Strategies to Recruit and Represent Low Incidence Population Groups and Engender Sample Diversity in a Probability Panel

2022 FCSM Research & Policy Conference
Session B-6: Leveraging Probability Panels for Federally Sponsored Statistical Data Collections

David Dutwin, SVP, NORC
Ipek Bilgen, Sr. Research Methodologist, NORC

AmeriSpeak

Federal surveys have a "higher calling" which includes deep representivity

- Particularly today, a focus on diversity in surveys means sample and consequently respondent diversity

- Particularly today, a focus on equity in surveys means:
  - Ensuring that all persons are equally able to participate
  - Ensuring that our survey designs encourage as many people, equally and equitably as possible, to, in fact, participate

- This study serves as a case study in this regard: how the AmeriSpeak panel has made substantial efforts toward sample and panelist diversity and equity
  - What have we done with regard to sampling?
  - Special efforts to increase Spanish speakers, AANHPI
  - Impact of a new "equitable" design of recruiting materials.
  - Efforts to engage respondents we fear may disengage

Big Data to find hard to recruit populations

AmeriSpeak

Big Data Classifying

Predicting ahead of time whether a household has a person who is "x" (young, poor, Spanish speaking, smokes, etc.)

Potential Utility:

- To develop indictors far more effective than sample appends/vendor flags and/or geographic targeting.
- Significantly higher incidence = lower survey costs.
- Significantly higher coverage = lower weighting variances.

Anything can be modelled *so long as there is an available training dataset with self-reporting on the variable of interest, address, and some indicator of respondent (name, or demo combinations like gender + age).*

Models will vary in effectiveness: many explain more than half of indictor variance, but others have not been shown to be effective.

What Big Data?

There are a range of Big Data available, and at differing levels of measurement:

- **Governmental data** including Census Planning Database, the ACS, data from FEMA, the EPA, etc.

- Voter Registration data

- Consumer data:

  - Financial data:
    - Investment activities
    - Credit statistics
    - Personal finances and spending
  - Lifestyle indicators (hobbies)
  - Automobile purchases
  - Channel indicators (media use, subscriptions, etc)
  - Housing/Real estate data

  - Demographics
  - Travel
  - Sports use/interest
  - Music interests
  - Retail transactional data
  - Health indicators
  - Social media use

Process

- **Acquire and Match** Big Data

- **Merge** onto a large, high-quality survey

- Apply Data Science to **Train** a model

- **Test** the model

- **Apply (Score)** to a survey project
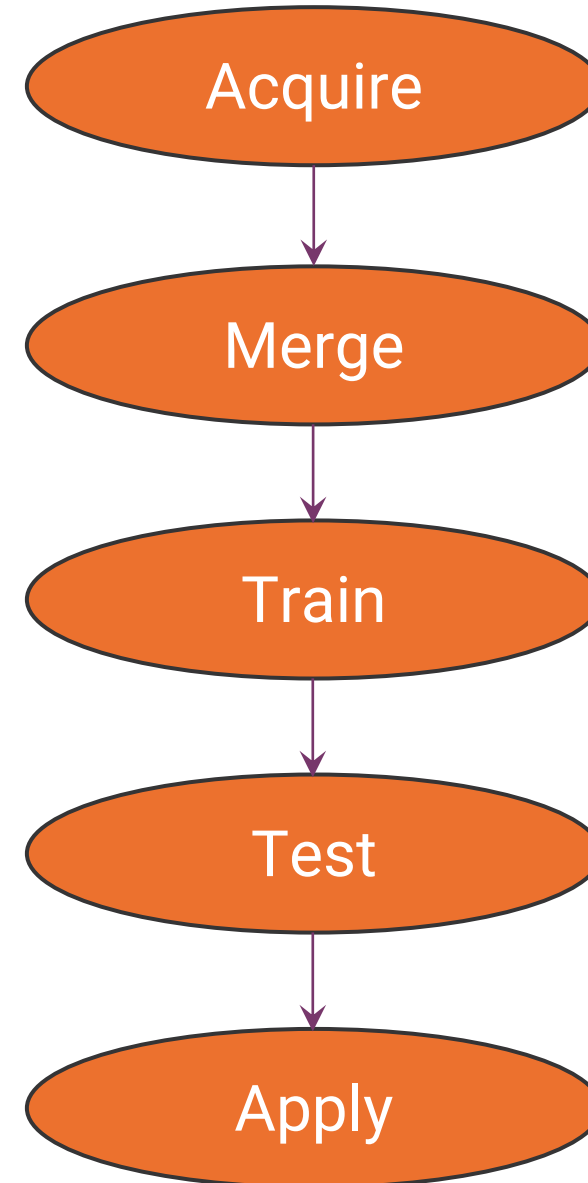
Impact of Big Data Classifier Sample Stratification on Recruiting

2021 Results w/ Big Data Sampling vs. 2020 Recruits
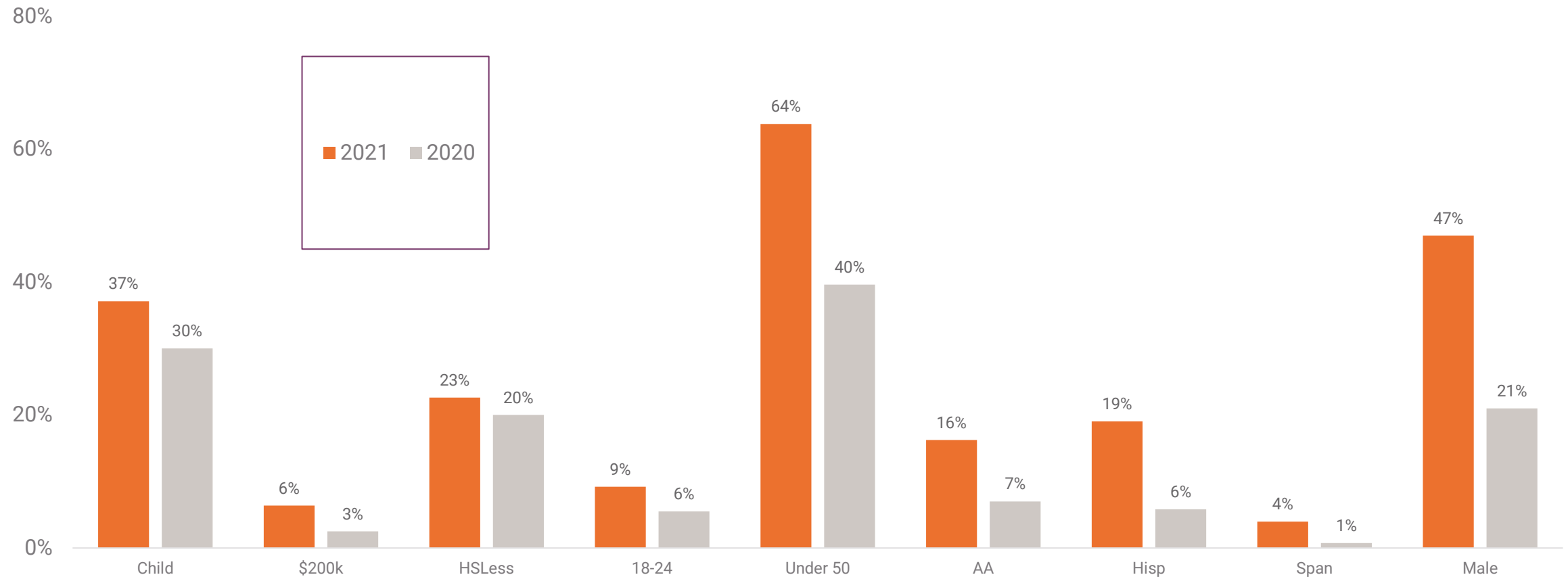
Increasing Spanish Language

AmeriSpeak

The Challenge of Spanish Language

In short, you can target such households with Big Data classifiers, but they still may no choose to join as a Spanish speaker:

- Multiple studies find that Spanish speakers participate less, and when they do, they often strive to complete in English (for ex: CHIS 2019).

- Added 2,000 addresses identified as having the highest likelihood of containing Spanish speaking residents.

- Specifically in CA (LA, Fresno, San Diego, Modesto areas), TX (Houston, Dallas), FL (Miami, Orlando), NY (New York), and IL (Chicago) (approx. 60% of predicted Spanish speakers nationwide)

- Sent a 2-day FedEx packet followed by in-person visits in the fall of 2021. Hired 20 additional Bilingual interviewers (~50 in total)

- An average of 2.5 visits were made to each household

Spanish Dominant FedEx Packet Brochure:

Bilingual/Spanish Leave Behinds: Door hangars, Sorry I Missed You cards, & other

Results of Spanish Language Recruiting Sample 2021

Case Status	N	%
TOTAL	**2,000**	**100%**
Complete	432	22%
Hispanic	1,801	90%
Spanish: Fluent Read/Speak	1,702	84%
Spanish survey takers	980	49%

Percent Spanish Recruit w/in Spanish Big Data Classifer by Modality of Recruiting
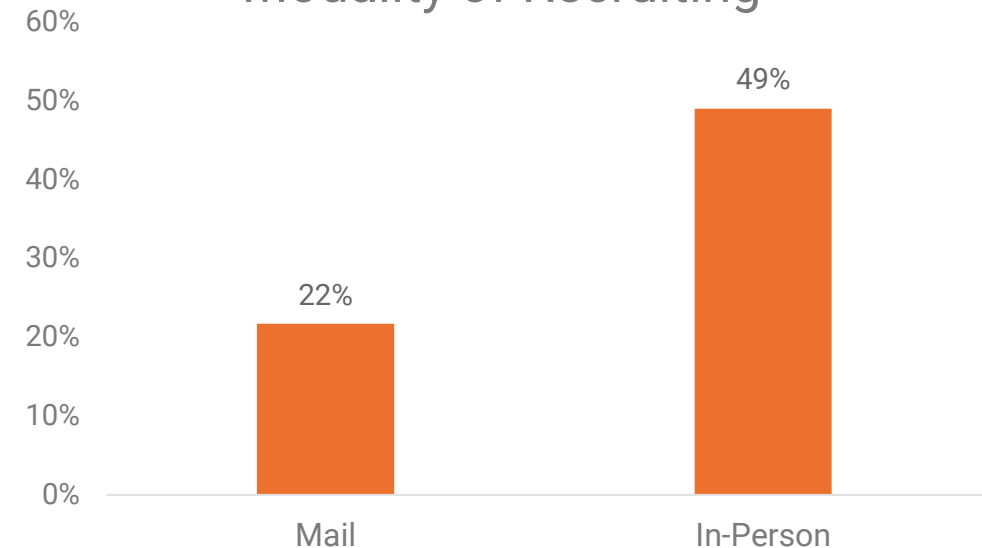
Increasing AANHPI

AmeriSpeak

Turning AANHPI from "Other" to a Robustly Represented Group

- AmeriSpeak working with AARP to build an AANHPI Panel that would supplement AmeriSpeak but also be able to support its own standalone surveys of the AANHPI population

- Aiming to support 1,000 panel interviews in early 2023 and 2,000 by end of 2023; 3,000 end of 2024.

- Pilot in 2022 resulted in 341 screened AANHPI interviews of which 292 agreed to future survey participation/empanelment

- Sample design leverages Big Data Classifiers + medium and high incidence CGBs.

- Multiple major funders already committed; continuing to seek additional funders to realize the full build

Largely representative of the AANHPI population:

Speak Non-English Language At Home

	ACS	S1 Realloc
Yes	75.7%	65.0%
No	24.3%	34.0%

Age

	ACS	S1 Realloc
18 - 34	32%	25%
35 - 49	29%	29%
50 - 64	22%	29%
65 Plus	16%	17%

Heritage

	ACS	S1 Reallocated
Asian Indian	21.2%	18.3%
Chinese	23.7%	24.4%
Filipino	16.6%	13.3%
Japanese	4.4%	11.9%
Korean	8.0%	7.1%
Vietnamese	9.8%	7.3%
Other Asian	11.7%	4.0%
Native Hawaiian	0.9%	2.7%
Samoan	0.6%	0.6%
Fijian	0.2%	0.9%
Other Pacific Islander	1.1%	0.3%
Multi-AANHPI	1.8%	9.2%
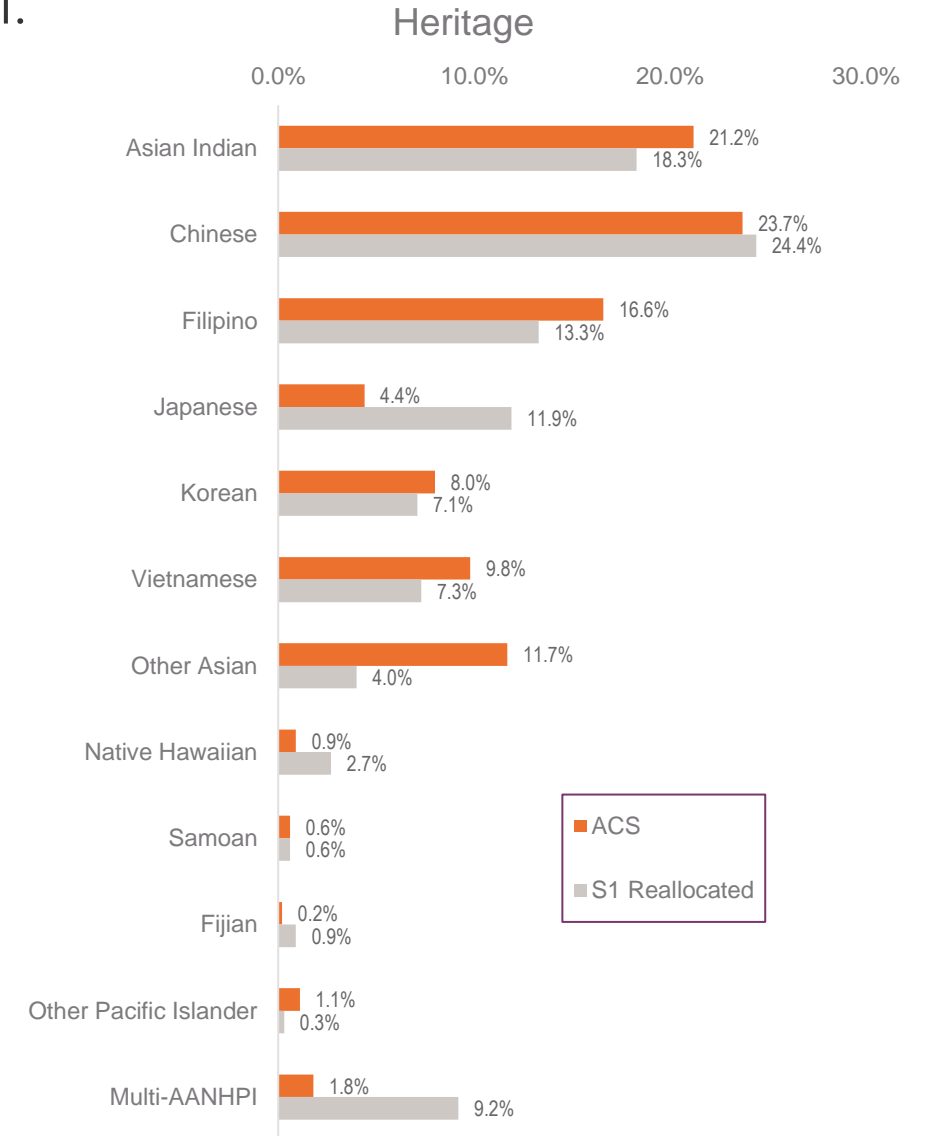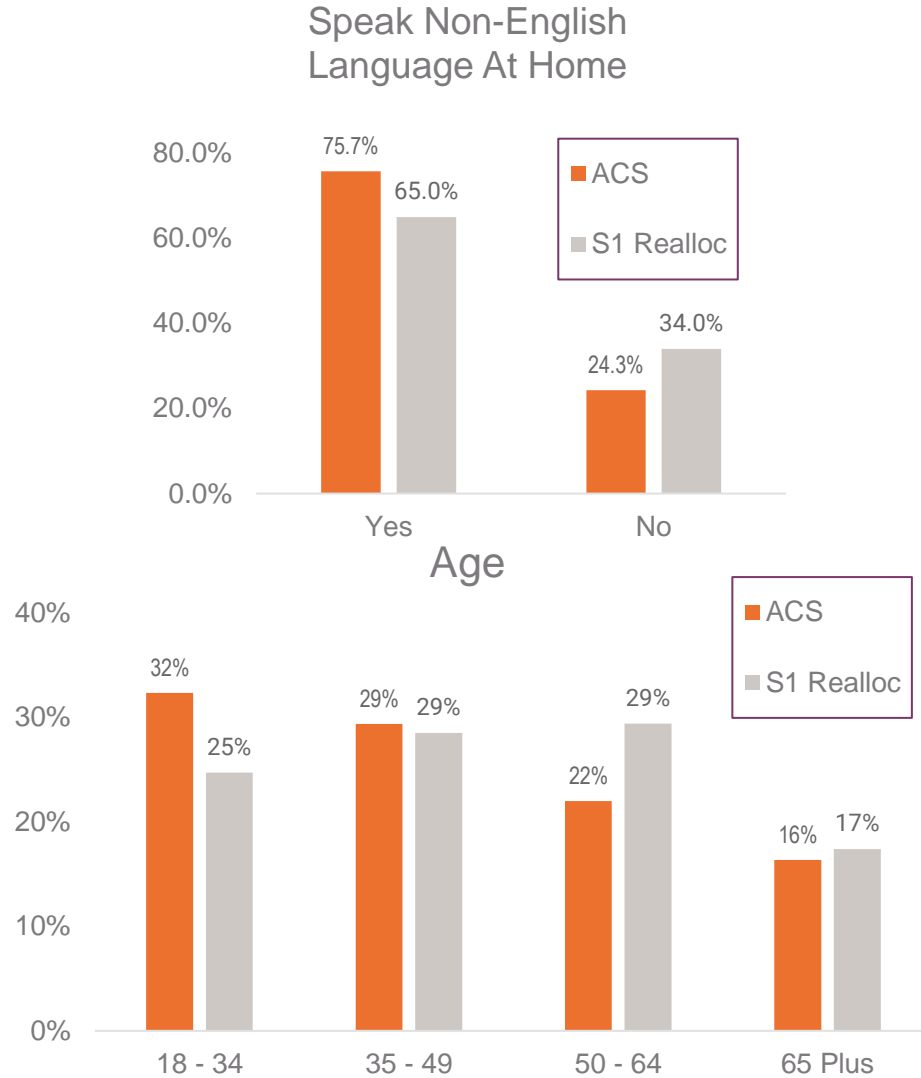
Creating Equitable Materials

