

# STATISTICAL NEWS

PA Department of Health ♦ Bureau of Health Statistics and Research ♦ Vol. 28 No. 1 ♦ January 2005

## New County Assessment Modules Enhance EpiQMS

*Provides More Comprehensive Tool for Analyzing Health Statistics*

The latest enhancement to EpiQMS, the Bureau's interactive health statistics web tool, includes the addition of county assessment modules for the Death, Birth, Cancer Incidence, and Behavioral Risk Factors Surveillance System (BRFSS) datasets. The addition of these modules provides users with an even more comprehensive tool for analyzing health statistics online.

EpiQMS (Epidemiologic Query and Mapping System) was added to our web pages in September 2003. The EpiQMS web tool allows users to create customized data tables, charts, maps, county profiles, and county assessments of birth, death, cancer incidence, population, and BRFSS statistics online.

Within the EpiQMS county assessment modules, users will be able to create data tables of birth, death, cancer incidence, or BRFSS statistics that show the results of significance testing (at the 95% confidence level) for a specific county compared to the state. Please note that under the BRFSS dataset the module is called regional assessment since only some data are available for

...users will be able to create data tables... that show the results of significance testing... for a specific county compared to the state.

single counties. Most of the sub-state level BRFSS data are available for groups of 2-3 neighboring counties. This was necessary because the BRFSS data are sample-based and, therefore, less populated neighboring counties need to be grouped together in order to obtain a sample size large enough to produce reliable statistics.

In the EpiQMS county assessment screen, users will be presented with a series of menus along with instructions/steps to follow in order to create data tables that show results of significance testing at the 95% confidence level. The menus and menu selections may vary for each dataset. The data items listed within the data tables produced may also vary for each dataset.

*Go to Page 3 or click here...*

## Retailers Continue to Deny Cigarettes to Minors

*Federally-Mandated Synar Survey Tracks Attempts by Minors to Buy*

Retailers are selling fewer cigarettes to minors. According to a recent study conducted by the Department of Health, it was estimated that only 7% ( $\pm 2\%$ ) of the 14,796 Pennsylvania outlets licensed to sell cigarettes sold cigarettes to children under 18. The survey did not reveal a statistically significant difference between the 2003 and 2004 estimates but those estimates were significantly lower than the estimates measured between 1996 and 2002. Chart 1 on page 4 shows the 95% confidence intervals for the statewide violation rate from 1996 to 2004.

These results are based on the federally-mandated Synar survey. The survey protocol consisted of an under-age person entering an outlet to attempt to buy cigarettes. The survey was conducted by male and female inspectors between the ages of 14 and 17. The gender breakdown of the youth inspectors was 56% (646) male and 44% (508) female. Almost half (567 or 49%) of the inspectors were 15 year-olds, 47% (543) 16 year-old inspectors, 3% (35) 17 year-olds, and less than 1% (9) were 14 year-old inspectors.

...only 7% of the... outlets licensed to sell cigarettes sold...to children under 18.

### INSIDE THIS ISSUE

What is Geocoding and How Does it Work? ... 2

Web Site Updates ..... 3

HP2010 Objectives: Hospitalization Rate for Pediatric Asthma ..... 7

DEPARTMENT OF HEALTH

Edward G. Rendell, Governor  
Calvin B. Johnson, M.D., M.P.H.  
Secretary of Health

*Go to Page 4 or click here...*

# What Is Geocoding And How Does It Work?

## Part 1 of a Five Part Geocoding Series Appearing in Statistical News

**G**eocoding is the process of taking information about a location and assigning latitude and longitude coordinates to it. This can be accomplished by converting a table of addresses to points with latitude and longitude or by recording the geographic coordinates from a Global Positioning System (GPS) unit. Once a latitude and longitude has been assigned, points can be placed on a map that represents these locations and different methods of spatial analysis can be performed.

### Background:

Before Geographic Information Systems (GIS) and hand held GPS units, colored pins were pushed into paper maps. This was the way that different locations were plotted. The military would push pins into maps that signified their own battalions or enemy troops' locations. The police would push pins into maps showing crime or accident scenes. This all changed with the growth of technology. Today, most maps are created and geocoded locations are plotted electronically on a computer or an electronic device.

All of us, at one time or another, have most likely taken a local address, pulled out a city streets map, found the name of the street on the map, hopped in the car, drove to the street, and found the address by watching the house numbers as we drove up and down the street. We learned that all of the odd house numbers are on one side of the street and even numbers are on the other side. The numbers increase or decrease depending on the direction that we move up

or down the street. This example shows that most people can understand the verbal description of an address and how it relates to locations on the earth's surface. However, a computer cannot do this. In order for a computer to display locations on a map, it must be given some type of geometric representation for these address locations, such as points containing latitude and longitude coordinates. This is why geocoding is important.

There are three basic methods for geocoding addresses or assigning geographic coordinates to a location:

- Address Matching
- ZIP Code Centroid Plotting, or
- Global Positioning Systems (GPS) Reading.

### Address Matching:

Geocoding using address matching methods requires specialized software and a streets reference file. The Pennsylvania Department of Health has recently used two different software applications to do address matching—ESRI's ArcGIS and Matchmaker. The ArcGIS software that is currently used takes a table of addresses, parses out the four components of an address (shown in Figure 1), takes these address components along with the ZIP code, and attempts to match the same address components to a streets reference file. The streets

---

...points (for addresses) can be placed on a map...and... spatial analysis can be performed.

---

reference file contains actual lines depicting streets, roads, or highways and a data table that houses information about those lines, such as beginning and ending address numbers on each side of a street, street name, street direction, street type, and where this street segment is located on the earth's surface. There are multiple suppliers of street reference files including: the US Census Bureau (TIGER files), ESRI (StreetMapUSA extension), and Geographic Data Technologies (GDT files). The Pennsylvania Department of Health currently uses GDT files to perform geocoding because GDT files are considered to be one of the most complete and accurate sets of street and boundary files in the GIS industry.

After the software attempts to match the address to a street segment in the reference file, it estimates an address location by evenly distributing all possible street numbers across a street segment and places a point in the location that corresponds to the address number. Figure 2 (page 6) shows a fictitious segment of

Main St between 1<sup>st</sup> Avenue and 2<sup>nd</sup> Avenue with street numbers ranging from 1 to 100. Odd numbers are on one side of the street and the even numbers on the opposite side of the street. If we were to geocode 50 Main St in this ZIP code area, a point would be placed in the exact middle of the street segment on the side of the street containing the even street numbers (2-100). This is an estimated location, not an exact physical location for an address. Although there is some inaccuracy, this method of geocoding provides a quick and efficient way to approximate thousands of address locations in a short period of time. This method of address matching is called address interpolation.

There is a second method of address matching called intersection matching. This method of geocoding is probably the most simple. If an address file contains an intersection of two streets, like 1<sup>st</sup> Avenue and Main Street, a geocoding service will simply place a point on the node where these two streets intersect as shown in Figure 3. In other words, it would place a point at the intersection of 1<sup>st</sup> Avenue and Main Street.

---

(Address interpolation or geocoding) is an estimated location, not an exact physical location for an address.

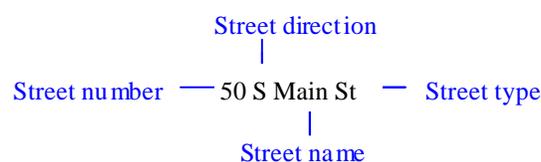
---

### ZIP Code Centroid:

Another method of geocoding a location is known as plotting the

*Go to Page 6 or click here...*

Figure 1



---

# Web Site Updates:

## *Cancer Facts and Figures for African Americans, 2002 Cancer Incidence Statistics, 2003 Local Behavioral Health Risk Data*

### ***Cancer Facts & Figures for African Americans:***

A special report entitled *Cancer Facts and Figures for African Americans Pennsylvania 2004* is now available on the Department of Health web site. The overall findings are summarized in the Data Highlights section.

This informative report not only contains projections for the number of cancer cases and deaths by sex and by major site (colorectal, breast, lung, and prostate) for 2004 but also includes trend data for cases and deaths by sex and race. Pennsylvania trends for African Americans were compared to the United States trends based on the National Cancer Institute Surveillance, Epidemiology, and End Result (SEER) Program. Additional sections include discussions of stage data and behavioral risk factors.

The Bureau of Health Statistics and Research was inspired to generate this report by the American Cancer Society's publication *Cancer Facts and Figures for African Americans*. Although the basic concepts were similar, this report focuses on Pennsylvania cancer statistics in more detail. In particular, race disparities were identified by the type of cancer, stage of disease at diagnosis, and behavioral risk factors.

To access this new report, go to [www.health.state.pa.us/stats/](http://www.health.state.pa.us/stats/) and select **Cancer Incidence and Mortality**.

### ***2002 Cancer Incidence Statistics Available:***

A large volume and variety of cancer incidence data tables, including statistics for all counties and municipalities in Pennsylvania have been updated with 2002

data on the Health Statistics web pages at [www.health.state.pa.us/stats/](http://www.health.state.pa.us/stats/). To access the data online, click on **Cancer Incidence and Mortality** and then select **Cancer Incidence and Mortality Statistics 1990-2003**.

Each year when a new cancer incidence data file is finalized and released, staff in the Bureau of Health Statistics and Research create and update thousands of pages of numerous crosstabulations that are used to respond to the hundreds of data requests we receive every year.

The 2002 Adobe PDF files have been added to the Health Statistics web pages and provide all data users with an abundant variety of cancer incidence and mortality statistics. Cancer data users can now easily access the latest available single (2002) and five-year (1998-2002) summary statistics online.

### ***2003 Local Behavioral Health Risk Data:***

The 2003 Local Behavioral Health Risk Data web page has been updated to include the results of locally added questions selected by the 2003 Pennsylvania Behavioral Risk Factor Surveillance System (BRFSS) Over Sampling Program participants for inclusion in their county surveys. These surveys were conducted via telephone in eight specific counties selected to be sampled by program participants.

Some of the topics chosen by the program participants include oral health, women's health, influenza, adult asthma, childhood asthma, tobacco indicators, heart attack and stroke, and many other topics. To access these reports online go to [www.health.state.pa.us/stats](http://www.health.state.pa.us/stats) and select **Behavioral Risk Data (BRFSS)**.

---

*Continued from Page 1...*

## County Assessments Added to EpiQMS

Through the menu selections, users will be able to output all rates/ratios/percentages that are significantly higher, significantly lower or not significantly different than the corresponding state statistic.

Data can easily be downloaded from EpiQMS by highlighting and then using copy and paste. All data rows are column-formatted for easy insertion into most spreadsheet software packages.

An additional enhancement

to the assessment modules planned for later this year is to include a link to a trend line for each statistic that is output. This trend line will be based on the three-year summary periods available through EpiQMS.

The next enhancement to EpiQMS will be the addition of municipality level data for birth, death, and cancer incidence datasets. The output at the municipality level will be limited to data tables and will be available for single years back to 1990.

The output will also be limited to birth, death, cancer incidence or population counts. Rates will not be calculated or displayed at the municipality level. Municipality level data are currently only available for the population dataset. However, the work involved with adding municipality level data to EpiQMS has already begun and will be available to EpiQMS users in the very near future.

To access EpiQMS, go to the Bureau of Health Statistics

and Research web pages at: [www.health.state.pa.us/stats](http://www.health.state.pa.us/stats) and click on the **EpiQMS** logo. Step-by-step instructions for each screen and detailed help sections were added to assist users with understanding the system.

Please contact the Bureau of Health Statistics and Research at 717-783-2548 for questions about EpiQMS. Staff are also available to provide EpiQMS presentations/demonstrations at your organization or professional meetings.

# Retailers Continue to Deny Cigarettes to Minors

The line in Chart 1 (below) indicating the 20 percent level of noncompliance is highlighted because states which fail to remain below that line become liable for substantial penalties assessed against their federal funds for substance abuse treatment.

Outlets were selected from a sampling frame created from the Department of Revenue's cigarette license file. The license file contained the name and address of every outlet that purchased a license to sell cigarettes. These surveys have been conducted each year since 1996. The

**...the most prevalent type (of outlet selling cigarettes) was the convenience store.**

2004 survey was conducted from July 1<sup>st</sup> to August 31<sup>st</sup>.

Of the 1,652 outlets contained in the Synar sample, 1,154 were inspected, eligible, and did not serve alcohol. The types of outlets selling cigarettes varied from beer distributors to tobacco outlets, but the most prevalent type was the convenience store. It was estimated that convenience stores, gas stations, grocery/produce stores, drug stores, and supermarkets comprise 72% ( $\pm 3\%$ ) of the statewide cigarette outlet types. On the next page, Chart 2 displays the distribution along with the 95% confidence interval (vertical bar).

Fast food establishments, such as delis, pizza shops, bagel shops, and donut/coffee shops were estimated to have a significantly higher violation rate than beer distributors, drug stores,

**The survey results showed that most retailers are displaying warning signs...that they do not sell to persons under the age of 18.**

grocery/produce stores, supermarkets, and tobacco outlets but were not significantly different from the rest of the outlet types. Chart 3 displays the violation rate of the various types of outlets.

Outlets that served alcohol by the glass were defined as ineligible for the 2004 survey. It was estimated that 7% ( $\pm 2\%$ ) of accessible outlets (i.e. accessible to minors) licensed to sell cigarettes also serve alcohol by the glass.

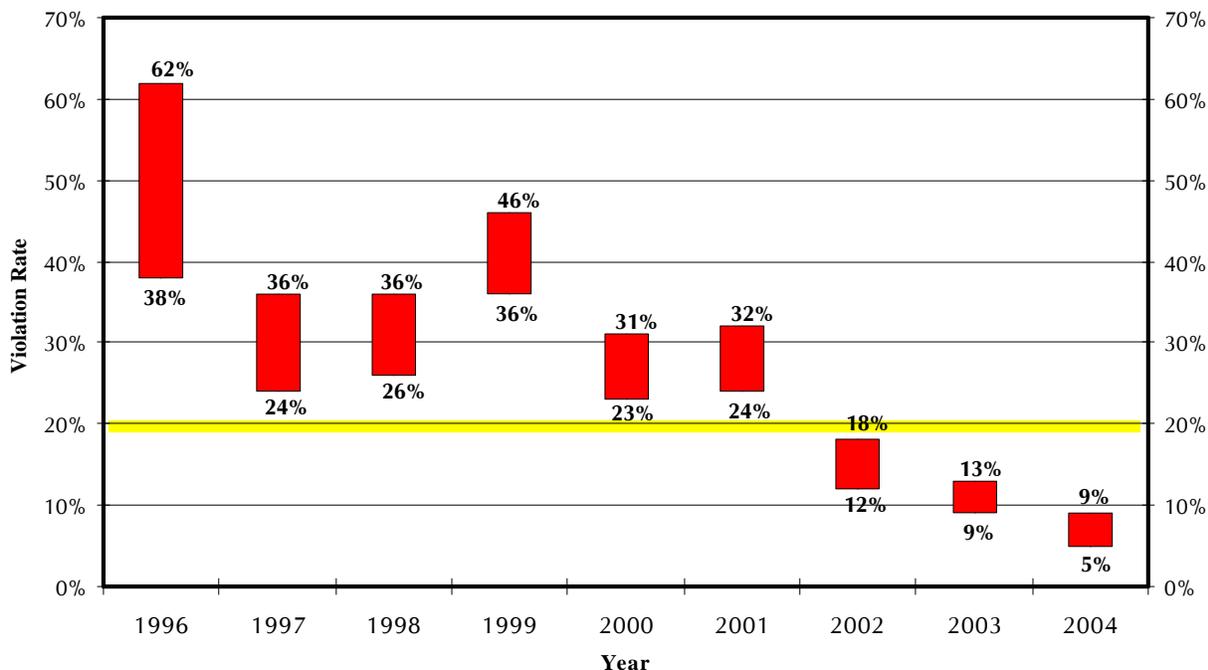
The survey results showed

that most retailers are displaying warning signs. It was estimated that 80% ( $\pm 3\%$ ) of the target population display signs that warn potential cigarette buyers that they do not sell to persons under the age of 18.

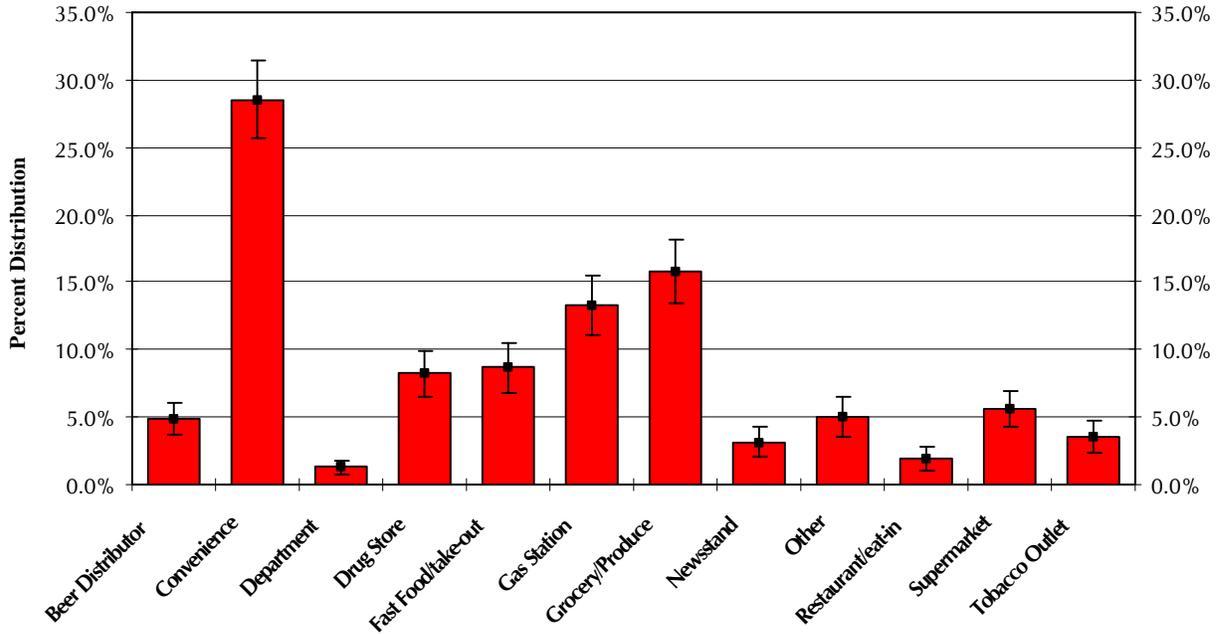
For every outlet entered, the survey participants were asked to classify the clerk that served them as either male or female. It was estimated that 42% ( $\pm 3\%$ ) of the outlets have a male clerk and 58% ( $\pm 3\%$ ) have female clerks. Male clerks sold cigarettes to minors 9% ( $\pm 3\%$ ) of the time, compared to 6% ( $\pm 2\%$ ) for female clerks.

To access federal statistics online, go to <http://prevention.samhsa.gov/tobacco>. For questions regarding this article, please contact the Bureau of Health Statistics and Research at 717-783-2548.

