# Framework for Data Collection and Analysis

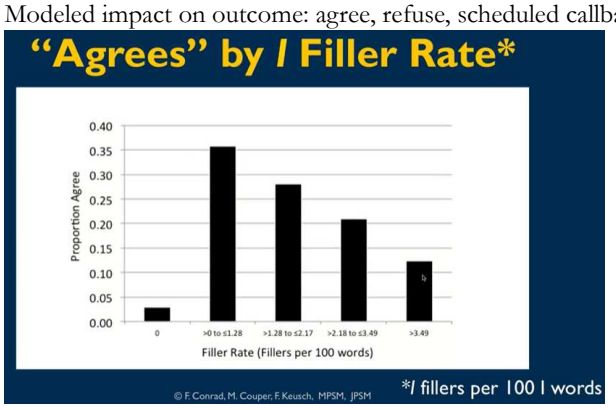
## Week 3 Interviewers and Interviewing

### 3.1 Interviewers impact on participation and sampling

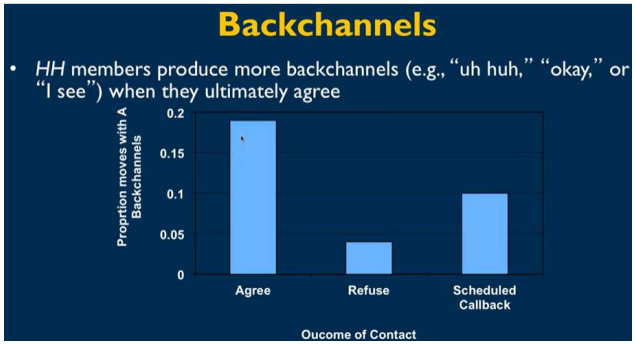
*Interviewer roles, Obtaining Interviews*

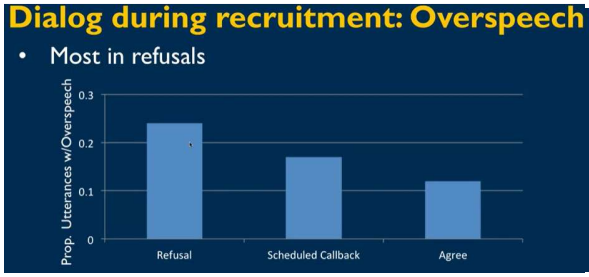
• Survey Interviewers play many roles
- each can affect one or more sources of error
- sampling and coverage error
** locate and contact sample households
** determine eligibility
** Sample within household sampling: R selection
- non response error
** contact sample households
** gain cooperation, including persuading reluctant Rs
- measurement error
** they ask the survey questions (e.g. ask differently, unable to clarify misunderstanding etc)
** probe for incomplete answers, clarify questions
** record answers
- interviewers also have an enormous impact on costs and timeliness (usually the single biggest component of costs)

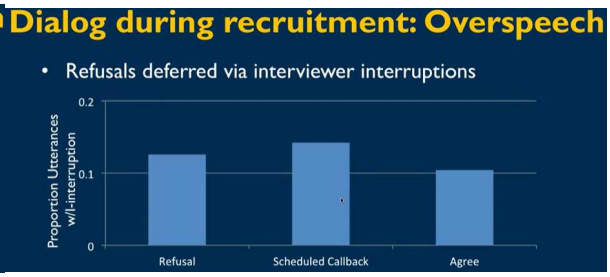
• Interviewer Effects on NonResponse
- recruitment and invitation
** some interviewers are more successful than others in obtaining interviews (lifetime RR)
** one explanation may lie in how they speak and interact with Rs
➔ Effects of Interviewer speech and interaction with R on participation (example 1)
Interviewers vary in their success obtaining interviews + how interviewers speak can affect success + Conrad (2013) coded speech behaviors of interviewers in 1380 telephone survey introductions
Disfluencies / backchannels (e.g. uh huh) by HH member / overspeech and interruptions
Modeled impact on outcome: agree, refuse, scheduled callback




Agreement rate higher when having a bit of filler than not having filler at all cuz it sounds robotic and unnatural if there is no filler words at all