AmeriSpeak

AmeriSpeak recognizes the need for equitable designs

- Most Americans don't read or speak like survey researchers do!

- Surveys historically fail getting at low education, young, and non-white respondents

- More currently, some concern with strong Republicans

AmeriSpeak recognizes the need for equitable designs

- Most Americans don't read or speak like survey researchers do!

- Surveys historically fail getting at low education, young, and non-white respondents

- More currently, some concern with strong Republicans

Results: Improvements in race/ethnicity and age

Race Ethnicity

Yield by Education Category

Yield by Ideological Category

Yield by Age Category
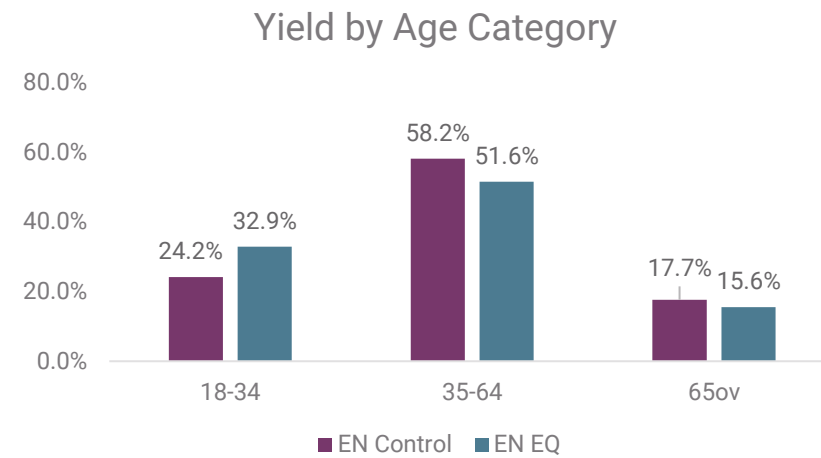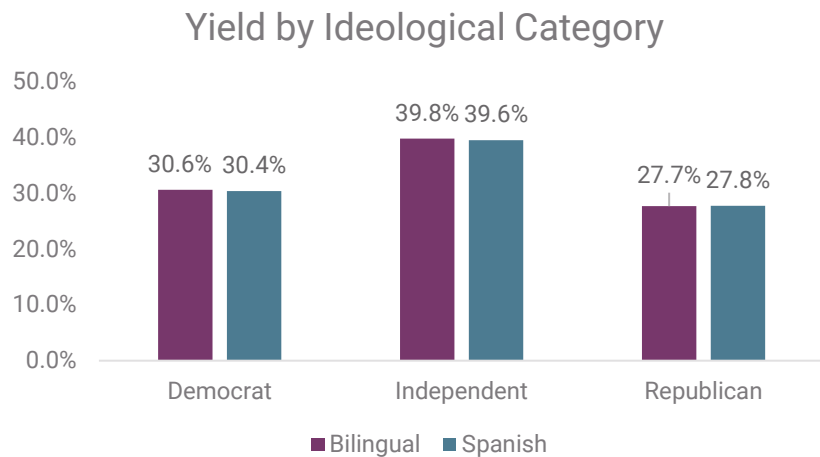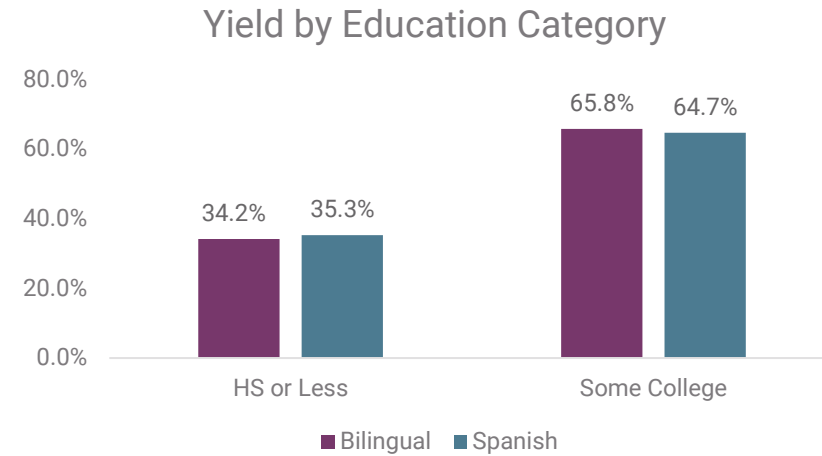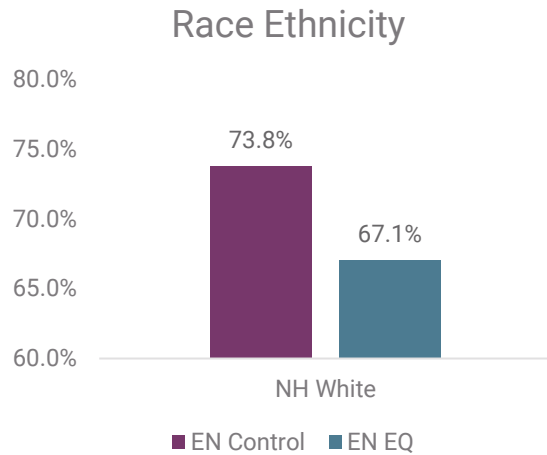
Retainment is Just as Important and Recruitment

AmeriSpeak has made significant efforts to increase retainment and participation over time

- 1st 6 months are critical: 50% of attrition occurs at this time.

- Therefore, testing 1st 6th month mail and e-mail retainment communications

- Specific campaign to engage Hispanics, who attrit the most in the 1st 6th months

- Additional mailings

- Modelling future participation (see Stas!)

Questions?

Thank you.

Get Your Research Right

AmeriSpeak