbered strata in NHANES 1999-2000 and later. Otherwise, the statistical analysis would incorrectly treat all the data from the stratum numbered 1 as having being selected from the same statistical stratum, and similarly for the other stratum numbers, which would lead to incorrect estimates of the variances and p-values, although the annual statistics would not be affected. For this reason, 1,000 should be added to the original stratum numbers for the NHANES II survey; instead of 1,000, any number that avoids overlaps between the NHANES II stratum numbers and the stratum numbers for 1988 and later can be chosen. For the same reason, 2,000 should be added to the original stratum numbers for the NHANES III survey; instead of 2,000, any number that avoids overlaps between the NHANES III stratum numbers and the stratum numbers for 1976-1980 and 1999-2000 and later can be chosen.

Unadjusted Test for Trend

Table 7 shows the percentages of obese children ages 2 to 17, by NHANES period.

Table 7. Percentages of obese children ages 2 to 17 years, from NHANES.
.

NHANES Period	Percentage of Obese Children
1976-1980	5.4%
1988-1991	9.4%
1991-1994	11.0%
1999-2000	13.8%
2001-2002	15.2%
2003-2004	16.8%
2005-2006	15.3%
2007-2008	16.9%

The statistical test for trend tests whether the percentage is constant from year to year using a logistic model. The logarithm of the odds for year y is equal to log(p/(1-p)) = logit(p), the logit function, where the proportion of obese children, p, depends upon the year, y. If the log odds are all equal, then logit(p) must be the same for all years, so that the percentage is constant.

The logistic model has the general formulation:

Logit {Prob(obesity for child c)} = function of explanatory variables for child c

In general this function can be any parametric function of the year and of the child's explanatory variables, such as age, other demographic variables, state, county, other questionnaire variables, etc. These variables can be numerical variables or categorical variables. Categorical variables are represented by dummy indicator variables, one for each level. For example, for the categorical variable sex, this is a linear combination of the indicator I_b that equals 1 for boys and 0 for girls, and the indicator I_g that equals 1 for girls and 0 for boys.[5] The data used to fit this model consists of the binary responses (obese or not obese) for each surveyed child.

For the unadjusted trend model, the percentages are assumed to have been generated from the statistical model:

Logit {Prob(obesity for child c)} = Intercept + Trend $\times$ Year

[5] To avoid an over-parameterized model, if there is an intercept, then the coefficient for the last category is usually set to be zero, so that in effect the last category is not used in the regression model. Thus for sex we only need the I_b indicator.

In this equation, “Year” is assumed to be a numerical variable rather than a categorical variable and the observed values are the middle years of each NHANES period. The Intercept and Trend are estimated using logistic regression applied to the binary (obese or not obese) responses for each surveyed child.[6] The SUDAAN procedure RLOGIST was used to fit this logistic regression model, taking into account the complex survey design and survey weights. The estimated value of Trend is the predicted annual change in the logit from one year to the following year, which is one half of the predicted change in the logit between two consecutive two-year NHANES periods. If the estimated value of Trend is statistically significantly different from zero at the 5 percent level, then a statistically significant trend has been found.

Adjusted Test for Trend

In a similar manner to the NHAMCS and NHDS adjusted test for trend, we also use an adjusted test for trend for the NHANES obesity data. This is a test for trend adjusted for possible confounding effects of the demographic variables.

The adjusted analysis is restricted to the following model formulation:

$$\text{Logit }\{\text{Prob(obesity for child c)}\} = \text{Intercept} + \text{Trend} \times \text{Year} + \text{f(age group)} + \text{g(sex)} + \text{h(race/ethnicity)} + \text{k(income)}$$

In this equation, “Year” is assumed to be a numerical variable rather than a categorical variable and the observed values are the middle years of each NHANES period. The functions f, g, h, and k define the categorical variables for age group, sex, race/ethnicity, and income, represented by linear combinations of dummy variables. This model assumes there are no interactions between these four demographic effects. The p-value for Trend tests for a trend adjusted for the demographic variables.

For analyses of the percentages of obese children ages 2 to 17, we use the following categories:

Age groups

- 2-5 years
- 6-10 years
- 11-15 years
- 16-17 years

Sex groups

- Males
- Females

[6] The fitted model for the population of US children has the following likelihood:

$$\prod_{\text{Obese children}} \frac{e^{\text{Intercept}+\text{Trend}\times\text{Year}}}{1+e^{\text{Intercept}+\text{Trend}\times\text{Year}}} \prod_{\text{Non-obese children}} \frac{1}{1+e^{\text{Intercept}+\text{Trend}\times\text{Year}}}$$

The first expression is a product of the estimated probabilities of being obese for each obese child. The second expression is a product of the estimated probabilities of not being obese for each non-obese child. The values of Intercept and Trend are estimated for the population by maximizing this likelihood. For the survey data, these parameters are estimated by maximizing a survey-weighted estimate of this likelihood.

Race/ethnicity groups

- White non-Hispanic
- Black non-Hispanic
- Mexican-American
- Other

Income groups

- Below Poverty Level
- At or Above Poverty Level
- Unknown Income

The race/ethnicity groups are based on the RIDRETH1 NHANES variable and are the same as the groups used in the current ACE. The “Other” race/ethnicity includes 1) individuals whose racial or ethnic identity was not White non-Hispanic, Black non-Hispanic, or Mexican-American, and 2) multiracial persons. Persons of “Other” race/ethnicity are selected into the survey with a probability that is very much lower than White non-Hispanic, Black non-Hispanic, and Mexican-American individuals, and as a group they are not representative of all other races and ethnicities in the United States.
A significant proportion of the respondents did not report their family income. Among the sampled children ages 2 to 17 years with known obesity status, 7% had an unknown income over the entire period 1976 to 2008, 5% had an unknown income in 2005-2006, 7% had an unknown income in 2007-2008, and 6% had an unknown income in 2005-2008. We therefore include the Unknown Income category as one of the income groups, since grouping the unknown income cases with one of the other groups would lead to unrealistic models for the effects of income.

A statistically significant adjusted trend would show that the trend in the percentage of obese (children is statistically significant after adjusting for possible confounding effects of age, sex, race/ethnicity, and income.

Stratified Tests for Trend

To investigate possible interactions between the trend and the demographic group, we use a stratified analysis that separately tests for a trend in each demographic group. The analysis adjusts for possible confounding by the other demographic groups.

The following logistic regression models are used, one for each demographic subset:

Adjusted trend test by age group:

Logit {Prob(obesity for child c)} =
Intercept + Trend × Year + g(sex) + h(race/ethnicity) + k(income)
(model fitted to children in age group a only)

Adjusted trend test by gender:

Logit {Prob(obesity for child c)} =
Intercept + Trend × Year + f(age) + h(race/ethnicity) + k(income)
(model fitted to children in sex group s only)

Adjusted trend test by race/ethnicity:

Logit {Prob(obesity for child c)} =
Intercept + Trend × Year + f(age) + g(sex) + k(income)
(model fitted to children in race/ethnicity group r only)

Adjusted trend test by income:

Logit {Prob(obesity for child c)} =
Intercept + Trend × Year + f(age) + g(sex) + h(race/ethnicity)
(model fitted to children in income group i only)

A statistically significant adjusted trend would show that the trend in the percentage of obese children in the particular demographic group is statistically significant after adjusting for the effects of the other demographic groups.

Year-to-Year Change

To compare the percentages of obese children in different NHANES periods, the set of trend analyses are repeated using only the data for the two periods of interest. For consistency, the year effect is again assumed to be numerical, so that the estimated Trend is the estimated annual change in the prevalence. To estimate the period-to-period change for two consecutive two-year periods, the estimated Trend coefficient should be doubled; the p-value is not affected.

Results for Trends and Year-to-Year Changes in Percentages of Obese Children

These logistic regression tests were applied to the obesity data from NHANES. The trend test was applied to the trend over the period 1976 to 2008. The year-to-year change test was applied to the change from 2005 to 2008, i.e., between the NHANES periods 2005-2006 and 2007-2008. Table 8 gives the statistical significance tests. From and To give the range of years analyzed. Against is the variable of interest, which is year for the trend and year-to-year change tests. Subset is the demographic group subset for the stratified tests for trend. The P-VALUES columns give the p-value of the statistical test, so that p-values less than 0.05 are statistically significant at the 5 percent significance level. “Unadjusted” gives the p-values for the trend and year-to-year change test without adjusting for potential demographic confounders. “Adjusted” gives the p-values for the trend and year-to-year change test after adjusting for potential demographic confounders, in this case for age, sex, race/ethnicity, and income. For example, from row 1, we can see that the overall trend is statistically significant (unadjusted p-value < 0.001, and adjusted p-value < 0.001). From row 2, we can see that the trend for Males is statistically significant (adjusted p-value < 0.001). All the p-values in this table show statistically significant trends from 1976 to 2008. The year-to-year change from 2005-2006 to 2007-2008 is not statistically significant.