➔ Effects of Interviewer speech and interaction with R on participation (example 2)
Schaeffer, Garbarski, Freese & Maynard (2013)
Selected 257 pairs of cases, with same propensities to participate (based on prior wave* data)
Outcome for one case in pair was acceptance and one declination
Classified both I and R speech using conversational analytic techniques

| | Outcome | |
|---|---|---|
| Declination or acceptance occurs... | Declination | Acceptance |
| During opening | 57 | 0 |
| During study description | 95 | 4 |
| After request for participation | 105 | 253 |

Declination happens in various stages (altho numbers not evenly distributed) while acceptance mostly happened only after request for participation… shows **harder to elicit acceptance and requires explicit request**

| Action | Feature of action | Odds ratio | P-value |
|---|---|---|---|
| Greeting token | Hello (vs. hi) | 1.44 | 0.06 |
| Request to speak to sample member | Very polite (vs. not) | 1.60 | 0.07 |
| Order of interviewer identification and request to speak | Interviewer identification first (vs. second) | 1.80 | 0.04 |

Odd ratio increasing with more polite and intimate (first person instead of second) expressions, so higher likelihood of acceptance

| Feature | Odds ratio | P-value |
|---|---|---|
| Sample member asks at least one question | 2.07 | 0.00 |
| Sample member initiates statement or question about the length of the interview | 5.44 | 0.00 |
| Sample member asks at least one wh-type question | 0.59 | 0.02 |
| Sample member asks at least one wh-type question before the request to participate | 0.35 | 0.00 |
| Sample member asks at least one wh-type question after the request to participate | 1.64 | 0.14 |

General questions from sample member usually mean interest in survey, so associated with higher acceptance probability but wh-questions from sample member usually mean skepticism, so is associated with lower acceptance probability but this doesn't hold after the request to participate stage cuz by this time the sample member would have already decided whether to take part in the study or not, so doesn't matter as much

*Respondent Selection, within Household sampling*

• Respondent Selection
➔ Most popular 3 methods….
- Kish Method (gold standard; oldest; equal probability random selection; most intrusive cuz it requires respondents or household members to reveal considerable information about the household)
- Birthday methods (simpler but doesn't show probability of selection cuz X require # of people in the household to be indicated by the respondent)
- Rizzo, Brick & Park Hybrid Method (either Kish or Birthday method with some sort of coin toss

techniques)

➔ No sampling – interview the person who opens door or answers the phone (how might this bias the sample?)
➔ Knowledgeable respondent (use for household or family level surveys e.g. to create roster)
➔ Quota methods (e.g. designate some HHs for female R, others for male R)
➔ Probability (quasi-prob) methods: Kish selection table, next and last birthday methods and other

• Kish (1949) Roster Method
- List all members of housing unit/household, then check eligible members
- number eligible males from oldest to youngest then eligible females from oldest to youngest then eligible females from oldest to youngest
- use selection table to select respondent (8 selection tables randomly allocated to each household)

Example of one of the 8 selection tables

| Selection Table D | |
|---|---|
| **If the number of eligible persons is:** | **Interview the person numbered:** |
| 1 | 1 |
| 2 | 2 |
| 3 | 2 |
| 4 or more | 3 |

- advantage: most precise selection method (each HH member has equal and known probability of selection)
- disadvantages
  ** requires gender and age of all eligible HH members
  ** intrusive, time consuming
  ** subject to interviewer error cuz of its complexity
- can be automated (and extended to oversample groups of interest such as teenagers, males etc.)

• Birthday Methods
- Last Birthday: May I speak to the person in the household who is 18 years of age or older and whose birthday was most recent?
- Next Birthday: Interviewer asks to speak to the adult member of the household who has the next b day
- relatively easy to administer
- equal probability of selection within household
- less intrusive and brief; low effort HH member
- generates that sample is random

BUT….
- the exact chance of selection is unknown hence QUASI-probability based
- subject to reporting error; difficult to verify
- Biemer & Lyberg(2003) recommend against use in rigorous, scientific studies

• Kish v.s. B day Methods
- Binson, Canchola & Catania (2000) compared dropouts in Kish and b day methods in RDD study
- Overall, Kish performed worst (more dropouts) and NB performed best
  % Dropout: LB(34.1), NB(32), Kish(37)
- But much of the effect occurred before R could know what type of screener questions would be asked
  % Dropout: LB(16.1), NB(12.2), Kish(16.6)
- Authors attribute to interviewers "telegraphing" what will follow as if they are somehow indicating that Kish method will be more burdensome, so Rs terminate before the method was administered

