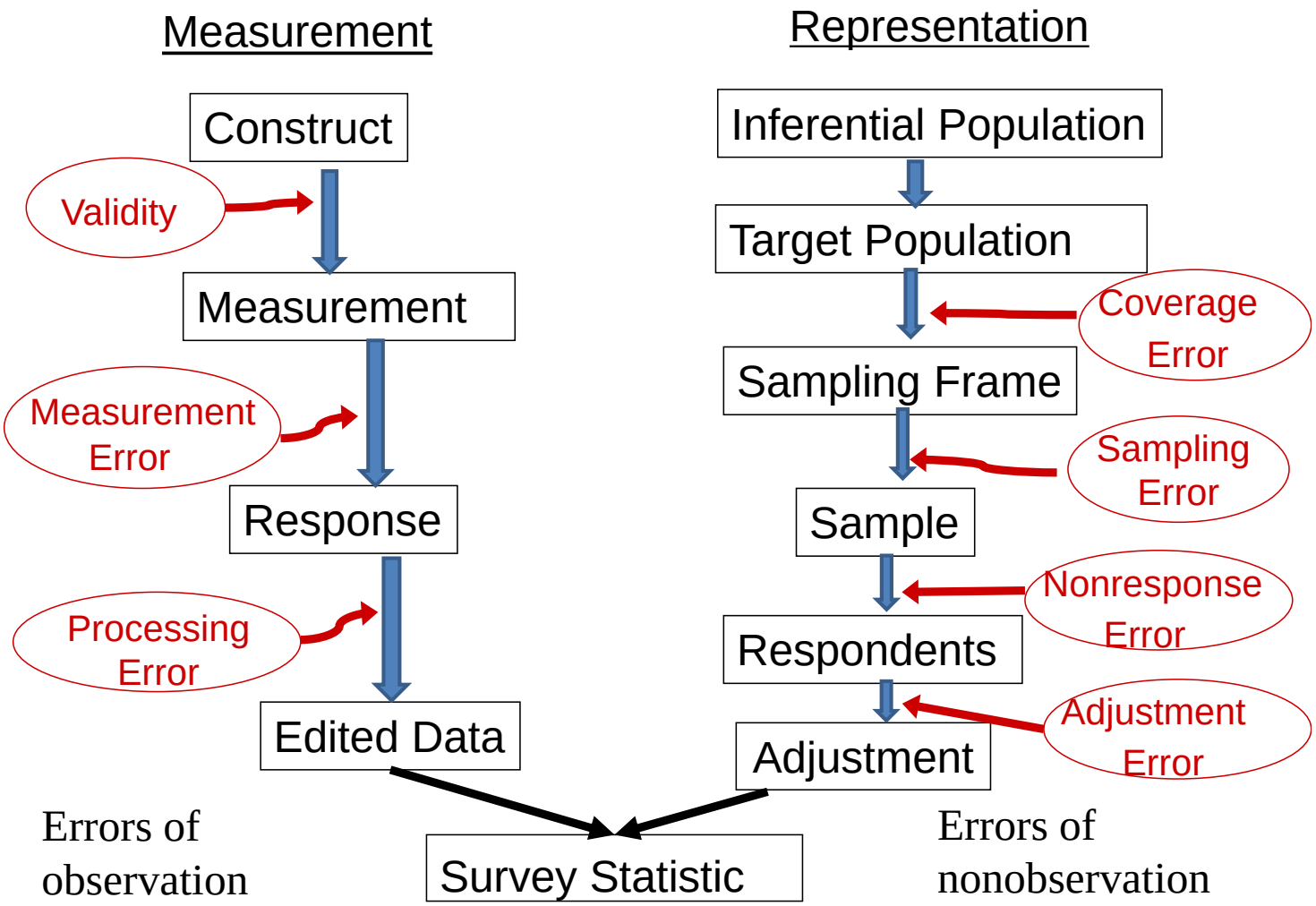
Reduction of Survey Nonobservation Errors Through Adaptive Sampling Design

Shin-Jung Julia Lee

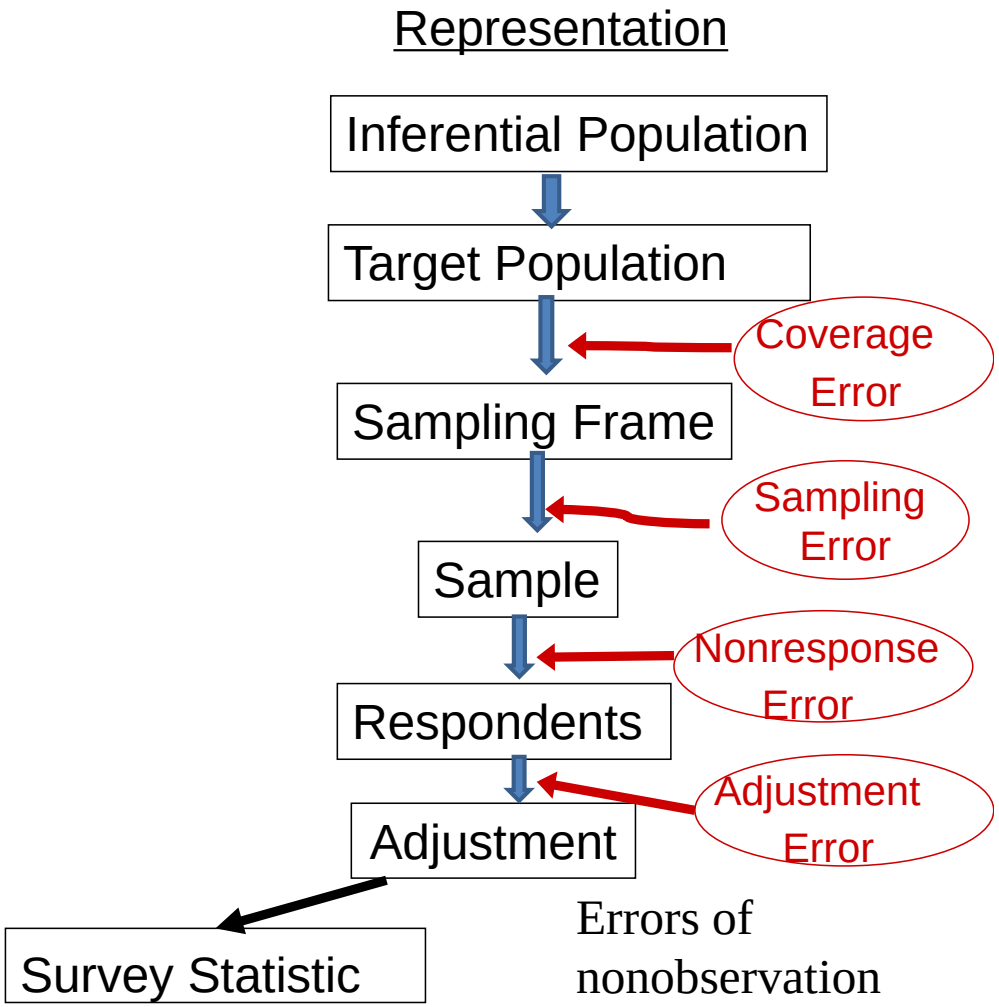
Oct 3, 2014

Total Survey Error Framework

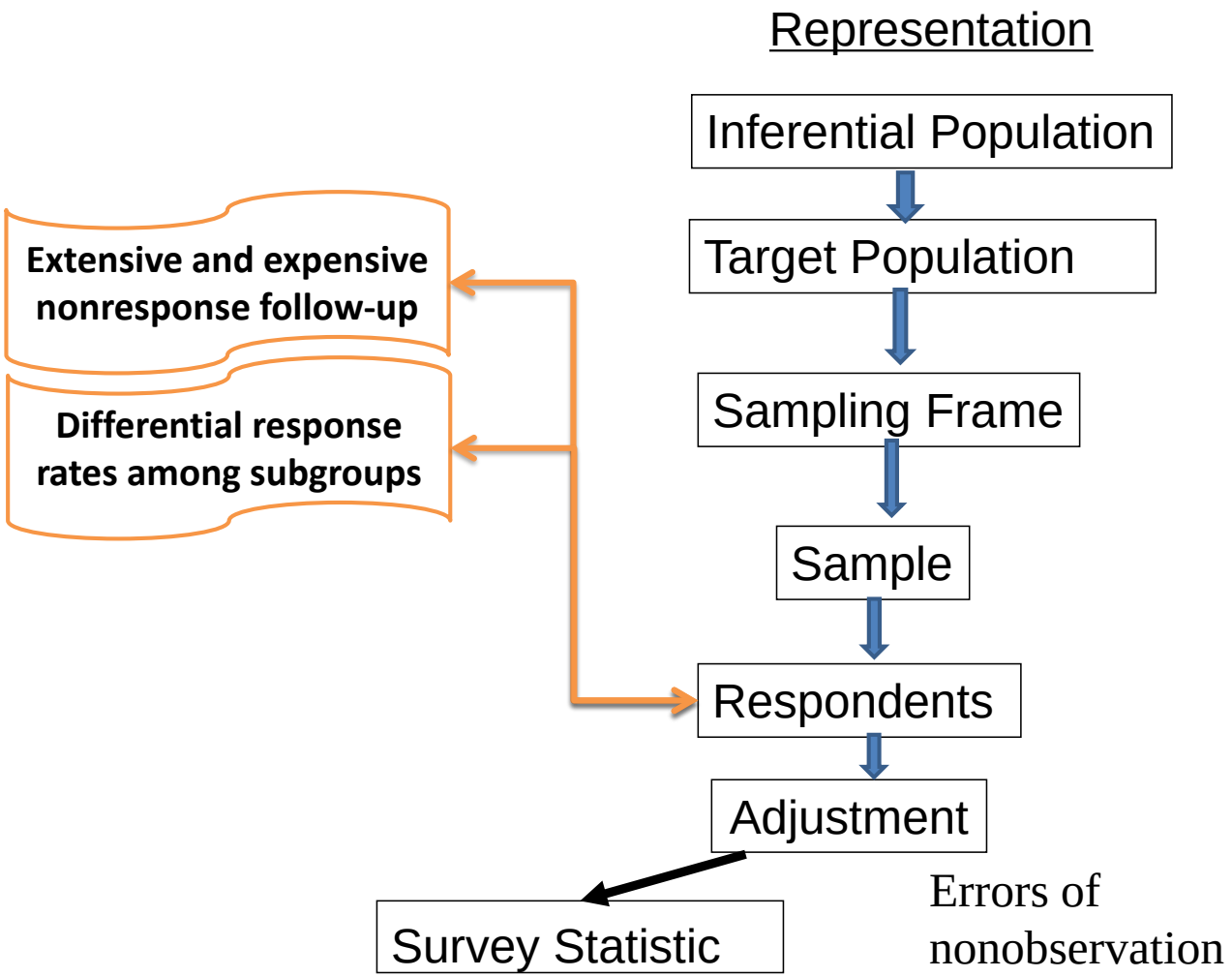


Reference: Groves, et al. (2004)

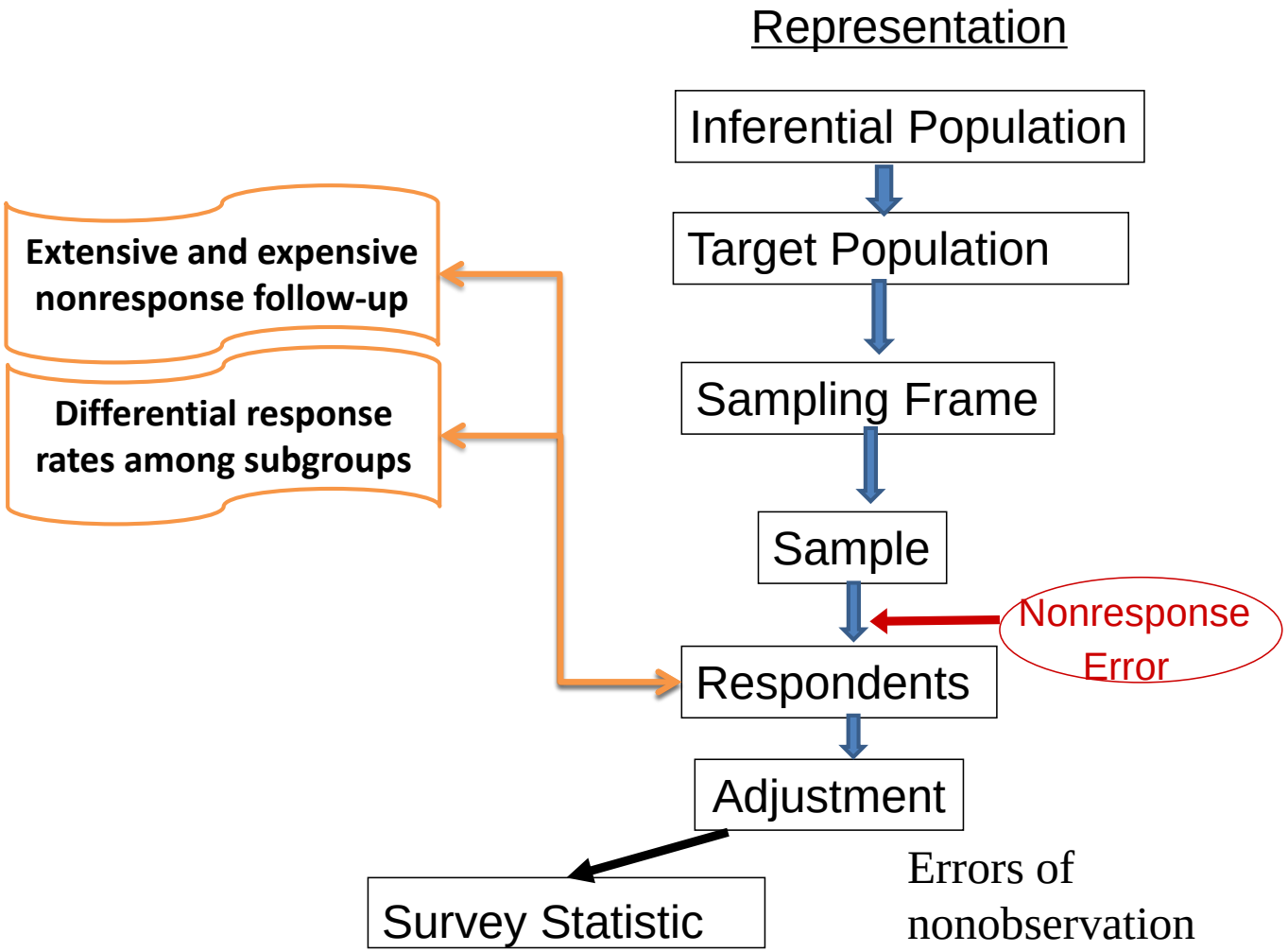
Errors Of Nonobservation



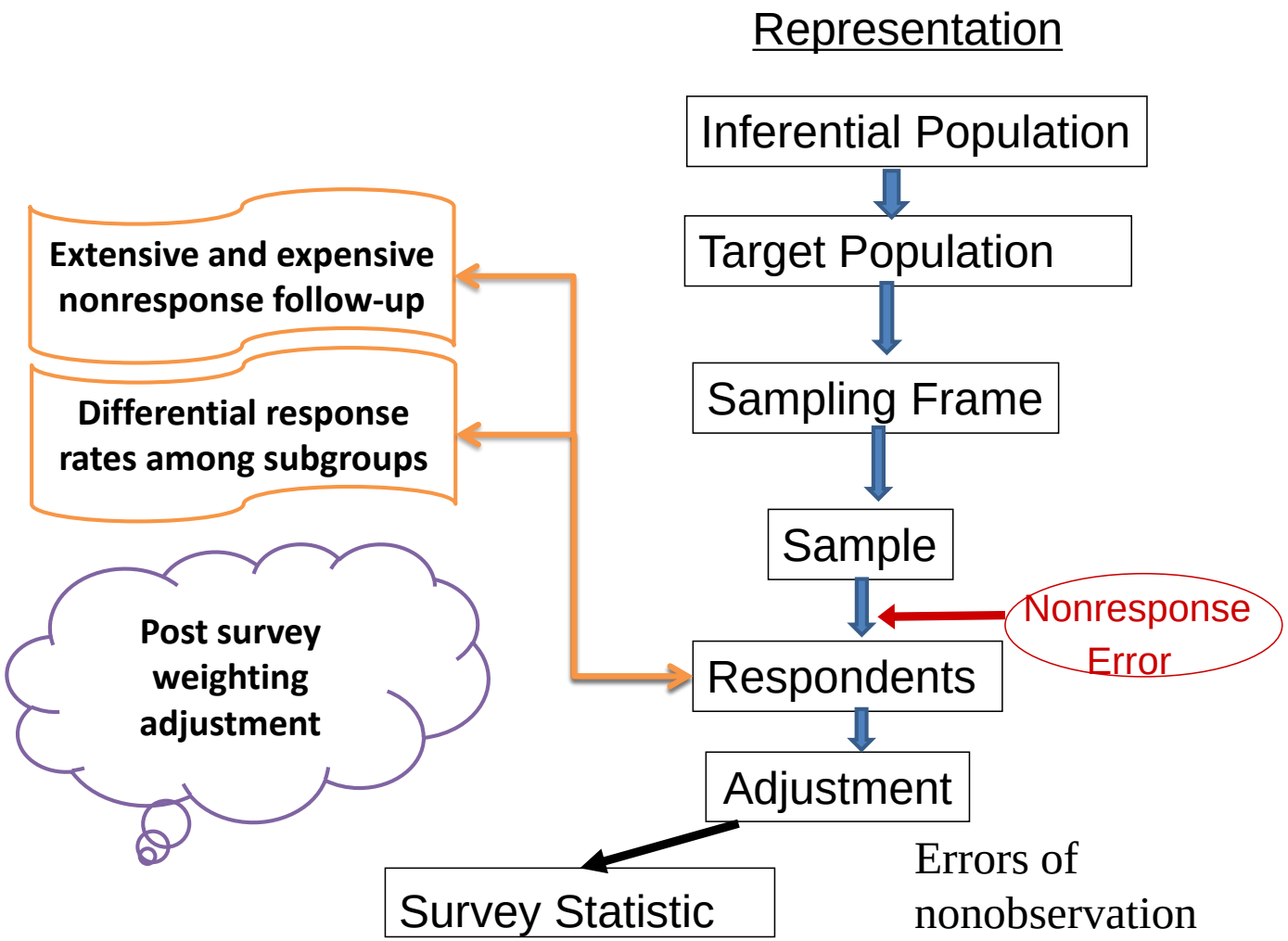
Nonresponse Error



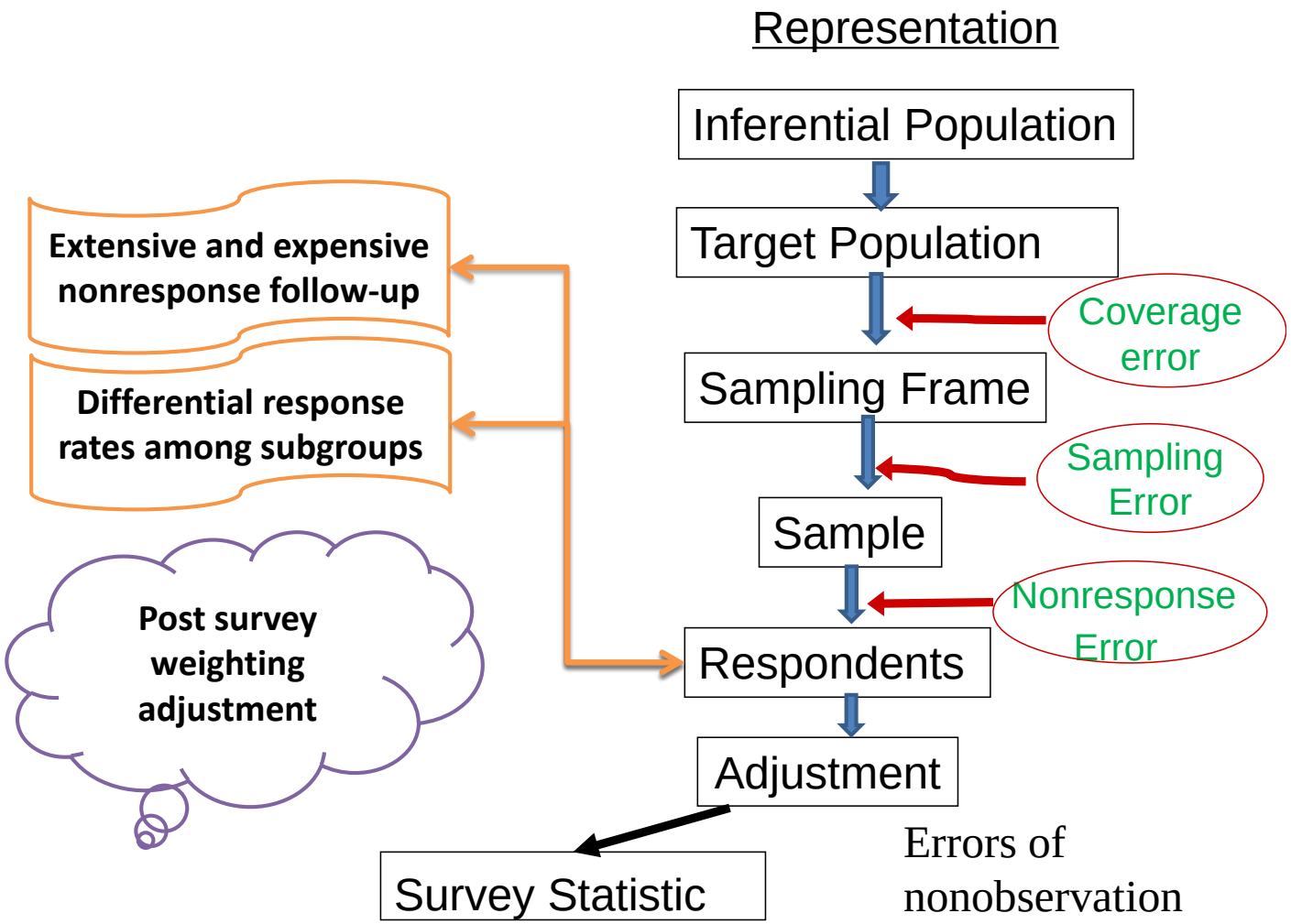
Nonresponse Error



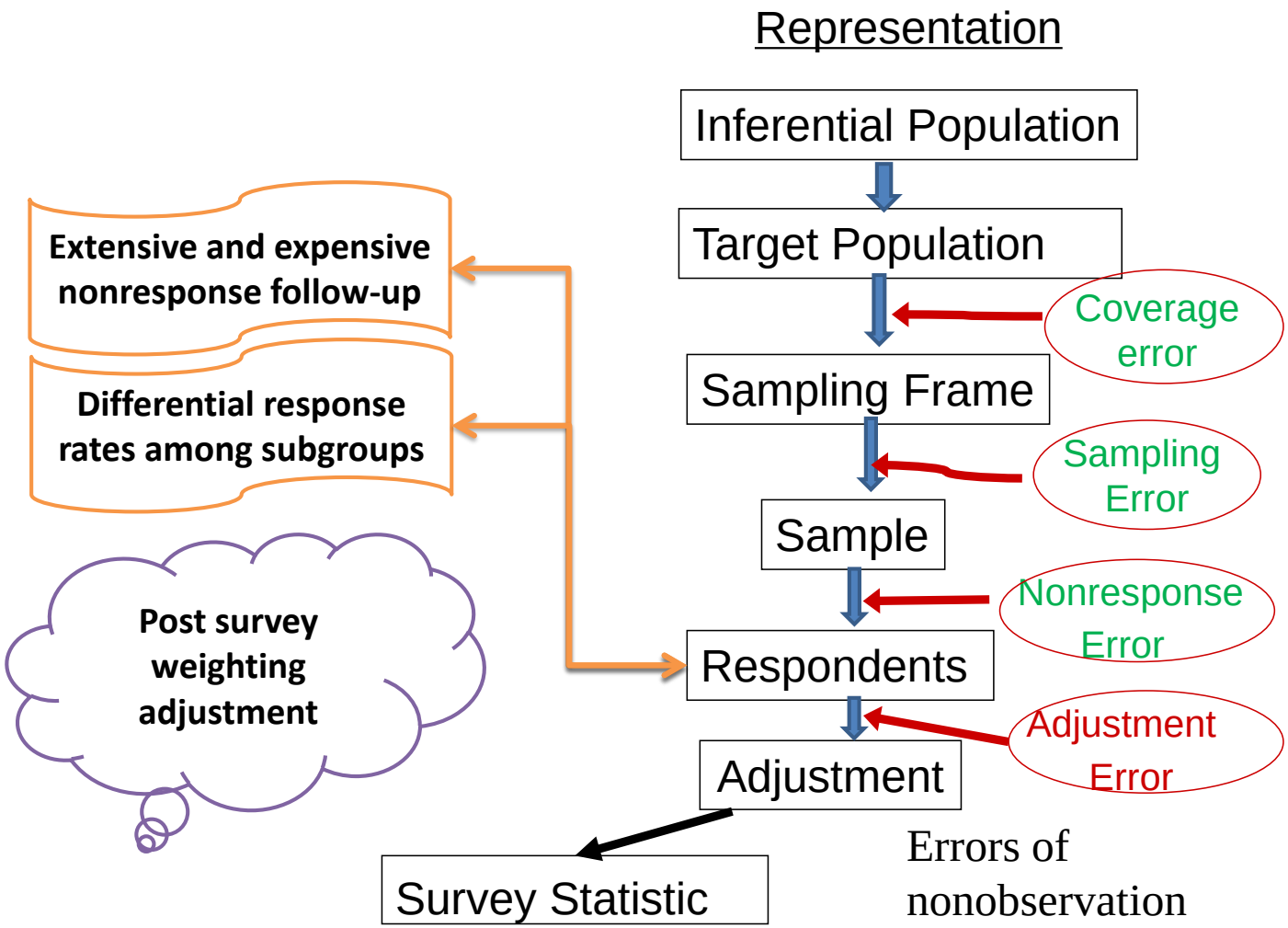
Common Practice



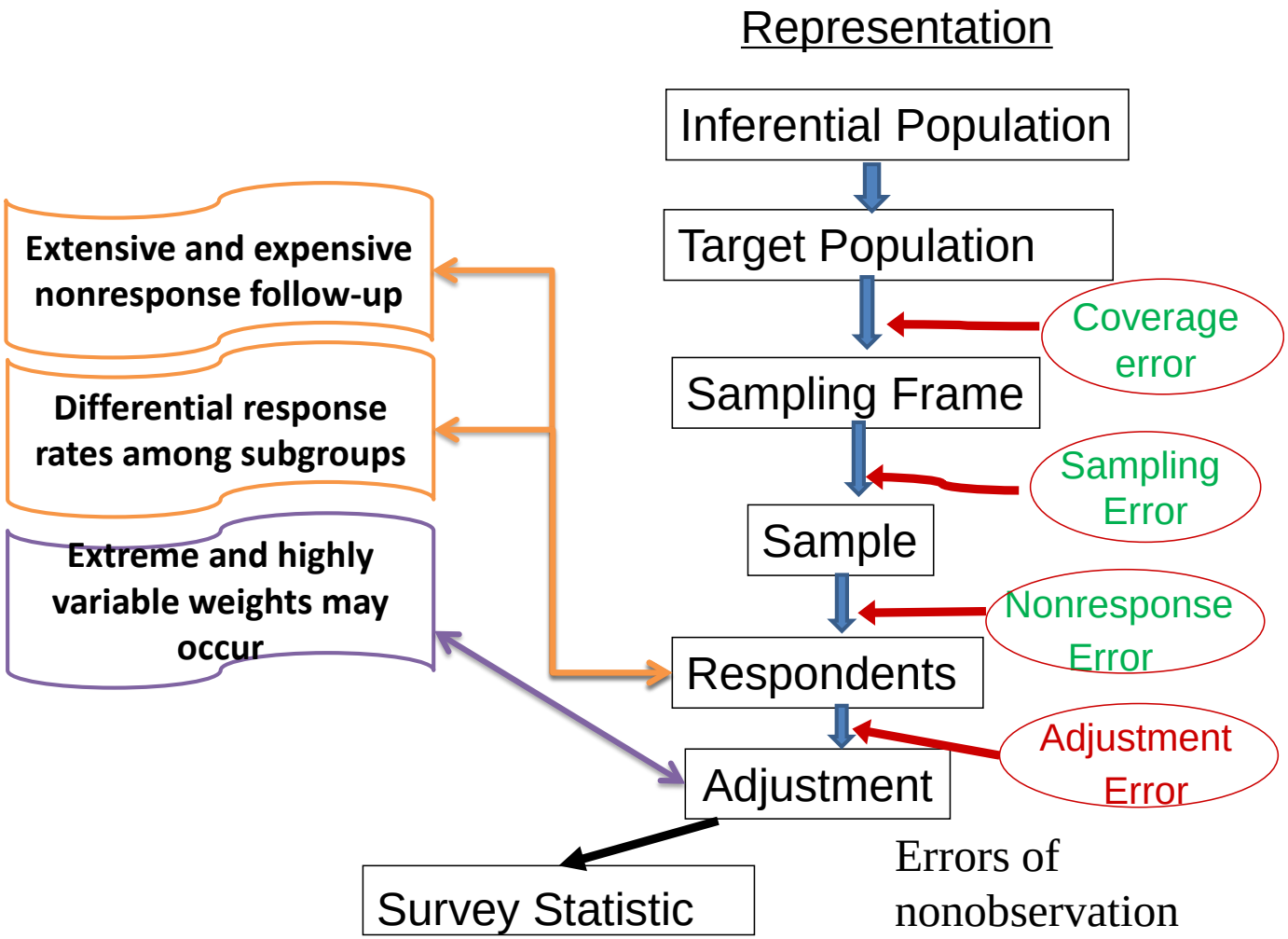
Weighting


Representation
Inferential Population
Target Population
Coverage error
Sampling Frame
Sampling Error
Sample
Nonresponse Error
Respondents
Adjustment
Survey Statistic
Errors of nonobservation
Extensive and expensive nonresponse follow-up
Differential response rates among subgroups
Post survey weighting adjustment

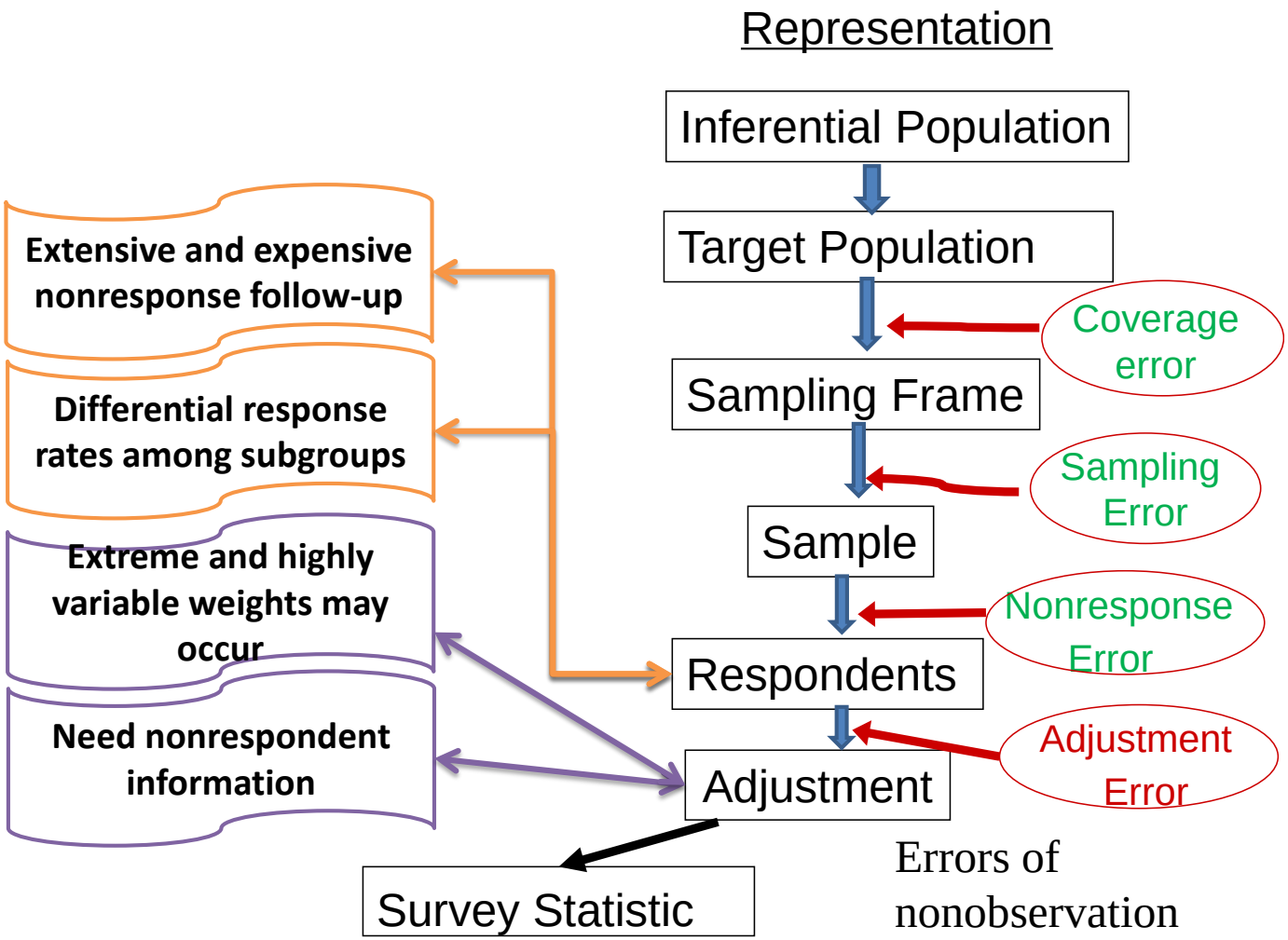
Adjustment Error



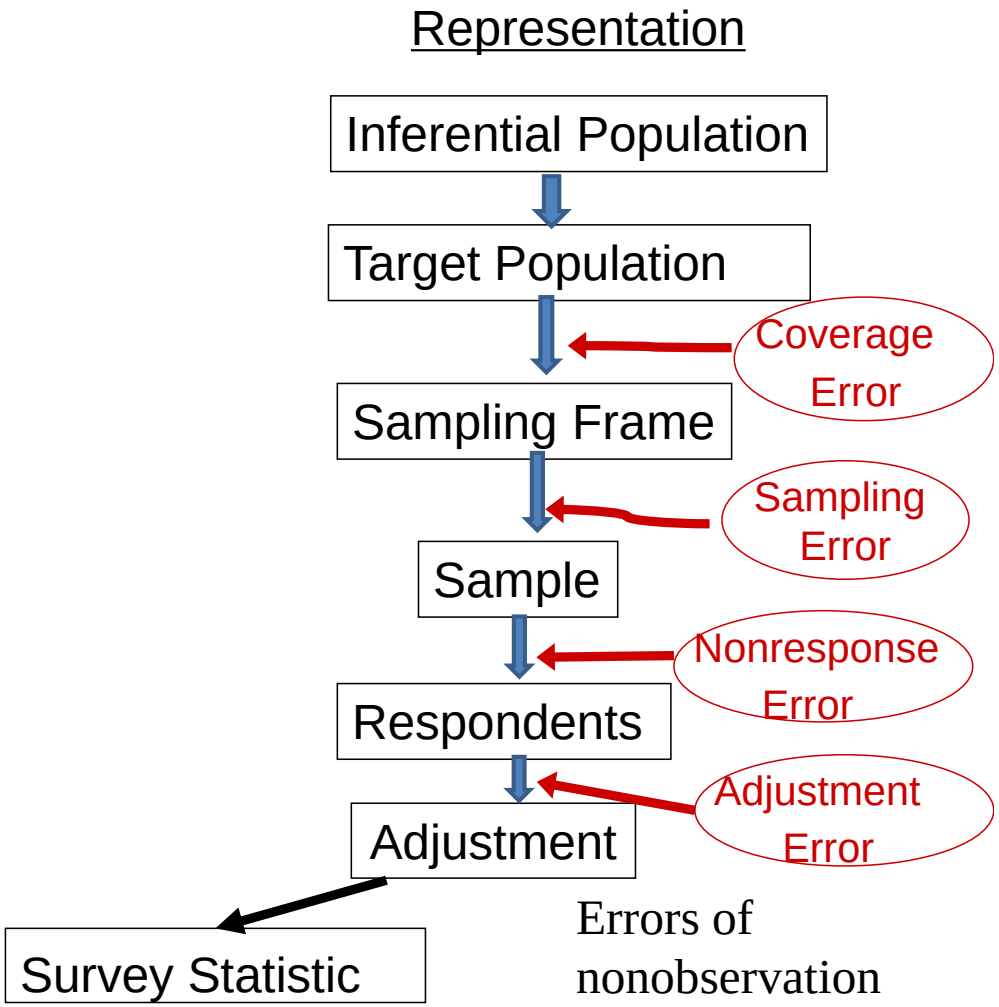
Adjustment Error



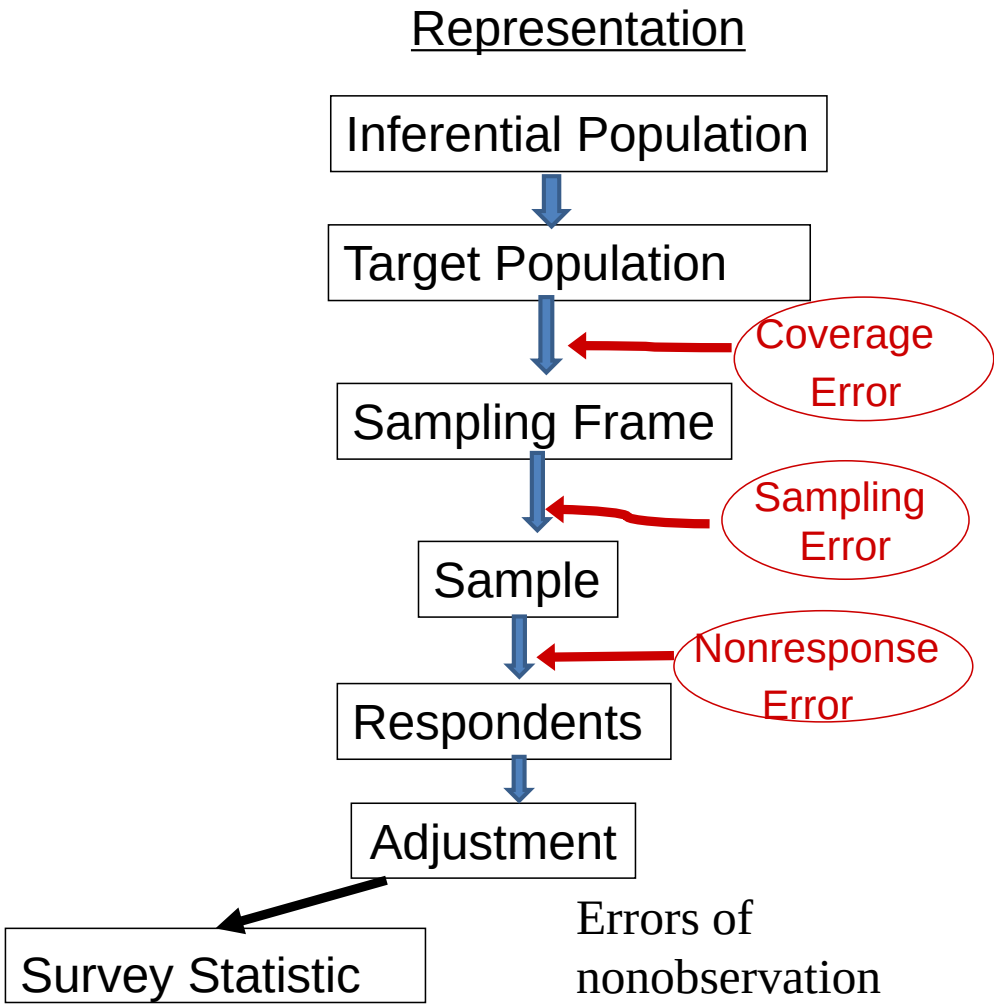
Adjustment Error



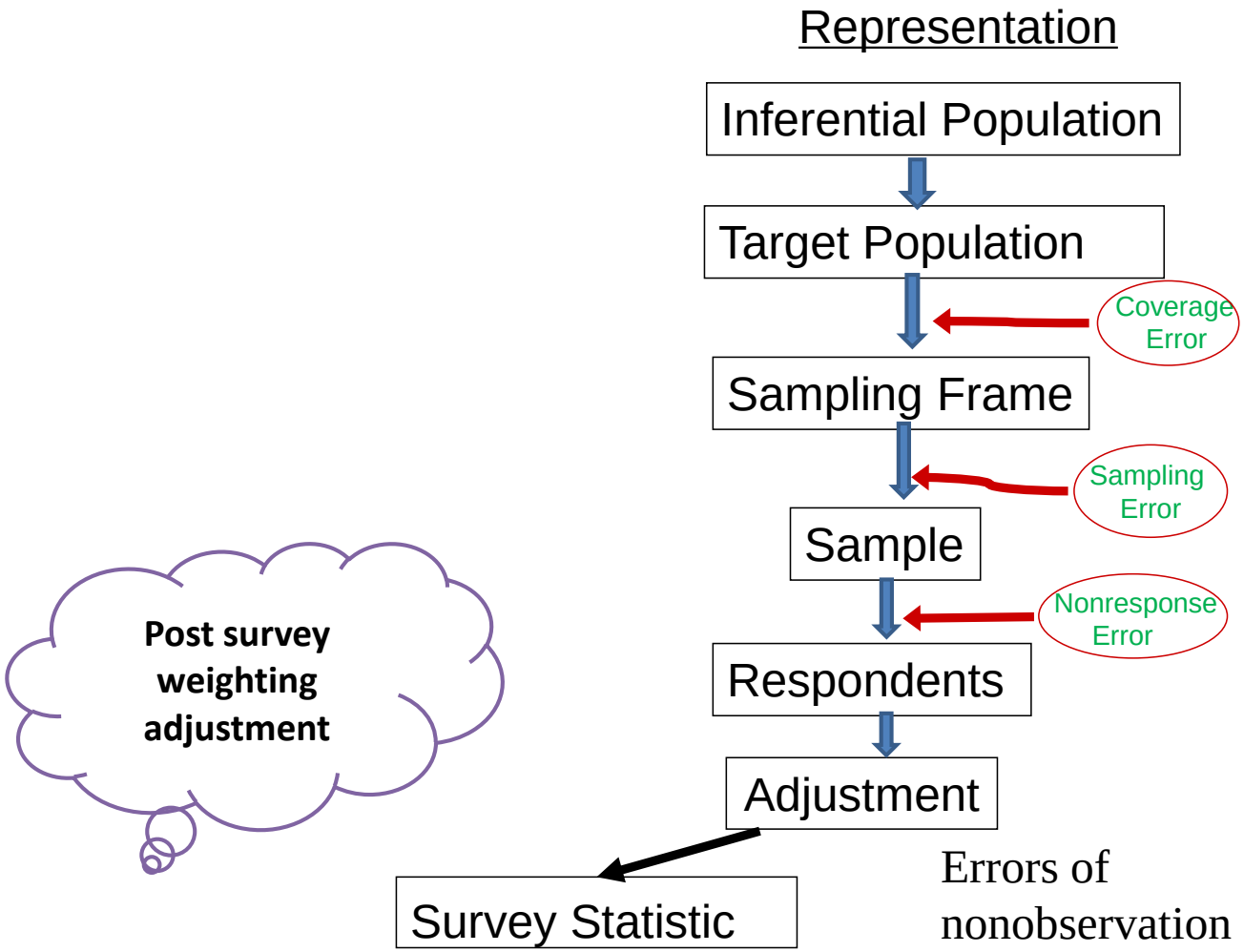
Errors of Nonobservation



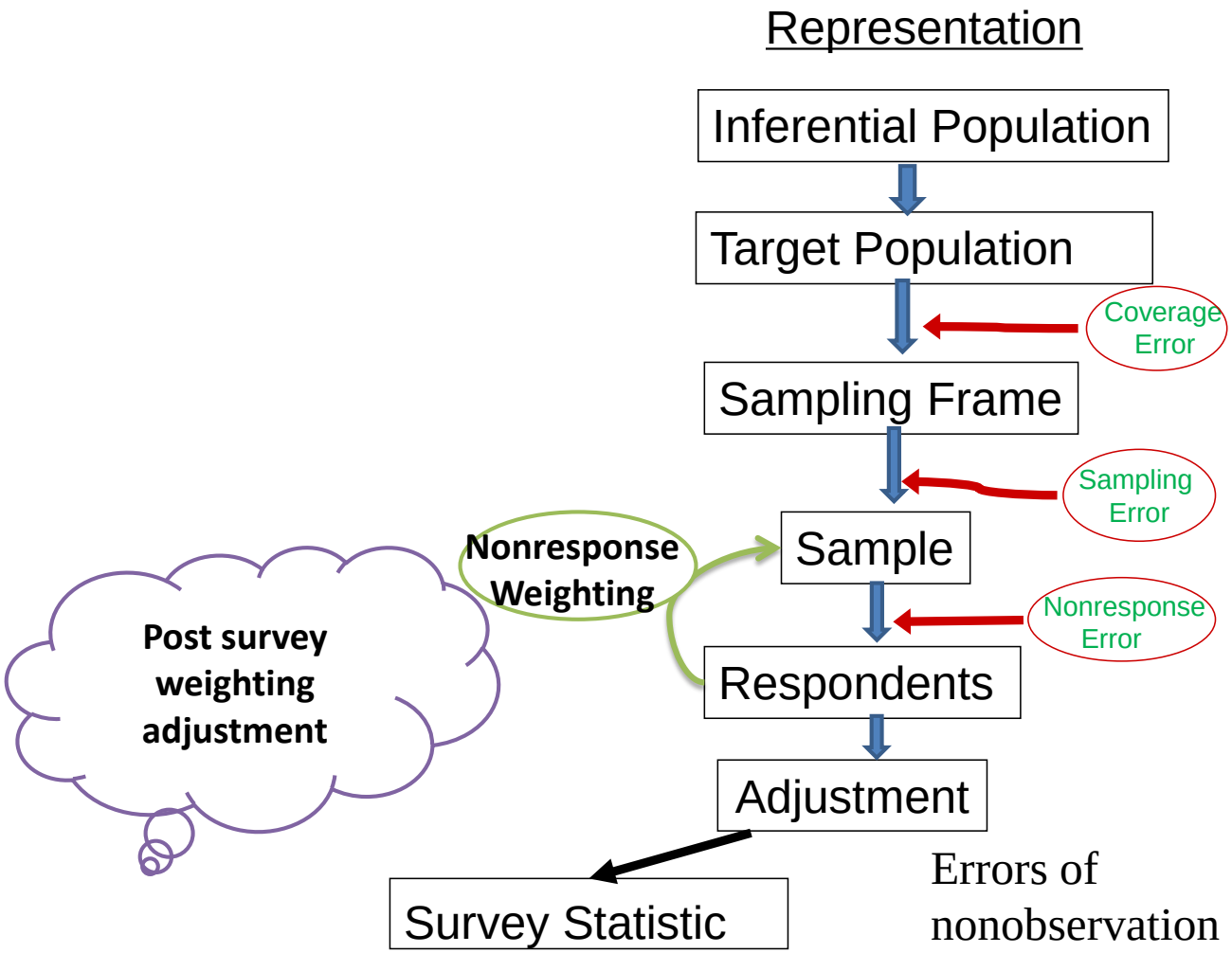
Before The Adjustment Error



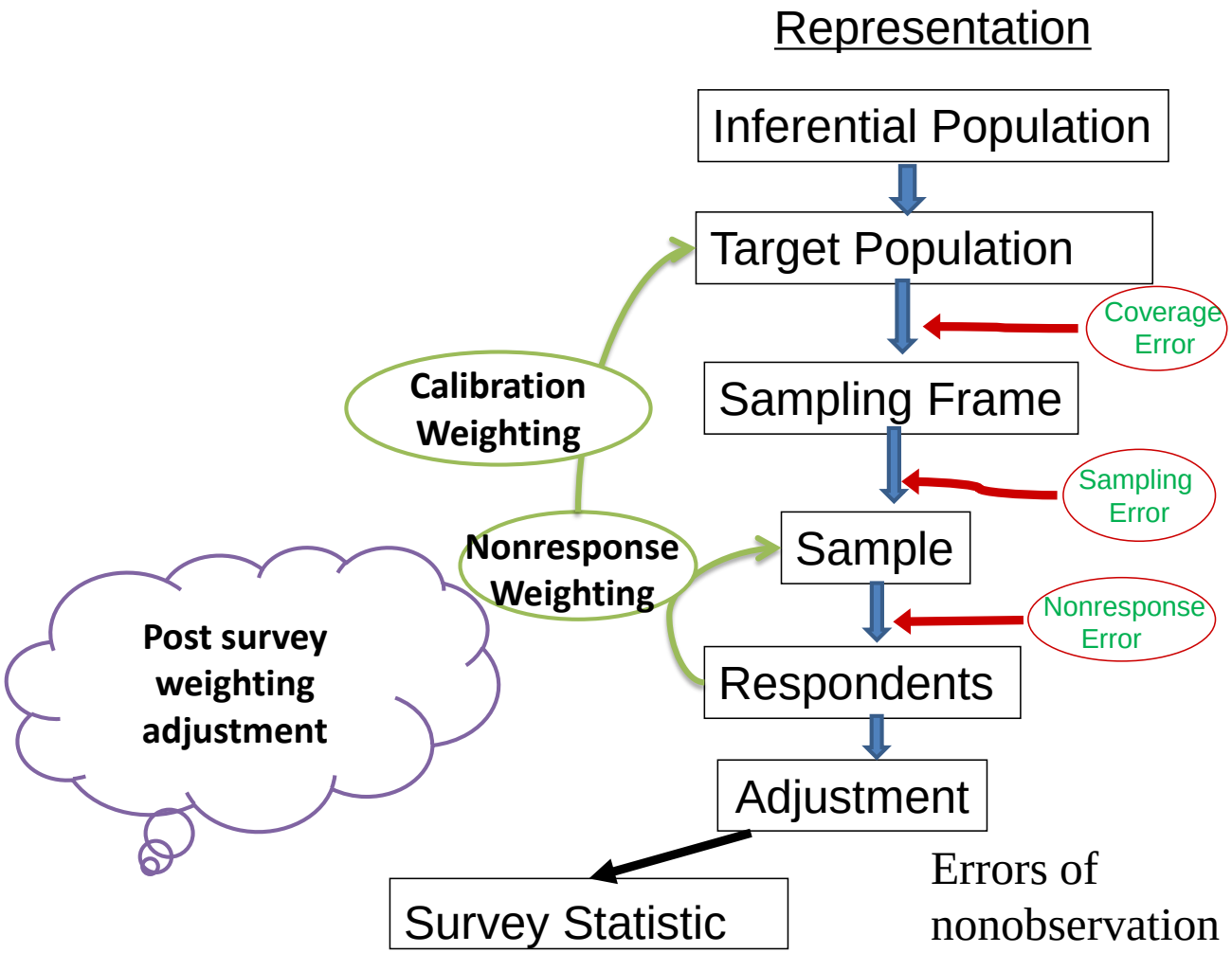
Before The Adjustment Error



Nonresponse Weighting



Calibration Weighting



Information Used For Weighting

Calibration Weighting

- **Census**
- **Benchmark**

Nonresponse Weighting

- **Frame information**
- **Contextual information**
- **Paradata**

Post survey weighting adjustment

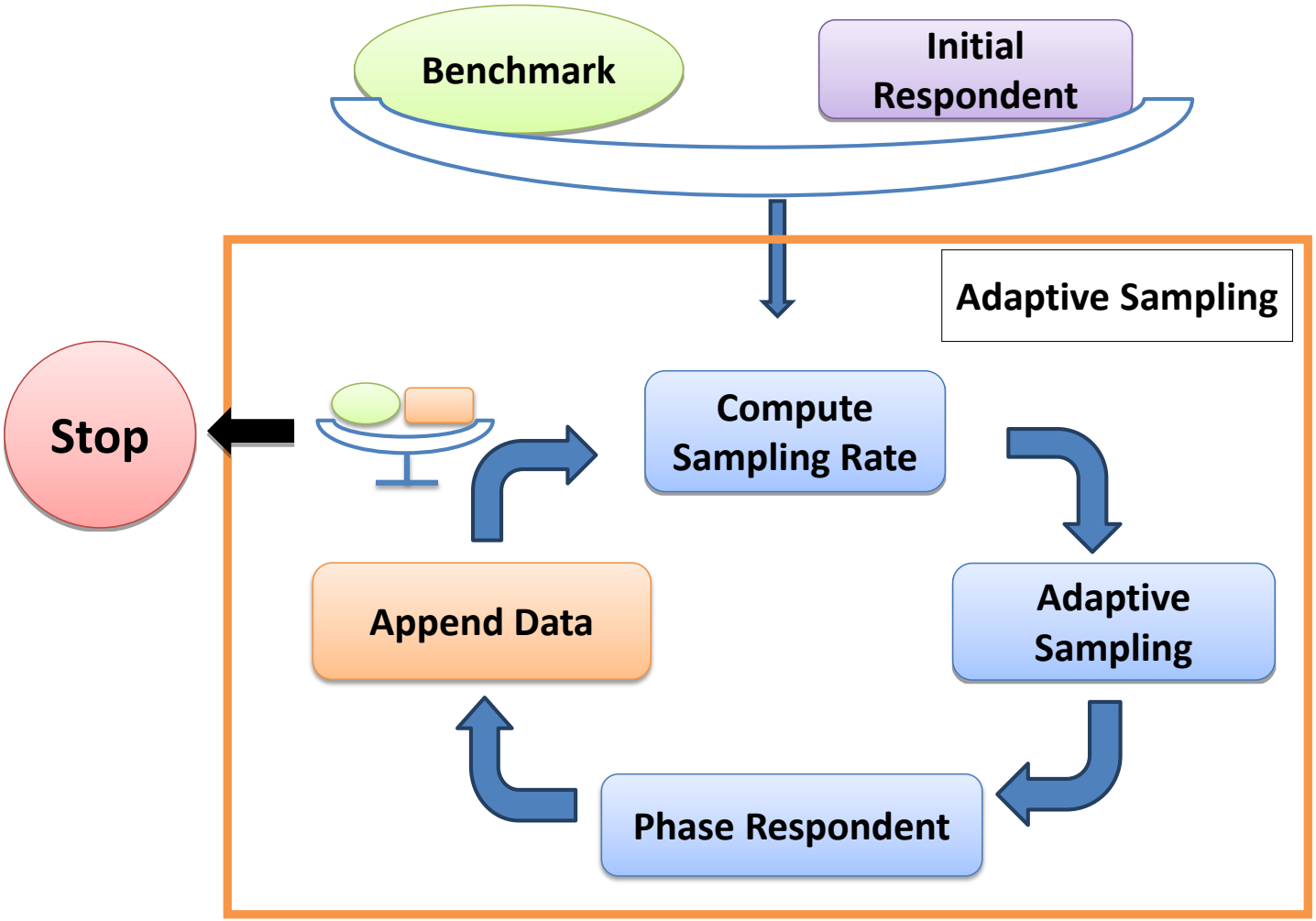
Alternative Strategy

Information used for post-survey adjustment

is used at

the **sampling design and data collection stage**

Alternative Strategy: Adaptive Sampling



Alternative Strategy: Adaptive Sampling

Improve sampling to obtain representative respondents:

- Actively control sampling at design stage and data collection stage by adaptively benchmarking to target popualtion.

- Effectively use frame data, contextual data, paradata, and benchmark information to improve respondent representativeness.

Objectives And Tools

Objective: Reduce errors of nonobservation

- Obtain a more representative respondent pool
- Minimize post-survey adjustments

Tools:

- Adaptive sampling strategy
- An external high quality benchmark

Terminology

Terminology:

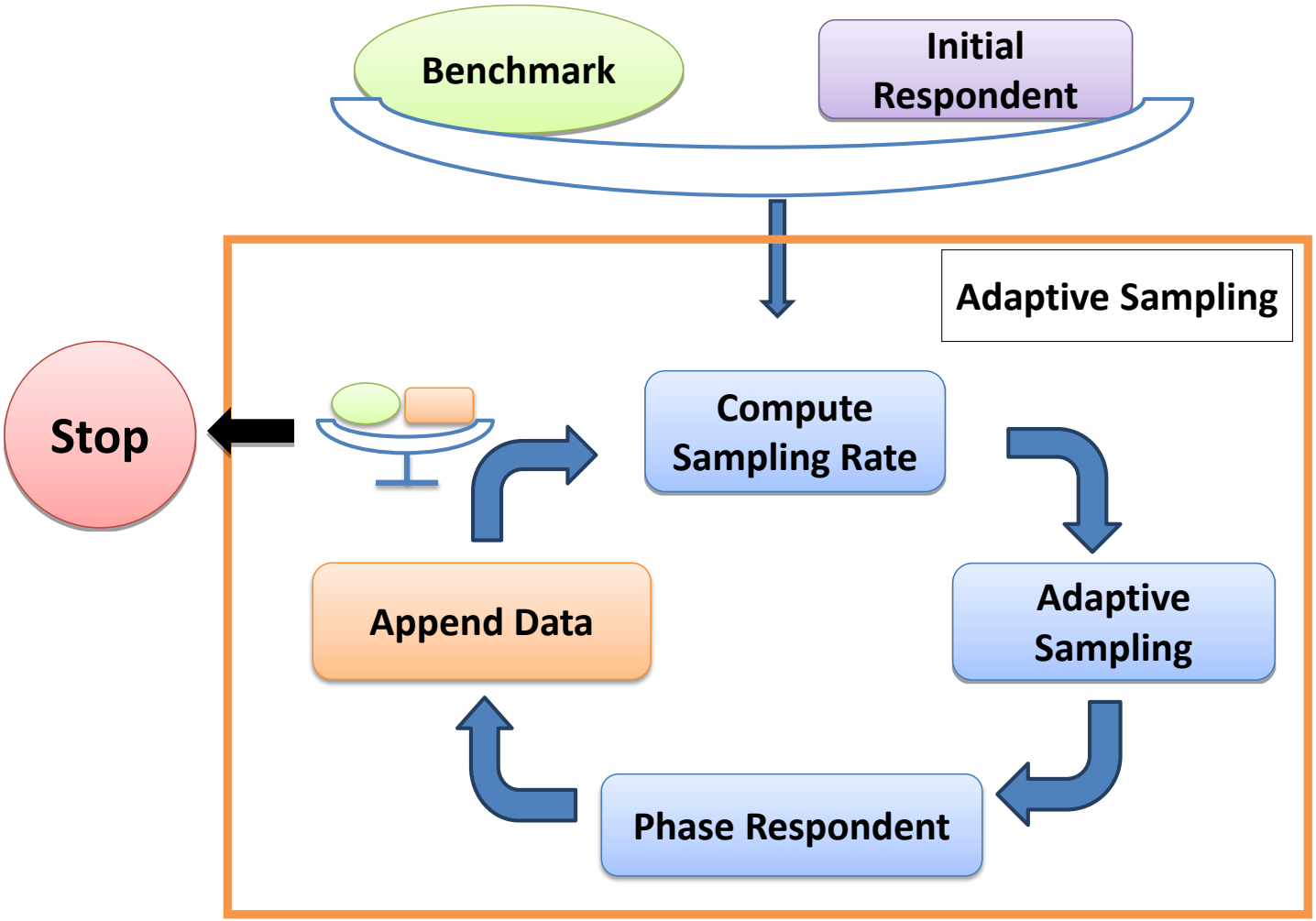
- Benchmark survey: census or a survey capturing the target population, such as the National Health Interview Survey (NHIS).
- Focal Survey: survey that you are conducting, e.g. BRFSS in this talk.

Sampling Rate

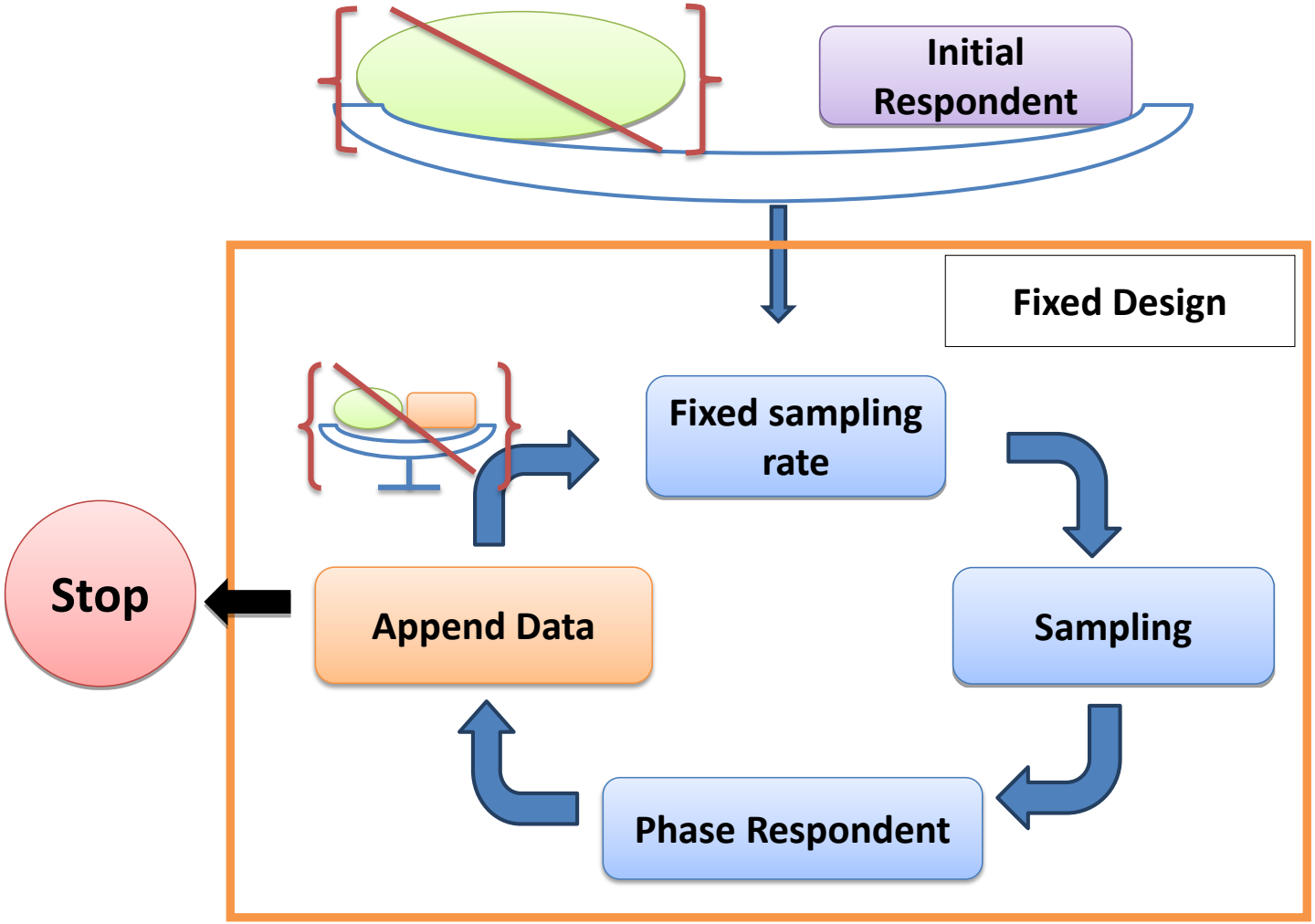
Adaptive Sampling Procedures:

1. Model the origin of each data point (1=benchmark, 0= focal survey) in terms of covariates.
2. Compute ratio of propensity score density between benchmark and focal survey.
3. Sample next phase subjects using the proposed sampling rate.

Adaptive Sampling



Traditional Sampling



Simulation Data

Data set:

- Benchmark: NHIS
- Focal Survey: BRFSS

Variables:

- Auxiliary variables: Race, Sex, Education and Age
- Survey variable: Body mass index

Simulation Data

Nonresponse model:
Race, Sex, Education and Age

Response rate:
Mimic BRFSS response data assuming NHIS is true population

Weighting methods:

- Design weights
- Nonresponse weighting
- Calibration

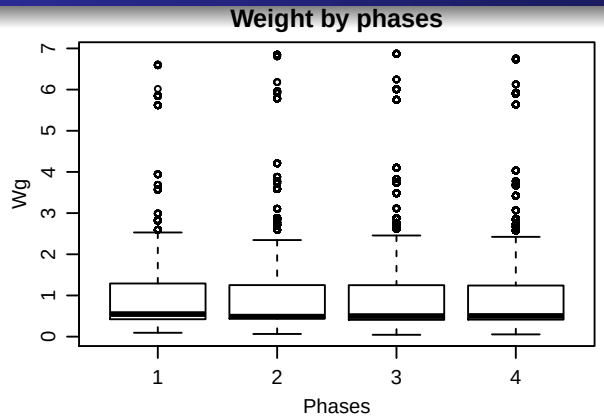
Weighting variables:
Race, Sex, Education and **Age groups**

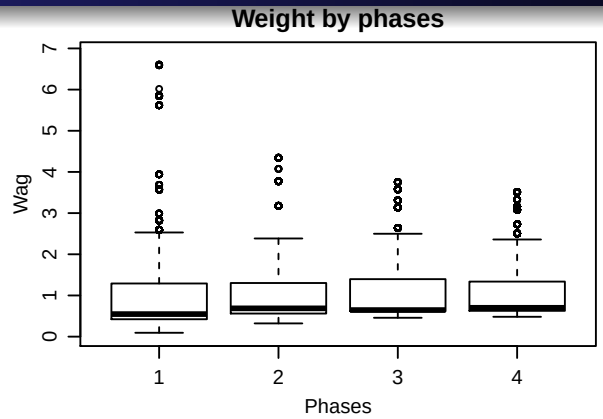
Simulation Experiments

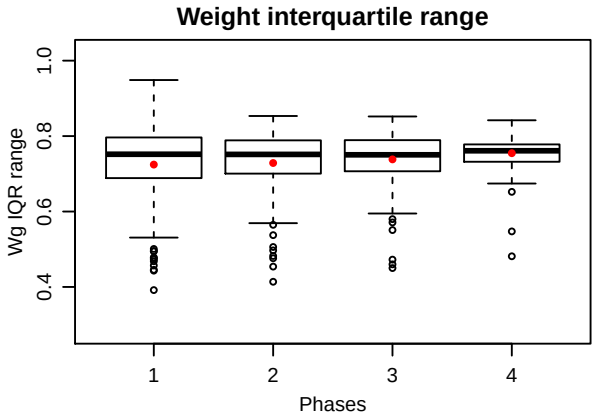
Adaptive sampling v.s. fixed sampling design

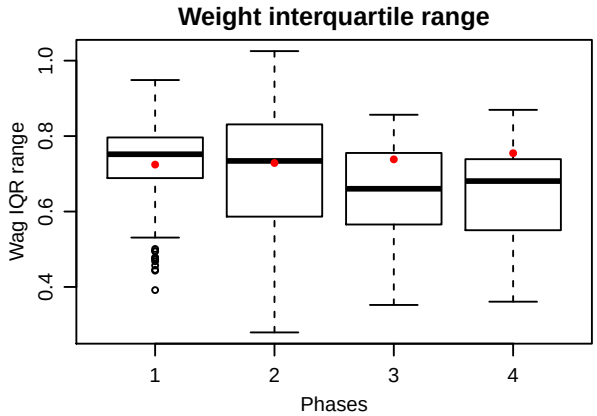
- Four phases of data collection
- Identical nonresponse model
- Identical weighting procedures and variables
- Identical first phase respondents
- Scenario 1: Small coverage error
- Scenario 2: Large coverage error

