Sampling to Balance Community Engagement and Representativeness

October 7, 2025

Presented by: Brett Close, TRC
Donna Whitsett and Christy Zook, TRC

IEPEC

TRCCOMPANIES.COM

The Randomness Mirage and its Challenges

One key idea and two pieces of context

Randomized contact lists do not make a random sample

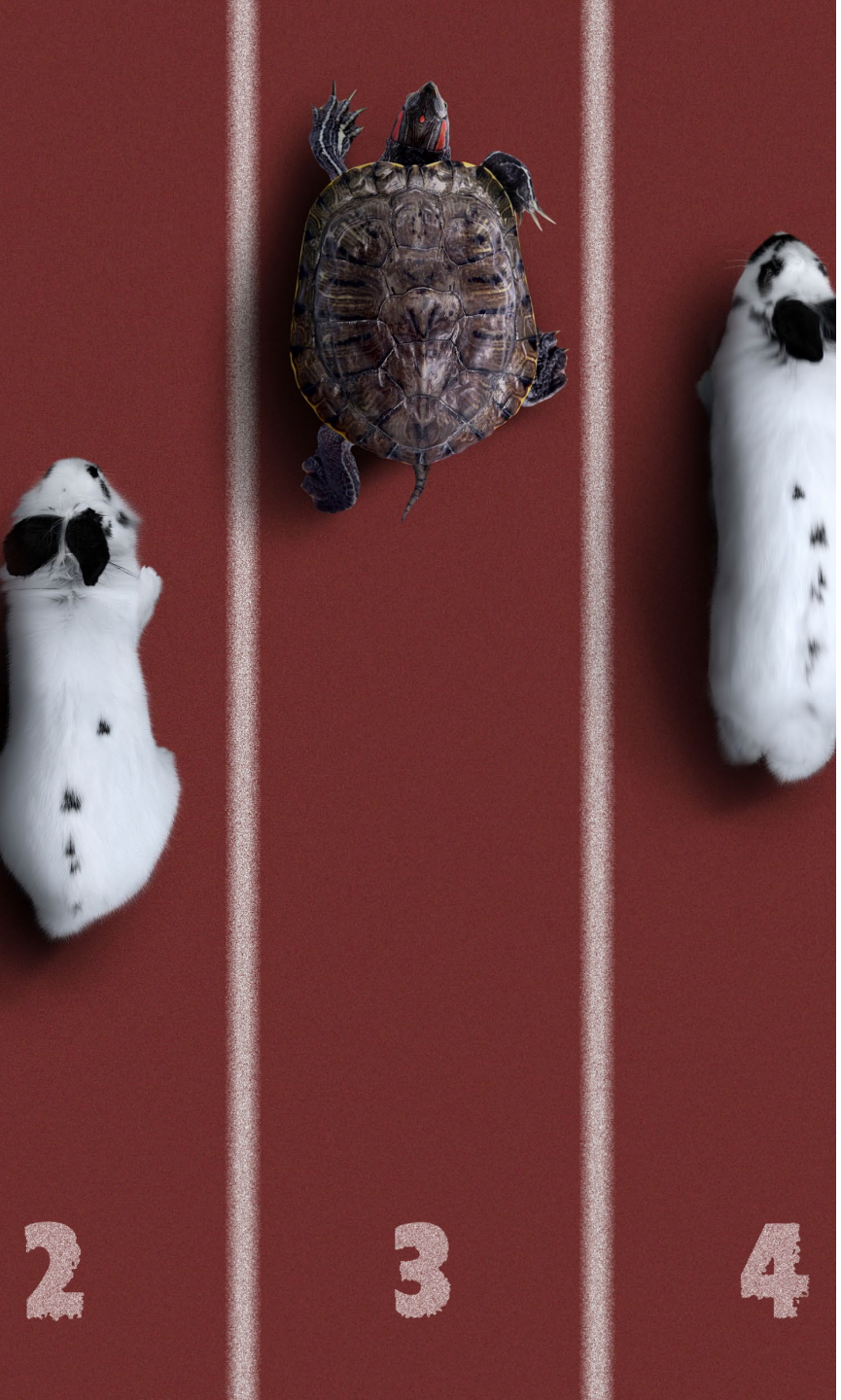
A random sample requires that all members of the population have
equal probability of being in the sample.

Non-response bias tends to leave out the perspective of specific groups

In particular, people who are lower-income, non-white, and whose primarily language is not English

Approaches to address non-response bias are expensive (and may be intrusive)

Simultaneously, the industry is experiencing pressure to deliver lower-cost programs and studies.



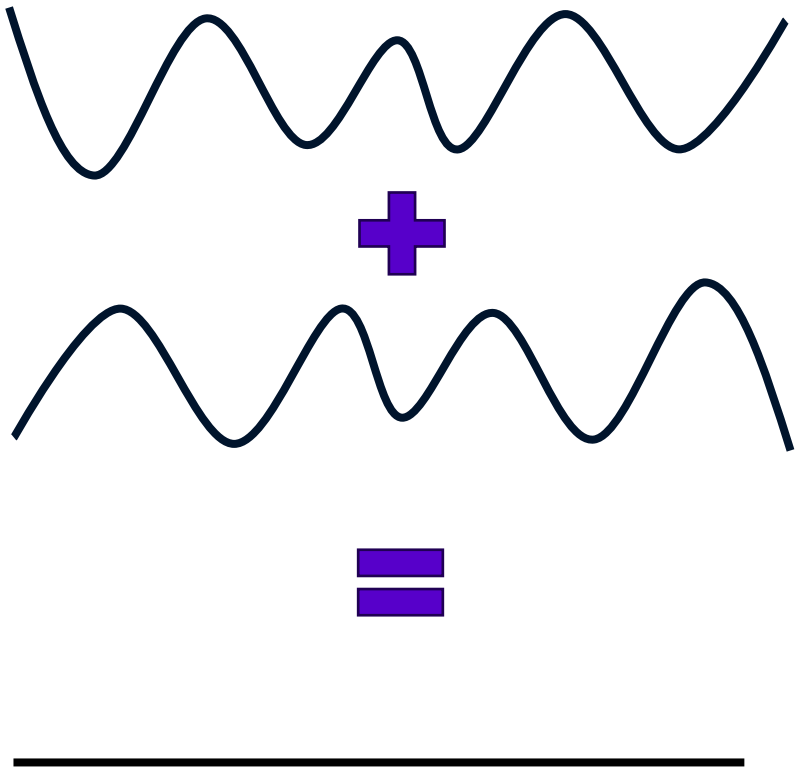
Summary: We are not in the realm of random sampling and are likely to have results that reflect white, well-off respondents more than others

We are not in a “first best” vs. “second best” scenario, we are looking for something feasible.

One Approach:

“Noise-canceling headphones” sampling

Introduce targeted non-randomness to cancel-out existing non-randomness by partnering with Community-based Organizations



Our problem is a systematic non-randomness. Can we introduce a targeted non-randomness to cancel it?

We partnered with community-based organizations (CBOs) with strong local connections in under-represented communities to increase response rates

A large market characterization study presented a sampling challenge, and an opportunity to try a creative solution

- **Study:** A market characterization for California Air and Resources Board (CARB)
- **Population:** Households and business in California that use propane for space and water heating
- **Known Challenges:**
  - Respondent recruitment
  - No well-defined population list

A hybrid sampling approach:

Combine a traditional contact list approach with CBO engagement

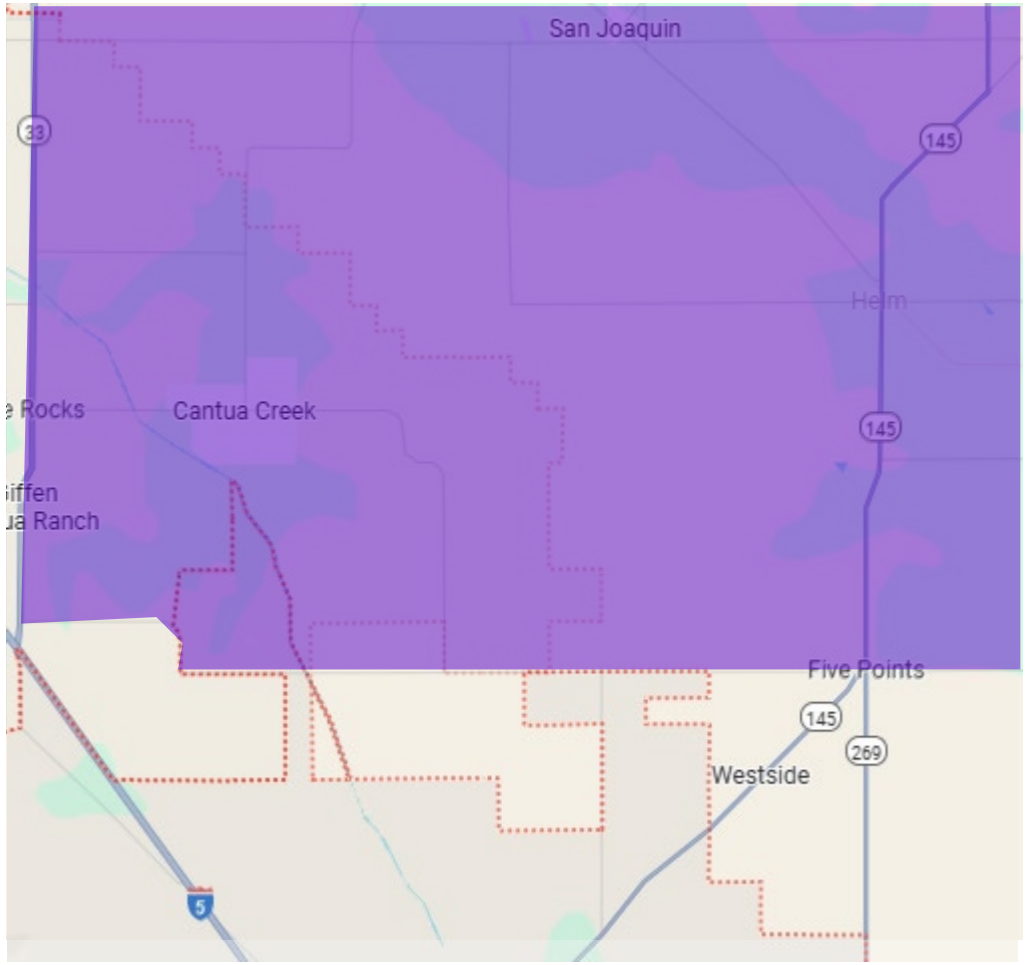
1. **Typical approach: contact lists**
   - Propane use tends to be spatially correlated, so we developed a geographically-based approach
   - We then randomly recruited from contact lists for areas with high rates of propane and wood space heating

2. **Supplemental approach: engagement from community-based organizations (CBOs)**
   - Worked with CBOs that have strong community connections to obtain additional responses from disadvantaged communities

Design and implementation challenges:

Defining Units and Balancing Strata

Meaningful geographic designations do not map neatly to datasets

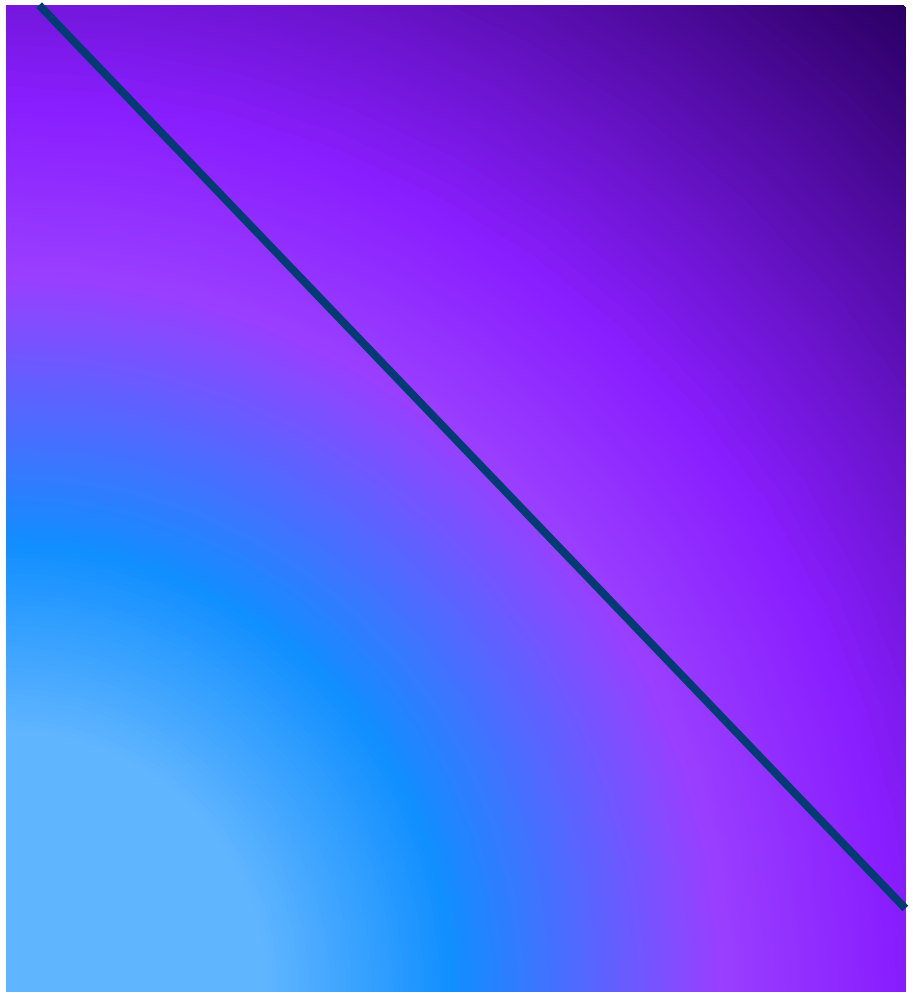


Block Groups vs. Mailing Addresses

- Demographic data from the Census and ACS are reported for **block groups**
- But most people do not know their block group
- Most recruiters cannot match an address to block group
- ZIP codes and block groups do not match neatly

- We aggregated to USPS Cities, groups of ZIP codes that correspond to addresses, not to municipal boundaries

Researchers have less control over sampling balance when working with external partners than when sampling from a list



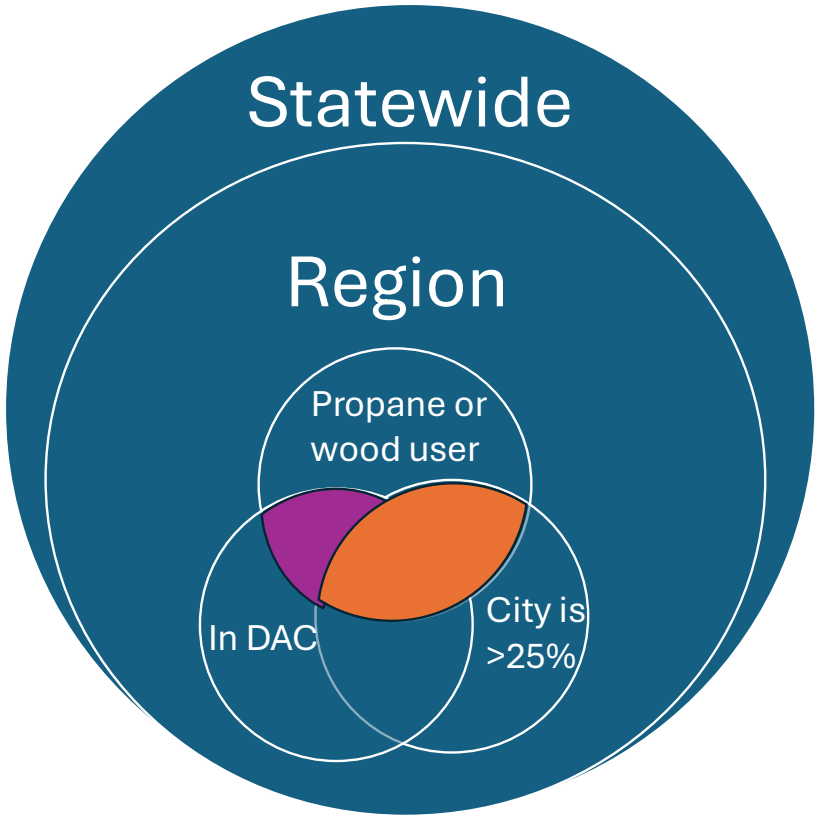
- We wanted to avoid a situation where too many sample points were within a specific area, so we **provided quotas** to our CBO partners
- To develop quotas, we treated the problem as a constrained minimization of the expected variance using the method of **Lagrange multipliers***
- In this setting, the results differ from traditional Neyman allocations

*If you love multivariable calculus, see the paper!

Analysis challenges:

Weighting

With the hybrid sample we had additional considerations for determining the weights for each respondent



What population we are targeting?

- Ideally: all propane and wood users
- In practice: some subset

What population is the CBO recruitment is targeting?

- Ideally: individuals in disadvantaged communities
- Reality: individuals in specific disadvantaged communities

What is the population our list-base recruiting is targeting?

- Ideally: all propane and wood users
- In practice: propane and wood users in communities with greater than 25% propane and wood

With the hybrid sample we also needed an approach for reporting weighted vs. unweighted results

- Weighting makes a non-random sample more like a random sample
  - Makes the sample characteristics more like the population
  - But what are the characteristics to focus on?
- If we weighting on demographic characteristics, it can **improve representativeness**, but can also lead to **unstable results** with small cell weights.
- **Solution:** use weighted results for reporting state-wide or regional values and unweighted results for reporting on specific demographic groups.

We are working in an imperfect setting. Hopefully, this approach moved in the right direction.

There were many challenges with this approach and there is room for improvement.

Thank You

For more detail, questions, or discussion contact:

Brett Close, bclose@trccompanies.com