

Using Secondary Data

Presented by: Robert Martiniano, DrPH, MPA
Senior Program Manager
(518) 402-0250
rmartiniano@albany.edu

Center for Health Workforce Studies
School of Public Health | University at Albany, SUNY



The Center for Health Workforce Studies at the University at Albany, SUNY

- Established in 1996
- Based at the University at Albany School of Public Health
- Committed to collecting and analyzing data to understand workforce dynamics and trends
- Goal to inform public policies, the health and education sectors and the public
- Broad array of funders in support of health workforce research

Today's Topics

- Why Use Secondary Data
- What Needs to be Considered When Using Secondary Data
- Cleaning Raw Secondary Data
- How to Use Secondary Data
- Potential Secondary Data Sources
 - Licensure Data
 - Utilization Data
 - Decennial Census/American Community Survey
 - IPEDS
- Telling the Story

Why Use Secondary Data

- Why use secondary data:
 - Cost
 - Convenience
 - Expertise
 - ✓ Sampling issues
 - ✓ Creating the questions
 - Time
 - ✓ Staff time
 - ✓ Data collection

What Needs to be Considered When Using Secondary Data

- What we need to consider:
 - Time since data was collected
 - ✓ How old
 - ✓ Potential interventions influencing the data
 - Response rate
 - Size of the sample
 - Whether it has been cleaned

What Needs to be Considered When Using Secondary Data

- Who collected the data?
 - Government
 - Researchers
 - Advocacy groups
- What was the purpose of the initial data collection – how does it relate to what you are doing?
- What were the questions asked?
- How were the questions asked?
 - Education
 - Practice setting

What Needs to be Considered When Using Secondary Data

- How are data presented?
 - Rates per 1,000, 10,000, 100,000
 - Percentages
 - Counts
 - Raw data
- Level of geography
 - Census division
 - State
 - County
 - Sub county
- What does the data describe?
 - The entire population
 - A sample

Cleaning Raw Secondary Data

- Once data collection has been completed:
 - What to do with incomplete responses
 - Coding non-responses to questions
 - Data imputation
- Coding the data:
 - Assigning numeric values to text responses
 - Creating categories (age, years of service)
 - Combining responses (race/ethnicity)
 - Separating responses to “check all that apply”
- Does the data make sense?

How to Use Secondary Data

- Need to consider how to use the data:
 - Basic
 - Intermediate
 - Advanced
- Need to consider how to tell the story:
 - What level of geography
 - How to develop the narrative
 - What type of figures to use
 - ✓ Maps
 - ✓ Tables
 - ✓ Graphs

Potential Secondary Data Sources

- State licensure data
- Health care utilization data
 - Hospitalizations
 - Emergency departments
 - Outpatient settings
- Census Data
 - Decennial census
 - American Community Survey
- Educational data
 - IPEDS

What Is Included in Licensure Data

- Questions to consider:
 - Which health care occupations are licensed?
 - How many licenses are required?
 - How often is it collected?
 - What is asked for licensure purposes?
 - ✓ What data is available
 - ✓ Is there a public vs. protected data set

What to Consider When Using Licensure Data

- What to consider when using licensure data:
 - What practice information is available?
 - ✓ Specialty (physicians, NPs)?
 - ✓ Those actively practicing or not practicing?
 - ✓ Number of offices?
 - ✓ Is the address home or office?
 - What demographic information is available?
 - What levels of education are available (RNs)?