Table 8. Statistical significance tests for trends and year-to-year changes in the percentage of children ages 2 to 17 years who were obese for 1976-2008.

					P-VALUES	
Variable	**Against**	**From**	**To**	**Subset**	**Unadjusted**	**Adjusted***
Obese	year	1976	2008		< 0.001	< 0.001
Obese	year	1976	2008	Males	< 0.001	< 0.001
Obese	year	1976	2008	Females	< 0.001	< 0.001

					P-VALUES	
Variable	**Against**	**From**	**To**	**Subset**	**Unadjusted**	**Adjusted***
Obese	year	1976	2008	White non-Hispanic	< 0.001	< 0.001
Obese	year	1976	2008	Black non-Hispanic	< 0.001	< 0.001
Obese	year	1976	2008	Mexican	< 0.001	< 0.001
Obese	year	1976	2008	Other	< 0.001	< 0.001
Obese	year	1976	2008	2-5 years	< 0.001	< 0.001
Obese	year	1976	2008	6-10 years	< 0.001	< 0.001
Obese	year	1976	2008	11-17 years	< 0.001	< 0.001
Obese	year	1976	2008	Below Poverty Level	< 0.001	< 0.001
Obese	year	1976	2008	At or Above Poverty Level	< 0.001	< 0.001
Obese	year	1976	2008	Unknown Income	0.004	0.006
Obese	year	2005	2008		0.389	0.380

*For Against = “year,” the p-values are adjusted for age, sex, race/ethnicity, and income.

NHANES Obese Children Indicator-–Compare Demographic Groups

To compare the percentages of obese children among different demographic groups, an analogous approach to the NHAMCS/NHDS analysis is applied. The analysis combines the data from two NHANES periods. The same demographic group categories as the trend analyses are used. For example, Table 9 shows the percentages of obese children for different race/ethnicity groups.

Table 9. Percentage of children ages 2 to 17 years who were obese, by race/ethnicity, for 2005-2008.

Race/Ethnicity	Percentage of Obese Children
White non-Hispanic	14.2%
Black non-Hispanic	20.0%
Mexican-American	21.9%
Other	14.5%

The most important comparisons are for differences among demographic groups. The unadjusted tests tabulated in Table 10 below use the models:

Logit {Prob(obesity for child c)} = Intercept + f(age),
Logit {Prob(obesity for child c)} = Intercept + g(sex), or
Logit {Prob(obesity for child c)} = Intercept + k(income) (Children with Known Income)

The adjusted tests tabulated in Table 10 use the model:

Logit {Prob(obesity for child c)} = Intercept + f(age) + g(sex) +
h(race/ethnicity) + k(income)

Test for age: Test f(age) = 0
Test for gender: Test g(sex) = 0
Test for income: Test k(income) = 0

The test for income was applied to the subset of those with known incomes only, which was approximately 94% of the sampled children, and thus this test compares children below the poverty level with children at or above the poverty level.

Table 11 below tabulates contrast comparisons between pairs of race/ethnicity or race/ethnicity/income groups. These comparisons answer questions about whether two specific race/ethnicity or race/ethnicity/income groups are statistically significantly different, and can be used to evaluate or confirm statements that a particular race/ethnicity or race/ethnicity/income group has the highest percentage, i.e., that the differences are not attributable to random variation.

For example, consider the unadjusted test for race/ethnicity differences, based on the model:

Logit {Prob(obesity for child c)} = Intercept + h(race/ethnicity)

The overall test is for the hypothesis that there are no race/ethnicity differences. Number the four race/ethnicity groups from 1 to 4 and define the four associated dummy variables so that I_j is the dummy variable that equals 1 for members of group j and equals 0 otherwise. Then the unadjusted logistic regression model for race/ethnicity is of the form:

Logit {Prob(obesity for child c)} = Intercept + $h(1)I_1 + h(2)I_2 + h(3)I_3 + h(4)I_4$

To compare groups j and k, the appropriate statistical test is of the null hypothesis h(j) = h(k), which tests the null hypothesis that the two groups have the same regression parameters, and hence the same predicted percentage. This test is easily performed in the SUDAAN RLOGIST procedure using an effects contrast. For example, to compare groups 1 and 3, the contrast statement is coded as:

effects race = (1, 0, −1, 0);

A p-value is calculated for each pair of race/ethnicity groups. The same set of contrast comparisons are used for the adjusted model, with additional terms for age, sex, and income. As for the NHAMCS/NHDS analyses, we choose not to make an adjustment for the multiple paired comparisons.

As mentioned above, we would like to use the results to evaluate statements about the race/ethnicity group with the highest percentage. For example, Table 9 shows that Mexican-American children had the highest percentage of obesity among the four race/ethnicity groups in 2005-2008. A statistical test for the specific alternative hypothesis that Mexican-American children had the highest percentage among the four race/ethnicity groups is not available. Instead we can examine the paired comparisons between the Mexican-American group and each of the other three race/ethnicity groups and if we find that all the differences are statistically significant, then we can confirm that the Mexican-American children had the highest percentage among the four race/ethnicity groups since the differences are not attributable to random variation..

Also of interest are evaluating differences between pairs of race/ethnicity groups at the same income levels and evaluatingdifferences between children in a given race/ethnicity group below poverty and children in the same group at or above poverty . . For these analyses, the unadjusted model was of the form:

Logit {Prob(obesity for child c)}= Intercept + q(race/ethnicity and income)

The categorical variable q has 12 categories, including the four combinations of race/ethnicity with unknown income. Thus this model was fitted to all the data including the 6% of sampled children with unknown income. For each of the other eight categories, we calculated the statistical significance of each of the 12 paired differences of interest in the above model (i.e., excluding comparisons with children of unknown income and excluding comparisons between children below poverty in a given race/ethnicity group and children at or above poverty in a different race/ethnicity group). This model allows for possible interactions between race/ethnicity and income. The results are tabulated in Tables 11 and 12 below.

Results for NHANES Tests on Demographic Group Differences in Percentages of Obese Children

The demographic comparisons presented in Table 10 show significant differences for age, both unadjusted and adjusted, but non-significant differences for sex. The differences in the percentages of obese children between those below poverty level and at or above poverty are also significant.

Table 10. Statistical significance tests of demographic group differences in percentages of children who were obese for 2005-2008.

					P-VALUES	
Variable	From	To	Against	Subset	Unadjusted	Adjusted*
Obese	2005	2008	age		< 0.001	< 0.001
Obese	2005	2008	sex		0.153	0.148
Obese	2005	2008	income	Known Income	0.005	0.020

*For Against = "age," the p-values are adjusted for sex, race/ethnicity, and income.
For Against = "sex," the p-values are adjusted for age, race/ethnicity, and income.
For Against = "income," the p-values are adjusted for age, sex, and race/ethnicity.

The paired race/ethnicity comparisons in Table 11 compare different race/ethnicity groups. The p-values for "All incomes" are for statistical significance tests using the race/ethnicity contrasts to compare the two race/ethnicity groups. The p-values for "Below Poverty Level" or "At or Above Poverty Level" are for statistical significance tests using the race/ethnicity/income contrasts to compare the two race/ethnicity groups for the subsets "Below Poverty Level" or "At or Above Poverty Level," respectively. For example, from row 1 the unadjusted analysis shows that the percentages for White non-Hispanic and Black non-Hispanic are statistically significantly different (p-value = 0.003), the percentages for White non-Hispanic Below Poverty Level and Black non-Hispanic Below Poverty Level are not statistically significantly different (p-value = 0.471), and the percentages for White non-Hispanic At or Above Poverty Level and Black non-Hispanic At or Above Poverty Level are statistically significantly different (p-value = 0.003).

Table 11. Statistical significance tests comparing the percentages of children ages 2 to 17 years who were obese, between pairs of race/ethnicity groups, for 2005-2008.

			P-VALUES					
Variable	First race/ethnicity group	Second race/ethnicity group	All incomes	All incomes (adjusted for age, sex, income)	Below Poverty Level	Below Poverty Level (adjusted for age, sex)	At or Above Poverty Level	At or Above Poverty Level (adjusted for age, sex)
Obese	White non-Hispanic	Black non-Hispanic	0.003	0.005	0.471	0.434	0.003	0.002
Obese	White non-Hispanic	Mexican-American	< 0.001	< 0.001	0.045	0.022	0.002	0.001
Obese	White non-Hispanic	Other	0.882	0.939	0.104	0.084	0.436	0.483
Obese	Black non-Hispanic	Mexican	0.247	0.183	0.328	0.218	0.421	0.356
Obese	Black non-Hispanic	Other	0.010	0.015	0.503	0.462	0.002	0.002
Obese	Mexican-American	Other	0.003	0.003	0.933	0.994	0.003	0.002

Table 12 compares the percentages for the same race/ethnicity group at different income levels. The p-values are for statistical significance tests using the race/ethnicity/income contrasts to compare the race/ethnicity group for the subset "Below Poverty Level" with the same race/ethnicity group for the subset "At or Above Poverty Level." The row for Population = "All" compares the percentages for children of all race/ethnicity groups at different income levels, exactly as in Table 10.

Table 12. Statistical significance tests comparing the percentages of children ages 2 to 17 who were obese, between those below poverty level and those at or above poverty level, for 2005-2008.

		P-VALUES	
Variable	Population	Unadjusted	Adjusted (for age, sex)*
Obese	All	0.005	0.020
Obese	White non-Hispanic	0.115	0.093
Obese	Black non-Hispanic	0.932	0.954
Obese	Mexican-American	0.770	0.518
Obese	Other	0.006	0.005

* Comparison for "All" is adjusted for age, sex, and race/ethnicity; comparisons for race/ethnicity categories are adjusted for age and sex.

SEER Indicators

The Surveillance, Epidemiology, and End Results (SEER) Program includes data on cancer incidence and mortality. To test for trends and demographic group differences, a Poisson regression method was applied. The general approach is similar to the logistic regression method applied for the NHANES data on the percentages of obese children. Therefore, to avoid repetition, we will simply describe the general statistical model formulation and the demographic groups used in these analyses,

The logistic regression models used for the NHANES data on the percentages of obese children are of the general form:

Logit {Prob(obesity for child c)} = function of explanatory variables for child c

For example, the adjusted trend analysis model is of the form:

Logit {Prob(obesity for child c)} = Intercept + Trend × Year +
f(age group) + g(sex) + h(race/ethnicity) + k(income)

For the SEER cancer data, the SEER*Stat software was used to obtain total counts of cancer incidence or cancer mortality in the SEER registries, together with the associated total populations. Assume that the number of children in the demographic group G for year Y is Pop (G, Y) and that the number of cancer cases or cancer deaths is Cases (G, Y). We assume that for children in a demographic group G and year Y, the number of cancer cases or cancer deaths, Cases (G, Y), has a discrete distribution with a mean given by

Expected number of cases for demographic group G and year Y
= Pop(G, Y) × Rate (G, Y)

and has a variance given by

Variance of number of cases for demographic group G and year Y
= φ × Expected number of cases

The parameter φ is called the dispersion parameter. If the count data has a Poisson distribution, then the dispersion parameter equals 1 and so the variance equals the mean. In many cases, this type of count data is “over-dispersed” so that the dispersion parameter is greater than 1 and the variance exceeds the mean. For example, this would happen in this case if pairs of children in the same demographic group tended to share the propensity for cancer incidence or mortality, so that cancer incidence (or mortality) in different children is not statistically independent and is positively correlated. Under-dispersion ($\varphi < 1$) can also occur but is not very often found. This type of model is known as a Poisson regression model adjusted for dispersion.

We also assume that the logarithm of the rate is a linear function of the explanatory variables:

Log {Rate (G, Y)} = function of explanatory variables for group G and year Y

For example, the adjusted trend analysis model is of the form:

Log {Rate (G, Y)} = Intercept + Trend $\times$ Y +
f(age group) + g(sex) + h(race/ethnicity)

(For SEER, the demographic variables analyzed did not include income. Income data at the individual level are not included in SEER. SEER does include data on the median income for the county).

The Poisson regression statistical model for the SEER data also included the assumption that the counts of cancer cases in different demographic groups and/or different years are statistically independent. These assumptions are commonly made for modeling counts of rare events such as cancer incidence or mortality, and a Poisson model without dispersion[7] is used by the SEER*Stat software to compute standard errors and confidence intervals for the numbers of cases or deaths.

The SAS procedure GENMOD was used to fit the Poisson regression models adjusted for dispersion to the SEER cancer data. The Poisson regression model was fitted to the counts of cancer cases or deaths for each combination of calendar year, age group, sex, and race/ethnicity. The age group, sex, and race/ethnicity groups used were as follows:

Age groups

- 0-4 years
- 5-9 years
- 10-14 years
- 15-19 years

Sex groups

- Males
- Females

Race/ethnicity groups

[7] Since the SEER*Stat software estimates standard errors for a single count value instead of fitting a regression model to a set of count values, an adjustment for over- or under-dispersion cannot be calculated in that case. For the same reason, the relative standard errors used to flag unreliable estimated cancer incidence or mortality rates for the ACE childhood cancer indicators assume a Poisson model without a dispersion adjustment.

- White non-Hispanic
- Black non-Hispanic
- American Indian/Alaska Native, non-Hispanic
- Asian or Pacific Islander, non-Hispanic
- Hispanic

Although the cancer incidence and mortality data from SEER included a small number of cases for children that were not in any of the above race/ethnicity groups, the corresponding population numbers were all zero. For this reason, the cases in other race/ethnicity groups are not included in these statistical analyses.

As discussed above, the Poisson regression modeling included a dispersion parameter. Instead of assuming a Poisson distribution, the fitted distribution is not a specified probability distribution but is instead assumed to be any distribution that has the same mean and has a variance equal to φ multiplied by the mean. The dispersion parameter for each regression model was estimated as the deviance divided by its degrees of freedom. The deviance can be defined as twice the difference between the maximum achievable log likelihood and the log likelihood at the maximum likelihood estimates of the regression parameters when the dispersion parameter is exactly one. Compared to a Poisson regression model without a dispersion parameter, all the estimated parameters other than the dispersion parameter φ are exactly the same, but their estimated variances are all multiplied by φ and the standard errors are all multiplied by the square root of φ. Thus if the model is adjusted for over-dispersion, then all the p-values will be increased compared to the Poisson model. Without the adjustment for overdispersion, the Poisson model would be more likely to find false positives where a zero trend or difference in the population is wrongly determined to be statistically significant.

Of interest for these Poisson regression models is the distribution of the model residuals.Since the underlying model is a Poisson distribution for cancer case counts, normality of the residuals is not to be expected, but normality of the residuals should hold approximately when the counts are large. An analysis of the residuals for cancer incidence trends from 1992 to 2007 using a Shapiro-Wilk normality test at the 5% significance level, shows that the standardized Pearson residuals were not normally distributed for the unadjusted or adjusted trend model. This finding may suggest the need for alternative trend model formulations in this case. However, normality was not rejected for the standardized Pearson residuals from the unadjusted and adjusted models for demographic differences in cancer incidence over 2005 to 2007.

National Vital Statistics System (NVSS) Natality Data

The National Vital Statistics System (NVSS) includes data on adverse birth outcomes such as preterm and low birth weight births. To test for trends and demographic group differences, a logistic regression method was applied. The general approach is similar to the logistic regression method applied for the NHANES data on the percentages of obese or overweight children, except that instead of treating the data as complex survey data, we treat the adverse birth outcomes as a random sample of all births. Therefore, to avoid repetition, we will simply describe the general statistical model formulation and the demographic groups used in these analyses.

The logistic regression models used for the NHANES data on the percentages of obese children are of the general form:

Logit {Prob(obesity for child c)} = function of explanatory variables for child c

For example, the adjusted trend analysis model is of the form:

$$\text{Logit }\{\text{Prob(obesity for child c)}\} = \text{Intercept} + \text{Trend} \times \text{Year} + f(\text{age group}) + g(\text{sex}) + h(\text{race/ethnicity}) + k(\text{income})$$

For the NVSS birth outcomes data, similar logistic regression models are used, but the only two demographic variables used in the models were the mother's age group and race/ethnicity. The mother's gender is always female, of course, and the NVSS birth data does not include the mother's income.

Thus the unadjusted trend analysis model for adverse birth outcomes is of the form:

$$\text{Logit }\{\text{Prob(adverse birth outcome for child c)}\} = \text{Intercept} + \text{Trend} \times \text{Year}$$

and the adjusted trend analysis model for adverse birth outcomes is of the form:

$$\text{Logit }\{\text{Prob(adverse birth outcome for child c)}\} = \text{Intercept} + \text{Trend} \times \text{Year} + f(\text{age group}) + h(\text{race/ethnicity})$$

The SAS procedure GENMOD was used to fit the logistic regression models to the NVSS natality data. The age group and race/ethnicity groups used were as follows:

Age groups for mother:

- < 20 years
- 20-39 years
- 40 or more years

Race/ethnicity groups for mother:

- White non-Hispanic
- Black non-Hispanic
- American Indian/Alaska Native, non-Hispanic
- Asian or Pacific Islander, non-Hispanic
- Hispanic
- Unknown ethnicity

The logistic regression model was fitted to all combinations of year, mother's age group, and mother's race/ethnicity group. The binomial model used assumes that the numbers of adverse birth outcomes for different years and/or age groups and/or race/ethnicity groups are statistically independent. The logistic model was fitted to the binary response data for each individual birth (adverse or non-adverse birth outcome). The demographic group comparisons for age, race/ethnicity, and the contrast comparisons between pairs of race/ethnicity groups also used logistic regression. These analyses only use the data for the most recent year and therefore test for demographic differences in that year only.

The unadjusted logistic regression model for age differences is of the form:

$$\text{Logit }\{\text{Prob(adverse birth outcome for child c)}\} = \text{Intercept} + f(\text{age})$$

and the adjusted logistic regression model for age differences is of the form:

$$\text{Logit }\{\text{Prob(adverse birth outcome for child c)}\} = \text{Intercept} +$$

f(age) + h(race/ethnicity)

In both cases the test for age group differences tests if f(age) = 0.

The unadjusted logistic regression model for race/ethnicity differences is of the form:

Logit {Prob(adverse birth outcome for child c)} = Intercept + h(race/ethnicity)

and the adjusted logistic regression model for race/ethnicity differences is of the form:

Logit {Prob(adverse birth outcome for child c)} = Intercept +
f(age) + h(race/ethnicity)

In both cases the test for race/ethnicity group differences tests if h(race/ethnicity) = 0.

Conclusion

This memorandum has presented statistical methods to test for trends, year-to-year differences, and demographic group differences in various health indicator data. The NHAMCS and NHDS statistical methods use weighted regression models for the rates of emergency room visits or hospitalizations for asthma or other respiratory diseases. The NHANES statistical methods discussed in this memorandum use logistic regression models applied to the percentages of obese children, taking into account the complex survey design. The SEER statistical methods use Poisson regression models adjusted for dispersion for the numbers of cancer cases or deaths. The NVSS birth outcome statistical methods use logistic regression models applied to the rates of adverse birth outcomes, assuming a simple random sample of births. The unadjusted analyses are compared with adjusted analyses that account for possible confounding effects of other demographic variables, and to stratified analyses that account for possible interactions between the trends (or year-to-year changes) and the demographic group effects.

A reviewer suggested a multilevel modeling approach for these indicators. For example, a multilevel model could add random effects terms to the regression models described above in order to model the statistical effects of the location (e.g., region, state, or county) and/or the demographic variables. Using this advanced modeling approach, responses in the same location group, or for the same demographic group, would have the same random effect term and so would be correlated. This approach has several appealing characteristics. However, this approach is not feasible for ACE for several reasons. First, for the NHAMCS, NHDS, and NHANES indicators, statistical theory on how to account for the survey design and survey weights when fitting a multilevel model does not appear to have been fully developed in the literature in cases like these where the strata and PSUs are not the same as the levels of the multilevel model and/or are masked for confidentiality. The model-based statistical approach of ignoring the survey design when doing statistical modeling is inconsistent with the use of the survey design in the main ACE report to unbiasedly estimate the rates and standard errors. In principle, for the NHAMCS, NHDS and SEER indicators, it would be possible to include geographical information in the statistical model, either as a fixed effect, or as a random effect in a multilevel model. However, potential geographical differences in rates are not of primary interest for ACE and they would not be expected to change the p-values very much unless the rate depended upon the geographical region, after adjusting for the effects of age and other demographic variables in the models. Geographic data are not publicly released for NHANES, or for the NVSS in recent years. In principle, a multilevel model could also treat the demographic variables as random effects, but since these variables only have a few categories, we prefer the more standard approach of treating these variables as fixed effects only. Finally we note the high level of effort needed to fit suitable multilevel models to these data and then to update these models each year when new data become available.