

Non-response

Ideal Situation: we prevent it

In Practice: ????

Two kinds of non-response:

- **Unit Non-response**: no information from a unit
- **Item Non-response**: no information from a unit for one (or more) particular unit; do get information from the unit for at least one other item

When might we get a high percentage of unit non-responses?
item non-responses??

*sensitive subject;
don't understand
can't locate people
etc.*

Nonresponse Bias: nonresponders are different from responders.

(think about it as dividing the population into two strata; you're missing membership of ^{the nonresponse} strata)

Types of Unit Non-response (and what we can do about them):

- Non-contact

Remedy? - repeated contact
(no success decreases)

- Refusal

Remedy? what types are more likely to be refused?
incentive? → maybe change method of request

- Unable to Participate

Remedy? speak different language
health problems
mentally incapable

try to get demographic information about non-respondents;

*age
gender
race
if person
characteristics of home/city
(rural/urban, etc)*

if not person, get information about area, time, location, etc.

Try to identify possible correlations w/

What can we do about item non-response?

nonresponse

Approaches for Non-response

- Ignore the nonresponse.

What assumption does this make about our sample and our population?

sample is representative of population; all estimates are "representative" as well

in reality, nonresponse likely tied to ~~some~~ one or more variables/characteristics of respondents

two "strata" in the populations: respondents/non-respondents

you're getting a sample of the respondents

What effect does this have on our estimates?? estimating for respondents

Ex:

N_R # of respondents

N_M # missing/non-respondents

$N_R + N_M = N$ # in pop

For the population $\bar{Y}_{pop} = \frac{N_R}{N} \bar{Y}_R + \frac{N_M}{N} \bar{Y}_M$ we ~~end~~ up trying to estimate $\bar{Y}_R$ ~~with~~ our sample mean $\bar{Y}_r$

Say Bias? $E[\bar{Y}_r] - \bar{Y}_{pop} \approx \bar{Y}_R - \frac{N_R}{N} \bar{Y}_R - \frac{N_M}{N} \bar{Y}_M = \frac{N_M}{N} (\bar{Y}_R - \bar{Y}_M)$
 $\bar{Y}_r$ approx unbiased est of $\bar{Y}_R$ likely
 Variance decreases as well - why?

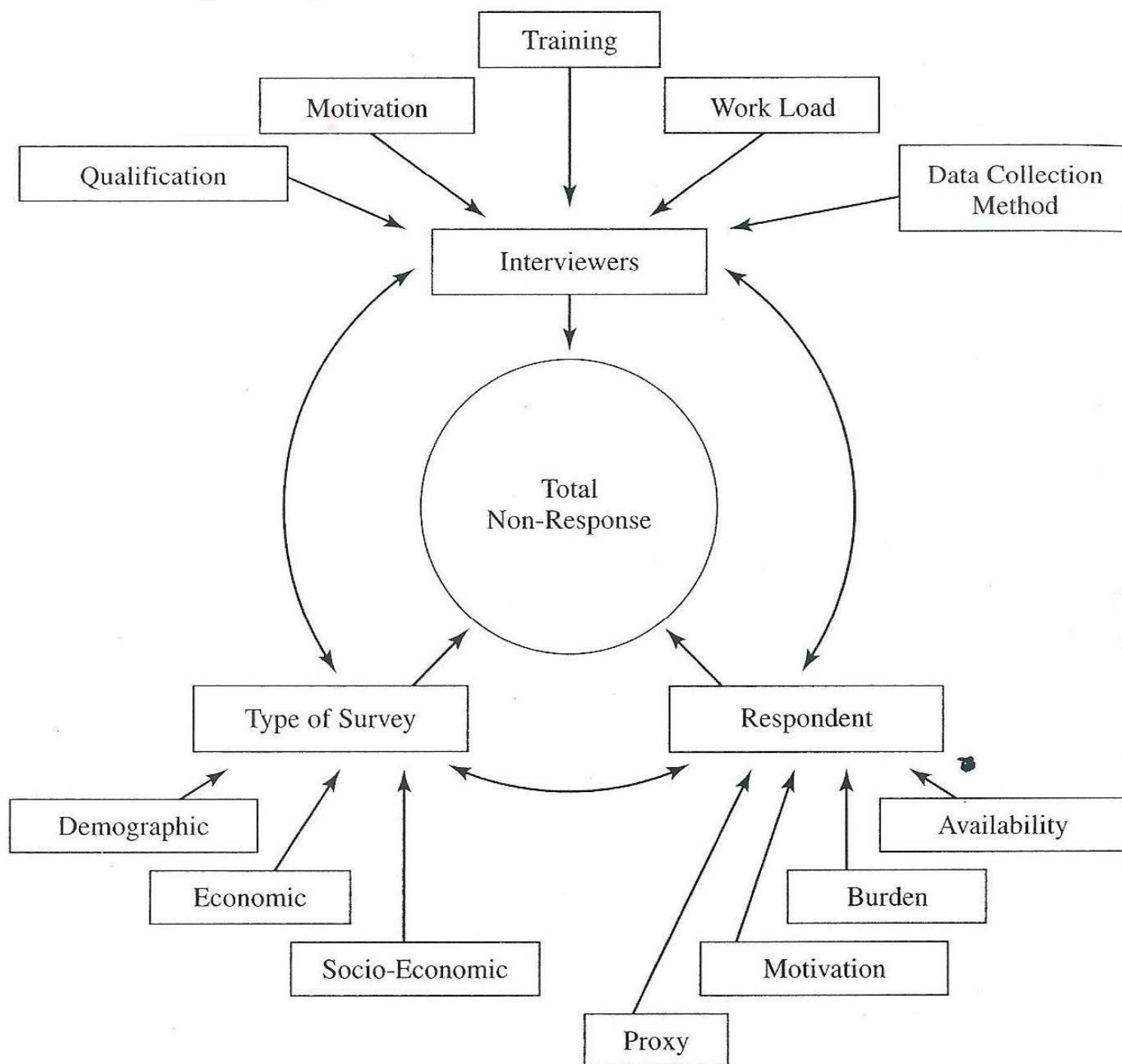
when is bias small?
 1) $\bar{Y}_R \approx \bar{Y}_M$
 2) $\frac{N_M}{N}$ small

