Evaluating a Propensity Score Adjustment for Combining Probability and Non-Probability Samples in a National Survey

FedCASIC 2015

Environment Canada

Outline

- **2012 Canadian Nature Survey**
  - Research questions
  - Survey design
- **Weighting Methodology**
- **Results (Comparison of weighted estimates)**
- **Conclusions**

Research Questions

- **National population survey of Canadian adults**

Connection to & awareness of nature

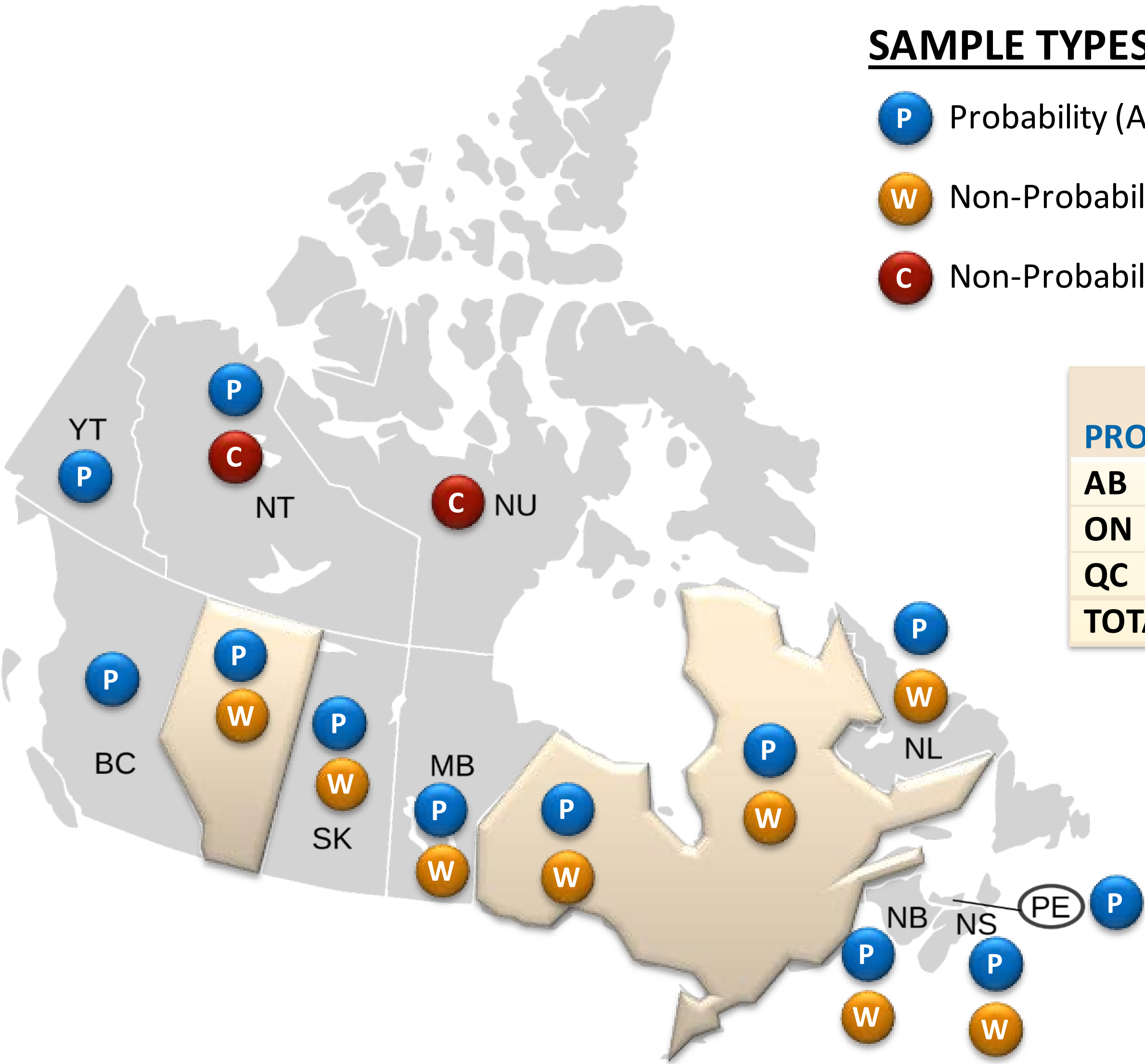
Nature-based activities, participation, and expenditures

