

Telephone, Face-to-Face or Self-Completion Survey: Horses for Courses

Peter Lynn,

Institute for Social and Economic Research (ISER)

University of Essex

Outline

- Survey Data Collection Modes
- Considerations in the Choice of Mode
- Mixed Mode and Multi-Mode Designs
- Conclusions: Some Principles and Guidelines

Some Relevant History

Social Survey Methods

1880s-1920s: FTF interviewing (often by proxy at first)

1920s-1960s: FTF or mail

1960s-1980s: FTF, mail or telephone

1980s onwards: Other possibilities due to computerisation

Modes / Methods

A 'mode' is defined by the **channel(s) of communication** used to perform a particular survey **task**.

Channels

Oral/ aural or visual

Computerised or paper

Private or not (various degrees)

Interviewer present or not

Tasks

Contacting sample member and enlisting co-operation

Delivering questions

Delivering response options

Communicating responses

Survey Method

Is more than a simple “mode”. It is potentially a combination of modes used in different ways for different tasks.

Even for a single survey question, several combinations of modes are possible.

Examples of Mode for a Survey Question I

Interviewer Present

Question	Response options	Response	Description
Aural	Aural	Oral	FTF (no card)
Aural	Visual	Oral	FTF (card)
Visual (CAI)	Visual (CAI)	Written (CAI)	CASI
Visual (paper)	Visual (paper)	Written (paper)	SAQ in-int'w
Aural (CAI)	Aural (CAI)	Written (CAI)	ACASI
Aural (CAI)	Visual (CAI)	Written (CAI)	ACASI

Examples of Mode for a Survey Question II

Interviewer on Phone

Question	Response options	Response	Description
Aural	Aural	Oral	Telephone
Aural	Aural	Written (CAI)	TDE
Aural	Visual	Oral	T'phone w cards
Aural (rec)	Aural (rec)	Oral (CAI)	IVR

Examples of Mode for a Survey Question III

No Interviewer

Question	Response options	Response	Description
Visual (paper)	Visual (paper)	Written (paper)	SAQ (e.g. mail)
Visual (CAI)	Visual (CAI)	Written (CAI)	Web / email
Aur & Vis (CAI)	Aur & Vis (CAI)	Written (CAI)	A-Web

Why So Many Modes?

Big Differences in Key Dimensions:

Why So Many Modes?

Big Differences in Key Dimensions:

- Cost

Why So Many Modes?

Big Differences in Key Dimensions:

- Cost
- Quality (several dimensions)

Why So Many Modes?

Big Differences in Key Dimensions:

- Cost
- Quality (several dimensions)
- Type of questions that can be asked

Why So Many Modes?

Big Differences in Key Dimensions:

- Cost
- Quality (several dimensions)
- Type of questions that can be asked
- Number of questions that can be asked

(See Groves et al 2004)

Costs

Fixed costs broadly similar, other things being equal (which they rarely are).

Marginal costs:

- Lowest for SAQ, esp. mail
- Higher for telephone
- Highest for FTF

Response Rates

Generally:

- Lowest for mail
- Higher for telephone if named/engaged sample; otherwise can be lower than mail
- Highest for FTF

But:

Depends on resources, reminder efforts, etc

Non-Response Bias

SAQ:

- RRs lower amongst less educated / less cognitively able

Non-Response Bias

SAQ:

- RRs lower amongst less educated / less cognitively able

FTF:

- RRs lower amongst persons in FT employment; young adults without children; persons in urban areas

Non-Response Bias

SAQ:

- RRs lower amongst less educated / less cognitively able

FTF:

- RRs lower amongst persons in FT employment; young adults without children; persons in urban areas

Telephone:

- Coverage issues (young / lower SES under-represented if mobiles excluded)

But: Depends on resources, reminder efforts, etc

Example: Mail Survey

Scottish School Leavers Survey;
Sampling frame is list of pupils and their school exam results.

Highest Qualification	Response rate	Selected sample	Responding sample
		%	%
5+ Higher grades	91.1%	18.0	21.4
3-4 Higher grades	85.1%	13.0	14.5
1-2 Higher grades	81.7%	15.0	16.1
5+ Standard grades 1-3	76.4%	8.1	8.1
3-4 Standard grades 1-3	74.1%	9.1	8.8
1-2 Standard grades 1-3	69.1%	14.5	13.1
Standard grades 4-7 only	62.6%	14.4	11.8
No qualifications	59.6%	7.8	6.1
<i>Base</i>		4,542	3,469

Source: Lynn (1996)

