Survey data analysis
Week 44:
“Nonresponse”

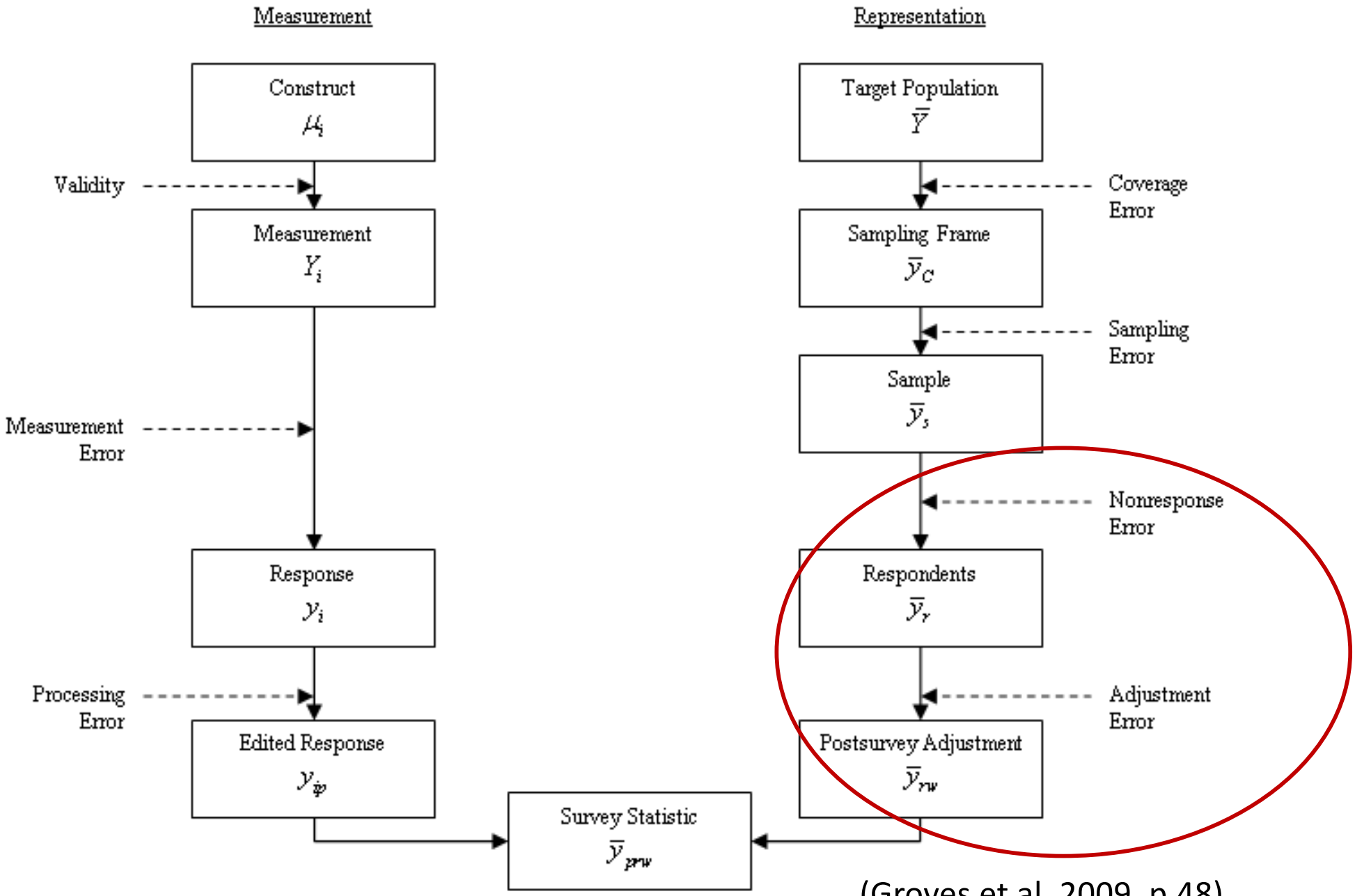
Today

- Lecture on NR
- Exercise working with weights

Literature today

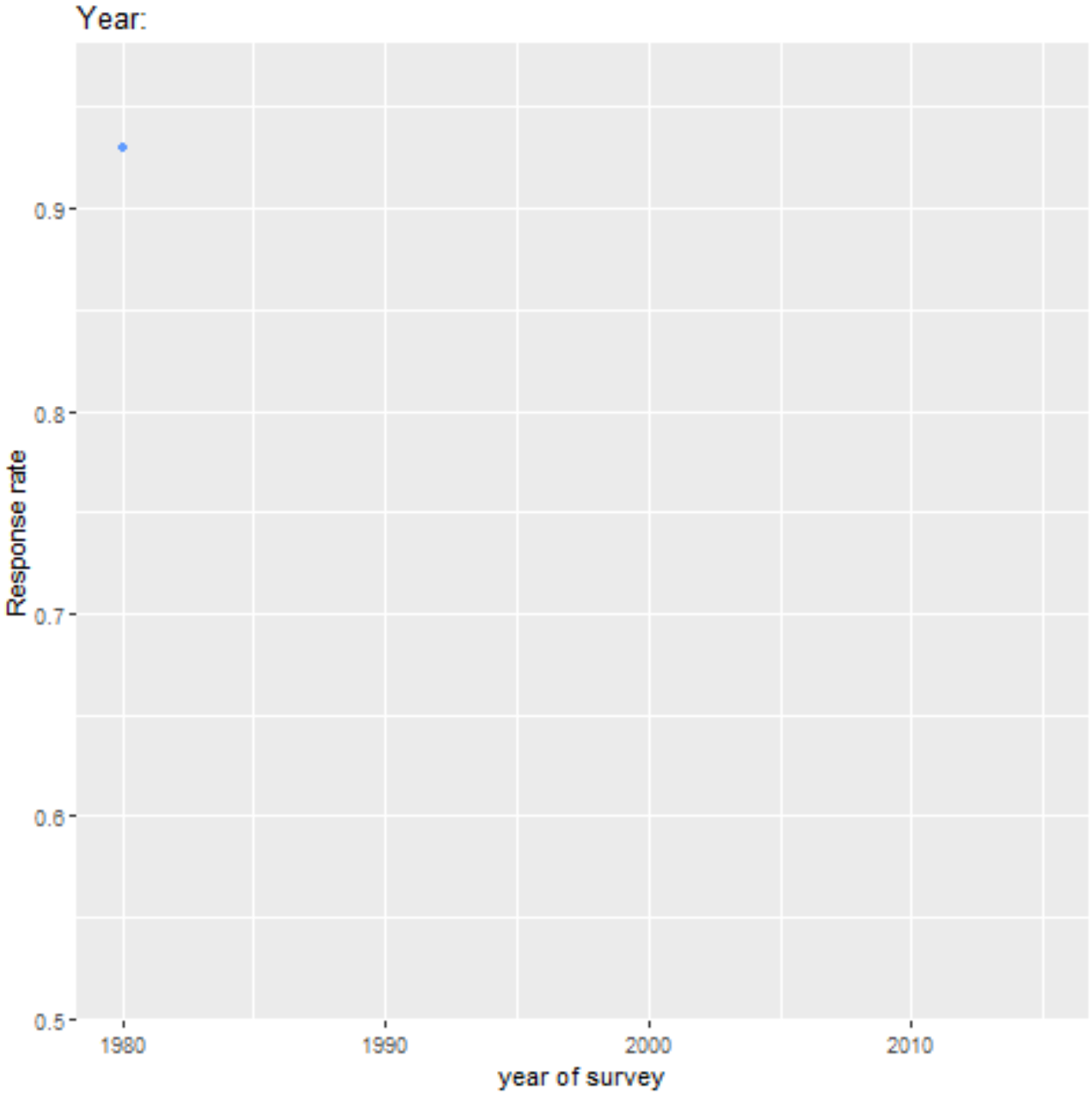
- The increasing problem of nonresponse
  - Luiten, de Leeuw & Hox (2018)
- Details of weighting methods
  - Kalton and Flores-cervantes (2003)

Total Survey Error (TSE) Framework



(Groves et al. 2009, p.48)

Nonresponse in LFS over time



Based on Luiten, De leeeuw & Hox (2018) nternational Nonresponse Trends across Countries and Years: An analysis of 36 years of Labour Force Survey data. Survey Insights: Methods from the Field. Retrieved from https://surveyinsights.org/?p=10452.