Human/wildlife conflict

Survey Design

- **Complex sample design with hybrid probability and non-probability samples**
- **Multi-mode administration (Paper + Web)**
- **For probability sample (nationally):**
  - 76,363 addresses sampled from ABS frame
  - 15,207 completes
  - 20% response rate (lower bound)
- **For non-probability samples (nationally):**
  - 8,897 completes

Survey Design



SAMPLE TYPES

- P Probability (ABS)
- W Non-Probability (Web Panel)
- C Non-Probability (Community)

PROVINCE	ABS RESPONSES	WEB PANEL RESPONSES
AB	1,511	818
ON	1,011	4,584
QC	1,029	2,986
TOTAL	**3,551**	**8,388**

Survey Design

Address-Based Sample of Canadian Adults

- Drawn from Canada Post address file
- Stratification:
  - Province/Territory (all except Nunavut)
  - Urban/Rural address (Canada Post frame variable)
- Mode of Administration:
  - Paper, with Web option
- Within-HH selection by Last Birthday Method
- Targeted 1,000 completes in each province and territory

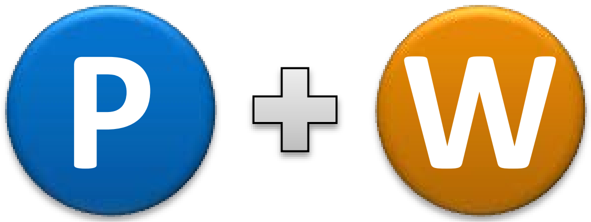
Survey Design

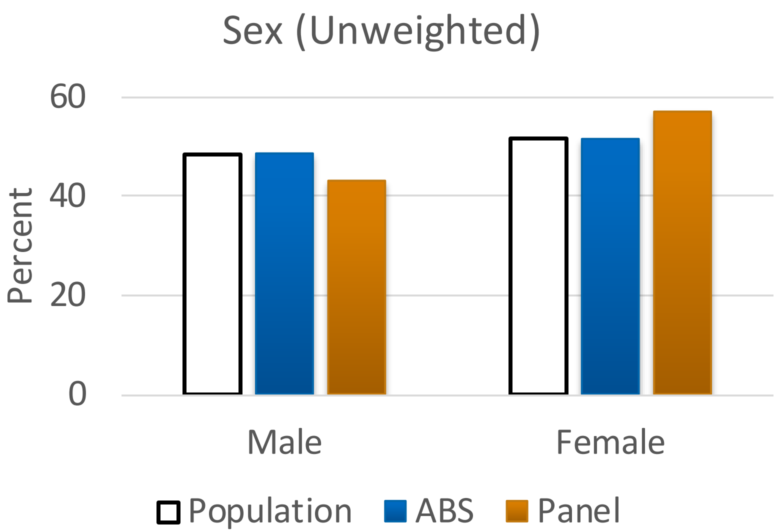
Web Panel Sample

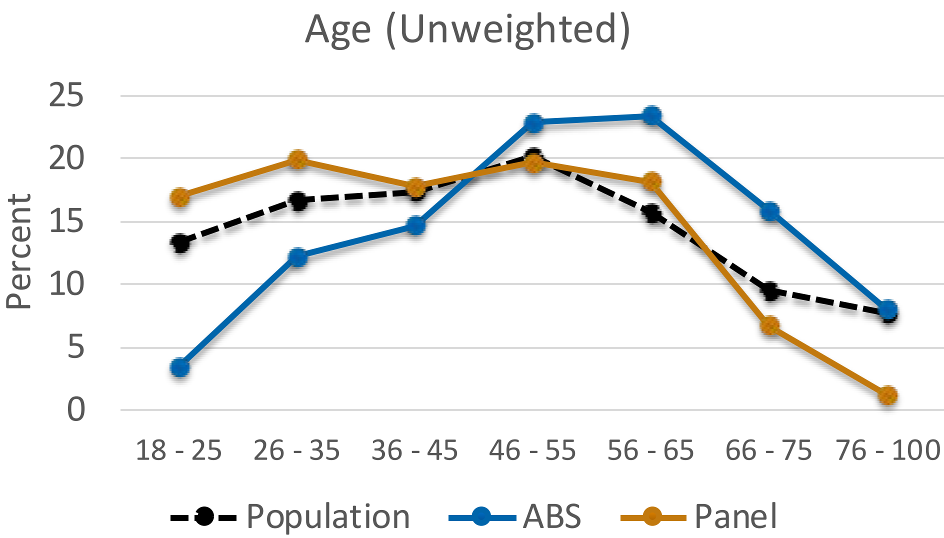
- Canadian adults recruited via social media and websites
- Recruited to match key demographic distributions (race, age, education, income)
- In each P/T, fielded until target number of completes was reached

Weighting Methodology

- **Focus of current research is evaluation of weighting to combine the probability (ABS) and non-probability (Web panel) datasets for analysis**







Weighting Methodology

- **An ABS analytic weight was developed for ABS respondents**
  - Standard probability-based selection weight adjusted for non-response and post-stratified to Census totals:
    - Province x Age x Sex
    - Province x Urban/Rural
    - Aboriginal/Non-Aboriginal

Weighting Methodology

- **The following approach was explored for combining the ABS and Panel respondents into a single weighted dataset:**
  1. Estimate probability of observation in Panel (vs. Population)
  2. Score all (Panel and ABS) cases to assign a probability of observation under Panel design
  3. Assign probability of observation under ABS design to Panel cases
  4. Combine ABS and Panel observation probabilities to compute combined weight

Weighting Methodology

- **Estimate probability of observation in Panel (vs. Population) using weighted logistic regression**
  - Outcome = Observation in Panel (vs. Population)
    - P(Observation) = P(Selection) * P(Response)
  - Weights:
    - For ABS cases, weight = ABS analytic weight (post-stratified to population)
    - For Panel cases, weight = 1

Weighting Methodology

- **Estimate probability of observation in Panel (vs. Population) using weighted logistic regression**
  - Predictors:

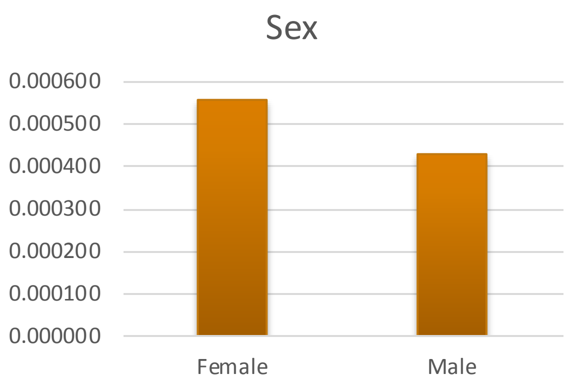
Effect	Comparison	Odds Ratio	
Province	AB vs QC	0.7	
	ON vs QC	1.1	*ns*
Age	18 - 25 vs 76 - 100	7.9	
	26 - 35 vs 76 - 100	8.7	
	36 - 45 vs 76 - 100	7.7	
	46 - 55 vs 76 - 100	6.4	
	56 - 65 vs 76 - 100	7.3	
	66 - 75 vs 76 - 100	4.3	
Sex	Female vs Male	1.2	
Urbanicity	1 vs 2	1.2	
Nature-related Profession	0 vs 1	0.9	*ns*
Aboriginal	0 vs 1	1.1	*ns*
Immigrant	0 vs 1	1.3	
Education (Highest)	1 vs 8	0.4	
	2 vs 8	1.4	
	3 vs 8	2.4	
	4 vs 8	1.8	
	5 vs 8	1.2	*ns*
	6 vs 8	1.1	*ns*
	7 vs 8	1.2	*ns*
HH Income		0.9	

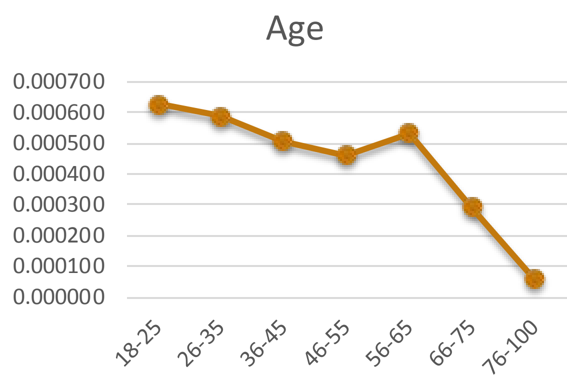
Base Model (Province through Urbanicity) $R^2 = .11$

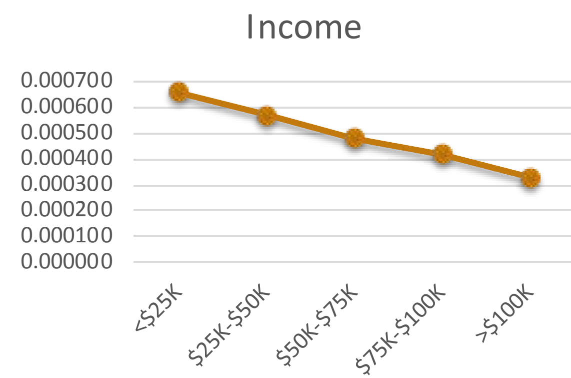
Full Model (all predictors) $R^2 = .19$

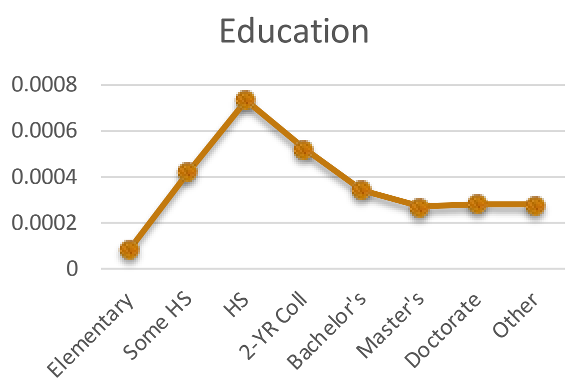
Weighting Methodology

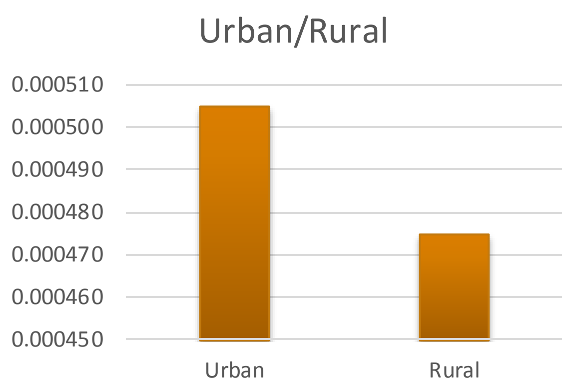
- **Score all (Panel and ABS) cases to assign a probability of observation in Panel**
  - Mean estimated probability of observation under Panel design:

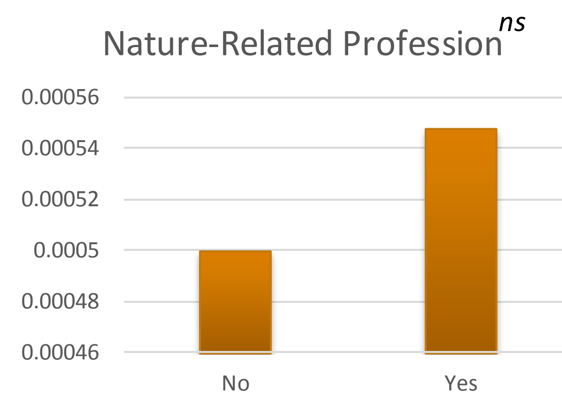












Weighting Methodology

- **Assign probability of observation under ABS design to Panel cases**
  - Probability of observation under ABS design computed as inverse of post-stratified ABS analytic weight
  - Within post-stratification classes, same ABS probability was assigned to Panel respondents
    - This assumes that ABS and Panel cases within these classes have the same probability of observation under ABS design
  - Result is that all cases in combined sample have a (true or estimated) probability of observation under both the ABS and Panel designs

P(Observation)

