

2022

Online Surveys in Latin America: Assessing Nonprobability Sampling Approaches

Oscar Castorena
Vanderbilt University

Noam Lupu
Vanderbilt University

Maita Schade
Vanderbilt University

Elizabeth Zechmeister
Vanderbilt University

Follow this and additional works at: https://ir.lib.uwo.ca/c-dem_conf_survey

Citation of this paper:

Castorena, Oscar; Lupu, Noam; Schade, Maita; and Zechmeister, Elizabeth, "Online Surveys in Latin America: Assessing Nonprobability Sampling Approaches" (2022). *Workshop: Survey Design Under Constraints*. 3.

https://ir.lib.uwo.ca/c-dem_conf_survey/3

Online Surveys in Latin America: Assessing Nonprobability Sampling Approaches

Oscar Castorena, Noam Lupu, Maita Schade, and Elizabeth J. Zechmeister



VANDERBILT
UNIVERSITY®



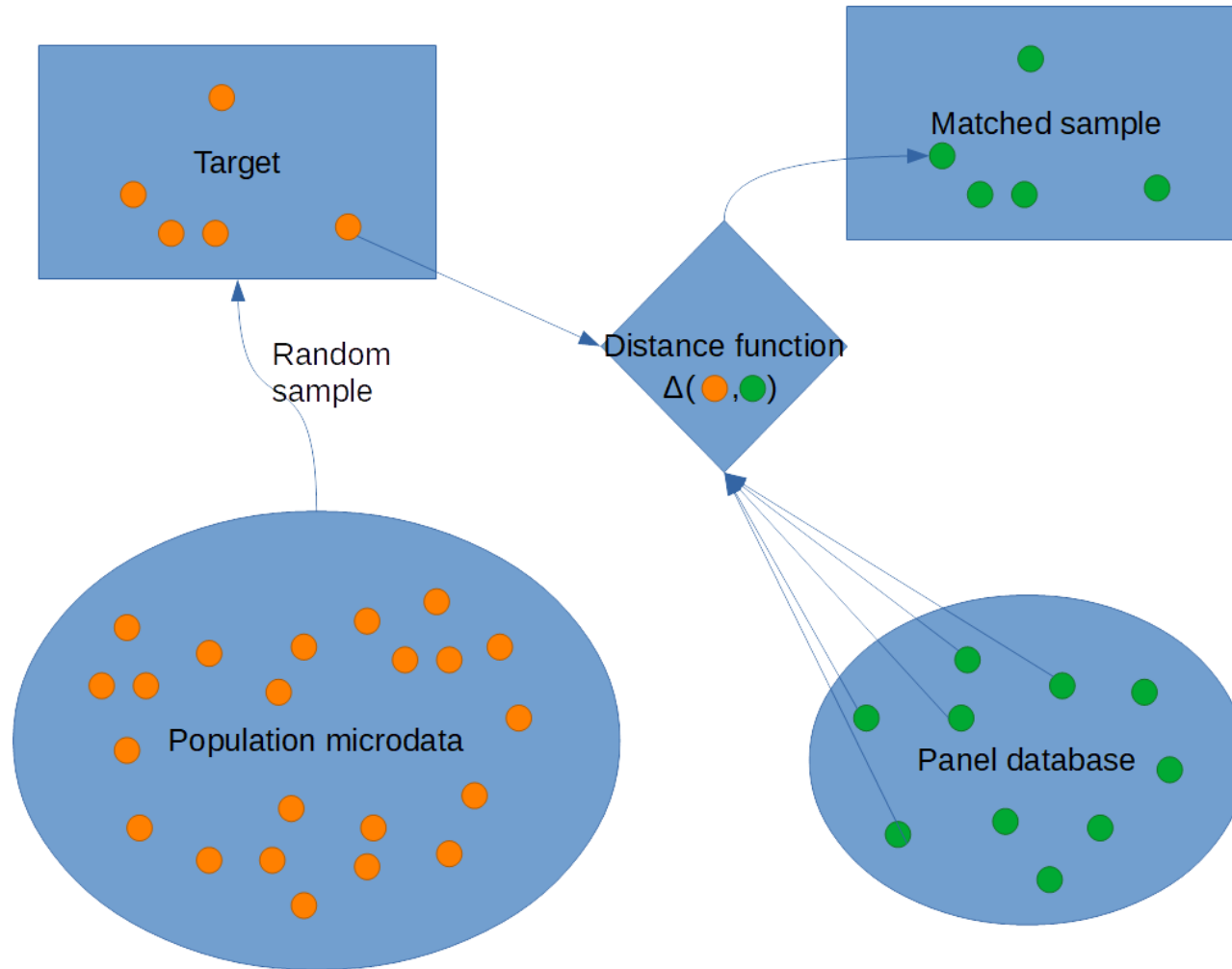
Motivation

- Internet surveys pose many opportunities
 - Increasing difficulty of standard telephone and face-to-face methods
 - Low cost and faster data collection
 - Increasing reach and availability of online panels
- Challenges
 - Coverage issues especially in the global South
 - Opaque sampling practices
- Sampling
 - Probability-based panels (e.g. Pew's American Trends Panel, GfK's KnowledgePanel)
 - Non-probability methods
 - Quota sampling (among other opaque processes; most common approach)
 - **Sample matching** using opt-in panels (best practice)

Research question

Would sample matching improve the quality of nonprobability online samples compared to the standard sample generated by a commercial panel provider?

Sample matching – what is it?



Research design

- Comparison study
 - Work with panel provider operating in Latin America
 - Survey in Brazil, Argentina, and Mexico
 - For each survey (N=2400), half sample collected using sample matching and other half collected using panel provider's standard method
- Sample matching
 - Reference population: census
 - IPUMS-International microdata census/intercensal survey samples
 - Dating from 2010/2015
 - Sampling frame: panel
 - Provider's proprietary panels
 - Recruited from invitational ads on social networks
 - Between 160k – 368k individuals per country
 - Enforced match on age decile, gender, and residence in capital
 - Nearest-neighbor matching algorithm
 - Using whichever variables available (education, car ownership, etc.)

Research design

- Benchmark questions
 - Drawn from official face-to-face survey in each country
 - Indicators such as:
 - Home ownership
 - Number of rooms in the household
 - Number of persons in household
 - Use of government assistance
 - Employment status
 - Question wording and response options identical in benchmark survey and comparison study
- Included among a core survey of approx. 100 questions about social and political attitudes (median survey duration 26-29 minutes)
- Spring of 2020

