

Comparing a Conventional and Machine-Learning Algorithm to Categorize Health Insurance Coverage

Joanne Pascale, US Census Bureau

Angela Fertig, University of Minnesota

Kathleen Call, State Health Access Data Assistance Center

11th Health Survey Research Methods Conference

Williamsburg, VA

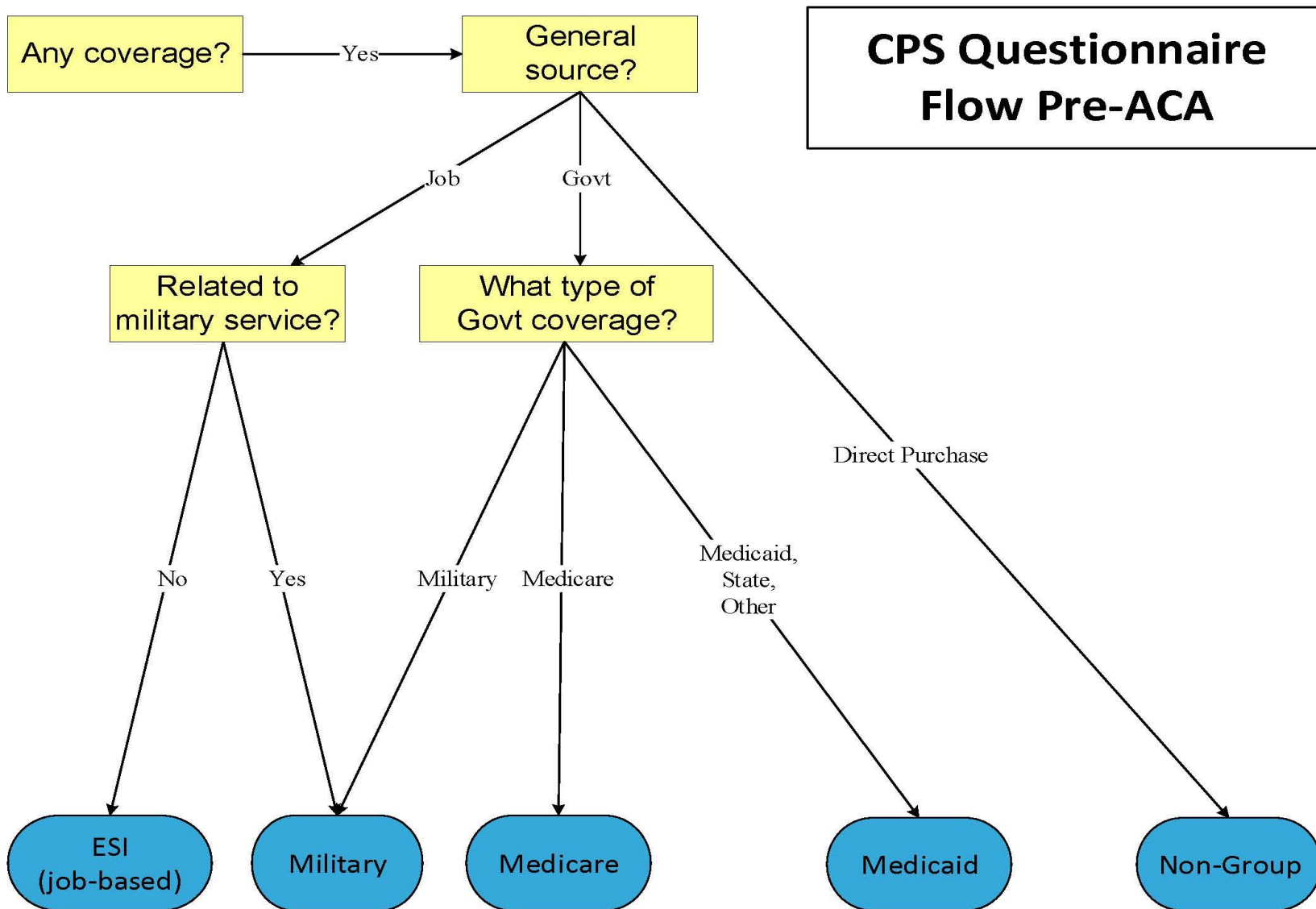
March 4-7, 2020

Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. This presentation has been reviewed for disclosure avoidance and approved under CBDRB-FY20-1555.



U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov





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

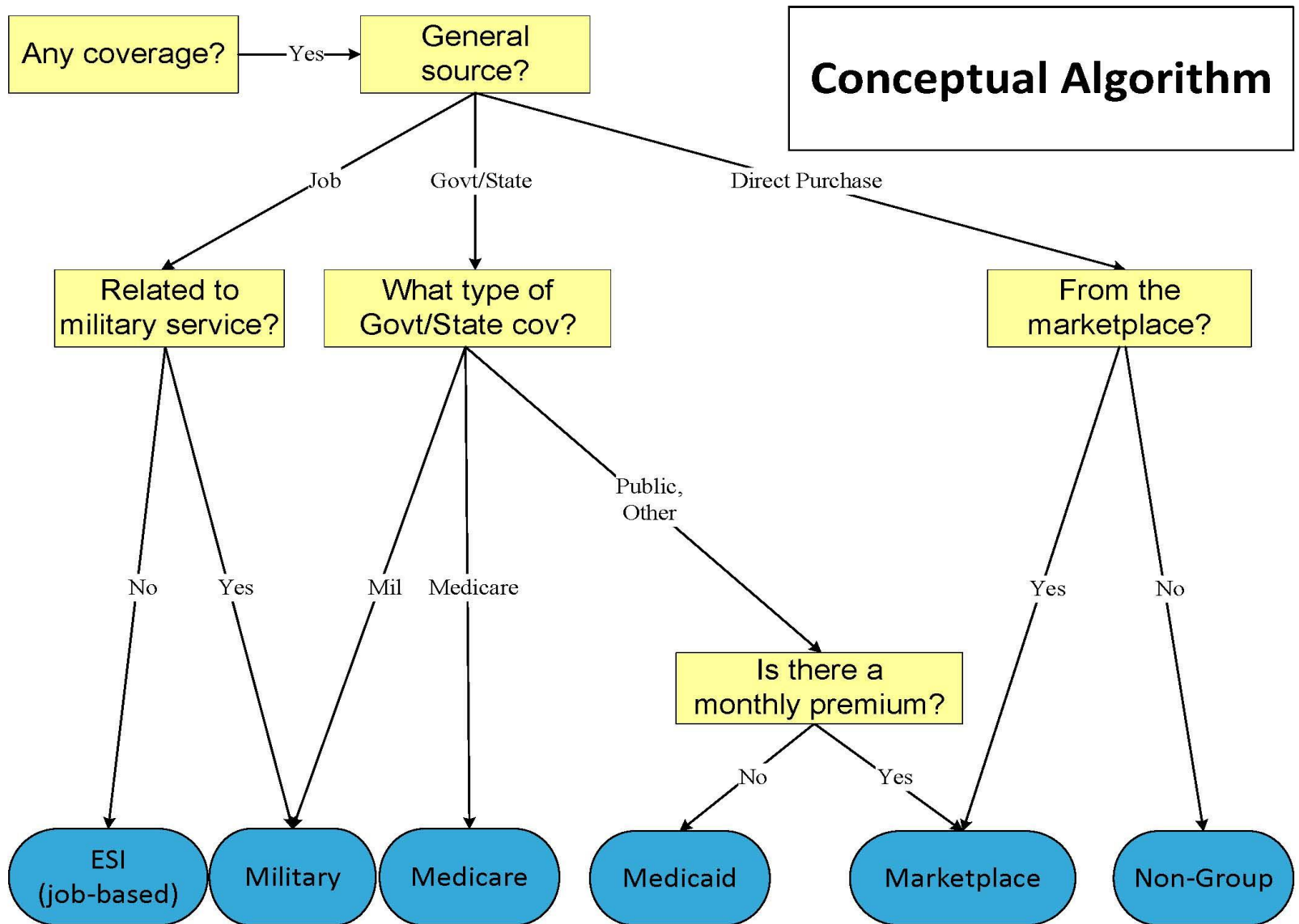
1. The term 'marketplace' has a dual meaning:
 - Portal for shopping for coverage (e.g.: healthcare.gov)
 - The coverage itself ('marketplace' aka 'ObamaCare')
2. Getting coverage on the portal does not define coverage type:
 - Broad spectrum of coverage is available, from fully-subsidized Medicaid to 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

