

Note: This presentation accompanied the Keynote Address by Dr. Tourangeau at the 2013 Federal Committee on Statistical Methodology Research Conference, Nov. 4, 2013

Confronting the Challenges of Household Surveys by Mixing Modes

Roger Tourangeau, Westat

Outline

- Why mix modes of data collection?
- Types of MM Designs
- Five Issues with MM Designs
 - Assessing mode effects
 - Impact of choice
 - Impact of order
 - Getting responses via the Internet
 - “Unimode” versus best practices
- Conclusions

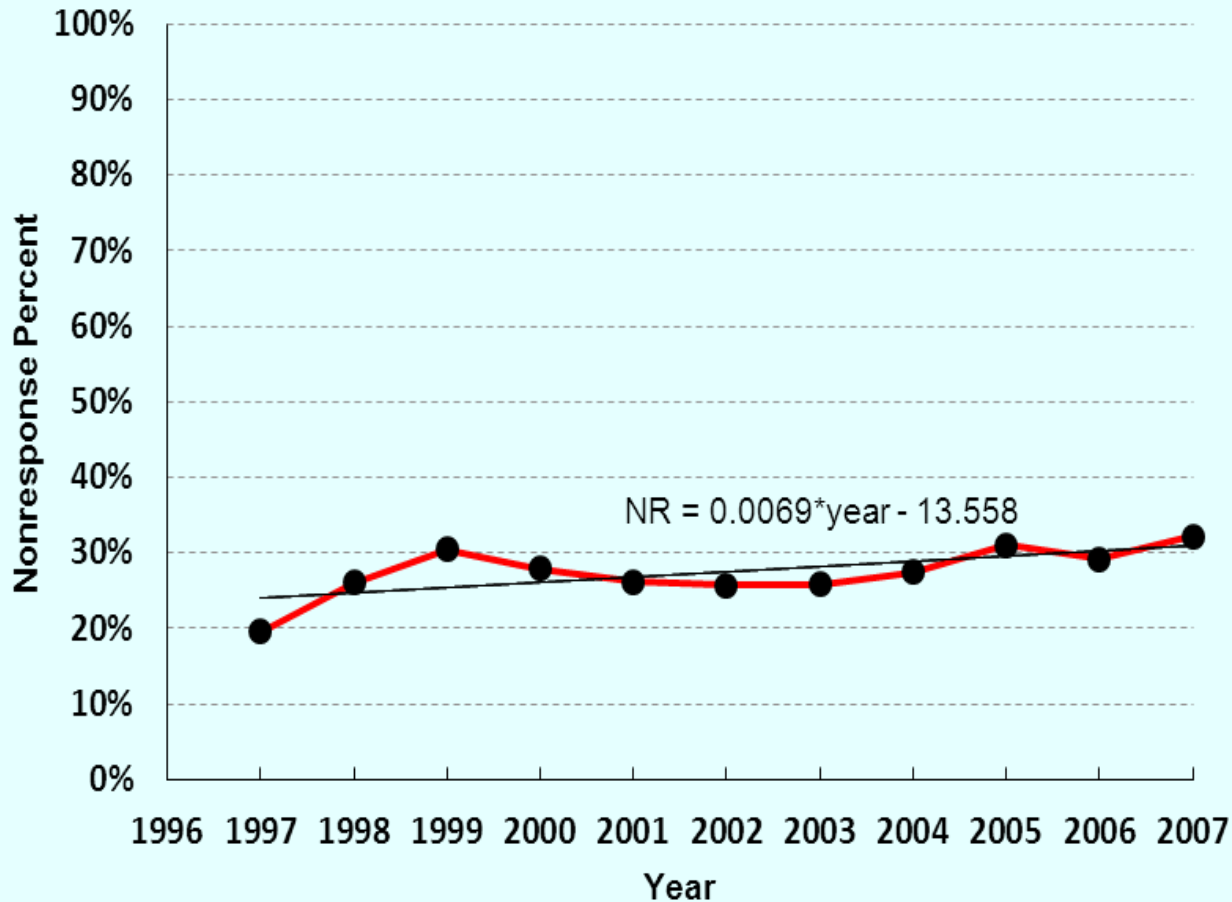
Why Mix Modes?

- Improve coverage
- Increase response rates
- Reduce costs
- Improve measurement?

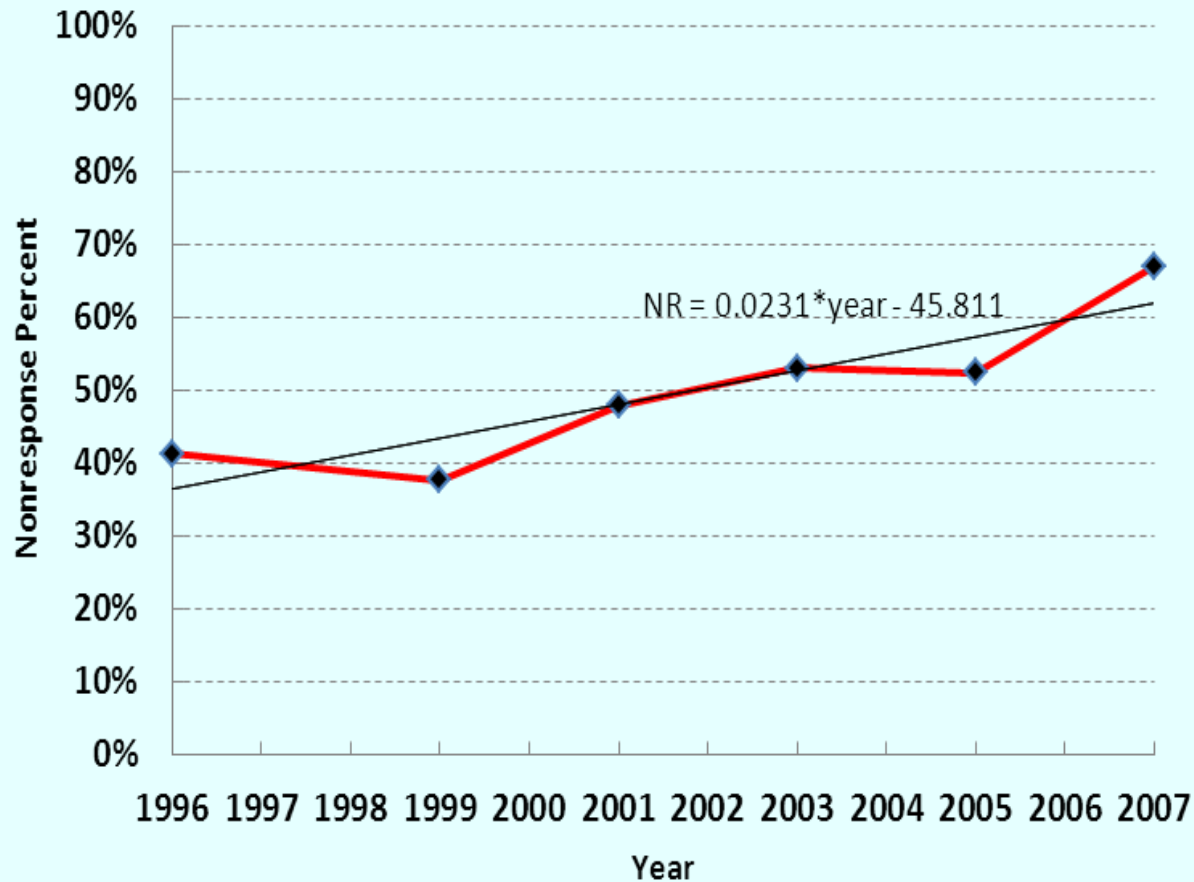
Why Mix Modes—II?

- Costs are rising and response rates falling
- Response rate problems found not just in government, but also in academic and commercial surveys (Tourangeau and Plewes, 2013)
- Polls often have response rates in the single digits

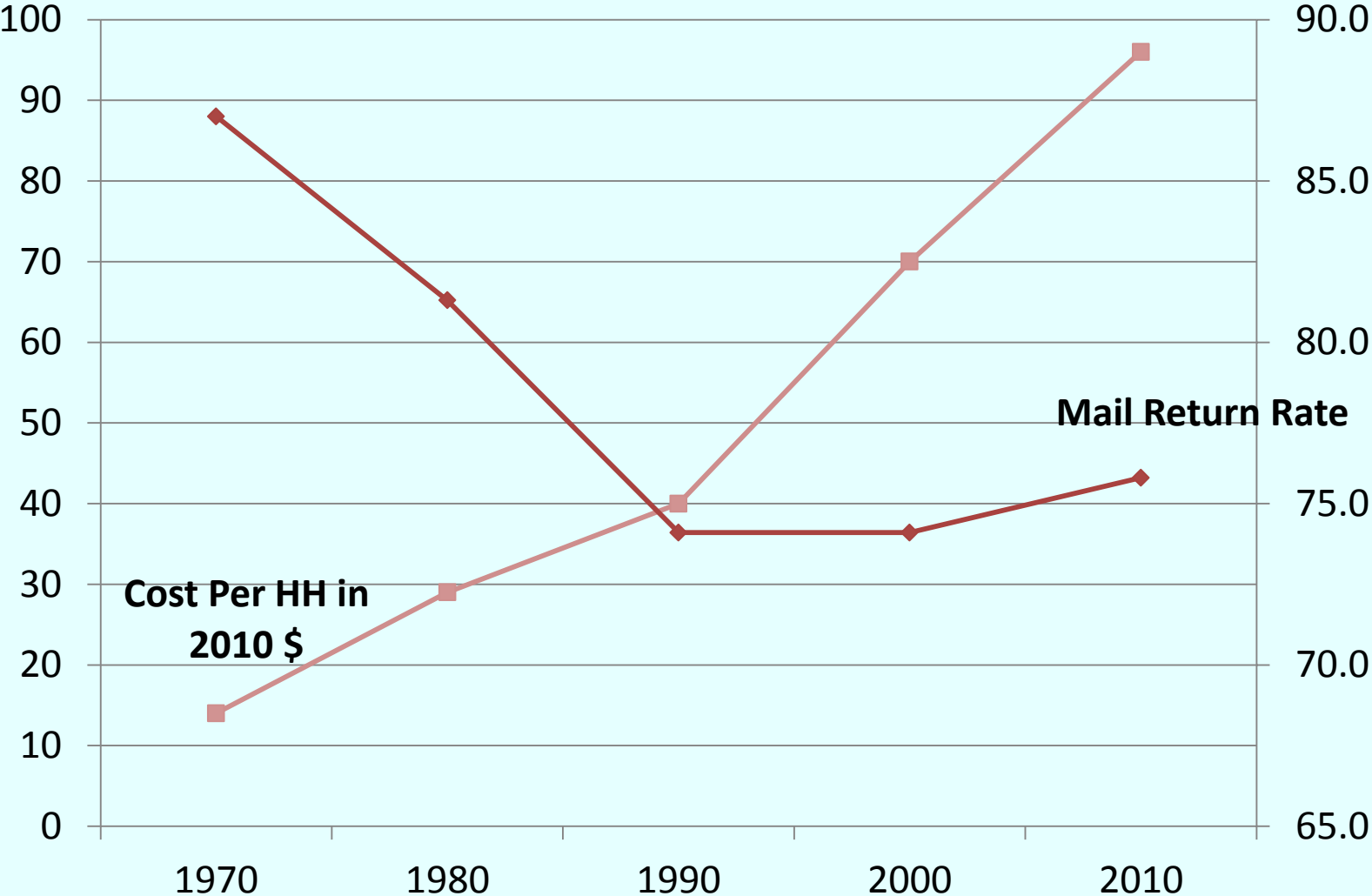
Sample Adult Nonresponse in NHIS (Brick and Williams, 2013)



Sample Adult Nonresponse in NHES (Brick and Williams, 2013)



Costs and Return Rates: Decennial Census



Can Mixing Modes Help?

Four common designs:

1. Cross-sectional mixed modes: Start with cheapest mode and follow up with more expensive modes to reduce NR bias; sometimes concurrent choice
 - American Community Survey (mail, CATI, CAPI)
 - U.S. population census, since 1969 (mail, FTF)
 - Canadian census (mail/Internet, FTF)
2. Different modes for different sections (NSFG)
3. Longitudinal mixed modes: Start with high RR mode, then follow-up waves done more cheaply
 - Current Population Survey—FTF with maximum telephone in Waves 2-4, 6-9)
4. Cross-national surveys: Use different modes in different countries

Five Issues Raised by MM Designs

1. Assessing mode effects: How do you tell whether the mode affects the measurement?
2. Impact of choice: Is it useful to let people choose their method responding?
3. Impact of order: If a sequential design is used, does the order of the options matter?
4. Getting responses via the Internet: How can one get responses in Web surveys? *Can* one get responses via the Web?
5. “Unimode” versus best practices: Should one minimize differences across mode?

Mode Effects

- Is a mode effect a form of error?
- Tourangeau, Conrad, and Couper (2013): No mode effects per se; two components contribute to mode differences
 - Differential non-observation error: Who responds?
 - Differential observation (measurement) error: What answers do they give?
- Estimate from mode A reflects three ingredients:
$$\hat{\theta}_A = \mu_A + b_A + \bar{e}_A$$
 - the mean true score among the respondents to the survey in that mode (μ_A)
 - the systematic impact of the mode of data collection on the answers (b_A)
 - and the average random error under that mode (e_A)
- Mode effect: Net effect of non-observation and measurement error differences by mode

Non-Observation Errors

- Impact of non-observation (non-coverage/non-response) for Mode A

$$\mu_A - \mu = P_{A0}(\mu_{A1} - \mu_{A0}) + \frac{Cov(p_{A1i}, \mu_i)}{\bar{p}_{1A}}$$

- Two components—those excluded entirely and those who might respond with different propensities

Non-Observation Effects--II

- Two modes ought to increase coverage, reduce non-response relative to one mode
- Consider the proportions with no chance of responding

$$\begin{aligned} P_{AB0} &\leq P_{A0} \\ &\leq P_{B0} . \end{aligned}$$

- The hope is that the mixed mode design likely to reduce the size of the excluded population as well as boosting mean response propensities

Observation Errors

- Overall level of measurement error, depends on average measurement effect in both modes:

$$E_m(\hat{\theta}_{AB}) = wb_A + (1-w)b_B$$

- Whether the quantity above is larger or smaller than the measurement error in the corresponding estimate from a single-mode survey depends on
 - the relative magnitudes of errors in two modes,
 - whether they are in same or the opposite directions.

Issue 1: Assessing Mode Effects

- Three common designs
 - Mode experiments
 - With true scores (Tourangeau, Groves, and Redline, 2010; Kreuter, Presser, and Tourangeau, 2008)
 - Without true scores (Tourangeau and Smith, 1996)
 - Mixed mode design compared with single-mode design (Vannieuwenhuyze and Loosveldt, 2012; Vannieuwenhuyze, Loosveldt, Molenberghs, 2010)
 - Experimental design—B with A follow-up versus A only
 - Assume coverage/nonresponse the same in two conditions
 - A respondents would give same answers regardless of whether they had B option
 - First assumption not very plausible (look at ESS data, Dutch experiment)

Vannieuwenhuyze and colleagues (2010, 2012)

- Two components (non-observation and observation error) may offset
- Advocate design that compares B then A versus A only
- Survey about surveys in the Netherlands
- Measurement and nonresponse effects offset each (FTF-only vs. mail with FTF)
 - People who dislike surveys less likely to respond to mail
 - People less like to report negative attitudes in FTF
- Different conclusions about mode effect unless FTF only included in analysis

Tourangeau, Groves, and Redline (2010)

- Sample of Maryland residents who are registered to vote; sample stratified by voter status
- Alternative strategy:
 - Experimentally varied mode
 - Have true scores (from frame on key variables)
- Response rates (overall 34%) reflect incentive (44% vs. 23%) and voter status (41% vs. 26%)

Bias Estimates

Estimated Percentage of Voters in 2006

Subgroup	Entire Sample (Frame Data)	Respondents (Frame Data)	Respondents (Survey Reports)	Bias	
				Nonresponse	Measurement
Overall	43.7 (2689)	57.0 (904)	76.0 (895)	13.3	19.0
Topic					
Politics	42.6 (1346)	58.5 (441)	77.4 (438)	15.9	19.4
Health	44.7 (1343)	55.5 (463)	74.6 (457)	10.8	18.9
Incentive					
\$5	43.4 (1349)	54.8 (591)	75.9 (586)	11.4	21.1
\$0	44.0 (1340)	61.0 (313)	76.0 (309)	17.0	15.0
Mode					
Telephone	43.2 (1020)	57.4 (350)	79.4 (345)	17.0	15.2
Mail	43.9 (1669)	56.7 (554)	73.8 (550)	14.2	22.0

Issue 2: Impact of Choice

- In cross-sectional design: Should you give people a choice? At first blush, it would seem to encourage higher response rates; let people respond by the mode they prefer
- ACS experiment (Griffin, Fischer, and Morgan, 2001) showed lower response rate for mail with Internet option than mail-only
- Medway and Fulton (2012): Examine 19 experiments comparing Web + mail versus mail only
 - Choice lowers RR by 3.8 percent on average
 - Only 10.2 percent use Web