• Hybrid R-Selection Method
- Rizzo, Brick & Park (2004) propose method that is minimally invasive and usually explicit about p(selection)
- Ask R in screener how many adults in household
  If N = 1, screener R selected; end of process
  If N > 1, randomly sample with p(selection) = 1/N
    If N = 2 and screener R not selected, then other selected
    If N > 2 and screener R not selected, then use LB method (or Kish when b days are unknown) to select from remaining adults

- Rizzo tested in HINTS (RDD survey of where American adults acquire health info)
  Male and female proportions close to what is expected from CPS

• Response errors affect all methods
- Kish requires enumeration of HH which may not be truthful
- B day method reveal purpose of questioning more directly, more subject to manipulation if e.g. HH member does not want to participate
• Full listings of households require explaining the rationale of R-selection process to household member
- may affect overall response rates in some cases
• Different organizational structures regrading preferences for different methods; different client preferences

*Proxy Responding*

• Proxy Respondents
- several advantages to collecting information proxy
  Less time and therefore expense to use proxy reports (e.g. CPS reduces cost by about 12%)
  Use of proxies result in higher response rates
- general belief that self-reports are better than proxy reports
  proxies' response process thought to rely more on general knowledge than actual events
- response differences
  Proxies more likely to use estimation than recall
  Dating information may not be accurate
  Self respondents more likely to use direct retrieval; answers more likely based on actual events than general knowledge
  - motivational differences
    Proxy reporters may be less motivated
    Proxy responses may be less subject to social desirability issues
  - proxy reports likely affected by level of communication between selected R and proxy

• Moore (1988) reviewed literature and found no consistent evidence that self-reports better quality than proxy reports, contrary to assumption that proxies less accurate
- self-selection problems in literature
  Ideally, the contacted HH member is randomly assigned to be self or proxy reporter
  In practice, if selected R not available, another HH member serves as proxy
  Finding in health studies that proxies report fewer health events for the "target R" than self (e.g. doctor visits) may not reflect lower quality data from proxies
  Could reflect self-selection bias; those at home likely to be in worse health and so when report for self, report more health events
- lack of validation data in the studies, so conclusion that no difference is tentative

• If observers asked to describe actors' behavior, rely more on underlying dispositions (traits) than situations (events), but opposite for own behavior
• Schwarz and Wellens (1997) tested these ideas in self proxy context
  - self reports based on memory for events while proxy reports draw more on general knowledge
  - self reports should reflect inconsistencies in behavior – departures from usual routine – more than proxy reports
  - 24 couples asked to report either their own or their partner's typical behavioral frequency and the highest and lowest frequency of the respective behavior during the preceding month

• Schwarz and Wellens (continued):
  – Self reports reflect greater deviation from "typical" behavioral frequencies than do proxy reports

| | Self | | Proxy | |
|---|---|---|---|---|
| | Highest | Lowest | Highest | Lowest |
| Library | +3.41 | -4.32 | +2.13 | -1.91 |
| Snacks | +7.13 | -6.14 | +4.67 | -4.93 |
| Being Late | +2.61 | -2.98 | +1.82 | -1.61 |

mean difference between typical and highest or lowest behavioral frequency

➔ Proxies not aware of the exceptions in the routine, so have to rely on routines or general knowledge instead of

specific traits

So to summarize….

• I-s can affect, for better or worse, all sources of survey error
  - sampling, coverage, non-response and measurement
• Recruitment (I-s vary in success; manner of speech may matter)
• Within household sampling (methods vary in explicitness about p(selection), intrusiveness, efficiency)
• self-proxy issues
  - response processes differ between self and proxy Rs
  - cost, non response, and measurement error trade-offs

### 3.2 Interviewing Technique

*Standardizing Debate: Wording v.s. Meaning*

Standardized Interviewing (wording)
• To minimize interviewer related error (rho_int)
  - ask Qs slowly and exactly as they are worded
  - If R says anything that is not an acceptable response, use neutral probes such as "Let me repeat the question…"
    Intended to make sure any substantive info is provided to all Rs
  - Ask Qs the order they appear in questionnaire
    Q order must be standardized for responses to particular Qs to be comparable
  - Ask every Q even if R has previously volunteered relevant info
    If actually ask Q, R might answer differently
  - Maintain a professional stance: avoid statements that R might interpret as criticism, surprise, approval or disapproval

• Examples of question-wording alterations
  - original: where do you get most of your news about current events in this country – from the radio, the newspapers, TV or talking to people?
  - interviewer revision: where do you get most of your news about current events?
  - interviewer revision: where do you get most of your news about current events in this country – from the radio the newspapers, TV or talking to people? That is, which one do you rely on most?

• Examples of Non Directive probes Initiated by I-s
  - Is there anything else?
  - yes, I see or uh-huh in an expectant manner and followed by a pause
  - could I read back what I have take down to be sure I have exactly what you wanted to say?
  - what do you mean by that?
  - could you be a little bit more specific?
  - which comes closest to the way you feel?

• Examples of non-directive probes after respondent requests clarification
  - I'm sorry but I don't have that info
  - it's important that the question be answered as best you can, in terms of the way it's stated. Perhaps I could read it to you again
  - whatever it means to you

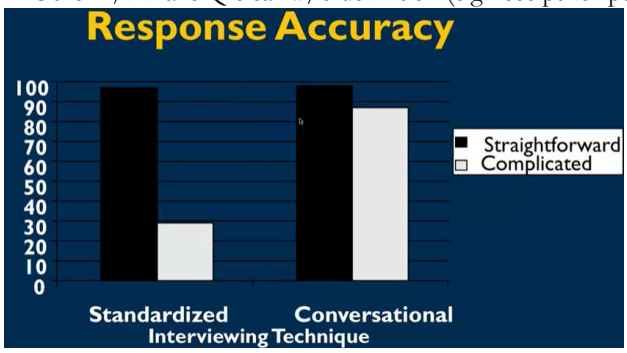
• Criticism of Standardized wording
  - chief criticism is that standardized wording <u>prevents conversational grounding</u> (e.g. Clark 1996)
  - Grounding
    Speaker and listener exchange turns until agree understand each other well enough for current task
    If allowed in standardized interviews, would lead to variable wording across interviews
  - standardized meaning (uniform interpretation) – not just wording – should be the goal ➔ will promote **valid**, not just reliable, responses
  - Suchman and Jordan (1990), Schaeffer (1991) and others note that the survey interviewing style conflicts with "normal conversational" practices
    Only one actor (I-s) determine topic of conversation
    And is not able to engage in conversational processes like grounding

R is not free to ask about anything beyond the questionnaire
The interviewer determines the ending of the conversation
- crux of issue: even if "normal conversation" sometimes promotes more accurate understanding and answers, does this outweigh increased variation in wording

• When is Standardized wording sufficient?
- when all Rs interpret question as intended
Everyone would agree that if a married couple with two young children live in a house, and no one else ever sleeps there, then four people live there
- But what about situations in which Rs do not know what to include or exclude?
What if one child is away at college?
What if a nanny sleeps in the house most nights?

*Different approaches to standardized Interviewing*

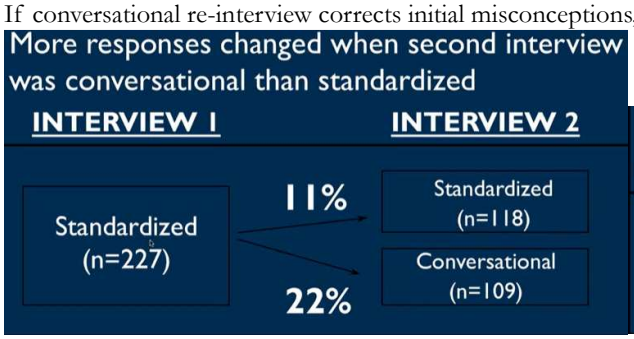
Standardized v.s. Conversational Interviews
• Schober & Conad (1997): interviewers asked laboratory Rs 12 US gov't surveys Qs using one of two techniques:
- standardized procedure where interviewer was not allowed to deviate from the script at all
- flexible* procedure
Encouraged R to ask questions if did not understand something and
I-s able to clarify world meaning e.g. whether someone who baby sits for >1 family has one job or > one job, verbatim or paraphrased (* later called conversational interviewing)
- Rs answer fictional scenarios for which true value known
1/2 scenarios make Q ambiguous w/o definition (e.g. Q asks if any purchases of furniture and scenario is receipt for buying floor lamp)
Other 1/2 make Q clear w/o definition (e.g. receipt for purchase of end table)




**Minutes per Interview**

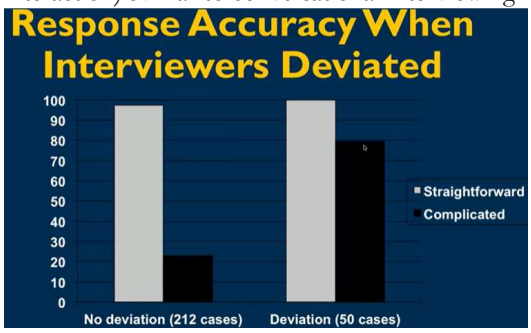
| | Standardized | Conversat'al |
|---|---|---|
| Median | 3.41 | 11.47 |
| Min | 2.48 | 6.10 |
| Max | 5.99 | 35.44 |

Conversational technique definitely leads to increased response accuracy especially for complicated scenarios but comes with cost (longer time required… following graph)
• Tradeoffs between standardized and conversational approaches
- conversational Interviewing can produce greater response accuracy than Standardized Interviewing
- But also produces longer interviews (may result in fewer cases for fixed cost)
- how frequent are complicated situations requiring clarification and how much error is tolerable?
- researchers need to determine for each study
• How often is unscripted clarification helpful?
- Conrad & Schober (2000) assessed this through by re-interviewing Rs in a national sample who had initially been interviewed with standardized techniques
1/2 Rs re-interviewed in a standardized interview, 1/2 in conversational interview
If conversational re-interview corrects initial misconceptions, these Rs should change their answers more




| Interview 1 (Standardized) | Interview 2 | |
|---|---|---|
| 5 mins | 5 mins | Standardized |
| | 9 mins | Conversational |

• How standardized are standardized interviews?
- schober, Conrad & fricker (2004) instructed interviewers to do what they ordinarily do ("natural" condition)
Same Qs and scenarios as 1997 laboratory study
Response accuracy similar to what was observed in strictly standardized interviewing
Except… when examined separately for cases in which interviewers departed from script, accuracy (and I-R interaction) similar to conversational interviewing




➔ I-s almost spontaneously providing clarification when they are able to even though their training would require them to use non-directive or neutral probes when a respondent asks what the question is intended to mean

• Summary
- conversational interviews can substantially improve response accuracy compared to standardized interviews
When questions are ambiguous
Cuz conversational interviewers can ground the meaning of the terms and phrases in the questions
- grounding takes time so duration of conversational interviews greater than strictly standardized interviews
- when standardized interviewers clarify meaning, benefits comparable to when explicitly instructed to do so

*Personal v.s. Formal Style, I-R Rapport*

• Which interviewing approach produces data of higher quality: business like or interpersonally-engaged?
• Dijkstra (1987) explored formal and personal interviewing styles
- not equivalent to conversational v.s. standardized distinction
• Both approaches require learning basics of interviewing e.g. reading questions, probing
- formal style includes unlearning person-oriented behavior
- personal style includes training in person-oriented behavior (e.g. that is nice for you or I have similar feelings)
• Is randomly assigned to one style or the other
• behavior coded interaction

Personal v.s. Formal Interviewing: Accuracy and Social Desirability
• Dijkstra (1987) examined impact of interviewing style on accuracy and social desirability
- motivation v.s. ingratiation: will personal style lead Rs to answer as accurately as possible or will it lead them to answer in ways they believe will please I-s
• accuracy: asked Rs draw map of their town
- coded for proportional correctness and inclusion of key landmarks, among other attributes
• social desirability: reduced version of Marlowe-Crowne impression management scale
- higher scores indicate more social desirability

**Accuracy of Drawings**
**Higher scores indicates more accurate (max=37)**

| | Respondent Gender | |
|---|---|---|
| Interviewing Style | Male | Female |
| Personal | 16.3 | 14.5 |
| Formal | 14.6 | 12.9 |

**Social Desirability**
**Higher scores indicate more social desirability**

| | Respondent Gender | |
|---|---|---|
| Interviewing Style | Male | Female |
| Personal | 11.4 | 12.5 |
| Formal | 12.1 | 13.0 |

Rapport between I and R
• Sheatsley (1951): rapport = demographic similarity
  - opposite of social distance
• Long assumed to promote honest and thoughtful survey responding
  - Weiss (1968) found lower validity (record check) when rapport was higher
  - Goudy & Potter (1976) find no relation between rapport and Interviewer performance
    Rapport measured on several-item scale e.g. During the interview how ill at ease did the respondent feel?
• may be due to I-s' behavior (e.g. non verbal) more than fixed attributes
• level of sensitivity: increased rapport may increase disclosure of moderately but not highly sensitive behavior

• Social Distance between Interviewer and respondent
  - Johnson (2000) operationalized social distance as overlap of demographic features between I and R
    more overlap = less social distance
    age, gender, race/ethnicity, education
  - How does I-R distance affect disclosure of undesirable behavior
  - telephone Rs report more drug use than those less distant from I-s

| Social Distance Index | Lifetime Use | Recent Use |
|---|---|---|
| 0 (low similarity) | 23.3 | 3.0 |
| 1 | 26.9 | 4.9 |
| 2 | 33.1 | 7.0 |
| 3 | 36.0 | 8.6 |
| 4 (high similarity) | 49.7 | 9.3 |

• Rapport: I's nonverbal behavior
  - Foucault (2013) analyzed facial behavior of 4 NSFG I-s and obtained R ratings of rapport with I
    Ratings based on agreement that particular objectives characterized interview situation (e.g. relaxed, cooperative, unfriendly) and interviewer (e.g. cooperative, friendly, cold)
  - Rapport (based on ratings) associated
    Positively with interviewer smiling and nodding
    Negatively with interviewer direct gaze
  - No direct measure of data quality (but in Foucault & Sun, under review), more acceptable answers with more rapport
• Rapport and disclosure
  - possible that increased rapport will deter disclosure of very sensitive info because R cares more about interviewer's impression of him/her
  - Sun (2014, under review) varied the sensitivity of questions
    ** questions rated (in separate study) as high, moderate or low in sensitivity
    ** collected Rs ratings of rapport immediately after interview
    ** High rapport increased disclosure of sensitive information including highly sensitive behavior (contrary to hypothesis)
    ** but increased item non-response (consistent with hypothesis)

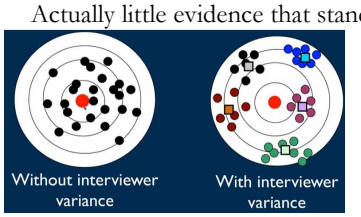
Summary
• personal style motivate respondents to give accurate and candid responses relative to formal style without increasing ingratiation
• In fact, leads to report of more (moderately) socially undesirable behaviors
• Seems to work by increasing rapport between I and R
  - Note: Personal != Conversational interviewing
• Increased rapport seems to increased disclosure of moderately and extremely sensitive behavior
• but also increases non-response for extremely sensitive questions
• less social distance between I and R seems to increase disclosure

## 3.3 Interviewer effects

*Variance: Interviewer Behavior*

• Interviewers can add value to the survey enterprise (e.g. increase response rates, explain Q meaning to R, conduct within HH sampling)
• But can also introduce error – interviewer effects
• 2 main kinds of interviewer effects, either due to:
- interviewer behavior, associated with variance due to interviewer
- interviewer fixed attributes, e.g. race, gender, age, associated with directional response error (bias)

• Interviewer Effects: Behavior
- Generally assumed differences in Interviewer behavior inflate interviewer-related variance (rho_int)
Differences between the answers collected by different interviewers
- standardized interviewing conducted to reduce diff's in interviewer behavior and so reduce rho_int
Actually little evidence that standardization has this effect




Estimating Interviewer variance
• Kish (1962) introduced rho_int to capture the correlation between interviewers and the answers they elicit (sometimes called interclass correlation)

$$\rho_{int} = \frac{\text{between – interviewer variance}}{\text{between – interviewer variance + within-interviewer variance}}$$

• Between interviewer variance is ideally zero
- But never is
- Assumes random assignment of interviewers to respondents i.e. interpretation
• Within interviewer variance is (essentially) range of values for variable of interest
- usually answers to a particular question
• rho_int permits computation of design effects due to interviewers

Some Question types more prone to interviewer effects
• O' Muircheartaigh & Campanelli (1999) observed significant rho_int in:
- attitude questions (26%)
With Likert scales (33%)
- factual questions (26%)
• Interviewer effects inflate variance about as much as geographical clustering
• Schnell * Kreuter (2005) examine 108 questions about crime
- Interviewer effects greater for
- sensitive than non-sensitive Q
- nonfactual than factual Qs
- Open than closed Qs
- Difficult than easy Qs
- Interviewer effects increased with index (0-4) of harmful question properties

Interviewer behavior and Interviewer Effects
• Some interviewer behavior may affect rho_int
• Mangione, Fowler & Louis (1992) identified interviewer behaviors in recorded interviews
• Computed correlation between each behavior and rho_int
• Most correlations very small but those involving probes are significant

| | |
|---|---|
| Correct Probe | .23 |
| Directive Probe | .20 |
| Failed to Probe | .49 |
| Recording Error | .39 |

Significant correlations between the incidence of specific interviewer behaviors and ρ
Source: Mangione, Fowler & Louis (1992)

Interviewer Variance (IV) and Interviewing Technique
• Proponents of standardized interviewing argue that departure from scripted interaction will increase IV (rho_int)
• West, Conrad, Kreuter & Mittereder (under review) compared IV in Standardized Interviewing (SI) and Conversational Interviewing(CI) ➔ In FTF interviews, national sample
• CI improved data quality in 25% of Qs but increased IV in only 10%
- even when IV increased in CI, did not offset improved quality
• IV similar in SI and CI
• CI led to more variance in duration but not longer interviews

Interviewer Effects on Non Response
• What appears to be interviewer effect on measurement can reflect different I-s recruiting different types of Rs
- despite interpenetration
• West and Olsen (2013) used administrative records (divorce certificates) to determine if interviewer effects due to non-response or measurement in responses to Wisconsin Divorce Study
- Two Qs produced rho_int reliably > 0: age when married and age when divorced
- Based on records, effect for age at marriage due to measurement error, but effect for age at divorce due to significant nonresponse error
- For age at divorce, seems some I-s recruited younger Rs than other I-s

Impact of Interviewer Effects on Precision
• Design Effect (deff)
- deff_int = 1 + rho_int(m-1)
- inflation of the variance due to clustering
- m = average interviewer workload (# interviews)
- m moderates (i.e. amplifies/diminishes) impact of rho_int
More interviewers each conducting fewer interviews will reduce deff_int
- design factor (deft) is sqrt(deff)
• Effective Sample size : n / deff
- e.g. if n = 1000 and deff = 1.26 effective sample size = 793

**Relationship Between $\rho_{int}$, Workload, and Design Effects**

| | Telephone Methodology | Health in America | Consumer Attitudes |
|---|---|---|---|
| Response Rate | 59% | 80% | 74% |
| Number of interviewers | 37 | 33 | 26 |
| Average workload | 41.3 | 58.1 | 14.1 |
| Mean values: | | | |
| $\rho_{int}$ | .0089 | .0018 | .0067 |
| $deff_{int}$ | 1.36 | 1.10 | 1.09 |

➔ impact of clustering is diluted interviewers are conducting less interviews each, so deff_int is about the same
(If you look at Health in America v.s. Consumer attitudes, health survey has way lower rho_int values but deff_int is about the same because consumer survey has way less average workload)

• Interviewer-related variance (rho_int) is a correlation between interviewers and the responses they elicit
- it should not matter who asks the questions
- so association between I-s and responses suggests variation in quality of answers
• But rho_int is an indirect measure of data quality
- possible for rho_int to be low (good) and response validity to be low (bad) if all interviewers consistently collect wrong answer
- e.g. Rs asked if have recently visited doctor; survey intended to measure MD visits but interviewers not instructed to or not permitted to define "doctor"; many Rs include visits to non-MDs, e.g. Pas, podiatrists
- rho_int would be low (uniform response distributions across interviewers) and accuracy would also be low (many answers > true value)

Interviewer Effects due to behavior: Summary
• Interviewer variance related to Interviewer Behavior

- e.g. probing
- but no greater in CI than SI (despite reductions in measurement error with CI)
- rho_int can be inflated by nonresponse error
  If I-s recruit Rs whose true values differ from those recruited by other I-s

• Impact of rho_int "diluted" by small average workload (fewer interviews per interviewer)
• rho_int generally not a measure of response accuracy – measures variance so indirect measure of data quality

*Bias: Interviewers' Fixed Attributes*

Gender of Interviewer
• Kane & Macaulay (1993) report that both male and female Rs gave more egalitarian answers to female than male I-s but for different teams
- more men reported sharing child care when I was female
- more women supported feminist policy when I was female
- no impact of interviewer gender when topic of questions not relevant to gender

• Huddy (1997) report numerous gender of interviewer effects in two pre-election surveys
- male and female Rs show same gender of interviewer effects

Race of Interviewer (ROI) effects
• Analogous to gender of interviewer effects
- only observed for Qs for which race is relevant
- even for those items, only observe effects some of the time

• Effects generally larger in face-to-face than telephones studies but ROI reported to affect responses in phone interviews

| Category | Black interviewers | White interviewers |
|---|---|---|
| Would not mind if relative married a Negro | 71.7% | 25.5% |
| Believes Negro and white students should go to same school | 90.7% | 56.0% |
| Would not be disturbed if Negro of same class moved into block | 100.0% | 68.6% |
| Believe Negro and white children should play together freely | 92.5% | 84.0% |

| | | % in Category | |
|---|---|---|---|
| Question | Category | White Intw'r | Black Intw'r |
| Can trust white people? | Trust most whites | 35 | 7 |
| Negro parents work best with Negro teacher? | Yes | 14 | 29 |
| Favorite entertainers? | Named only black entertainers | 16 | 43 |

Race of Actual and Video Recorded Interviewers
• Krysan & Couper (2003) compared impact of live and video recorded I-s (V-ACASI) on answers to race-related Qs
- allows comparison of social desirability in live interviews to stereotype activation in video-recorded interviews

• African American Rs reported less liberal racial attitudes to white than black I-s in both live and recorded interviews
- social desirability live interviews, stereotype activation for recorded interviews

• White Rs reported more conservative racial attitudes to black than white recorded I-s
- stereotype activation (priming of existing attitudes)

ROI effects in telephone interviews
• Davis (1997) examined race of interviewer effects in the 1984 National Black Election Study
• Telephone survey of African Americans
• 48 white interviewers and 27 African American interviewers
• Correlation between racial consciousness and support for Jesse Jackson:
- 0.15 when interviewer is African American
- 0.04 when interviewer is White

Perceived v.s. actual race of interviewer
• Telephone Rs have only auditory cues to interviewer's race
• Davis and Silver (2003) found
- R's ability to judge race varies widely "correct" race identification ranged from 14% - 82%
- Black Rs answered fewer knowledge questions correctly – stereotype threat – when they **perceived I-s to be White**
- But no significant difference based on I-s' **self-reported race**

Social distance between Interviewer and respondent – Johnson (2000)

• In CATI survey (no visual info), Rs more likely report drug use if more similar to I-s
  - social distance score based on I-R overlap on age, gender, race/ ethnicity, education
  - more overlap = less social distance

| Social Distance Index | Lifetime Use | Recent Use |
|---|---|---|
| 0 (low similarity) | 23.3 | 3.0 |
| 1 | 26.9 | 4.9 |
| 2 | 33.1 | 7.0 |
| 3 | 36.0 | 8.6 |
| 4 (high similarity) | 49.7 | 9.3 |

Interviewer effects due to fixed attributes: Summary
• Attributes such as race and gender can lead to biased responses when topic relevant to attribute
  - little effect when topics unrelated to attribute
  - stronger FTF but effects of (perceived) race observed on phone
• Social distance based on demographic overlap can affect disclosure: the more overlap the more disclosure
• Beyond training interviewers to administer questionnaire as professionally as possible, probably little one can do to reduce these kinds of interviewer effects
• Interviewers recruit units (e.g. households) and select respondents from within units
  - recruiting can be affected by how interviewers interact with sample members
  - selection can be affected by type of procedures e.g. Kish v.s. B day methods
• Interviewing technique can affect response accuracy – and may not affect interviewer variance
  - CI can substantially increase response accuracy but takes more time than SI
• Interviewer effects, both due to I-s' behavior and attributes, add error to survey estimate
• But interviewers continue to add value to survey enterprise by increasing response rates, clarifying question meaning and motivating Rs and keeping them on task

### 3.4. Interview with Nora Cate Schaeffer