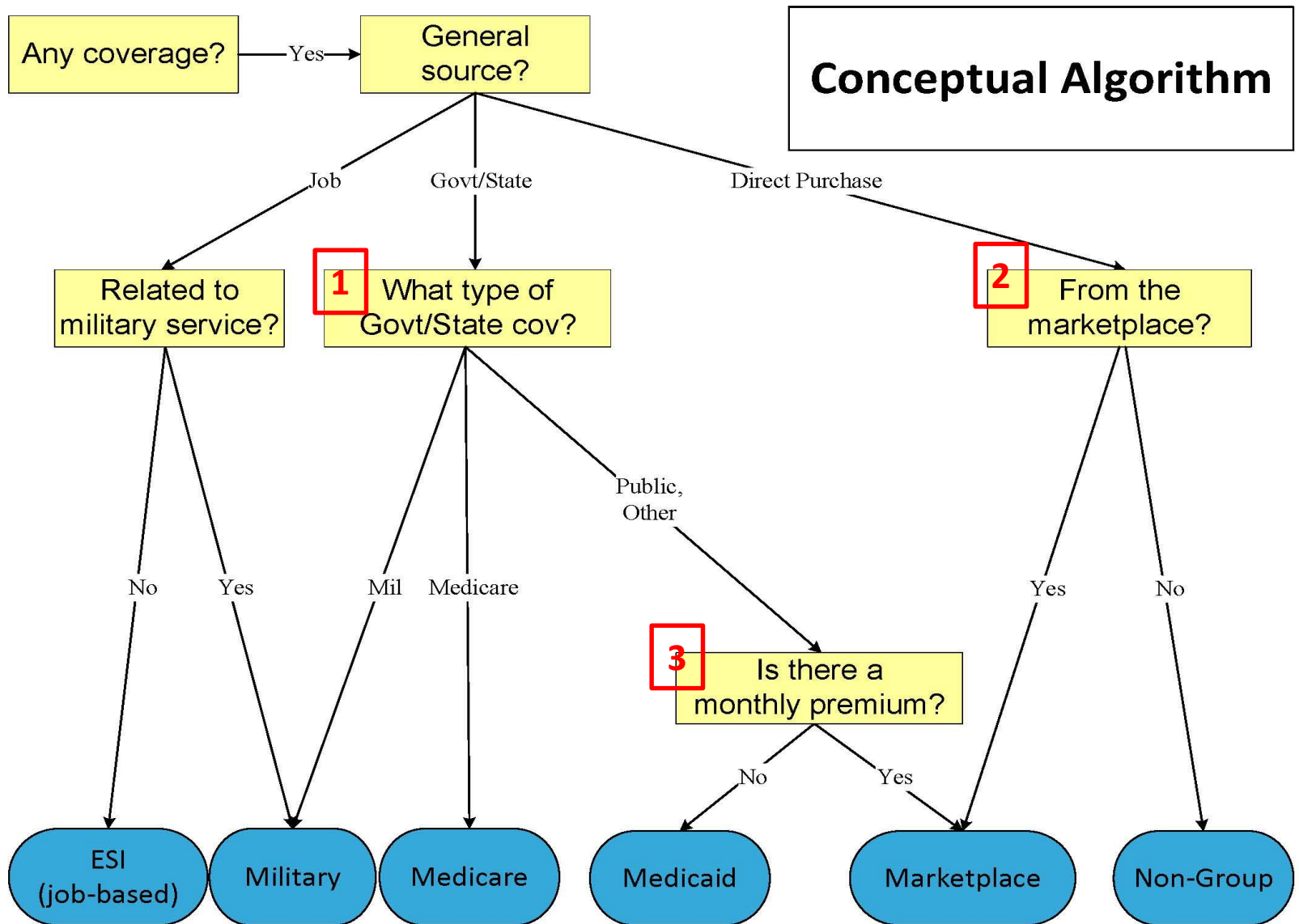
Upshot

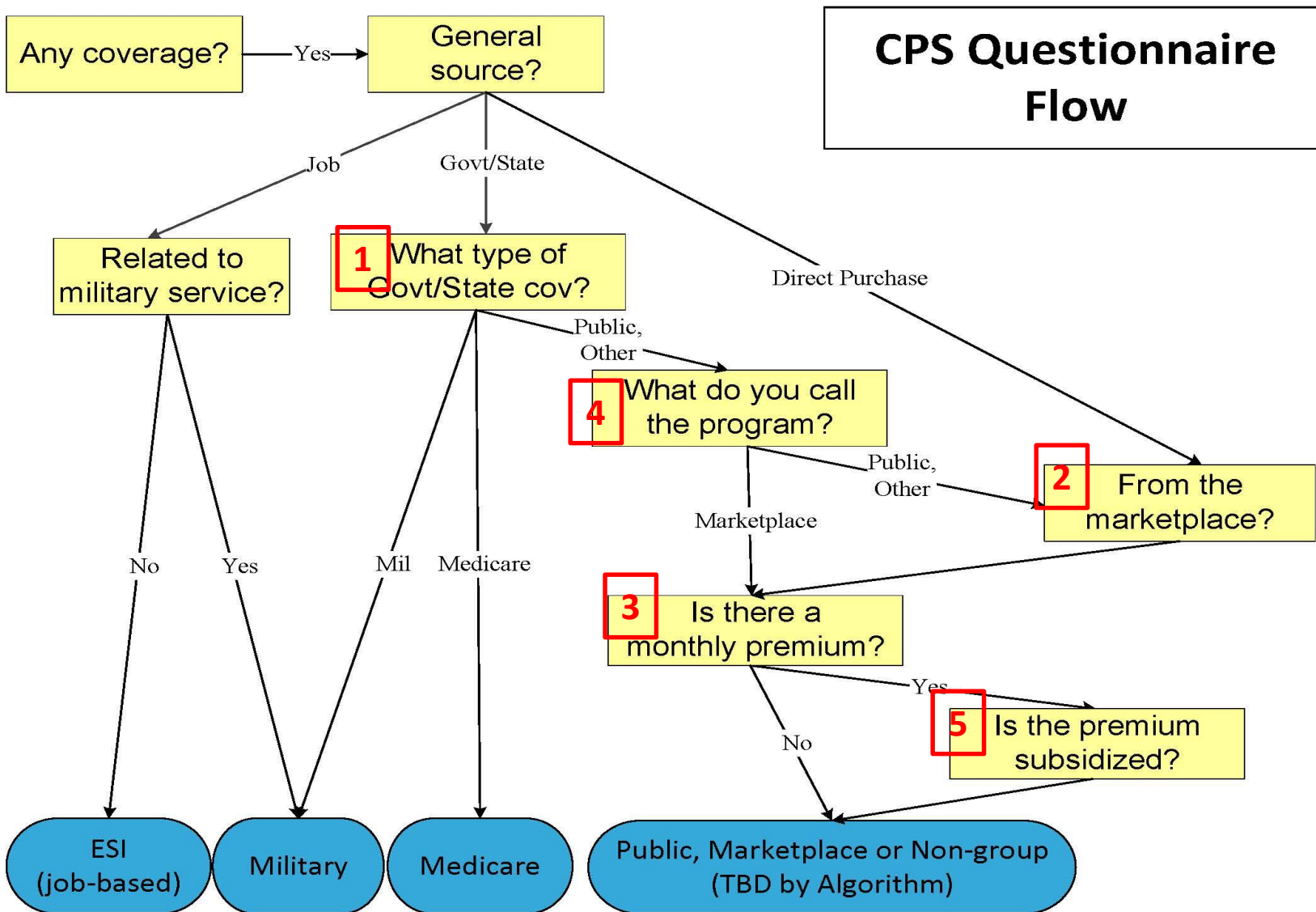
- Ambiguity:
 - No one question is sufficient to categorize coverage
 - Some questions have more than one correct answer
- Separating subsidized Marketplace from Medicaid is especially difficult “on paper”
- Need to use multiple data points and create an algorithm to classify coverage type
- RESEARCH QUESTIONS:
 - What is the ‘ideal’ algorithm?
 - How should it be developed?
 - How should it be evaluated?