European Social Survey Nonresponse

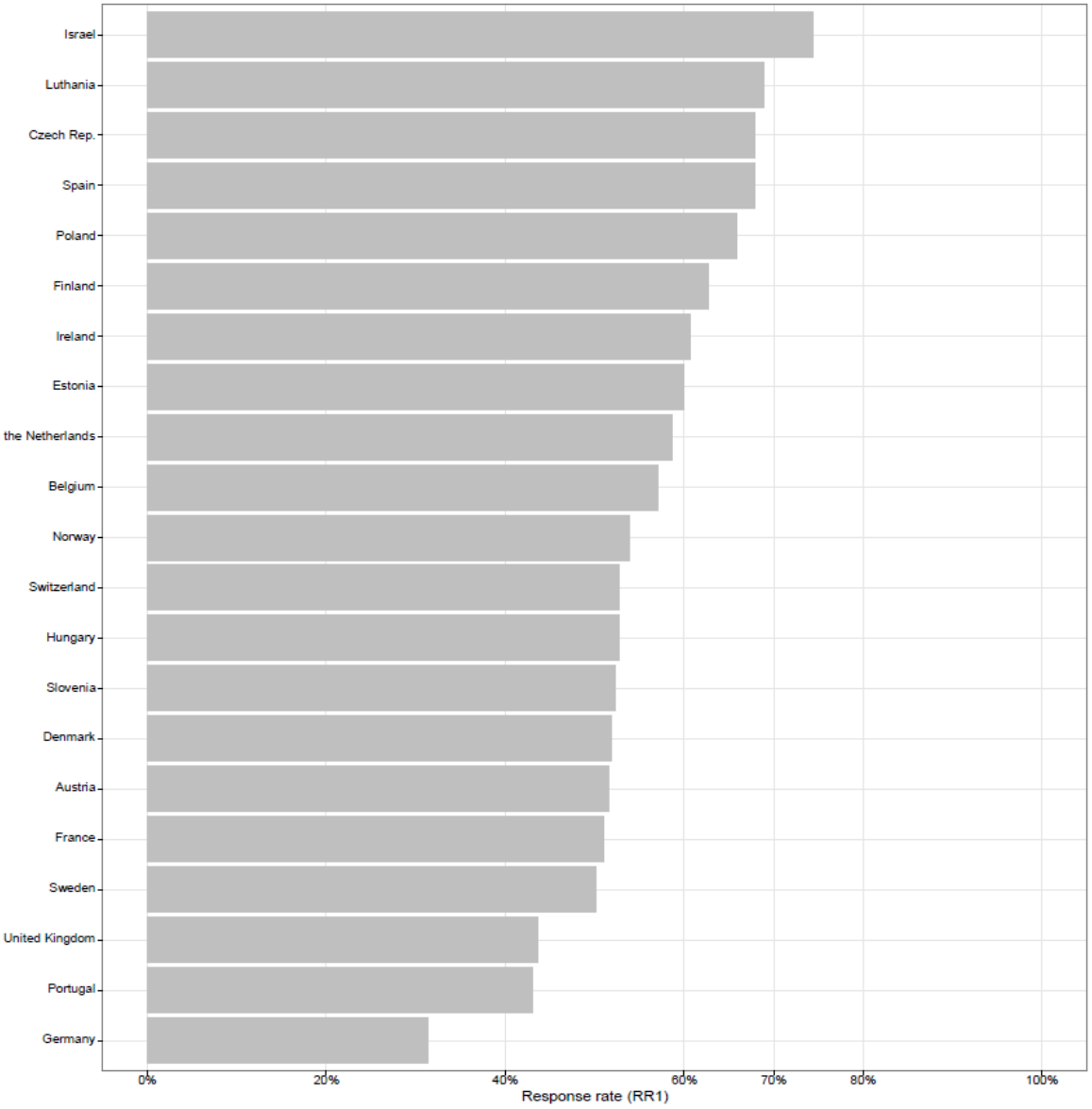
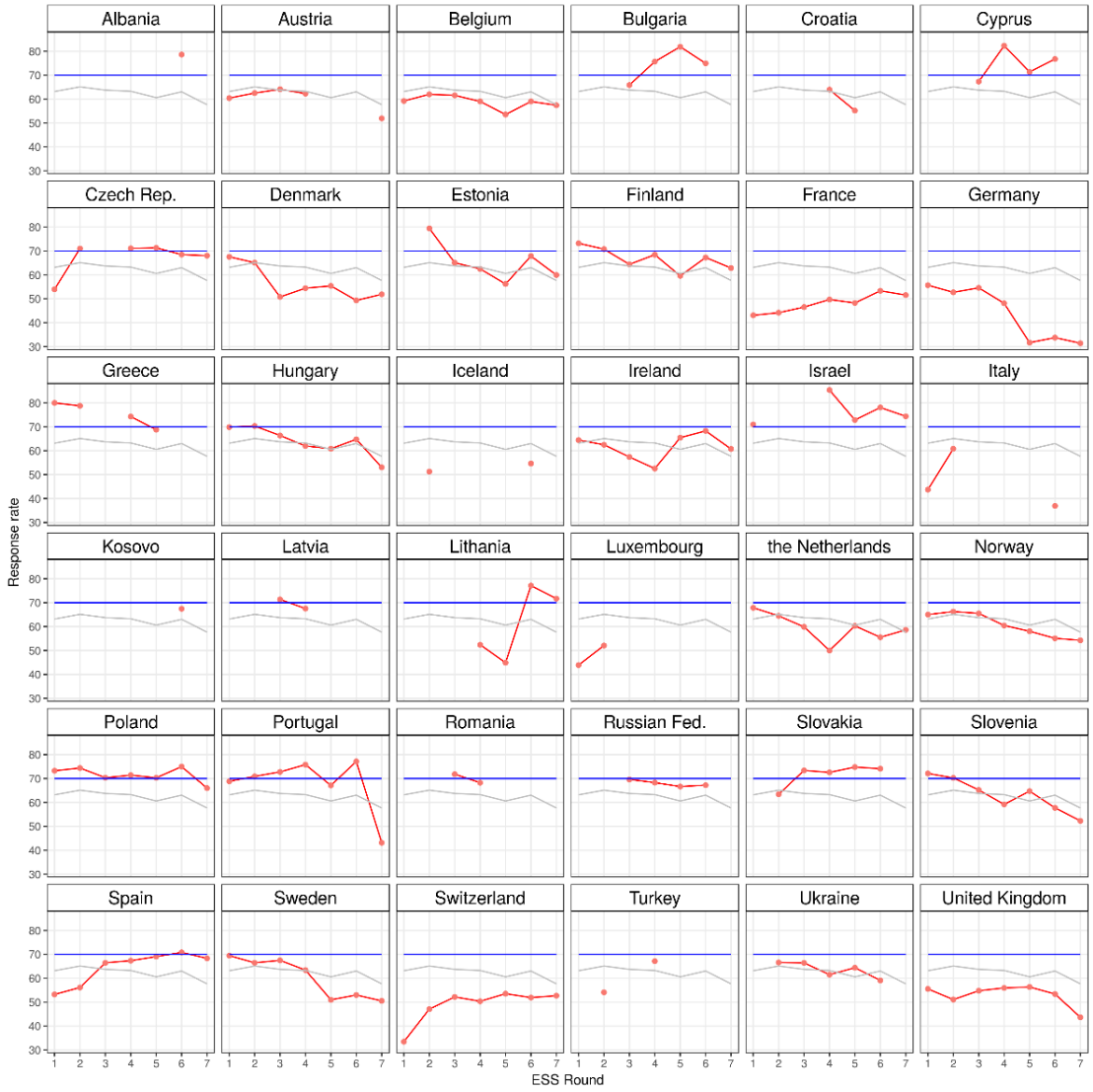


Figure 1: Response rates per country, ESS7[1]

From: Beullens, K., Loosveldt G., Vandenplas C. & Stoop I. (2018).Response Rates in the European Social Survey: Increasing, Decreasing, or a Matter of Fieldwork Efforts? *Survey Methods: Insights from the Field. Retrieved from* https://surveyinsights.org/?p=9673

ESS: RR variation in response rate



Four main types of nonresponse

In survey research we typically distinguish four types of nonresponse:

- **Unit nonresponse**
  The sample unit (e.g. person, household, institution) was sampled, i.e. belonged to the gross sample, but did not participate in the survey.

- **Item nonresponse**