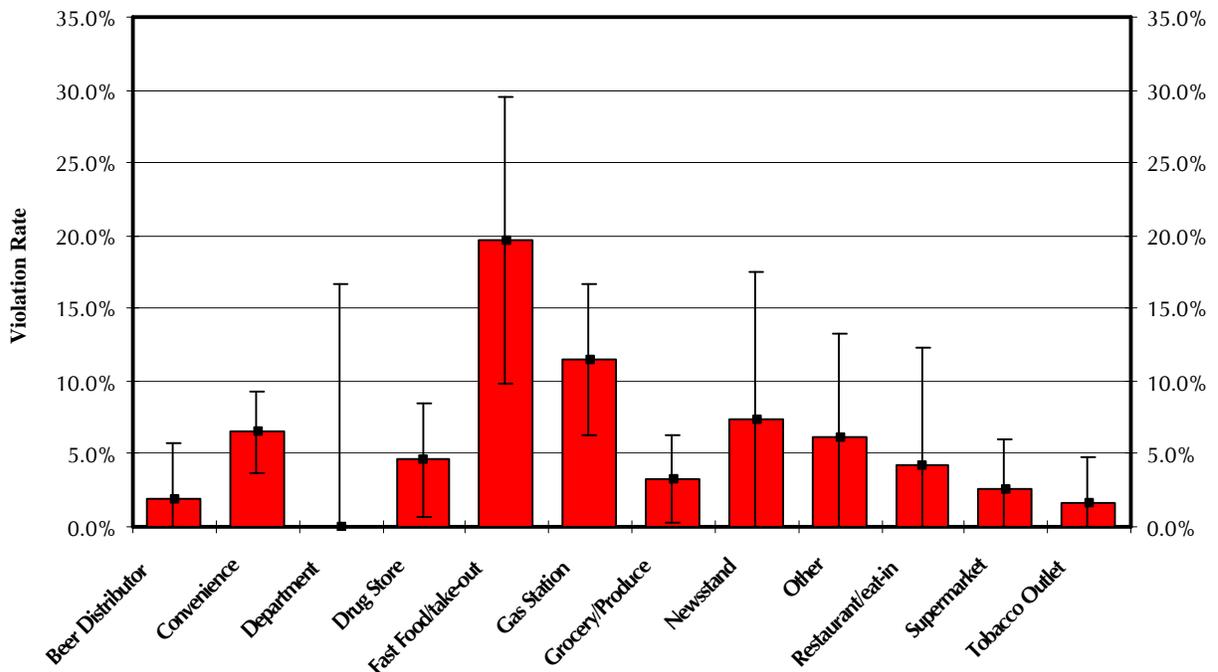
**Chart 1**  
Range of 95% Confidence Intervals for Estimated Rates of Cigarette Sales to Minors  
Pennsylvania Synar Survey, 1996-2004



**Chart 2**  
**Weighted Distribution of Outlets Sampled by Outlet Type**  
**Pennsylvania Synar Survey, 2004**



**Chart 3**  
**Weighted Violation Rates of Cigarette Sales to Minors by Outlet Type**  
**Pennsylvania Synar Survey, 2004**



# What Is Geocoding And How Does It Work

ZIP code centroid. If an address cannot be geocoded using address matching because an address contains a post office box, a rural postal route or is unrecognizable in a streets reference file, geocoding services can be set up to place a point in the exact geometric center of the respective ZIP code area. This is known as the ZIP code centroid. As is shown in Figure 4 below, the five-digit number represents the ZIP code, the polygons represent the ZIP code area, and the red dots are the ZIP code centroids. This method is not very accurate because the actual physical location of the address could be many miles away from the ZIP code centroid.

Although the map in Figure 4 shows polygons that represent ZIP code areas, ZIP codes are really just a group of postal delivery routes. The United States Postal Service (USPS) does not

create a boundary file for ZIP code areas, but they do produce a street segment file that is updated monthly which contains the latitude and longitude of the beginning and ending points of street segments in ZIP code areas. Private vendors of geospatial data layers take these updates and create polygons from this information. Some vendors create these boundaries using advanced algorithms. Other vendors, like GDT, draw the boundaries and ensure that they are accurate and visually appealing. GDT performs monthly testing after they receive the files from the USPS to ensure that their ZIP code boundaries are correct and will enclose only the addresses in a specific ZIP code.

The Pennsylvania Department of Health uses both address matching and ZIP code centroid plotting in its geocoding efforts. If the ZIP code centroid method-

ology is used, there is a quality code assigned to a field in the related database table to indicate this. An analyst can then differentiate between the geocoded data that is most likely very accurate and the data that is not so accurate.