Research Questions

Algorithms	Development of Algorithms	Evaluation of Algorithms
Conceptual	<ul style="list-style-type: none"> Assumptions based on predominant characteristics of health coverage Ignores atypical scenarios (e.g., Medicaid that charges a premium) 	<ol style="list-style-type: none"> Under-reporting: among those who have Cov Type X according to records, how many report it? Over-reporting: among those who report Cov Type X, how many are validated in records to have it? Prevalence: how does prevalence of Cov Type X in records compare to survey estimate?
Machine Learning	<ul style="list-style-type: none"> Develop algorithm using 'supervised machine learning' approach Explore: <ul style="list-style-type: none"> Additional data points Combinations of data points 	







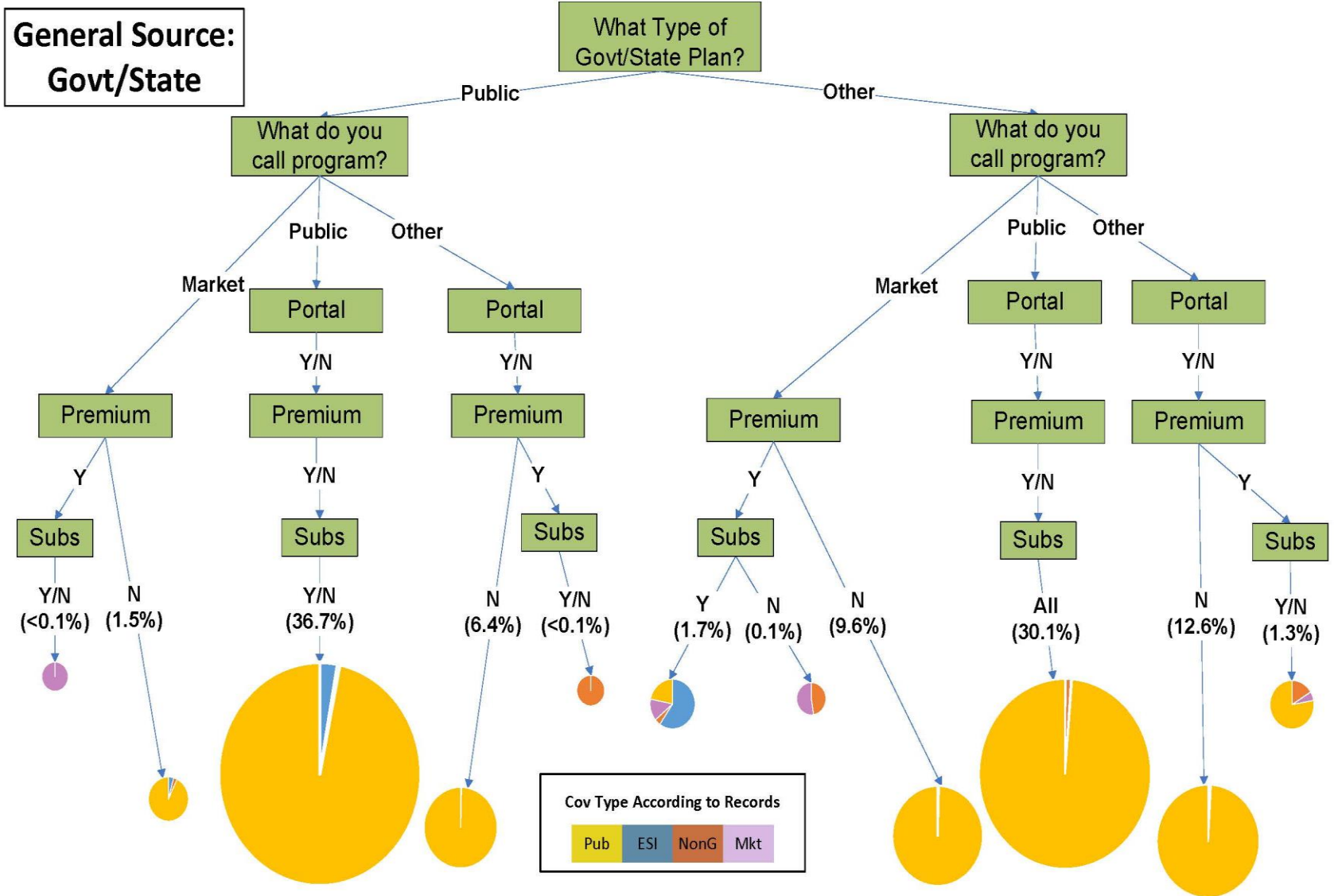
Machine Learning Approach

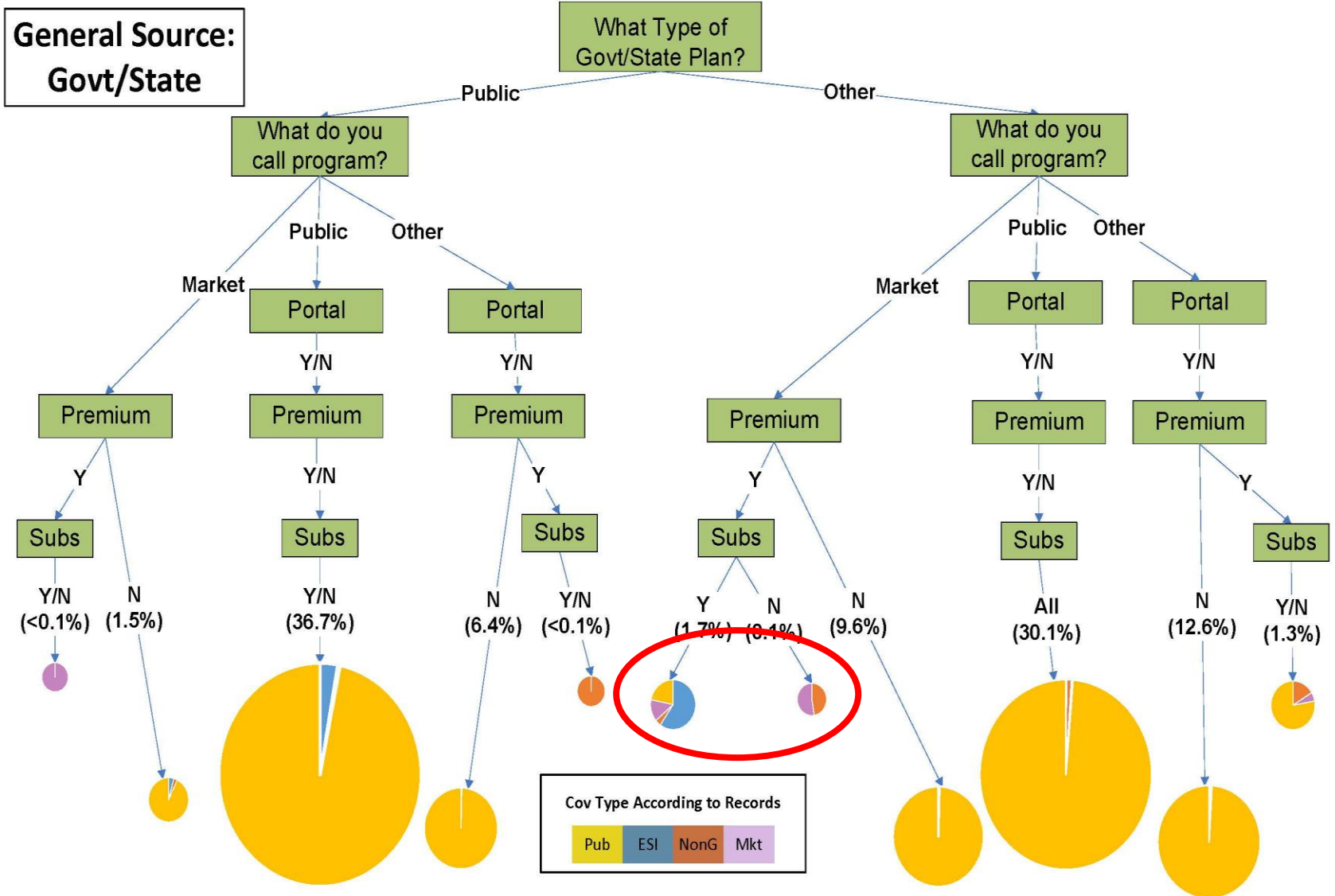
- Given that we know coverage type, can we determine what survey reporting patterns best match actual enrollees of each type of coverage?
- Steps:
 - Identify key variables or “features” in survey (n=5)
 - Create all permutations of answers to key variables (n=157)
 - Data reduction:
 - Collapse response categories
 - Collapse permutations where enrollee distribution is similar
 - Pattern recognition based on actual enrollment
 - Classification of coverage type

Reverse Record Check Study

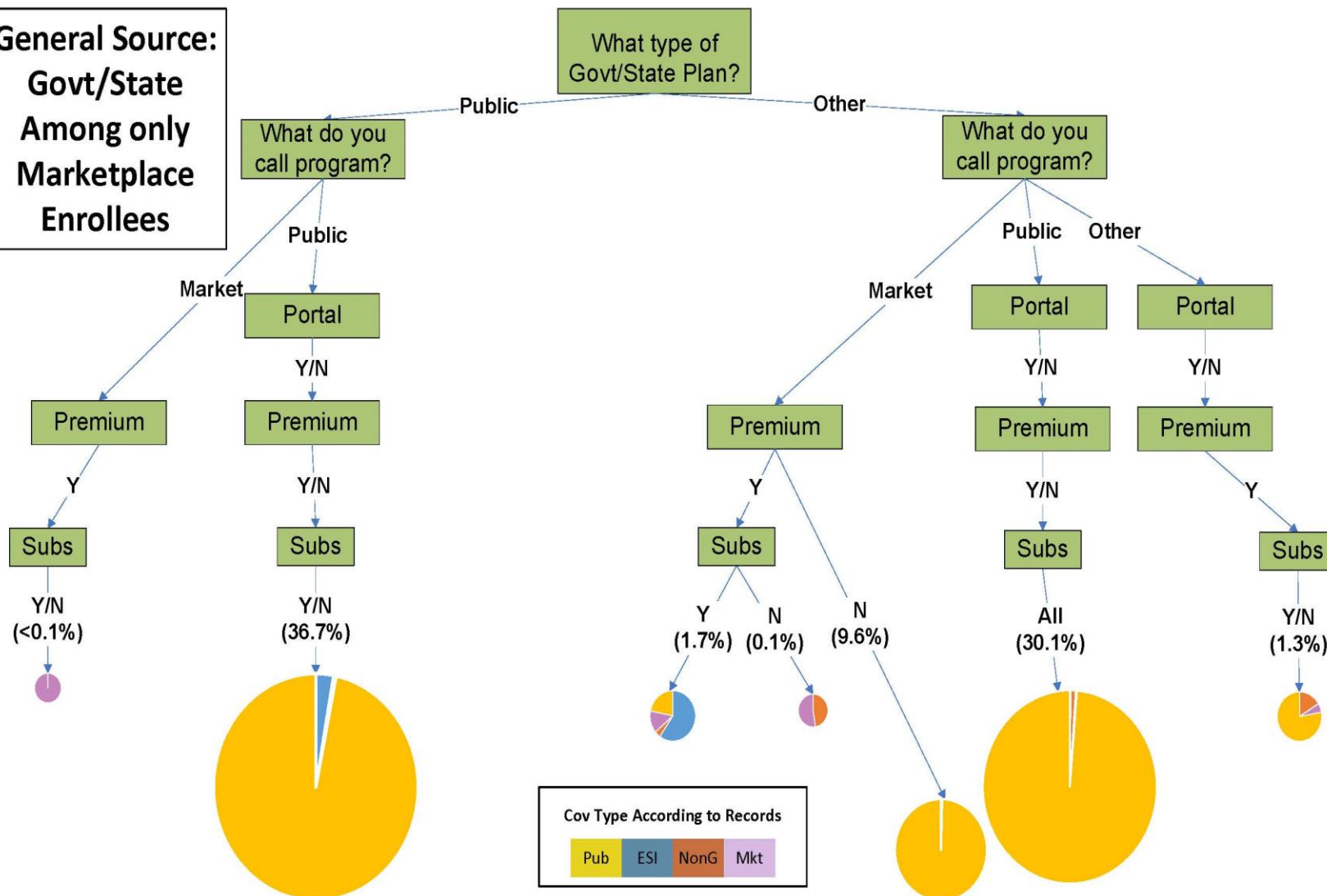
Data Collection Methods

- 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 $n \sim 2,000$
- Weighted data to health plan population totals

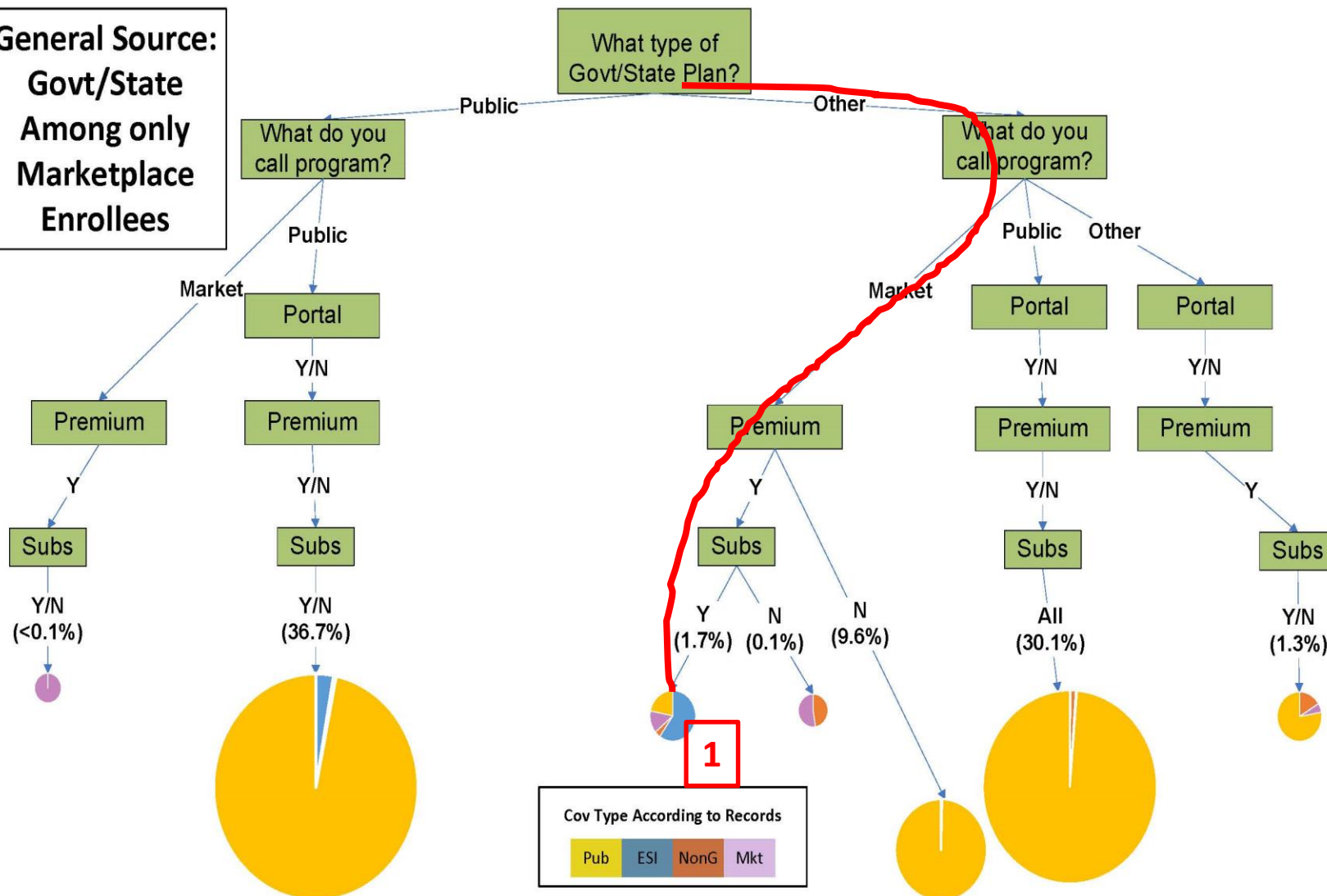


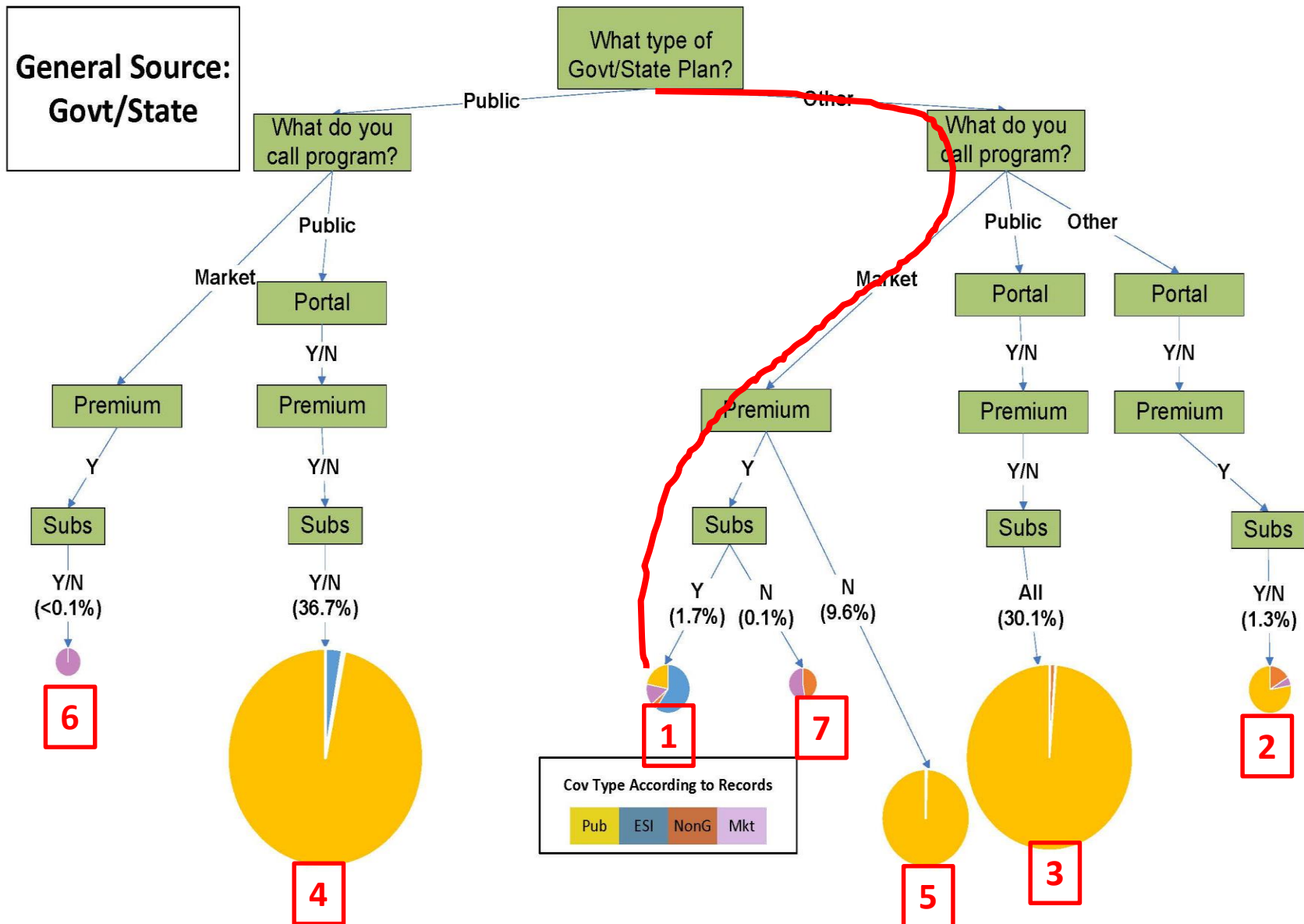


**General Source:
Govt/State
Among only
Marketplace
Enrollees**



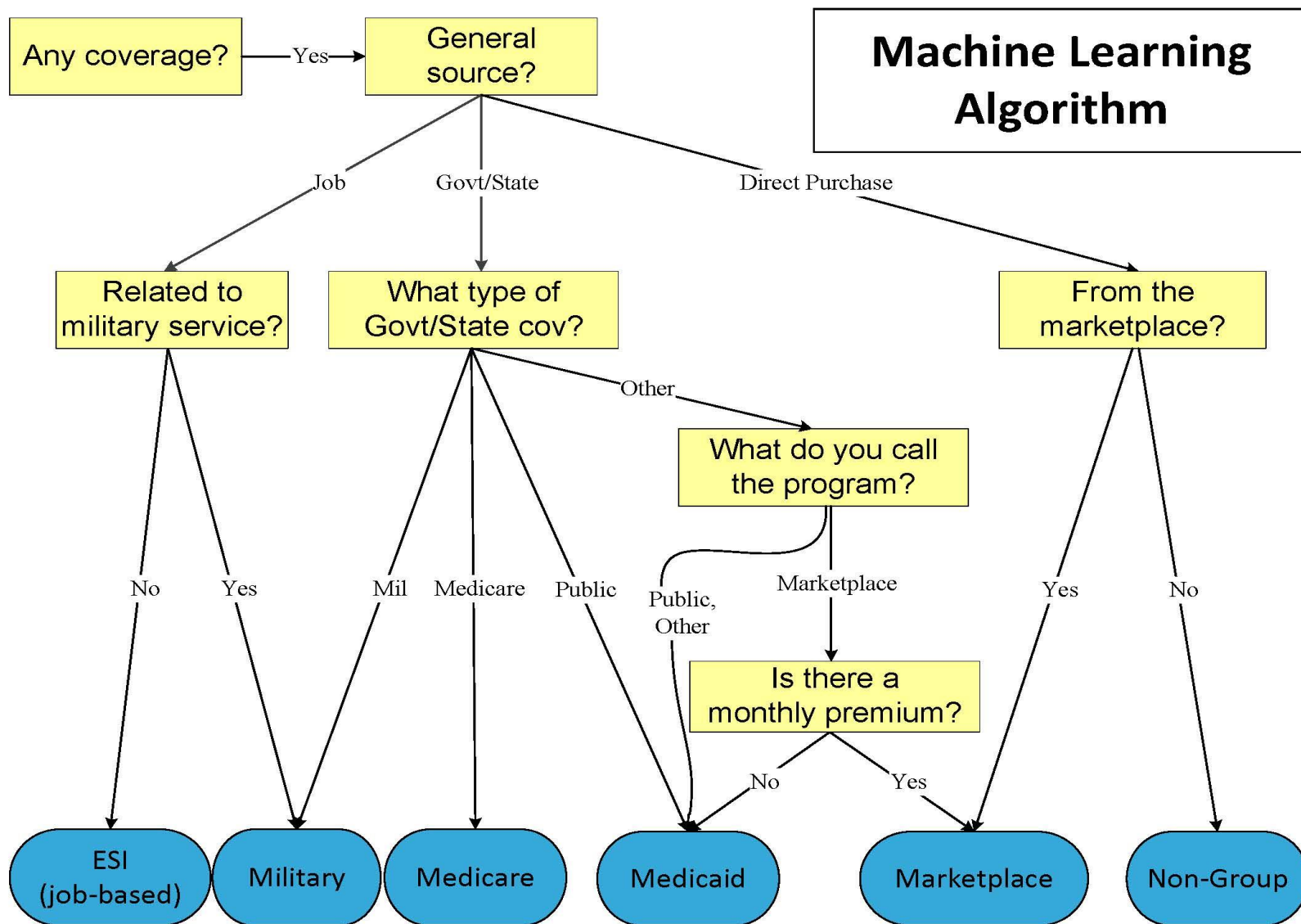
**General Source:
Govt/State
Among only
Marketplace
Enrollees**

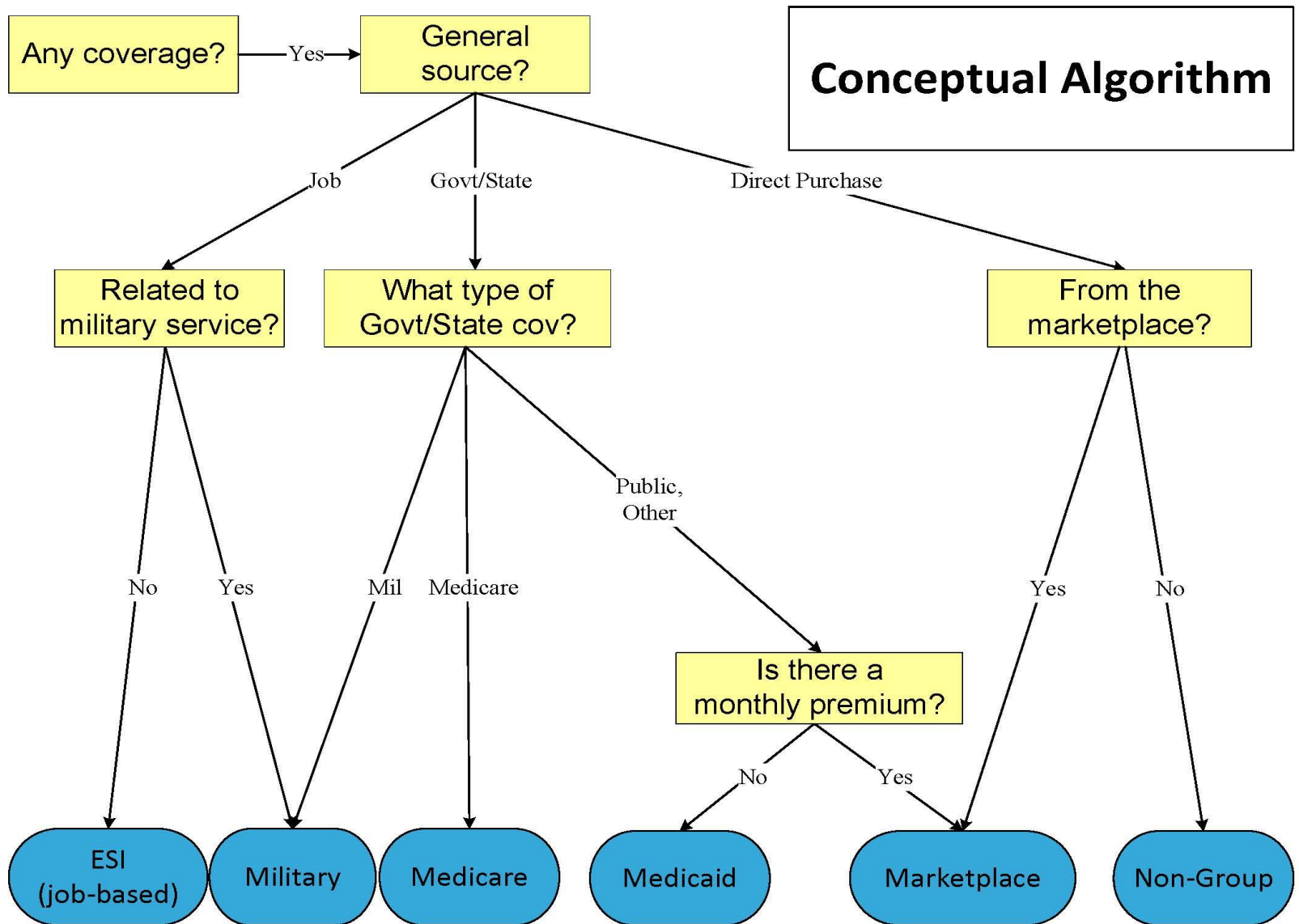




Exploratory Algorithms

- 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

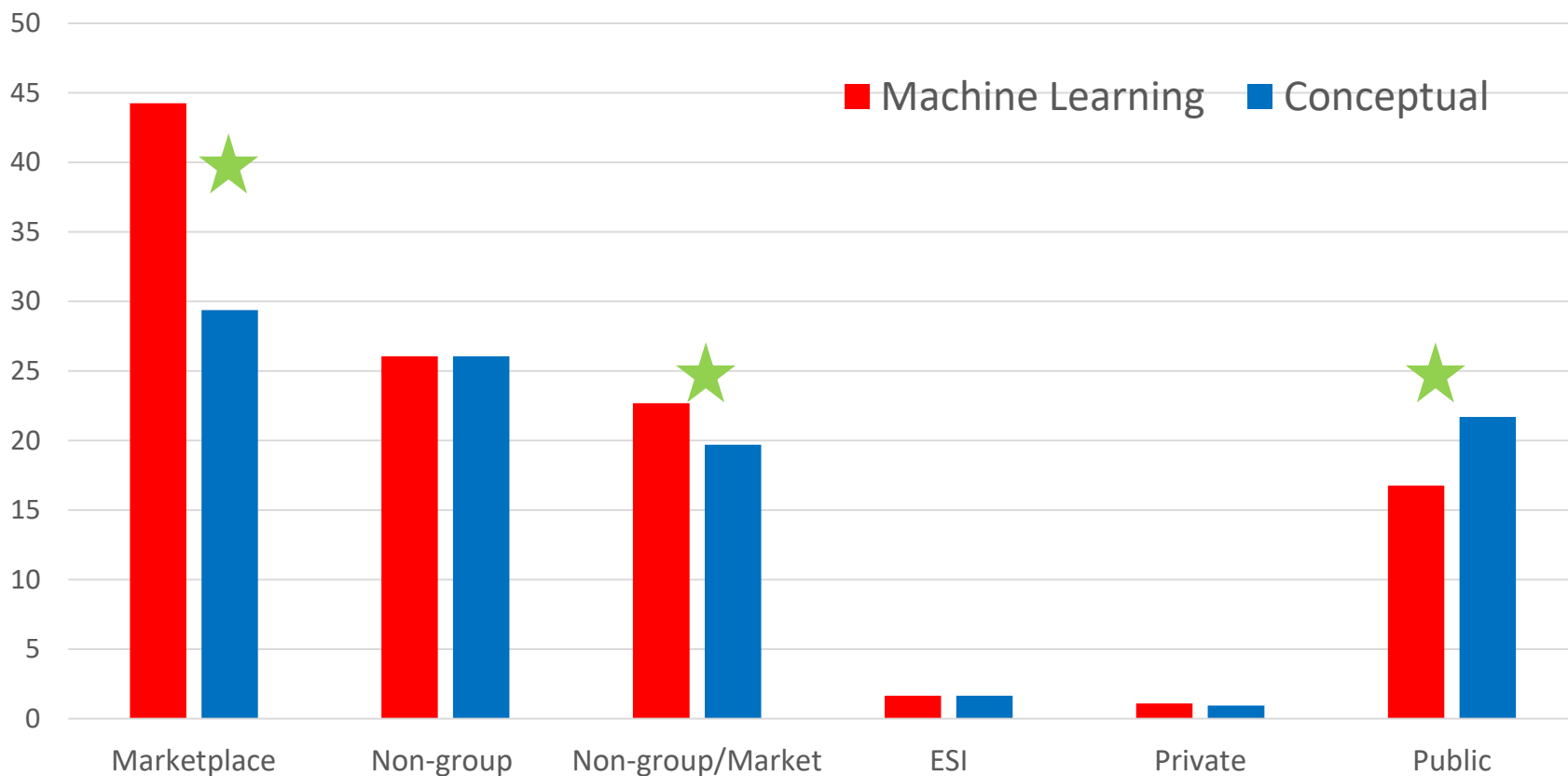




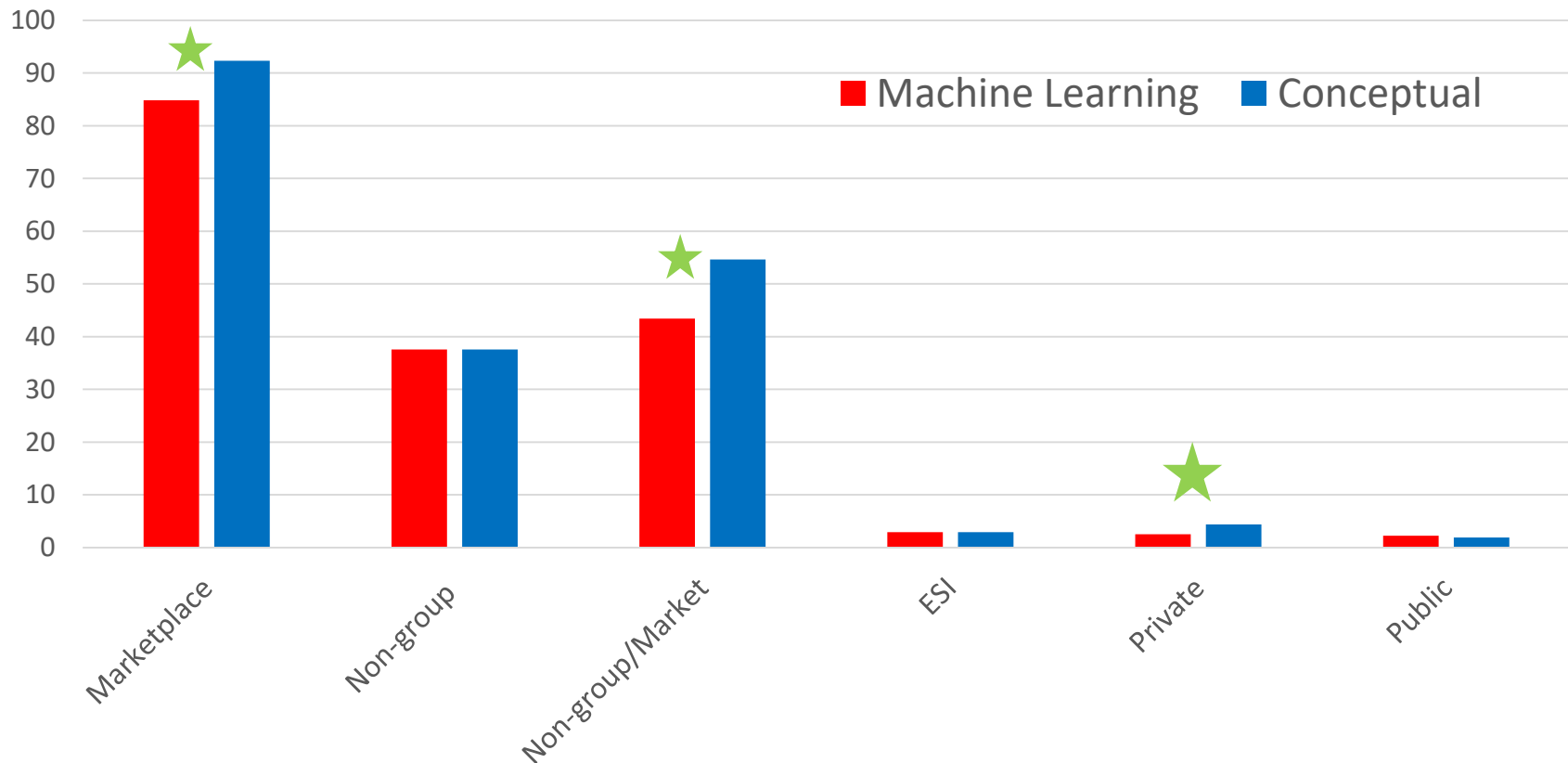
Results



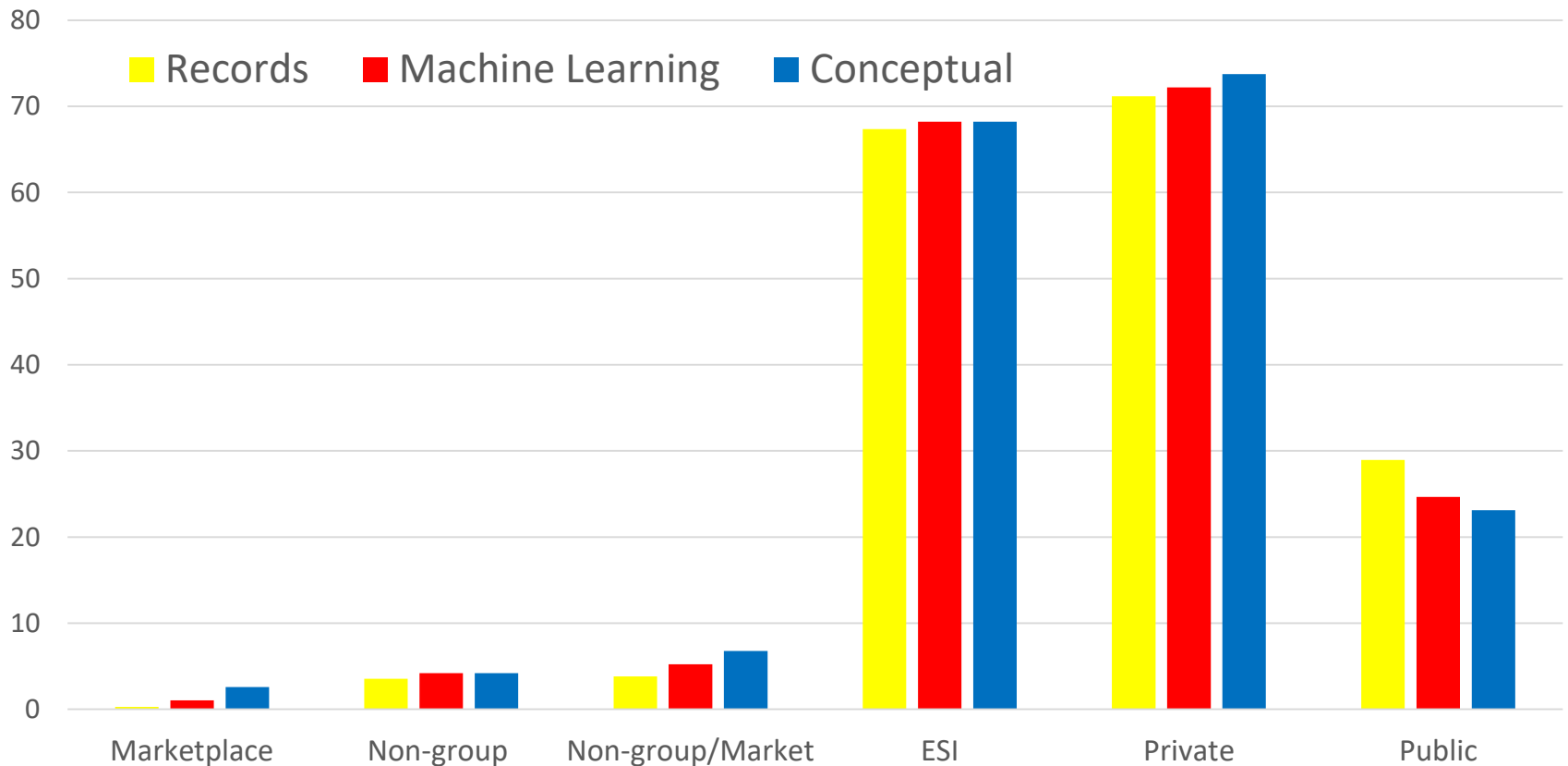
Under-Reporting



Over-Reporting



Point Estimates versus Records Prevalence



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

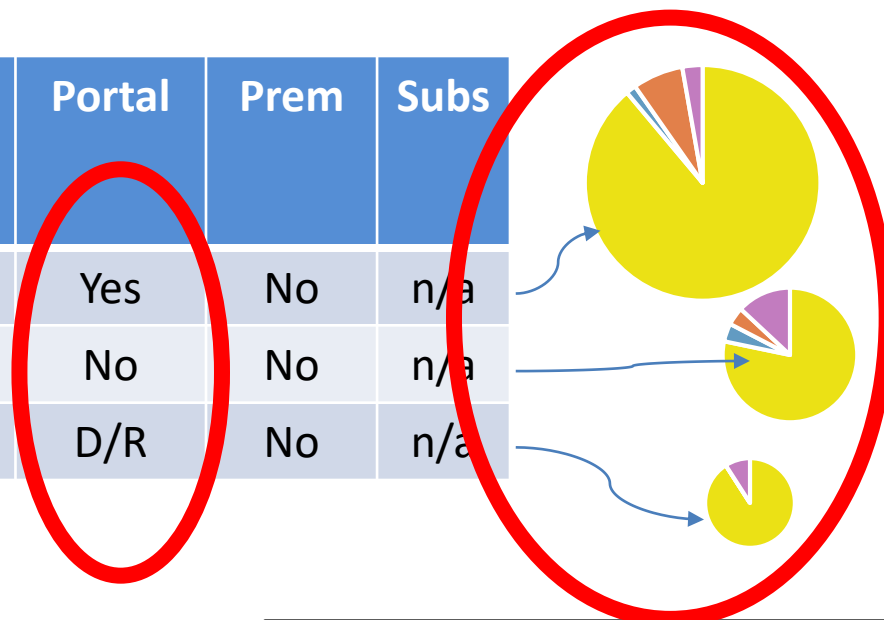
Joanne.Pascale@census.gov

Data Reduction Example

Perm	Freq	Type of Govt/State Cov	What do you call program	Portal	Prem	Subs
A	70	Public	Medicaid	Yes	No	n/a
B	22	Public	Medicaid	No	No	n/a
C	13	Public	Medicaid	D/R	No	n/a

Data Reduction Example

Perm	Freq	Type of Govt/State Cov	What do you call Program	Portal	Prem	Subs
A	70	Public	Medicaid	Yes	No	n/a
B	22	Public	Medicaid	No	No	n/a
C	13	Public	Medicaid	D/R	No	n/a

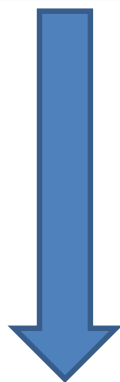
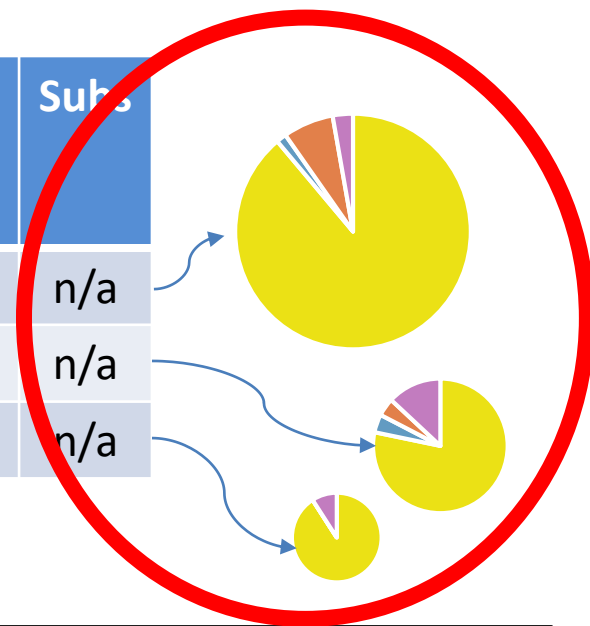


Cov Type According to Records



Data Reduction Example

Perm	Freq	Type of Govt/State Cov	What do you call program	Portal	Prem	Subs
A	70	Public	Medicaid	Yes	No	n/a
B	22	Public	Medicaid	No	No	n/a
C	13	Public	Medicaid	D/R	No	n/a



Cov Type According to Records



Perm	Freq	Type	Program	Portal	Prem	Subs
ABC	103	Public	Medicaid	Y,N,D/R	No	n/a

Premium and Subsidy

Verbatim Questions

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.