  The sample unit was sampled and interviewed, but failed to provide answers to all of the survey items.

- **Attrition**
  The sample unit was sampled and initially interviewed for a longitudinal surveys, but did not complete all waves of the survey.

- **Partial (household) nonresponse**
  The sample unit was sampled and at least one member of the unit interviewed. However, at least one member of the unit did not participate.

Main causes of nonresponse

- **Unit nonresponse**

  non-contact, refusal, unable

- **Item nonresponse**

  refusal, don’t know, breakoff

- **Attrition**

  non-location, non-contact, refusal, unable

- **Partial (household) nonresponse**

  non-contact, refusal, unable

How to prevent nonresponse

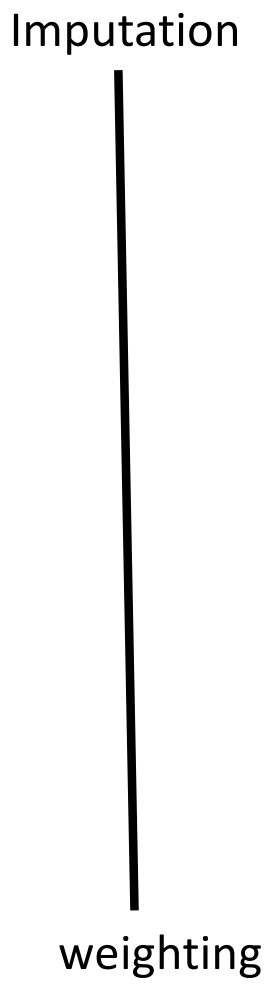
- Things you noticed in your adopted survey?

How to prevent nonresponse

1. A good questionnaire, invitation letter, etc.
   - keep it simple, keep it simple, test it
2. Incentives
   - Preferable unconditional, and cash
3. Multiple contact attempts
4. Multiple modes (e-mail, mail, phone, f2f)
5. Refusal conversion
   - Interviewer training
6. Be responsive to questions/remarks/problems

Correction for nonresponse

- Item nonresponse
  - Rich information on individual
- Partial (household) nonresponse
  - Proxy-answers, information on household
- Attrition
  - Information from earlier waves
- Unit nonresponse
  - Weak individual information (only frame)



What is nonresponse bias?

- Nonresponse bias occurs when the sampled units (e.g. individual, household, business …) are not or only partially observed (e.g. interviewed)
- **AND** observed units are systematically different from unobserved units.

MCAR, MAR, NMAR

- Missing Completely At Random (MCAR):

  The responding units are a random subsample of the gross sample.

- Missing At Random (MAR):

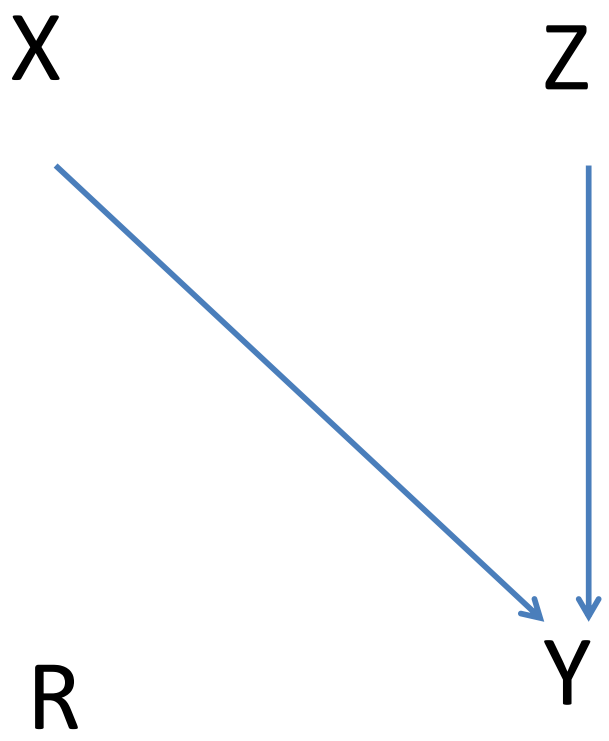
  The responding units are not a random subsample of the gross sample. However, the auxiliary information *x* renders the relationship between *y* and response *r* independent.

- Not Missing At Random (NMAR):

  The responding units are not a random subsample of the gross sample. In addition, the auxiliary information x does not render the relationship between *y* and response *r* independent.