Choice (cont'd)

- Medway and Fulton did not examine two large ACS experiments
- ACS 2011 test (Tancreto et al., 2012): No impact of choice per se

	Overall		Internet	
	Targeted	Non-Targeted	Targeted	Non-Targeted
Mail Only	38.1	29.7	--	--
Prominent Choice	38.3	30.4	9.8	6.3
Non-Prominent Choice	37.6	29.8	3.5	2.0
Push (Regular Schedule)	31.1	19.8	28.6	17.1
Push (Accelerated Schedule)	40.6	29.8	28.1	17.3

Choice (cont'd)

- If you want people to respond by the cheaper mode, why give them a choice?
- Does the order of modes matter?

Issue 3: Sequence of Modes

- Having refused by one mode, are they more likely to refuse in a second mode?
- Lynn (2013): Offering low propensity/low cost mode first may lower overall RR
 - Messer and Dillman (in Washington State): Mail followed by Web higher RR than Web followed by mail
 - Holmberg, Lorenc, and Warner (2010): Web followed by mail nearly as high as mail only, but 65 percent complete by mail
 - Olson, Smyth, and Wood (2012): Web followed by mail no different from mail followed by web
- ACS tests in 2011 suggest that sequence not so crucial (see also Matthews et al., 2012)

Issue 4: Getting Responses via the Web

- Can you get people to respond via the Web?
- The 2011 Canadian experience (Dolson, 2013)
 - Experiment in mail-out areas (80 percent of total population)
 - 75 percent get letter, 25 percent mailed questionnaire

	Letter	Questionnaire
Mail	16.3	50.1
Internet	71.6	25.8
Help Line	0.7	0.8
Nonresponse Follow-Up	9.1	20.0
Nonresponse	2.3	3.5

- The majority of Canadians responded by the Internet (53 + percent)
- Four mailings: Prenotification, reminder letter, mail questionnaire, voice reminder

Getting Responses via the Web (cont'd)

- Mag (2013): Hungarian experience

	Mode
Online	18.6
Paper	16.2
FTF	65.1

- Three lessons
 1. Don't give people a choice
 2. Don't let them procrastinate
 3. Give them an incentive

Issue 5: Unimode versus Best Practices

- Should one try to remove mode effects (measurement effects by design) or attempt to reduce measurement effects within each mode group?
- Two issues relevant here:
 - Is the estimate an overall population estimate or an estimated difference across groups?
 - Does the estimate involve an attitudinal or factual variable?

Overall Estimates

- Combined estimate from MM survey:

$$\begin{aligned}\hat{\theta}_{AB} &= w\hat{\theta}_A + (1-w)\hat{\theta}_B \\ &= w(\mu_A + b_A + \bar{e}_A) + (1-w)(\mu_B + b_B + \bar{e}_B)\end{aligned}$$

- Minimize error in overall estimate by minimizing measurement error (not by maximizing comparability)

Making Comparisons

- Minimize difference in mode effects
 - Use the same mode of data collection to collect within each population (e.g., satisfaction ratings from patients at two hospitals)
 - If more than one mode were used to collect the data (say, a combination of web and mail), then use the same mix of modes in each population (at each hospital)
 - Third, if neither the same mode nor same mix of modes can be used, then use the unimode approach (designing the questions to minimize mode differences)
- Differences often in attitudinal variables

Conclusions

- 1) Mode differences are not themselves a form of error
Instead, mode differences reflect differences in coverage/nonresponse and differences in measurement errors
- 2) Some danger that offering a choice can lower response rates
 - Don't give concurrent choice
 - Push people to the cheapest option
- 3) With a cross-section survey, it seems possible that refusing in one mode increases likelihood of refusal in second mode

Conclusions (cont'd)

- 4) People will respond by the Web—Don't give them a choice, do multiple follow-ups
- 5) Minimize error, not mode differences
 - With factual items and overall estimates, use the best methods in each mode
 - With comparisons, especially comparisons involving attitudinal judgments (such as satisfaction ratings), reduce mode differences
 - Use single mode,
 - Use same mix of modes
 - Use unimode design

Thank you!

RogerTourangeau@Westat.Com