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
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.
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

- **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**
- **Relevant to Objectives?**
- **Accessible?**
- **Appropriate Source?**
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**

- Relevant to Objectives?
- Accessible?
- Appropriate Source?

**Data Dimensions Guide the Search**

**Universe of Datasets**

Datasets are defined by six key dimensions.
“Datasets to Evaluate the Impact of National Healthcare Policy”

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

**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?

- Must address access, costs, and/or healthiness
- Must address all persons, including uninsured and those w/o claims
- Must be available to the public, pre and post 2014
- Must be national in scope
Datasets to Evaluate the Impact of National Healthcare Policy

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.

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

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

Case Study
Selection Process

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.

**“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.**
“Datasets to Evaluate the Impact of National Healthcare Policy”

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

Content
- Access
- Charges/Cost of Care
- Utilization
- Health Status
- Quality/Satisfaction
- Clinical Classification
- 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.

Accessible?

Relevant to Objectives?

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.

Datasets to Evaluate the Impact of National Healthcare Policy

Project Screening Criteria

- Relevant to Objectives?
- Accessible?
- Appropriate Source?
- Must address access, costs, and/or healthiness
- Must be available to the public, pre and post 2014
- Must be national in scope

Dimensions

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

- Granularity
  - Microdata
  - Macrodta
    - Tables
    - Online Queries
  - Other

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

- 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.
## Dimensions of Data

### Note

1. "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.)

2. 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.

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

4. 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.

### Case Study

**Selection Process**

**“Datasets to Evaluate the Impact of National Healthcare Policy”**

**SelectedDatasets by Key Dimensions**

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Notes</th>
<th>Mechanism</th>
<th>Sponsor (Data Collector)</th>
<th>Content Variable Counts:†</th>
<th>Primary Unit of Analysis</th>
<th>Granularity□</th>
<th>Constraints &amp; Use*</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACS American Community Survey</td>
<td>Extensive geographic and demographic drill downs on disability and health insurance.</td>
<td>Survey</td>
<td>Census Bureau</td>
<td>Healthcare: 17, Non-Healthcare: 199</td>
<td>Person</td>
<td>Micro</td>
<td>Requires Skills**</td>
</tr>
<tr>
<td>MEPS-HC Medical Expenditure Panel Survey, Household Component</td>
<td>Person-level health expenditures with longitudinal capabilities.</td>
<td>Survey</td>
<td>AHRQ (Westat)</td>
<td>Healthcare: 1,252, Non-Healthcare: 330</td>
<td>Person</td>
<td>Micro</td>
<td>Requires Skills**</td>
</tr>
<tr>
<td>NHANES National Health and Nutrition Examination Survey</td>
<td>Survey combines interviews and physical examination, including lab tests.</td>
<td>Survey</td>
<td>NCHS/CDC</td>
<td>Healthcare: 1,733, Non-Healthcare: 147</td>
<td>Person</td>
<td>Micro</td>
<td>Requires Skills**</td>
</tr>
<tr>
<td>NHIS National Health Interview Survey</td>
<td>Principal source of information on health of U.S. population. Robust demographic, socioeconomic data.</td>
<td>Survey</td>
<td>CDC (NCHS)</td>
<td>Healthcare: 1,388, Non-Healthcare: 212</td>
<td>Person</td>
<td>Micro</td>
<td>Requires Skills**</td>
</tr>
<tr>
<td>Medicaid (various program data)</td>
<td>Actual enrollment data. Breakouts of new eligibility categories created by ACA.</td>
<td>Admin</td>
<td>CMS</td>
<td>Healthcare: 30, Non-Healthcare: 8</td>
<td>Person</td>
<td>Macro</td>
<td>Ready-to-use</td>
</tr>
</tbody>
</table>

---

**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  

---

© 2020 Stern Consulting LLC
### Selected Datasets by Key Dimensions

#### Dimensions of Data

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Notes</th>
<th>Mechanism</th>
<th>Sponsor (Data Collector)</th>
<th>Content Variable Counts:</th>
<th>Primary Unit of Analysis</th>
<th>Granularity</th>
<th>Constraints &amp; Use*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ACS</strong></td>
<td>American Community Survey</td>
<td>Extensive geographic and demographic drill downs on disability and health insurance.</td>
<td>Survey</td>
<td>Census Bureau</td>
<td>Healthcare 17, Non-Healthcare 199</td>
<td>Person</td>
<td>Micro Macrodata</td>
</tr>
<tr>
<td><strong>MEPS-HC</strong></td>
<td>Medical Expenditure Panel Survey, Household Component</td>
<td>Person-level health expenditures with longitudinal capabilities.</td>
<td>Survey</td>
<td>AHRQ (Westat)</td>
<td>Healthcare 1,252, Non-Healthcare 330</td>
<td>Person</td>
<td>Micro Macrodata</td>
</tr>
<tr>
<td><strong>NHANES</strong></td>
<td>National Health and Nutrition Examination Survey</td>
<td>Survey combines interviews and physical examination, including lab tests.</td>
<td>Survey</td>
<td>NCHS/CDC</td>
<td>Healthcare 1,733, Non-Healthcare 147</td>
<td>Person</td>
<td>Microdata</td>
</tr>
<tr>
<td><strong>NHIS</strong></td>
<td>National Health Interview Survey</td>
<td>Principal source of information on health of U.S. population. Robust demographic, socioeconomic data.</td>
<td>Survey</td>
<td>CDC (NCHS)</td>
<td>Healthcare 1,388, Non-Healthcare 212</td>
<td>Person</td>
<td>Micro Macrodata</td>
</tr>
<tr>
<td><strong>MEPS-IC</strong></td>
<td>Medical Expenditure Panel Survey, Insurance Component</td>
<td>Factors contributing to use of employer sponsored insurance. Premiums and cost sharing.</td>
<td>Survey</td>
<td>AHRQ (Census Bureau)</td>
<td>Healthcare 153, Non-Healthcare 22</td>
<td>Employers, Employees</td>
<td>Macrodata</td>
</tr>
<tr>
<td><strong>Medicaid</strong></td>
<td>(various program data)</td>
<td>Actual enrollment data. Breakouts of new eligibility categories created by ACA.</td>
<td>Admin</td>
<td>CMS</td>
<td>Healthcare 30, Non-Healthcare 8</td>
<td>Person</td>
<td>Macrodata</td>
</tr>
<tr>
<td><strong>NHEA</strong></td>
<td>National Health Expenditure Accounts</td>
<td>Official estimates of healthcare spending in U.S. Includes care, admin, research and infrastructure.</td>
<td>Multiple Sources</td>
<td>HHS</td>
<td>Healthcare 640, Non-Healthcare 14</td>
<td>Services, Payers, Sponsors</td>
<td>Macrodata</td>
</tr>
</tbody>
</table>

---

**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

1. “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.)

2. 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.

© 2020 Stern Consulting LLC
### Case Study

#### “Datasets to Evaluate the Impact of National Healthcare Policy”

#### Selected Datasets by Key Dimensions

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Notes</th>
<th>Mechanism</th>
<th>Sponsor (Data Collector)</th>
<th>Content Variable Counts:¹</th>
<th>Primary Unit of Analysis</th>
<th>Constraints &amp; Use*</th>
<th>Granularity¹²</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACS</td>
<td>American Community Survey Extensive geographic and demographic drill downs on disability and health insurance.</td>
<td>Survey</td>
<td>Census Bureau</td>
<td>Healthcare: 17; Non-Healthcare: 199</td>
<td>Person</td>
<td>Micro Data</td>
<td>Requires Skills**</td>
</tr>
<tr>
<td>MEPS-HC</td>
<td>Medical Expenditure Panel Survey, Household Component Person-level health expenditures with longitudinal capabilities.</td>
<td>Survey</td>
<td>AHRQ (Westat)</td>
<td>Healthcare: 1,252; Non-Healthcare: 330</td>
<td>Person</td>
<td>Micro Data</td>
<td>Requires Skills**</td>
</tr>
<tr>
<td>NHANES</td>
<td>National Health and Nutrition Examination Survey Survey combines interviews and physical examination, including lab tests.</td>
<td>Survey</td>
<td>NCHS/CDC</td>
<td>Healthcare: 1,733; Non-Healthcare: 147</td>
<td>Person</td>
<td>Micro Data</td>
<td>Requires Skills**</td>
</tr>
<tr>
<td>NHIS</td>
<td>National Health Interview Survey Principal source of information on health of U.S. population. Robust demographic, socioeconomic data.</td>
<td>Survey</td>
<td>CDC (NCHS)</td>
<td>Healthcare: 1,388; Non-Healthcare: 212</td>
<td>Person</td>
<td>Micro Data</td>
<td>Requires Skills**</td>
</tr>
<tr>
<td>Medicaid (various program data)</td>
<td>Actual enrollment data. Breakouts of new eligibility categories created by ACA.</td>
<td>Admin</td>
<td>CMS</td>
<td>Healthcare: 30; Non-Healthcare: 8</td>
<td>Person</td>
<td>Macro Data</td>
<td>Ready-to-use</td>
</tr>
<tr>
<td>NHEA</td>
<td>National Health Expenditure Accounts Official estimates of healthcare spending in U.S. Includes care, admin, research and infrastructure.</td>
<td>Multiple Sources</td>
<td>HHS</td>
<td>Healthcare: 640; Non-Healthcare: 14</td>
<td>Services, Payers, Sponsors</td>
<td>Macro Data</td>
<td>Ready-to-use</td>
</tr>
</tbody>
</table>

**Notes:**

¹ 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.)

² Microdata consist of information at the unit level and provide the highest degree of analytic flexibility. Microdata 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.

---

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

1. 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.)

2. Microdata consist of information at the unit level and provide the highest degree of analytic flexibility. Microdata 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.

© 2020 Stern Consulting LLC
## Case Study

### Selected Datasets by Key Dimensions

### Dimensions of Data

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Notes</th>
<th>Mechanism</th>
<th>Sponsor (Data Collector)</th>
<th>Content Variable Counts:†</th>
<th>Primary Unit of Analysis</th>
<th>Granularity#</th>
<th>Constraints &amp; Use*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ACs</strong> American Community Survey</td>
<td>Extensive geographic and demographic drill downs on disability and health insurance.</td>
<td>Survey</td>
<td>Census Bureau</td>
<td>Healthcare 17, 199, Non-Healthcare 199</td>
<td>Person</td>
<td>Micro</td>
<td>Requires Skills**</td>
</tr>
<tr>
<td><strong>MEPS-HC</strong> Medical Expenditure Panel Survey, Household Component</td>
<td>Person-level health expenditures with longitudinal capabilities.</td>
<td>Survey</td>
<td>AHRQ (Westat)</td>
<td>Healthcare 1,252, Non-Healthcare 330</td>
<td>Person</td>
<td>Micro</td>
<td>Requires Skills**</td>
</tr>
<tr>
<td><strong>NHANES</strong> National Health and Nutrition Examination Survey</td>
<td>Survey combines interviews and physical examination, including lab tests.</td>
<td>Survey</td>
<td>NCHS/CDC</td>
<td>Healthcare 1,733, Non-Healthcare 147</td>
<td>Person</td>
<td>Micro</td>
<td>Requires Skills**</td>
</tr>
<tr>
<td><strong>NHIS</strong> National Health Interview Survey</td>
<td>Principal source of information on health of U.S. population. Robust demographic, socioeconomic data.</td>
<td>Survey</td>
<td>CDC (NCHS)</td>
<td>Healthcare 1,388, Non-Healthcare 212</td>
<td>Person</td>
<td>Micro</td>
<td>Requires Skills**</td>
</tr>
<tr>
<td><strong>MEPS-IC</strong> Medical Expenditure Panel Survey, Insurance Component</td>
<td>Factors contributing to use of employer sponsored insurance. Premiums and cost sharing.</td>
<td>Survey</td>
<td>AHRQ (Census Bureau)</td>
<td>Healthcare 153, Non-Healthcare 22</td>
<td>Employers, Employees</td>
<td>Macrodata</td>
<td>Ready-to-use</td>
</tr>
<tr>
<td><strong>Medicaid</strong> (various program data)</td>
<td>Actual enrollment data. Breakouts of new eligibility categories created by ACA.</td>
<td>Admin</td>
<td>CMS</td>
<td>Healthcare 30, Non-Healthcare 8</td>
<td>Person</td>
<td>Macrodata</td>
<td>Ready-to-use</td>
</tr>
<tr>
<td><strong>NHEA</strong> National Health Expenditure Accounts</td>
<td>Official estimates of healthcare spending in U.S. Includes care, admin, research and infrastructure.</td>
<td>Multiple Sources</td>
<td>HHS/Department of Health and Human Services</td>
<td>Healthcare 640, Non-Healthcare 14</td>
<td>Services, Payers, Sponsors</td>
<td>Macrodata</td>
<td>Ready-to-use</td>
</tr>
</tbody>
</table>

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

† “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.)

‡ Microdata consist of information at the unit level and provide the highest degree of analytic flexibility. Microdata 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.

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.
## Categories of Variables

### Healthiness
- Ability to Pay for Care
- Ability to Get Care
- Charges
- Encounters
- Expenditures
- Expenditures by Sponsor
- Behavior/Attitude
- Body Composition
- Child-Specific Problems
- Clinical Results
- Condition
- Days Lost Due to Illness
- Diet
- Functional Limitation
- Status

### Economic
- Income
- Other Benefits
- Taxes
- Food Security
- Job Characteristics
- Labor Force
- Economic Indicators/Population

### Demographic
- Age
- Race/Ethnicity
- Sex

### Geographic
- Specified Areas

### Housing
- Financial
- Physical

### Social
- Family/Household Composition
- Marital Status
- Military Status
- Neighborhood Characteristics
- Child Care
- Education
- Heritage
- Migration
- Internet/Computer Use

### Access

### Non-Healthcare

The "non-healthcare" variables are likewise more comprehensive than those typically provided in administrative data.

### Additional detail on the “Content” dimension

<table>
<thead>
<tr>
<th>Variables</th>
<th>ACS</th>
<th>ASEC</th>
<th>MEPS-HC</th>
<th>NHANES</th>
<th>NHIS</th>
<th>MEPS-IC</th>
<th>M’caid</th>
<th>NHEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>1,053</td>
<td>10</td>
<td>157</td>
<td>546</td>
<td>17</td>
<td>203</td>
<td>87</td>
<td>22</td>
</tr>
<tr>
<td>Cost</td>
<td>19</td>
<td>19</td>
<td>21</td>
<td>69</td>
<td>17</td>
<td>262</td>
<td>226</td>
<td>586</td>
</tr>
<tr>
<td>Healthiness</td>
<td>290</td>
<td>4</td>
<td>224</td>
<td>62</td>
<td>11</td>
<td>8</td>
<td>1</td>
<td>79</td>
</tr>
<tr>
<td>Non-Healthcare</td>
<td>303</td>
<td>13</td>
<td>211</td>
<td>25</td>
<td>5</td>
<td>51</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Economic</td>
<td>10</td>
<td>20</td>
<td>6</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Demographic</td>
<td>20</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Geographic</td>
<td>47</td>
<td>8*</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Housing</td>
<td>32</td>
<td>27</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>306</td>
<td>33</td>
<td>78</td>
<td>65</td>
<td>88</td>
<td>42</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>6,806</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ACS** contains extensive additional geographic detail available only in the macrodata renderings.

**NHANES** is the only dataset to include clinical results. **NHIS** and **MEPS-HC** are also strong on healthiness.

**MEPS-HC** and **NHIS** are strong on "getting" care.

**ASEC** and **MEPS-HC** are strong on economic issues, particularly income, labor force participation and job characteristics.
Sample ACS Data

Percent of Population with a Disability
Selected Geographies
2010-2018

ACS is administered to over 3 million persons. The large sample size permits finer levels of granularity than other data sources.

There is wide variation in disability rates across New York counties as compared to the state and national averages.

*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.

Source: ACS One-Year Estimates, Table S1810. Chart created by Stern Consulting using data from U.S. Census Bureau’s data dissemination platform, data.census.gov/cedsci/
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.*

*Days lost across the total working population steadily decreased.*

*Days lost due to illness varies widely across various combinations of ethnicity, race and sex.*

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
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
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

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

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.
### Dataset Directories

<table>
<thead>
<tr>
<th>List</th>
<th>Sponsor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data, Tools and Statistics</td>
<td>National Institutes of Health, National Library of Medicine</td>
</tr>
<tr>
<td></td>
<td>National Information Center on Health Services Research and Health Care Technology (NICHSR)</td>
</tr>
<tr>
<td>Directory of Health and Human Services</td>
<td>U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation</td>
</tr>
<tr>
<td>Data Resources</td>
<td><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></td>
</tr>
<tr>
<td>Guide to HHS Surveys and Data Resources</td>
<td>U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation</td>
</tr>
<tr>
<td>HealthData.gov</td>
<td>U.S. Department of Health and Human Services, Office of the Chief Technology Officer</td>
</tr>
<tr>
<td></td>
<td><a href="https://healthdata.gov/">https://healthdata.gov/</a></td>
</tr>
<tr>
<td>HealthyPeople.gov Data Sources</td>
<td>U.S. Department of Health and Human Services</td>
</tr>
<tr>
<td></td>
<td><a href="https://www.healthypeople.gov/2020/data-search/Data-Sources">https://www.healthypeople.gov/2020/data-search/Data-Sources</a></td>
</tr>
<tr>
<td>Data Sources</td>
<td>Deloitte, DATA USA</td>
</tr>
<tr>
<td></td>
<td><a href="https://datausa.io/about/datasets/">https://datausa.io/about/datasets/</a></td>
</tr>
</tbody>
</table>

Additional Resources, 2 of 2

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

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)
- USA.gov (www.usa.gov/federal-agencies).
Sample NHEA Data

Expenditures by Type of Sponsor (Percent Distribution)
1987-2018

**State & Local Government**

15 14.9 14.9 15.1 15.4 15.3 15.6 16.3 16.2 16 16.3 16.4 16.5 16.7 17 16.8 16.9 17.2 17.2 17.3 17.1 16.2 16.2 17.1 17.5 17.7 17.3 17.2 17 16.9 16.5

**Federal Government**

The relative contribution by the federal government, primarily via payroll taxes for Medicare (both as collector and payer), general revenues, and premiums for employee health insurance, has steadily increased.

16.5 15.8 16.2 17.2 18.4 20 20.9 21 21.1 21.4 21.1 19.6 19.1 19 20.6 21.3 21.9 22.4 22.3 23 23 24.3 27.3 28.2 27.2 26.2 26.2 27.6 28.4 28.1 28.3

**Households**

The relative contribution of households, primarily via payroll taxes for Medicare, premiums for private insurance and out-of-pocket expenditures, has decreased.

38.1 38.2 37.3 36.2 35.2 33.8 32.9 32.3 32.1 31.4 31.8 32.5 32.4 32.4 31.4 31.2 30.9 30.7 30.6 30.4 30.2 29.2 28.9 28.9 29 28.5 28.4 28.4 28.5 28.4

**Private Business**

22.4 22.8 23.5 23.6 23.3 22.3 22.9 22.8 23.3 22.8 23.2 23.8 24.5 24.3 23.8 23.7 23.3 23.2 22.4 22 21.3 20.5 19.9 20.2 20.2 20 19.8 19.5 19.5 19.6 19.9

**Other Private Revenue**

8 8.4 8 7.8 7.7 7.8 7.7 7.6 7.7 7.8 8.1 8.4 8.3 7.6 6.9 6.7 6.7 6.8 7 7.4 7.2 6.7 6.8 6.7 7.1 7.2 6.8 6.6 6.7 7 6.9

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

© 2020 Stern Consulting LLC
Sample MEPS-IC Data

Factors That Determine Employer Sponsored Insurance (ESI) Enrollment
Private Sector Employees, United States, 1996-2018

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.

Source: MEPS-IC. Chart created by Stern Consulting based on data from “MEPSnet Query Tools,” meps.ahrq.gov/mepsweb/data_stats/meps_query.jsp

© 2020 Stern Consulting LLC
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.

Persons working less than full-time or not year-round, are more likely to be uninsured...

...than those working full-time year-round

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>All workers</td>
<td>11.6%</td>
</tr>
<tr>
<td>Did not work at least one week</td>
<td>13.7%</td>
</tr>
<tr>
<td>Less than full-time year-round</td>
<td>15.6%</td>
</tr>
<tr>
<td>Full-time year-round</td>
<td>10%</td>
</tr>
</tbody>
</table>

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
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.

The Medicaid Budget and Expenditure System (MBES) provides monthly figures for total enrollment. It provides separate figures for those persons that fall within the new category of eligibility called the “VIII Group” created by the Affordable Care Act. The VIII group is broken out by those that meet the new federal guidelines but were already covered under their state-specific guidelines and those gaining coverage through the ACA. The former group is referred to as “Not Newly Eligible” and the latter as “Newly Eligible.” The “Original Medicaid” figures reported herein are calculated as the difference between the “totals” provided by MBES and the figures for the VIII Groups.

Is this drop real? Administrative data may typically be more precise than other sources such as surveys, which are based on samples. However, administrative data may have its own limitations. Note the “drop” in enrollment for 2014. This decrease is evidently due to state reporting issues. The data for some states for Q3 2014 (as well as for other periods such as late 2017) are flagged as “Awaiting state reporting/enrollment reasonableness review in progress.” Indeed as recently as this year, Medicaid continues to update the 2014 numbers, suggesting the “drop” is a reporting issue and is not real.
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.

**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

**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

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.

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