Missing data mechanisms

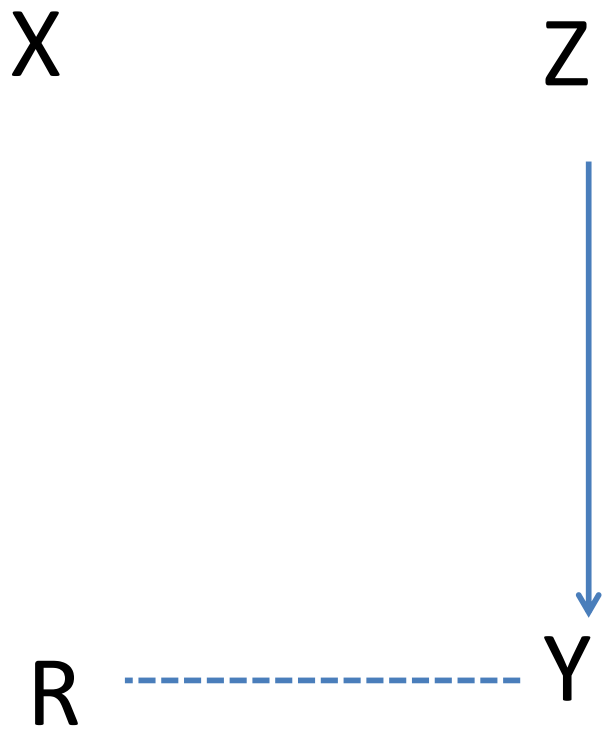
MCAR



Missing data mechanisms

MCAR

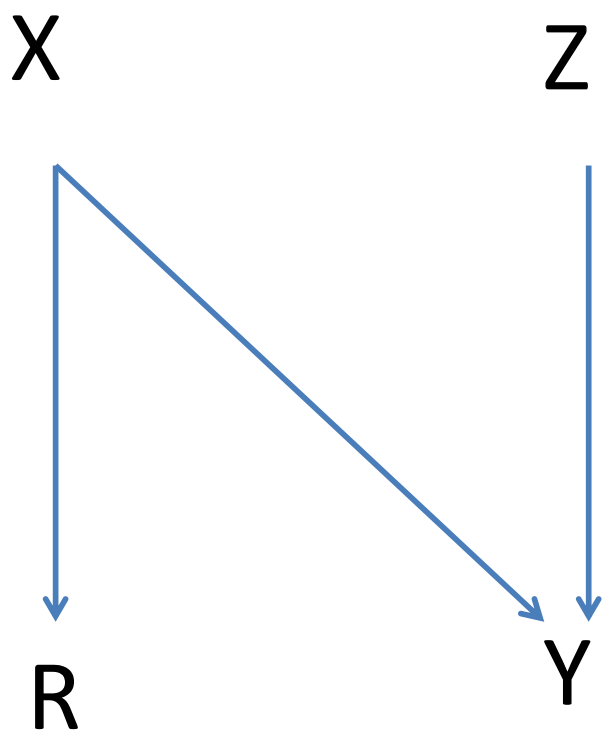
MAR



Missing data mechanisms

MCAR

MAR

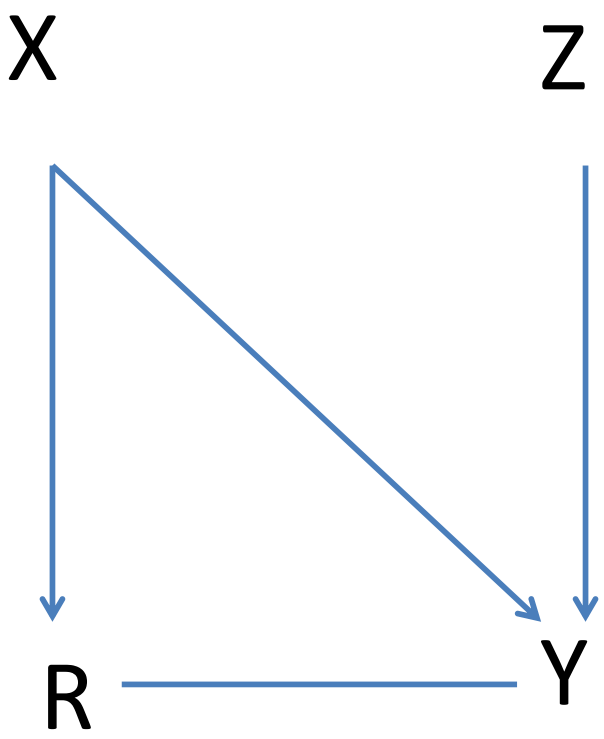


Missing data mechanisms

MCAR

MAR

MNAR



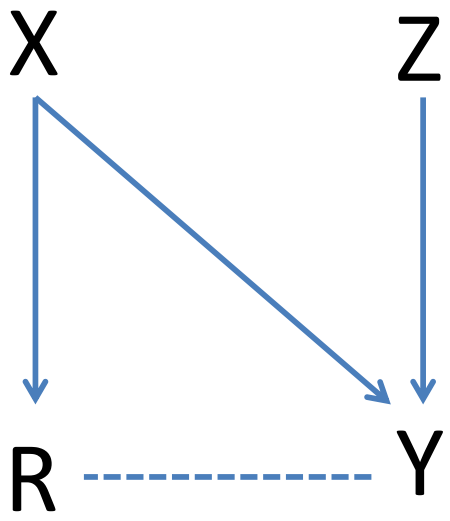
Example: income

- Unit response rates in surveys:
  - ~5-50%
  - Nonresponse: 50-95%!
- Item-nonresponse for income question:
  - ~25%

- What do we have: MCAR, MAR, or MNAR?
  - Discuss!

Item nonresponse (weeks 46,47)

- Use covariates (x) at level of respondent
- Strongly related to both response (R) and Y
- MCAR, MAR, MNAR models

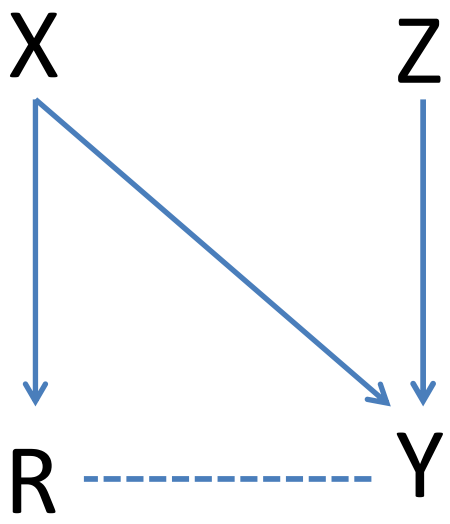


Example item missings in income:

- Education, wealth, age, gender, value of house (X)
- To predict income (Y) and
- Take away relation R-Y

Unit nonresponse

- Few covariates (x) at level of respondent
  - Often only address, or e-mail
- Weakly related to both response (R) and Y



Example Unit Nonresponse:

- Only use address (house price)
- Can predict income (Y), but
- Cannot explain relation R-Y
- Not successful in NR correction

Why weight?

Sampling: selection probabilities may differ

-> design weights

Coverage: sampling list may not cover target population

Nonresponse: not all people in sample will end up in data

-> adjustment weights for coverage/NR

Design weights (repeat from weeks 39-42)

SRS: equal probabilities

-> no design weights

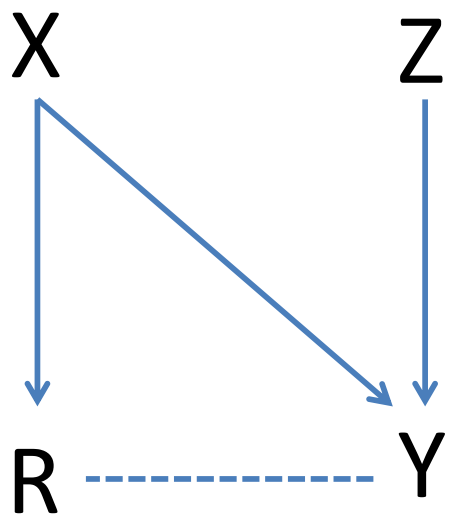
Stratified, cluster, multistage

-> need design weights for unbiased estimates

- See slides for those weeks
- Weights not needed if:
  - you specify correct svydesign (ids=~, strata=~, ….)
  - You use a HT-estimator (fpc = ~inclusionprobabilities)

The idea behind weighting

- X values at level of (sub)population
- One weighting model for all substantive analysis
  - In imputation model often Y-specific.



Population level data?