What if we just asked more people? non-response mechanism still there;

continue to sample people from respondent strata;

do not get a "representative" sample automatically by increasing # people in sample

FIGURE 8.2
Factors affecting nonresponse



SOURCE: "Some Factors Affecting Non-Response," by R. Platek, 1977, *Survey Methodology*, 3, 191–214. Copyright © 1977 Survey Methodology. Reprinted with permission.

as related to (1) survey content, (2) methods of data collection, and (3) respondent characteristics, and illustrates various sources using the diagram in Figure 8.2. Groves

- Prevent it. (Best method)

How?? Survey Design

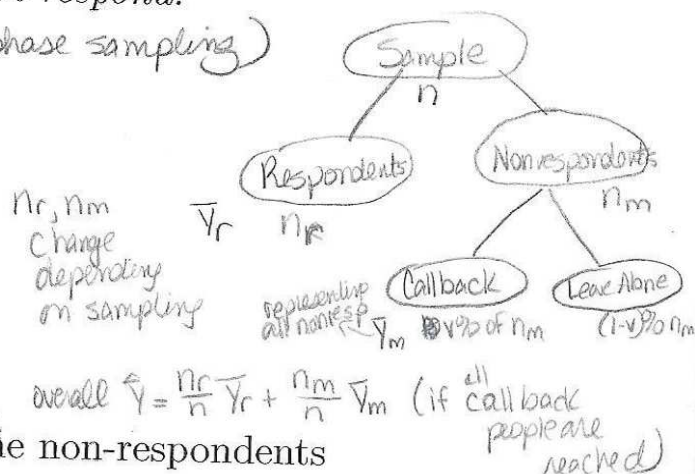
- Survey content: some sensitive subjects may get large non-response. Ex?
 may increase response rate by ordering questions carefully
 randomized response technique } flip coin to get question; interviewer doesn't know which question you're answering
- Time of survey:
 think about what time of year/time of day etc will increase your response rate given your subject content
- Interviewers: improve skills; consistent training; qualification
 need to motivate; not overload
 consistent data-collection (easy to do)
- Data-collection Method: consistent; which type of survey?
 what types of surveys would you more readily answer?
 computers? written?
- Questionnaire Design: question wording, order
- Burden on Respondent: should be non-intrusive; short, easy,
 stratification may help
 b/c need smaller sample gives desired precision
 ↳ what's easy?
- Survey Introduction:
 introduction should explain clearly & motivate the respondent
 assure confidentiality
 ↳ what will their response contribute to? what purpose?
- Incentives/Disincentives:
 financial? raffle?
 removing privilege or suspending a privilege if don't respond
- Follow-up:
 initial contact less costly per unit than followups of non-respondents
 ↳ expensive
 need to think about increasing responses
 vs. nonresponse

ALWAYS TRY TO GET SOME CHARACTERISTICS OF NON-RESPONDENTS

- Take a representative subsample of the non-respondents; use it to make inference about the other nonrespondents

How can we do this? They didn't respond.

Callbacks (Multi-phase sampling)



- Predict the missing values for the non-respondents

What type of missing values can we have?

- Missing completely at random - chance of non-response for any unit doesn't depend on any variable / characteristics of sample / people.
 respondents are "representative" of pop; nonrespondents are a random sample
 estimates are "unaffected" → i.e. they're ~~also~~ considered approx unbiased
- Missing at random (given the questions asked); not dependent on response; nonresponse depends ^{only} on observed variables; can account for patterns in missingness.
 nonresponse is random if conditioned on covariates
 income cannot

Asking about political party affiliation / education

INCOME → high incomes tend to have higher nonresponse
what would high income be affiliated with?

- condition on Income (split high income / low income brackets) Missingness random within the brackets
- Missing dependent on response; NOT IGNORABLE, why??
 can't be explained by conditioning on covariates
 Survey of crime victims → whether or not been victim

How could we predict or fill in the missing values? mean? median?

How would this imputation affect our estimates?? Can maybe deduce

what effect on mean?

what effect on variance?

cluster group
respondents into
similar classes; ←
assign class value

hotdeck:
random,
nearest
neighbor

Response Rates

Can have several different response rates, depending on type of survey, non-response etc. Some rates are inflated because people remove units that could not be located from the denominator.

Specify clearly what your response rate is.

Define what response means for your survey.

Possible rates:

- no. of completed interviews / no. of units in sample
 $\downarrow$
 no unit non-response?
 no item non response? includes everybody
~~respondents & non-respondents~~
(even those who you couldn't contact)
- no. of completed interviews / no. of units contacted
 $\downarrow$
 respondents & non-respondents
 for refusal & unable to answer
- (completed interviews + ineligible units) / contacted units
 $\downarrow$
 interviews that are excluded
- completed interviews / (contacted units - ineligible units)
 $\downarrow$
 resp & non-resp
 for refusal
~~non-respondents~~
- completed interviews / (contacted units - ineligible units - refusals)
 $\downarrow$
~~non-respondents~~
 respondents

Compare/Contrast: