Insurance Coverage Types Before 2014 Health Reform

1. Private
   a. Employer-sponsored insurance (ESI)
   b. Non-group purchased on the individual market

2. Public
   a. Medicaid (for low income)
   b. Medicare (for 65+)
   c. Military
Post-Health Reform: Marketplace in the Mix

1. Private
   a. Employer-sponsored insurance (ESI)
   b. Non-group purchased on the individual market
      1. Outside the marketplace
      2. On the marketplace (aka ObamaCare)

2. Public
   a. Medicaid (for low income)
   b. Medicare (for 65+)
   c. Military*
Measuring Health Insurance Got More Complicated Post-ACA

1. Private
   a. Employer-sponsored insurance (ESI)
   b. Non-group purchased on the individual market
      1. Outside the marketplace
      2. On the marketplace (aka ObamaCare)

2. Public
   a. Medicaid (for low income)
   b. Medicare (for 65+)
   c. Military
Ambiguity Between Marketplace and Medicaid

1. The term ‘marketplace’ has a dual meaning:
   - Portal for shopping for coverage (e.g.: healthcare.gov)
   - The coverage itself (‘marketplace’ aka ‘ObamaCare’)

2. Getting coverage on the portal does not define coverage type:
   - Broad spectrum of coverage is available, from fully-subsidized Medicaid to unsubsidized private
   - Brokers can sell Marketplace coverage; not required to go thru portal

3. Private/public blurry line:
   - Some marketplace coverage has $0 premium
   - Some Medicaid requires enrollees to pay part of premium
Ambiguity:
- No one question is sufficient to categorize coverage
- Some questions have more than one correct answer

Separating subsidized Marketplace from Medicaid is especially difficult “on paper”

Need to use multiple data points and create an algorithm to classify coverage type

RESEARCH QUESTIONS:
- What is the ‘ideal’ algorithm?
- How should it be developed?
- How should it be evaluated?
# Research Questions

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Development of Algorithms</th>
<th>Evaluation of Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptual</td>
<td>• Assumptions based on predominant characteristics of health coverage</td>
<td>1. Under-reporting: among those who have Cov Type X according to records, how many report it?</td>
</tr>
<tr>
<td></td>
<td>• Ignores atypical scenarios (e.g., Medicaid that charges a premium)</td>
<td>2. Over-reporting: among those who report Cov Type X, how many are validated in records to have it?</td>
</tr>
<tr>
<td>Machine Learning</td>
<td>• Develop algorithm using ‘supervised machine learning’ approach</td>
<td>3. Prevalence: how does prevalence of Cov Type X in records compare to survey estimate?</td>
</tr>
<tr>
<td></td>
<td>• Explore:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Additional data points</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Combinations of data points</td>
<td></td>
</tr>
</tbody>
</table>
Conceptual Algorithm

Any coverage? → Yes → General source?

Job

Govt/State

Direct Purchase

Related to military service?

What type of Govt/State cov?

No

Yes

Mil

Medicare

Public, Other

From the marketplace?

Yes

No

Is there a monthly premium?

No

Yes

ESI (job-based)

Military

Medicare

Medicaid

Marketplace

Non-Group
Any coverage? → Yes → General source?

Job
Govt/State

Related to military service?

No
Yes

What type of Govt/State cov?

Public, Other
Direct Purchase

What do you call the program?

Public, Other
Marketplace

Is there a monthly premium?

Yes
No

Is the premium subsidized?

ESI (job-based)
Military
Medicare

Public, Marketplace or Non-group (TBD by Algorithm)
Machine Learning Approach

- Given that we know coverage type, can we determine what survey reporting patterns best match actual enrollees of each type of coverage?

- Steps:
  - Identify key variables or “features” in survey (n=5)
  - Create all permutations of answers to key variables (n=157)
  - Data reduction:
    - Collapse response categories
    - Collapse permutations where enrollee distribution is similar
  - Pattern recognition based on actual enrollment
  - Classification of coverage type
Reverse Record Check Study
Data Collection Methods

- Sample: phone numbers of enrollees from US-based private health insurance company records; random sample drawn from multiple strata:
  - Employer-sponsored insurance (ESI)
  - Non-group (direct purchase/outside marketplace)
  - Marketplace (unsubsidized and subsidized)
  - Medicaid
- 15-minute phone survey conducted in Spring, 2015
- Content: abbreviated CPS (demos, labor force, health insurance)
- Data collected on all household members
- Response rate = 22%
- Health plan enrollment file sent post-data collection
- Month-level data from enrollment file matches survey reports
- Records matched to survey at person-level; final matched file n ~ 2,000
- Weighted data to health plan population totals
Exploratory Algorithms

- Explored multiple algorithms:
  - GovSkew: classify all ambiguous cases as public
  - MktSkew: classify all ambiguous cases as market
  - Hybrid: split the difference

- Ran all three accuracy metrics:
  - Under-Reporting
  - Over-Reporting
  - Prevalence (records versus survey estimate)

- Compared empirical results, which take into account net effects of all three accuracy metrics AND prevalence of coverage type simultaneously
Results
Under-Reporting

[Bar chart showing under-reporting in different categories with Machine Learning and Conceptual categories compared.]

Categories include:
- Marketplace
- Non-group
- Non-group/Market
- ESI
- Private
- Public

Legend:
- Red: Machine Learning
- Blue: Conceptual
Over-Reporting

Machine Learning

Conceptual
Point Estimates versus Records Prevalence

- Records
- Machine Learning
- Conceptual

Categories: Marketplace, Non-group, Non-group/Market, ESI, Private, Public
Summary of Results:
Under- and Over-Reporting

- Conceptual prioritizes Marketplace over Public

Benefits:
- Less under-reporting of Marketplace (by 15 ppt)
- Less under-reporting of non-group/mkt (by 3 ppt)
- [BUT no difference in aggregated private because marketplace prevalence is so low]

Costs:
- Higher under-reporting of public by almost 5 ppt
- Higher over-reporting of marketplace by 7.5 ppt
- Higher over-reporting of private by 2 ppt
Summary of Prevalence Results

- Both algorithms over-estimate private and under-estimate public
- For all coverage type categories where algorithms are different:
  - Marketplace
  - Non-group/marketplace
  - Private
  - Public
  ➔ the machine learning estimate is closer to the records prevalence than the conceptual algorithm
Thank you!

Contact Information:
Joanne Pascale
Joanne.Pascale@census.gov
## Data Reduction Example

<table>
<thead>
<tr>
<th>Perm</th>
<th>Freq</th>
<th>Type of Govt/State Cov</th>
<th>What do you call program</th>
<th>Portal</th>
<th>Prem</th>
<th>Subs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>70</td>
<td>Public</td>
<td>Medicaid</td>
<td>Yes</td>
<td>No</td>
<td>n/a</td>
</tr>
<tr>
<td>B</td>
<td>22</td>
<td>Public</td>
<td>Medicaid</td>
<td>No</td>
<td>No</td>
<td>n/a</td>
</tr>
<tr>
<td>C</td>
<td>13</td>
<td>Public</td>
<td>Medicaid</td>
<td>D/R</td>
<td>No</td>
<td>n/a</td>
</tr>
</tbody>
</table>
### Data Reduction Example

<table>
<thead>
<tr>
<th>Perm</th>
<th>Freq</th>
<th>Type of Govt/State Cov</th>
<th>What do you call Program</th>
<th>Portal</th>
<th>Prem</th>
<th>Subs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>70</td>
<td>Public</td>
<td>Medicaid</td>
<td>Yes</td>
<td>No</td>
<td>n/a</td>
</tr>
<tr>
<td>B</td>
<td>22</td>
<td>Public</td>
<td>Medicaid</td>
<td>No</td>
<td>No</td>
<td>n/a</td>
</tr>
<tr>
<td>C</td>
<td>13</td>
<td>Public</td>
<td>Medicaid</td>
<td>D/R</td>
<td>No</td>
<td>n/a</td>
</tr>
</tbody>
</table>

**Cov Type According to Records**

- Pub
- ESI
- NonG
- Mkt
### Data Reduction Example

<table>
<thead>
<tr>
<th>Perm</th>
<th>Freq</th>
<th>Type of Govt/State Cov</th>
<th>What do you call program</th>
<th>Portal</th>
<th>Prem</th>
<th>Subs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>70</td>
<td>Public</td>
<td>Medicaid</td>
<td>Yes</td>
<td>No</td>
<td>n/a</td>
</tr>
<tr>
<td>B</td>
<td>22</td>
<td>Public</td>
<td>Medicaid</td>
<td>No</td>
<td>No</td>
<td>n/a</td>
</tr>
<tr>
<td>C</td>
<td>13</td>
<td>Public</td>
<td>Medicaid</td>
<td>D/R</td>
<td>No</td>
<td>n/a</td>
</tr>
</tbody>
</table>

#### Cov Type According to Records

- Pub
- ESI
- NonG
- Mkt

---

**Note:** The data reduction example illustrates the process of analyzing and summarizing data to reduce redundancy and improve efficiency. The table shows a summary of different program types and their characteristics, with a focus on Medicaid coverage and portal access. The pie charts indicate the distribution of Cov Type according to records.
Is there a monthly premium for this plan?

**READ IF NECESSARY:** A monthly premium is a fixed amount of money people pay each month to have health coverage. It does not include copays or other expenses such as prescription costs.

Yes

Is the cost of the premium subsidized based on family income?

**READ IF NECESSARY:** A monthly premium is a fixed amount of money people pay each month to have health coverage. It does not include copays or other expenses such as prescription costs.

**READ IF NECESSARY:** Subsidized health coverage is insurance with a reduced premium. Low and middle income families are eligible to receive tax credits that allow them to pay lower premiums for insurance bought through healthcare exchanges or marketplaces.