

**REPORT OF THE
HEALTH OUTCOME
DATA
WORK GROUP**

January 2006

**NEW YORK STATE DEPARTMENT
OF ENVIRONMENTAL CONSERVATION
NEW YORK STATE DEPARTMENT
OF HEALTH**

TABLE OF CONTENTS

	Page
I. Introduction and Background	1
II. Health Outcome Data and Demographics	1
A. Selection of Health Outcome Data	2
B. Data Quality	3
C. Data Display	4
D. Demographics	6
E. Conclusions	6
III. Health Outcome Comparisons	8
A. Selection of Comparison Areas	8
A.1. Potential Health Outcome Comparison Areas	8
A.2. Pros and Cons of Some Potential Comparison Areas	9
A.3. Discussion of Urban/Rural Classifications or Population Density	11
A.4. Determination of Population Density by ZIP Code	11
A.5. Criteria for Selecting Comparison Areas	12
B. Methods for Comparing Health Outcome Data Between Areas	13
B.1. Spatial Scan Statistics	13
B.2. Ranking of Geographic Units by Rates of Health Outcomes	14
B.3. Calculation of Rate Ratios and Confidence Intervals	14
B.4. Example I	15
Selection of Comparison Areas for the COC	15
Tabulation of Asthma Hospitalization Data	16
Tabulation of Cancer Incidence Data	17
B.5. Example II.	22
Selection of Comparison Areas for the COC	22
Tabulation of Asthma Hospitalization Data	22
Tabulation of Cancer Incidence Data	22
IV. Conclusions and Recommendations	27
Appendices	31
Appendix A – Health Outcome Data Work Group Members	31
Appendix B – Glossary	35
Appendix C - Evaluation of Data Sets	41

1. Cancer Registry	43
2. SPARCS	44
3. Vital Records	45
4. Developmental Disabilities.	45
Appendix D – Maps Showing Differences Between City and ZIP Code Boundaries	53
Appendix E – Methods for Creating Population Density Groupings	57
Appendix F – Age-Adjusted Rates and Confidence Intervals	63

LIST OF TABLES

<u>Table</u>	<u>Page</u>
<u>Table 1.</u> Example I: demographic data for ZIP Code X and four comparison areas.	19
<u>Table 2.</u> Example I: asthma hospitalizations for 1998-2000 by age group.	20
<u>Table 3.</u> Example I: cancer incidence data for ZIP Code X and for three other areas that are referred to as comparison areas in the text.	21
<u>Table 4.</u> Example II: demographic data for ZIP Codes A and B and three comparison areas.	24
<u>Table 5.</u> Example II: asthma hospitalizations for 1998-2000 for ZIP Codes A and B and three comparison areas.	25
<u>Table 6.</u> Example II: cancer incidence data for an area including ZIP Codes A and B and for three other areas that are referred to as comparison areas in the text.	26

I. INTRODUCTION AND BACKGROUND

The Environmental Justice Advisory Group (Advisory Group) was established by the New York State Department of Environmental Conservation (DEC) in 1989. After many discussions and public meetings, the Advisory Group issued a report in January 2002 to the DEC Commissioner entitled *Recommendations for the New York State Department of Environmental Conservation Environmental Justice Program*. The report, which focuses on the environmental permit process and contains recommendations for creating an effective environmental justice program, was followed by DEC Commissioner Policy-29, Environmental Justice and Permitting (CP-29, issued March 19, 2003). This policy established two work groups to assist DEC in developing and incorporating critical environmental justice information into the DEC environmental review process: the Disproportionate Adverse Environmental Impact (DAEI) Work Group and the Health Outcome Data (HOD) Work Group.

In its report, the Advisory Group recommended that closer scrutiny be given to environmental decisions in minority and low-income communities and that DEC expand the universe of information used in the permit approval process to address environmental justice concerns. The charge of the HOD work group as specified in CP-29 is to “identify reliable sources of existing human health data and recommend means to incorporate such data into the environmental review process.” Twelve individuals with technical knowledge and experience in the areas of environmental health, toxicology, epidemiology, risk assessment, or environmental sciences were appointed to the HOD work group (**Appendix A**). The work group began meeting in March 2003.

The work group identified the following as the scope of its work:

- To identify available health outcome data,
- To develop a protocol to describe and present health outcome data for use in DEC’s permit review process, and
- To provide recommendations to DEC

II. HEALTH OUTCOME DATA AND DEMOGRAPHICS

This section will define health outcome data and describe important considerations for selecting and displaying these data as part of the permitting process. Factors relating to data availability and quality and the need to include demographic information will be included, as well as a list of health outcomes recommended for inclusion.

Health outcome data are counts and rates of health-related events, such as cases of disease or deaths, in a population. It is important to note that 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 some geographic standard such as the city or state as a whole, 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, and exposures to chemicals in the air or water. However, if the population of a community has low health status, it may be more vulnerable to the effects of environmental exposures.

The term environmental exposures is used throughout this document. The term environmental exposures can be very broad and could, for example, include physical aspects of the

environment that could be related to a health outcome. For example, skiing accidents could be related to icy slope conditions. For this document, the term environmental exposures refers to contaminants in the air or water to which people are exposed and which might be related to health outcomes. Some examples of sources of environmental exposures include emissions from industrial facilities, motor vehicle emissions, spills, run-off from agricultural fields or roadways, emissions from power plants, furnaces, and woodstoves, etc.

A. Selection of Health Outcome Data

Health outcome data can provide information on the current health status of the community of concern. The community of concern will be identified by DEC staff. It is defined as an area whose population is likely to be affected by at least one potentially significant adverse environmental and/or human health impact related to a proposed action, for which a permit application has been submitted.

An important indicator of health status is the rate at which a particular disease or condition occurs in the population. The disease rate can be measured as incidence (new cases of a disease) or prevalence (existing cases of a disease) in a given population (see glossary in **Appendix B** for definitions). A population-based disease registry contains records for people diagnosed with a specific type of disease who reside within a defined geographic region. The New York State Department of Health (NYSDOH) maintains several population-based disease registries, including the Cancer Registry and the Congenital Malformations Registry.

Another important source of health data is the New York State Vital Records system, which includes information from birth and death certificates. Mortality data and information on birth outcomes, such as number of low-birth weight infants and infant deaths, can be obtained from Vital Records. Some data are available at the public web sites of the NYSDOH (<http://www.health.state.ny.us>) and the New York City Department of Health and Mental Hygiene (NYSDHMH) (<http://www.ci.nyc.ny.us/html/doh/home.html>). Data are also available from Federal agencies, other state agencies, and local sources such as county health departments, clinics and organizations serving local populations.

Population-based registries do not exist for every health outcome. The rate of occurrence of some health outcomes is estimated using other data sources, such as surveys and hospital discharge data. However, when data sets that have been created for another purpose, such as hospital billing or to record receipt of services, are used to estimate disease rates, it is important to recognize that factors such as how the data are collected and what data are collected will have an effect on the estimate.

When health outcomes are being selected for display, those that result from an acute exposure, such as asthma, may be more relevant to current place of residence than those that result from a chronic exposure, such as cancer. Because of the mobility of our population, the exposures that are important for an outcome such as cancer may have occurred decades earlier when the individual lived at a different location.

Since a vast amount of health information exists, the HOD work group developed the following criteria to provide guidance for the applicant in choosing health outcomes to be displayed in preparing an application that falls under the DEC environmental justice policy (CP-29).

1. The health outcome should be one that has some plausible relation to environmental exposure. The relation may exist because the exposure (1) is a potential cause of the health outcome, or (2) increases vulnerability to the effects of other causes of the health outcome. (This criterion is intended to distinguish a health outcome such as asthma that may be related to environmental exposures from a health outcome such as head injuries that is unlikely to be related to environmental exposures as defined above.)
2. The data for the health outcome should be readily available to the public at the ZIP Code level or smaller geographic level. As of July 2004, some health data are available on the NYSDOH public web site at the ZIP Code level or at a smaller geographical level. However, we anticipate that much of the data listed below will be available by ZIP Code in the future.
3. The data for the health outcome should be able to characterize both the population of interest and a comparison population, e.g., data should be collected on a countywide or statewide basis.
4. Sufficient information should be available to describe the quality of the data. The quality of the data for timeliness and completeness should be rated low, medium, or high. Accuracy and consistency of data collection across the state or county are difficult to rate on this type of scale. Available information on how accuracy and consistency of data collection are assessed should be summarized and provided.

Reliable sources of data that meet these criteria do not exist for every health outcome. Therefore, there may be health outcomes of concern to a community for which it may not be possible to use the methods described in this report. To gain more information, a community group may be interested in using another source of data such as the results of a local survey or information from a health clinic, but it is important that issues such as accuracy, completeness, and the availability of the data to characterize both the population of interest and a comparison population be discussed for every data source (see next section on Data Quality).

B. Data Quality

Those who develop and/or use the health outcome data should be aware that the various health outcome databases differ with respect to their timeliness, completeness, accuracy, and consistency across the state. The work group reviewed information on specific databases whose data may be useful for this project. These reviews can be used as models for applicants considering other databases. Each database was given qualitative ratings for the specific purpose of displaying health outcome data for consideration by DEC in the environmental review process (**Appendix C**). The databases were evaluated as follows.

- 1) Timeliness is a consideration of how current the data are or how soon after the end of the year for which the data are collected the data are available for use. For the purpose of this report, timeliness is ranked as:
 - a. High = 2 years or less
 - b. Medium = 5 years or less
 - c. Low = 10 years or less

- 2) Completeness is an assessment of the proportion of cases of the outcome that is actually captured by the database within the time frame and geographic areas of interest. A related question is whether reports are received from sources throughout the state (statewide coverage) or only from selected areas. To assess completeness of reporting from individual reporting sources, tracking systems, audits or other procedures must be in place. In some cases another data source may be compared to the database being evaluated. For the purpose of this report, completeness is ranked as:
 - a. High = estimated 90 % or greater completeness over the time frame and areas of interest
 - b. Medium = estimated 80% or greater completeness over the time frame and areas of interest
 - c. Low = estimated less than 80% completeness over the time frame and areas of interest

Completeness of reporting may vary by region of the state. The program maintaining the database may have information specific to the region of interest.

- 3) Accuracy is a determination of whether key fields are completed (not missing) and are completed correctly. To assess accuracy, the program must have quality assurance procedures in place, such as checks for presence of key fields, consistency, out-of-range values, duplicate reports, etc. Some databases have been evaluated for accuracy by an organization using specific numeric standards, but others have not. When the database has been assessed for accuracy, information is provided on these ratings. In other cases, available information relating to data accuracy has been summarized. The accuracy of reporting may vary by region of the state. The program maintaining the database may have information specific to the region of interest.
- 4) Information on the consistency of the data across the state or county should be available so that health outcome rates can be compared between different areas. Health outcome databases are not necessarily established for evaluating geographic differences but may have been developed for purposes such as recording vital statistics, rate-setting for hospitals, etc. One important issue is how cases or events (e.g., births, deaths) are identified and recorded. Equally important is that objective guidelines or criteria should exist that define an event. Information on factors that influence how cases or events are identified and counted as well as regional differences should be discussed.

C. Data Display

When health outcome data are presented in tables, the cases or events are frequently displayed in categories, such as gender, race/ethnicity, and age groups. The appropriate categories for display of a particular outcome depend on the characteristics of that outcome, i.e., whether the outcome is more common in a particular age group, ethnic group, etc. There is not a single set of display categories that is appropriate for all health outcomes. For example, Lyme disease is not generally displayed separately for males and females, but lung cancer is displayed separately for males and females because of interest in gender differences in lung cancer rates.

Tables of health outcome data generally show numbers of events, which can be births, deaths, or cases of a health outcome, and rates based on these numbers. A rate is a measure of the

frequency of an event per unit of population and time period. Rates are used in comparing populations since the number of events depends partly on the size of the population.

When data are broken down into many different categories, the number of events in each cell of the table decreases. 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. At the NYSDOH web site, <http://www.health.state.ny.us/nysdoh/chronic/ratesmall.htm>, there is a good discussion of the issue. 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%. Rates based on small numbers of events should be interpreted with caution since it is nearly impossible to distinguish random variability from true changes in the underlying risk of disease. For example, in Essex County, which in 1999 had a population of children 0-4 years old of about 2,400, the number of asthma hospitalizations of children 0-4 years old was 6 in 1998 and 2 in 1999, which appears to be a 67% decrease, but in fact may be due to random variability. Comparisons over time or between communities based on unstable rates can lead to invalid conclusions. NYSDOH is in the process of developing guidelines for presenting health data for small geographic areas. It is recommended that these guidelines be followed when they become available.

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 1 or 2 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, AIDS/HIV infection, etc., 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.¹ When possible, combine the number of cases or deaths 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 and the data are available by sub-categories, the data can be displayed in these 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, mortality.
 - 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.

¹ See <http://www.health.state.ny.us/nysdoh/chronic/ratesmall.htm> for a discussion of unstable rates due to small numbers.

- The categories used to display data for specific outcomes at the New York State Department of Health web site (<http://www.health.state.ny.us>) or the U.S. Centers for Disease Control and Prevention web site (<http://www.cdc.gov>) may be used as a guide.

D. Demographics

Demographic information describes a population in terms of the number of people and characteristics such as age, gender, race/ethnicity, income, etc. CP-29 (DEC's environmental justice policy) describes how potential environmental justice areas are identified using a Geographic Information System (GIS) application that takes into account the proportion of the community meeting the GIS application threshold for a low-income or minority community. The HOD work group believes that it is important to identify what demographic information would be important from a health standpoint and what breakdown groupings should be used.

The source for demographic information is the U.S. Census Bureau, which collects information on the U.S. population every 10 years; the most recent information available is from the 2000 Census accessible at <http://www.census.gov>. A vast amount of information is available from the U.S. Census, but for the purposes of this report the work group recommends focusing on age, gender, race/ethnicity, and a measure of income such as median household income or percent of persons under poverty. Gender is important when considering health outcomes because certain health outcomes affect only one gender (e.g., prostate cancer) or are more common in one gender than the other (e.g., breast cancer, cardiovascular disease). It is important to look at specific age groups because some health outcomes are more common in certain age groups; for example, asthma hospitalization rates are highest in young children. Similarly, data for women of child-bearing age would be reviewed if there were a concern about adverse reproductive outcomes. The U.S. Census Bureau displays information on age by 10-year or smaller age groups, but these age groups are frequently combined into fewer groups in published reports. Income level or the proportion of the population living in poverty is relevant because certain factors that can affect rates of disease are higher or lower in poor populations. These factors include smoking and exposure to second-hand smoke, quality of housing, adequacy of nutrition, and access to and source of medical care. Rates of some health outcomes vary among racial and ethnic groups, for example, the infant mortality rate and mortality rates due to cardiovascular disease, breast cancer and prostate cancer.

Table 1 displays demographic information for a ZIP Code that meets DEC's environmental justice criteria for an urban area and four comparison areas. The data can be obtained from the U.S. Census Bureau web site (<http://www.census.gov>) through AmericanFactFinder Quick Tables. Information on age, gender and race/ethnicity are from Summary File 1 in table DP-1 Profile of General Demographic Characteristics 2000. Information on median household income and percentage of individuals under poverty are from Summary File 3 in table DP-3 Profile of Selected Economic Characteristics 2000.

E. Conclusions

Based on the criteria outlined in Part A of this section, the following potential list of health outcomes is recommended for consideration whenever health outcome data are displayed under DEC's environmental justice policy. The most recent data available should be included in the health outcome data display. (The most recent years for which data were available as of July 2004 are in parentheses.) The applicant may consider it appropriate to include additional

health outcomes based on the specific application or geographic area. This list of potential outcomes could be revised in the future based on new information.

Only health outcome data that are publicly available at the ZIP Code or smaller geographic level are being recommended for inclusion. As of July 2004, some health data are available on the NYSDOH public web site at the ZIP Code or smaller geographical level; these data are shown in bold type. However, we anticipate that much of the data listed below will be available by ZIP Code in the future.

1. Respiratory diseases
 - a. **Hospitalization (SPARCS 1992-2000) (codes from ICD-9-CM)²: asthma (ICD 493)³, chronic bronchitis (ICD 491) and chronic obstructive pulmonary disease (ICD 490-496)**
 - b. Mortality (Vital Records 1998-2000): asthma, chronic obstructive pulmonary disease

2. Cardiovascular diseases
 - a. Hospitalization (SPARCS 1992-2000): cardiovascular and other diseases of the circulatory system (SPARCS) (ICD 390-459), all diseases of the heart (ICD 390-398, 402, 404-429)
 - b. Mortality (Vital Records 1998-2000): cardiovascular and other diseases of the circulatory system, all diseases of the heart

3. Cancer
 - a. **Incidence (Cancer Registry 1993-1997): breast, lung, prostate, colorectal⁴**
 - b. Mortality (Vital Records 1996-2000): total cancer deaths, deaths due to specific cancers

4. Perinatal health (County/Zip Perinatal Data Profiles; Vital Records 1998-2000)
 - a. Number of births
 - b. Percent of births that are low birth weight (<2500 grams)
 - c. Number and rate of infant deaths

5. Lead
 - a. **Incidence of blood lead levels greater than or equal to 10 micrograms per deciliter in children younger than 6 years screened 1997-1999, New York State, excluding New York City, ZIP Code map showing 3 rate ranges⁵**
 - b. **Numbers and rates of children with blood lead levels greater than or equal to 10 micrograms per deciliter, ages 6 months to less than 6 years, New York City, by ZIP Code, 2000⁶**

² ICD-9-CM is the *International Classification of Diseases, 9th Revision, Clinical Modification*. This classification is designed for coding of morbidity and mortality information for statistical purposes, and for the indexing of hospital records by disease and operations, for data storage and retrieval. The 10th Revision is currently under development. The codes for outcomes of interest may change for data released in the future as subsequent revisions of the classification come into use.

³ Available by ZIP Code for New York City for children 0-14 years at <http://www.nyc.gov/html/doh/pdf/asthma/facts.pdf>

⁴ Available by ZIP Code at <http://www.health.state.ny.us/nysdoh/cancer/csii/nyscsii.htm>

⁵ Available by ZIP Code at <http://www.health.state.ny.us/nysdoh/lead/childlead.pdf>, see figure 5.

⁶ Available by ZIP Code at <http://www.nyc.gov/html/doh/html/lead/l2002t4.html>

Some data sets were reviewed but are not included in the recommended list. Although NYSDOH has a Congenital Malformations Registry, congenital malformations are rare events, and the data are not available below the county level because of confidentiality issues when there are small numbers of cases. Information on students with developmental disabilities receiving special education services is reviewed in **Appendix C**. These data are also not available at the ZIP Code level or smaller geographic level, and at this time there are concerns about comparing rates in one region to those in another. As more knowledge is gained on the relationship of special education classification rates to the prevalence of developmental disabilities, the usefulness of these data for comparing regions may be re-evaluated in the future.

III. HEALTH OUTCOME COMPARISONS

A. Selection of Comparison Areas

In order to provide a context for evaluating the health outcome data for the community of concern (COC), data for the same health outcomes should also be displayed for comparison areas. It is recommended that multiple comparison areas be used to provide a broad context in which to view the health outcome data. Some comparison areas are large areas that should be chosen to represent some “average” health status. Some comparison areas should be smaller and located in fairly close proximity to the COC. An advantage of having a small local comparison area is that the community will be familiar with the area including the land use in the area; however, a small area, especially in a rural setting, may have a small number of cases leading to the problem of unstable rates (see Section II.C). A comparison area may also be chosen because it is similar to the COC in terms of its degree of urbanization; population density (number of people per square mile) is frequently used to denote how urban or rural an area is.

For some health outcomes such as cancer, the way in which the data are provided might have an effect on the comparison areas that can be used. For example, the cancer data available by ZIP Code provide observed and expected numbers of cancers, and the expected numbers are based on the New York State population. In this case, it is recommended that indirect comparisons be shown; more detail is provided in Examples I and II of Section III.B.

A.1. Potential health outcome comparison areas

- New York State
- New York City if the COC is located in New York City
- New York State excluding New York City if the COC is located outside of New York City
- County in which COC is located (or county minus the COC)⁷
- ZIP Codes approximating the city or town in which COC is located (or city minus the COC)⁷
- ZIP Codes surrounding the COC
- ZIP Codes within the county with population density similar to the COC

⁷ If the COC contains 5%-10% of the population of the county, city, or town being used as the comparison area, it is recommended that the COC be excluded from the comparison area since its population may affect the analysis.

A.2. Pros and cons of some potential comparison areas

- Pro: Statewide health outcome data are readily available since NYSDOH routinely compiles data on a statewide basis. Federal agencies such as the U.S. Centers for Disease Control and Prevention collect health data from the states for inclusion in national reports that allow comparisons with other parts of the country.
- Pro: Comparison with the statewide average provides a sense of whether the COC is more severely or less severely affected in terms of the health outcomes selected than the state as a whole.
- Pro: Because New York State has a large population, health outcome rates for this comparison areas would be stable.
- Con: The state has a large area and a large and diverse population; using a statewide average makes it difficult to see differences in rates between parts of the state and subgroups of the population.

New York City for a COC in New York City

- Pro: Data for New York City are readily available since NYCDHMH routinely compiles health data for the city.
- Pro: The five New York City counties are generally more similar to each other than to the counties outside New York City. This comparison provides information on whether the COC is more severely or less severely affected in terms of these health outcomes than the city as a whole.
- Pro: The existence of health-related programs administered through the NYCDHMH provides some baseline of health programs and services throughout the city.
- Pro: Because New York City has a large population, health outcome rates for this comparison area would be stable.
- Con: New York City includes a large and diverse population; for example, the demographics of Richmond County are quite different from those of the other counties in the city. Using a New York City average makes it difficult to see differences between parts of the city and subgroups of the population.

New York State excluding New York City for a COC located outside of New York City

- Pro: Data for New York State excluding New York City can be for any database with county level data.
- Pro: The counties outside of New York City are generally more similar to each other than to the New York City counties. This comparison provides information on whether the COC is more severely or less severely affected in terms of these health outcomes than the state as a whole, excluding New York City.
- Pro: Because New York State excluding New York City has a large population, health outcome rates for this population would be stable.
- Con: The counties outside of New York City include a large area and a large and diverse population, encompassing for example both the Adirondack Park and the City of Buffalo. Using this comparison makes it difficult to see differences between parts of the state and subgroups of the population

The county in which the COC is located (or county minus the COC)⁷

- Pro: County data are generally readily available since health data are routinely compiled at the county level.
- Pro: A county is a small area relative to the area of New York City or New York State excluding New York City. This comparison provides information on whether the COC is

more severely or less severely affected in terms of these health outcomes than the county as a whole.

- Pro: The existence of health-related programs administered through the county health departments provides some baseline of health programs and services throughout the county.
- Con: If the COC is located at the border of a county, some of the surrounding area will be in the adjacent county(ies).
- Con: A county may include a diverse population particularly if it includes an urban area. Using a county average makes it difficult to see differences between parts of the county and subgroups of the population.

ZIP Codes approximating the city or town (other than New York City) in which the COC is located (or city minus the COC) ⁷

- Pro: Data are currently available by ZIP Code for some health outcomes; it is anticipated that more data will be available by ZIP Code in the future.
- Pro: A city has a small area relative to the state or the county. The COC is likely to be more similar to the city or town than to larger, more distant areas. Local differences in diagnosis and reporting of health outcomes are likely to be minimized. This comparison provides information on whether the COC is more severely or less severely affected in terms of these health outcomes than the city or town as a whole.
- Con: The ZIP Code boundaries may not conform well to city boundaries. ZIP Codes outside of New York City can be very large; some ZIP Codes include urban and suburban areas. Maps providing examples of the differences between city and ZIP Code boundaries are included in **Appendix D**.

ZIP Codes surrounding the COC

- Pro: Data are currently available by ZIP Code for some health outcomes; it is anticipated that more data will be available by ZIP Code in the future.
- Pro: The COC is likely to be similar to the ZIP Codes surrounding it. Local differences in diagnosis and reporting of health outcomes are likely to be minimized. This comparison provides information on whether the COC is more severely or less severely affected in terms of these health outcomes than its immediate surroundings.
- Con: If the COC is on the edge of an urban area (either inside or outside), the ZIP Codes surrounding the COC may be quite diverse. Using an average for this diverse area may make it difficult to see differences between parts of the area and subgroups of the population.

ZIP Codes in the county with population density similar to the COC

- Pro: Data are currently available by ZIP Code for some health outcomes; it is anticipated that more data will be available by ZIP Code in the future.
- Pro: The COC is likely to be more similar in terms of land use to ZIP Codes in the county with similar population density, e.g., if the COC is in an area with low population density because the proposal requires a certain amount of space or for other reasons related to land use, the comparison area chosen by this method will be similar in this aspect. As such, these areas could be other areas where the facility could be sited. This comparison provides information on whether the COC is more severely or less severely affected in terms of these health outcomes than an area in the county with similar urban/rural characteristics.
- Pro or con: If the COC is in a densely populated urban area, it may be a low-income and minority community. A comparison area selected by this method may or may not be

similar to the COC in socioeconomic status. See the discussion on pages 12-13 on the pros and cons of comparison areas similar to the COC in socioeconomic status.

- Con: Some ZIP Codes are very large and include densely populated areas and sparsely populated areas. Using an average population density for these ZIP Codes makes it difficult to see differences between the densely populated and sparsely populated areas.
- Con: In some rural areas, it may not be feasible to establish population density groupings of ZIP Codes because there may be very few ZIP Codes within a county or the ZIP Codes may all be similar in population density.

A.3. Discussion of urban/rural classifications or population density

Population density is the number of people per square mile (or square kilometer). It indicates whether a community is urban or rural. Urban areas contain more people per square mile and thus are more densely populated than suburban areas, which are more densely populated than rural areas. Urban areas differ from rural areas in land use and environmental factors (e.g., density of permitted facilities, traffic, etc.). A comparison area with population density similar to that of the COC is likely also to be similar in land use. For example, if the facility is proposed for an area with low population density because a certain amount of space is required, comparison areas chosen by this method will be similar in this aspect. As such, these areas could be other areas where the facility could be sited, and population density could serve as a proxy for land use.

Sometimes areas with different population densities also differ in demographic factors such as income, race and ethnicity, and age distribution. Some urban areas and some very rural areas have a lower median household income and a higher percentage of the population living in poverty than suburban areas. The large urban centers in New York State tend to have a higher percentage of minorities than less densely populated areas. Some factors that can affect rates of disease are higher or lower in different income groups. These factors include smoking and exposure to second-hand smoke, quality of housing, adequacy of nutrition, access to and source of medical care, and other environmental exposures.

Because of the work group's interest in a comparison area similar to the COC in urban/rural characteristics or population density, four urban/rural classification schemes to define such an area were explored: determination of population density groupings by ZIP Code, a classification method used by the Behavioral Risk Factor Surveillance System, a method used by the U.S. Environmental Protection Agency, and the U.S. Census urban/rural classification scheme. The most usable method was found to be a determination of population density groupings by ZIP Code. The other three classification schemes either resulted in cruder groupings than the calculation of population density or focused on geographic units other than ZIP Codes. These other methods are described in **Appendix E**.

A.4. Determination of population density groupings by ZIP Code

The population within each ZIP Code in the county should be estimated by adding together the population of 2000 census blocks whose geographic centers (centroids) fall within the ZIP Code. The area of each ZIP Code in square miles should be calculated using the area function of commercial mapping software. Population density should be calculated as persons per square mile. The ZIP Code population groups can be determined by the natural break method,

which is based on the procedure described by Jenks and Caspall.⁸ This method divides the data into a specified number of groups (6 in Example I of Part B of this section) and then examines the means of each group. The data are divided in such a way as to maximize the difference between the means of the groups. This method is used by some commercial mapping software. The number of groups that should be selected depends somewhat on the range of population density and the number of ZIP Codes. With a wide range or a greater number of ZIP Codes, more groups might be needed, while a narrow range or a small number of ZIP Codes might call for fewer groups. A general rule of thumb might be a minimum of 3-4 groups and a maximum of 8 groups.