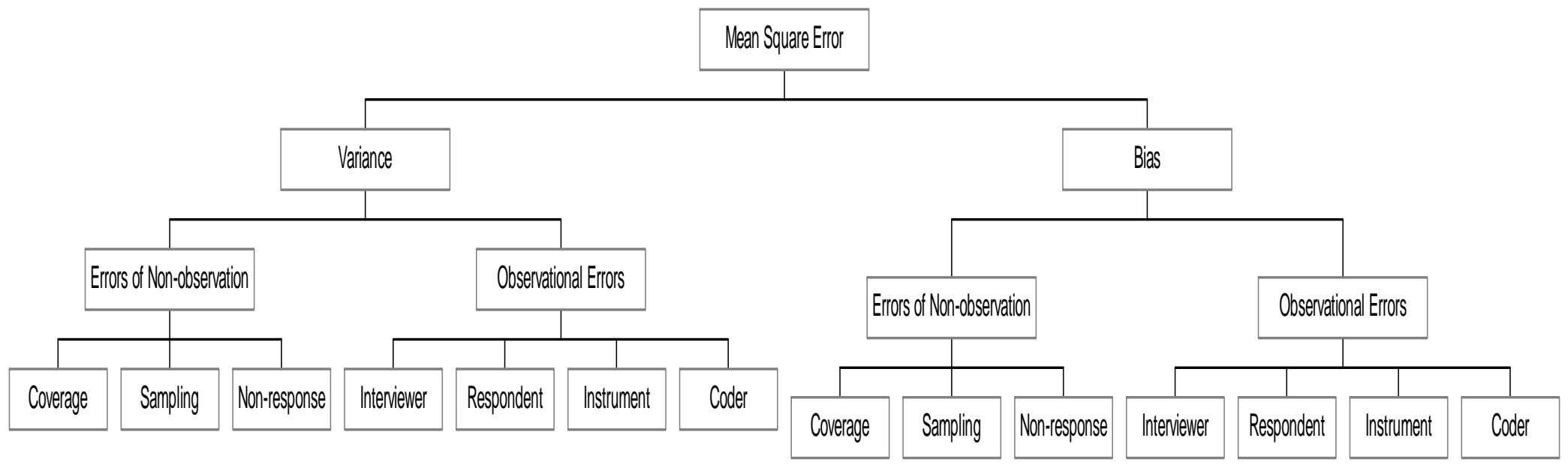
Example: FTF Survey

1996 British Crime Survey; Data from interviewer observation

	Response rate	Selected sample	Responding sample
		%	%
House: detached	82.6%	19.5	21.0
semi-detached	79.6%	30.2	31.3
end terrace	79.2%	7.3	7.6
mid-terrace	77.7%	20.4	20.6
Maisonette	74.9%	1.7	1.7
Flat: converted	72.3%	2.9	2.7
purpose-built	70.3%	11.7	10.7
Rooms/bedsit	75.6%	0.3	0.3
Unable to code	51.2%	6.0	4.0
<i>Base</i>		<i>13,117</i>	<i>10,059</i>

Source: Lynn (1996)

Survey Error



Observational Errors

Error is estimate-specific

$$\begin{aligned}\text{MSE}(y) &= E(y - Y)^2 \\ &= \text{Var}(y) + \text{Bias}^2(y)\end{aligned}$$

Estimates of levels: Mode matters if it biases measurement;

Comparisons: Mode matters only if it biases the comparison;

This is unlikely (but not impossible) unless different modes used for the groups being compared, e.g. surveys, survey rounds, sub-samples, etc.

Mixed-Mode v Multi-Mode Surveys

Multi-mode: different modes for different sets of survey items, but each survey item is collected by the same mode for all sample members.

Mixed mode: the same item might be collected by different modes for different sample members.

Example 1: UKHLS

UK Household Longitudinal Survey

New panel survey: Approx. 40,000 households

Annual interviews with all household members

Multi-topic and multi-disciplinary

Major social science research resource for academia and government

Wave 1 face-to-face; budget constraints demand consideration of mixed/other modes for subsequent waves

UKHLS Example Ctd

Issues:

Subsequent wave(s) entirely by a mode other than f2f:

- Low(er)/(more) biased response? And differential measurement error?

Mixed mode within wave:

- Sequential could maintain response rate while even reducing non-response bias and/or reducing costs
- But, differential measurement error??
- Which sequence of modes? Optimal strategies?

Example 2: ESS

European Social Survey

Repeating cross-sectional survey

Approx. 2,000 persons per country per round

Face-to-face interviews

Rounds every two years: 2002-03, 2004-05, 2006-07, ...

Approx. 23 countries per round; 32 countries at least once

Major social science research resource for academia

ESS Example Ctd

Not considering changing the mode of the entire survey

To allow a country to carry out ESS entirely by a mode other than f2f?

- Differential coverage/sampling/non-response/measurement error relative to other countries? (for between-country comparisons)
- Differential error between modes within country? (for between-years comparisons)

To allow a country to use mixed modes?

- Same Q as above with respect to modes other than f2f
- Differential measurement error between modes within the country?

Illustration of Mode Effect on a Comparison

Simple difference in proportions: $P1-P2$ (two countries)

True values: $P1=0.58$; $P2=0.44$; $P1-P2=0.14$

Illustration of Mode Effect on a Comparison

Simple difference in proportions: P1-P2 (two countries)

True values: P1=0.58; P2=0.44; P1-P2=0.14

Scenario 1: Mode Effect Constant Across Domains

Produces $P^* = 0.9P_{f2f}$; $P^* = P_{web}$

Country 1: 100% F2f; Country 2: 20% Web

Illustration of Mode Effect on a Comparison

Simple difference in proportions: $P1-P2$ (two countries)

True values: $P1=0.58$; $P2=0.44$; $P1-P2=0.14$

Scenario 1: Mode Effect Constant Across Domains

Produces $P^*=0.9P_{f2f}$; $P^*=P_{web}$

Country 1: 100% F2f; Country 2: 20% Web

Find $P1^*=0.52$; $P2^*=0.405$; $P1^*-P2^*=0.115$

(cf. 0.126 if both f2f, 0.14 if both web)

Illustration of Mode Effect II

Recall $P1 - P2 = 0.14$

Scenario 2: Mode Effect Differs Across Domains

$P1^* = 0.9P1$ f2f; $P1^* = P1$ web

$P2^* = P2$ f2f; $P2^* = P2$ web (less error than scenario 1)

Illustration of Mode Effect II

Recall $P1-P2=0.14$

Scenario 2: Mode Effect Differs Across Domains

$P1^*=0.9P1$ f2f; $P1^*=P1$ web

$P2^*=P2$ f2f; $P2^*=P2$ web (less error than scenario 1)

Country 1: 100% F2f; Country 2: 20% Web (as scen.1)

Find $P1^*=0.522$; $P2^*=0.44$; $P1^*-P2^*=0.082$

Illustration of Mode Effect III

Summary:

In Scenario 2, for country 1, mode effect was same as scenario 1. For country 2, no error in scenario 2 (= no mode effect)

But error in estimate of P1-P2 is larger in scenario 2

True P1-P2=0.140

Scenario 1 estimate 0.115

Scenario 2 estimate 0.082

Conclusion from Illustration

Mode effect in between-country comparisons depends on:

- a) Extent of mode effects
- b) Extent to which these effects differ between countries
- c) Extent to which mode mix differs between countries

Conceptual Framework for Mode Effects

(Jäckle, Roberts & Lynn, 2007)

Start from model of question response process (Tourangeau *et al*): Comprehension – Retrieval - Judgement – Response;

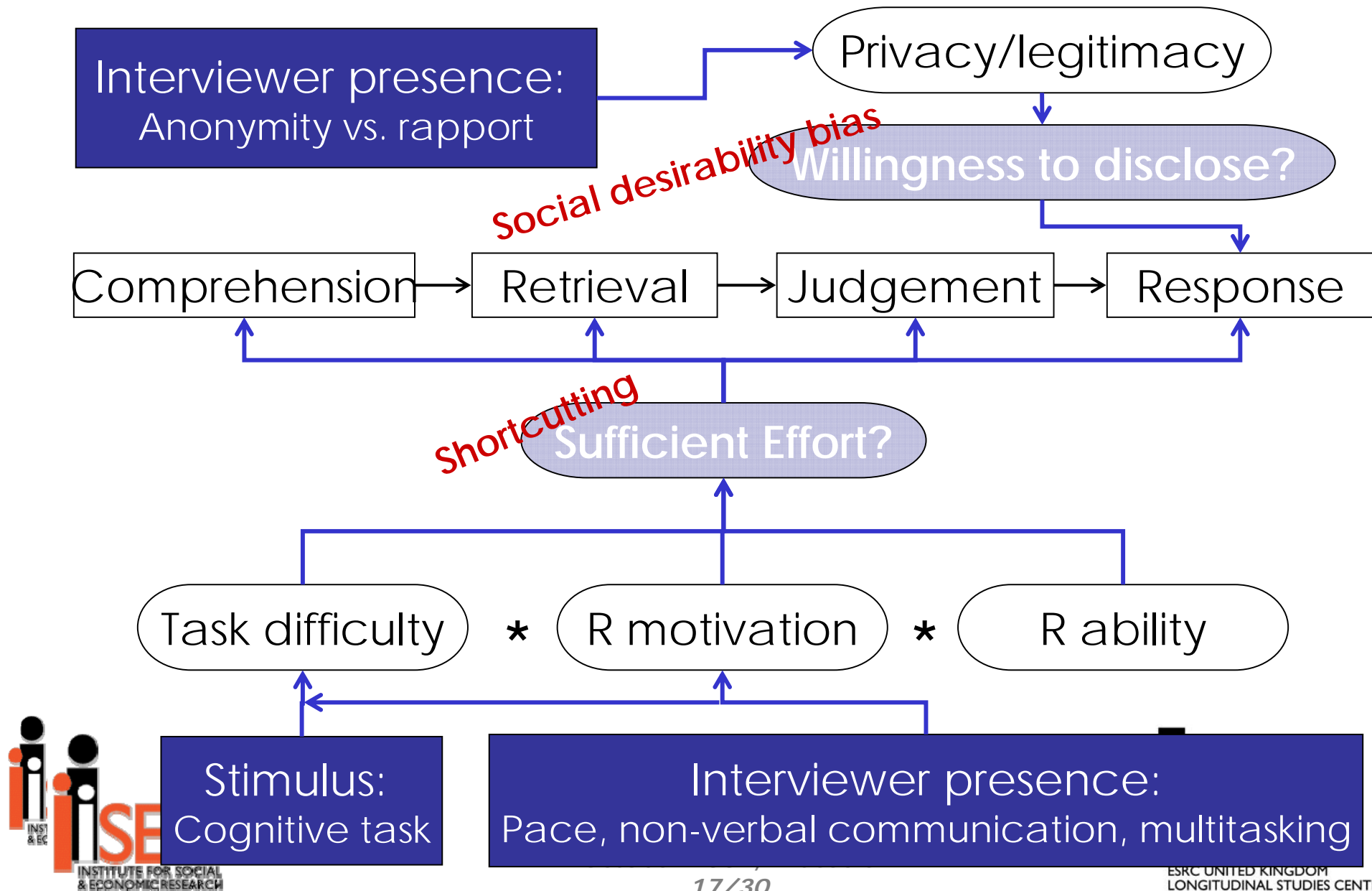
Identify how respondent behaviour may influence each of these stages;

Consider factors which may influence these behaviours;

And survey design features which may influence those factors

Focus should be on how modes differ in terms of these features

Causes of Mode Effects on Measurement



Satisficing (Shortcutting)

Failure of respondent to execute response task optimally

Result of interaction between task difficulty, respondent motivation and respondent ability (Krosnick 1991)

Can take various forms, e.g.:

- Non-differentiation
- Incomplete answers
- Acquiescence
- Primacy / recency effects

Social Desirability Bias

Respondent reports a more socially desirable answer in order to portray themselves in a good light (Demaio 1984)

Can be conscious or sub-conscious

May be influenced by perceived privacy, perceived legitimacy, and rapport/trust

Examples of Mode Differences in Satisficing

Example 1: Holbrook *et al* (2003, POQ)

Telephone vs. F2F; Randomised experiments; USA;

3 different years; 2 different survey organisations

Election surveys: opinion and political behaviour questions

Holbrook *et al* II

Acquiescence:

- More common on telephone on all three studies
- Particularly for less educated respondents
- Size of effect modest

e.g. 2000 NES:

Mean 'agree'	FTF	Telephone
Total sample	.31	.33
High education	.34	.35
Low education	.29	.32

Holbrook *et al* III

No opinion responding:

- More common on telephone on all three studies
- Particularly for less educated respondents
- Size of effect medium

e.g. 1982 NES:

Mean “no opinion”	F2F	Telephone
Total sample	.17	.24
High education	.14	.18
Low education	.32	.46

Holbrook *et al* IV

Non-differentiation:

- More common on telephone on all three studies
- Particularly for less educated respondents
- Size of effect modest

e.g. 1982 NES:

ND level	F2F	Telephone
Total sample	.37	.41
High education	.37	.41
Low education	.38	.44

Satisficing Example 2: ESS/Gallup-Europe Studies

Phase 1: Hall test in Hungary, 2004

Telephone vs. F2F vs. Web vs. Paper SC

Randomised allocation subsequent to agreement to participate

Complete control of coverage, sampling and non-response errors

But arguably low on realism of survey situation

Plus, telephone vs. F2F comparison confounded visual stimuli (show cards) with presence/absence of interviewer