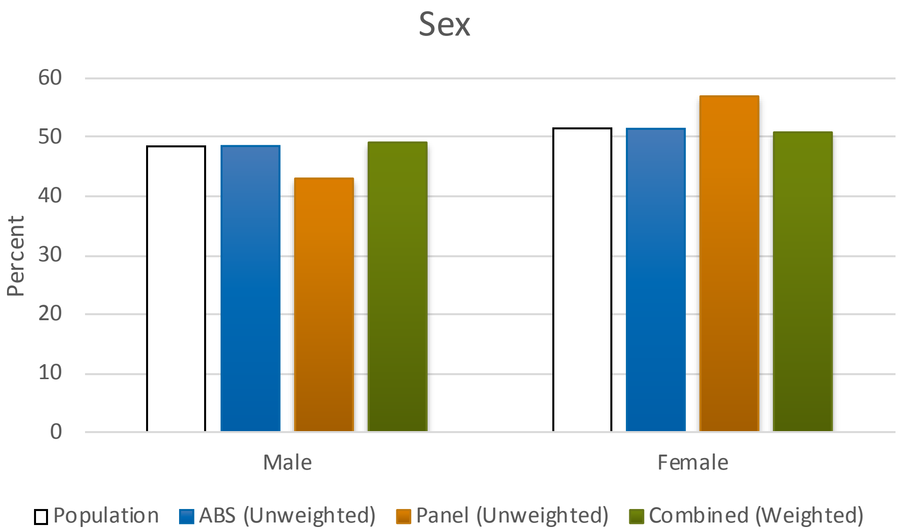
		ABS	***Panel***
Sample Source	***ABS***	Inverse of post-stratified, NR-adjusted ABS sampling weight ➡	Matched by post-stratification class
	Panel	Estimated Panel probability	Estimated Panel probability

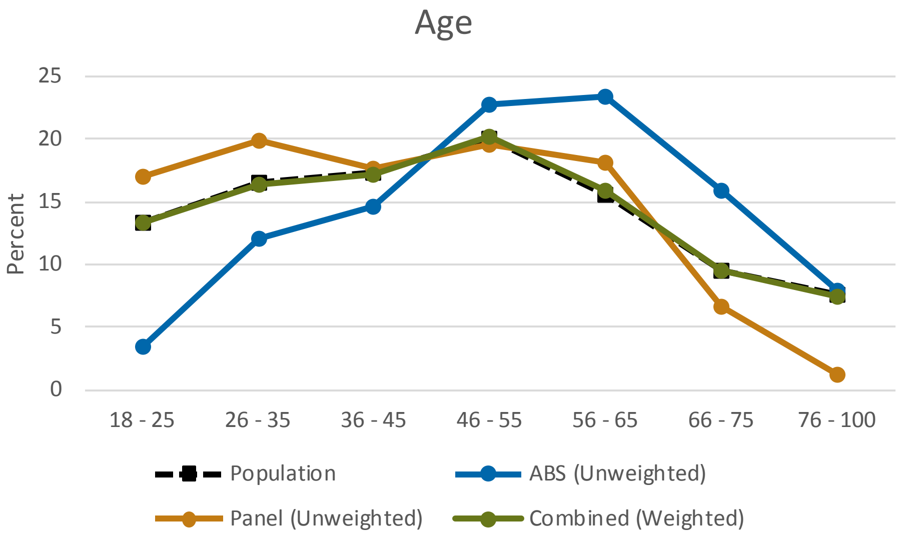
Weighting Methodology

- **Combine ABS and Panel probabilities to compute combined weight**
  - $p(ABS \cup Panel) = p(ABS) + p(Panel) - p(ABS) * p(Panel)$
  - $w_{combined} = 1/p(ABS \cup Panel)$