1. Sampling frame (nonresponse)
   - Address.
     - Can be enriched (e.g. use google streetview)
   - Statistics Netherlands: admin data
- 2. Population level statistics (coverage + NR)
   - Gender (50/50), age, income, region, nationality

Weighting methods

- 1. Propensity score weighting
  - Often using frame information from sample

- 2. post-stratification
- 3. linear weighting (GREG)
- 4. Raking
  - 2-4 often based on population statistics

Nonresponse bias

Deterministic

It is a function of the nonresponse rate *M/N* and the difference between the respondents' *r* and the nonrespondents' *m* population values.

$$B(\bar{y}_r) = \left(\frac{M}{N}\right)\left(\bar{Y}_r - \bar{Y}_m\right)$$

Probabilistic

It is a function of the correlation σ of the survey outcome y with the response propensity ρ and the mean response propensity measured in the target population (Bethlehem 2002).

$$B(\bar{y}_r) \approx \frac{\sigma_{y\rho}}{\bar{\rho}}$$

Note that nonresponse bias is always estimate-specific!

Propensity weighting

Deterministic

It is a function of the nonresponse rate *M/N* and the difference between the respondents' *r* and the nonrespondents' *m* population values.

$$B(\bar{y}_r) = \left(\frac{M}{N}\right)\left(\bar{Y}_r - \bar{Y}_m\right)$$

Probabilistic

It is a function of the correlation σ of the survey outcome y with the response propensity ρ and the mean response propensity measured in the target population (Bethlehem 2002).

$$B(\bar{y}_r) \approx \frac{\sigma_{y\rho}}{\bar{\rho}}$$

Propensity-score weights

For propensity-score weights (logistic regression) models estimate the response propensity (predicted probability) of each sample unit given a set of covariates.

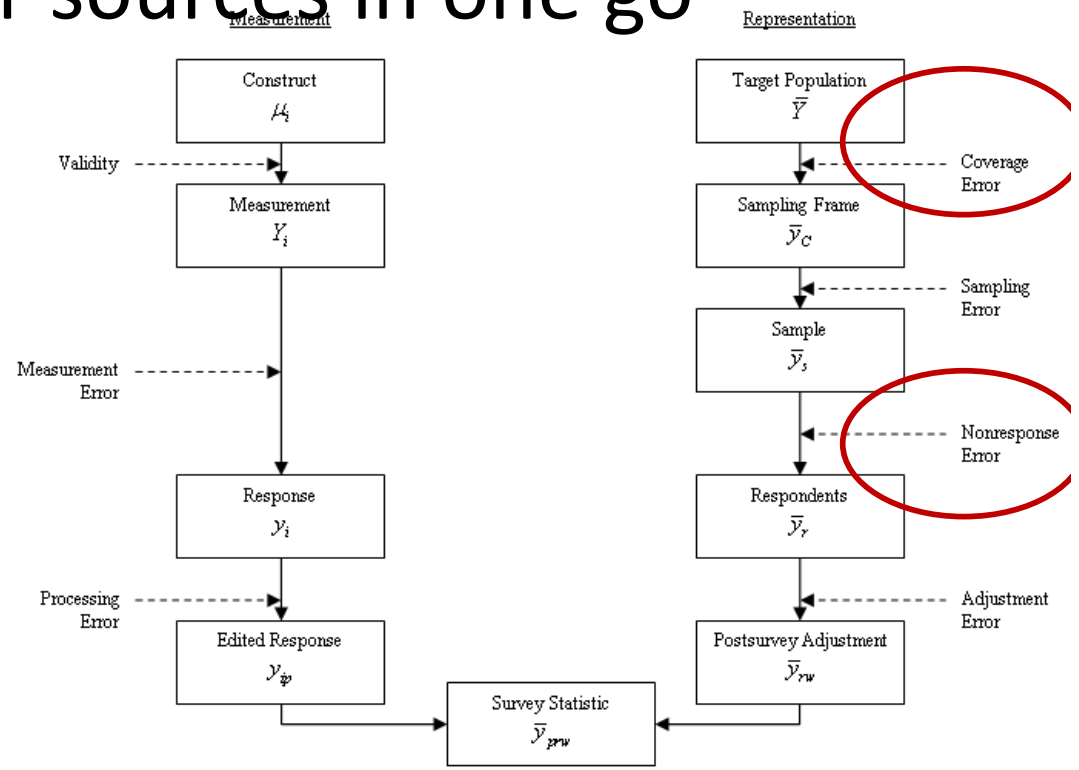
- Response rate for all linear combinations of for example:
  - response[0;1] ~ gender+age+region+typehouse

Weight is the scaled inverse of the predicted response propensity of each sample unit.

- Design weight = sample inclusion probability
- Propensity weight = participation probability

Post-stratification

- Deterministic approach
- Uses population statistics
- Correct multiple error sources in one go
  - Nonresponse
  - Coverage



Post-stratification: fictitious example

Survey distribution

Age	Gender	
	Male	Female
16-25	6.1%	5.9%
26-35	12.8%	10.7%
36-45	19.5%	16.9%
46-55	10.0%	9.5%
56-65	3.9%	4.7%

Population distribution

Age	Gender	
	Male	Female
16-25	10.0%	10.0%
26-35	10.0%	10.0%
36-45	10.0%	10.0%
46-55	10.0%	10.0%
56-65	10.0%	10.0%

Male respondents aged 16-25 receive a weight of
$w = 0.1/0.061 = 1.639$

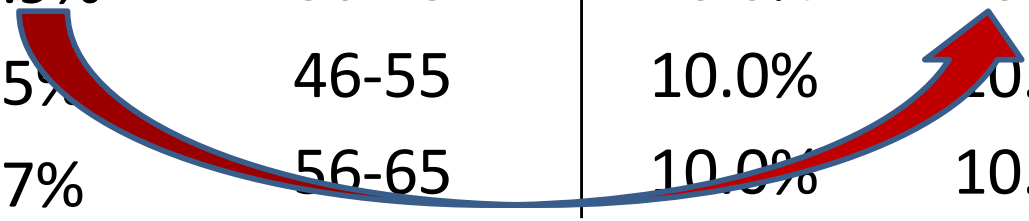
Post-stratification: fictitious example

Survey distribution

Age	Gender	
	Male	Female
16-25	6.1%	5.9%
26-35	12.8%	10.7%
36-45	19.5%	16.9%
46-55	10.0%	9.5%
56-65	3.9%	4.7%

Population distribution

Age	Gender	
	Male	Female
16-25	10.0%	10.0%
26-35	10.0%	10.0%
36-45	10.0%	10.0%
46-55	10.0%	10.0%
56-65	10.0%	10.0%



Female respondents aged 36-45 receive a weight of
$w = 0.1/0.169 = 0.592$

When does weighting work?