## Global Positioning System (GPS):

If there was a need for the identification of a very specific geographic location where address matching would not work, someone would have to hop in the car, drive to the location, and take a GPS reading. This is not a method of geocoding that the Pennsylvania Department of Health uses very frequently at this time.

GPS has become a more common term since the marketing of consumer GPS hand held units and the installation of GPS systems in today's vehicles but many people are still unclear

about how the technology works. A GPS receiver sends radio signals into space and those signals bounce off of satellites orbiting the earth and return to the GPS unit on the earth's surface. The GPS unit measures its distance to the multiple satellites to which it sends a signal by recording the time it takes for the radio signal to reach each satellite and return to earth. A complex method of triangulation is computed and the GPS unit returns a latitude and longitude coordinate to the GPS user. There is some inaccuracy with this methodology, although the error is usually measured in feet and not yards or miles as could be the case with address matching or ZIP code centroid plotting methods.

If you have any questions concerning geocoding concepts covered in this article, please contact the Bureau at 717-783-2548.

Figure 2

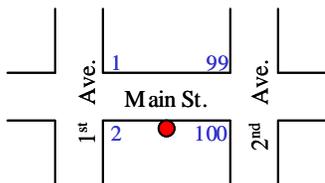


Figure 3

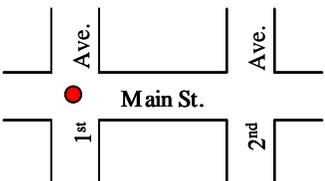
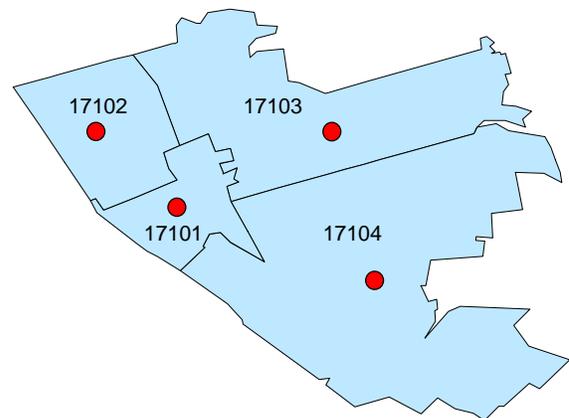


Figure 4



# Update: Healthy People 2010 Objectives

## Focus Area 01: Access to Quality Health Services

### 01-09a - Reduce the hospitalization rate for pediatric asthma (persons under 18).

#### 2010 Target: 17.3 discharges per 10,000

#### All Persons Under 18 and By Sex:

The hospitalization rate for pediatric asthma among Pennsylvania residents was 23.7 per 10,000 population under age 18 in 2002. There were 6,882 discharges with a primary diagnosis of asthma to residents under 18 that year from Pennsylvania hospitals.

The pediatric asthma hospitalization rate in 2002 was over 50 percent higher among males under 18 (28.5 per 10,000), compared to females under 18 (18.6). This difference in the rates by sex was consistent between 1998 and 2002.

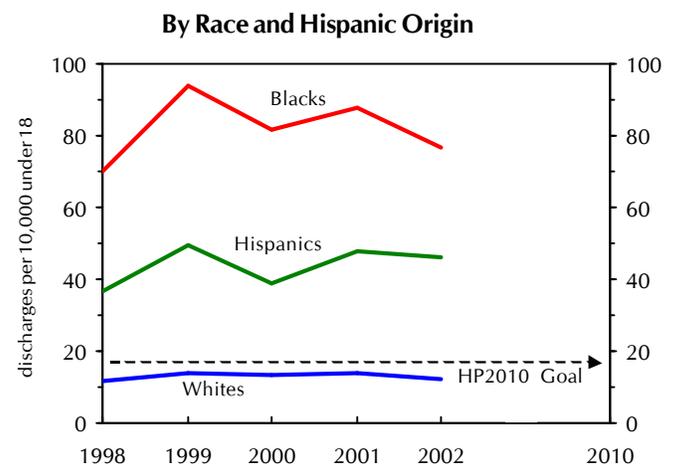
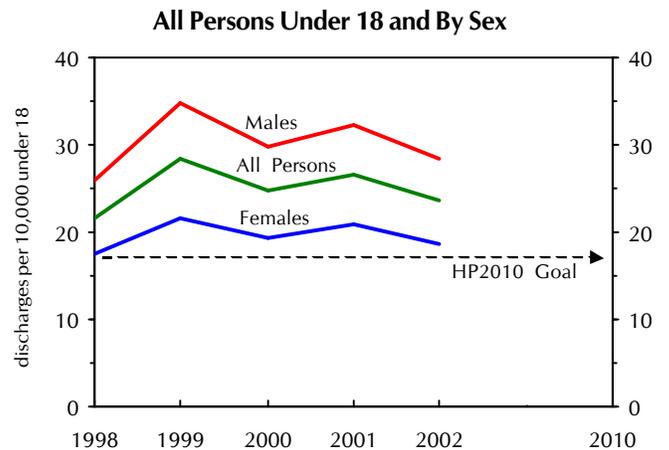
The Healthy People 2010 objective is a rate of 17.3 per 10,000. The rates for all persons under 18 and for males under 18 have been much higher than the 2010 goal and historical data show no trends. The rates for females under 18 are closer to the 2010 goal but, again, no trend is evident.

#### Race and Hispanic Origin:

The 2002 pediatric asthma hospitalization rates by race and Hispanic Origin show that the rate for blacks (76.8 per 10,000) was over six times higher than the rate for whites (12.3) and the Hispanic rate (46.1) was almost four times higher than the white rate. In addition, the rates for blacks between 1998 and 2002 have remained much higher than the goal while the rates for whites have been consistently lower. The rates among the white and Hispanic populations under 18 displayed no evident trends between 1998 and 2002.

The rates for whites have been below the Healthy People objective of 17.3 throughout the five-year period of 1998-2002. A substantial decline in the Hispanic rates are necessary in order to meet the HP 2010 goal. The rates for blacks show no trend and are so high now that it seems very unlikely that they will meet the goal.

Hospitalization Rates For Pediatric Asthma  
Pennsylvania Residents Under Age 18, 1998-2002



**Hospitalization Rate\* for Pediatric Asthma  
By Sex and Race, Pennsylvania Residents, 1998-2002**

	2002	2001	2000	1999	1998
All Persons Under 18 .....	23.7	26.7	24.7	28.3	21.7
Males Under 18 .....	28.5	32.2	29.8	34.7	25.8
Females Under 18 .....	18.6	20.9	19.4	21.7	17.4
Whites Under 18 .....	12.3	14.1	13.3	14.1	11.5
Blacks Under 18 .....	76.8	87.9	81.4	94.0	70.2
Hispanics** Under 18 .....	46.1	47.7	38.7	49.2	36.5

\*discharges per 10,000 under 18 \*\*Hispanics can be of any race

#### HP2010 State and County Data on the Web

To access the Department of Health's web page of Healthy People 2010 statistics for the state and counties, go to [www.health.state.pa.us/stats](http://www.health.state.pa.us/stats). The latest available statistics as well as trend data are shown. You can view data for the state, all counties, a specific demographic element (age, sex, race, etc.) or just for a specific county. Complete data sets for the state and counties can be downloaded. There is also a link to the national HP2010 web site.

---

*Statistical News* is published bimonthly by the Bureau of Health Statistics and Research, Pennsylvania Department of Health, 555 Walnut St., 6th Floor, Harrisburg, PA, 17101. Please write, telephone (717-783-2548) or FAX (717-772-3258) us if you have any questions regarding the contents of this newsletter. Visit the Health Statistics section of the Department's web site at [www.health.state.pa.us/stats](http://www.health.state.pa.us/stats) to access additional health statistics and reports.

*The Department of Health is an equal opportunity provider of grants, contracts, services and employment.*



Click on the EpiQMS logo above to access our interactive health statistics web site.

H106.830P