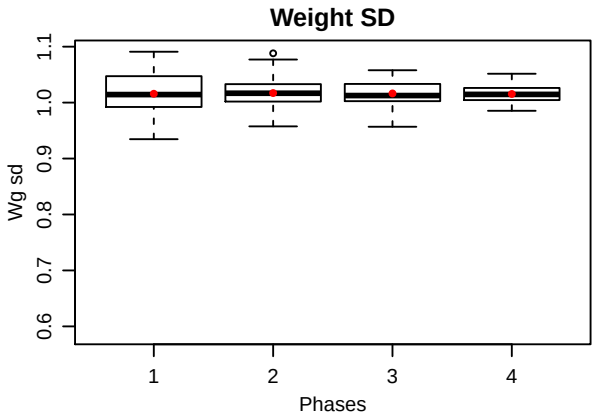
Final weight by phase

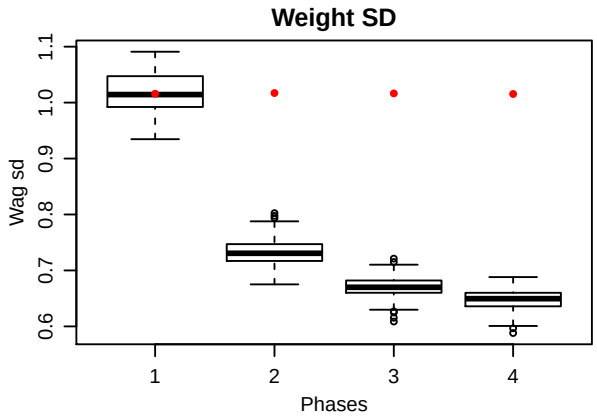




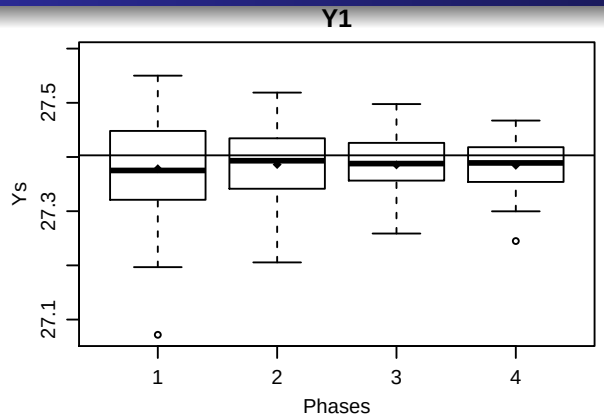


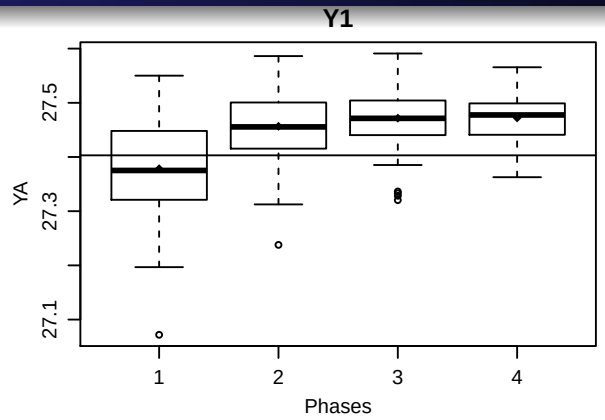


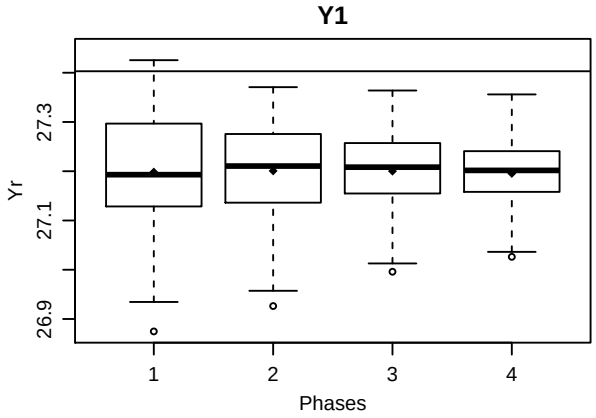


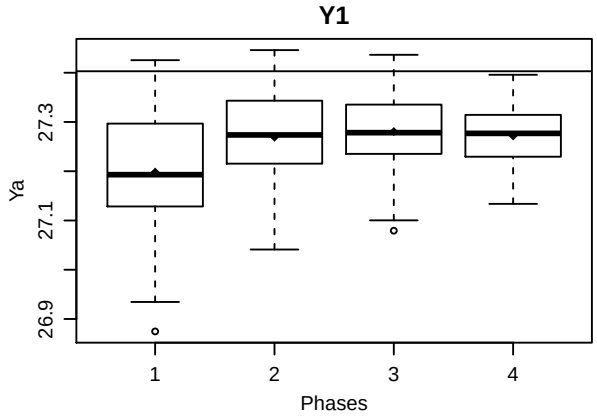


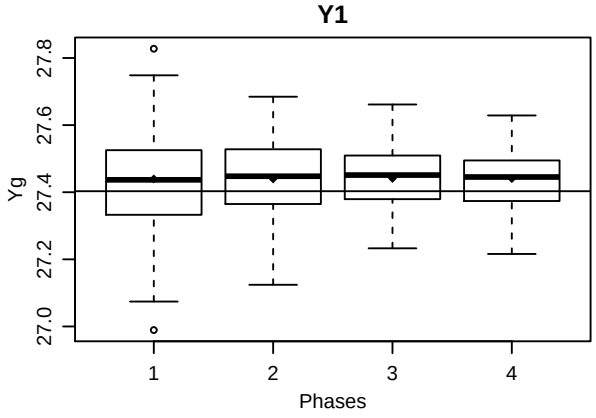
Scenario 1 BMI mean by phase

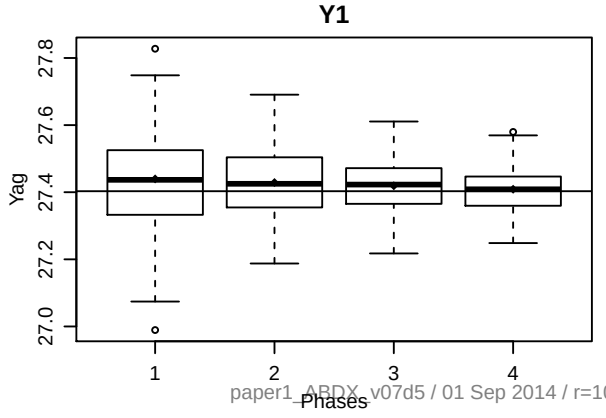




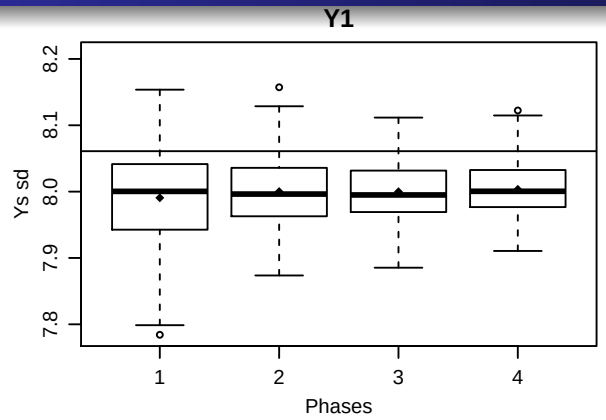


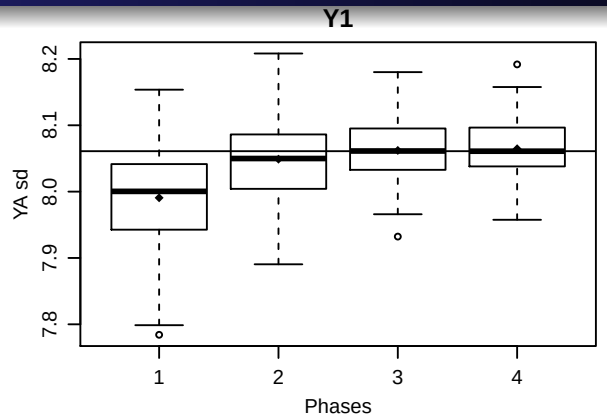


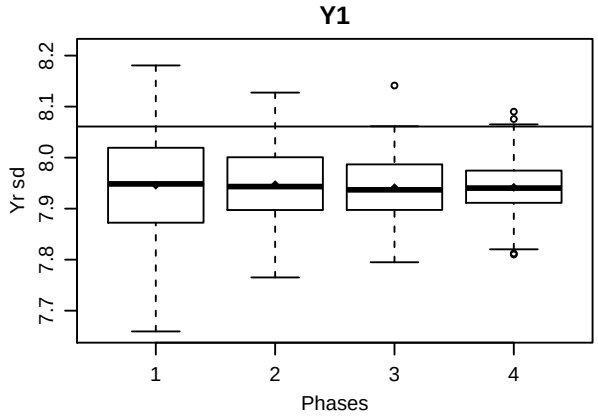


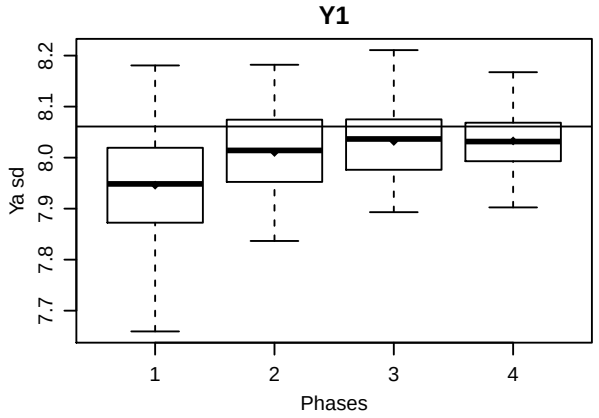


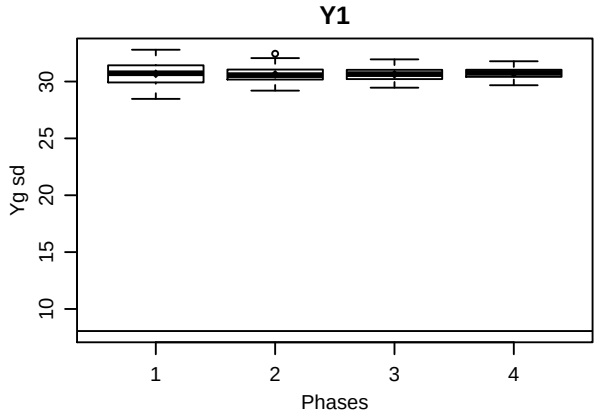
Scenario 1 BMI SD by phase

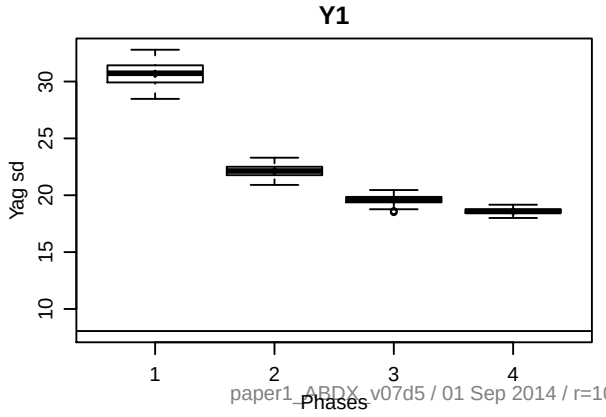






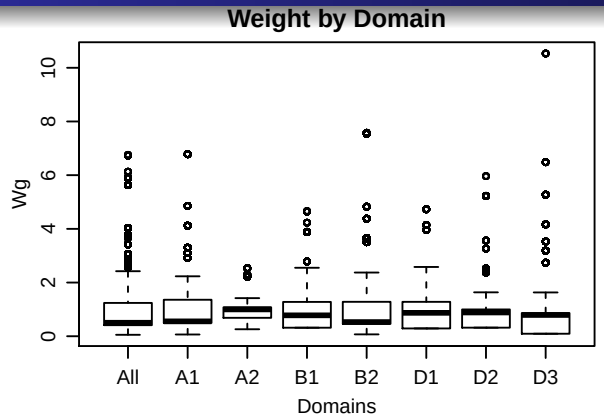


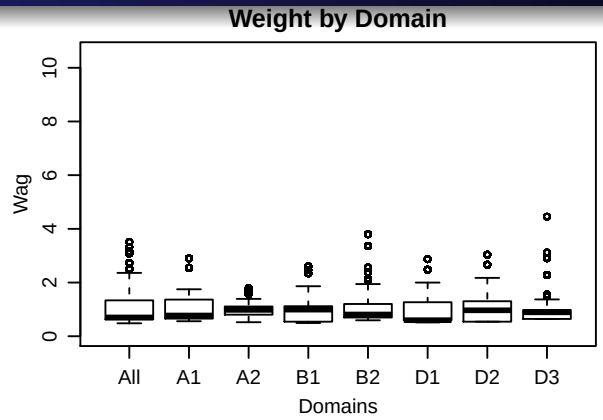


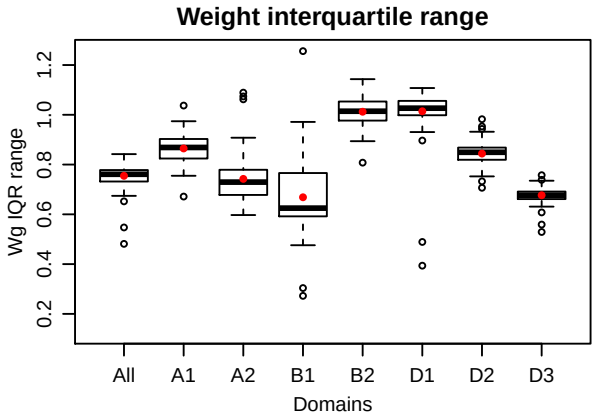


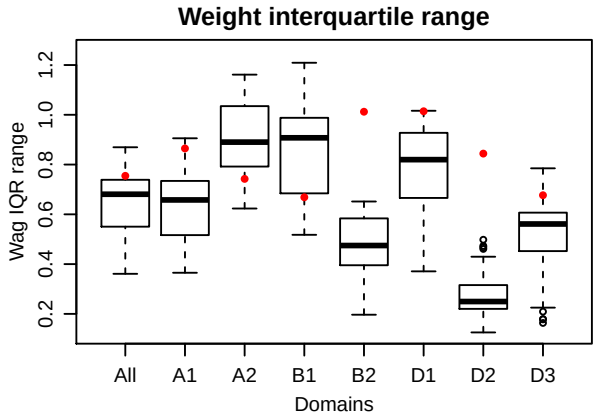
paper1_ABDX_v07d5 / 01 Sep 2014 / r=100

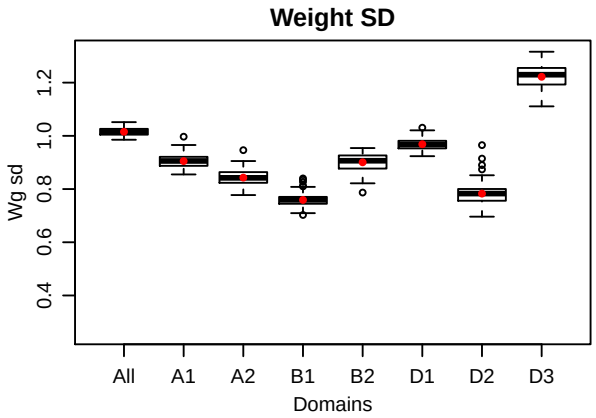
Final weight by domain

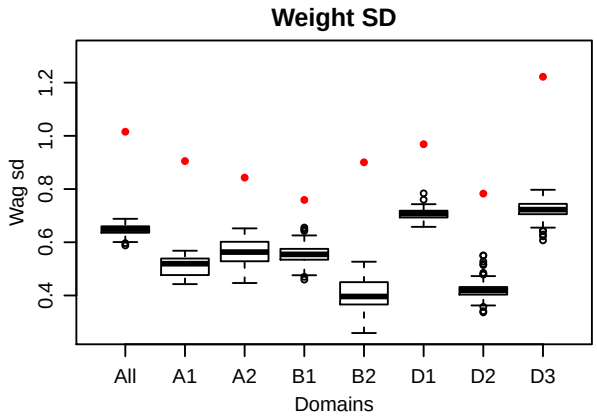




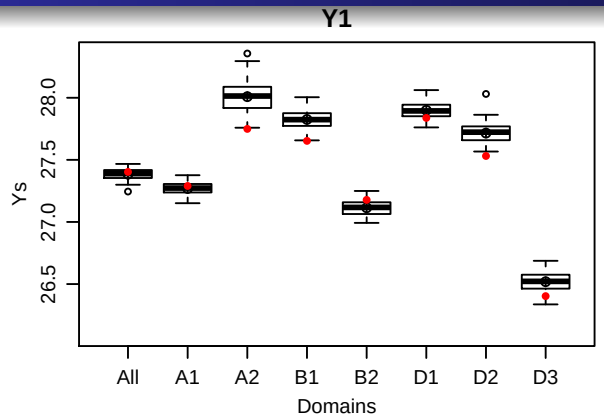


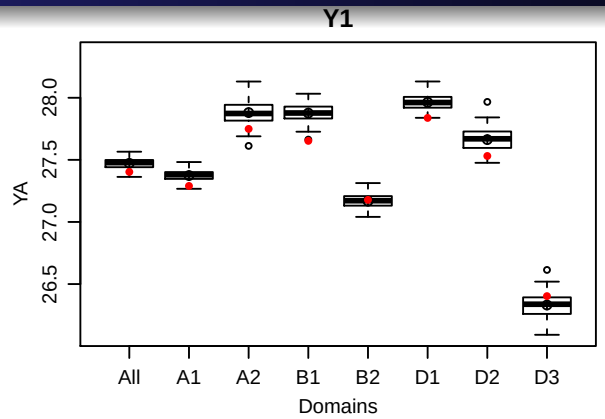


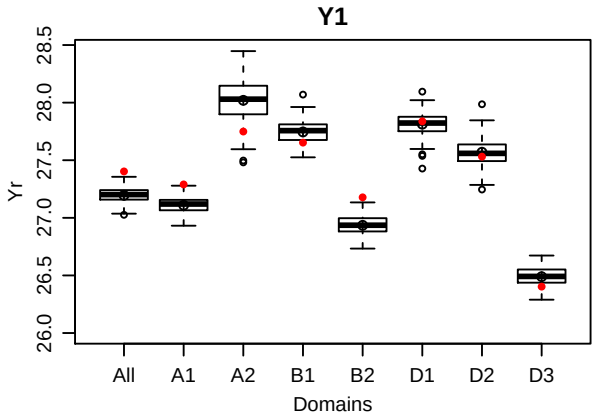


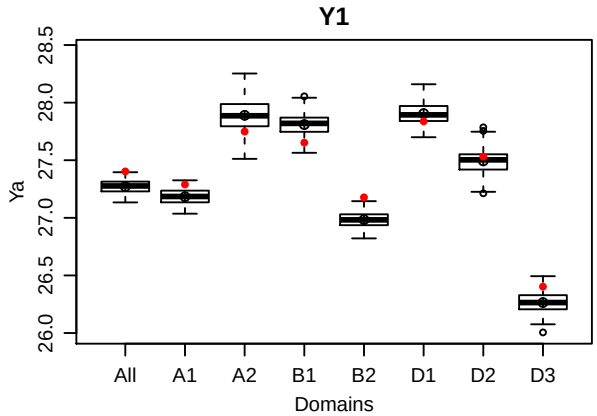


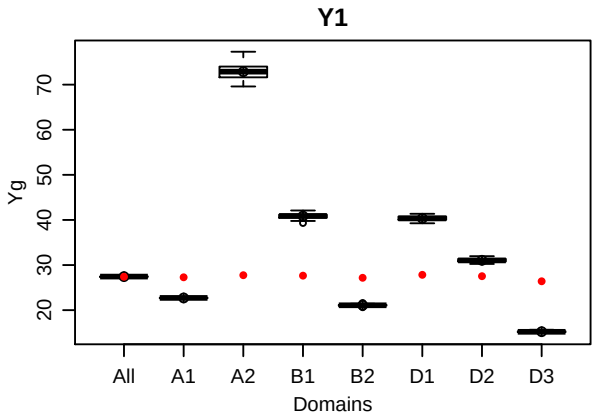
Scenario 1 BMI mean by domain

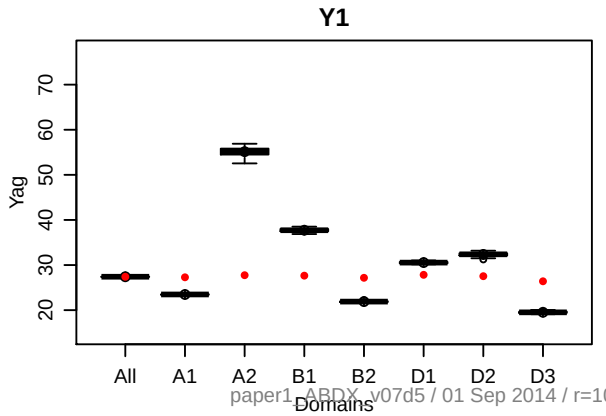




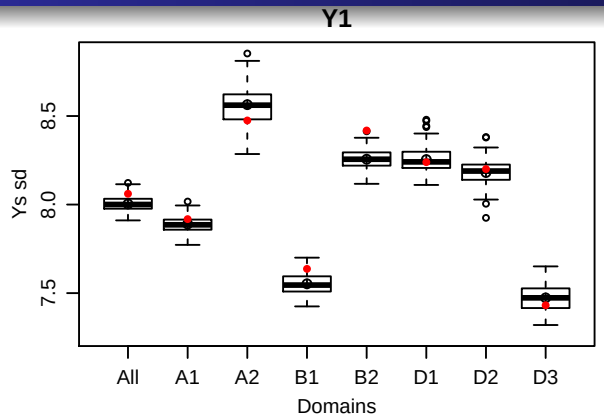


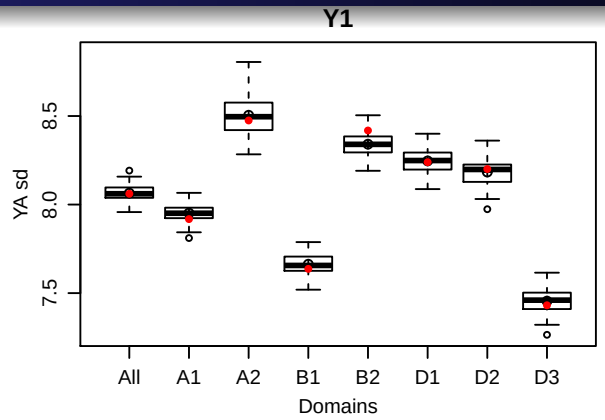


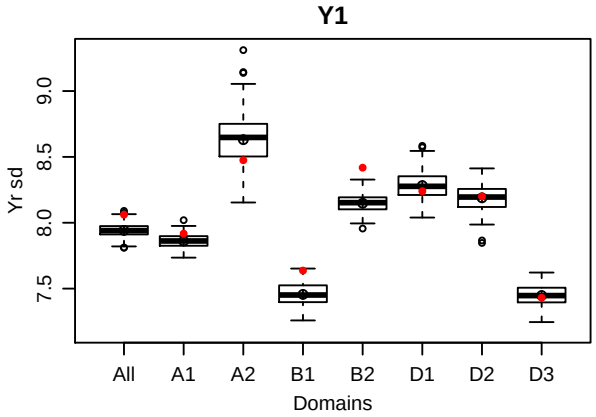


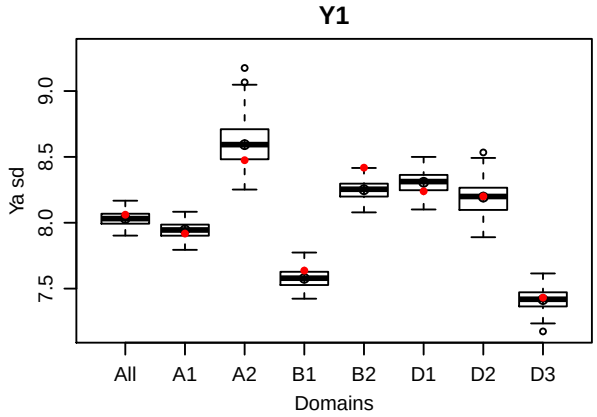


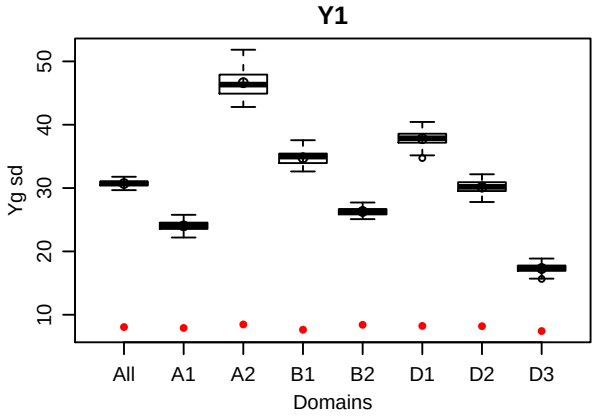
Scenario 1 BMI SD by domain

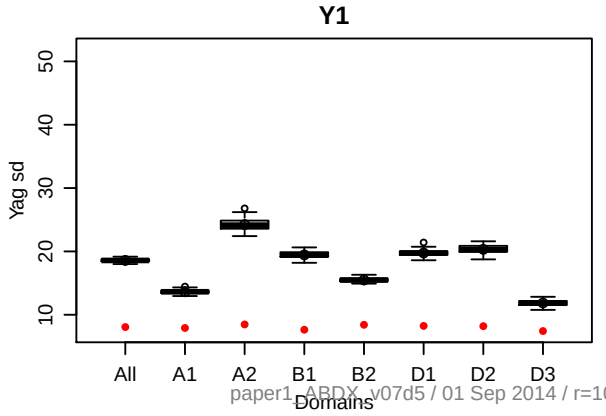




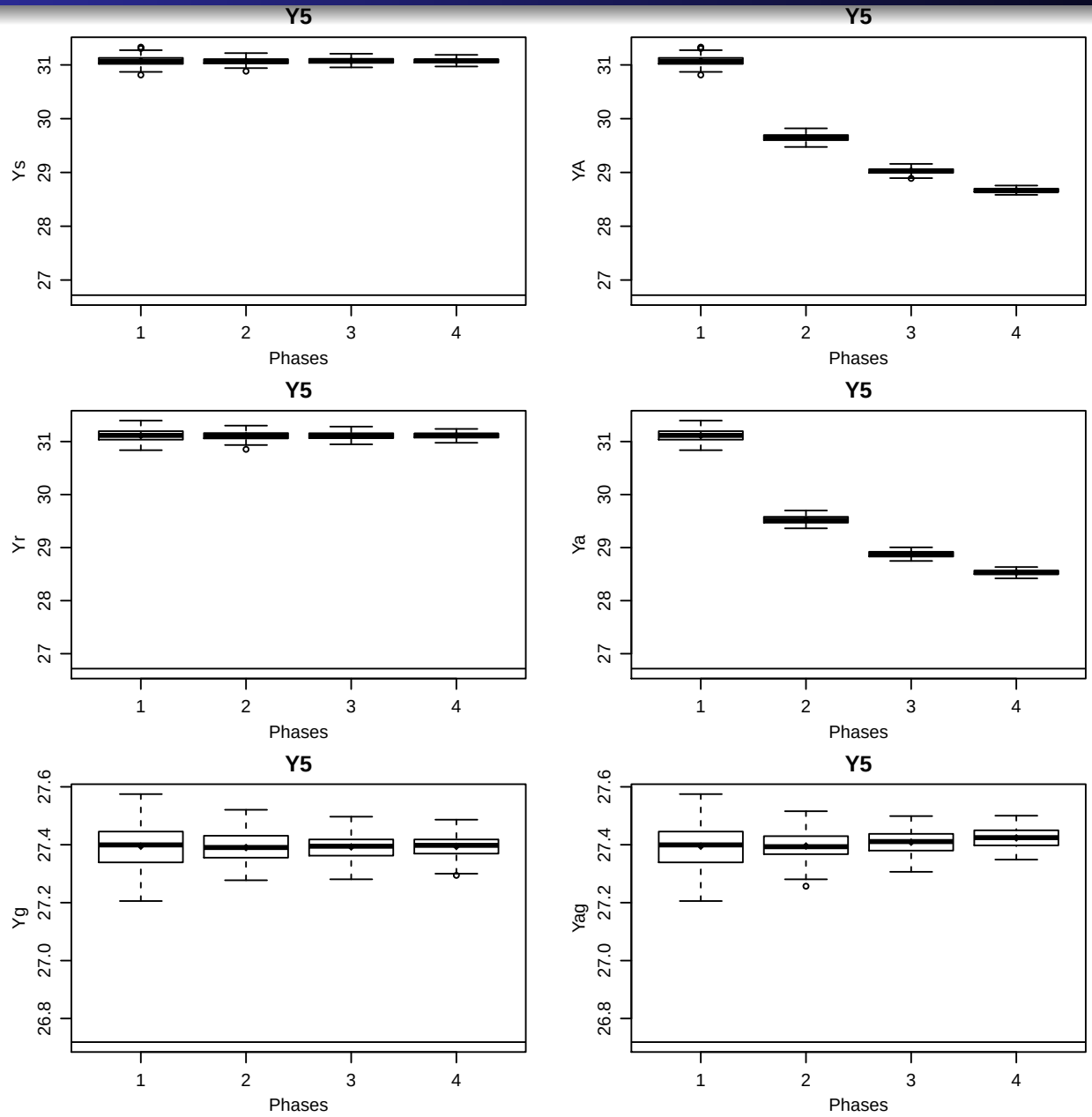




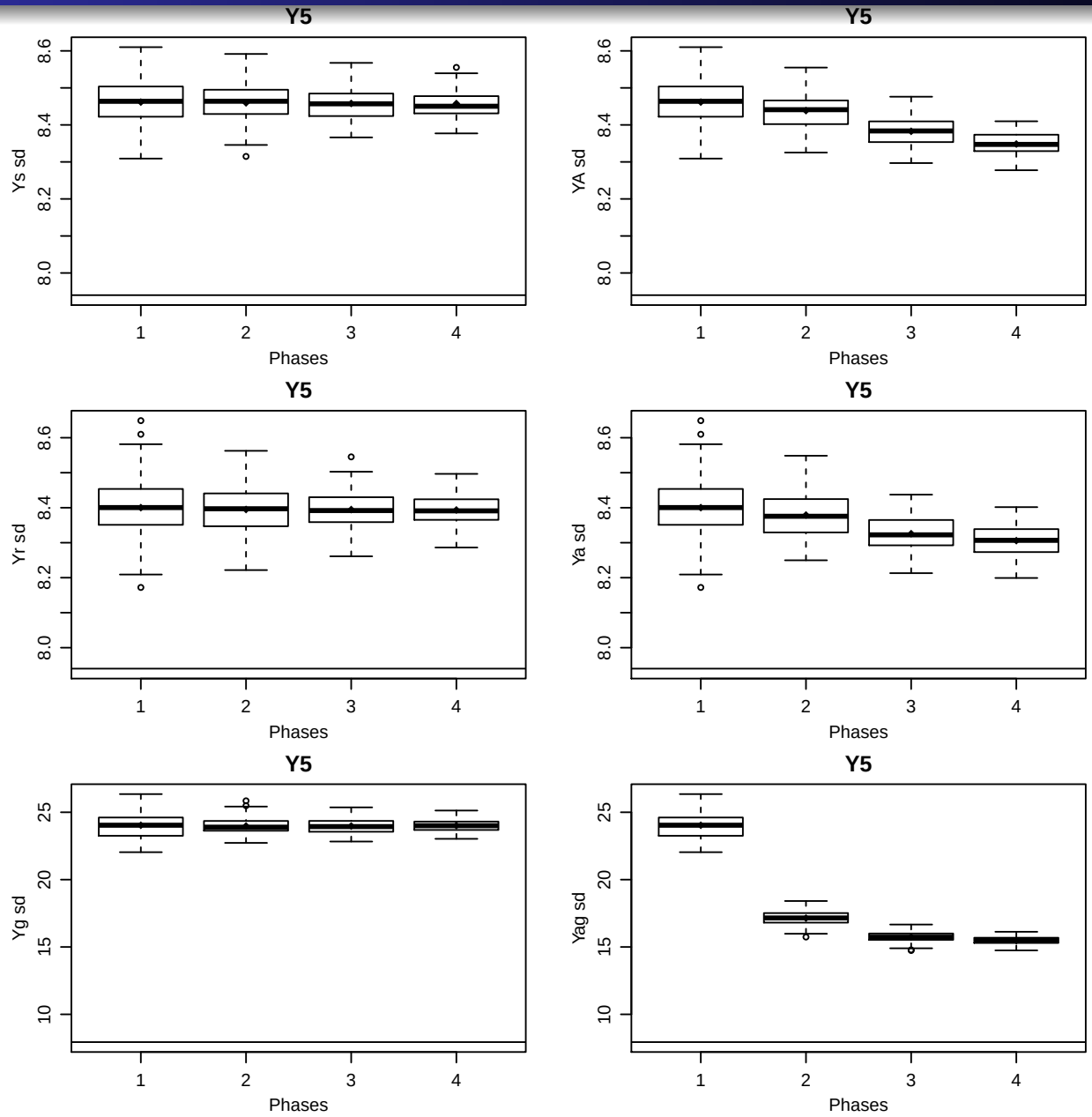




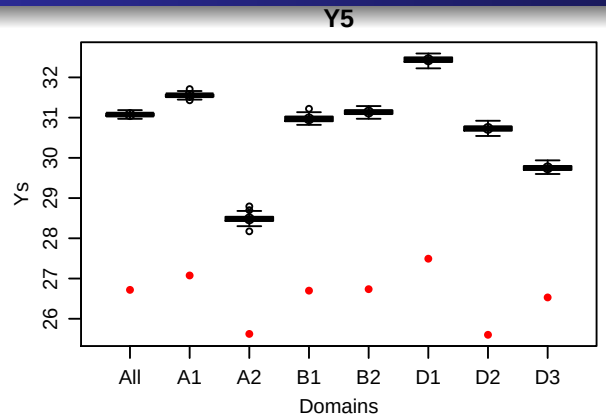
Scenario 2 BMI mean by phase

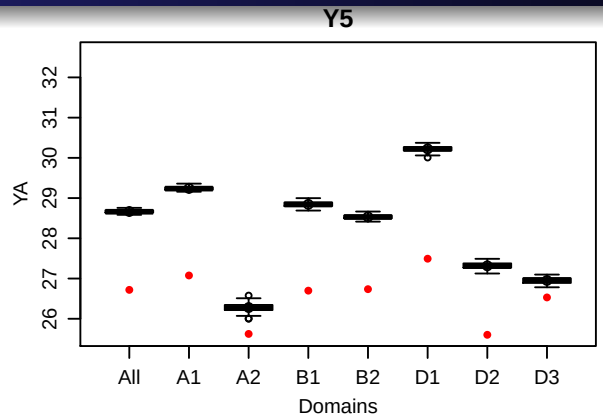


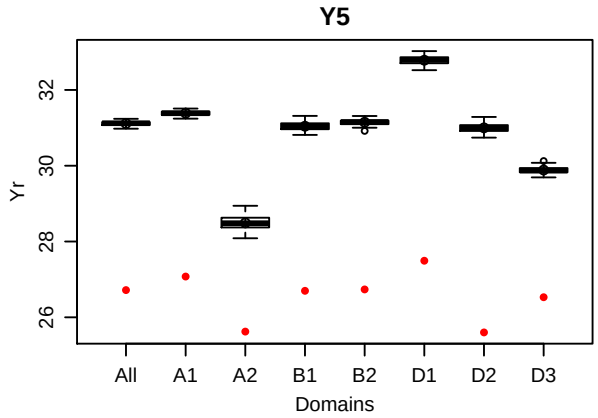
Scenario 2 BMI SD by phase

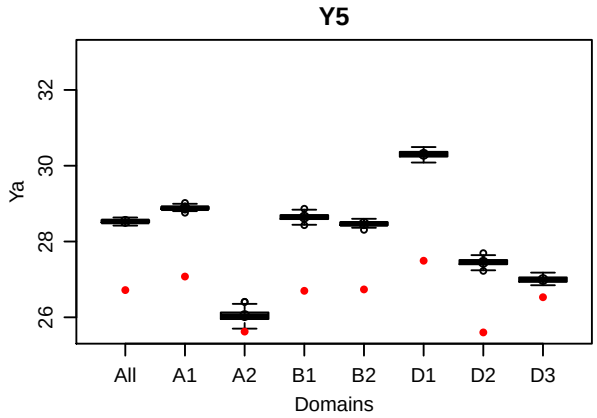


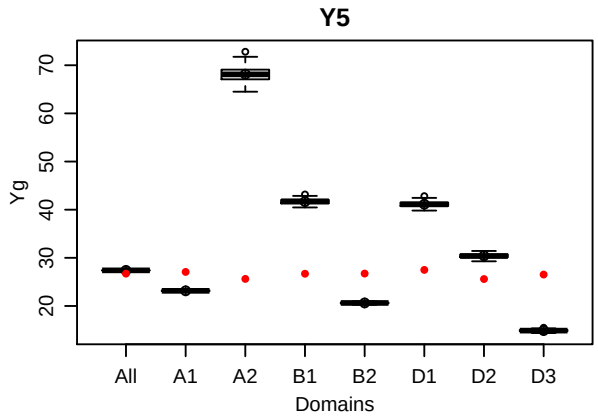
Scenario 2 BMI mean by domain

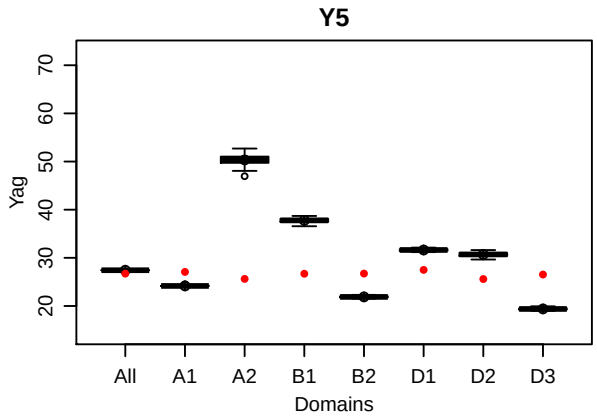




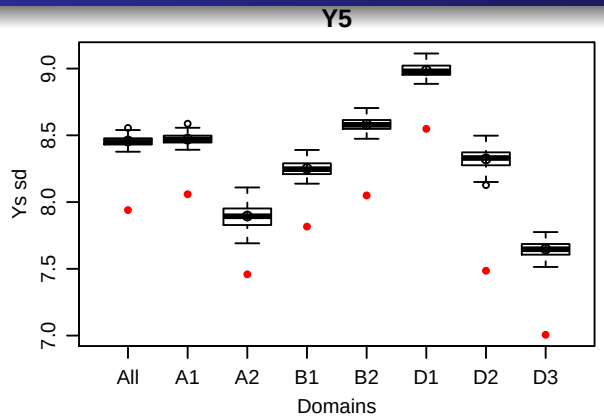


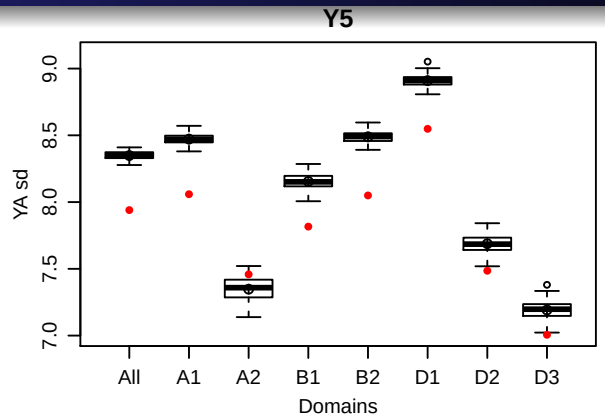


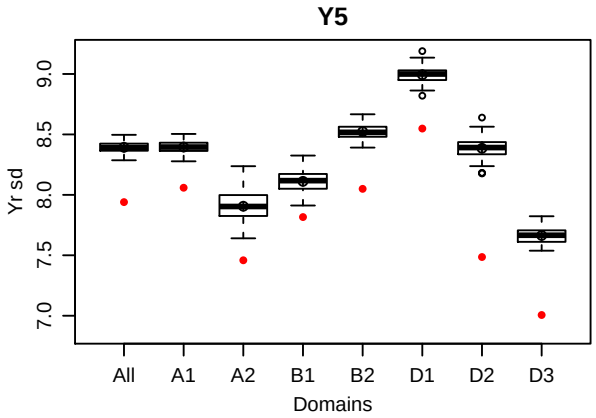


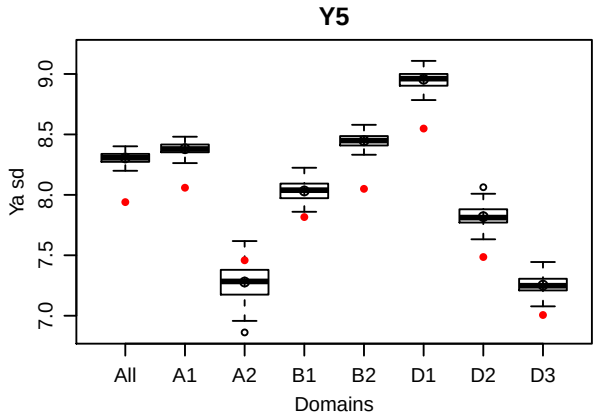


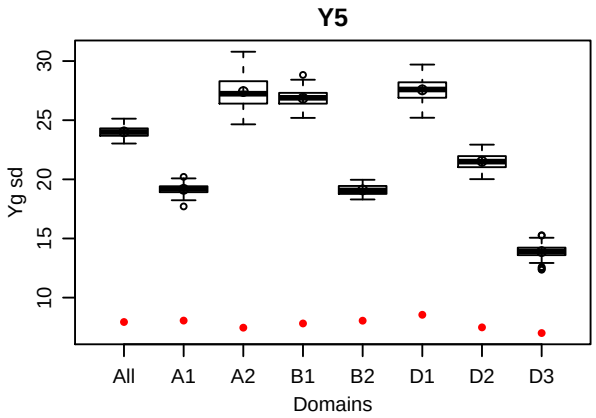
Scenario 2 BMI SD by domain

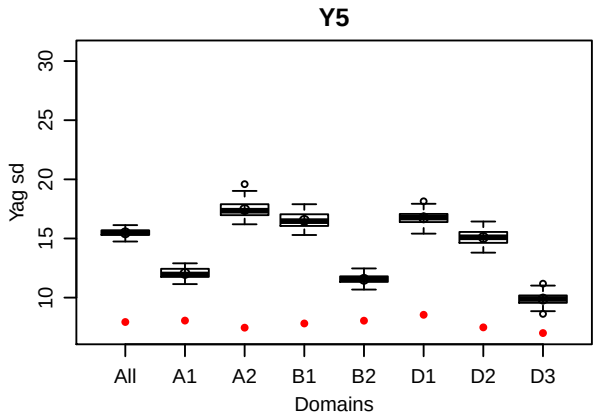












Conclusion

- **Sampling**
  Properly target over- and under-represented subjects
- **Data collection**
  Simultaneously correct for frame imperfection and nonresponse bias
- **Analysis**
  Successfully reduce errors of nonobservation for overall and domain estimates

Thank you!

Comments are highly appreciated!

Contact: julialee@umich.edu