Results

- **Demographics**

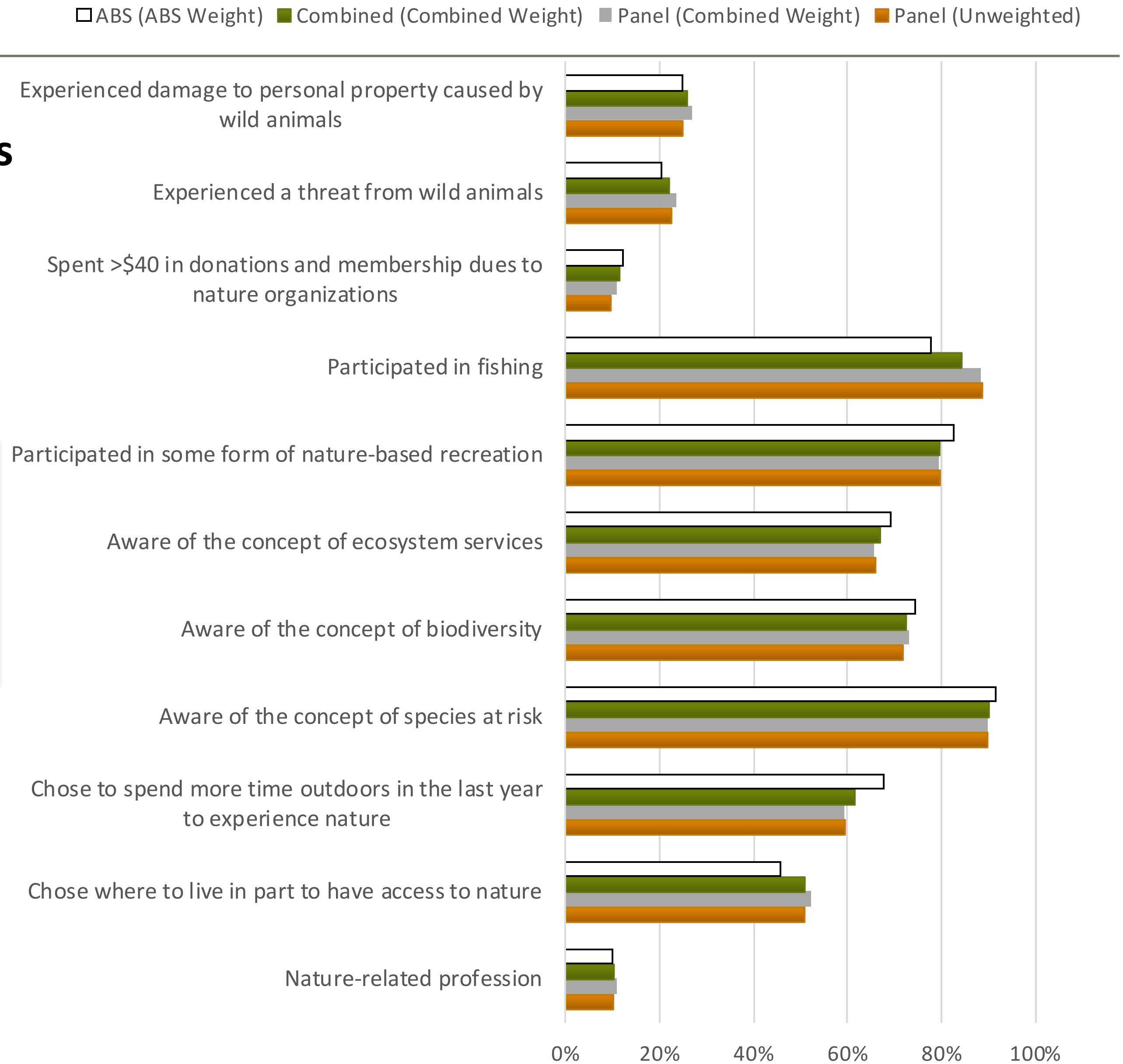




Results

- **Key Survey Outcomes**

MAD of Panel from ABS population estimates is 10% lower after weighting, and ~40% lower with combined weighted sample



Conclusions

- **Unweighted panel data differed from benchmarks**
  - Demographics: More female, younger, lower income, less educated, more urban
  - Outcomes:
    - Accurate (±2 points):
      - Nature-related profession
      - Aware of the concept of species at risk
      - Experienced a threat from wild animals
      - Experienced damage to personal property caused by wild animals
    - Overestimates (>2 points over):
      - Chose where to live in part to have access to nature
      - Participated in fishing
    - Underestimates (>2 points under):
      - Chose to spend more time outdoors in the last year to experience nature
      - Aware of the concept of biodiversity
      - Aware of the concept of ecosystem services
      - Participated in some form of nature-based recreation
      - Spent >$40 in donations and membership dues to nature organizations

Conclusions

- **Propensity score model was used to estimate probability of being observed in the panel compared to general population**
  - Model explained only some of the variance ($R^2 = .19$) – room for improvement
  - Nevertheless, estimated probability of observation
    - Brought panel demographics in line with population
    - Reduced bias in panel estimates for key survey outcomes
    - Made possible the combination of probability (ABS) and non-probability (Panel) data into a single, weighted dataset

Conclusions

- **Next steps…**
  - Building a more comprehensive model of P(Observation) under panel design
  - Does reduction in bias via panel weight come at the price of increased variance? How accurate are estimates of sampling error from modeled probabilities of selection?

Thank You!