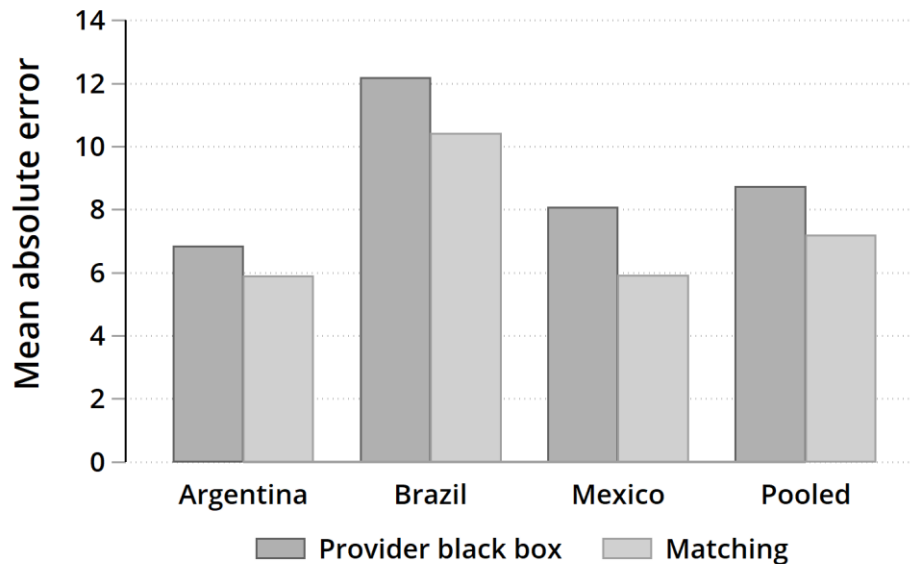
Comparison analysis

Mean Absolute Error (MAE): the average difference, across a number of benchmark questions, between the proportion of respondents falling into the modal response in the population and that proportion in the sample.

- N = number of comparison questions
- r_i = response to question i (e.g., reported number of rooms in house)
- \tilde{r}_i = modal response to question i in benchmark survey
- $p(\tilde{r}_i)$ = proportion of respondents in benchmark survey choosing \tilde{r}_i
- $p'(r_i)$ = proportion of respondents in sample for whom $r_i = \tilde{r}_i$

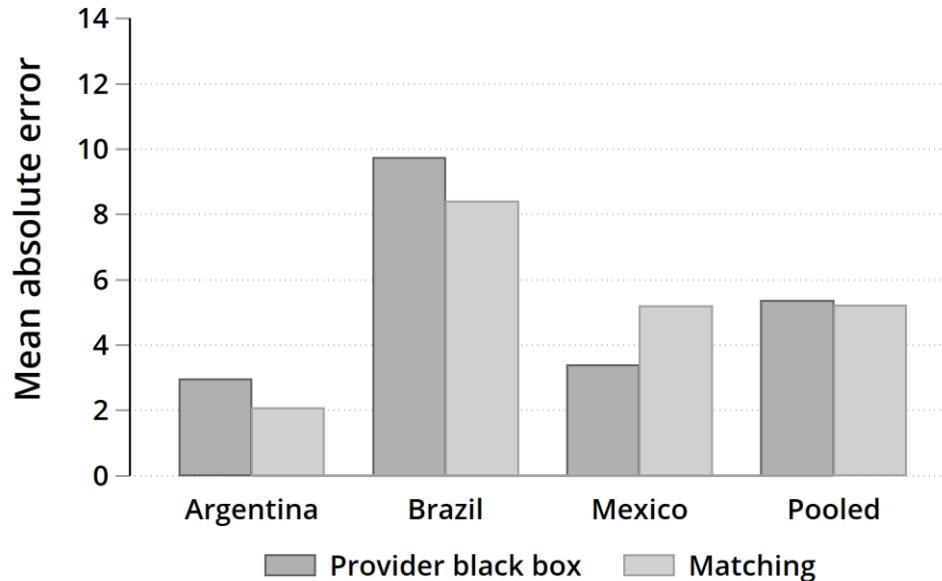
$$MAE = \frac{1}{N} \sum_{i=1}^N |p'(r_i) - p(\tilde{r}_i)|$$

Comparison analysis



- Matched sample has consistently smaller MAE
- Difference in MAE is statistically significant when pooling all three countries (7.2 %-pts in matched vs 8.7 %-pts in black box sample)
- Overall, matched sample outperforms provider sample on benchmark questions

Comparison analysis (demographics)



- Samples perform similarly with respect to demographic variables
- Provider samples are typically generated by quotas on these variables

Tradeoffs

- Improvements
 - Sample-mating marginally outperforms “standard” black-box approach
 - Transparency and reproducibility of sampling procedures
- Drawbacks
 - Iterative process takes time (6 weeks as opposed to 1-2 weeks per sample)
 - Significant demands on staff on both sides
 - Data requirements

Conclusions

- Both transparency and sample quality can be improved
- Significant costs in terms of time and effort
- Default to panel provider methods in recent online studies
- Additional avenues for improving sample and data quality, specifically questionnaire design

Thank you!

We look forward to your comments and questions!

Contact info:

Oscar Castorena

oscar.castorena@vanderbilt.edu