1. Target variables vary little within strata
2. Response probabilities vary little within strata
3. Target variable and response probabilities are not correlated within strata (=MAR)

Example Unit Nonresponse:

- Income varies little within gender and age (male/female)
- Given gender and age, there is little variation in response rates
- Given gender and age, there is no correlation between probability of response and income

When does weighting work?

1. Target variables vary little within strata
2. Response probabilities vary little within strata
3. Target variable and response probabilities are not correlated within strata (=MAR)

Example Unit Nonresponse:

- Income varies little within gender and age (male/female)
- Given gender and age, there is little variation in response rates
- Given gender and age, there is no correlation between probability of response and income

Unlikely

How to weight better…

- Use more variables in weighting
  - Gender, age, …
- To improve relation X-> R and X->Y

- Problems:
- 1. There may be empty strata
  - Combine empty strata
  - Use fewer variables
- 2. The population distribution across all variables may not be available

Raking

aka multiplicative weighting

Raked weights adjust the individual distributions of key survey variables to known joint population distributions.

Raking is an iterative process.

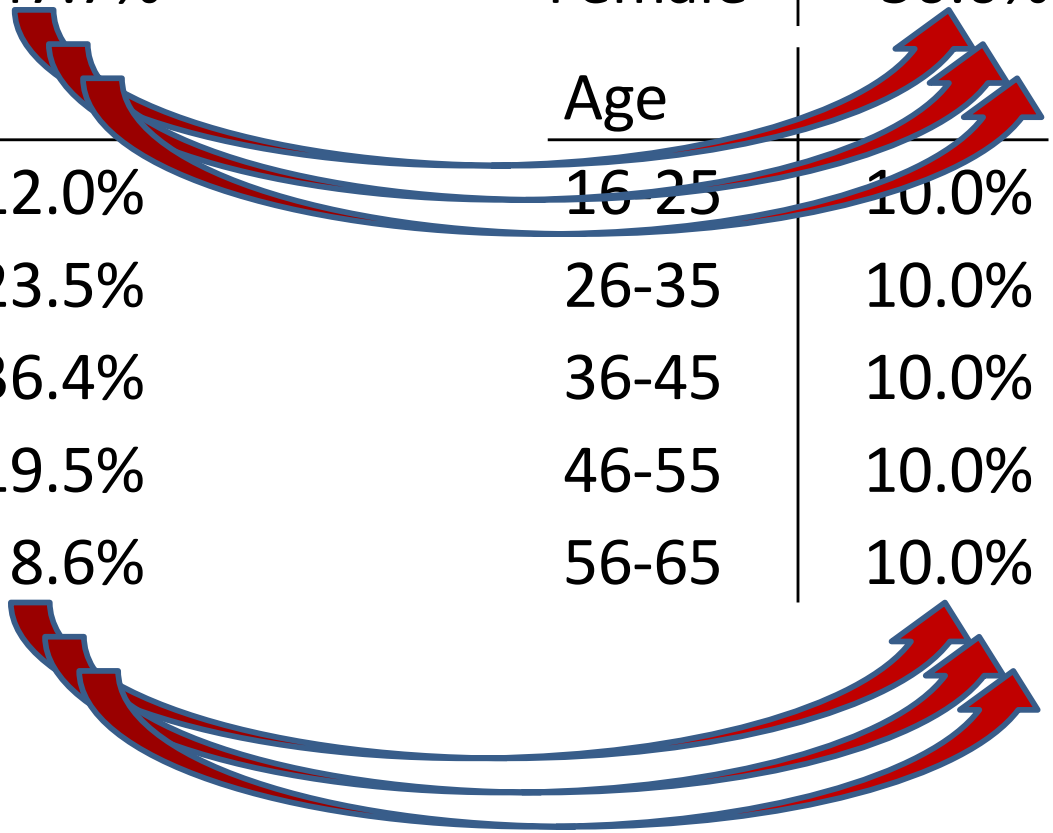
Raking: fictitious example

Survey distribution

Gender	
Male	52.3%
Female	47.7%
Age	
16-25	12.0%
26-35	23.5%
36-45	36.4%
46-55	19.5%
56-65	8.6%

Population distribution

Gender	
Male	50.0%
Female	50.0%
Age	
16-25	10.0%
26-35	10.0%
36-45	10.0%
46-55	10.0%
56-65	10.0%



Raking: fictitious example

Survey distribution

Gender	
Male	52.3%
Female	47.7%
Age	
16-25	12.0%
26-35	23.5%
36-45	36.4%
46-55	19.5%
56-65	8.6%

Population distribution

Gender	
Male	50.0%
Female	50.0%
Age	
16-25	10.0%
26-35	10.0%
36-45	10.0%
46-55	10.0%
56-65	10.0%

The survey distributions are iteratively adjusted to the age and gender population distributions until an equilibrium is reached.

Alternative: Linear weighting

- Aka *Generalized Regression Estimator (GREG) or calibration*
- Fixes problem of empty population cells
- Fixes problem of population distribution

	Male	Female	Total
Young	23	15	38
Middle	16	17	33
Old	13	16	39
Total	52	48	100

Sample

	Male	Female	Total
Young	?	?	43
Middle	?	?	30
Old	?	?	27
Total	51	49	100%

Population

With thanks to Jelke Bethlehem for this example

Linear weighting

Assume there are p continuous auxiliary variables

- Similar to propensity score weighting (other link-function)
- Required: vector of population means
- Estimate: conditional RR (most likely) for combinations of categories of p

- Best value for B (least squares): $$B = \left(\sum_{k=1}^{N} X_k X_k^{'} \right)^{-1} \left(\sum_{k=1}^{N} X_k Y_k \right)$$

- Sample-based estimate (full response): $$b = \left(\sum_{i=1}^{n} x_i x_i^{'} \right)^{-1} \left(\sum_{i=1}^{n} x_i y_i \right)$$

With thanks to Jelke Bethlehem for this example

Linear weighting

Linear weighting computations

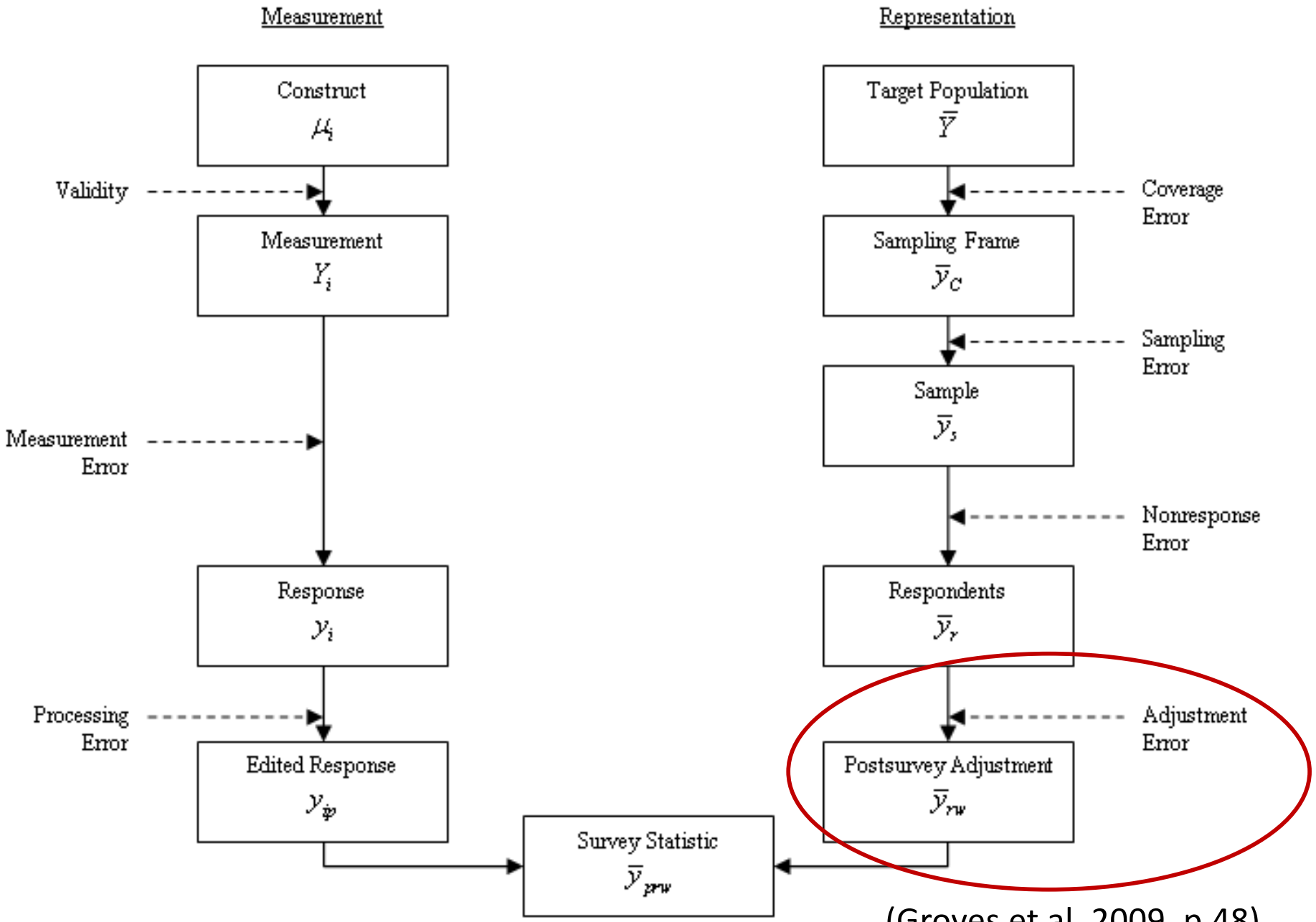
Gender	Age	X1	X2	X3	X4	X5	X6
Male	Young	1	1	0	1	0	0
Male	Middle	1	1	0	0	1	0
Male	Old	1	1	0	0	0	1
Female	Young	1	0	1	1	0	0
Female	Middle	1	0	1	0	1	0
Female	Old	1	0	1	0	0	1
Sample prop		-	.52	.48	.38	.33	.39
Population prop.		1.00	.51	.49	.43	.30	.27
Weights		1.00	-.033	.033	.161	-.095	-.066

- Weight for young female = 1 +.033+.161=1.185

GREG or Raking?

- GREG
  - Clear model
  - Variance of estimators can be computed
  - Can include continuous weighting variables (e.g. age)
  - Assumptions of regression model (normality, linearity)
  - Weights may become negative (causing computational problems)
- Raking
  - No model, no assumptions
  - No variance estimation
  - Only categorical variables

Total Survey Error (TSE) Framework



(Groves et al. 2009, p.48)

Combining weights

Designing weights can have several stages. For example:

1. The design weight is derived from the sampling frame (and the household selection during fieldwork).
2. Propensity-score weights are calculated using information from the sampling frame and the response indicator.
3. The data are post-stratified to known population distributions.

Weights can be combined : W1 * W2 * W3 = Wt

- Often only the total Weight included in public datasets
  - No detailed information on sampling design or nonresponse correlates
- Or svydesign for sampling plan, use weights for nonresponse and coverage

Next weeks

- Next week: designing weights
- In 2+3 weeks -> imputation (by Stef van Buuren)

- Assignment 2