ESS/ Gallup-Europe Studies II

Phase 2: Field experiment in Hungary and Portugal, 2005

Telephone vs. F2F without showcards vs. F2F with showcards

Randomised allocation prior to fieldwork

High on realism of survey situation

Complete control of coverage and sampling errors

But potential confounding of non-response errors

Three treatment groups enable separation of visual stimuli effects from presence of interviewer effects

ESS/ Gallup-Europe Studies III

Non-differentiation, Phase 1:

- Slightly more common on telephone than F2F
- More common in SC modes than either interview mode
- Size of effect modest

	F2F	Telephone	Paper SC	Web
ND level	.32	.34	.40	.39

ESS/ Gallup-Europe Studies IV

Non-differentiation, Phase 2:

- More common on telephone than F2F
- F2F, more common with visual stimuli
- Size of effect modest

	F2F w SC	F2F, no SC	Telephone
ND level	.45	.41	.44

Examples of Mode Differences in Social Desirability Bias

Holbrook *et al* :

- Socially desirable answers more common on phone than F2F

ND level	F2F	Telephone
Total sample	.37	.41
High education	.37	.41
Low education	.38	.44

ESS/ Gallup-Europe Phase 2:

- Socially desirable answers more common on phone than F2F

Conclusions Re. Mode Effects on Measurement I

Mode effects on measurement should be taken seriously

However, knowledge of their nature is surprisingly limited

Importance of careful experimental design highlighted by Holbrook et al

Few studies have truly isolated relevant effects

Studies largely limited to certain survey topics

No studies have addressed the complexities of mixed-mode designs

Mode effects on bias more important the larger the sample size
(smaller the variance)

Conclusions: Mode Effects on Measurement II

Some consistent evidence of mode differences in:

- Satisficing/ shortcutting
- Social desirability bias

Many studies that appear to show mode effects are on survey questions/ topics likely to be relatively susceptible to one or other of these phenomena

Mode effects likely to be less important for topics and populations that are less susceptible to satisficing or social desirability bias

Some consistent evidence that SC modes are better for revealing sensitive data

Conclusions: Mode Effects on Coverage and Non-Response Bias

Also important

In general, less scope for bias with FTF

- both because RRs generally higher overall, and because less differential w.r.t. literacy / cognitive ability
- but cf. availability

Mode effects on bias more important the larger the sample size (smaller the variance)

Conclusions: Guidance

Choice of mode depends on balance of factors, many of which will be survey-specific, including:

- Key survey estimates (variable(s), statistic, subgroups)
- Likely non-response bias in those estimates in each mode;
- Likely measurement error in those estimates in each mode;
- Survey costs in each mode

Combining modes can bring advantages

- e.g. cost reduction and response rate increases simultaneously;

But can also bring disadvantages:

- e.g. potential for biased comparisons

If you want to compare your results with another survey, use the same mode as that survey (unless you are really sure that your survey measures are insensitive to mode effects in either non-response error or measurement error)

Similarly, avoid mixed mode designs (unless)

But multi- mode designs can have advantages, e.g. if you want to collect some items that are known to have better error properties in SAQ modes and other that are better in FTF or telephone.

Design to the strengths of each mode, e.g. do not try to ask questions that require visual stimuli by telephone!

References

- Groves R M et al (2004) *Survey Methodology*, New York and Chichester: Wiley. Especially Chapter 5: 'Methods of Data Collection'
- Holbrook A L, Green M C and Krosnick J A (2003) Telephone vs. face-to-face interviewing of national probability samples with long questionnaires: comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly*, 67, 79-125.
- Krosnick J (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236
- Lynn P (1996) Weighting for non-response, in *Survey and Statistical Computing 1996*, Chesham: Association for Statistical Computing.
- Lynn P (2007) Sequential mixed mode survey designs. Presented at International Workshop on Household Survey NonResponse, Southampton, September, and in preparation for journal submission.
- Jäckle A, Roberts C and Lynn P (2006) Telephone versus Face-to-Face Interviewing: Mode Effects on Data Quality and Likely Causes (Report on Phase II of the ESS-Gallup Mixed Mode Methodology Project). Institute for Social and Economic Research Working Paper 2006-41. Colchester: University of Essex. <http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2006-41.pdf>
- Jäckle A, Roberts C and Lynn P (2007) Assessing the Effect of Data Collection Mode on Measurement. Presented at International Statistical Institute Session, Lisbon, August, and in preparation for journal submission.

**Telephone, Face-to-Face or Self-Completion
Survey:
Horses for Courses**

Peter Lynn,

Institute for Social and Economic Research (ISER)

University of Essex