

**New York State Department of Health**  
**Updated Guidance for Health Data Review and Analysis Relating to**  
**NYSDEC Environmental Justice Requirements for CP-29 and 6 NYCRR 487**  
(updated October 2014, revised 1-26-15, updated June 2017; links updated January 2021)

**Table of Contents**

Section I.

- A. Introduction
- B. Health Outcome Data Analysis and Environmental Review: Overview of the Method
- C. Demographics
- D. Impact Study Area and Comparison Areas
  - a. Selecting ZIP codes for the impact study area
  - b. Comparison areas
  - c. Selecting ZIP codes for the population density comparison area
  - d. Selecting ZIP codes for the surrounding ZIP codes comparison area
- E. Selection of Health Outcomes for Display
- F. Health Outcome Data Displays
- G. Data Comparisons
  - a. Asthma emergency department visits
  - b. Cancer incidence
  - c. Low birth weight
- H. Discussion and Conclusions

Section II. Generating the Table of Demographic Data

- A. Elements in the table
- B. Instructions for a single ZIP code tabulation area (ZCTA)
- C. Instructions for multiple ZCTAs
- D. County and regional comparison areas
- E. Summary table

Section III. Examples

- A. Example 1
  - a. Impact study area
  - b. Comparison areas
  - c. Table of demographic data
  - d. Asthma emergency department visits
  - e. Cancer incidence
  - f. Low birth weight
- B. Example 2
  - a. Impact study area
  - b. Comparison areas
  - c. Table of demographic data
  - d. Asthma emergency department visits
  - e. Cancer incidence
  - f. Low birth weight

Section IV. Calculating Confidence Intervals (including links to pre-programmed spreadsheets)

## SECTION I.

### A. Introduction

New York State Department of Environmental Conservation (NYSDEC) regulations relating to Environmental Justice analysis under 6 NYCRR 487.9(b)(4)(i) require that a comprehensive description of information on public health be included in applications for a Certificate of Environmental Compatibility and Public Need pursuant to Public Service Law Article 10. 6 NYCRR 487.9(c) indicates that in presenting the required health data, applicants shall use a protocol approved by NYSDEC and the New York State Department of Health (NYSDOH). This document contains the protocol to be followed.

This document should also be followed in fulfilling the requirements of NYSDEC [Commissioner Policy 29 Environmental Justice and Permitting \(CP-29\)](#). This document updates previous guidance entitled “Guidance for Health Outcome Data Review and Analysis Relating to NYSDEC Environmental Justice and Permitting” and the earlier [Report of the Health Outcome Data Work Group](#).

### B. Health Outcome Data Analysis and Environmental Review: Overview of the Method

NYSDEC CP-29 specifies that health outcome data review be conducted when it is determined that a proposed action may affect the community of concern, and an environmental justice area is included. 6 NYCRR Part 487.9 requires that an applicant prepare comprehensive demographic, economic, and physical descriptions of the impact study area and comparison areas, including descriptions of health data, with asthma and cancer being noted. (Note that in this document the term “impact study area” will be used for both the community of concern in CP-29 and the impact study area in 6 NYCRR Part 487.9). A review of health outcome data can provide information on the current health status of the study area. If the population of an impact study area has a higher rate of health-related outcomes than comparison areas, it may be more vulnerable to the effects of environmental exposures. A separate guidance document for [screening](#) an impact study area for the presence of an environmental justice area as mentioned in 6 NYCRR Part 487.5 is available at the NYSDOH website.

This document provides a method for comparing the health status of the study area with that of selected comparison areas. A more detailed discussion about the method and its development can be found in the [Report of the Health Outcome Data Work Group](#). For this method, the study area and comparison areas must be identified as one or more ZIP codes because health outcome data for small areas are tabulated at the NYSDOH website by ZIP codes. The method specifies tabular displays of demographic and health outcome data for the study area and at least four comparison areas. (The Work Group defined health outcome data as counts and rates of health-related events.) The rates of health outcomes in the study area are compared to rates in comparison areas by calculation of rate ratios and confidence intervals. The results of these analyses are summarized and discussed.

Although a display of health outcome data can provide information on the number and percentage of people in a community who have a variety of diseases or health outcomes, and on how this community compares in disease rates to other areas, it cannot tell us what is causing the disease in the community. Many factors influence the risk of disease including heredity, age, lifestyle factors such as smoking and diet, adequacy of nutrition, housing, health care, health behavior, and exposures to chemicals in environmental media. However, if the population of a community has low health status, it may be more vulnerable to the effects of environmental exposures.

### **C. Demographics**

A table of demographic data describing the populations of the study area and each comparison area must be created. Instructions for preparing this table are provided in **Section II**. This table must include information on age, sex, race/ethnicity, population density, and poverty. This information is available for ZIP Code Tabulation Areas (ZCTAs) defined by the U.S. Census Bureau. The elements to be shown in the demographic table and the specific Census files that contain the needed elements are described in Section II. Note that for these analyses the definition of a minority population included in NYSDEC's [CP-29](#) is used: a population that is identified or recognized by the U.S. Census Bureau as Hispanic, African American or Black, Asian and Pacific Islander, or American Indian. This definition includes both race and ethnicity.

The U.S. Census Bureau's five-year [American Community Survey](#) (ACS) was the source of information on poverty and income in the demographic tables. This survey is conducted on a sample of the nation's population each month. For the five-year ACS, five years of data are combined and averaged. For areas with small populations, such as some ZCTAs, there are wide margins of error around the estimates (which are included in the ACS data tables), sometimes leading to questions about the accuracy of the estimates. NYSDEC and NYSDOH are exploring other indicators of social vulnerability and social determinants of health, and in the future other measures may be used.

### **D. Impact Study Area and Comparison Areas**

- a. *Selecting ZIP codes for the impact study area.* Because the health outcome data at the NYSDOH public website are displayed for ZIP codes, the impact study area must be approximated by selecting ZIP Code Tabulation Areas (ZCTAs) defined by the U.S. Census Bureau. Overlay a map of ZIP codes onto a map of the study area using commercial mapping software. Some ZIP codes may clearly be within the study area. Other ZIP codes may be partially within the study area and partially outside the study area. For each of these latter ZIP codes, the proportion of the population that lies within the study area should be estimated by summing the population of the 2010 census blocks whose geographic centers (centroids) fall within the study area. ZIP codes with greater than 50% of the population within the study area should be included. (If necessary [e.g., in rural areas with large ZIP codes], you may adjust the criterion to 40-45% of the population.)

- b. *Comparison areas.* Use all of the following comparison areas. Areas 1-3 and 5 are included in the [Report of the Health Outcome Data Work Group](#) and in the previous guidance; area 4 was added in 6 NYCRR Part 487.
- 1) The county in which the study area is located. If the study area is located in more than one county, use each of those counties as a separate comparison area; there will be multiple county comparison areas.
  - 2) A large regional comparison area. If the study area is in New York City, use New York City. If the study area is located outside of New York City, use New York State excluding New York City.
  - 3) An area, composed of ZIP codes, with population density similar to that of the impact study area and located in the same general geographic area (e.g., county or contiguous counties). See method below.
  - 4) A comparison area composed of the ZIP codes that surround the study area. These ZIP codes will be contiguous with the study area.
  - 5) If specific alternative locations for the facility are being considered, each alternative location should be a comparison area composed of ZIP codes.
- c. *Selecting ZIP codes for the population density comparison area.* Population per square mile is one of the fields in the table of demographic data to be prepared for the impact study area and each comparison area (see Section II. Generating the table of demographic data). While accessing data for this table, you should also download area (land) and population for every ZIP code (ZCTA) in the county (or counties) containing the impact study area. Calculate the population per square mile for each ZIP code. Choose ZIP codes (ZCTAs) within the relevant county (or counties) that have population per square mile that is  $\pm 10\%$  of that of the impact study area. (If there are no ZIP codes in that range, you may have to use  $\pm 15\%$  or  $\pm 20\%$ .) A reasonable number of ZIP codes is six to eight, but in some areas there may be more or fewer.
- d. *Selecting ZIP codes for the surrounding ZIP codes comparison area.* Choose those ZIP codes that surround the impact study area ZIP codes.

## **E. Selection of Health Outcomes for Display**

The following health outcome data should be used in the data display and analyses since they are currently available as counts and rates at the ZIP code level on the NYSDOH web site. (These data meet the criteria discussed in the [Report of the Health Outcome Data Work Group](#).) The most recent data available are posted on the web site and should be used. Since the last guidance document was issued, data on low birth weight and asthma emergency department visits have become available by ZIP code. It is recommended to use emergency department visit data rather than asthma hospital discharge data because asthma emergency department visits are more

frequent than asthma hospital discharges, leading to potentially more stable rates for statistical testing. It is anticipated that additional health outcome data at the ZIP code level will become available in the future and will be added to the list.

1. Emergency department visits for asthma from SPARCS available at [New York State Asthma Dashboard \(ny.gov\)](https://www.health.ny.gov/statistics/asthma/emergency_department_visits/) (find sub-county data links under data views column) Asthma Emergency Department Visits in New York State by County and ZIP code); includes admission to the hospital directly from the emergency department with admitting diagnosis of asthma (ICD-9-CM code 493) and outpatient emergency department visits with principal diagnosis of asthma (ICD-9-CM code 493)
2. Cancer incidence for male and female colorectal, female breast, male and female lung and bronchus, and male prostate cancer from the New York State Cancer Registry available at <http://www.health.ny.gov/statistics/cancer/registry/zipcode/index.htm>
3. Low birth weight from the New York State County/ZIP Code Perinatal Data Profile available at <http://www.health.ny.gov/statistics/chac/perinatal/index.htm>

#### **F. Health Outcome Data Displays**

The tables in the examples in **Section III** provide examples of health outcome data displays with the most recently available data. Use the age and gender groupings shown for each health outcome when creating health outcome data tables.

When data are divided into many categories, the number of events in each cell of the table becomes smaller. In highly populated areas or when the health outcome is relatively common, this may not present a problem. However, if the outcome is rare or the population is small, the number of events in some cells of the table may become so small that rates based on these events may be unstable and fluctuate dramatically from year to year. A discussion of the issue can be found at the NYSDOH web site at [Rates Based on Small Numbers](https://www.health.ny.gov/statistics/cancer/registry/zipcode/index.htm). A graph is included which shows that with 20 cases, the relative standard error (a measure of statistical uncertainty or random variability around the rate estimate) is 20% and with 10 cases it is 30%. One should be cautious about drawing conclusions from rates based on small numbers of events because, due to the influence of random variability, they may not reliably represent the underlying risk of disease.

These general guidelines should be followed when displaying data:

- Do not display information in a way that could identify an individual and constitute a breach of confidentiality. A count of one or two events in a small population could unintentionally disclose confidential information. For example, if there is one person in a small community who is frequently hospitalized, a table that shows one case of a serious illness, such as cancer, in this community could inadvertently reveal confidential information.

- Avoid the presentation of rates for cells with 10 or fewer events. Cells with 10 or fewer events should be displayed, but rates based on 10 or fewer cases may be unstable and should be interpreted cautiously. When possible, combine the number of cases over several years so that the rates are based on a larger number of cases. For example, use three-year or five-year average annual rates instead of single-year rates. It may also be possible to combine the number of cases across geographic areas to obtain a larger number of cases, for example, by combining ZIP codes or using the rate for a county; however, in the process the ability to characterize the health status of the specific geographic area of interest may be lost. If rates based on 10 or fewer cases are displayed, highlight these rates with a footnote stating that these rates may be unstable.
- If there are enough cases that small numbers are not an issue, the data can be displayed in sub-categories as appropriate:
  - If the outcome has differences in rates between males and females, the data should be displayed separately for males and females, e.g., cancer.
  - If there are known differences in rates by specific age groups, these age groups should be used when displaying the data, e.g., asthma rates are higher in children 0-4 years of age.
  - If reliable data are available by race/ethnicity and there are known differences in rates by race/ethnicity, the data should be displayed by race/ethnicity.
- The categories used to display data for specific outcomes at the New York State Department of Health web site (<http://www.health.ny.gov>) or the U.S. Centers for Disease Control and Prevention web site (<http://www.cdc.gov>) may be used as a guide.

## G. Data Comparisons

In accordance with the recommendations of the [Report of the Health Outcome Data Work Group](#), rate ratios and confidence intervals are calculated to compare disease rates between the impact study area and comparison areas. The rate ratio is the ratio of the rate in the impact study area to the rate in the comparison area. When the two rates being compared are the same, then the ratio is equal to 1.0. When the rate is higher in the impact study area than in the comparison area, the ratio is greater than 1.0, and when the rate is lower in the impact study area than in the comparison area, the ratio is less than 1.0.

When the two rates are compared, a confidence interval can be calculated to gain a sense of certainty about the estimated difference between the rates. The confidence interval is the range around the ratio in which the true measurement lies with a certain degree of confidence. The confidence interval is a measure of the variability in the data; in this case variability is contributed by the two rates in the ratio. One reason for variability is because there are random fluctuations in the number of cases in an area over time or between different communities. If there were little variability (i.e., the rates were relatively stable), the value of the ratio would be close to the same if the measurement were repeated.

Confidence intervals can be calculated using commercial statistical software or spreadsheet software. **Section IV** contains the formulas for calculating confidence intervals and a link to pre-

programmed spreadsheets for calculating confidence intervals when working with the asthma, low birth weight, and cancer data described in this document.

A confidence interval that does not include the number 1 provides an indication that the difference between the rates being compared is not likely to be due to the random-like variability mentioned above. If the confidence interval includes the number 1, then the difference in the two rates is likely to be due to random variability. (More information on confidence intervals can be found at the web sites

<http://www.doh.wa.gov/DataandStatisticalReports/DataSystems/HealthyYouthSurvey/TechnicalNotes/ConfidenceIntervals> and <http://www.health.ny.gov/diseases/chronic/confint.htm> and in Appendix F of the 2006 [Report of the Health Outcome Data Work Group](#)).

The use of rate ratios and confidence intervals is best shown in the examples in **Section III**; the impact study area in Example 1 is in an urban area, and the impact study area in Example 2 is in a rural area. In the examples, comparison areas for the impact study area are selected, and tables for asthma emergency department visits, four types of cancer, and low birth weight in the impact study area and the comparison areas are developed. For asthma and low birth weight, rates in the impact study area are compared to those in the comparison areas. For cancer, a different type of comparison is conducted.

The following sections describe the methods for displaying and comparing data for asthma emergency department visits, cancer incidence, and low birth weight.

*a. Asthma emergency department visits.*

Impact study area and ZIP code comparison areas. ZIP code data for asthma ED visits are collected as part of SPARCS and are available at [New York State Asthma Dashboard \(ny.gov\)](#) (find sub-county data links under data views column). The asthma ED visit data include patients admitted to the hospital from the ED and patients seen in the ED but not hospitalized. At the DOH website, data are presented for age groups and the total population of the ZIP code. Use the following age groups: 0-4 years, 0-17 years, 18-64 years, 65+ years, and total population. The sum of the last three age groups is the **total population**; the 0-4 years age group is included because asthma ED visit rates are higher in this age group, particularly in low income and minority areas. The DOH data tables display, for each ZIP code, the number of asthma ED visits during a three-year period and the average annual rate of asthma ED visits per 10,000 population per year. Select these fields for each ZIP code in the impact study area and in each ZIP code comparison area and copy data to a spreadsheet or database. For each ZIP code calculate the average annual population for each age group and the total population as follows and create a column for average annual population:

$$\text{Average annual population} = \frac{\text{no. of asthma ED visits over 3 years}/3}{\text{asthma ED visit rate}} \times 10,000$$

For the impact study area and each ZIP code comparison area, calculate the number of ED visits and the population for each age group and the total population by summing the data for the

individual ZIP codes. For the impact study area and each ZIP code comparison area, calculate the average annual rate of ED visits per 10,000 population for each age group and the total population as follows:

$\text{Average annual rate of asthma ED visits per 10,000 population} = \frac{\text{no. of asthma ED visits over 3 years}}{\text{average population}} \times 10,000$
--

The rate ratio is the rate in the impact study area divided by the rate in the comparison area. Calculate confidence intervals using the method described in **Section IV**, using the appropriate spreadsheet provided.

County and regional comparison areas. For the county comparison area, do not use the county total in the asthma ZIP code data tables because data may not be displayed for ZIP codes with small numbers of cases and there may be problems with ZIP codes that cross county lines. Access data tables on Asthma Emergency Department Visits by region and county at New York State Asthma Dashboard (ny.gov) (find sub-county data links under data views column) for the county or counties of interest. Use the following age groups: 0-4 years, 0-17 years, 18-64 years, 65+ years, and [total population](#). The DOH data tables display, for each county, the number of asthma ED visits for each of three years and the three-year total, the average population for the three years, and the average annual rate of asthma ED visits per 10,000 population per year. Select and copy data for the relevant counties to your spreadsheet or database.

If New York State is the regional comparison area, select and copy the data for New York State, which is the last line of the region/county table. If New York City is the regional comparison area, select the Region 7 New York City, Region Total, and copy the data from the table. If New York State exclusive of New York City is the regional comparison area, you will have to deduct the number of ED visits in New York City from the number of ED visits in New York State to calculate the number of ED visits in New York State exclusive of New York City. Perform a similar calculation to get the average population for the three-year period for New York State exclusive of New York City. Calculate the average annual rate of asthma ED visits per 10,000 population per year for New York State exclusive of New York City for each age group and the total population as follows:

$\text{Average annual rate of asthma ED visits per 10,000 population for New York State exclusive of New York City} = \frac{\text{no. of asthma ED visits over 3 years}}{\text{average population}} \times 10,000$
--

The rate ratio is the rate in the impact study area divided by the rate in the comparison area. Calculate confidence intervals using the method described in **Section IV**, using the appropriate spreadsheet provided.

***b. Cancer incidence.***

Tables showing cancer incidence by ZIP code for female breast, lung, colorectal, and prostate cancers are available at the NYSDOH web site <http://www.health.ny.gov/statistics/cancer/registry/zipcode/index.htm>. These data are tabulated in a different way from the asthma emergency department visit data discussed in the previous section. With the asthma data, the data are presented by age group, and the number of visits and rate are provided. For the cancer data, the number of cases of the specific type of cancer in the total population of the ZIP code (all ages) is presented as the observed number. For each ZIP code an expected number of cases of all ages is shown. The cancer rate for the entire state of New York and the number of people in a ZIP code are used to estimate the number of people in each ZIP code that would be expected to develop cancer within a specified five-year period if the ZIP code had the same rate of cancer as the state (for more information, see “Frequently Asked Questions” at <http://www.health.ny.gov/statistics/cancer/registry/zipcode/faq.htm>). Age and population size are taken into consideration when determining the expected number; this process is called [age-adjustment](#).

For some ZIP codes in the cancer tables, there were too few cases to be shown for confidentiality reasons. These ZIP codes are combined with neighboring ZIP codes, and data are provided for the combined groups of ZIP codes. In sparsely populated areas, ZIP codes are frequently combined. If a ZIP code in the study area or a comparison area has been combined with other ZIP codes, include these additional ZIP codes in the study area or comparison area and make note in the text and table notes of these additional ZIP codes.

Because of the way the cancer data are presented, the cancer data for the impact study area cannot be compared directly to the cancer data for the comparison areas; instead, since the expected number of cases is based on the cancer rate for New York State, the state is the comparison area for the impact study area and for the three areas that are referred to as comparison areas. From the cancer incidence by ZIP code table, for each site of cancer, copy to your spreadsheet the number of cases observed and the number of cases expected for each ZIP code in the impact study area and ZIP code comparison areas. Calculate the total numbers of cases observed and total number of cases expected for the impact study area, the population density ZIP code comparison area, and the surrounding ZIP codes comparison area. Calculate the standardized incidence ratio (SIR), which is the ratio of the number of cases observed to the number of cases expected. Calculate confidence intervals using the method described in **Section IV**, using the appropriate spreadsheet provided.

To obtain the observed and expected numbers for the county, you must sum the observed and expected numbers for the ZIP codes. If there is a notation that one or more ZIP codes cross county boundaries, the ZIP code(s) should be included in your county total. However, if you are summing ZIP codes to create a New York City comparison area (five counties), you must be careful to count a ZIP code that crosses county lines only once in the New York City total. (It will be listed in more than one county.) The same will be true when summing ZIP codes to create a comparison area that is New York State excluding New York City. For assistance with summing ZIP codes to create these large comparison areas, contact DOH at [beoe@health.ny.gov](mailto:beoe@health.ny.gov) (reference EJ health data review and facility name).

**c. Low birth weight (LBW)**

Impact study area and ZIP code comparison areas. The percent of LBW births during a three-year period by ZIP code is included in Perinatal Data Profiles at the DOH website <http://www.health.ny.gov/statistics/chac/perinatal/index.htm>. Some ZIP codes are not included in the table because there are fewer than 10 births in the three-year period. If a ZIP code in the study area or in a comparison area is not included in the table, leave this ZIP code out, noting the exclusion in the text and table notes, and use the data for the other ZIP codes.

The Perinatal Data Profiles also include the total number of births in each ZIP code during the three-year period, allowing for the calculation of the number of LBW births in each ZIP code. Using this information, ZIP codes can be combined and the percentage of LBW births can be calculated for the impact study area and each ZIP code comparison area. A rate ratio can then be calculated for each comparison.

From the Perinatal Data Profile for each relevant county, copy to a spreadsheet the number of total births and the percentage of LBW births for each ZIP code in the impact study area and in each ZIP code comparison area. From these two numbers, calculate the number of LBW births in each ZIP code, and round to the nearest whole number:

$$\text{No. LBW births} = \frac{\text{Percent LBW} \times \text{number of births}}{100}$$

Calculate the total number of LBW births and total number of births in the impact study area and in each ZIP code comparison area by summing data for the relevant ZIP codes. Calculate the percent LBW in the impact study area and each ZIP code comparison area by dividing the number of LBW births by the number of births, and multiplying by 100. Calculate the rate ratio for each ZIP code comparison area by dividing the percent LBW in the impact study area by the percent LBW in the comparison area. Calculate confidence intervals using the method described in **Section IV**, using the appropriate spreadsheet provided.

County comparison area. The last row of each County Perinatal Data Profile provides the number of total births and percentage of LBW births for the county; from these data calculate the number of LBW births as described above. Calculate the rate ratio and the 95% confidence intervals as described above.

Regional comparison area. For New York City and New York State exclusive of New York City, the number of births during a three-year period and percent LBW births can be found in the New York State Regional Perinatal Data Profile at <http://www.health.ny.gov/statistics/chac/perinatal/index.htm>. Calculate the rate ratio by dividing the percent LBW in the impact study by the percent LBW in the relevant regional comparison area. Calculate confidence intervals using the method described in **Section IV**, using the appropriate spreadsheet provided.

## H. Discussion and Conclusions

When evaluating the results of the analyses, we focus on the rate ratios (greater than 1.0 or less than 1.0), the number of comparisons for which rate ratios are greater than 1.0, and the confidence intervals around the rate ratios (whether or not 1.0 is excluded). We also look at the number of health outcomes for which we see higher rates in the study area and consider the specific types of health outcomes that show elevations.

The 2006 [Report of the Health Outcome Data Work Group](#) indicates that the health outcome data are to be considered as part of the permitting process, recognizing that the data provide no information about the causes of any increase or decrease in rates between the study area and comparison area populations. The more often the observations fall into the same pattern, the greater the likelihood that the observations suggest a real difference in health status between the study area and comparison area populations. The report states that, if any of the following conditions are met, consideration of additional options for the permitting conditions should be reviewed as part of the permitting process because of the health outcome data displays and comparisons. The greater the number of conditions that are met, the greater the likelihood is that the health status of the study area is actually lower than that found in other areas.

- 1) A disease rate is higher in the study area than in any comparison area population for any health outcome;
- 2) A disease rate is higher in the study area than in multiple comparison area populations for any health outcome;
- 3) The confidence intervals are greater than 1.0 (A greater elevation in health outcome in the study area compared to the comparison area population and a larger number of events will increase the likelihood that the confidence interval will not include 1.0.);
- 4) There is a pattern of higher rates of multiple health outcomes in the study area; and
- 5) Health outcomes that result from an acute exposure (e.g., asthma exacerbations) are elevated rather than those that result from a chronic exposure (e.g., cancer). Health outcomes resulting from an acute exposure may be more relevant to current place of residence than those that result from a chronic exposure. For a chronic effect such as cancer, a crucial exposure or risk factor may have occurred decades earlier when the individual resided at a different location.

The information on the health status of the community provides a more complete picture of the area under study and may suggest the need for action, including more stringent permit conditions, voluntary pollution reduction, or other corrective measures. Therefore, the health outcome data displays and comparisons should be used in making a permitting decision along with other considerations such as regulatory standards, environmental impacts, mitigation, benefits, needs, and costs. The significance of the difference between the study area and the comparison area populations should be considered in determining which action is appropriate. Possible actions may include, but are not limited to:

- Considering alternative siting locations, especially outside of the study area;

- Applicant evaluation and implementation of pollution prevention options, such as chemical substitution; changes in work practice standards, such as evaluating ways to reduce fugitive emissions; emission reductions by the installation of best available control technology to achieve the lowest achievable emission rates (LAER) possible; the implementation of holistic environmental management system; or the purchase of emission reduction credits.
- Providing assistance beyond the permit review process by using agency regulatory authority to evaluate the feasibility of reducing existing exposures from other sources that may be contributing to a health outcome of concern, such as obtaining offsets from other sources of air pollution in the study area;
- Providing assistance that may be beyond the permit review process by using agency regulatory authority to prevent, diagnose, monitor or treat the health outcome of concern or improve health status of the study area; and
- Taking no action based on the results of the health outcome comparisons.

Outside of these recommendations, there are state and federal environmental justice grant programs for community groups that could enter into partnerships with government, business, and the academic sector to work on projects to address local environmental and/or public health concerns (see <http://www.dec.ny.gov/public/31226.html> and <https://www.epa.gov/environmentaljustice/environmental-justice-grants-funding-and-technical-assistance>). Regardless of which options are considered, the local and state health departments may be able to identify services that might improve the health status of the study area.

## **SECTION II. Generating the Table of Demographic Data**

***ELEMENTS IN THE TABLE*** (see sample demographic tables in examples in **Section III** and summary table at the end of this section)

- Total population

- Population per square mile

- Number of households

- Sex: percent male and percent female

- Age distribution (%): less than five, 5-14, 15-19, 20-44, 45-64, and 65+ years

- Race/ethnicity (%):

  - One race

    - White

    - Black or African-American

    - American Indian/Alaska Native

    - Asian

    - Hawaiian/Pacific Islander

    - Some other race

  - Two or more races

  - Hispanic or Latino (of any race) (%)

  - Total minority (%)

- Income

  - Median household income (\$)

  - Persons below poverty (%)

### ***INSTRUCTIONS FOR A SINGLE ZIP CODE TABULATION AREA (ZCTA)***

**Step 1:** Go to the U.S. Census website at <https://www.census.gov/> . On the top menu bar, select Data > Data Tools and Apps > American FactFinder

**Step 2:** Use Community Facts and enter the ZCTA in the search box.

Under Population (2010 Census), select General Population and Housing Characteristics (Population, Age, Sex, Race, Households and Housing...). From this table you should acquire:

Total population

Sex: percent male and percent female

Age distribution (%): **combine age groups** so that you have these groups – less than five, 5-14, 15-19, 20- 44, 45-64, and 65+ years

Race/ethnicity (%):

One race

White

Black or African-American

American Indian/Alaska Native

Asian

Hawaiian/Pacific Islander

Some other race

Two or more races

Hispanic or Latino (of any race) (%)

Total minority (%): calculate by subtracting the percentage of “Not Hispanic or Latino-White Alone” from 100%

**Step 3:** Return to Community Facts (if you have to, enter the ZIP code again). Under Income, select American Community Survey, **Income in the Past 12 Months (Households, Families...)** (table S1901 ACS 5-year estimates). Find Total (number of households) and Median Income (dollars) – Households

**Step 4:** Return to Community Facts (if you have to, enter the ZIP code again). Under Poverty, select American Community Survey, **Poverty Status in the Past 12 Months (Age, Sex, Race, Education, Employment...)** (table 1701 ACS 5-year estimates). Find Population for whom poverty status is determined – percent below poverty level.

**Step 5:** To calculate population per square mile, you will acquire ZCTA land areas. Go to American FactFinder >**Advanced Search**, click “show me all”

**Step 6:** In the box for “Topic or Table name,” enter G001

Select the table listed below and click View

Table	ID	Dataset
<i>Geographic Identifiers</i>	G001	<i>2010 SF1 100% Data</i>

**Step 7:** Click the **Geographies** tab on the left

Under **Geographic type:** select 5 Digit ZIP Code Tabulation Area

Under **State:** select NY

Under **One or more geographic areas:** select the ZCTA of interest, click “add to your selections,” and close the box

**Step 8:** Click View. Find the row for Area (land) under AREA CHARACTERISTICS. This will give you the area in square meters. To convert it to square miles, **divide by 2,589,988**. Divide the Total population of the ZCTA (acquired previously) by the area to get population per square mile.

### ***INSTRUCTIONS FOR MULTIPLE ZCTAs***

Demographic information on population age, race, sex

**Step 1:** Go to the U.S. Census website at <https://www.census.gov/>. On the top menu bar, select Data > Data Tools and Apps > American FactFinder. Select Guided Search.

**Step 2:** Select Get Me Started and choose **I'm looking for information about people**

**Step 3:** From **Topics** Select >Basic Counts/Estimates > Population Total

**Step 4:** From **Geographies**, use the second option.

Under **Geographic type:** select 5 Digit ZIP Code Tabulation Area

Under **State:** select NY

Under One or more geographic areas: Select the ZCTAs of interest (use <Ctrl> + <Caps Lock> to select multiple ZCTAs)

Click **"Add to Your Selections,"** then **"Next"**

**Step 5:** Click **"Skip this step"**

**Step 6:** Select the table listed below and click view

Table	ID	Dataset
<i>Profile of General Population and Housing Chars:2010</i>	DP-1	2010 SF1 100% Data

**Step 7: To download the data**

Click Modify table tab and uncheck the Percent box at the top of the table

Click on the Download tab

Select Comma delimited (.csv) format

- Data and annotations in a single file
- ✓ Include descriptive data element names

The file will be downloaded as .zip file. Unzip the file and open the file named

"DEC\_10\_SF1\_SF1DP1\_with\_ann.csv" in MS Excel. Rename and save as an MS Excel file since the census download reuses the same name for subsequent downloads.

**Step 8:** The columns of interest can now be summarized to tabulate the total population, number of households, and the percentages of the age/race/sex groups. To calculate total minority (%), subtract the percentage of "Not Hispanic or Latino-White Alone" from 100%.

### Information on poverty and median household income

**Step 1:** Go to American FactFinder> **Advanced Search**, click “show me all”

**Step 2:** In the box for “Topic or Table name,” enter both S1701 and S1901, click Go. From the list, select the following tables/datasets

Table	ID	Dataset
Poverty Status in Past 12 Months	S1701	<i>2011 ACS 5-Year Estimates</i>
Income in the past 12 Months	S1901	<i>2011 ACS 5-Year Estimates</i>

**Step 3:** Click the **Geographies** tab on the left

Under **Geographic type:** select 5 Digit ZIP Code Tabulation Area

Under **State:** select NY

Under **One or more geographic areas:** select the ZCTAs of interest (use <Ctrl> + <Caps Lock> to select multiple ZCTAs)

Click **Add to Selections**, then close this box

The tables and ZCTAs should be showing up under the “Your Selections” box in the upper left-hand corner.

**Step 4:** Select the check boxes next to each of the two tables listed in Step 2 and select View from the tabs on top

**Step 5:** For the Poverty table

- Select the modify table tab on top and uncheck all Margin of error and Percent below poverty estimate columns
- You should end up with two columns under each ZCTA – One with the “Total” Population and one with the “Below poverty level” population
- You are interested in the line “Population for whom poverty status is determined”
- At this point you can copy and paste the column headings and data into a spreadsheet or you can go through the download procedure outlined above

**Step 6:** For the Household income

- In the same table viewer Select result 2 of 2 (Arrows on top right of Viewer)
- Select the modify table tab on top and uncheck columns except for household estimate (the first column) for each ZCTA
- You can also unselect some rows so you just have “Total” (number of households) and “Median Income - households”
- You should end up with one columns under each ZCTA with the number “Households” and median income
- At this point you can copy and paste the column headings and data into a spreadsheet or you can go through the download procedure outlined above
- To calculate percent below poverty of a group of ZCTAs, sum the Population below poverty level for the individual ZCTAs and divide by the Total population for which poverty was determined for the group of ZCTAs (multiply by 100)
- To calculate the median household income of a group of ZCTAs, take a weighted average of the median household income of each ZCTA weighted by the number of households in the ZCTA (the weights are the number of households in a ZCTA divided by the total number of households in all ZCTAs – these weights should add up to 1.0).

Calculation of population/square mile

**Step 1:** Go to American FactFinder> **Advanced Search**, click “show me all”

**Step 2:** In the box for “Topic or Table name,” enter G001  
Select the table listed below and click view

Table	ID	Dataset
<i>Geographic Identifiers</i>	G001	<i>2010 SF1 100% Data</i>

**Step 3:** Click the **Geographies** tab on the left  
Under **Geographic type**: select 5 Digit ZIP Code Tabulation Area  
Under **State**: select NY  
Under **One or more geographic areas**: select the ZCTAs of interest (use <Ctrl> + <Caps Lock> to select multiple ZCTAs)  
Click “**Add to Selections**,” then close this box

**Step 3:** Click View. Find the row for Area (land) under AREA CHARACTERISTICS. This will give you the area in square meters. Copy and paste the appropriate rows or download as described above. (The population per square mile will be used individually for each ZCTA in the county to determine the ones most similar to the impact study area. They will also be combined to determine the overall population per square mile of the comparison areas.)

**Step 4:** To get the area per square mile, divide the area per square meter **by 2,589,988**. To calculate the population/square mile of a group of ZCTAs, find the sum of the populations of the individual ZCTAs to find the Total population. Sum the area in square miles of the individual ZCTAs to get the Total area in square miles. Divide the Total population by the total area to get the Total population per square mile.

***COUNTY AND REGIONAL COMPARISON AREAS***

The same process can be followed to obtain demographic data for the county and regional comparison areas, except that the geographic unit will be the county. For New York City or New York State exclusive of New York City, counties will have to be combined.

**SUMMARY TABLE:** additional notes for generating the table of demographic data.

Element	Source (U.S. Census American FactFinder)	Notes
Total population	Table DP-1 (SF-1)	
Population per square mile	Table G001 (SF-1)	To convert square meters to square miles divide by 2,589,988
Number of households	Table S1901 - 2011 ACS 5-Year Estimates	
Sex (%)	Table DP-1 (SF-1)	
Age distribution (%)	Table DP-1 (SF-1)	Combine age groups to display six groups: less than 5, 5-14, 15-19, 20-44, 45-64, 65+
Race/ethnicity (%)	Table DP-1 (SF-1) for one race (White, African-American, American Indian/Alaskan, Asian, Hawaiian/Pacific Islander, some other race), two or more races, and Hispanic or Latino (of any race).	Show percent one race (subdivided into White, African-American, American Indian/Alaskan, Asian, Hawaiian/Pacific Islander, and some other race) and two or more races (this will yield 100%, except for rounding). Show percent Hispanic or Latino (of any race). Calculate percent minority by subtracting the percentage of “Not Hispanic or Latino-White Alone” from 100%.
Median household income (\$)	Table S1901 - 2011 ACS 5-Year Estimates	For multiple ZCTAs - weight the median household income of each ZCTA by the # of households in the ZCTA/ total # of households in all ZCTAs
Persons below poverty (%)	Table S1701 - 2011 ACS 5-Year Estimates	Use the “Population for whom poverty status is determined” field

## SECTION III. Examples

### A. Example 1

- a. *Impact study area.*** The impact study area chosen for this example is ZIP code 11105, which is located in Astoria in Queens County. This ZIP code includes areas identified by NYSDEC as Potential Environmental Justice Areas (see [CP-29](#) and [County Maps Showing Potential Environmental Justice Areas](#)). This impact study area is in New York City.
- b. *Comparison areas.*** Four comparison areas were selected to evaluate the disease rates in the impact study area in the context of a number of different settings (*see Section I.D Impact Study Area and Comparison Areas*).
1. A comparison area made up of ZIP codes in Queens County that are similar to ZIP code 11105 in population density. This area may be similar to ZIP code 11105 in land use and urban/rural characteristics.
  2. A comparison area that is composed of the ZIP codes surrounding ZIP code 11105.
  3. Queens County, which is the county that contains ZIP code 11105. This is a larger area and represents average health status, but it is smaller than New York City and will be familiar to the community because it is in close proximity to the study area.
  4. New York City is a large area chosen to represent the average health status.

Two of these comparison areas are made up of ZIP codes. The ZIP codes for the population density comparison area were chosen by the method described in Section II.D Impact Study Area and Comparison Areas: ZIP codes 11412, 11418, 11420, 11423, 11428, 11429, and 11436. The following ZIP codes are in the surrounding ZIP codes comparison area: 11102, 11103, 11370, and 11377. The other two comparison areas are Queens County and New York City.

- c. *Table of demographic data.*** **Table A1** displays demographic information for the impact study area and the four comparison areas (*Section II. Generating the Table of Demographic Data*). The table at the end of Section II indicates the specific Census files to be used to complete the table. Note that the definition of a minority population included in NYSDEC's [CP-29](#) is used: a population that is identified or recognized by the U.S. Census Bureau as Hispanic, African American or Black, Asian and Pacific Islander, or American Indian. This definition includes both race and ethnicity.
- d. *Asthma emergency department (ED) visits.*** Instructions for display and analysis of asthma ED visits are included in Section I.G. **Table A2** displays asthma ED visit data for 2010-2012 for four age groups and the total population (0-65+ years) for ZIP code 11105 (part 1) and four comparison areas (parts 2-5). Parts 2-5 also provide rate ratios and 95% confidence intervals (CI). The rate ratio compares the rate in ZIP code 11105 to that in the specified comparison area. In the last row of Table A2, showing the data for all age groups combined, all rate ratios are less than 1.0, indicating that the asthma ED visit rate is lower in the study area than in each comparison area. For some age group comparisons,

rate ratios are greater than 1 (ZIP codes surrounding ZIP 11105, 65+ years; Queens County, 65+ years). However, the CIs include 1.0, indicating that the difference between the rates may be due to random fluctuation.

- e. Cancer incidence.* Instructions for display and analysis of cancer incidence data are included in Section I.G. Data are for the years 2005-2009. **Table A3** displays incidence of breast cancer in females, colorectal cancer in males and females, lung cancer in males and females, and prostate cancer in males for the impact study area ZIP code 11105, the two ZIP code comparison areas, and Queens County. The comparison for cancer data is different from that used for asthma ED visit data in Table A2 in that the cancer data for ZIP code 11105 are not compared directly to the cancer data for the comparison areas; instead the number of observed cancer cases in ZIP code 11105 is being compared to the expected number of cases for ZIP code 11105, which is calculated based on the cancer rate for New York State. Similarly, the number of observed cases in each comparison area is compared to the number of cases expected in that comparison area, which is calculated based on the cancer rate for New York State. In this table, the ratio represents the ratio of the observed number of cases to the expected number of cases. Because the data have been age adjusted, the ratios are standardized incidence ratios (SIRs). (The SIRs in the table should not be compared with each other because of the different age distributions of the populations of each area.)

The SIRs in Table A3 for the study area (ZIP code 11105) show that the SIRs for all the cancer sites are less than 1.0, indicating that the number of cases observed was lower than the number expected based on cancer rates in New York State. Some of the CIs exclude 1.0, and others include 1.0, indicating that some of the differences may be due to random variation. When the SIRs for the other areas (ZIP codes with similar population density to ZIP code 11105, ZIP codes surrounding ZIP code 11105, and Queens County) are examined, almost all of the SIRs are less than 1.0 (or close to 1.0) and the CIs include 1.0. For prostate cancer in ZIP codes with similar population density to ZIP code 11105 (shown in bold type), the SIR is greater than 1.0 and the CI excludes 1.0, indicating that the number of prostate cancer cases observed in this area was greater than the number expected based on the prostate cancer rate in New York State, and the difference is not likely due to chance.

- f. Low birth weight (LBW).* Instructions for display and analysis of LBW data are in Section I.G. **Table A4** shows the LBW data for 2009-2011 for ZIP code 11105 and the four comparison areas. The rate ratio compares the rate of LBW births in ZIP code 11105 to that in the specified comparison area. All of the rate ratios are less than or close to 1.0. The CIs include 1.0, indicating that the differences could be due to random variation.

## **B. Example 2**

- a. Impact study area.* For the second example, an impact study area in Allegany County in western New York State was chosen. The study area is composed of ZIP codes 14735 and 14744. These ZIP codes includes areas identified by NYSDEC as Potential

Environmental Justice Areas (see [CP-29](#) and [County Maps Showing Potential Environmental Justice Areas](#)). This impact study area is outside of New York City.

- b. Comparison areas.** Four comparison areas were selected to have enough information to evaluate the disease rates in the study area in the context of a number of different settings (see *Section I.D Impact Study Area and Comparison Areas*).
1. A comparison area made up of ZIP codes in Allegany County and contiguous counties that are similar to ZIP codes 14735 and 14744 in population density. This area may be similar to ZIP codes 14735 and 14744 in land use and urban/rural characteristics.
  2. A comparison area that is composed of ZIP codes surrounding ZIP codes 14735 and 14744.
  3. Allegany County, which is the county that contains ZIP codes 14735 and 14744. This is a larger area and represents average health status, but is a smaller area than New York State excluding New York City and will be familiar to the community because it is in close proximity to the study area.
  4. New York State excluding New York City is a large area chosen to represent the average health status.

Two of these comparison areas are made up of ZIP codes. The ZIP codes for the population density comparison area were chosen by the method described in *Section II.D Impact Study Area and Comparison Areas*: ZIP codes 14065, 14711, 14737, and 14777. The following ZIP codes are in the surrounding ZIP codes comparison area: 14024, 14060, 14065, 14066, 14536, 14709, 14711, 14717, 14777, 14836, and 14846. The other two comparison areas are Allegany County and New York State excluding New York City.

- c. Table of demographic data.** **Table B1** displays demographic information for the impact study area and the four comparison areas (*Section II. Generating the Table of Demographic Data*). The table at the end of *Section II* indicates the specific Census files to be used to complete the table. Note that the definition of a minority population included in NYSDEC's [CP-29](#) is used: a population that is identified or recognized by the U.S. Census Bureau as Hispanic, African American or Black, Asian and Pacific Islander, or American Indian. This definition includes both race and ethnicity.
- d. Asthma emergency department (ED) visit data.** Instructions for display and analysis of asthma ED visits are included in *Section I.G*. **Table B2** displays asthma ED visit data for 2009-2011 for the impact study area and the four comparison areas. ZIP codes 14735 and 14744 are sparsely populated, with only 45 asthma ED visits among residents of [all ages](#) during the three-year period 2009-2011. Data are not broken down into age groups because of the small number of ED visits. **Table B2** also displays rates, rate ratios, and 95% CIs. The rate of asthma ED visits is lower in the impact study area than the rate in each of the four comparison areas. All of the CIs except for one exclude the numeral 1.0 indicating that, for most of the comparisons, the difference between the rates is not likely due to random fluctuation. The conclusion that can be drawn from **Table B2** is that the

asthma ED visit rate in the study area during this period was lower than the rate in an area with similar population density, the rate in the surrounding ZIP codes, the rate in Allegany County, and the rate in New York State excluding New York City.

- e. **Cancer incidence.** Instructions for display and analysis of cancer incidence data are included in Section I.G. Data are for the years 2005-2009. As discussed in Example 1, observed and expected numbers of cancer cases are tabulated by ZIP code for certain cancer sites at the NYSDOH web site (<http://www.health.ny.gov/statistics/cancer/registry/zipcode/index.htm>). When there are too few cases to be shown for confidentiality reasons, the New York State Cancer Registry combines ZIP codes with neighboring ZIP codes and provides the data in maps and tables for the combined groups of ZIP codes.

**Table B3** displays incidence of breast cancer in females, colorectal cancer in males and females, lung cancer in males and females, and prostate cancer for the years 2005-2009. In the NYSDOH cancer data tables, an additional ZIP code has been combined with ZIP code 14735; therefore, the study area for this analysis includes ZIP codes 14735 and 14744 plus the additional ZIP code 14745. The population density comparison area includes seven ZIP codes rather than four because three additional ZIP codes were combined with a ZIP code that was selected for this comparison area. For the surrounding ZIP codes comparison area, 11 ZIP codes had been selected; because some ZIP codes were combined in the NYSDOH cancer data tables, this comparison area now includes 16 ZIP codes.

As discussed in the previous example, Table B3 is different from the table of asthma ED visits in that the cancer data for the study are not compared directly to the cancer data for the comparison areas. Instead the number of observed cancer cases in the study area is being compared to the expected number of cases for the study area, which is calculated based on the cancer rate for New York State. Similarly, the number of observed cases in each comparison area is compared to the number of cases expected in that comparison area, which is calculated based on the cancer rate for New York State. In Table B3, the ratio represents the ratio of the observed number of cases to the expected number of cases. Because the data have been age-adjusted, the ratios are SIRs. (The SIRs in the table should not be compared with each other because of the different age distributions of the populations of each area.)

The SIRs in Table B3 for the study area (ZIP codes 14735 and 14744 plus another ZIP code) show that the SIRs for all the cancer sites are less than 1.0 or close to 1.0, indicating that the number of cases observed was lower than (or similar to) the number expected based on cancer rates in New York State. There were very few colorectal cancers in males or females and very few lung cancers in females in the study area during this period. With such small numbers of cases, an increase or decrease of one or two cases per year can cause a dramatic fluctuation in the rate; thus, the rates for these sites of cancer are considered too unstable for analysis and should be interpreted with caution.

In the second part of Table B3 (ZIP codes with similar population density to the impact study area), the SIRs are greater than 1.0 but the CIs include 1.0. This indicates that the observed numbers of cases are greater than those expected based on the cancer rates in New York State, but the differences could be due to random variation.

In the third part of Table B3 (ZIP codes surrounding the impact study area), some SIRs are greater than 1.0, but only one CI excludes 1.0 (in bold type). For female breast cancer, the number of cases observed is greater than the number expected based on the breast cancer rate in New York State, and this difference is unlikely due to random variation.

In the fourth part of Table B3 (Allegany County), some SIRs are greater than 1.0, and three CIs exclude 1.0. For colorectal cancer in males and lung and bronchus cancer in males and females, the observed numbers of cases are greater than those expected based on the cancer rates in New York State, and the differences are unlikely due to random variation.

- f. Low birth weight (LBW).* Instructions for display and analysis of LBW data are in Section I.G. **Table B4** shows the LBW data for 2009-2011 for the impact study area (ZIP codes 14744 and 14735) and the four comparison areas. All of the rate ratios are greater than 1.0, indicating that the rate of LBW is greater in the impact study area than in all of the four comparison areas. The CIs for the two ZIP code comparison areas exclude 1.0, indicating that these differences are unlikely due to random fluctuation. However, the number of LBW births is quite small in both the impact study area and these two ZIP code comparison areas, leading to less stable rates.

**Table A1.** Example 1: demographic data for ZIP code 11105 in Astoria, NY, and four comparison areas.

Group	ZIP code 11105	ZIP codes in Queens County with population density similar to ZIP code 11105*	ZIP codes surrounding 11105†	Queens County	New York City
Total population	36,688	207,701	202,431	2,230,772	8,175,133
Population/mi <sup>2</sup>	22,508	21,455	36,701	19,935	26,419
No. of households§	16,120	62,382	71,800	780,117	3,109,784
<b>Sex (%)</b>					
Male	48.7	47.5	48.7	47.5	48.7
Female	51.3	52.5	51.3	52.5	51.3
<b>Age distribution (%)</b>					
Less than 5 years	4.5	5.9	5.1	5.9	6.3
5 - 14	7.4	12.8	9.2	11.1	11.5
15 - 19	3.9	7.5	5.6	6.2	6.6
20 - 44	48.8	35.6	47.2	38.1	39.0
45 - 64	21.9	26.8	22.6	25.8	24.4
65+	13.6	11.4	10.3	12.8	12.1
<b>Race/ethnicity (%)</b>					
One race	96.5	93.0	95.8	95.5	96.0
White	75.5	15.4	49.5	39.7	44.0
African - American	2.2	49.0	7.0	19.1	25.5
American Indian / Alaskan	0.3	1.1	0.8	0.7	0.7
Asian	11.2	14.4	24.9	22.9	12.7
Hawaiian/Pacific Islander	0.1	0.2	0.1	0.1	0.1
Some other race	7.2	13.0	15.7	12.9	13.0
Two or more races	3.5	7.0	4.2	4.5	4.0
Hispanic or Latino (%)	19.7	20.5	34.6	27.5	28.6
Total minority (%)‡	35.5	92.4	68.3	72.4	66.7
<b>Income/poverty</b>					
Median household income (\$)§	57,576	64,390	51,145	56,780	51,865
Persons below poverty (%)§	13.6	11.4	14.7	14.4	19.9

Source: U.S. Census Bureau; unless otherwise noted, Census 2010.

\*ZIP codes with similar population density include ZIP codes 11412, 11418, 11420, 11423, 11428, 11429, 11436.

†Surrounding ZIP codes include ZIP codes 11102, 11103, 11370, 11377.

‡Minority includes Hispanic, African American or Black, Asian and Pacific Islander, or American Indian (see [New York State Department of Environmental Conservation Commissioner Policy CP-29](#)).

§From 2012 American Community Survey 5-year estimates, 2008-2012.

**Table A2.** Example 1: asthma emergency department (ED) visits by age group, 2010-2012, for ZIP code 11105 and four comparison areas.

Age group (years)	Part 1: ZIP code 11105			Part 2: ZIP codes in Queens County with similar population density to ZIP code 11105					Part 3: ZIP codes surrounding ZIP 11105				
	ED visits	2010- 2012 average population	Rate	ED visits	2010- 2012 average population	Rate	Rate ratio	95% CI	ED visits	2010- 2012 average population	Rate	Rate ratio	95% CI
0-4	61	2,019	100.7	1,437	15,606	306.9	0.33	(0.25, 0.42)	639	11,864	179.5	0.56	(0.43, 0.72)
0-17	161	6,925	77.5	3,301	52,326	210.3	0.37	(0.31, 0.43)	1,621	39,952	135.2	0.57	(0.49, 0.67)
18-64	340	27,508	41.2	3,954	142,306	92.6	0.44	(0.40, 0.49)	1,843	136,508	45.0	0.92	(0.82, 1.02)
65+	64	5,644	37.8	375	25,532	49.0	0.77	(0.59, 0.99)	255	22,833	37.2	1.02	(0.78, 1.30)
TOTAL	565	40,071	47.0	7,507	225,059	111.2	0.42*	(0.38, 0.45)	3,815	202,053	62.9	0.75*	(0.69, 0.81)

Age group (years)	Part 4: Queens County					Part 5: New York City				
	ED visits	2010- 2012 average population	Rate	Rate ratio	95% CI	ED visits	2010- 2012 average population	Rate	Rate ratio	95% CI
0-4	10,371	135,839	254.5	0.40	(0.30, 0.51)	56,458	532,274	353.6	0.28	(0.22, 0.37)
0-17	24,105	463,809	173.2	0.45	(0.38, 0.52)	134,469	1,776,157	252.4	0.31	(0.26, 0.36)
18-64	27,081	1,494,202	60.4	0.68	(0.61, 0.76)	178,547	5,459,200	109.0	0.38	(0.31, 0.39)
65+	2,914	292,646	33.2	1.14	(0.88, 1.45)	16,838	1,007,636	55.7	0.68	(0.52, 0.87)
TOTAL	53,847	2,261,761	79.4	0.59*	(0.54, 0.64)	321,539	8,270,641	129.6	0.36*	(0.33, 0.39)

NOTE. Rate = average annual rate of asthma ED visits per 10,000 population. Rate ratio = rate of ED visits in the impact study area (ZIP code 11105) divided by the rate in the comparison area. CI = 95% confidence interval. Source: Statewide Planning and Research Cooperative System (SPARCS).

\* Age-adjusted standardized incidence ratio using three age groups (0-17, 18-64, 65+ years).

**Table A3.** Example 1: incidence of four types of cancer for ZIP code 11105 and for three areas that are referred to as comparison areas, 2005-2009. Cancer rates in the study area are not compared directly to rates in these three areas.

Cancer site	No. of cases observed	No. of cases expected*	SIR	95% CI
<b>ZIP code 11105</b>				
Breast (female)	125	154.0	0.81	(0.68, 0.97)
Colorectal (male)	49	56.3	0.87	(0.64, 1.15)
Colorectal (female)	49	54.7	0.88	(0.65, 1.16)
Lung (male)	62	78.9	0.79	(0.60, 1.01)
Lung (female)	28	71.1	0.39	(0.26, 0.57)
Prostate	106	177.3	0.60	(0.49, 0.72)
<b>Zip codes with similar population density to ZIP code 11105</b>				
Breast (female)	573	776.3	0.74	(0.68, 0.80)
Colorectal (male)	237	233.9	1.01	(0.89, 1.15)
Colorectal (female)	226	253.1	0.89	(0.78, 1.02)
Lung (male)	231	326.1	0.71	(0.62, 0.81)
Lung (female)	188	336.3	0.56	(0.48, 0.64)
Prostate	897	763.5	<b>1.17</b>	<b>(1.10, 1.25)</b>
<b>ZIP codes surrounding ZIP code 11105</b>				
Breast (female)	517	666.6	0.78	(0.71, 0.85)
Colorectal (male)	201	219.2	0.92	(0.79, 1.05)
Colorectal (female)	210	224.3	0.94	(0.81, 1.07)
Lung (male)	276	300.1	0.92	(0.81, 1.03)
Lung (female)	195	292.2	0.67	(0.58, 0.77)
Prostate	465	689.6	0.67	(0.61, 0.74)
<b>Queens County</b>				
Breast (female)	7,489	8,781.3	0.85	(0.83, 0.87)
Colorectal (male)	2,785	2,888.4	0.96	(0.93, 1.00)
Colorectal (female)	2,930	3,071.1	0.95	(0.92, 0.99)
Lung (male)	3,207	4,022.4	0.80	(0.77, 0.83)
Lung (female)	2854	3,993.1	0.71	(0.69, 0.74)
Prostate	8,134	9,111.8	0.89	(0.87, 0.91)

NOTE. SIR = standardized incidence ratio: ratio of the number of cases observed to the number of cases expected. Because the data have been age-adjusted, the ratios are standard incidence ratios. CI = confidence interval. Rate ratios greater than 1.0 with confidence intervals that exclude 1.0 are shown in bold type. Source: New York State Cancer Registry.

\*The cancer rate for the entire state of New York and the number of people in a ZIP code are used to estimate the number of people in each ZIP code that would be expected to develop cancer within the five-year period 2005-2009 if the ZIP code had the same rate of cancer as the state.

**Table A4.** Example 1: rates of low birth weight (LBW) births for ZIP code 11105 and four comparison areas, 2009 – 2011.

<b>Area</b>	<b>No. LBW</b>	<b>No. births</b>	<b>Rate LBW (%)</b>	<b>Rate ratio</b>	<b>95 % CI</b>
ZIP code 11105	85	1,153	7.37	-	-
ZIP codes with similar population density to ZIP code 11105	856	8,087	10.6	0.70	(0.56, 0.85)
ZIP codes surrounding ZIP code 11105	535	7,304	7.33	1.01	(0.80, 1.23)
Queens County	7,393	91,264	8.1	0.91	(0.73, 1.11)
New York City	30,729	357,635	8.59	0.86	(0.69, 1.05)

NOTE. Rate ratio = rate of LBW births in the impact study area (ZIP code 11105) divided by the rate in the comparison area. CI = confidence interval. Source: New York State Vital Statistics.

**Table B1.** Example 2: demographic data for the impact study area in Allegany County, NY (ZIP codes 14744 & 14735), and four comparison areas.

Group	ZIP codes 14744 & 14735	ZIP codes in Allegany County with population density similar to 14744 & 14735*	ZIP codes surrounding ZIP codes 14744 & 14735†	Allegany County	New York State excluding New York City
Total population	4,845	4,252	11,770	48,946	11,202,969
Population/mi	48.3	38.6	31.4	47.3	207
No. of households§	1,425	2,653	11,764	26,140	4,207,971
<b>Sex (%)</b>					
Male	47.3	50.1	51.3	50.7	47.5
Female	52.7	49.9	48.7	49.3	52.5
<b>Age distribution</b>					
Less than 5 years	5.9	6.6	6.2	5.3	5.7
5 - 14	11.7	13.3	13.1	12.0	12.8
15 - 19	14.6	7.1	7.5	11.0	7.4
20 - 44	32.7	27.7	27.4	30.1	31.2
45 - 64	21.6	30.5	31.1	26.6	28.4
65+	13.5	14.8	14.6	15.0	14.5
<b>Race/ethnicity (%)</b>					
One race	99.2	99.0	99.2	99.2	97.7
White	95.9	97.8	98.1	96.3	83.6
African - American	0.9	0.3	0.3	1.2	8.8
American Indian / Alaskan	0.4	0.3	0.2	0.3	0.4
Asian	1.7	0.4	0.3	0.9	3.4
Hawaiian/Pacific Islander	0.0	0.0	0.0	0.0	0.0
Some other race	0.3	0.3	0.2	0.3	3.4
Two or more races	0.8	1.0	0.8	0.8	2.3
Hispanic or Latino (%)	1.5	1.2	0.9	1.4	9.6
Total minority (%)‡	5.2	3.1	2.5	4.5	23.4
<b>Income</b>					
Median household income (\$)§	42,498	39,755	43,667	41,900	61,110
Persons below poverty (%)§	18.2	14.7	9.7	16.6	10.9

Source: U.S. Census Bureau; unless otherwise noted, 2010 Census.

\*ZIP codes with similar population density include ZIP codes 14065, 14711, 14737, and 14777.

†Surrounding ZIP codes include ZIP codes 14024, 14060, 14065, 14066, 14536, 14709, 14711, 14717, 14777, 14836, and 14846.

‡Minority includes Hispanic, African American or Black, Asian and Pacific Islander, or American Indian (see [New York State Department of Environmental Conservation Commissioner Policy CP-29](#)).

§From 2012 American Community Survey, 2008-2012 5 year estimates.

**Table B2.** Example 2: asthma emergency department (ED) visits by age group, 2009-2011, for the impact study area in Allegany County, NY (ZIP codes 14744 & 14735), and four comparison areas.

Area	ED visits	2009-2011 average population	Rate	Rate ratio	95% CI
ZIP codes 14744 & 14735	45	4,759	31.5	-	-
ZIP codes in Allegany County with population density similar to ZIP codes 14744 & 14735	132	7,941	55.4	0.57	(0.41, 0.76)
ZIP codes surrounding 14744 & 14735	110	8,780	41.8	0.75	(0.55, 1.01)
Allegany County	642	48,694	43.9	0.72	(0.52, 0.96)
New York State excluding New York City	168,774	11,218,940	50.1	0.63	(0.46, 0.84)

NOTE. Rate = average annual rate of asthma ED visits per 10,000 population. Rate ratio = rate of ED visits in the impact study area (ZIP codes 14744 & 14735) divided by the rate in the comparison area. CI = 95% confidence interval. Source: Statewide Planning and Research Cooperative System (SPARCS).

**Table B3.** Example 2: incidence of four types of cancer for a study area including ZIP codes 14744, 14745 & 14735, and for three areas that are referred to as comparison areas, 2005-2009. Cancer rates in the study area are not compared directly to those in these two areas.

Cancer site	No. of cases observed	No. of cases expected*	SIR	95% CI
<b>Impact study area (3 ZIP codes including 14744, 14745 &amp; 14735)</b>				
Breast (female)	15	15.2	0.99	(0.55, 1.63)
Colorectal (male)	3	6.3	0.48†	(0.10, 1.39)
Colorectal (female)	1	5.7	0.17†	(0.00, 0.98)
Lung and Bronchus (male)	9	9.1	0.99†	(0.45, 1.88)
Lung and Bronchus (female)	4	7.3	0.55†	(0.15, 1.4)
Prostate	18	19.9	0.90	(0.54, 1.43)
<b>Zip codes with similar population density to ZIP codes 14744, 14745 &amp; 14735 (7 ZIP codes)</b>				
Breast (female)	37	31.0	1.19	(0.84, 1.65)
Colorectal (male)	18	12.2	1.48	(0.87, 2.33)
Colorectal (female)	14	10.4	1.35	(0.74, 2.26)
Lung and Bronchus (male)	22	17.4	1.26	(0.79, 1.91)
Lung and Bronchus (female)	20	14	1.43	(0.87, 2.21)
Prostate	42	40.1	1.05	(0.75, 1.42)
<b>ZIP codes surrounding ZIP codes 14744, 14745 &amp; 14735 (16 ZIP codes)</b>				
Breast (female)	65	48.2	<b>1.35</b>	<b>(1.04, 1.72)</b>
Colorectal (male)	28	19.4	1.44	(0.96, 2.09)
Colorectal (female)	14	16	0.88	(0.48, 1.47)
Lung and Bronchus (male)	30	27.4	1.09	(0.74, 1.56)
Lung and Bronchus (female)	22	21.1	1.03	(0.64, 1.56)
Prostate	67	64.1	1.05	(0.81, 1.33)
<b>Allegany County</b>				
Breast (female)	277	281.3	0.98	(0.87, 1.11)
Colorectal (male)	139	109	<b>1.28</b>	<b>(1.08, 1.51)</b>
Colorectal (female)	118	102.9	1.15	(0.95, 1.37)
Lung and Bronchus (male)	185	156.1	<b>1.19</b>	<b>(1.02, 1.37)</b>
Lung and Bronchus (female)	171	133.7	<b>1.28</b>	<b>(1.09, 1.49)</b>
Prostate	293	351.4	0.83	(0.74, 0.93)

NOTE. Because of small numbers of cases, some ZIP codes have been combined; see text. SIR = ratio of the number of cases observed to the number of cases expected. Because the data have been age-adjusted, the ratios are standardized incidence ratios. CI = confidence interval. Rate ratios greater than 1.0 with confidence intervals that exclude 1.0 are shown in bold type. Source: New York State Cancer Registry.

\*The cancer rate for the entire state of New York and the number of people in a ZIP code are used to estimate the number of people in each ZIP code that would be expected to develop cancer within the five-year period 2005-2009 if the ZIP code had the same rate of cancer as the state.

†SIRs are based on fewer than 10 cases; rates may be unstable and should be interpreted with caution.

**Table B4.** Example 2: rates of low birth weight (LBW) births for ZIP codes 14744 & 14735 and four comparison areas, 2009 – 2011.

Area	No. LBW	No. births	Rate LBW (%)	Rate ratio	95 % CI
ZIP code 14744 & 14735	16	172	9.3	-	-
ZIP codes with similar population density to ZIP code 14744 & 14735	12	317	3.8	<b>2.46</b>	<b>(1.40, 3.99)</b>
ZIP codes surrounding ZIP code 14744 & 14735	21	439	4.8	<b>1.95</b>	<b>(1.11, 3.16)</b>
Allegany County	122	1,609	7.6	1.23	(0.70, 1.99)
New York State excluding New York City	29,501	375,385	7.9	1.18	(0.69, 1.92)

NOTE. Rate ratio = rate of LBW births in the impact study area (ZIP code 11105) divided by the rate in the comparison area. CI = confidence interval. Rate ratios greater than 1.0 with confidence intervals that exclude 1.0 are shown in bold type. Source: New York State Vital Statistics.

## Section IV. Calculating Confidence Intervals

A basic spreadsheet such as Microsoft's Excel can be used to calculate confidence intervals using standard equations found in many epidemiology textbooks. This is especially useful if you will be repeating the calculation numerous times. There are also many web-based statistical calculators which can calculate confidence intervals, although the results may vary slightly depending on which method is used. While Excel does have the capability to calculate a number of descriptive statistics and statistical functions itself, including confidence intervals, many of these assume you have access to the raw data needed to calculate the variance and standard deviation. Since we are only given the either the observed and expected cases or rates among the study area and comparison populations, preprogrammed spread sheets with formulas that allow the user to calculate confidence intervals can be used without having all of the data points. Selection of the appropriate spreadsheet will depend on how your data are organized and the measure around which the confidence interval is being developed. So far there are two specific types of examples that have been given, asthma rates and cancer incidence. Each one needs to be approached slightly differently; however, the same general formula can be applied to both. The confidence limits are based on the Byar's approximation of the exact Poisson distribution which is extremely accurate even with small numbers.<sup>1</sup>

$$SIR = \frac{O}{E}$$

$$SIR_L = SIR \left( 1 - \frac{1}{9O} - \frac{Z_{\alpha/2}}{\sqrt{3O}} \right)^3$$

$$SIR_U = SIR \left( \frac{O+1}{O} \right) \left( 1 - \frac{1}{9(O+1)} + \frac{Z_{\alpha/2}}{\sqrt{3(O+1)}} \right)^3$$

SIR = Standardized Incidence Ratio

O = Observed number of cases

E = Expected number of cases

$Z_{\alpha/2}$  = Value of the standard normal distribution for a given significance level (alpha).

(For a 95% confidence interval alpha = 0.05 and  $Z_{\alpha/2} = 1.96$ )

**Links to spreadsheets to calculate confidence intervals are included in the sections below.**

**Asthma Hospitalizations or Emergency Department Visits:** For asthma hospitalization or emergency department visit data we are comparing two rates against each other using a *rate ratio* (RR - the ratio of the study area hospitalization or emergency department visit rate to a comparison area's rate). We have information on the number of hospitalizations or emergency

---

<sup>1</sup> Breslow NE, Day NE. Statistical methods in cancer research: Volume II. The design and analysis of cohort studies. International Agency for Research on Cancer, Lyon, France, 1987.

department visits and the rate per 10,000. With this information the population size for each area; the rate ratio; and its corresponding confidence interval can be calculated. The same equations can be used simply by applying the comparison area's rate to the study area population size to determine the expected number of cases. However, to calculate an expected number for the standardized (age-adjusted) rate in the final row it will be necessary to weight each age group and sum the individual age specific expected values. The formula for this is given below. The [spreadsheet for asthma](#) automatically makes these calculations plus it calculates the overall age-adjusted rate ratio for the comparison.

$$E^* = \sum_j n_j \lambda_j$$

$E^*$  = Age-adjusted expected number of cases

$n_j$  = Study area population of age group  $j$

$\lambda_j$  = Hospitalization rate of comparison population in age group  $j$

**Cancer Incidence:** For cancer data the expected rates have already been calculated and the observed rates for an area are supplied. Thus all that you need to do to calculate the *standardized incidence ratio* (SIR - the ratio to the observed vs expected number of cases) is to enter the observed and expected numbers in the spreadsheet for your study area. The [spreadsheet for cancer incidence](#) will calculate the standardized incidence ratio and generate 95% confidence intervals.

**Low Birth Weight:** For low birth weight (LBW) births, the percent of LBW births during a three-year period by ZIP code is included in Perinatal Data Profiles at the DOH website along with the total number of births in each ZIP code during the three-year period. With this information, the number of LBW births in each ZIP code can be calculated then summed to calculate the number of LBW births and the total number of births in each comparison area. These numbers can then be used to calculate the rates of LBW in the comparison areas; the rate ratios comparing the study area to the comparison areas; and the confidence intervals around each rate ratio. There is no age adjustment done for LBW births, but the [spreadsheet for LBW](#) will calculate the confidence intervals.