A.5. Criteria for selecting comparison areas

1. Use multiple comparison areas (at least three) to have enough information to evaluate the disease rates in the COC in the context of a number of different settings. Include numbers 2, 3, and 4 below. If the way in which the data are provided has an effect on the comparison areas that can be used, as with ZIP Code level cancer data, show indirect comparisons if possible (see Examples I and II in Section III.B).
2. At least one of the comparison areas should be a local area (e.g., within the same city or county) since the community will be familiar with the type of land use in local areas.
3. At least one of the comparison areas should include a population within the same general geographic area (e.g., county or contiguous counties) that is similar in population density to the COC. This comparison would provide information about whether the COC is more severely or less severely affected in terms of the health outcomes displayed than a community that is similar in urban/rural characteristics.
4. At least one of the comparison areas should be a larger area such as the county, New York State exclusive of New York City, or New York City to obtain a wider perspective.
5. If specific alternative locations for the facility are being considered, these specific alternative locations should be included as comparison areas.
6. Other comparison areas may also be included with explanations for why they were selected.

The applicant also has the option to consider including a comparison area that is similar to the COC in socioeconomic status (SES), which is generally measured by income or educational level. In this report, two variables used as indicators of SES are the percentage of the population living in poverty and the percentage of the population that is minority. Higher levels of poverty have been reported among minority populations than among whites (http://www.healthypeople.gov/Document/html/uih/uih_bw/uih_2.htm#goals). In general, higher disease and mortality rates are found among low-income and minority populations. Specific health outcomes that are higher among low-income and minority populations include heart disease, low birth weight, and infant mortality. Factors that can affect rates of disease in low-income populations include quality of housing, adequacy of nutrition, smoking and exposure to second-hand smoke, and access to and source of medical care.

The work group discussed the pros and cons of including a comparison area that is similar in SES to the COC. The process recommended in Section III.A for selecting multiple comparison areas may result in comparison areas that have a range of SES. However, none of the comparison areas may be of SES similar to the COC (see Example I of Section III.B). The

⁸ Jenks, Caspall. "Error on Choroplethic Maps: Definition, Measurement, Reduction." *Annals of American Geographers*, June, 1971.

discussion is summarized to assist the applicant in deciding whether to include this type of comparison area among the multiple areas that are recommended.

- Pro: Use of a similar SES comparison area would provide information on how health outcome rates in the COC compare with those in a population with similar risk factors such as quality of housing, smoking, nutrition, access to and source of medical care, etc. This information is traditionally sought by epidemiologists.
- Con: Rates of some diseases are known to be higher in low-income and minority communities. If the disease rates in a low SES COC were found to be the same as those in the comparison area of similar SES, a possible interpretation could be that the COC does not have elevated disease rates even though the disease rates in the COC and the comparison area are elevated above the “average” levels represented by a larger area such as New York State excluding New York City or New York City.
- Pro: With multiple comparison areas with different SES levels, there would be an opportunity to discuss differences in disease rates between the COC and comparison areas with similar, higher or lower SES. A finding that a low SES COC has higher disease rates than a comparison community with similar SES might suggest that other factors in this community may be contributing to the elevated rates of disease and could suggest that this community be given greater consideration under DEC's environmental justice policy (CP-29).
- Con: From the environmental justice perspective, comparing a low-income and minority community to another low-income and minority community may not be appropriate because the rates of certain health outcomes are known to be higher in low-income and minority communities. When one such community is compared to another, disease rates may be high in both communities with no difference seen between the two.

B. Methods for Comparing Health Outcome Data Between Areas

Three options were explored for comparing health outcome data from the COC with data from the comparison areas to aid in the interpretation of the differences between the disease rates: use of spatial scan statistics, ranking, and calculation of rate ratios and confidence intervals.

B.1. Spatial scan statistics

SaTScan™ software was developed at the National Cancer Institute for cluster detection and geographic surveillance.⁹ The method looks at the number of cases and the population at risk in a geographic unit such as a ZIP Code and compares rates inside and outside the geographic unit. The program then generates multiple circles around each geographic unit in which the radius of the circle varies continuously from zero to a predefined upper limit (based on maximum population size or geographic size) each time capturing additional neighboring geographic units and comparing the rates inside and outside the circle. Statistical tests that take into account multiple testing are applied for significance. The user must set a number of parameters including the model (Bernoulli, Poisson or space/time permutations), the scale (cluster size from 1 to 50% of the population), shape (circle or ellipse and shape of ellipse), and

⁹ See <http://srab.cancer.gov/satscan/>

degree of overlap of the shapes. There are no “correct” parameters, and the choice depends on the disease, the distribution of disease, and the questions being asked. The method is not very sensitive to rare diseases or diseases occurring in a small population. The method may be best for finding areas of the state with elevated disease rates and may be more complex than needed for the health outcome displays described in this report. For the purpose of comparing disease rates between one area and another, calculation of rate ratios and confidence intervals is more straightforward.

B.2. Ranking of geographic units by rates of health outcomes

All ZIP Codes in a particular county could be placed in rank order by rate of a specified health outcome such as asthma hospitalizations. A statement could be made as to the ranking number of the ZIP Code containing the COC out of the total number of ZIP Codes in the county. A similar ranking by rate of a health outcome could also be done for the counties in New York State, and a statement could be made about the rank order of the county containing the COC. This type of comparison could provide some information about the rate of the health outcome in the ZIP Code containing the COC relative to the rest of the county (or the county containing the COC relative to the rest of the state). However, in ZIP Codes with small populations, there may be small numbers of cases of the health outcome of interest, leading to rates that may be high one year and low the next year (see Section II.C for a discussion of unstable rates due to small numbers). Therefore, it would be necessary to include information on population size with any ranking of ZIP Codes. For the purpose of comparing disease rates between one area and another, more information would be provided by the calculation of rate ratios and confidence intervals.

B.3. Calculation of rate ratios and confidence intervals

Data on health outcomes are generally provided as the number of cases in a specified geographic area, the population of that area, and the rate of the disease (the number of cases during a time period in a specified population, e.g., per 10,000 or 100,000 population). It is important always to compare rates rather than numbers of cases since different areas will have different population sizes. As stated previously, areas with small populations may have small numbers of cases of the health outcome of interest, leading to rates that may be high one year and low the next year (see Section II.C for a discussion of unstable rates due to small numbers). When possible, annual rates should be calculated. In order to assess the difference between the rate of the health outcome in the COC and the rate in the comparison area, a rate ratio can be calculated and a confidence interval can be found. The rate ratio is the ratio of the rate in the COC to the rate in the comparison population. When the two rates being compared are the same, then the ratio is equal to 1. When the two rates are being 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.

A confidence interval that does not include 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. (More information can be found at <http://www.doh.wa.gov/Data/Guidelines/ConfIntguide.htm> and <http://www.health.state.ny.us/nysdoh/cancer/nyscr/age.htm> and in **Appendix F**). The use

of rate ratios and confidence intervals is best shown in examples of health outcome data display. Two examples follow, one using a COC in an urban area and the other using a COC in a rural area. Comparison areas for the COC will be selected, tables of health outcomes (asthma hospitalizations and two types of cancer) in the COC and the comparison areas will be developed, and disease rates in the COC will be compared to those in the comparison areas.

B.4. Example I

For the first example, a COC in an urban area not located in New York City was chosen. The COC is ZIP Code X, which meets the urban environmental justice criteria and is located in City Y and County Z.

Selection of comparison areas for the COC. Four comparison areas were selected to evaluate the disease rates in the COC in the context of a number of different settings.

- 1) A comparison area made up of ZIP Codes in County Z that are similar to ZIP Code X in population density. This area will be similar to ZIP Code X in land use and urban/rural characteristics.
- 2) A comparison area that is the remainder of the city that contains ZIP Code X (i.e., City Y minus ZIP Code X). This comparison area is in the local area and was chosen because the community will be familiar with the area and the type of land use in this area.
- 3) County Z, which is the county that contains ZIP Code X. This is a larger area and represents average health status, but it has 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 COC.
- 4) New York State excluding New York City is a large area chosen to represent the average health status.

To determine the ZIP Codes that approximate the remainder of City Y, a map of ZIP Codes was overlaid onto a map of the city using commercial mapping software. Some ZIP Codes were clearly within the city. Other ZIP Codes lay partially within the city and partially outside the city. For each of these ZIP Codes, the proportion of the population that lay within the city boundaries was estimated by summing the population of the 2000 census blocks whose geographic centers (centroids) fell within the ZIP Code. ZIP Codes with greater than 50% of the population within the city boundaries were included in the remainder of City Y comparison group; six ZIP Codes were included. ZIP Code X was not included in this comparison group because the population of ZIP Code X was about 10% of the population of City Y, which may be great enough to affect the results of the analysis.

For the population density comparison area, the ZIP Codes in County Y with population density similar to ZIP Code X were selected using the methods described in Section III.A.4 of this report. Six population density groupings were created. ZIP Code X was categorized by population density into Group II (2,816 – 7,467 people per square mile), which has the second highest population density of the six groups in the county. The comparison group is made up of the six ZIP Codes in Group II with population density similar to that of ZIP Code X, but does not include ZIP Code X. The population density comparison area is similar to the remainder of City Y comparison area but not identical; the two comparison areas had five ZIP Codes in common.

Demographic data for ZIP Code X and the four comparison areas are shown in Table 1. The remainder of City Y and the population density comparison areas are closer to the COC than the larger comparison areas (County Z and New York State excluding New York City) in characteristics such as median household income and percentage minority.

Tabulation of asthma hospitalization data. Data for many health outcomes are made available by age groups. The specific age breakdowns vary by the health outcome and by the agency reporting the data, but the categories generally include an age group for children, one for adults, and one for the elderly. These three groups are sometimes broken down into smaller age groups. Young children are an important focus when asthma hospitalizations are reviewed since asthma hospitalization rates are known to be higher in young children. The age groups available for asthma hospitalizations when this report was being prepared were 0-4, 5-14, 15-24, 25-44, 45-64, and 65+ years, but different age breakdowns may be available at the time an applicant is preparing health outcome data displays. For example, to focus on childhood asthma but also to cover the complete age distribution, asthma hospitalization data could be displayed by these age groups: 0-4 years, 0-17 years, 18-64 years, and 65+ years.

The NYSDOH data used for this example included the number of asthma hospitalizations during 1998-2000, the 1999 population, and the average annual rate of asthma hospitalizations per year. Asthma hospitalization data for ZIP Code X and the four comparison areas are shown in Table 2. As an example, the average annual rate of asthma hospitalizations per 100,000 population is calculated as follows for the 0-4 year old group in ZIP Code X:

$$\frac{\text{No. hospitalizations over 3 years}}{\text{1999 population X 3 years}} = \frac{19 \text{ hospitalizations over 3 years}}{782 \times 3 \text{ years}} = 0.008099 \text{ hospitalizations per person per year}$$

$$\text{Average annual rate of hospitalization per 100,000 population} = 0.008099 \times 100,000 = 809.9 \text{ hospitalizations per 100,000 population per year}$$

The data in Table 2 are shown for six age groups and the total population for ZIP Code X and the four comparison areas (Parts 1-5). Table 2 Parts 2-5 have an additional column marked "rate ratio (95% CI)." This column compares the data for ZIP Code X to that for the specified comparison area. For each age group, the rate ratio is the hospitalization rate for ZIP Code X divided by the hospitalization rate for the comparison area. As an example, for the 0-4 year old age group, the following rate ratio is calculated to compare the asthma hospitalization rate in ZIP Code X to that in the remainder of City Y (part 3):

$$\text{Rate ratio} = \frac{\text{Rate in ZIP Code X}}{\text{Rate in remainder of City Y}} = \frac{809.9}{616.9} = 1.31$$

For the total population in parts 2-5, the different age distributions in the COC and the comparison areas have been taken into account in calculating a ratio called the age-adjusted standardized incidence ratio (SIR). It can be seen that every rate ratio and SIR in Table 2 is greater than 1, which indicates that for each age group the asthma hospitalization rate in ZIP Code X is higher than the rate in each of the four comparison areas. The ratios are considerably higher when ZIP Code X is compared with the larger comparison areas (County Z and New York State excluding New York City, parts 4 and 5) than when the COC is compared with the local comparison areas (parts 2 and 3), indicating that the rates in the COC are more similar to those in the local comparison areas.

As described above, the confidence interval (CI) provides a range around the ratio in which the true measurement lies with a certain degree of confidence. The 95% CI was calculated by the exact method of Daly.¹⁰ In this case the 95% CI indicates that there is a 95% chance that the true value of the ratio is included in the range. In parts 2-5 of Table 2 for age groups 5-14 years and above, as well as the total, all of the CIs exclude the numeral 1 and appear in bold type; for these comparisons, there is 95% confidence that the asthma hospitalization rate in ZIP Code X is greater than the rate in the comparison area and that the difference is not due to chance.

Of interest in Table 2 is that the asthma hospitalization rates, especially for children, are much higher in the remainder of City Y than in County Z, indicating that rates in the county outside of City Y are considerably lower than in City Y. This is a good example of how differences in disease rates among portions of a larger area such as a county might not be seen if one looks only at the rate for a larger area such as a county.

For the 0-4 year old group, the asthma hospitalization rate for ZIP Code X was higher than that in all four comparison areas, but the difference did not reach the 95% confidence level for the remainder of City Y and population density comparison areas because the rates in these areas were also quite high (616.9 and 573.9, respectively, compared with 359.6 for County Z).

The conclusion that can be drawn from Table 2 is that asthma hospitalization rates in the COC for 1998-2000 were higher than those in the city surrounding the COC, in the county in which the COC is located, and in New York State excluding New York City. Because the same result is seen with multiple comparisons, there is added confidence in the results.

Although review of asthma hospitalizations is important in asthma surveillance, people who are hospitalized with asthma are only a subset of people with asthma. Asthma hospitalization rates differ from region to region and are known to be higher among poor and minority populations. National data from the U.S. CDC for 1999 showed that asthma hospitalization rates were about three times higher among blacks than whites.¹¹ Poverty appears to be an important contributing factor to asthma disability.¹² ZIP Code X, the remainder of City Y, and the population density comparison areas (Table 1) have higher percentages of minorities and persons living in poverty than County Z and New York State excluding New York City. Currently, research is being conducted on factors that may contribute to higher asthma hospitalization rates among poor and minority populations, including differences in access to medical care, source of care, medical management including type of medications, and provider-patient interactions.^{13 14 15}

Tabulation of cancer incidence data. Maps and tables showing cancer incidence data by ZIP Code for breast, lung, colorectal and prostate cancers are available at the NYSDOH web site

¹⁰ Daly, L. Simple SAS macros for the calculation of exact binomial and Poisson confidence limits. *Comput. Biol. Med.* 22(5):351-361, 1992.

¹¹ U.S. Centers for Disease Control and Prevention. Surveillance for asthma – United States, 1980-1999. *Morbidity and Mortality Weekly Reports* 51(SS01):1-13, 2002.

¹² Healthy People 2010. Respiratory Diseases.

<http://www.healthypeople.gov/document/html/volume2/24Respiratory.htm>

¹³ Haas, JS et al. The impact of socioeconomic status on the intensity of ambulatory treatment and health outcomes after hospital discharge for adults with asthma. *J. Gen. Intern. Med.* 9(3):121-126, 1994.

¹⁴ Finkelstein, JA et al. Quality of care for preschool children with asthma; the role of social factors and practice setting. *Pediatrics* 95(3):389-394, 1995.

¹⁵ Lieu, T et al. Racial/ethnic variation in asthma status and management practices among children in managed Medicaid. *Pediatrics* 109(5):857-865, 2002.

(<http://www.health.state.ny.us/nysdoh/cancer/csii/nyscsii.htm>). Cancer data that are available from NYSDOH are tabulated in a different way from the asthma hospitalization data discussed in the previous section. The asthma data are presented by age group, including number of hospitalizations, population and hospitalization rate. 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, but the rate and number of people by age group are not provided. Instead, 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 the five-year period 1993-1997 if the ZIP Code had the same rate of cancer as the state (for more information, see "Frequently Asked Questions" at <http://www.health.state.ny.us/nysdoh/cancer/csii/nyscsii.htm>). Age and population size are taken into consideration when determining the expected number; this process is called age-adjustment (**Appendix F**).

For some ZIP Codes in the cancer maps and tables, there were too few cases to be shown for confidentiality reasons. These ZIP Codes are combined with neighboring ZIP Codes, with the data provided for the combined groups of ZIP Codes.

In Table 3 data on breast and colorectal cancers are shown for ZIP Code X and three comparison areas. This table is different from the table of asthma hospitalizations (Table 2) in that the cancer data for ZIP Code X are not 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 ZIP Code X and for the three comparison areas. (Therefore, the fourth comparison area used in the asthma hospitalization area, New York State excluding New York City, is not being used in this example because ZIP Code X is already being compared to a large area representing average health status.) 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 different populations.) If the observed number were equal to the expected number, the SIR would be equal to 1. As discussed previously, the CI is a range around the ratio in which the true measurement lies within a certain degree of confidence, in this case, 95%. When the CI excludes 1, there is 95% probability that the difference is not due to random variation. The 95% confidence interval is generally chosen by convention; however, additional confidence intervals could be displayed.

The SIRs in Table 3 for ZIP Code X compared with New York State show that the breast cancer rate in ZIP Code X over this time period was essentially the same as that in New York State, while the colorectal cancer rate was somewhat higher. For both breast and colorectal cancer, the 95% CIs around the SIRs include 1, which indicates that the difference in the observed and expected numbers may be due to random variation. When the observed numbers of breast and colorectal cancers in the three comparison areas were compared to the numbers expected based on the rates in New York State, the SIRs were all close to 1. The 95% CIs all include 1, indicating that the differences could be due to random variation. The conclusion that can be drawn from the data in Table 3 is that the breast and colorectal cancer rates for 1993-1997 in the COC were similar to the rates in New York State excluding New York City. In addition, breast and colorectal cancer rates in the city surrounding the COC and in the county in which the COC is located were also similar to those in New York State excluding New York City.

Table 1. Example I: demographic data for ZIP Code X and four comparison areas.

Group	ZIP Code X	ZIP Codes in County Z with population density similar to ZIP Code X	ZIP Codes approximating the remainder of City Y	County Z	New York State excluding New York City
Total population ¹	10,021	84,967	85,841	294,565	10,968,179
Sex (%) ¹					
Male	46.0	47.5	47.7	47.8	51.2
Female	54.0	52.5	52.3	52.2	48.8
Age distribution (%) ¹					
Less than 5	8.5	5.0	5.3	5.7	6.4
5 – 14	15.8	10.6	10.7	13.0	14.5
15 - 19	6.8	9.5	9.4	7.5	7.0
20 - 44	42.0	41.3	42.6	36.6	35.1
45 - 64	18.3	19	18.1	22.8	22.7
65+	8.6	14.5	13.8	14.5	14.0
Race/ethnicity (%) ¹					
One race					
White	35.6	73.2	66.3	83.2	85
African-American	54.5	18.7	25.1	11.1	8
American Indian/Alaskan	0.5	0.2	0.3	0.2	<1
Asian	1.0	3.6	3.5	2.7	2
Hawaiian/Pacific Islander	0.0	0.0	0	0.0	<1
Some other race	4.3	1.7	1.9	1.1	2
More than one race	3.9	2.5	2.9	1.7	2
Hispanic or Latino ¹	9.4	4.5	5.2	3.1	6
Minority (%) ^{*1}	66.7	28.6	35.7	18.2	18
Median household income 1999 (\$) ²	21,250	36,278	31,499	42,935	47,517
Persons below poverty 1999 (%) ²	33.7	16.4	20.3	10.6	10.0

* Minority includes Hispanics, African-Americans, Asian-Americans, Pacific Islanders and Native Americans, and others.

¹U.S. Bureau of the Census. *2000 Census of population and housing summary file 1 (SF1)*. U.S. Department of Commerce. 2001.

²U.S. Bureau of the Census. *2000 Census of population and housing summary file 3 (SF3)*. U.S. Department of Commerce. 2002.

Table 2. Example I: asthma hospitalizations for 1998-2000 by age group. (CIs that exclude 1 are shown in bold type.)

Age group (years)	Part 1: ZIP Code X			Part 2: ZIP Codes in County Z with population density similar to ZIP Code X				Part 3: ZIP Codes approximating the remainder of the City Y			
	Hospitalizations	1999 Pop.	Rate*	Hospitalizations	1999 Pop.	Rate*	Rate ratio † (95% CI)	Hospitalizations	1999 Pop.	Rate*	Rate ratio † (95% CI)
0 - 4	19	782	809.9	74	4,298	573.9	1.41 (0.85-2.20)	90	4,863	616.9	1.31 (0.79-2.05)
5 - 14	21	1,497	467.6	70	8,772	266.0	1.76 (1.09-2.69)	81	9,756	276.8	1.69 (1.05-2.58)
15 - 24	10	1,162	286.9	25	15,184	54.9	5.23 (2.51-9.61)	31	16,306	63.4	4.52 (2.17-8.32)
25 - 44	19	3,336	189.8	63	25,396	82.7	2.30 (1.39-3.60)	80	28,311	94.2	2.02 (1.21-3.15)
45 - 64	38	1,673	757.1	65	15,526	139.6	5.42 (3.84-7.44)	67	16,807	132.9	5.70 (4.03-7.82)
65+	12	980	408.2	55	13,070	140.3	2.91 (1.50-5.08)	58	13,468	143.6	2.84 (1.47-4.97)
TOTAL	119	9,430	420.6	352	82,246	142.7	2.62‡ (2.17-3.13)	407	89,511	151.6	2.41‡ (1.99-2.88)

Age group (years)	Part 4: County Z				Part 5: New York State excluding New York City			
	Hospitalizations	1999 Pop.	Rate*	Rate ratio † (95% CI)	Hospitalizations	1999 Pop.	Rate*	Rate ratio † (95% CI)
0 - 4	194	17,981	359.6	2.25 (1.36-3.52)	7,653	710,992	358.8	2.26 (1.36-3.52)
5 - 14	170	40,634	139.5	3.35 (2.07-5.12)	5,687	1,644,788	115.3	4.06 (2.51-6.20)
15 - 24	78	43,430	59.9	4.79 (2.30-8.81)	2,921	1,397,425	69.7	4.12 (1.97-7.57)
25 - 44	210	94,248	74.3	2.56 (1.54-3.99)	8,943	3,321,742	89.7	2.12 (1.27-3.31)
45 - 64	201	63,435	105.6	7.17 (5.07-9.84)	8,737	2,482,673	117.3	6.45 (4.57-8.86)
65+	162	47,073	114.7	3.56 (1.84-6.22)	6,884	1,504,271	152.5	2.68 (1.38-4.68)
TOTAL	1,018	306,801	110.6	3.62‡ (3.00-4.33)	40,825	11,061,891	123.0	3.36‡ (2.79-4.02)

Source: SPARCS, 1998-2000.

CI = confidence interval.

*Average annual rate of asthma hospitalizations per 100,000 population.

†Rate in ZIP Code X is numerator; rate in comparison area is denominator.

‡Age-adjusted standardized incidence ratio.

Table 3. Example I: cancer Incidence data for ZIP Code X and for three other areas that are referred to as “comparison” areas in the text. The study area is not compared directly to these three areas (see footnotes below).

Cancer site	No. of cases observed	No. of cases expected*	Ratio†	95% CI
ZIP Code X				
Breast (female)	26	26.1	1.00	(0.65,1.46)
Colorectal (male)	13	9.1	1.43	(0.76,2.44)
Colorectal (female)	14	11.4	1.23	(0.67,2.06)
ZIP Codes in County Z with population density similar to ZIP Code X				
Breast (female)	295	287	1.03	(0.91,1.15)
Colorectal (male)	133	114.3	1.16	(0.97,1.38)
Colorectal (female)	146	150.4	0.97	(0.82,1.14)
ZIP Codes approximating the remainder of City Y				
Breast (female)	313	299.8	1.04	(0.93,1.17)
Colorectal (male)	139	120	1.16	(0.97,1.37)
Colorectal (female)	157	154.6	1.02	(0.86,1.19)
County Z				
Breast (female)	1379	1373.1	1.00	(0.95,1.06)
Colorectal (male)	595	587.8	1.01	(0.93,1.10)
Colorectal (female)	651	660.9	0.99	(0.91,1.06)

Source: New York State Cancer Registry’s Cancer Surveillance Improvement Initiative, 1993-1997 (<http://www.health.state.ny.us/nysdoh/cancer/csii/nyscsii.htm>). CI = confidence interval.

*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 1993-1997 if the ZIP Code had the same rate of cancer as the state.

†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 (SIRs).

B.5. Example II

For the second example, a COC in a rural area in western New York State was chosen. The COC comprises two ZIP Codes (A and B) and meets the rural environmental justice criteria for poverty.

Selection of comparison areas for the COC. Three comparison areas were selected to have enough information to evaluate the disease rates in the COC in the context of a number of different settings. One comparison area is made up of ZIP Codes in County C that are similar to ZIP Codes A and B in population density and is thus likely to be similar in urban/rural characteristics. This comparison area is located in the same county as the COC; the community is likely to be familiar with the type of land use in this area. Two comparison areas (County C and New York State excluding New York City) are larger areas and represent the average health status. Of these two areas, however, County C is smaller and is in closer proximity to the COC than New York State excluding New York City.

For the population density comparison area, the ZIP Codes in County C with population density similar to ZIP Codes A and B were selected using the methods described in Section III.A.4 of this report. Four population density groupings were created, but the entire county is sparsely populated except for one village. ZIP Codes A and B were categorized by population density into Group III (38–68 people per square mile), which has the third highest population density of the four groups. The comparison group is made up of eight ZIP Codes in Group III with population density similar to that of ZIP Codes A and B, but does not include ZIP Codes A and B. Demographic data for ZIP Codes A and B and three comparison areas are shown in Table 4. The population density comparison area and County C are more similar to the COC than New York State excluding New York City in characteristics such as median household income and percent minority.

Tabulation of asthma hospitalization data. ZIP Codes A and B are sparsely populated, with only 9 asthma hospitalizations among residents of all ages during the three-year period. In each age group, there were 2 or fewer hospitalizations. As discussed in Part C of Section II of this report, the display of 1 or 2 cases should be avoided for confidentiality reasons. In the same section of this report, the problem of unstable rates due to small numbers of cases is discussed. Rates based on 2 or fewer cases would be very unstable. For these reasons, asthma hospitalizations are not shown for age groups for ZIP Codes A and B. In Table 5 asthma hospitalizations are presented for the total population of ZIP Codes A and B as well as for the three comparison areas. The asthma hospitalization rate in ZIP Codes A and B is lower than that in all three comparison areas. The CIs all contain the numeral 1 indicating that the difference between the rates may be due to random fluctuation. Since the asthma hospitalization rate for ZIP Codes A and B is based on fewer than 10 hospitalizations, it is noted in the table that the data should be interpreted with caution. (See Example I for the methods used to calculate rates, ratios, CIs, and age-adjusted ratios.)

The conclusion that can be drawn from Table 5 is that the asthma hospitalization rate for 1998-2000 in the COC did not differ significantly from the rate in an area in the county with similar population density, from the rate in the county as a whole, or from the rate in New York State excluding New York City. However, these results should be viewed with caution because of the small number of asthma hospitalizations in the COC.

Tabulation of Cancer Incidence Data. As discussed in Example I, observed and expected numbers of cancer cases are tabulated by ZIP Code for certain cancer sites at the NYSDOH

web site (<http://www.health.state.ny.us/nysdoh/cancer/csii/nyscsii.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. Since the area around the COC is sparsely populated, many ZIP codes have been combined. Information on breast and colorectal cancers is shown in Table 6. The study area and comparison areas presented in Table 6 do not match exactly the areas for asthma hospitalizations in Table 5. A third ZIP Code has been combined with ZIP Codes A and B. Of the eight ZIP Codes in the population density comparison area in the asthma table, two were dropped because they were combined for the cancer data with a ZIP Code not originally selected for the comparison area, and four were added because they were combined with a ZIP Code that was selected for the comparison area. There are a total of 10 ZIP Codes in the population density comparison area in Table 6.

As discussed in the previous example, Table 6 is different from the table of asthma hospitalizations in that the cancer data for the study area (which includes ZIP Codes A and B) are not compared directly to the cancer data for the comparison areas. Instead, the expected number of cases is based on the cancer rate for New York State, i.e., the state is the comparison area for the study area and for the three comparison areas. In Table 6, 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. If the observed number were equal to the expected number, the SIR would be equal to 1. As discussed previously, the CI is a range around the ratio in which the true measurement lays with a certain degree of confidence, in this case 95%. When the CI excludes 1, there is 95% chance that the difference between the rates is not due to random variation.

The SIRs in Table 6 for the study area (ZIP Codes A and B plus another ZIP Code) compared with New York State show that the breast cancer rate was essentially the same as that in New York State, while the colorectal cancer rate was somewhat lower. The colorectal cancer rates are based on very small numbers and are thus unstable and should be interpreted with caution. The CIs for these comparisons include 1, indicating that the differences could be due to random variation.

When the observed numbers of breast and colorectal cancers in the two comparison areas were compared to the numbers expected based on the rates in New York State, the rate ratios were all close to 1. All of the CIs include 1, indicating that the differences could be due to random variation.

Because of the small numbers of cases in shown in Table 6, conclusions must be drawn with caution. Since all of the CIs included 1, the breast and colorectal cancer rates for 1993-1997 in an area approximating the COC were similar to those in New York State excluding New York City. The breast and colorectal cancer rates in an area in the county with similar population density and in the county as a whole were also similar to those in New York State excluding New York City.

Table 4. Example II: demographic data for ZIP Codes A and B and three comparison areas.

Group	ZIP Codes A and B	ZIP Codes in County C with population density similar to ZIP Codes A and B	County C	New York State excluding NYC
Total population ¹	5,225	18,685	49,927	10,968,179
Sex (%) ¹				
Female	52.5	50.4	50.0	51.2
Male	47.5	49.6	50.0	48.8
Age distribution (%) ¹				
Less than 5	5.7	6.0	5.6	6.4
5-14	14	15.8	14.3	14.5
15-19	12.8	7.7	10.5	7.0
20-44	37.0	32.3	33.4	35.1
45-64	18.8	24.5	22.2	22.7
65+	11.7	13.8	14.0	14.0
Race/ethnicity (%) ¹				
One race				
White	97.4	97.5	97.0	85
Black	<1	<1	<1	8
American Indian/ Alaskan	<1	<1	<1	<1
Asian	<1	<1	<1	2
Hawaiian/Pacific Islander	0	0	0	<1
Some other race	<1	<1	<1	2
More than one race	<1	<1	<1	2
Hispanic or Latino ¹	1.3	<1	<1	6
Minority (%) ^{*1}	3.4	3.0	3.4	18
Median household income 1999 (\$) ²	31,440	33,306	32,106	47,517
Persons below poverty 1999 (%) ²	21.8	14.2	15.5	10.0

* Minority includes Hispanics, African-Americans, Asian-Americans, Pacific Islanders and Native Americans, and others.

¹U.S. Bureau of the Census. *2000 Census of population and housing summary file 1 (SF1)*. U.S. Department of Commerce. 2001.

²U.S. Bureau of the Census. *2000 Census of population and housing summary file 3 (SF3)*. U.S. Department of Commerce. 2002.

Table 5. Example II: asthma hospitalizations for 1998-2000 for ZIP Codes A and B and three comparison areas. The data in this table should be interpreted with caution since the rate in ZIP Codes A and B is based on fewer than 10 hospitalizations.

Area	Hospitalizations	1999 population	Rate*	Age-adjusted standardized incidence ratio (95% CI) †
ZIP Codes A and B	9	3,972	75.5	-
ZIP Codes in County C with population density similar to ZIP Codes A and B	54	17,649	102.0	0.79 (0.36,1.49)
County C	171	50,200	113.5	0.68 (0.31,1.29)
New York State excluding New York City	40,825	11,061,891	123.0	0.63 (0.29,1.19)

Source: SPARCS, 1998-2000. 1999 Population from Claritas Corp. CI = confidence interval.

*Average annual rate of asthma hospitalizations per 100,000 population.

†As compared with ZIP Codes A and B.

Table 6. Example II: cancer Incidence data for an area including ZIP Codes A and B and for three other areas that are referred to as “comparison” areas in the text. The study area is not compared directly to these three areas (see footnotes below). Rates based on fewer than 10 cases should be interpreted with caution.

Cancer site	No. of cases observed	No. of cases expected*	Ratio†	95% CI
Study area (3 ZIP Codes including ZIP Codes A and B)				
Breast (female)	14	12.5	1.12	(0.61,1.88)
Colorectal (male)	2	6.0	0.33	(0.04,1.20)
Colorectal (female)	3	5.9	0.51	(0.10,1.49)
Population density comparison area (14 ZIP Codes)				
Breast (female)	45	54.6	0.82	(0.60,1.10)
Colorectal (male)	25	26.9	0.93	(0.60,1.37)
Colorectal (female)	24	24.7	0.97	(0.62,1.45)
County C				
Breast (female)	248	240.4	1.03	(0.91,1.17)
Colorectal (male)	122	115.7	1.05	(0.88,1.26)
Colorectal (female)	118	115.3	1.02	(0.85,1.23)

Source: New York State Cancer Registry’s Cancer Surveillance Improvement Initiative, 1993-1997 (<http://www.health.state.ny.us/nysdoh/cancer/csii/nyscsii.htm>). CI = confidence interval. If a ZIP code in the study or comparison area includes an additional ZIP code, then data from the additional ZIP code are also included in the observed and expected tabulations. However, if a study or comparison area ZIP code is included in another ZIP code that is out of the study area, then data from that ZIP code are excluded from the tabulations.

*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 1993-1997 if the ZIP Code had the same rate of cancer as the state.

†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 (SIRs).

IV. CONCLUSIONS AND RECOMMENDATIONS

Recommendations about selecting, displaying and comparing health outcome data are provided throughout the body of this report. A critical factor to recognize when using health outcome data displays in the permitting process is that health outcome data are providing information on a disease rate in a community compared to other communities (i.e., county, state, etc.). These data cannot tell what is causing the disease. Many factors, including the way the health outcome data are collected, may affect the estimated disease rates. Other factors that influence the risk of disease include heredity, age, lifestyle factors such as smoking and diet, adequacy of nutrition, housing and health care. The characteristics of the population, including any environmental exposures or other risk factors such as lifestyle factors, were not considered when the health outcome data were collected or displayed. However, if the population of a community has low health status, it may be more vulnerable to the effects of environmental exposures. The important recommendations in the report are summarized below.

- Demographic information, in addition to health outcome data, should be provided for the COC (see Section II.D.) and any comparison community. An environmental justice community is identified by its race/ethnicity or income distribution and this information will already be known. However, other demographic information such as the number of women of childbearing age may be important from a health perspective.
- Four criteria (see Section II.A.) are provided to guide the selection of health outcome data for preparing an application that falls under CP-29 (DEC's environmental justice policy). One criterion states that the health outcome data should be readily available at the ZIP Code level or lower, and very few health outcome data meet this requirement. County data are available for many more outcomes; however, those data do not provide information at a geographic level that would correspond to a community being affected by emissions from a facility. At this time, data on five health outcomes are available from NYSDOH at the ZIP Code level or lower: breast, colorectal, lung and prostate cancer rates and asthma hospitalization data. The New York City Department of Health and Mental Hygiene has ZIP Code level data available at its web site for New York City lead poisonings.
- If a database is going to be used by an applicant, sufficient data should be available to describe the data quality. Understanding the data quality could lead to the determination that a database should not be used for this purpose (see Section II.B. and Appendix C).
- General guidelines are provided for describing how the data should be displayed (see Section II.C). Data displays should include numbers of events and disease rates (a measure of the frequency of an event per unit of population). The diseases should be displayed in categories, such as gender, race/ethnicity, and age groups that are appropriate for the health outcome of interest. When 10 or fewer events are displayed, these rates may be unstable and not useful. Although these data can be displayed, any conclusion about how rates compare among different populations is less certain than when rates are calculated from a greater number of events.
- Comparison areas should be selected for the COC (see Section III.A.). At least three comparison areas should be used: one local area (e.g., within the same city or county), one area in the same general geographic area with the same population density, and one larger area (county, New York City, or New York State minus New York City). Using

multiple comparison areas provides a context for understanding and interpreting the disease rates for the COC. For some health outcome data such as the cancer data, the way in which the data are provided affects the comparison areas that can be used. The ZIP Code cancer data are provided as observed and expected numbers of cases; the expected number is based on the New York State population. In such instances, indirect comparisons should be used; more detail is provided in the examples in Section III.B.

- The health outcome data for the COC should be compared to the health outcome data for the comparison areas to aid in interpreting differences in disease rates. In general, rate ratios and confidence intervals, or the ratios of the number of cases observed to the number of cases expected along with confidence intervals, should be used. The method for the comparison may depend on data availability. As long as the results from one of these two methods are provided, other methods, along with a description and discussion of the method and results, can also be provided.

The data tables and analyses in the examples in this report were prepared by NYSDOH staff familiar with accessing and manipulating health outcome data, use of geographic information systems, and use of SAS software for statistical analysis. Using publicly available ZIP Code data for asthma and cancer, the analyses for each of the examples required a day or less of work by one staff person.

The person conducting the health outcome comparisons for the permit application process should have the following qualifications:

Master's degree and at least three years of applied research experience in epidemiology and statistics, including experience in management of computerized health data. A related Doctoral degree may substitute for one year of experience. Experience in community environmental health research is preferred.

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 COC and comparison 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 COC and comparison populations. 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 summaries. The greater the number of conditions that are met, the greater the likelihood is that the health status of the COC is actually lower than that found in other areas.

- A disease rate is higher in the COC than in any comparison population for any health outcome;
- A disease rate is higher in the COC than in multiple comparison populations for any health outcome;
- The confidence intervals do not include 1 (A greater elevation in health outcome in the COC compared to the comparison population and a larger number of events will increase the likelihood that the confidence interval will not include 1.);
- There is a pattern of higher rates of multiple health outcomes in the COC; and
- 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 health outcome data should be used along with other considerations (need, benefits, cost, impact, etc.) in making a permitting decision and should not necessarily be the determining factor in any decision. If the health outcome data suggest that the health status of the COC is lower than that of the comparison populations, possible options can be identified for the permitting process. The significance of the difference between the COC and the comparison populations should be considered in determining which options may be more appropriate than others. Possible options may include, but are not limited to:

- Considering alternative siting locations outside of the COC and/or options that would improve environmental and health conditions in the COC;
- Applicant or DEC evaluation and inclusion of options that will reduce pollution into the community to the maximum extent feasible (Examples include 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 or lowest achievable control technology; or the implementation of holistic environmental management system. The applicant may provide a wide array of recommendations to the Department on how it can reduce pollution from the facility.);
- Evaluating the feasibility of reducing existing exposures from other sources that may be contributing to a health outcome of concern, determining the feasibility of obtaining emission offsets from other sources of air pollution in the COC, and implementing the options deemed feasible;
- Providing assistance through CP-29 (DEC's Environmental Justice policy) to the COC to prevent, diagnose, or treat the health outcome of concern or improve health status; and
- Taking no action based on the results of the health outcome comparisons.

Outside of these recommendations, there are 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.epa.gov/compliance/environmentaljustice/grants/index.html>). Groups may be interested in developing community health report cards and may find the guidelines and examples in this report helpful. 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 COC.

The work group was unable to complete some of the tasks it identified. DEC, in conjunction with DOH, may want to implement additional suggestions the work group has.

- Compile a bibliography that could be used to help guide people interested in exploring the option of conducting community health surveys to provide health data not readily available.

- As additional health outcome data sets become available to use for this policy, develop examples (similar to those provided in Section IV.B.) for applicants to follow. Developing examples for lead exposures, perinatal health outcomes, respiratory diseases, and cardiovascular diseases should be initiated as soon as feasible as those outcomes are particularly relevant to CP-29 (DEC's Environmental Justice policy).

APPENDIX A

Appendix A

HEALTH OUTCOME DATA WORK GROUP ENVIRONMENTAL JUSTICE ADVISORY GROUP

Nancy K. Kim, Ph.D., Chair
Director
Division of Environmental Health Assessment
New York State Department of Health
Flanigan Square
547 River Street
Troy, NY 12180-2216

Lung Chi Chen, Ph.D., Associate Professor
School of Medicine, Environmental Medicine
New York University
Sterling Forest
57 Old Forge Road
Tuxedo, NY 10987

Thomas Gentile, MPH
Chief, Air Toxics Section
Division of Air Resources
New York State Department of
Environmental Conservation
625 Broadway
Albany, NY 12233

Kazuhiko Ito, Ph.D.
School of Medicine, Environmental Medicine
New York University
Sterling Forest
57 Old Forge Road
Tuxedo, NY 10987

Carole Ju, M.S.
Center for Environmental Health
New York State Department of Health
Flanigan Square
547 River Street
Troy, NY 12180-2216

Patrick Kinney, Sc.D.
Department of Environmental Sciences
Mailman School of Public Health at
Columbia University
60 Haven Avenue, B-1
New York, NY 10032

Jessica Leighton, PhD
Environmental Disease Prevention
New York City Department of Health
and Mental Hygiene
253 Broadway
New York, NY 10013

Carlos Montes
New York State Department of
Environmental Conservation
625 Broadway
Albany, NY 12233

David Morgott, Ph.D.
Health and Environment Laboratories
Kodak Park, B-320
Eastman Kodak Company
1100 Ridgeway Avenue
Rochester, NY 14652-6272

Scott Stoner, M.S.
Chief, Standards and Classifications Unit
Division of Water
New York State Department of
Environmental Conservation
625 Broadway
Albany, NY 12233

David Strogatz, Ph.D.
Department of Epidemiology
School of Public Health
State University of New York at Albany
1 University Place
Rensselaer, NY 12144

Brenda Nordenstam, Ph.D., State University of New York, Environmental Science and Forestry, and John Vena, Ph.D., State University of New York, Buffalo, were on the committee in the beginning but could not continue to participate because of other priorities.

APPENDIX B

Appendix B

GLOSSARY

Behavioral Risk Factor Surveillance System (BRFSS)

An on-going data collection program designed to measure behavioral risk factors in the adult, noninstitutionalized, civilian population in the U.S. Data are collected from a random sample of adults through a telephone survey. The program is administered by individual states and supported by U.S. Centers for Disease Control and Prevention.

Census Block

The smallest geographic area for which data are provided in the U.S. Census. Other census geographic areas (tract, block group, county) are made up of two or more groups of census blocks. The area of a block may be bounded by any physical feature such as a street, railroad track or stream, and by municipal boundaries such as a city, town, and county limits. In a city, a census block is generally a city block.

Centroid

The geographic center or midpoint. In a census block, the block centroid is the geographic center.

Chronic Disease

A disease that shows little change, or progresses slowly and persists over a long period of time.

COC

Community of concern, defined by the New York State Department of Environmental Conservation as an area whose population is likely to be affected by at least one potentially significant adverse environmental and/or human health impact related to a proposed action, for which a permit application has been submitted.

Comparison Population

Subjects with whom comparison is made in a case control study, randomized controlled trial, or other variety of epidemiologic study. In a case control study, researchers compare a population with a disease or condition of interest (case population) to a suitable group of persons without the disease (comparison or control population) to see if past exposure to chemicals or other risk factors were different. The comparison population resembles the case population in such respects as age and sex but do not have the disease or condition of interest.

Congenital Malformation

Any structural, functional, or biochemical abnormality determined genetically or induced during gestation and not due to birthing events. The **Congenital Malformations Registry** of the New York State Department of Health is a repository for case reports on children who are born or reside in New York State and are diagnosed before the age of two years.

Database

An organized set of data or collection of files that can be used for a specified purpose.

Demographic Information

The "person" characteristics—age, sex, race, and occupation—of descriptive epidemiology used to characterize populations at risk.

Disease Registry

A special database that contains information about people diagnosed with a specific type of disease. Most disease registries are either hospital-based or population-based.

Environmental Exposures

For the purposes of this report, contaminants such as chemicals in the air or water to which people are exposed and which might be related to health outcomes.

Environmental Justice

As defined by the New York State Department of Environmental Conservation, the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies. Environmental justice efforts focus on improving the environment in under-served communities, and addressing disproportionate adverse environmental impacts that may exist in those communities.

Epidemiology

The study of the occurrence and causes of health effects in specified human populations. An epidemiological study often compares two groups of people who are alike except for one factor such as exposure to a chemical or the presence of a health effect. The investigators try to determine if the factor is associated with the health effect.

Gender

The category to which an individual is assigned, in an epidemiological study for example, based on sex.

Health Outcome Data

Information on the current health status of people in a community, such as the number and proportion of people who have a variety of diseases or health outcomes.

Hospitalization

Admission of a patient to and confinement in a hospital.

Incidence of Disease

The number of newly diagnosed cases of a particular disease in a defined population over a specified time. This measure of disease helps researchers determine a person's probability of being diagnosed with a disease during a given period. **Incidence Rate** is calculated by dividing the number of newly diagnosed cases of a disease by the number of persons at risk for the disease.

International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM)

A manual of the complete list of specific conditions and groups of conditions determined by an internationally representative group of experts who advise the World Health Organization.

Low Birth Weight

Weight of a newborn that is less than 2500 grams (approximately 5.5 pounds).

Micrograms per Deciliter (mcg/dl or µg/dl)

A unit that expresses the concentration of a substance in a solution in some medical tests, such as the amount of lead in blood. A gram is 1/28 of an ounce; a microgram is one millionth of a gram. A deciliter is a measure of volume that is one tenth of a liter, about half a cup of liquid. One microgram per deciliter would be about the same as ten parts per billion.

Mortality Rate

An estimate of the portion of a population that dies during a specified period. The rate is calculated by dividing the number of deaths due to a disease by the total population.

Natural Break Method

A means of breaking data points into a specified number of groups and then examining the means (averages) of the groups. The data are divided in such a way as to maximize the difference between categories' means while minimizing the range of each category.

Perinatal

Pertaining to or occurring in the period shortly before and after birth. In medical statistics, it is generally considered to begin with the completion of 28 weeks of gestation and variously defined as ending one to four weeks after birth.

Population-based Disease Registry

A systematic collection of information about people diagnosed with a specific type of disease who reside within a defined geographic region.

Population Density

Number of people per square mile or square kilometer.

Prevalence of Disease

The total number of disease cases that exist in a defined population during a specified time. This measure of disease helps researchers determine a person's likelihood of having a disease. **Prevalence Rate** is calculated by dividing the total number of cases of a disease existing in a population by the total population.

Protocol

The plan or set of steps to be followed in a study, investigation, or intervention program.

Race/Ethnicity

Race refers to distinct characteristics or traits that are transmitted through the offspring and originated from a common ancestor. Some diseases are strongly correlated with biological aspects of race.

Ethnicity refers to the characteristics of a group having a distinctive social and cultural tradition, a common history and origin, and often a common genetic heritage. These features may be reflected in their health and disease experience.

Rate vs. Prevalence

Rate is an expression of the frequency with which an event occurs in a defined population in a specified period of time. The use of rates rather than raw numbers, such as disease prevalence data, is essential for comparison of experience between populations at different times, different places, or among different classes of persons.

Relative Standard Error

A measure of statistical uncertainty or random variability around the rate estimate.

Risk Assessment

A process that estimates the likelihood that exposed people may have health effects. The four steps of a risk assessment are 1) Hazard Identification (Can this substance damage health?); 2) Dose-response Assessment (What dose causes what effect); 3) Exposure Assessment (How and how much do people contact it?); and 4) Risk Characterization (Combines the other three steps to characterize risk and describe the limitations and uncertainties).

State Environmental Quality Review Act (SEQR)

This act requires all state and local government agencies to consider environmental impacts equally with social and economic factors during discretionary decision-making. Agencies must assess the environmental significance of all actions they have discretion to approve, fund, or directly undertake. SEQR requires the agencies to balance the environmental impacts with social and economic factors when deciding to approve or undertake an action.

Statewide Planning and Research Cooperative System (SPARCS)

A database containing information on hospitalizations in all non-federally regulated acute care hospitals in New York State. Veteran's Affairs and mental health institutions are not included in SPARCS. The database contains over 120 data elements describing the patient, hospital, health care provider and the hospitalization stay.

Toxicology

The study of actual or potential danger presented by the harmful effects of chemicals (poisons) on living organisms and ecosystems, of the relationship of such harmful effects to exposure, and of the mechanisms of action, diagnosis, prevention and treatment of intoxications.

UA

Urbanized area – as defined by the U.S. Census Bureau, an urbanized area consists of contiguous, densely settled census block groups and census blocks that meet minimum population density requirements, along with adjacent densely settled census blocks that together encompass a population of at least 50,000 people.

UC

Urbanized cluster - as defined by the U.S. Census Bureau, an urbanized center consists of contiguous, densely settled census block groups and census blocks that meet minimum population density requirements, along with adjacent densely settled census blocks that together encompass a population of at least 2,500 people, but fewer than 25,000 people.

APPENDIX C

Appendix C

EVALUATION OF DATA SETS

1. New York State Cancer Registry

Under the New York State Public Health Law, all individuals diagnosed with cancer in New York State must be reported to the New York State Cancer Registry. Cancer reports are mainly received from hospitals, but are also received from pathology laboratories, physicians, and ambulatory care centers. The registry contains information both on the tumor (e.g., type of cancer, anatomic site, stage at diagnosis, cell type, etc.) and on the individual diagnosed with cancer (i.e., age, gender, residence, race, ethnicity, etc.). More information on the Cancer Registry is available at www.health.state.ny.us/nysdoh/cancer/nyscr/about.htm. Cancer incidence data are available from NYSDOH at the ZIP Code level for breast, lung, prostate, and colorectal cancers. The NYS cancer registry participates in the certification process of the North American Association of Central Cancer Registries (NAACCR). As part of this process the NYS Cancer Registry submits data annually to be evaluated for timeliness, completeness and quality. The cancer registry received the Gold level of certification, the highest level of certification, for 1998, 1999 and 2000 data. Currently, data for 1995 through 2000 meet or exceed the Silver level certification standards.

- 1) Timeliness: available less than 23 months after the end of the year. Rating = high
- 2) Completeness: the completeness of the cancer registry is estimated to be over 95% for 1995 through 2000, based on the North American Association of Central Cancer Registries incidence to mortality rate ratio method. (Completeness of data for years prior to 1995 has not been evaluated.) Case finding audits are routinely conducted on hospitals throughout New York State as part of standard ongoing registry operations. Fewer than 3% cases were found to be reported by death certificate only for the years 1997 to 2000. The percent of cases ascertained from death certificates only is less than 4% for the entire database. Rating = high
- 3) Accuracy: The cancer registry uses a standard set of edit checks for consistency, and more than 99% of cases from 1995 through 2000 pass these edit checks. The prevalence of unresolved duplicate reports on the registry database is fewer than 1 per 1000 for 1995 through 2000, and less than 3% of cases were missing key data fields in for 1995 through 2000. Rating = high
- 4) Consistency of the data across the state for the purpose of comparing health outcome rates between different areas: The cancer registry has standard procedures in place to assure consistency of reporting of data across the state. Reporting is mandated by law, and enforced in all hospitals (except Veterans Administration [VA] Hospitals) that diagnosis and treat cancer patients in the state. Although VA Hospitals are exempt from the state law mandating reporting, many report voluntarily. All non-VA hospitals that render services to 100 or more cancer patients per year are required to employ a Certified Tumor Registrar to abstract medical records for reporting purposes. Training in cancer registration is offered to all facilities routinely. Each reporting facility is assigned to a cancer registry field staff representative, who individually monitor facilities for completeness, timeliness and accuracy. Data received from reporting facilities are checked for accuracy as they are received, and feedback procedures are in place to assure ongoing quality improvement. Reporting audits are routinely conducted on a sample of reporting facilities across the state annually. Other sources of information

about completeness of reporting from individual hospitals that are used in monitoring include the percent of cases ascertained from death certificates only and the unduplicated number of cancer patients determined from the SPARCS inpatient and ambulatory surgery data. The cancer registry has interstate data exchange agreements with all surrounding states and many other states, so that individuals diagnosed and treated at out-of-state facilities are registered. There is currently no indication that completeness of reporting varies across the state. Completeness of individual data items may vary due to differences in medical care standards and demographics. For example, a slightly higher percent of cases are missing race information in New York City compared to the remainder of the state, probably due to the higher overall percentage of the case who are races other than white in New York City. The differences, however, are likely to be small and, in general, have little effect on the cancer rates.

2. Statewide Planning and Research Cooperative System (SPARCS)

SPARCS is a database that contains information about hospitalizations in New York State. All non-federally regulated acute care hospitals in the state are required to report selected medical, socioeconomic and financial information for each inpatient discharge from its facility. (Veteran's Affairs hospitals are not included.) The system was developed for financial purposes and for health care resource planning. The data file is updated monthly with the most current information available. SPARCS data for asthma hospitalizations are available by ZIP Code from the NYSDOH.

- 1) Timeliness: available less than one year after the end of the year. Rating = high
- 2) Completeness: statewide coverage but excludes Veterans Administration hospitals; processes in place to audit completeness of reporting by institution and to summarize findings periodically. The summary contains a listing of individual facilities currently experiencing delays in reporting as well as an estimate of the overall level of reporting for all of the data years available. Rating = high.
- 3) Accuracy: edit checks in place to determine whether data elements within a record are consistent within itself and checks on whether the distribution of reported values for certain data fields is reasonable when compared to the statewide distribution of values. In addition, the database has been compared with other databases reporting the same event, such as the Vital Records birth certificate file and the death certificate file. Related data fields from two data sources were compared for consistency. In addition, staff reviews the database for irregular patterns over time. Use of race/ethnicity from the SPARCS database is not advised. The NYCDHMH found SPARCS data on race and ethnicity of individual patients to be imprecise. The quality of the data varied depending on the reporting facility. Some facilities were more conscientious than others. Data regarding Hispanic origin were reported as "unknown" for 25% of asthma cases, and large numbers of records had race listed as "other."¹⁶
- 4) Consistency of the data across the state for the purpose of comparing health outcome rates between different areas: Review of hospitalization data is an important component of chronic disease surveillance, but people who are hospitalized with a chronic disease represent only a subset of people with the disease. Since most states have statewide hospital discharge databases, the data are relatively easy to obtain and are available for the entire state. However, factors such as access to health care, utilization of health care and issues of medical management can influence chronic disease hospitalization rates. These issues have received much attention in relation to

¹⁶ Asthma Facts. New York City Department of Health, 1999.

asthma because considerable variation in asthma hospitalization rates has been seen. Asthma hospitalization rates are known to be higher among poor and minority populations, and differences in access to medical care, source of care, and medical management may be important factors. This issue should be noted in the data analysis discussion when areas are compared that differ in the proportion of minorities and persons living in poverty.

3. Vital Records

NYSDOH processes data from live birth, death, fetal death and marriage certificates recorded in New York State exclusive of New York City. Through a cooperative agreement, NYSDOH receives data on live births, deaths, fetal deaths and marriages recorded in New York City and on live births and deaths recorded outside of New York State to residents of New York State. Currently, Vital Statistics information is available from NYSDOH by county. As Vital Records data become available at the ZIP Code level, the data will be reviewed for data quality.

4. Developmental Disabilities

Reporting Requirements

- 1) The Federal Individuals with Disabilities Education Act (IDEA) Part B requires States receiving Federal funds for this program to report annually the number of children aged 3 to 21 with disabilities receiving special education and related services. Counts are provided by race, ethnicity and disability category and are reported separately for preschool children (aged 3-5 years) and students aged 6-21 years.
- 2) IDEA Part C requires States to report annually the number of infants and toddlers (birth through 2 years of age) with disabilities who are receiving early intervention services. The data are reported by race and ethnicity.

Infants and Toddlers (birth through two years of age) Receiving Early Intervention Services.

Children in this age group who are found to have a significant delay or disability or to have a condition with a high probability of developmental delay are eligible to receive services under IDEA Part C through NYSDOH's Early Intervention Program. The program is administered through 57 counties and New York City. The program's most recent annual report indicates that 24,077 children in the state were receiving services as of December 1, 1999, for a population-based rate of 3.32%.¹⁷ The report also provides the number of children receiving services and the population-based rate by county; these data are not publicly available at a geographic level smaller than a county. Information is also provided for the developmental domains under which the children have been found eligible (e.g., communication, physical, cognitive, etc.) for New York State, New York City and New York State excluding New York City.

Children Ages 3-5 Receiving Educational Services for Disabilities. The New York State Education Department (NYSED) Office of Vocational and Educational Services for Individuals with Disabilities (VESID) is responsible for the Federal IDEA reporting requirements for preschoolers with disabilities. States report yearly the number of preschoolers with disabilities who are receiving special education services. Since the 1997 Reauthorization of IDEA, reporting is by race, ethnicity, and the following disability categories:

¹⁷ Early Intervention Annual Report for July 1, 1999 – June 30, 2000 Program Year. Children Served and Services Authorized. New York State Early Intervention Coordinating Council, New York State Department of Health.

Mental retardation
Hearing impairments including deafness
Speech or language impairments
Visual impairments including blindness
Emotional disturbance
Orthopedic impairments
Autism

Traumatic brain injury
Other health impairments
Multiple disabilities
Deaf-blindness
Specific learning disabilities
Developmental delay

Although data are available by request from VESID for preschoolers, this report is focusing on data available on students aged 6-21 years (school-age group) since the preschool population of students receiving special education services is small compared to the school-age population of students receiving services. (As of December 1, 2000, there were 34,492 preschoolers and 406,841 students aged 6-21 years in New York State receiving special education programs and services: <http://www.vesid.nysed.gov/sedcar/goal2data.htm>).

Students Ages 6-21 Receiving Educational Services for Disabilities. NYSED is also responsible for the Federal IDEA reporting requirements for this age group. These data are available from NYSED by school district, which is a geographic level larger than a ZIP Code. The Federal government requires that states report yearly the number of children with disabilities who are receiving special education services, by race, ethnicity, and disability category. The disability categories for this age group are:

Mental retardation
Hearing impairments including deafness
Speech or language impairments
Visual impairments including blindness
Emotional disturbance
Orthopedic impairments
Autism

Traumatic brain injury
Other health impairments
Multiple disabilities
Deaf-blindness
Specific learning disabilities
Developmental delay (optional category for children 6-9 years old)

VESID produces a report each year that is available by request entitled "Report of Students with Disabilities Provided Special Education as of December 1st. Disability Category Total for All Public Schools by District and County" (PD-1/4). The report gives the number of students receiving special education services by disability category (not including developmental delay) for each school district and county. Summary tables of the number of children receiving special education programs and services statewide by disability category and age group are available at <http://www.vesid.nysed.gov/sedcar/goal2data.htm>. In addition, VESID produces annually the "Performance Report of Educational and Vocational Services and Results for Individuals with Disabilities." Table 1 of Volume 2 provides the number of preschool and school-age students with disabilities who were provided special education services and were resident to each school district as of December 1st, as well as a classification rate for each district. Copies of the performance reports are available from the NYSED Office of Vocational and Educational Services for Individuals with Disabilities by e-mailing VESIDCAR@mail.nysed.gov.

There is interest in viewing the percentage of students receiving special education services (special education classification rate) as an estimate of the prevalence of developmental disabilities. Variation in this estimate among different geographic areas reflects not only true differences in prevalence, but also differences in how school districts identify, record and provide services for disabilities in different areas of the state. It is known that classification rates vary widely from state to state and among the school districts of New York State. NYSED monitors classification rates and studies factors related to differences in these rates; some of

these factors are discussed under #3 and #4 below. NYSED has been identifying school districts with high rates of identification of students with disabilities as part of the provisions of Chapter 405 of the Laws of 1999 (<http://www.vesid.nysed.gov/specialed/chapter405/home.html>) and has been working with these districts to determine underlying causes and develop corrective action plans. The difference between the special education classification rate and the true prevalence of developmental disabilities is unknown and is the subject of on-going research through the National Center on Birth Defects and Developmental Disabilities of the CDC (<http://www.cdc.gov/ncbddd/ddlist.htm>) and the National Institutes of Child Health and Human Development (<http://www.nichd.nih.gov/>).

Factors relevant to comparing classification rates from place to place.

- 1) The process to determine a student's eligibility for special education services is by definition highly individualized; as a result, it is reasonable to expect variation among school districts. By state law, the determination of whether a student receives special education services is made at the school district level by the Committee on Special Education (CSE) for school-age children or the Committee on Preschool Special Education (CPSE) for pre-school children. There are 687 school districts in New York State. The first step in the process is an individual evaluation that is conducted by a multi-disciplinary team to comprehensively identify the student's present level of educational performance and the extent of special education needs. Federal and State regulations require that school districts have established procedures for conducting an individual evaluation. The specific assessments or tests to be included in an individual evaluation must be chosen on an individual basis, and there is considerable emphasis in the guidance documents¹⁸ on selecting the evaluations appropriate for each individual student. In order to find a student eligible for special education services, the multi-disciplinary team must decide that a student has a disability that affects his/her ability to learn. Local norms will affect the level of performance or behavior that is considered acceptable. The team also determines a disability classification. Many students have more than one impairment. This decision is frequently not a clear-cut choice given the complexity of the process, and a judgment by the team is required. For example, a student with emotional disturbance may also have speech impairment, and a significant proportion of autistic students are also affected by mental retardation.

It has long been recognized that states differ considerably in the percentage of students assigned to the various disability categories, and it is also acknowledged that school districts within states differ in classification patterns. However, there have been few studies that document these differences. Social variables such as ethnicity and language proficiency, fiscal variables such as wealth of the school district and per pupil expenditure rates, and migration of families of children with disabilities to areas with medical facilities have all been discussed as factors affecting the variability of classification rates.¹⁹

- 2) The disability categories can be broken down into general groupings. The sensory disabilities include visual impairments, hearing impairments including deafness, and deaf-blindness. Physical and neurological disabilities include orthopedic impairments,

¹⁸ Individual Evaluations and Eligibility Determinations for Students with Disabilities. Office of Vocational and Educational Services for Individuals with Disabilities, New York State Education Department, August 2000.

¹⁹ Special Education: Study of Incidence of Disabilities. Report to the California Department of Education. American Institutes for Research, 1998. <http://csef.air.org/publications/related/incidence.pdf>.

traumatic brain injury, multiple disabilities and autism. Together these disabilities represent about 10% of all children served under IDEA and are sometimes referred to as “low-incidence” disabilities. In general, these students are identified through medical history, physical examination, speech and hearing tests, and neurological tests,²⁰ and many are identified before they reach school age. Four categories (specific learning disabilities, speech or language impairments, mental retardation and emotional disturbance) account for about 90% of students served by IDEA and are sometimes referred to as “high-incidence” disabilities. These students are likely to be referred for assessment because of difficulties with school performance or social interaction. There is considerable discussion in the literature about greater subjectivity in the identification process for high-incidence disabilities. The category of specific learning disabilities constitutes the largest proportion of students receiving special education services (about 50% nationally and in New York State in 2000). There is wide variation in the classification rate for this category and the other high-incidence disabilities among states and also among school districts within a state. Concern about the lack of consistently applied diagnostic criteria for the category of specific learning disabilities has been discussed in the literature.²⁰

- 3) Although it is generally acknowledged that poverty has an effect on special education classification rates, there is considerable debate about the role that poverty plays. NYSED states that some school districts “serve exceptional numbers of educationally disadvantaged children ... who have been placed at risk by poverty, homelessness, poor nutrition or inadequate care...” (p. 66)²¹ NYSED has developed a Need/Resource Capacity (N/RC) index that is used to categorize school districts into similar groups based on their ability to meet the needs of their students with local resources. The index takes into account indicators of relative district wealth, population density, and student needs assessed partly on estimates of poverty.^{22 23} NYSED uses the following categories to present data on resources, student performance and attendance, as well as special education classification rates:²¹ New York City, Large City Districts (Buffalo, Rochester, Syracuse, Yonkers), High Need Relative to Resource Capacity (N/RC) Districts Urban-Suburban, High N/RC Districts-Rural, Average N/RC Districts, and Low N/RC Districts. (New York City’s annual school reports present similar school group comparison statistics based on quintiles.) New York City and the Large City Districts are treated as separate groups because of the large number of students they serve and because of the special challenges associated with these large urban districts (p.66).²¹ In New York State, classification rates in High N/RC school districts tend to be higher than those in Average N/RC school districts, which tend to be higher than those in Low N/RC school districts (p. VI.84).²²

²⁰ Assessment and Identification. In *A New Era: Revitalizing Special Education for Children and Their Families*. President’s Commission on Excellence in Special Education, 2002.

²¹ *New York: the State of Learning. A Report to the Governor and the Legislature on the Educational Status of the State’s Schools. Statewide Profile of the Educational System.* The University of the State of New York/The State Education Department, June 2001.

²² “Performance Report of Educational and Vocational Services and Results for Individuals with Disabilities, 1999-2000, Volume 2.” Office of Vocational and Educational Services for Individuals with Disabilities, New York State Education Department.

²³ “What is a Similar School?” <http://www.emsc.nysed.gov/repcrd2003/information/similar-schools/guide.html>

- 4) A large body of literature since the 1970s focuses on the disproportionate representation of minorities in special education.^{24 25} The 1997 Reauthorization of IDEA required states to collect data regarding the race/ethnicity of students with disabilities to determine whether “significant disproportionality” exists for overall identification of students as students with disabilities, identification of students as having particular types of disabilities, and placement of students with disabilities in separate or restrictive educational settings (p. III.16).²² Nationally, for the 1999-2000 school year, American Indian/Alaska Native students and Black students exceeded their representation as students with disabilities compared to the resident population (p. II-27).²⁶ The greatest disparities are seen in the categories of mental retardation and emotional disturbance.

In New York State, Chapter 405 of the Laws of 1999 requires NYSED to examine special education data in relation to a number of key areas including disproportionate identification and placement of minority students in special education. In New York State, American Indian/Alaskan Native, Black, and Hispanic students were identified as students with disabilities in greater proportions than their representation in the total school population (as of December 1, 2000).²⁷ The percentage of White and Asian/Pacific Islander students with disabilities was lower than their percentage in the total public school enrollment. In a 2002 report NYSED noted that the largest numbers of racially and ethnically diverse students are concentrated in high need school districts.²⁸ The report states that these students enter school districts where adequate support services in general education are rarely available, greater numbers of teachers are uncertified, and the lack of resources makes it more difficult to provide quality instruction and early intervention for these students. NYSED has identified these areas as requiring special attention in addressing disproportionality: 1) disproportionate placement of students in special education is a result of lack of supports and services in the general education environment; 2) research-based instruction must be available to all students who are struggling to learn to read; and 3) the limited availability of a comprehensive special education continuum within high need districts has a direct impact on disproportionality, especially as it relates to placement in separate sites. In addition, NYSED has been identifying school districts with the greatest need of support in relation to disproportionality and has been working with some of these districts to determine underlying causes and develop corrective action plans.

Summary of quality of the NYSED data on the number of children 6 – 21 years old with disabilities who are receiving special education services

- 1) Timeliness: data reflecting special education placements by school district as of December 1 of each calendar year are available by request to the NYSED Office of Vocational and Educational Services for Individuals with Disabilities less than two years from that date. Rating = high.

²⁴ Minority Students in Special and Gifted Education. Committee on Minority Representation in Special Education. M.S. Donovan and C.J. Cross, Eds., National Research Council, 2002.

²⁵ Racial Inequity in Special Education. D.J. Losen and G. Orfield, Eds., The Civil Rights Project at Harvard University, Harvard Education Press, 2002.

²⁶ To Assure the Free Appropriate Public Education of All Children with Disabilities: Individuals with Disabilities Education Act, Section 618. 23rd Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act. U.S. Department of Education, 2001.

²⁷ 2002 Pocketbook of Goals and Results for Individuals with Disabilities.

<http://www.vesid.nysed.gov/pocketbook/2002/>

²⁸ Addressing Racial and Ethnic Disproportionality in Special Education.

<http://www.vesid.nysed.gov/specialed/publications/regents/1202raciaethnic.pdf>

- 2) Completeness: statewide coverage. School districts report yearly to NYSED either by hard copy or electronically through the NYSED web site. If all of the required reports are not received, repeated notices are sent to the contact person for the school district. Rating = high.
- 3) Accuracy: For the hard copy forms there are specific instructions and edit checks for verification of the data. For electronic data submission, specific instructions, help files, and edit checks are built in. Verification and error reports are provided online as data are entered. Before the data are finalized, each district is required to verify that the district's data are accurately recorded on the web site. Ongoing technical assistance is available to school districts.
- 4) Consistency of the data across the state for the purpose of comparing different areas to one another: For the data to be useful for this purpose, consistency is needed from district to district in terms of the identification of students with disabilities and the disability categories to which the students are assigned, which can also be referred to as the validity of the disability classification. In recent years, NYSED has made changes in the manner in which school districts are monitored. A collaborative Quality Assurance Review process has been implemented, and the issues identified under Chapter 405, which include disproportion in the identification of students as students with disabilities or in the identification of students having a particular disability, are included in this process. Although NYSED is working to determine the underlying causes of these problems with individual school districts, at this time the use of the special education data to compare rates from one area to another would be problematic. In particular, examining individual disability categories among different school districts may not provide meaningful comparisons. Determination of a student's disability category is a complex and highly individualized process. School districts have a great degree of flexibility in assigning students to a disability category; in most cases there is not a single clear-cut choice and a judgment is made by the team.

A study from Atlanta provides an example of the problem. In 1991 the U.S. Centers for Disease Control and Prevention (CDC) established the Metropolitan Atlanta Developmental Disabilities Surveillance Program (MADDSP), an ongoing system for monitoring the occurrence of some developmental disabilities (<http://www.cdc.gov/ncbddd/dd/ddsurv.htm>). The MADDSP uses the data reported under IDEA by public school districts as well as information from pediatric hospitals, clinics, diagnostic and evaluation centers and private physicians. In a recently released report on the prevalence of autism in Metropolitan Atlanta, public school district special education records were found to be a primary source for case identification.²⁹ However, for the students identified by MADDSP as having autism who received special education services, autism was the primary eligibility category for only 41%, while 59% were assigned to other categories. CDC has begun funding states to adopt the model developed in Atlanta to conduct surveillance of developmental disabilities in other states as part of the Autism and Developmental Disabilities Monitoring Network (ADDM Net). Through this network more information will be gained on the prevalence of developmental disabilities and the relationship of classification rates to the prevalence of disabilities. It may be possible in the future to evaluate more thoroughly the usefulness of the data on students receiving special education services for the purposes outlined in this report.

²⁹ Yeargin-Allsopp, M. et al. Prevalence of autism in a US Metropolitan Area. JAMA 289:49-55, 2003.

NYSDOH has been involved in a multi-year effort to produce clinical practice guidelines to support the Early Intervention Program. In 1999 the Clinical Practice Guideline for Autism/Pervasive Developmental Disorders was published as a tool to help assure that infants and young children with these disorders receive early intervention services consistent with their individual needs, resources priorities and the concerns of their families.

APPENDIX D

Figure 1. Map showing the boundaries of the City of Troy and of ZIP Codes in the city.

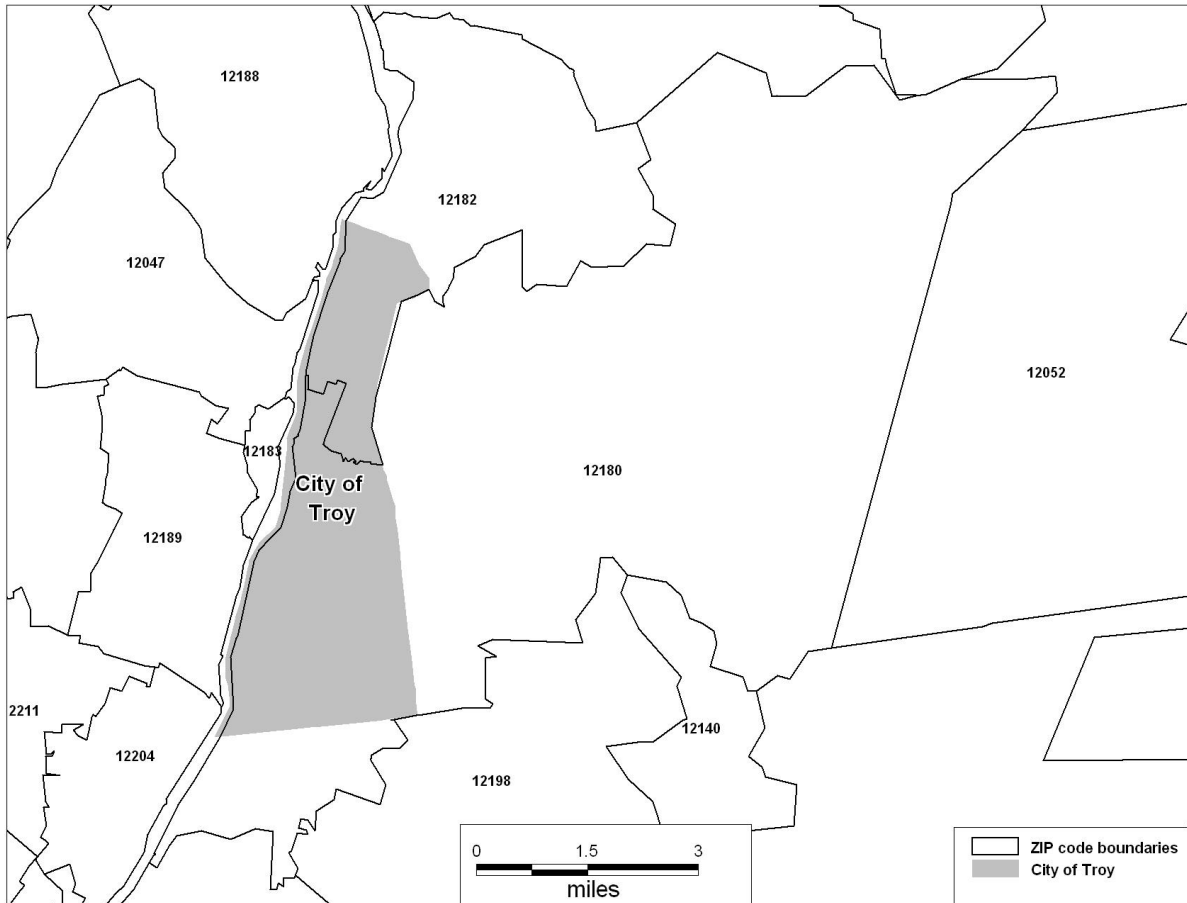
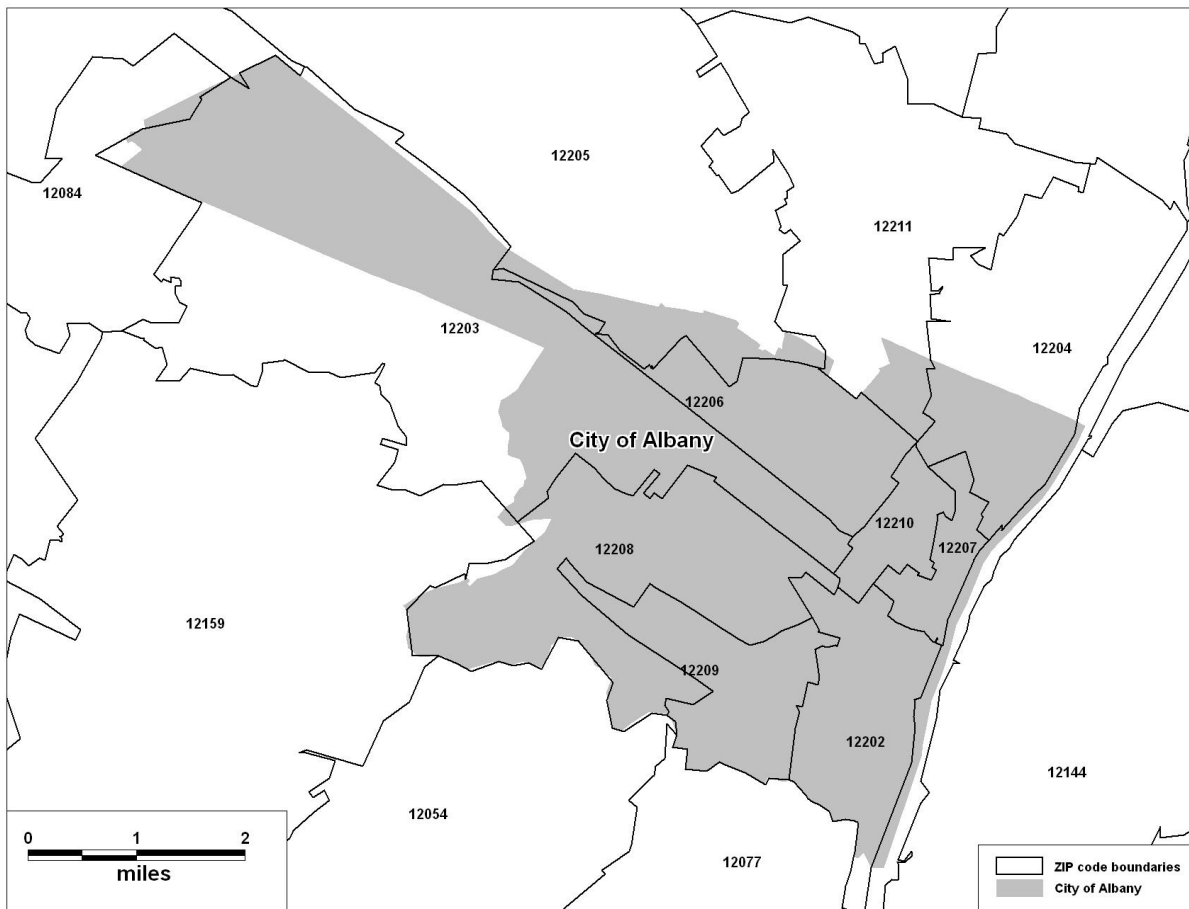


Figure 2. Map showing the boundaries of the City of Albany and the ZIP Codes in the city.



APPENDIX E

Appendix E

METHODS FOR CREATING POPULATION DENSITY GROUPINGS

The HOD work group explored several options for classifying ZIP Codes to create population density groupings. In addition to the calculation of population density that is discussed in Section IV.A of this report, methods described by the Behavioral Risk Factor Surveillance System, by the U.S. Environmental Protection Agency, and by the U.S. Census for urban/rural classification were also explored. These methods are described briefly below.

1) The Behavioral Risk Factor Surveillance System (BRFSS) is a program of the U.S. Centers for Disease Control and Prevention (CDC) that operates in all 50 states. States use a standard core questionnaire and standard procedures to collect data on health-related behaviors through a series of monthly telephone interviews with adults. Data can be compared across states. In New York, the program is administered by the Chronic Disease and Risk Factor Surveillance Unit of NYSDOH's Division of Chronic Disease Prevention and Adult Health. In order to generate county-specific prevalence estimates based on the state data, counties were grouped by demographic and socio-economic population characteristics into seven county clusters. The methodology and county cluster listings below are from the report "County Level Prevalence Estimates: New York State Behavioral Risk Factor Surveillance System."³⁰

- Cluster 1: Allegany, Cattaraugus, Cayuga, Chautauqua, Chenango, Columbia, Delaware, Franklin, Genesee, Herkimer, Lewis, Livingston, Madison, Montgomery, Orleans, Otsego, Schoharie, Schuyler, Seneca, Steuben, Washington, Wayne, Wyoming, Yates
- Cluster 2: Chemung, Clinton, Cortland, Essex, Fulton, Greene, Hamilton, Jefferson, Oneida, Ontario, Oswego, Rensselaer, Saratoga, St. Lawrence, Sullivan, Tioga, Tompkins, Ulster
- Cluster 3: Albany, Broome, Erie, Monroe, Niagara, Onondaga, Schenectady, Warren
- Cluster 4: Dutchess, Orange, Putnam, Richmond, Rockland, Suffolk
- Cluster 5: Bronx, Kings, Queens
- Cluster 6: Nassau, Westchester
- Cluster 7: New York

Because this system does not extend to a geographic level smaller than a county, it was not explored further.

2) For dispersion modeling purposes, the U.S. Environmental Protection Agency (USEPA) has categorized every census tract and county in the contiguous United States as urban or rural. A county is considered urban if, based on 1990 census, either it includes a metropolitan statistical area with a population greater than 250,000 or the U.S. Census Bureau designates more than 50% of the population as "urban." A county is considered rural if, based on 1990 census, it did not contain a metropolitan statistical area with a population greater than 250,000 or the U.S. Census Bureau did not designate more than 50% of the population as "urban."

³⁰ County Level Prevalence Estimates: New York State Behavioral Risk Factor Surveillance System. Chronic Disease and Risk Factor Surveillance Unit, Bureau of Chronic Disease Epidemiology and Surveillance, New York State Department of Health, December 2000.

Rural Counties	Urban Counties
Allegany	Albany
Cattaraugus	Bronx
Chautauqua	Broome
Chemung	Cayuga
Chenango	Dutchess
Clinton	Erie
Columbia	Genesee
Cortland	Herkimer
Delaware	Kings
Essex	Livingston
Franklin	Madison
Fulton	Monroe
Greene	Montgomery
Hamilton	Nassau
Jefferson	New York
Lewis	Niagara
Otsego	Oneida
St. Lawrence	Onondaga
Schuyler	Ontario
Seneca	Orange
Steuben	Orleans
Sullivan	Oswego
Tompkins	Putnam
Ulster	Queens
Warren	Rensselaer
Washington	Richmond
Wyoming	Rockland
Yates	Saratoga
	Schenectady
	Schoharie
	Suffolk
	Tioga

	Wayne
	Westchester

This method is similar to that of the U.S. Census Bureau but involves census tracts rather than ZIP Codes. Since the health outcome data are tabulated by ZIP Code, the U.S. Census Bureau method was explored instead.

U.S. Census 2000 classifies as “urban” all territory, population, and housing units located within an urbanized area (UA) or an urban cluster (UC). It delineates UA and UC boundaries to encompass densely settled territory, which consists of 1) core census block groups or blocks that have a population density of at least 1,000 people per square mile and 2) surrounding census blocks that have an overall density of at least 500 people per square mile. In addition, under certain conditions, less densely settled territory may be part of a UA or UC. UAs encompass a population of at least 50,000 people, while UCs encompass a population of at least 2,500 but fewer than 50,000 people (http://www.census.gov/geo/www/ua/ua_2k.html).

The Census classification of “rural” is all territory, population, and housing units located outside of UAs and UCs. Rural territory is classified as farm or nonfarm. Geographical entities, such as census tracts, counties, metropolitan areas, and the territory outside metropolitan areas, often are “split” between urban and rural territory, and the population they contain often are partly classified as urban and partly classified as rural.

There are 16 UAs in New York State, but two are areas where UAs in Danbury, CT and Bridgeport-Stamford, CT overlap into New York. A state sorted list of UAs with land area and population density can be found at <http://www.census.gov/geo/www/ua/st2kua.txt>. There are 116 UCs in New York state. A state sorted list of UCs with land area and population density can be found at <http://www.census.gov/geo/www/ua/st2kuc.txt>.

ZIP Code Tabulation Areas (ZCTAs) are a statistical entity developed by the U.S. Census Bureau for tabulating summary statistics from Census 2000. This new entity was developed to overcome the difficulties in precisely defining the land area covered by each ZIP Code. Defining the extent of an area is necessary to accurately tabulate census data for that area. ZCTAs are generalized area representations of U.S. Postal Service (USPS) ZIP Code service areas. Simply put, each ZCTA is built by aggregating the Census 2000 blocks, whose addresses use a given ZIP Code, into a ZCTA which gets that ZIP Code assigned as its ZCTA code. They represent the majority USPS five-digit ZIP Code found in a given area. For those areas where it is difficult to determine the prevailing five-digit ZIP Code, the higher-level three-digit ZIP Code is used for the ZCTA code. More information in ZCTAs is available at <http://www.census.gov/geo/ZCTA/zcta.html>.

U.S. Census file SF3 contains the number of people residing in a UA, in a UC, in rural farm territory and in rural nonfarm territory (American FactFinder, SF3, Table P5 Urban/Rural). These data can be tabulated by 5-digit ZCTA; therefore, the percentage of each ZCTA’s population residing in each of these types of territory can be calculated. For the two examples described in this report, an attempt was made to group ZCTAs into population density groupings by this scheme. The method yielded groupings that were much cruder than those achieved by the calculation of population density for ZIP Codes described in Section IV.A of the report. In one example almost the entire population of the county was in a UA; in the other example, almost the entire population of the county was in rural territory.

APPENDIX F

(link to “About Age Adjusted Rates, 95% Confidence Intervals and Unstable Rates”
<http://www.health.state.ny.us/nysdoh/cancer/nyscr/age.htm>)

