Can the Web "Work" for Public-Use Surveys of Mass Publics?

Simon Jackman
Stanford University

The Problem

- American National Election Studies

- a "near-canonical" source of data on political attitudes and behaviors

- every presidential election 1948 to 2012, plus most mid-terms

- **in-person interviews**, national coverage

- **long instrument**: 70 minutes pre-election, 70 minutes post-election

The Problem: Cost

- **Cost:** $4.2M contract with vendor for

  - 2K in-person interviews, incl 300+300 over-samples of African-Americans and Latinos; 70+70 mins; 125 PSUs (Census tracts)

  - **> $2K/complete**.

- National Science Foundation sole-funder

- NSF total budget for political science is (was?) approx $9-10M pa.

Web?

- web (self-complete) is a much cheaper mode

- potential for savings while dramatically expanding the sample size

- 20K-100K respondents quite feasible with a ANES-sized budget

Web?

- lively debate in the US as to quality of data produced by on-line survey vendors

- much variation within the industry

- "opt-in" panels: how are they recruited? Self-selection? Panel conditioning?

- coverage: poorer, minority households over-represented in households without broadband (or computer)

Web as Mode

- Can we get around extant anxieties re the Web?

- obvious costs-savings from Web as mode

- but utilize an acceptable sampling scheme?

- or even a superior sampling scheme?: in-person interviewing + national coverage = multi-stage area probability samples.

Quality

- for a study like ANES, "opt-in" - even very good "opt-in" - may be not tolerated by scientific community

- high response rate required (currently 60% via in-person)

- some kind of sampling required

Ideas from Dual-Frame Designs

- higher-quality Web vendors moving towards address-based sampling for recruitment

- take this idea even further, with dual frames (a la Vavreck NSF proposal, Doug Rivers)

- exploit **national lists** (each with varying degrees of coverage): union of consumer files, DSF, voter files

- individuals, not households

Dual Frame Design 1

- A: voter files (e.g., Catalist, approx 160M)

- B: consumer file (e.g., Acxiom, Experian): coverage thought to be very high (approx 90%)

- B\A is frame for unregistered

- stratification?

Dual Frame Design 2

- A: consumer file

- B: Postal Service Delivery Sequence File

- sample from A

- sample from B, but take only records unmatched in A

High Response Rate?

- aggressive pre-contact

- lists: may know a lot about R before contact

- incentives: currently spending $2K/case

- mixed modes?

- in-person as "mode of last resort"?

- trade-offs

Three Projects

- "Proof of concept" test on the sampling method

    - varying levels of incentives to assess compliance and costs (can we hit 60%?)

- A pure mode test to investigate differences between face-to-face and self-complete

- A beta-test alongside 2012 ANES