

35th Annual Conference  
National Association of Health Data Organizations  
Building a Bridge Between Data and Policy

# *Evaluating Healthcare: Comprehensive Data for a Complex System*

NAHDO Session Topic  
Building a Broader Vision of Healthcare Data: Thinking Outside the Box  
August 26, 2020

Brian Williams and David Stern  
Stern Consulting, LLC  
[www.sternconsulting.com](http://www.sternconsulting.com)

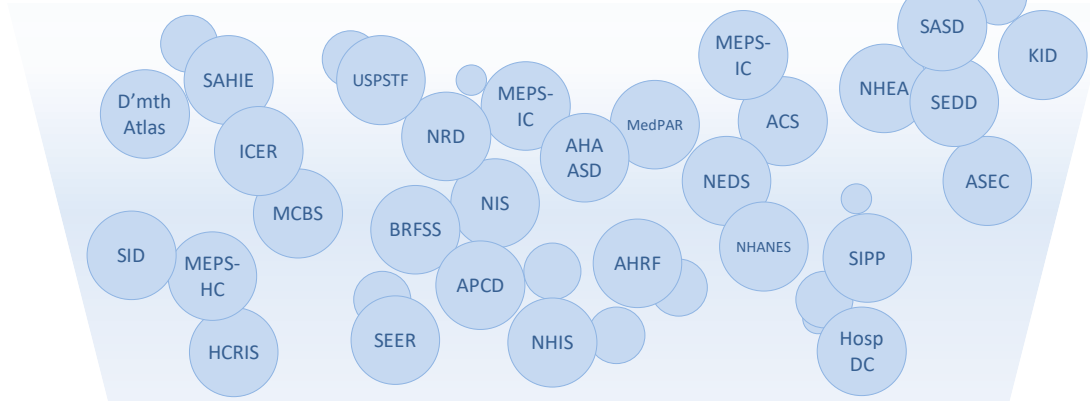


# The Challenge

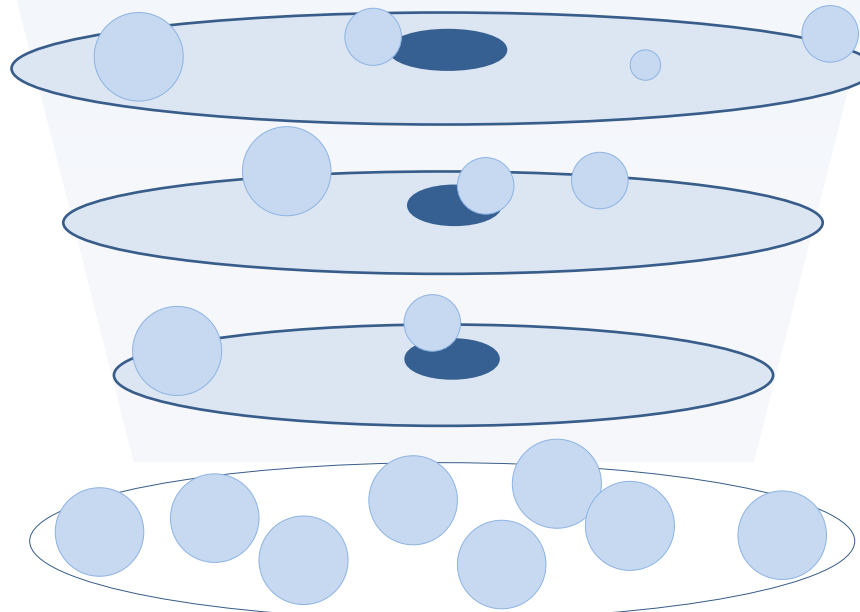
## Universe of Datasets

### Many are Called

*There are thousands of healthcare datasets.*



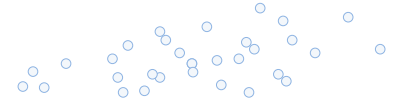
*Each project has selection criteria that determine the best resource(s) for the task.*



### Few are Chosen

*A handful may be suitable for a particular need.*

# A Framework for Meeting the Challenge



## Dimensions

### Mechanism

- Administrative
- Survey
- Disease Surveillance
- Evidence Based Healthcare
- Regulatory (e.g. Cost Reports)
- Medical Record Abstracts
- Vital Records
- Peer Reviewed Literature
- Gray Literature
- Directories/Code Books/Lists
- Other

### Granularity

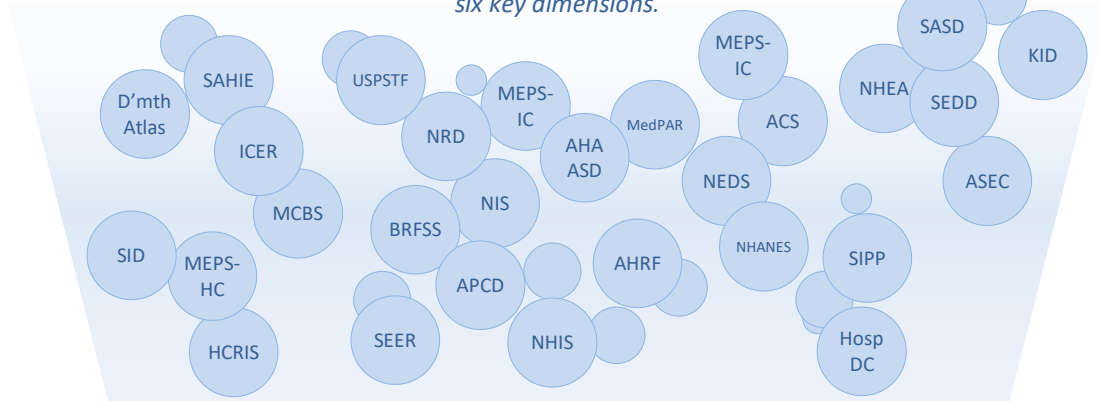
- Microdata
- Macrodata
  - Tables
  - Online Queries
  - Other

### Unit of Analysis

- Person
- Household
- Employer
- Encounter/Claim
- Diagnosis
- Procedure
- Provider
- Location
- Other

## Universe of Datasets

Datasets are defined by six key dimensions.



## Dimensions

### Content

- Access
- Charges/Cost of Care
- Utilization
- Health Status
- Quality/Satisfaction
- Clinical Results
- Clinical Classification
- Demographic
- Economic
- Social
- Other

### Constraints & Use

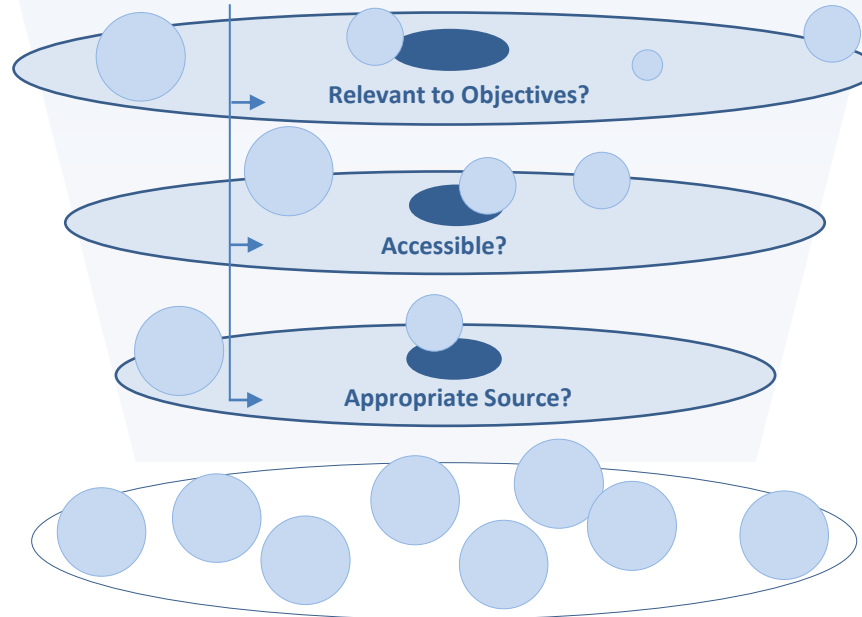
- Acquisition requirements
- Hardware and software
- Ease of use/Required expertise
- Restrictions on use
- Cost
- Time span availability
- Other

### Sponsor

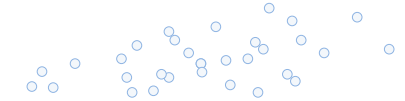
- Federal Government
- State Government
- Foundations
- Industry Associations
- Disease-Specific Organizations
- Academia
- Accreditation Bodies
- Commercial Entities
- Other

## Project Screening Criteria

Project criteria dictate data requirements.

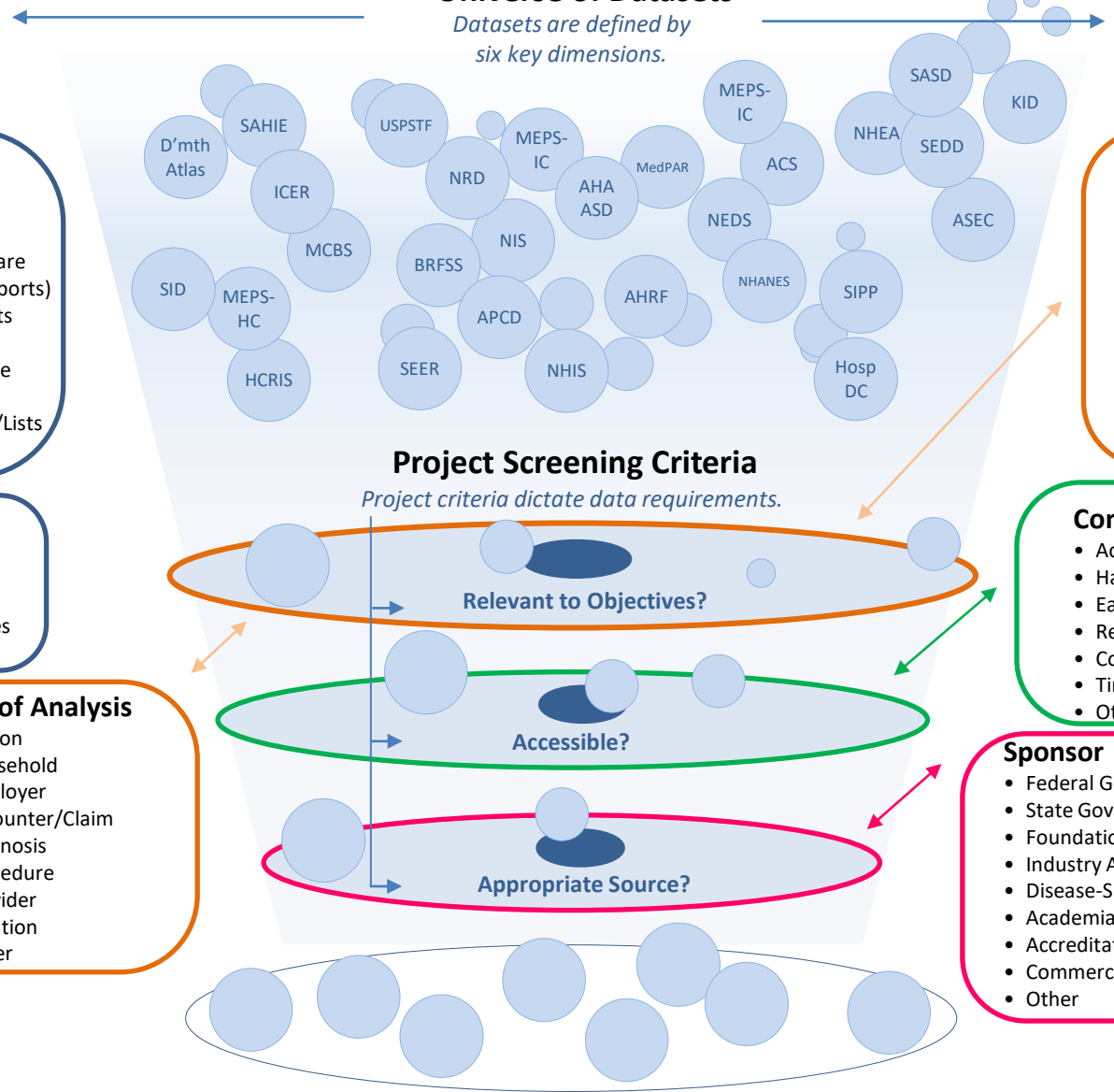


# Data Dimensions Guide the Search



## Universe of Datasets

Datasets are defined by six key dimensions.



### Dimensions

#### Mechanism

- Administrative
- Survey
- Disease Surveillance
- Evidence Based Healthcare
- Regulatory (e.g. Cost Reports)
- Medical Record Abstracts
- Vital Records
- Peer Reviewed Literature
- Gray Literature
- Directories/Code Books/Lists
- Other

#### Granularity

- Microdata
- Macrodata
  - Tables
  - Online Queries
  - Other

#### Unit of Analysis

- Person
- Household
- Employer
- Encounter/Claim
- Diagnosis
- Procedure
- Provider
- Location
- Other

### Dimensions

#### Content

- Access
- Charges/Cost of Care
- Utilization
- Health Status
- Quality/Satisfaction
- Clinical Results
- Clinical Classification
- Demographic
- Economic
- Social
- Other

#### Constraints & Use

- Acquisition requirements
- Hardware and software
- Ease of use/Required expertise
- Restrictions on use
- Cost
- Time span availability
- Other

#### Sponsor

- Federal Government
- State Government
- Foundations
- Industry Associations
- Disease-Specific Organizations
- Academia
- Accreditation Bodies
- Commercial Entities
- Other

### Project Screening Criteria

Project criteria dictate data requirements.

Relevant to Objectives?

Accessible?

Appropriate Source?

Our case study required publicly available datasets able to reflect the complex impacts of “National Healthcare Policy,” over time, for the U.S. population.

**Dimensions**

**Mechanism**

- Administrative
- Survey
- Disease Surveillance
- Evidence Based Healthcare
- Regulatory (e.g. Cost Reports)
- Medical Record Abstracts
- Vital Records
- Peer Reviewed Literature
- Gray Literature
- Directories/Code Books/Lists
- Other

**Granularity**

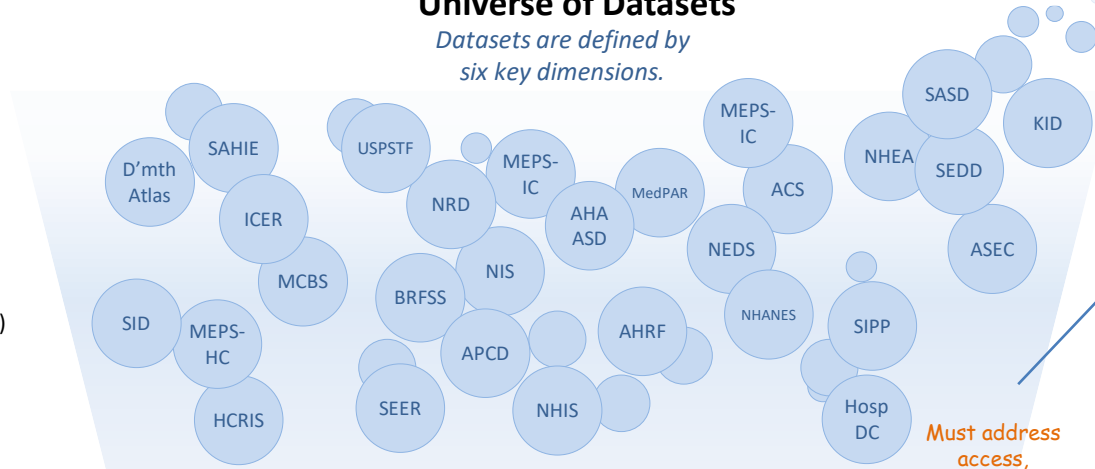
- Microdata
- Macrodata
  - Tables
  - Online Queries
  - Other

**Unit of Analysis**

- Person
- Household
- Employer
- Encounter/Claim
- Diagnosis
- Procedure
- Provider
- Location
- Other

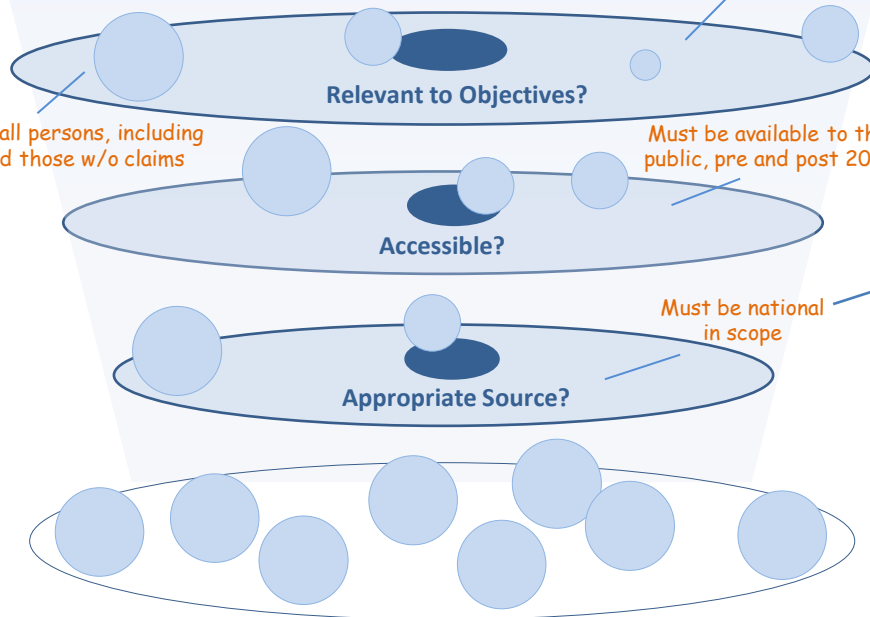
**Universe of Datasets**

Datasets are defined by six key dimensions.



**Project Screening Criteria**

Project criteria dictate data requirements.



**Dimensions**

**Content**

- Access
- Charges/Cost of Care
- Utilization
- Health Status
- Quality/Satisfaction
- Clinical Results
- Clinical Classification
- Demographic
- Economic
- Social
- Other

**Constraints & Use**

- Acquisition requirements
- Hardware and software
- Ease of use/Required expertise
- Restrictions on use
- Cost
- Time span availability
- Other

**Sponsor**

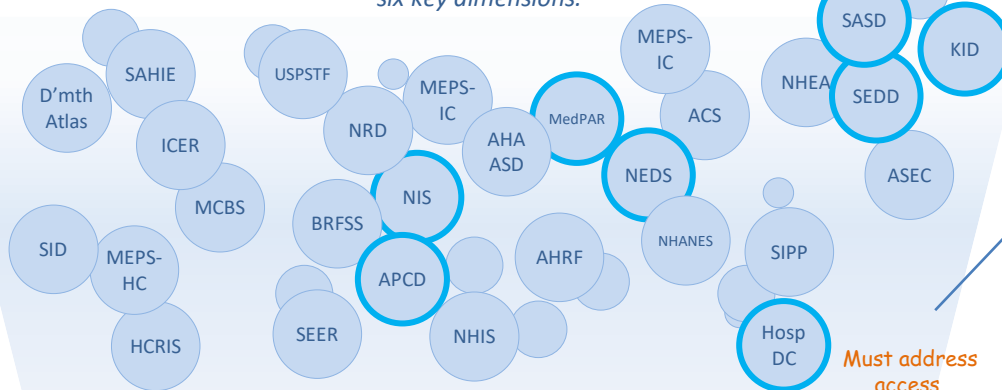
- Federal Government
- State Government
- Foundations
- Industry Associations
- Disease-Specific Organizations
- Academia
- Accreditation Bodies
- Commercial Entities
- Other

Note that typical administrative data do not address “access” or “healthiness.” Nor do they capture data on the uninsured or, aside from eligibility, persons without claims.

Our case study required publicly available datasets able to reflect the complex impacts of “National Healthcare Policy,” over time, for the U.S. population.

Universe of Datasets

Datasets are defined by six key dimensions.



Project Screening Criteria

Project criteria dictate data requirements.

Dimensions

Mechanism

- Administrative\*
- Survey
- Disease Surveillance
- Evidence Based Healthcare
- Regulatory (e.g. Cost Reports)
- Medical Record Abstracts
- Vital Records
- Peer Reviewed Literature
- Gray Literature
- Directories/Code Books/Lists
- Other

Granularity

- Microdata
- Macrodata
- Tables
- Online Queries
- Other

Unit of Analysis

- Person
- Household
- Employer
- Encounter/Claim
- Diagnosis
- Procedure
- Provider
- Location
- Other

Dimensions

Content

- Access
- Charges/Cost of Care
- Utilization
- Health Status
- Quality/Satisfaction
- Clinical Results
- Clinical Classification
- Demographic
- Economic
- Social
- Other

Constraints & Use

- Acquisition requirements
- Hardware and software
- Ease of use/Required expertise
- Restrictions on use
- Cost
- Time span availability
- Other

Sponsor

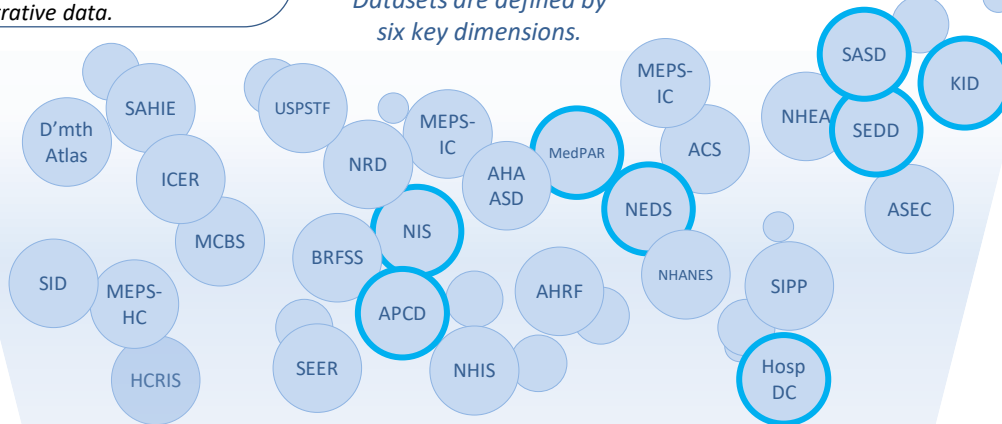
- Federal Government
- State Government
- Foundations
- Industry Associations
- Disease-Specific Organizations
- Academia
- Accreditation Bodies
- Commercial Entities
- Other

\*“Administrative” is defined herein as data generated pursuant to an encounter with the healthcare system or to an associated claim. It also includes eligibility and enrollment files such as those included in All Payer Claims Data (APCD) and Medicaid program data.

All of the circled data represent opportunities outside of the box of administrative data.

### Universe of Datasets

Datasets are defined by six key dimensions.



### Dimensions

#### Mechanism

- Administrative
- Survey
- Disease Surveillance
- Evidence Based Healthcare
- Regulatory (e.g. Cost Reports)
- Medical Record Abstracts
- Vital Records
- Peer Reviewed Literature
- Gray Literature
- Directories/Code Books/Lists
- Other

#### Granularity

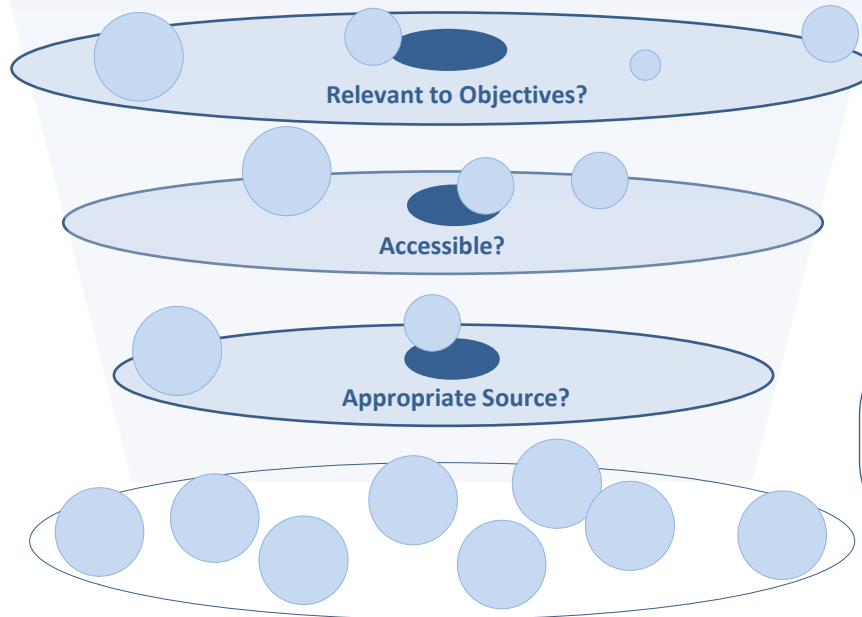
- Microdata
- Macrodata
- Tables
- Online Queries
- Other

#### Unit of Analysis

- Person
- Household
- Employer
- Encounter/Claim
- Diagnosis
- Procedure
- Provider
- Location
- Other

### Project Screening Criteria

Project criteria dictate data requirements.



### Dimensions

#### Content

- Access
- Charges/Cost of Care
- Utilization
- Health Status
- Quality/Satisfaction
- Clinical Results
- Clinical Classification
- Demographic
- Economic
- Social
- Other

#### Constraints & Use

- Acquisition requirements
- Hardware and software
- Ease of use/Required expertise
- Restrictions on use
- Cost
- Time span availability
- Other

#### Sponsor

- Federal Government
- State Government
- Foundations
- Industry Associations
- Disease-Specific Organizations
- Academia
- Accreditation Bodies
- Commercial Entities
- Other

Our case study required publicly available datasets able to reflect the complex impacts of “National Healthcare Policy,” over time, for the U.S. population.

**Dimensions**

**Mechanism**

- Administrative
- Survey
- Disease Surveillance
- Evidence Based Healthcare
- Regulatory (e.g. Cost Reports)
- Medical Record Abstracts
- Vital Records
- Peer Reviewed Literature
- Gray Literature
- Directories/Code Books/Lists
- Other

**Granularity**

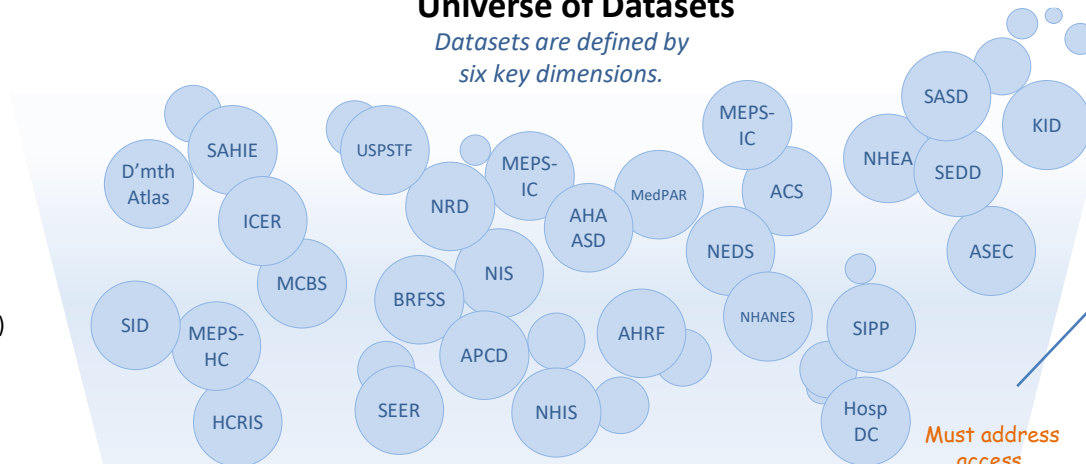
- Microdata
- Macrodata
- Tables
- Online Queries
- Other

**Unit of Analysis**

- Person
- Household
- Employer
- Encounter/Claim
- Diagnosis
- Procedure
- Provider
- Location
- Other

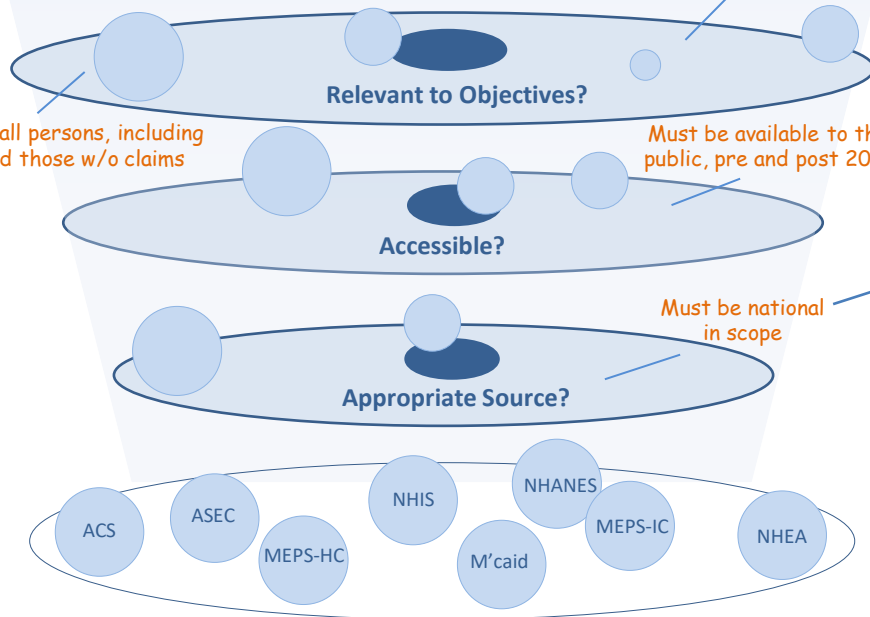
**Universe of Datasets**

Datasets are defined by six key dimensions.



**Project Screening Criteria**

Project criteria dictate data requirements.



**Dimensions**

**Content**

- Access
- Charges/Cost of Care
- Utilization
- Health Status
- Quality/Satisfaction
- Clinical Results
- Clinical Classification
- Demographic
- Economic
- Social
- Other

**Constraints & Use**

- Acquisition requirements
- Hardware and software
- Ease of use/Required expertise
- Restrictions on use
- Cost
- Time span availability
- Other

**Sponsor**

- Federal Government
- State Government
- Foundations
- Industry Associations
- Disease-Specific Organizations
- Academia
- Accreditation Bodies
- Commercial Entities
- Other

Eight datasets “made the cut” for our “Datasets to Evaluate the Impact of National Healthcare Policy” case study.



## Selected Datasets by Key Dimensions

|   |   | Dimensions of Data |                             |   |                 |                                  |                           |   |
|---|---|--------------------|-----------------------------|---|-----------------|----------------------------------|---------------------------|---|
| Datasets  | Notes   | Mechanism          | Sponsor<br>(Data Collector) | Content<br>Variable Counts: <sup>1</sup><br>Health Care<br>Non-Healthcare |                 | Primary Unit<br>of Analysis      | Granularity <sup>II</sup> | Constraints & Use*<br>Skill level/<br>Accessibility |
| <b>ACS</b><br>American Community Survey   | Extensive geographic and demographic drill downs on disability and health insurance.                  | Survey             | Census Bureau               | Healthcare  | 171             | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>ASEC</b><br>Annual Social and Economic Supplement to the Current Population Survey | Labor force data with health insurance, out-of-pocket \$ and health status fields.                    | Survey             | BLS<br>(Census Bureau)      | Healthcare  | 182             | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>MEPS-HC</b><br>Medical Expenditure Panel Survey, Household Component               | Person-level health expenditures with longitudinal capabilities.                                      | Survey             | AHRQ<br>(Westat)            | Healthcare  | 1,252           | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>NHANES</b><br>National Health and Nutrition Examination Survey                     | Survey combines interviews and physical examination, including lab tests.                             | Survey             | NCHS/CDC                    | Healthcare  | 1,733           | Person                           | Microdata                 | Requires Skills**                                   |
| <b>NHIS</b><br>National Health Interview Survey                                       | Principal source of information on health of U.S. population. Robust demographic, socioeconomic data. | Survey             | CDC<br>(NCHS)               | Healthcare  | 1,388           | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>MEPS-IC</b><br>Medical Expenditure Panel Survey, Insurance Component               | Factors contributing to use of employer sponsored insurance. Premiums and cost sharing.               | Survey             | AHRQ<br>(Census Bureau)     | Healthcare  | 153             | Employers,<br>Employees          | Macrodata                 | Ready-to-use  |
| <b>Medicaid</b><br>(various program data)   | Actual enrollment data. Breakouts of new eligibility categories created by ACA.                       | Admin              | CMS                         | Healthcare  | 30 <sup>I</sup> | Person                           | Macrodata                 | Ready-to-use  |
| <b>NHEA</b><br>National Health Expenditure Accounts                                   | Official estimates of healthcare spending in U.S. Includes care, admin, research and infrastructure.  | Multiple Sources   | HHS                         | Healthcare  | 640             | Services,<br>Payers,<br>Sponsors | Macrodata                 | Ready-to-use  |

BLS: Bureau of Labor Statistics  
 AHRQ: Agency for Healthcare Research and Quality  
 CDC: Centers for Disease Control and Prevention

NCHS: National Center for Health Statistics  
 CMS: Centers for Medicare and Medicaid Services  
 HHS: Department of Health and Human Services

<sup>1</sup> "Counts of variables" by topic is a reasonable method of determining a dataset's areas of focus. Each variable from the eight selected datasets has been categorized by subject matter. All ultimately roll up to either "non-healthcare" or "healthcare." Additional detail on content is provided below. (Counts exclude sample weights and variables related to survey administration.)

<sup>II</sup> Microdata consist of information at the unit level and provide the highest degree of analytic flexibility. Macrodata represent a broad range of pre-aggregated data, typically in ready-to-use format, as well as online utilities that enable users to generate their own queries against the underlying source data.

\* All are publicly available at no cost and include data pre- and post -2014.

\*\* For the microdata versions, users must have database management skills and ability to generate population estimates (relatively easy) and margins of error (more complicated) from raw survey data. Microdata are typically too large for MS Excel or MS Access.

Selected Datasets by Key Dimensions

Six of the eight selected datasets are surveys, one is based on administrative data (Medicaid enrollment) and the other on multiple mechanisms.

Dimensions of Data

| Datasets  | Notes   | Mechanism        | Sponsor<br>(Data Collector) | Content<br>Variable Counts: <sup>1</sup> |                       | Primary Unit<br>of Analysis      | Granularity <sup>II</sup> | Constraints & Use*<br>Skill level/<br>Accessibility |
|---|---|------------------|-----------------------------|--|-----------------------|----------------------------------|---------------------------|---|
|   |   |                  |                             | Health Care                              | Non-Healthcare        |                                  |                           |   |
| <b>ACS</b><br>American Community Survey   | Extensive geographic and demographic drill downs on disability and health insurance.                  | Survey           | Census Bureau               | Healthcare<br>171                        | Non-Healthcare<br>199 | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>ASEC</b><br>Annual Social and Economic Supplement to the Current Population Survey | Labor force data with health insurance, out-of-pocket \$ and health status fields.                    | Survey           | BLS<br>(Census Bureau)      | Healthcare<br>182                        | Non-Healthcare<br>479 | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>MEPS-HC</b><br>Medical Expenditure Panel Survey, Household Component               | Person-level health expenditures with longitudinal capabilities.                                      | Survey           | AHRQ<br>(Westat)            | Healthcare<br>1,252                      | Non-Healthcare<br>330 | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>NHANES</b><br>National Health and Nutrition Examination Survey                     | Survey combines interviews and physical examination, including lab tests.                             | Survey           | NCHS/CDC                    | Healthcare<br>1,733                      | Non-Healthcare<br>147 | Person                           | Microdata                 | Requires Skills**                                   |
| <b>NHIS</b><br>National Health Interview Survey                                       | Principal source of information on health of U.S. population. Robust demographic, socioeconomic data. | Survey           | CDC<br>(NCHS)               | Healthcare<br>1,388                      | Non-Healthcare<br>212 | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>MEPS-IC</b><br>Medical Expenditure Panel Survey, Insurance Component               | Factors contributing to use of employer sponsored insurance. Premiums and cost sharing.               | Survey           | AHRQ<br>(Census Bureau)     | Healthcare<br>153                        | Non-Healthcare<br>22  | Employers,<br>Employees          | Macrodata                 | Ready-to-use  |
| <b>Medicaid</b><br>(various program data)   | Actual enrollment data. Breakouts of new eligibility categories created by ACA.                       | Admin            | CMS                         | Healthcare<br>301                        | Non-Healthcare<br>8   | Person                           | Macrodata                 | Ready-to-use  |
| <b>NHEA</b><br>National Health Expenditure Accounts                                   | Official estimates of healthcare spending in U.S. Includes care, admin, research and infrastructure.  | Multiple Sources | HHS                         | Healthcare<br>640                        | Non-Healthcare<br>14  | Services,<br>Payers,<br>Sponsors | Macrodata                 | Ready-to-use  |

BLS: Bureau of Labor Statistics  
 AHRQ: Agency for Healthcare Research and Quality  
 CDC: Centers for Disease Control and Prevention  
 NCHS: National Center for Health Statistics  
 CMS: Centers for Medicare and Medicaid Services  
 HHS: Department of Health and Human Services

<sup>1</sup> “Counts of variables” by topic is a reasonable method of determining a dataset’s areas of focus. Each variable from the eight selected datasets has been categorized by subject matter. All ultimately roll up to either “non-healthcare” or “healthcare.” Additional detail on content is provided below. (Counts exclude sample weights and variables related to survey administration.)

<sup>II</sup> Microdata consist of information at the unit level and provide the highest degree of analytic flexibility. Macrodata represent a broad range of pre-aggregated data, typically in ready-to-use format, as well as online utilities that enable users to generate their own queries against the underlying source data.

\* All are publicly available at no cost and include data pre- and post -2014.

\*\* For the microdata versions, users must have database management skills and ability to generate population estimates (relatively easy) and margins of error (more complicated) from raw survey data. Microdata are typically too large for MS Excel or MS Access.

## Selected Datasets by Key Dimensions

Five of the datasets provide microdata. Three are macrodata only.

| Datasets  | Notes   | Mechanism        | Sponsor<br>(Data Collector) | Content<br>Variable Counts: <sup>1</sup> |                       | Primary Unit<br>of Analysis      | Granularity <sup>II</sup> | Constraints & Use*<br>Skill level/<br>Accessibility |
|---|---|------------------|-----------------------------|--|-----------------------|----------------------------------|---------------------------|---|
|   |   |                  |                             | Health Care                              | Non-Healthcare        |                                  |                           |   |
| <b>ACS</b><br>American Community Survey   | Extensive geographic and demographic drill downs on disability and health insurance.                  | Survey           | Census Bureau               | Healthcare<br>171                        | Non-Healthcare<br>199 | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>ASEC</b><br>Annual Social and Economic Supplement to the Current Population Survey | Labor force data with health insurance, out-of-pocket \$ and health status fields.                    | Survey           | BLS<br>(Census Bureau)      | Healthcare<br>182                        | Non-Healthcare<br>479 | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>MEPS-HC</b><br>Medical Expenditure Panel Survey, Household Component               | Person-level health expenditures with longitudinal capabilities.                                      | Survey           | AHRQ<br>(Westat)            | Healthcare<br>1,252                      | Non-Healthcare<br>330 | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>NHANES</b><br>National Health and Nutrition Examination Survey                     | Survey combines interviews and physical examination, including lab tests.                             | Survey           | NCHS/CDC                    | Healthcare<br>1,733                      | Non-Healthcare<br>147 | Person                           | Microdata                 | Requires Skills**                                   |
| <b>NHIS</b><br>National Health Interview Survey                                       | Principal source of information on health of U.S. population. Robust demographic, socioeconomic data. | Survey           | CDC<br>(NCHS)               | Healthcare<br>1,388                      | Non-Healthcare<br>212 | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>MEPS-IC</b><br>Medical Expenditure Panel Survey, Insurance Component               | Factors contributing to use of employer sponsored insurance. Premiums and cost sharing.               | Survey           | AHRQ<br>(Census Bureau)     | Healthcare<br>153                        | Non-Healthcare<br>22  | Employers,<br>Employees          | Macrodata                 | Ready-to-use  |
| <b>Medicaid</b><br>(various program data)   | Actual enrollment data. Breakouts of new eligibility categories created by ACA.                       | Admin            | CMS                         | Healthcare<br>301                        | Non-Healthcare<br>8   | Person                           | Macrodata                 | Ready-to-use  |
| <b>NHEA</b><br>National Health Expenditure Accounts                                   | Official estimates of healthcare spending in U.S. Includes care, admin, research and infrastructure.  | Multiple Sources | HHS                         | Healthcare<br>640                        | Non-Healthcare<br>14  | Services,<br>Payers,<br>Sponsors | Macrodata                 | Ready-to-use  |

BLS: Bureau of Labor Statistics

AHRQ: Agency for Healthcare Research and Quality

CDC: Centers for Disease Control and Prevention

NCHS: National Center for Health Statistics

CMS: Centers for Medicare and Medicaid Services

HHS: Department of Health and Human Services

<sup>1</sup> "Counts of variables" by topic is a reasonable method of determining a dataset's areas of focus. Each variable from the eight selected datasets has been categorized by subject matter. All ultimately roll up to either "non-healthcare" or "healthcare." Additional detail on content is provided below. (Counts exclude sample weights and variables related to survey administration.)

<sup>II</sup> Microdata consist of information at the unit level and provide the highest degree of analytic flexibility. Macrodata represent a broad range of pre-aggregated data, typically in ready-to-use format, as well as online utilities that enable users to generate their own queries against the underlying source data.

\* All are publicly available at no cost and include data pre- and post -2014.

\*\* For the microdata versions, users must have database management skills and ability to generate population estimates (relatively easy) and margins of error (more complicated) from raw survey data. Microdata are typically too large for MS Excel or MS Access.

## Selected Datasets by Key Dimensions

| Datasets  | Notes   | Mechanism        | Sponsor<br>(Data Collector) | Content<br>Variable Counts: <sup>1</sup> |                    | Primary Unit<br>of Analysis      | Granularity <sup>II</sup> | Constraints & Use*<br>Skill level/<br>Accessibility |
|---|---|------------------|-----------------------------|--|--------------------|----------------------------------|---------------------------|---|
|   |   |                  |                             | Health Care                              | Non-Healthcare     |                                  |                           |   |
| <b>ACS</b><br>American Community Survey   | Extensive geographic and demographic drill downs on disability and health insurance.                  | Survey           | Census Bureau               | Healthcare 171                           | Non-Healthcare 199 | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>ASEC</b><br>Annual Social and Economic Supplement to the Current Population Survey | Labor force data with health insurance, out-of-pocket \$ and health status fields.                    | Survey           | BLS<br>(Census Bureau)      | Healthcare 182                           | Non-Healthcare 479 | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>MEPS-HC</b><br>Medical Expenditure Panel Survey, Household Component               | Person-level health expenditures with longitudinal capabilities.                                      | Survey           | AHRQ<br>(Westat)            | Healthcare 1,252                         | Non-Healthcare 330 | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>NHANES</b><br>National Health and Nutrition Examination Survey                     | Survey combines interviews and physical examination, including lab tests.                             | Survey           | NCHS/CDC                    | Healthcare 1,733                         | Non-Healthcare 147 | Person                           | Microdata                 | Requires Skills**                                   |
| <b>NHIS</b><br>National Health Interview Survey                                       | Principal source of information on health of U.S. population. Robust demographic, socioeconomic data. | Survey           | CDC<br>(NCHS)               | Healthcare 1,388                         | Non-Healthcare 212 | Person                           | Micro<br>Macrodata        | Requires Skills**                                   |
| <b>MEPS-IC</b><br>Medical Expenditure Panel Survey, Insurance Component               | Factors contributing to use of employer sponsored insurance. Premiums and cost sharing.               | Survey           | AHRQ<br>(Census Bureau)     | Healthcare 153                           | Non-Healthcare 22  | Employers,<br>Employees          | Macrodata                 | Ready-to-use  |
| <b>Medicaid</b><br>(various program data)   | Actual enrollment data. Breakouts of new eligibility categories created by ACA.                       | Admin            | CMS                         | Healthcare 301                           | Non-Healthcare 8   | Person                           | Macrodata                 | Ready-to-use  |
| <b>NHEA</b><br>National Health Expenditure Accounts                                   | Official estimates of healthcare spending in U.S. Includes care, admin, research and infrastructure.  | Multiple Sources | HHS                         | Healthcare 640                           | Non-Healthcare 14  | Services,<br>Payers,<br>Sponsors | Macrodata                 | Ready-to-use  |

BLS: Bureau of Labor Statistics  
AHRQ: Agency for Healthcare Research and Quality  
CDC: Centers for Disease Control and Prevention

NCHS: National Center for Health Statistics  
CMS: Centers for Medicare and Medicaid Services  
HHS: Department of Health and Human Services

*Six of the datasets focus on healthcare. Two are "non-healthcare" but contain extensive demographic, social, economic and geographic capabilities against which their few healthcare variables may be analyzed.*

<sup>1</sup> "Counts of variables" by topic is a reasonable method of determining a dataset's areas of focus. Each variable from the eight selected datasets has been categorized by subject matter. All ultimately roll up to either "non-healthcare" or "healthcare." Additional detail on content is provided below. (Counts exclude sample weights and variables related to survey administration.)

<sup>II</sup> Microdata consist of information at the unit level and provide the highest degree of analytic flexibility. Macrodata represent a broad range of pre-aggregated data, typically in ready-to-use format, as well as online utilities that enable users to generate their own queries against the underlying source data.

\* All are publicly available at no cost and include data pre- and post -2014.

\*\* For the microdata versions, users must have database management skills and ability to generate population estimates (relatively easy) and margins of error (more complicated) from raw survey data. Microdata are typically too large for MS Excel or MS Access.

Additional detail on the “Content” dimension

Variable Counts by Content Category by Dataset

|  |         |
|--|---------|
|  | Present |
|  | Better  |
|  | Best    |

| Categories of Variables   |                       |                                | Total | Datasets  |      |         |        |       |  |   |      |     |
|---|-----------------------|--------------------------------|-------|---|------|---------|--------|-------|--|---|------|-----|
|   |                       |                                |       | Microdata (Macrodata also provided)             |      |         |        |       | Macrodata Only   |   |      |     |
| 1   | 2                     | 3                              | Total | ACS   | ASEC | MEPS-HC | NHANES | NHIS  | MEPS-IC  | M'caid  | NHEA |     |
|   |                       |                                | 6,806 | 216   | 661  | 1,582   | 1,880  | 1,600 | 175  | 38  | 654  |     |
| <b>Healthcare</b><br><br><i>The “Access” and “Healthiness” data are beyond what is available in administrative datasets...</i>                              | Access                | Ability to Pay for Care        | 1,053 | 10  | 157  | 546     | 17     | 203   | 87   | 22  | 11   |     |
|   |                       | Ability to Get Care            | 546   | <i>All address ability to pay for care.</i>     |      |         | 178    | 106   | 262  | <i>MEPS-HC and NHIS are strong on “getting” care.</i>   |      |     |
|   | Cost                  | Charges                        | 19    | <i>NHEA and MEPS-HC are strongest on costs.</i> |      |         | 19     |       |  |   |      |     |
|   |                       | Encounters                     | 107   |   |      | 21      | 69     | 17    |  |   |      |     |
|   |                       | Expenditures                   | 894   |   |      | 300     |        |       |  |   |      |     |
|   |                       | Expenditures by Sponsor        | 128   |   | 15   |         |        |       | 4  | 66  | 8    | 586 |
|   |                       | Behavior/Attitude              | 290   |   |      | 4       | 224    | 62    |  |   |      |     |
|   |                       | Body Composition               | 160   |   |      | 1       | 151    | 8     |  |   |      |     |
|   | Healthiness           | Child-Specific Problems        | 39    |   |      | 28      |        |       | 11   |   |      |     |
|   |                       | Clinical Results               | 376   |   |      |         | 376    |       |  | <i>NHANES is the only dataset to include clinical results. NHIS and MEPS-HC are also strong on healthiness.</i> |      |     |
|   |                       | Condition                      | 1,095 |   |      | 81      | 311    | 703   |  |   |      |     |
|   |                       | Days Lost Due to Illness       | 8     |   |      | 3       |        |       | 5  |   |      |     |
|   |                       | Diet                           | 406   |   |      |         | 406    |       |  |   |      |     |
|   |                       | Functional Limitation Status   | 189   | 7   | 9    | 39      | 55     | 79    |  |   |      |     |
|   |                       | 85                             |       | 1   | 32   | 18      | 34     |       |  |   |      |     |
| <b>Non-Healthcare</b><br><br><i>...The “non-healthcare” variables are likewise more comprehensive than those typically provided in administrative data.</i> | Demographic           | Age                            | 20    | 2   | 2    | 7       | 5      | 3     |  |   | 1    |     |
|   |                       | Race/Ethnicity                 | 44    | 12  | 7    | 8       | 2      | 15    |  |   |      |     |
|   |                       | Sex                            | 9     | 1   | 1    | 1       | 2      | 3     |  |   | 1    |     |
|   | Economic              | Income                         | 303   | 13  | 211  | 25      | 3      | 51    | <i>ASEC and MEPS-HC are strong on economic issues, particularly income, labor force participation and job characteristics.</i> |   |      |     |
|   |                       | Other Benefits                 | 38    | 1   | 19   |         | 10     | 8     |  |   |      |     |
|   |                       | Taxes                          | 26    |   | 20   | 6       |        |       |  |   |      |     |
|   |                       | Food Security                  | 10    |   |      |         |        | 10    |  |   |      |     |
|   |                       | Job Characteristics            | 229   | 15  | 47   | 139     | 4      | 12    | 12   |   |      |     |
|   |                       | Labor Force                    | 123   | 13  | 37   | 48      | 5      | 15    | 5  |   |      |     |
|   |                       | Economic Indicators/Population | 7     |   |      |         |        |       |  |   |      | 7   |
|   | Geographic            | Specified Areas                | 47    | 8*  | 13   | 4       |        | 4     | 5  | 8   | 5    |     |
|   | Housing               | Financial                      | 32    | 27  | 4    |         |        | 1     |  |   |      |     |
|   |                       | Physical                       | 28    | 17  | 5    |         |        | 6     |  |   |      |     |
|   | Social                | Family/Household Composition   | 306   | 33  | 78   | 65      | 88     | 42    |  |   |      |     |
|   |                       | Marital Status                 | 23    | 6   | 2    | 8       | 2      | 5     |  |   |      |     |
|   |                       | Military Status                | 37    | 13  | 5    | 6       | 2      | 11    |  |   |      |     |
|   |                       | Neighborhood Characteristics   | 5     |   |      |         |        | 5     |  |   |      |     |
| Child Care  |                       | 12                             |       | 9   |      |         | 3      |       |  |   |      |     |
| Education   |                       | 25                             | 7     | 4   | 6    | 4       | 4      |       |  |   |      |     |
| Heritage  |                       | 55                             | 17    | 5   | 7    | 19      | 7      |       |  |   |      |     |
| Migration   |                       | 13                             | 3     | 10  |      |         |        |       |  |   |      |     |
|   | Internet/Computer Use | 19                             | 11    |   |      | 1       | 7      |       |  |   |      |     |

**American Community Survey (ACS)**

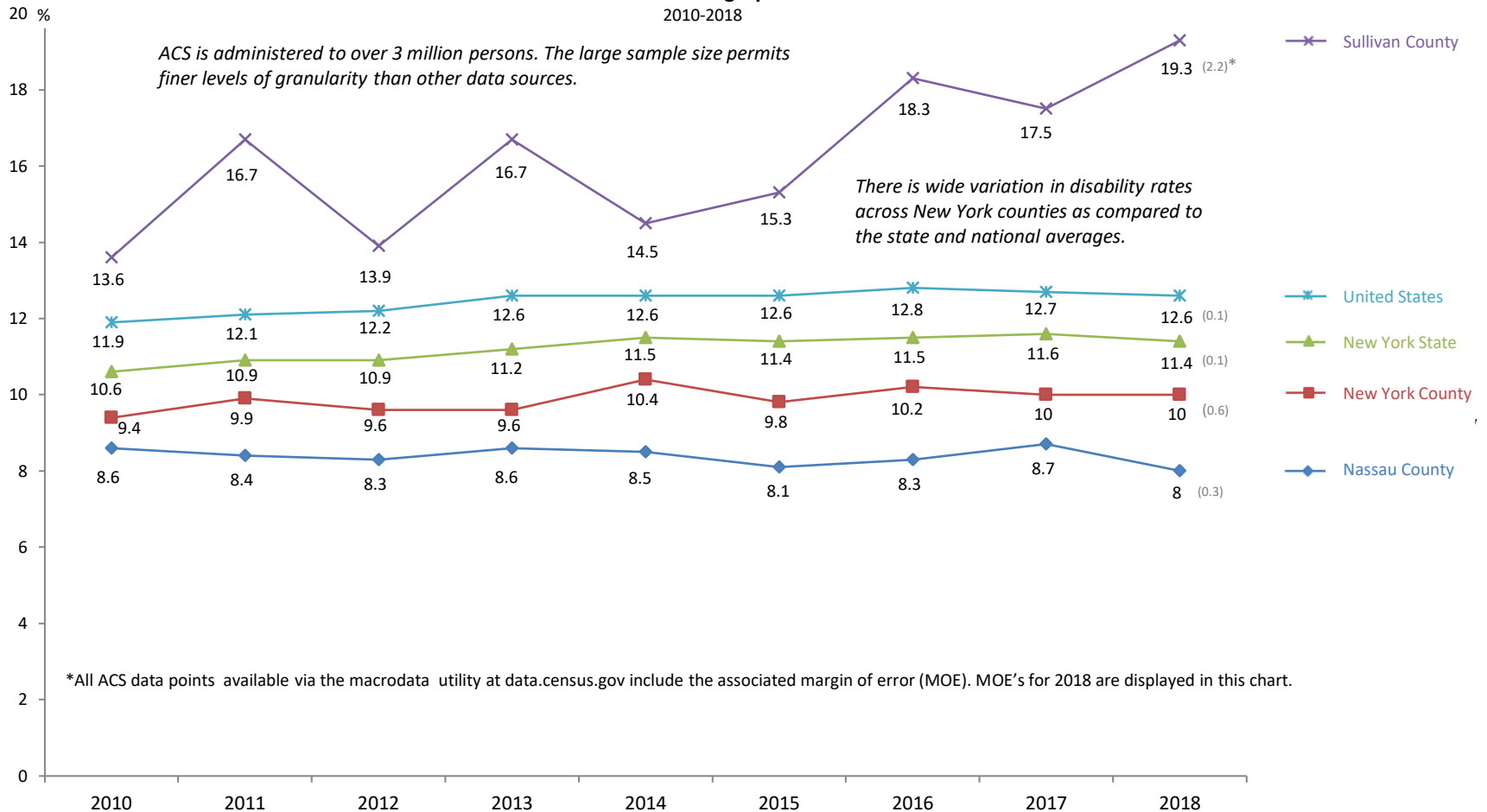
- Broad ranging survey addressing demographic, social, economic and housing information
  - Contains two healthcare variables: health insurance coverage and disabilities
- Robust geographic and demographic drill down capabilities for healthcare (and other) variables
- Conducted annually by US Census Bureau (designed to assist US Congress in designing and funding programs)

**Sample ACS Data**

**Percent of Population with a Disability**

**Selected Geographies**

2010-2018



\*All ACS data points available via the macrodata utility at data.census.gov include the associated margin of error (MOE). MOE's for 2018 are displayed in this chart.

**National Health Interview Survey (NHIS)**

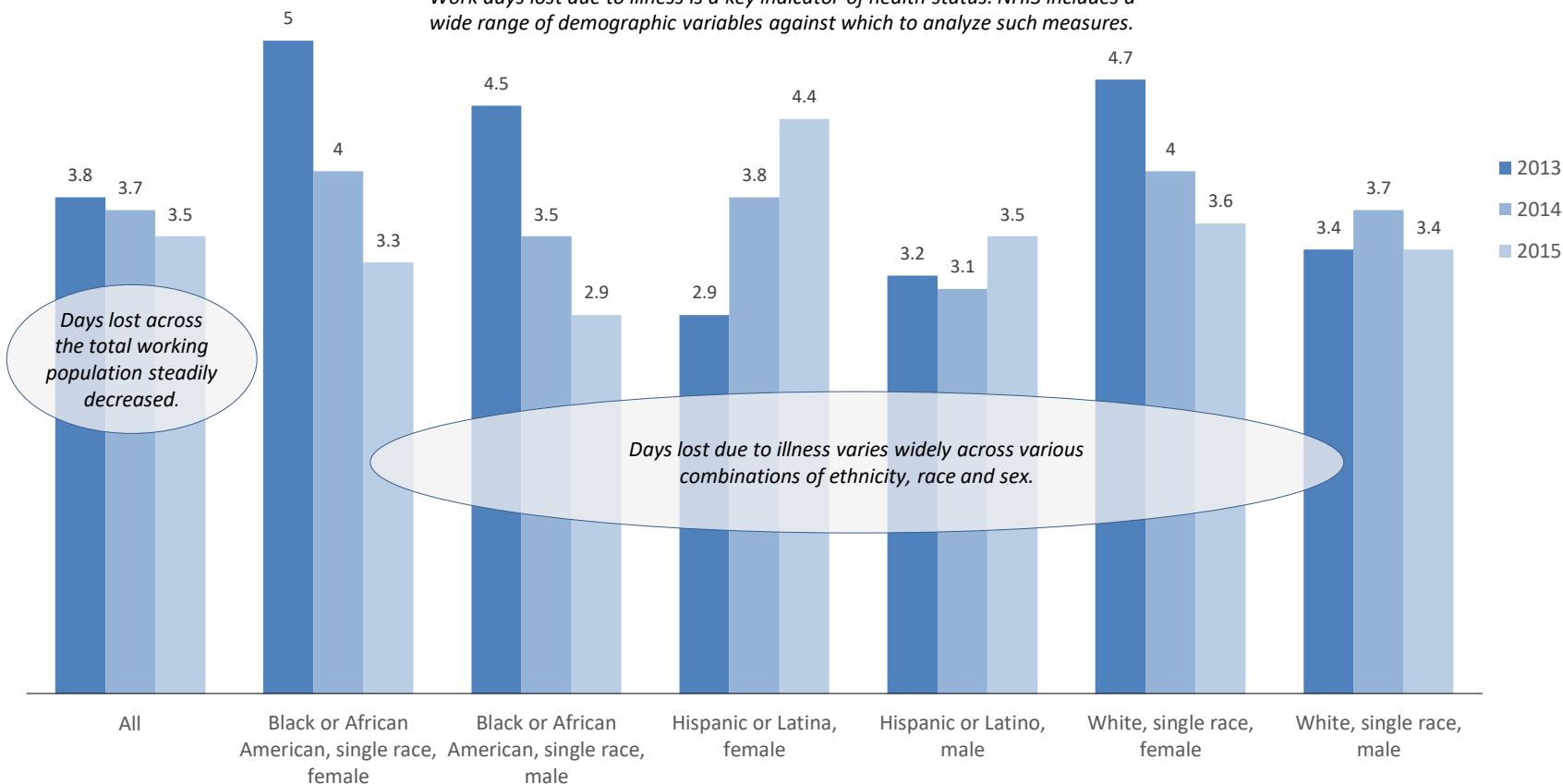
- Principal source of information on health of the U.S. population
- Addresses access, utilization, conditions, limitations, behaviors and insurance coverage
- Enables analysis of comprehensive healthcare data across wide range of demographic and socioeconomic variables
  - Produced annually by the Centers for Disease Control and Prevention (CDC)

**Sample NHIS Data**

**Average Number of Work Days Lost Due to Illness Over Past Twelve Months**

Employed Adults By Ethnicity, Race and Sex, 2013-2015

*Work days lost due to illness is a key indicator of health status. NHIS includes a wide range of demographic variables against which to analyze such measures.*



Source: NHIS. Chart created by Stern Consulting based on data from “Tables of Summary Health Statistics,” Table A-9, [cdc.gov/nchs/nhis/shs/tables.htm](http://cdc.gov/nchs/nhis/shs/tables.htm)

# Summary Thoughts On Selecting Data Sets

- **Think broadly:**

- Complex system
- Many players, all generate data



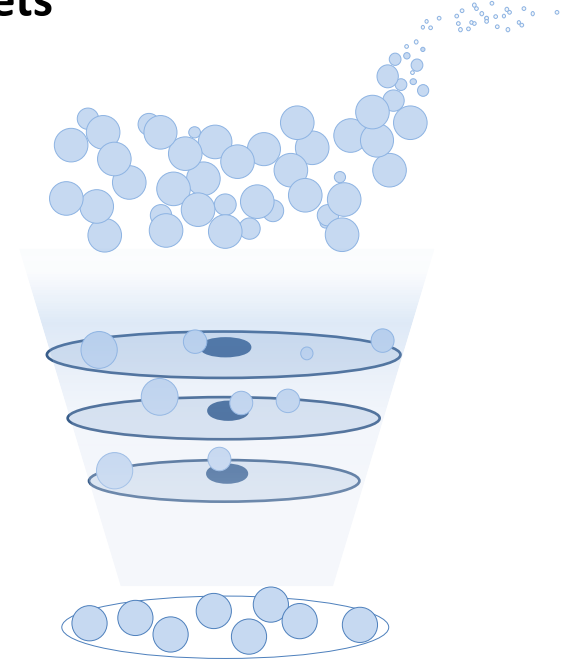
- **Filter wisely:**

- Don't write off "non-healthcare" data
- Micro- & macro-data
- Admin data, surveys, regulatory...
- Many sources: government, private, industry ...



- **Consider these eight:**

- **ACS** (American Community Survey)
- **ASEC** (Annual Social and Economic Supplement to the Current Population Survey)
- **Medicaid** (various program data)
- **MEPS-HC** (Medical Expenditure Panel Survey, Household Component)
- **MEPS-IC** (Medical Expenditure Panel Survey, Insurance Component)
- **NHANES** (National Health and Nutrition Examination Survey)
- **NHEA** (National Health Expenditure Accounts)
- **NHIS** (National Health Interview Survey)





# **Appendix**

I. Additional Resources

II. Sample Data

III. About Us

## Additional Resources, 1 of 2

*In addition to our eight profiled datasets, the following four resources each contribute a particular perspective on national healthcare policy.*

### Selected Other (Non-Administrative) Datasets by Key Dimensions

| Datasets   | Notes   | Dimensions of Data                         |                    |                                |                             |                 | Constraints & Use*<br>Skill level/<br>Accessibility |
|--|---|--|--------------------|--------------------------------|-----------------------------|-----------------|---|
|  |   | Sponsor                                    | Healthcare Content | Mechanism                      | Primary Unit of Analysis    | Granularity     |   |
| <b>BRFSS</b><br>Behavioral Risk Factor Surveillance System | State-based data on health-related risk factors, preventive services and conditions (with significant overlap with NHIS subject matter) | States in coordination with CDC            | Primary focus      | Survey                         | Person                      | Micro Macrodata | Requires Skills**                                   |
| <b>SAHIE</b><br>Small Area Health Insurance Estimates      | Health insurance coverage by county by demographic and economic characteristics   | Census Bureau                              | Primary focus      | Based on ACS/other Census data | County                      | Macrodata       | Ready-to-use  |
| <b>AHRF</b><br>Area Health Resources File                  | Healthcare workforce data, training, utilization, expenditures, shortage areas  | Health Resources & Services Administration | Primary focus      | Multiple Sources               | County, Congress'l District | Macrodata       | Requires Skills**                                   |
| <b>SIPP</b><br>Survey of Income and Program Participation  | Income and government transfer data and multiple health variables; longitudinal capabilities  | Census Bureau                              | Secondary Focus    | Survey                         | Person                      | Microdata       | Requires Skills**                                   |

CDC: Centers for Disease Control and Prevention

\* All are publicly available at no cost and include data pre- and post -2014.

\*\* For the microdata versions, users must have database management skills and ability to generate population estimates (relatively easy) and margins of error (more complicated) from raw survey data. Microdata are typically too large for MS Excel or MS Access.

## Additional Resources, 2 of 2

*“Data, Tools and Statistics”  
from the National Library of  
Medicine is particularly useful.*

### Dataset Directories

| List   | Sponsor   |
|--|---|
| <b>Data, Tools and Statistics</b>                            | National Institutes of Health, National Library of Medicine<br>National Information Center on Health Services Research and Health Care Technology (NICHSR)<br><a href="https://hsric.nlm.nih.gov/hsric_public/topic/datasites/">https://hsric.nlm.nih.gov/hsric_public/topic/datasites/</a> |
| <b>Directory of Health and Human Services Data Resources</b> | U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation<br><a href="https://aspe.hhs.gov/directory-health-and-human-services-data-resources">https://aspe.hhs.gov/directory-health-and-human-services-data-resources</a>                |
| <b>Guide to HHS Surveys and Data Resources</b>               | U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation<br><a href="http://aspe.hhs.gov/basic-report/guide-hhs-surveys-and-data-resources">http://aspe.hhs.gov/basic-report/guide-hhs-surveys-and-data-resources</a>                    |
| <b>HealthData.gov</b>  | U.S. Department of Health and Human Services, Office of the Chief Technology Officer<br><a href="https://healthdata.gov/">https://healthdata.gov/</a>   |
| <b>HealthyPeople.gov Data Sources</b>                        | U.S. Department of Health and Human Services<br><a href="https://www.healthypeople.gov/2020/data-search/Data-Sources">https://www.healthypeople.gov/2020/data-search/Data-Sources</a>   |
| <b>Data Sources</b>  | Deloitte, DATA USA<br><a href="https://datausa.io/about/datasets/">https://datausa.io/about/datasets/</a>   |

*Healthcare data may not always reside in obvious places. There are many federal and state agencies and departments, any of which may provide healthcare data. Lists of federal agencies may be found at:*

- Federal Register ([www.federalregister.gov/agencies](http://www.federalregister.gov/agencies))
  - USA.gov ([www.usa.gov/federal-agencies](http://www.usa.gov/federal-agencies)).

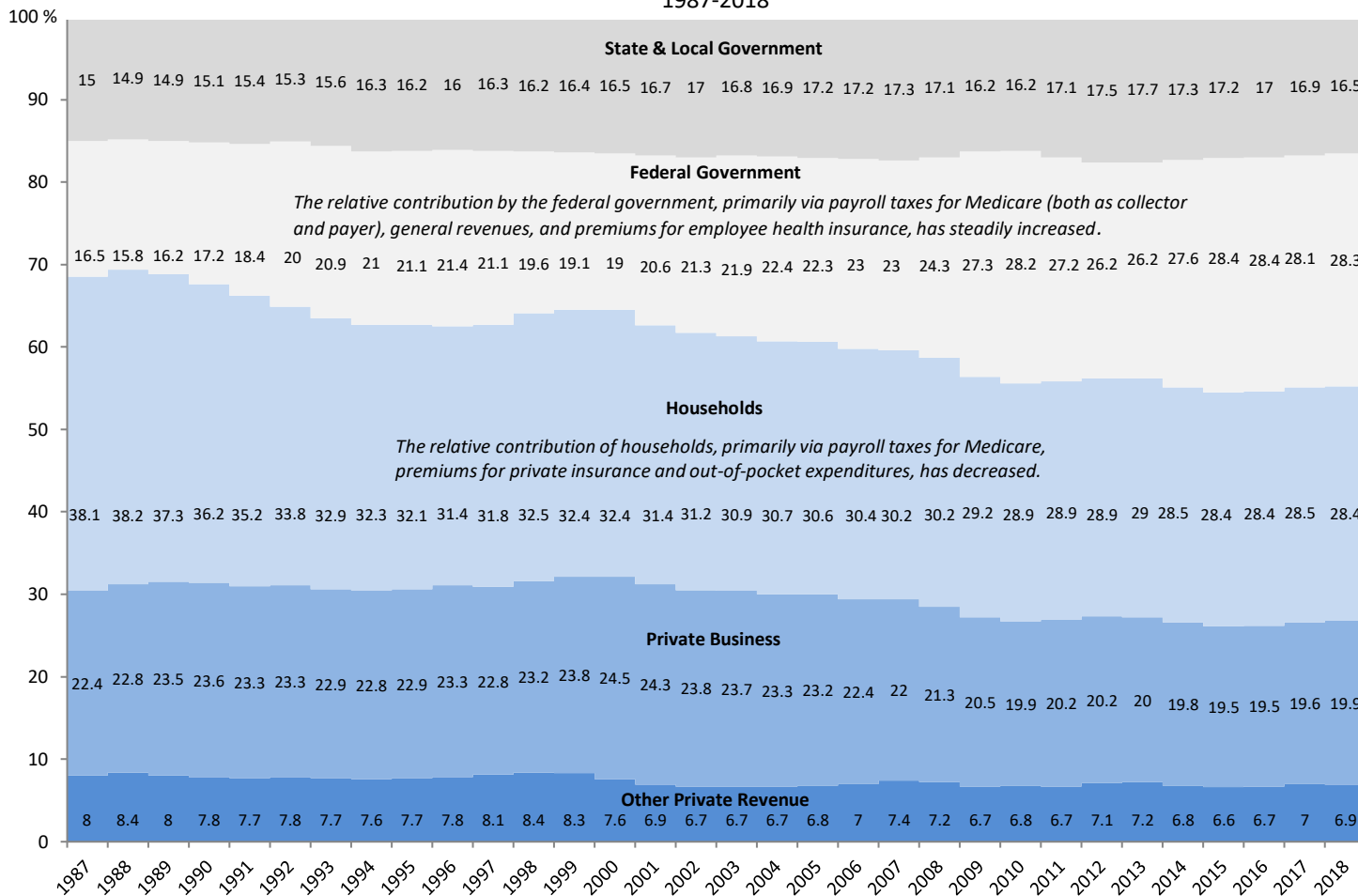
**National Health Expenditure Accounts (NHEA)**

- Official estimates of healthcare expenditures in the U.S.
- Reports personal care, administration, research and infrastructure spending by payer type
- Breaks out expenditures by sponsors of payers (e.g. households are “sponsors” of private insurance via premiums)
  - Produced annually by the U.S. Department of Health and Human Services (HHS)

**Sample NHEA Data**

**Expenditures by Type of Sponsor (Percent Distribution)**

1987-2018



**Sponsors vs. Payers**  
 NHEA addresses both payers and sponsors of healthcare expenditures. Payers are the entities, such as private insurance companies, that transfer funds to the providers of care. Sponsors are the entities, such as households and businesses, that finance those payments. The relative burden borne by sponsors fluctuates over time.

Source: NHEA. Chart created by Stern Consulting based on data from Table 5, National Health Expenditures by Type of Sponsor, cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical

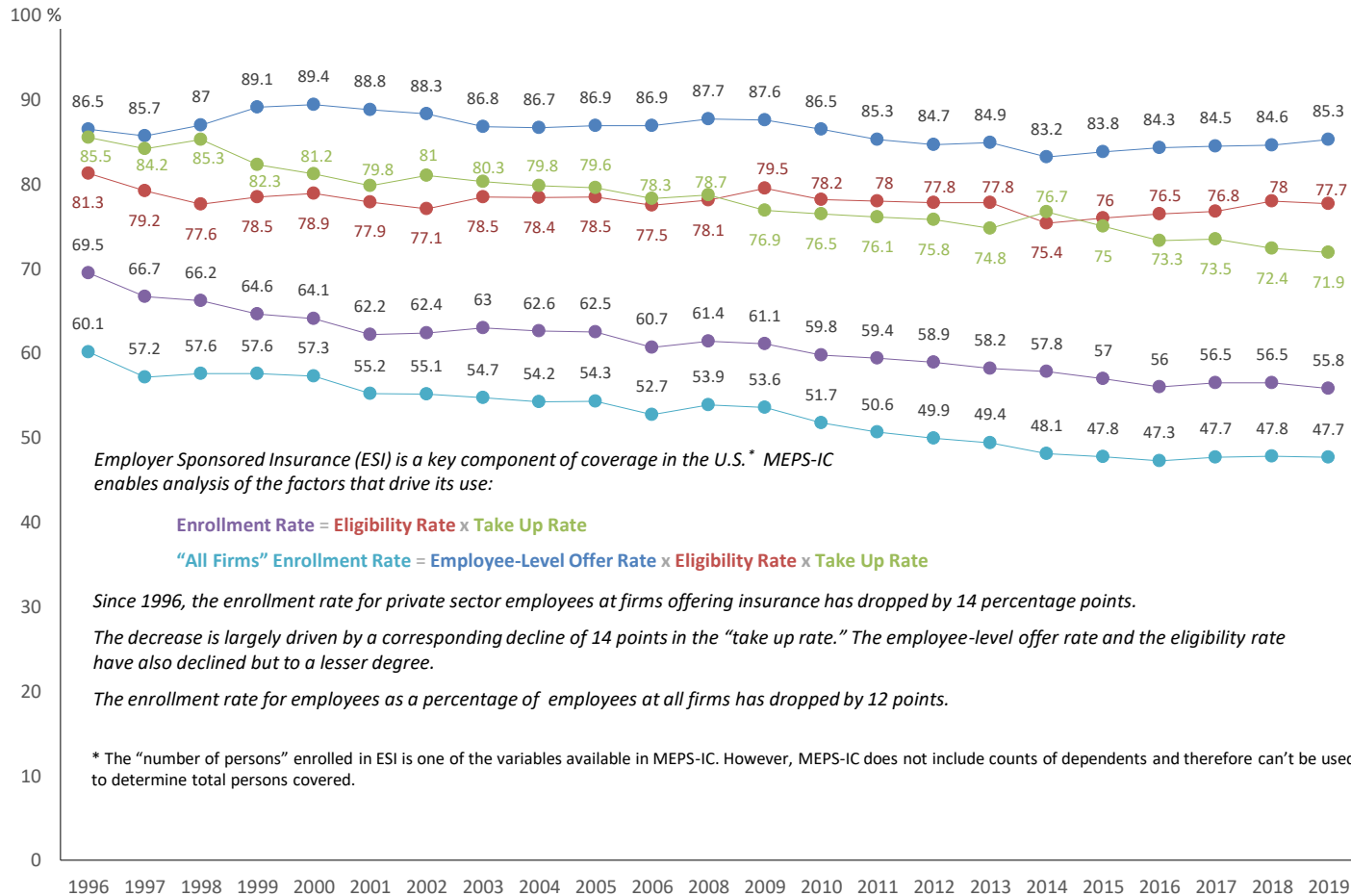
**Medical Expenditure Panel Survey-Insurance Component (MEPS-IC)**

- Deals exclusively with Employer-Sponsored Health Insurance (ESI)
- Administered to employers (private and public), not to persons
- Addresses premiums, cost-sharing, and extent to which employers offer, and employees accept, coverage
  - Produced annually by the Agency for Healthcare Research and Quality (AHRQ)

**Sample MEPS-IC Data**

**Factors That Determine Employer Sponsored Insurance (ESI) Enrollment**

Private Sector Employees, United States, 1996-2018



**Employee-Level Offer Rate:** Employees that work in firms offering health insurance as a percent of all employees

**Eligibility Rate:** Employees that are eligible as a percent of all employees that work in firms offering insurance

**Take Up Rate:** Employees that accept the insurance offered as a percent of eligible employees

**Enrollment Rate:** Employees that accept the insurance offered as a percent of all employees *at firms offering insurance*

**“All Firms” Enrollment Rate:** Employees that accept the insurance offered as a percent of all employees *at all firms* (Calculated by Stern Consulting based on MEPS-IC data)

Employer Sponsored Insurance (ESI) is a key component of coverage in the U.S.\* MEPS-IC enables analysis of the factors that drive its use:

**Enrollment Rate = Eligibility Rate x Take Up Rate**

**“All Firms” Enrollment Rate = Employee-Level Offer Rate x Eligibility Rate x Take Up Rate**

Since 1996, the enrollment rate for private sector employees at firms offering insurance has dropped by 14 percentage points.

The decrease is largely driven by a corresponding decline of 14 points in the “take up rate.” The employee-level offer rate and the eligibility rate have also declined but to a lesser degree.

The enrollment rate for employees as a percentage of employees at all firms has dropped by 12 points.

\* The “number of persons” enrolled in ESI is one of the variables available in MEPS-IC. However, MEPS-IC does not include counts of dependents and therefore can’t be used to determine total persons covered.

### Annual Social and Economic Supplement (ASEC) to the Current Population Survey

- Primary source of monthly employment statistics along with a broad range of other social and economic topics
  - Used by Census Bureau as primary source of health insurance coverage information
- Strong on coverage obtained via exchanges; Also addresses health status and out-of-pocket medical expenditures
  - Conducted by the Census Bureau on behalf of the Bureau of Labor Statistics

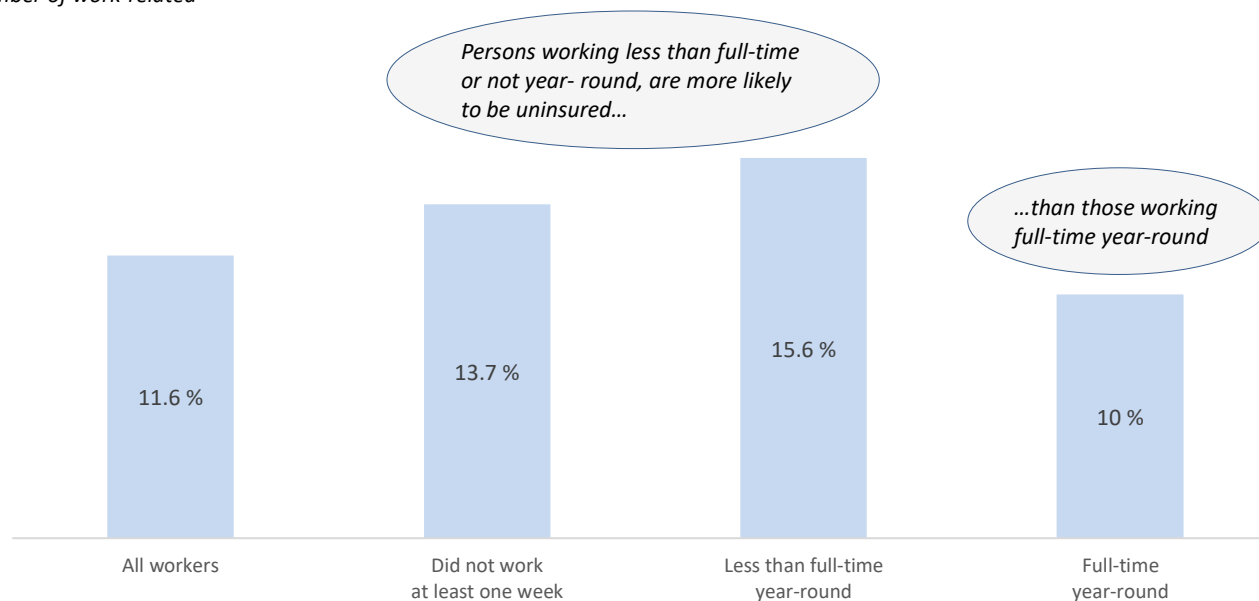
## Sample ASEC Data

### Percent of Persons With No Health Insurance Coverage

Broken out by Work Experience, 2019

Ages 15-64

*ASEC's focus on labor and employment enables analysis of healthcare issues across a number of work-related variables.*



Note: Coverage status is "as of date of interview."

Source: ASEC. Chart created by Stern Consulting based on data from Table H-02. Health Insurance Current Coverage Status and Type of Coverage by Selected Characteristics: 2019, [census.gov/data/tables/2019/demo/cps/health-insurance.html](https://www.census.gov/data/tables/2019/demo/cps/health-insurance.html)

**Medical Expenditure Panel Survey-Household Component (MEPS-HC)**

- Particularly broad range of healthcare and non-healthcare variables
- Person-level expenditures complement aggregate totals in National Health Expenditure Accounts (NHEA)
  - One of the few profiled datasets with longitudinal capabilities
- Produced annually by the Agency for Healthcare Research and Quality (AHRQ)
  - MEPS-HC is based on a subsample from the prior year’s National Health Interview Survey (NHIS)

**Sample MEPS-HC Data**

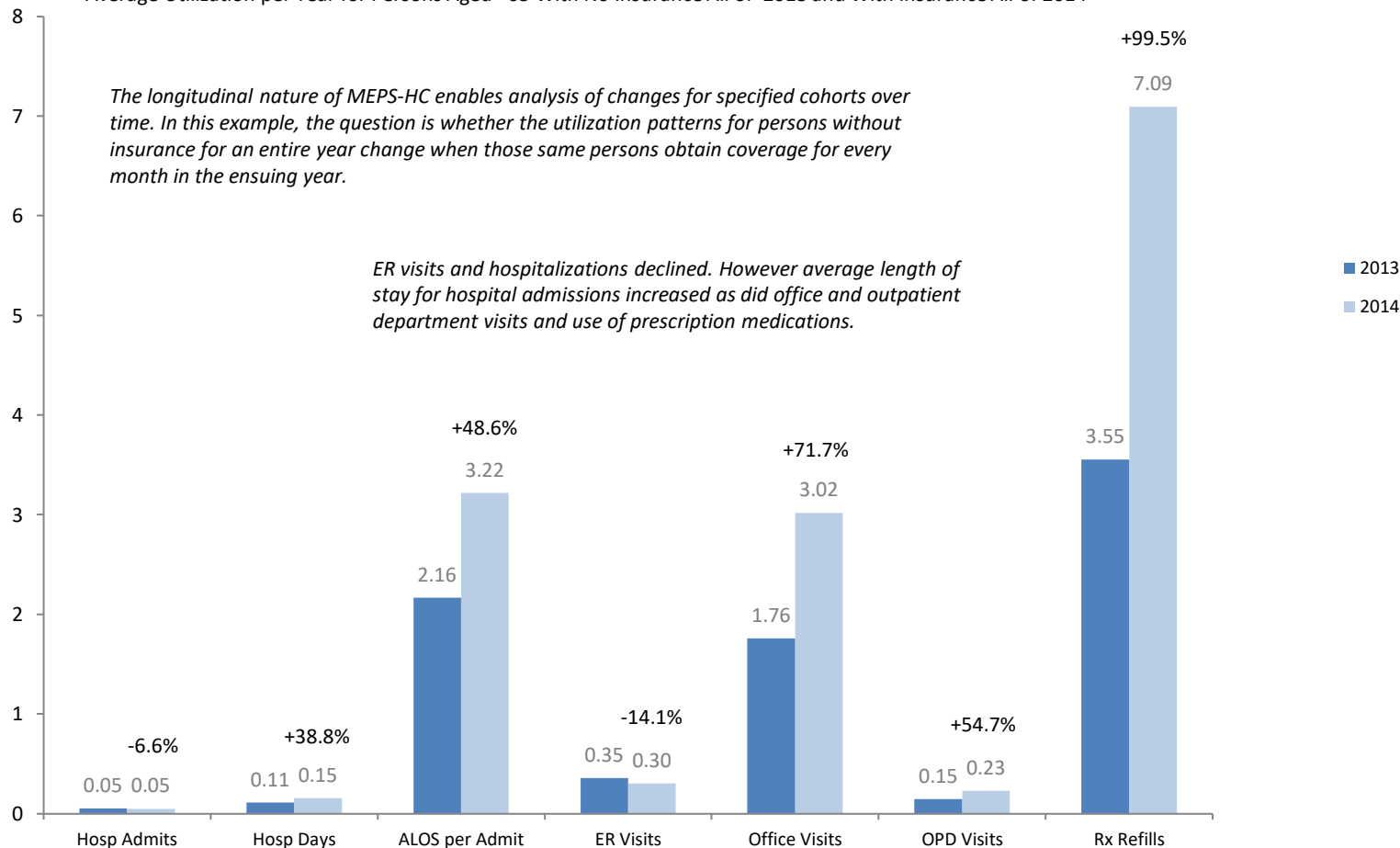
**Changes in Utilization Upon Obtaining Health Insurance**

Average Utilization per Year for Persons Aged <65 With No Insurance All of 2013 and With Insurance All of 2014

*The longitudinal nature of MEPS-HC enables analysis of changes for specified cohorts over time. In this example, the question is whether the utilization patterns for persons without insurance for an entire year change when those same persons obtain coverage for every month in the ensuing year.*

*ER visits and hospitalizations declined. However average length of stay for hospital admissions increased as did office and outpatient department visits and use of prescription medications.*

Average Number of Encounters



Source: MEPS-HC. Chart created by Stern Consulting based on data from Panel 18 Longitudinal File 2013-2014. [meps.ahrq.gov/mepsweb/data\\_stats/download\\_data\\_files.jsp](https://meps.ahrq.gov/mepsweb/data_stats/download_data_files.jsp)

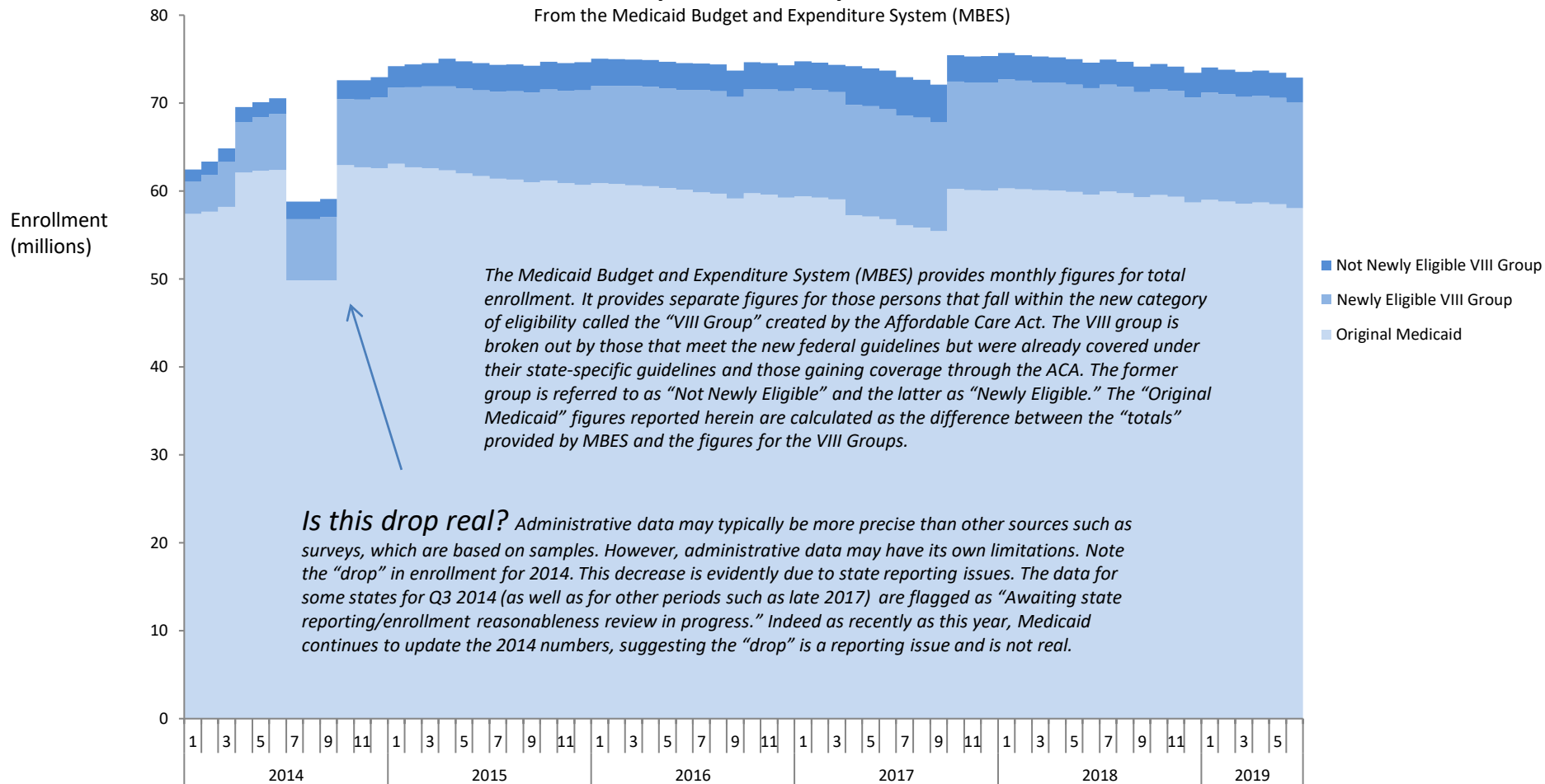
**Medicaid Program Data**

- Actual enrollment figures from multiple programs:  
[Medicaid Budget and Expenditure System (MBES), Performance Indicator Project (PIP), Statistical Enrollment Data System (SEDS)]
- Data available for various eligibility types including new Affordable Care Act (ACA) categories
- Reports provided by the Centers for Medicare and Medicaid Services (CMS) by month, quarter and year depending on program

**Sample Medicaid Program Data**

**Enrollment by Month, January 2014- June 2019**

From the Medicaid Budget and Expenditure System (MBES)



Source: Medicaid Budget and Expenditure System. Chart created by Stern Consulting based on Quarterly Medicaid MBES Enrollment (includes monthly breakouts). [medicaid.gov/medicaid/national-medicare-chip-program-information/medicaid-chip-enrollment-data/medicaid-enrollment-data-collected-through-mbes/index.html](https://www.medicare.gov/medicaid/national-medicare-chip-program-information/medicaid-chip-enrollment-data/medicaid-enrollment-data-collected-through-mbes/index.html)



## About Stern Consulting

**Stern Consulting LLC** provides specialty analytic and consulting services to healthcare leaders, hospital systems, healthcare companies, and investors. For more information, see [www.sternconsulting.com](http://www.sternconsulting.com).



**Brian Williams** oversees Stern Consulting's healthcare database and analysis functions. He has over 30 years of experience in healthcare decision support. Mr. Williams earned an MBA with concentration in healthcare management from Boston University and a BA in political science from Holy Cross College.  
[bwilliams@sternconsulting.com](mailto:bwilliams@sternconsulting.com)



**David M. C. Stern** is the founder and president of Stern Consulting. He has advised some of the pre-eminent organizations in the healthcare industry. He earned an MBA from the Yale School of Management and a BA in economics from Yale College.  
[dstern@sternconsulting.com](mailto:dstern@sternconsulting.com)

## About NAHDO

**The National Association of Health Data Organizations (NAHDO)** is a national non-profit membership and educational association dedicated to improving health care data collection and use. NAHDO's members include state and private health data organizations that maintain statewide health care databases and stakeholders of these databases. For more information, see [www.nahdo.org](http://www.nahdo.org).



The current slides were prepared for presentation at NAHDO's 2020 annual conference entitled *Building a Bridge Between Data and Policy*.