

Challenges of assessing the quality of a prerecruited probability-based panel of Internet users in Germany

GESIS Online Access Panel

Target: German-speaking Internet users 18+

Sampling: RDD

Mode: Mixed (CATI recruitment interview, online questionnaires)

Fieldwork: Feb 2011 - May 2012 AAPOR3 RR (CATI): 17.1%

Incentive: 0, 2, 5, 10, bonus 20 EUR (experimental variation)

% Internet users (unweighted): 71.5% Completion rate online: 55.6%

N for comparison (unweighted): 787 (long interviews)

Reference Surveys

ALLBUS (German GSS) Target: residents private hh 18+

Sampling: register Mode: face-to-face

Fieldwork: May - Nov 2010

RR: 34.4%

Incentive: 0, 10 (interviewer tailoring) Incentive: 20 EUR

% Internet users (weigh.): 67.1% N for comparison (unweigh.): 1869 ESS (German subsample) Target: residents private hh 15+

Sampling: register

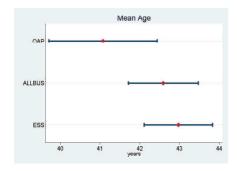
Mode: face-to-face

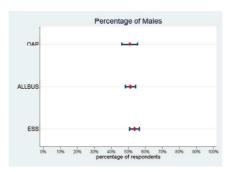
Fieldwork: Sep 2010 - Feb 2011

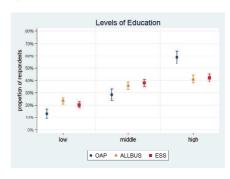
RR: 30.5%

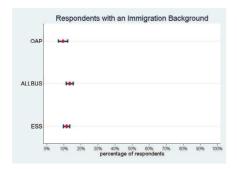
% Internet users (weigh.): 72.5% N for comparison (unweigh.): 2038

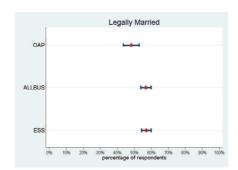
Comparison of Internet users 18+, all graphs based on design-weighted data, 99% confidence intervals

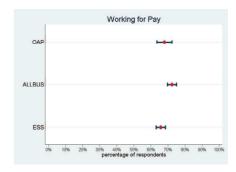


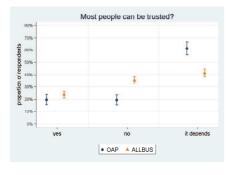


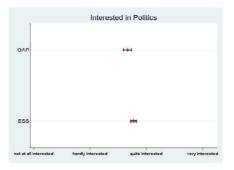


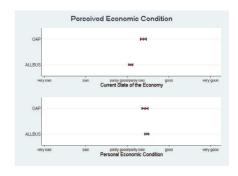












Summary

- Absolute average difference on selected demographics (age, sex, education, marital & employment status, immigration background:
 - OAP & ALLBUS: 5 percentage points
 - OAP & ESS: 5 percentage points
 - ALLBUS & ESS: 2 percentage points
- Statistically significant differences on most demographic & attitudinal variables (i.e. trust, life satisfaction, interest in politics, health status)
- OAP has wider confidence intervals

Discussion

- Despite the number of significant differences, judging by confidence intervals, OAP estimates don't seem "too far off" from those of f2f surveys
- Problem of choosing variables for comparison / alpha-inflation
- What to do with cases where two reference surveys significantly differ (e.g. working for pay, self-rated health, life satisfaction)?
- Are other surveys an appropriate benchmark in the absence of data from official statistics?
- Further research: disentangling nonresponse from mode and context effects