Displaying Licensure Data: Basic Counts

Health Professionals Per 100,000 in 2014

Professions	Counts	Per Capita
Clinical Laboratory Technician	1,803	9.1
Clinical Laboratory Technologist	12,662	64.1
Dental Hygienist	9,811	49.7
Dentist	16,866	85.5
Licensed Clinical Social Worker (R/P psychotherapy priv.)	24,966	126.4
Licensed Master Social Worker (no privileges)	25,047	126.8
Nurse Practitioners	17,085	86.5
Occupational Therapist	10,406	52.7
Registered Physician Assistant	11,245	56.9
Registered Professional Nurse	224,687	1137.9

Displaying Licensure Data: Per Capita

2014 Registered Professional Nurse Graduate Schools

County	School Sponsorship				
	NA*	SUNY	CUNY	Private	Total
Albany	693	14	3	4,527	5,237
Clinton	130	1	0	1,148	1,279
Erie	1,954	2	2	12,133	14,091
New York	1,012	10	40	8,949	10,011
Westchester	1,508	581	47	10,237	12,373
Total	5,297	608	92	36,994	42,991
*School Information is not available					

What Is Included in Utilization Data

- What is included in utilization data?
 - Diagnoses (primary and secondary)
 - Procedures (how many)
 - Insurance status/payer (primary and other)
 - Providers
- What patient information is available?
 - Age
 - Gender
 - Race/ethnicity

What to Consider When Using Utilization Data

- What to consider when using utilization data:
 - Is it population or sample data?
 - At what geographic level?
 - Is it a billing or service location address?
 - Is it possible to identify the actual provider of service?
 - Is it possible to identify specialty of provider (physicians, NPs)?
 - Public versus protected data set
 - What selection criteria to use

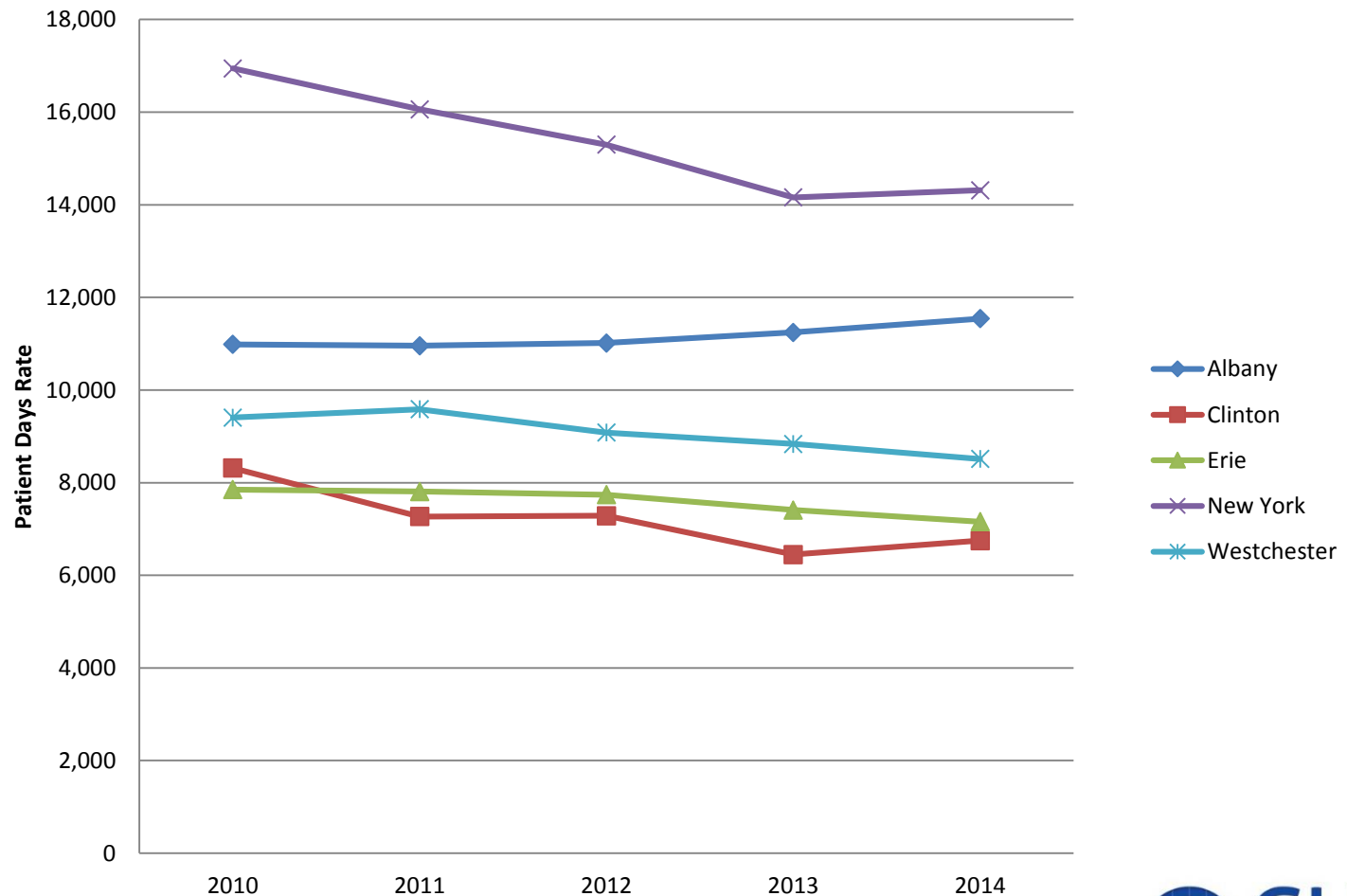
Displaying Utilization Data: Basic Analysis

2014 SPARCS Patient Days in Selected Counties

Counts	Patient Days	Percentage
Albany	355,683	2.7%
Clinton	55,100	0.4%
Erie	660,564	5.0%
New York	2,342,107	17.8%
Westchester	827,843	6.3%
New York State	13,147,046	100.0%

Displaying Utilization Data: Intermediate Analysis

Patient Days Rate per 10,000 population from 2010-2014



American Community Survey (ACS)

- Collected by Census Bureau
- Two components:
 - Household
 - Population
- 1-year data set (1% sample)
 - Geographic areas over 65,000
- 5-year Estimates (5% sample)
 - Any geographic area
- Data accessed through
 - American Factfinder
 - PUMS (Public Use Microdata Sample) Data

What to Consider When Using ACS Data

- What to consider when using ACS data:
 - Which is better to use?
 - ✓ American Factfinder
 - ✓ PUMS Data
 - At what geographic level?
 - ✓ 1-year estimates versus 5-year estimates
 - Does it get to the question asked?
 - ✓ Level of education
 - ✓ Work setting

Using American Factfinder for ACS Data

- Using Advanced Search for ACS or Census Data in American Factfinder
 - Identifying geography:
 - ✓ National
 - ✓ State
 - ✓ County
 - ✓ Sub county
 - Identifying topic:
 - ✓ People
 - ✓ Housing

Using American Factfinder for ACS Data

- Using Advanced Search for ACS or Census Data in American Factfinder
 - Using standardized reports:
 - ✓ DP-1 “General Population and House Characteristics” (2010 census – 100%)
 - Sex and age
 - Race and ethnicity
 - Household relationship
 - ✓ DP02 “Selected Social Characteristics”
 - Type of household
 - Marital status
 - Educational attainment
 - Citizenship status

Using American Factfinder for ACS Data

- Using Advanced Search for ACS or Census Data
 - Using standardized reports:
 - ✓ DP03 “Selected Economic Characteristics”
 - Employment status
 - Household income
 - Families and individuals below the federal poverty level
 - ✓ DP04 “Selected Housing Characteristics”
 - Number of rooms
 - Year structures built
 - Value
 - ✓ DP05 “ACS Demographic and Housing Estimates”
 - Sex and age
 - Race and ethnicity

Displaying ACS Data: Basic Analysis

<< 1 - 18 of 248 >>

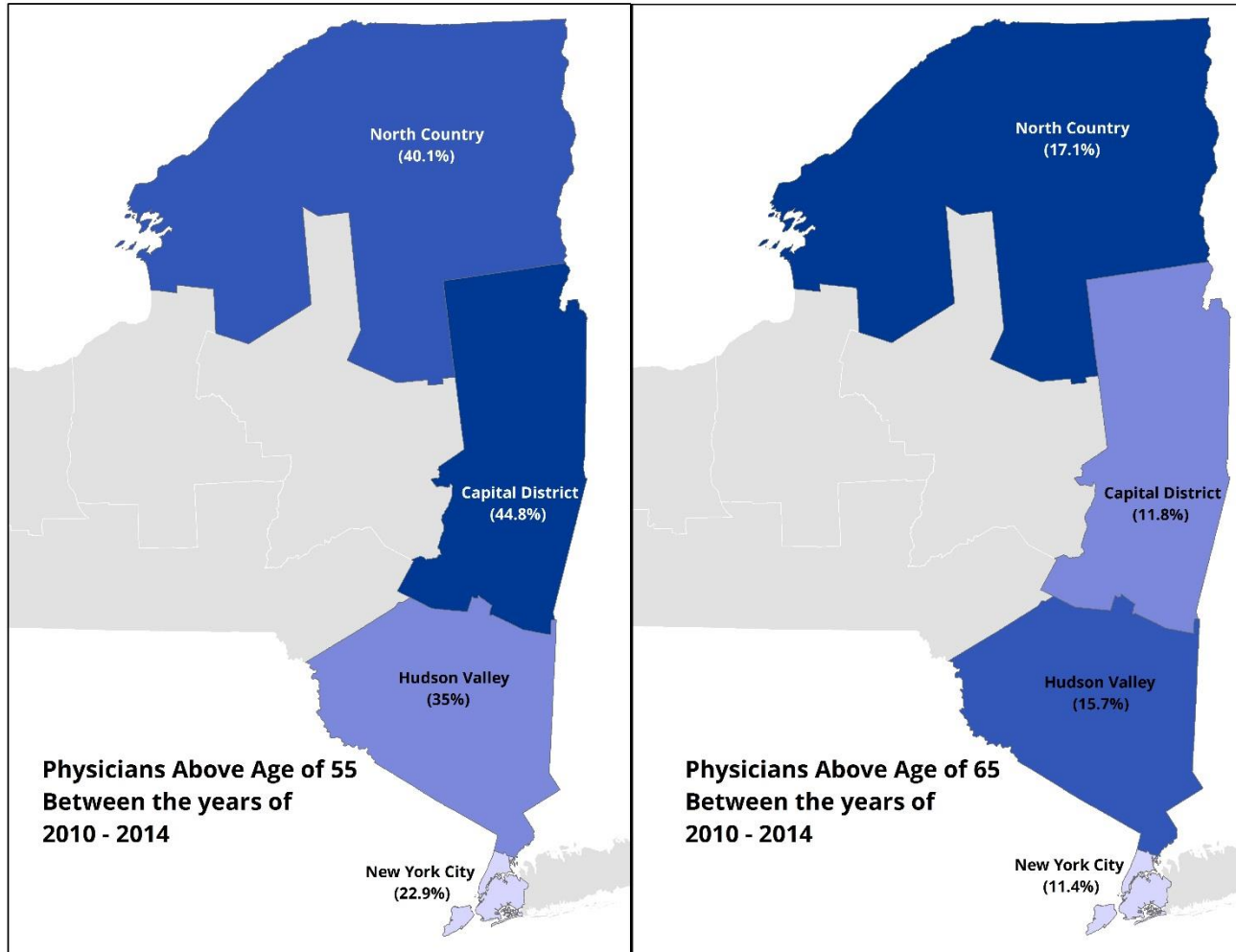
81 of 81

Subject	Albany County, New York				Allegany County, New York			
	Estimate	Margin of Error	Percent	Percent Margin of Error	Estimate	Margin of Error	Percent	Percent Margin of Error
SEX AND AGE								
Total population	306,124	*****	306,124	(X)	48,387	*****	48,387	(X)
Male	148,019	+/-29	48.4%	+/-0.1	24,481	+/-103	50.6%	+/-0.2
Female	158,105	+/-29	51.6%	+/-0.1	23,906	+/-103	49.4%	+/-0.2

HISPANIC OR LATINO AND RACE								
Total population	306,124	*****	306,124	(X)	48,387	*****	48,387	(X)
Hispanic or Latino (of any race)	16,336	*****	5.3%	*****	716	*****	1.5%	*****
Mexican	1,628	+/-405	0.5%	+/-0.1	112	+/-66	0.2%	+/-0.1
Puerto Rican	8,591	+/-717	2.8%	+/-0.2	267	+/-65	0.6%	+/-0.1
Cuban	546	+/-199	0.2%	+/-0.1	25	+/-18	0.1%	+/-0.1
Other Hispanic or Latino	5,571	+/-701	1.8%	+/-0.2	312	+/-90	0.6%	+/-0.2
Not Hispanic or Latino	289,788	*****	94.7%	*****	47,671	*****	98.5%	*****
White alone	229,001	+/-126	74.8%	+/-0.1	45,906	+/-69	94.9%	+/-0.1
Black or African American alone	36,544	+/-697	11.9%	+/-0.2	643	+/-70	1.3%	+/-0.1
American Indian and Alaska Native alone	386	+/-121	0.1%	+/-0.1	159	+/-33	0.3%	+/-0.1
Asian alone	16,572	+/-542	5.4%	+/-0.2	594	+/-50	1.2%	+/-0.1
Native Hawaiian and Other Pacific Islander alone	90	+/-77	0.0%	+/-0.1	0	+/-24	0.0%	+/-0.1
Some other race alone	297	+/-179	0.1%	+/-0.1	76	+/-67	0.2%	+/-0.1
Two or more races	6,898	+/-863	2.3%	+/-0.3	293	+/-84	0.6%	+/-0.2
Two races including Some other race	494	+/-187	0.2%	+/-0.1	10	+/-15	0.0%	+/-0.1
Two races excluding Some other race, and Three or more races	6,404	+/-854	2.1%	+/-0.3	283	+/-81	0.6%	+/-0.2
Total housing units	137,859	+/-500	(X)	(X)	26,106	+/-211	(X)	(X)

0-2014 American Community Survey 5-Year Estimates

Displaying ACS Data: Intermediate Analysis



Integrated Postsecondary Education Data System (IPEDS)

- Collected by the National Center for Education Statistics
- Collected annually
- Data accessed through:
 - Summary tables (user defined)
 - Data Files:
 - ✓ Institutional characteristics
 - Institutional information
 - Degree offerings
 - Student charges
 - ✓ 12-month enrollment
 - Unduplicated count
 - Duplicated count
 - ✓ Completions
 - Awards by CIP, award level, race/ethnicity, and gender

What to Consider When Using IPEDS Data

- Data fields may not be consistent over the years:
 - ✓ New fields added
 - ✓ Discontinuation
 - ✓ Definitions may change
 - ✓ Classification of Instructional Program (CIP) codes define fields of study in programs
 - CIP code system changed in 2010
- Each school may have multiple records corresponding to different programs
- Need to link files to institutional characteristics to analyze by geography, institution, and other features

Displaying IPEDS Data: Basic Analysis

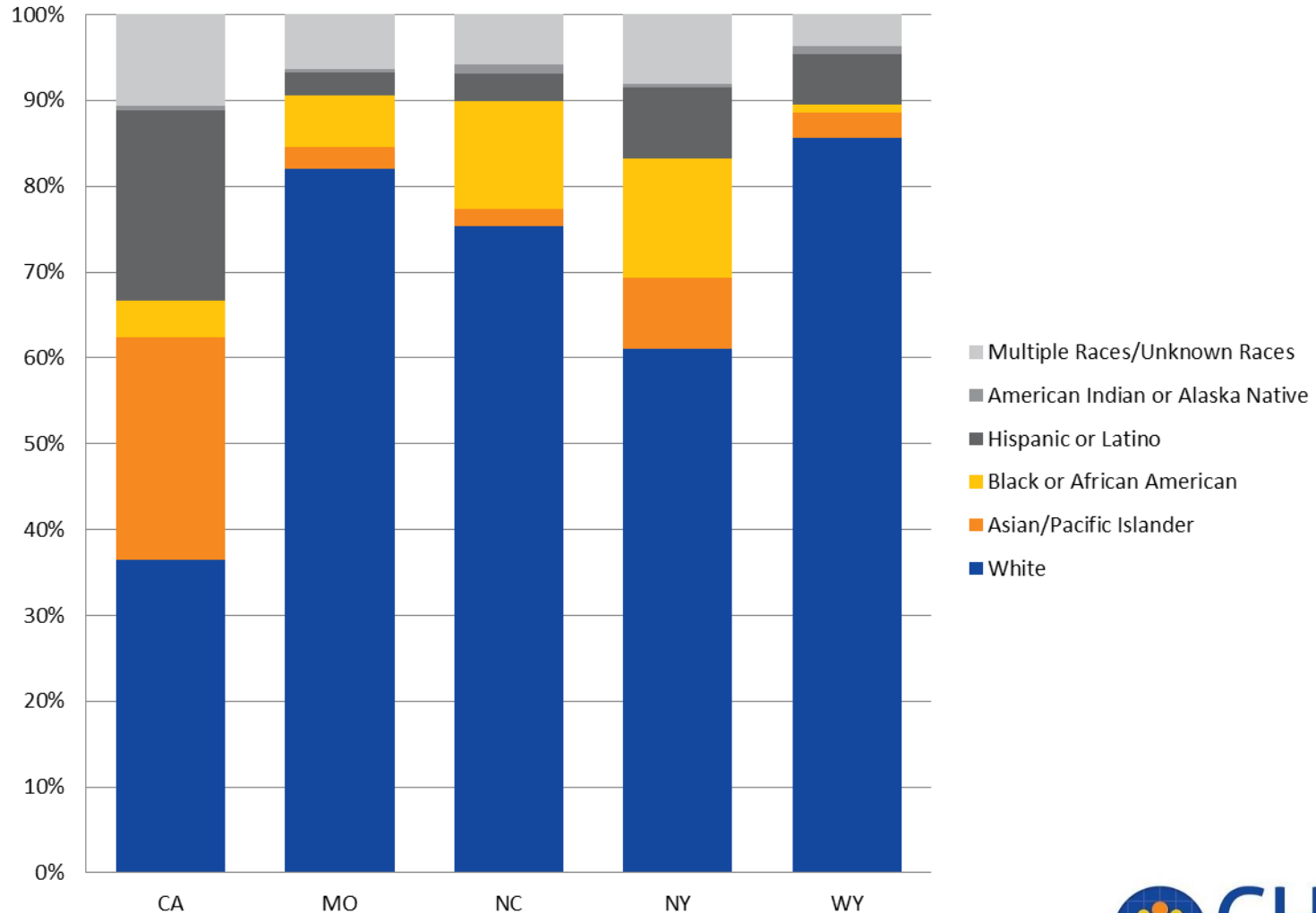
Number of Registered Nurse Graduates per 100,000
in Selected States in 2015

State	Total	Total Population	Per Capita
California	14,218	39,144,818	36.3
Missouri	5,243	6,083,672	86.2
North Carolina	5,595	10,042,802	55.7
New York	12,796	19,795,791	64.6
Wyoming	542	586,107	92.5

CIP Code for Registered Nurse Program is 51.3801.
State Populations are from 2015 American Community Survey.

Displaying IPEDS Data: Intermediate Analysis

Race Distribution of Registered Nurses graduates in Selected States in 2015



Telling the Story

NEW YORK CITY		Total Population: 8,128,980		Percent of Population Below FPL: 19%		Median Household Income: \$80,528			
Demographics	Percentage Distribution by Age Group				Race/Ethnicity		Education Level	Number	% of Total
	0 to 4	5 to 17	18 to 64	65 Plus	Number	% of Total	Less than High School	1,142,283	20.7%
White, non-Hispanic	5.4%	10.7%	66.6%	17.3%	2,724,300	33.5%	High School/GED	2,174,158	39.5%
Black, non-Hispanic	6.2%	18.0%	64.6%	11.3%	1,873,853	23.1%	Associate	334,602	6.1%
Hispanic/Latino	7.9%	19.3%	64.6%	8.3%	2,310,163	28.4%	Bachelor's	1,096,625	19.9%
Asian/Pacific Islander	5.5%	13.6%	71.4%	9.4%	1,027,392	12.6%	Master's or Above	758,212	13.8%
American Indian/Alaska Native	7.0%	19.1%	65.4%	8.6%	15,133	0.2%	Other		
Multiple Races/Other	10.8%	19.6%	62.7%	6.9%	178,139	2.2%	Unemployment	393,224	9.5%
Total	521,478	1,253,845	5,369,016	984,641	8,128,980	100%	Medicaid Eligible	3,074,232	15.9%
Percent Female	48.9%	49.1%	52.2%	60.3%					

Nativity	Number	Rate/Percent
Births per 1,000 Females, 15-44	121,168	63.9
Teen Births per 1,000 Females, 15-17	2,335	15.2
% Low Birth Weight (<2500 grams)	10,573	8.7%
% Premature Births (<37 Weeks)	15,087	12.7%
% With Early Prenatal Care	80,561	70.4%
Infant Mortality (per 1000 Live Births)	584	4.8

Health Status/Health Behaviors	Number	Percent
% Adults without Health Insurance	244,686	18.0%
% Adults with Usual Source of Care	1,045,351	76.9%
% Adults with Hypertension	324,888	23.9%
% Adults with Diabetes	130,499	9.6%
% Adults with Asthma	126,421	9.3%
% Adults Smoking	197,108	14.5%
% Adults Obese	289,545	21.3%
% Adults with Poor Physical Health	135,936	10.0%
% Adults with Poor Mental Health	125,061	9.2%

Cancer Mortality/Cases per 100,000 Population	Number	Rate
Cancer Deaths	12,321	147.6
All Cancer Cases	37,721	451.8
Lung and Bronchus Cancer Cases	4,257	51.0
Female Breast Cancer Cases	5,403	123.8
Prostate Cancer Cases	5,811	145.8

Mortality/Cases per 100,000 Population	Number	Rate
Overall Age Adjusted Mortality	51,767	602.5
Childhood Mortality, Ages 1-4	78	17.6
Childhood Mortality, Ages 5-14	112	11.3
Chronic Lower Respiratory Disease Mortality	1,627	19.6
Diseases of the Heart Mortality	19,482	234.4
Diabetes Mortality	1,644	19.8
Unintentional - Motor Vehicle Mortality	294	3.5
Unintentional - Non Motor Vehicle Mortality	1,164	14.0

Hospitalizations/ED Visits per 10,000 Population	Number	Rate
Total Hospitalizations	1,147,507	1,380.8
Total Patient Days	5,919,463	7,123.1
Total Preventable Hospitalizations	113,733	176.2
Total ED Visits	3,480,700	4,188.4
Chronic Lower Respiratory Disease Hospitalizations	37,142	44.7
Heart Disease Hospitalizations	101,451	122.1
Diabetes Hospitalizations	228,671	275.2
Asthma Hospitalizations, All Ages	25,777	31.0
Asthma Hospitalizations, Ages 0-4	4,719	84.6
Asthma Hospitalizations, Ages 65 Plus	5,195	54.0
Asthma ED Visits, Ages 0-17	42,997	231.5
Fall Related Hospitalizations, Ages Under 10	1,257	11.8
Fall Related Hospitalizations, Ages 65 Plus	18,354	180.6
Fall ED Visits, Ages 1-4	20,168	437.0

New York Example

124

Health Workforce Planning Data Guide

Thank You

Questions?

Visit Us On:

