DATA COLLECTION

STUDIES AND SURVEYS

THE GOAL OF GOOD STUDY/SAMPLING DESIGN

We need data that can:

- provide legitimate insight into our system of interest;
- provide correct, accurate answers to relevant questions;
- support the drawing of legitimate, valid conclusions, with the ability to qualify these conclusions in terms of scope and precision.

This starts with **study design** – what data to collect and how it should be collected

PATTERN FISHING / NON-PROBABILISTIC SAMPLING

Two separate issues can be combined to cause **problems** with data analysis:

- drawing conclusions (inferences) from a sample about a population that are not warranted by the sample collection method (symptomatic of NPS);
- looking for any available patterns in the data and then coming up with *post hoc* explanations for these patterns.

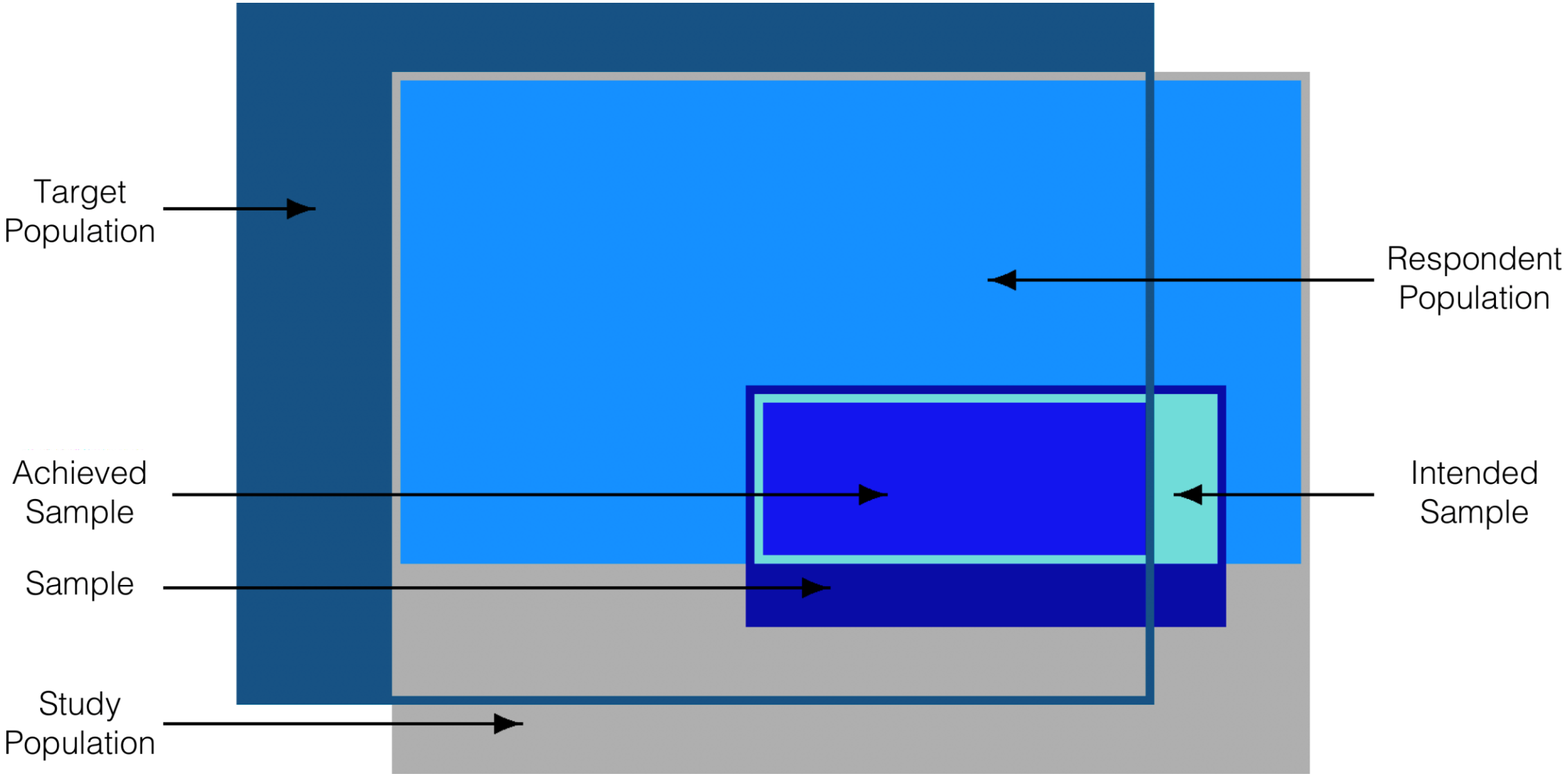
Alone or in combination, these lead to poor (and **potentially harmful**) conclusions.

STUDY/SURVEY STEPS

Studies or surveys follow the same general steps:

1. statement of objective
2. selection of survey frame
3. sampling design
4. questionnaire design
5. data collection
6. data capture and coding
7. data processing and imputation
8. estimation
9. data analysis
10. dissemination
11. documentation

The process is not always linear, but there is a definite movement from objective to dissemination.


Target Population
Respondent Population
Achieved Sample
Intended Sample
Sample
Study Population

SURVEY ERROR

Total Error = Sampling Error + Measurement Error + Non-Response Error + Coverage Error

- Sampling Error: survey, not census
- Measurement Error: observations not measured accurately
- Non-Response Error: non-respondents having systematic observation differences
- Coverage Error: frame decay and/or corruption

Statistical sampling can help provide estimates, but importantly, it can also provide some control over the **total error** (TE) of the estimates.

Ideally, TE $= 0$. In practice, there are two main contributions to TE: **sampling errors** (due to the choice of sampling scheme), and **nonsampling errors** (everything else).

DATA COLLECTION

STUDIES AND SURVEYS