Introduction to Design and Planning Longitudinal Studies: Sampling and Weights

Dr. Lauren Griffith

CLSA Associate Scientific Director

Dept. of Health Research Methods,

Evidence and Impact

McMaster University

**Summer Program on Aging
May 3, 2020**

clsa élcv
Canadian Longitudinal Study on Aging
Étude longitudinale canadienne sur le vieillissement

Talk Outline

- Key sampling concepts
- Why do we use sampling weights?
- CLSA sampling and use of sampling weights
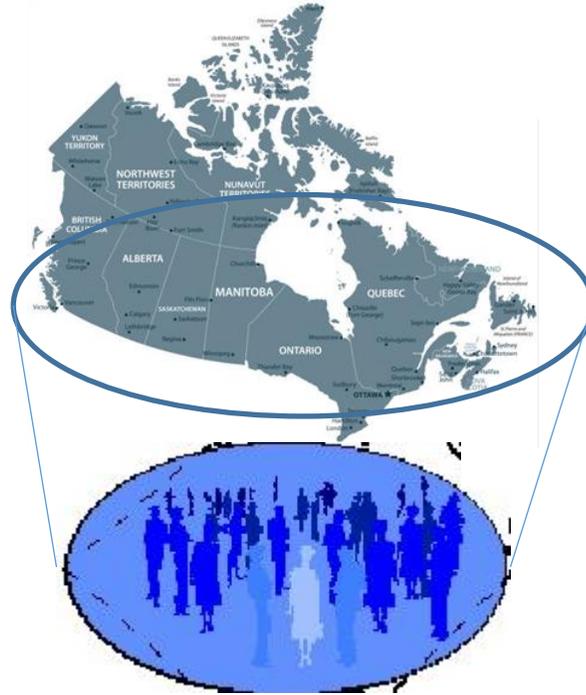- Special topic for sampling in longitudinal cohorts: replenishment sampling

clsa élcv
Canadian Longitudinal Study on Aging
Étude longitudinale canadienne sur le vieillissement

Key Concepts in Sampling

Who do you want to generalize to?

Theoretical Population (Target Population)

What population can you get access to?

The Study Population (Accessible Population)

How can you get access to them?

The Sampling Frame

Who is in your study?

The Sample

clsa élcv
Canadian Longitudinal Study on Aging
Étude longitudinale canadienne sur le vieillissement
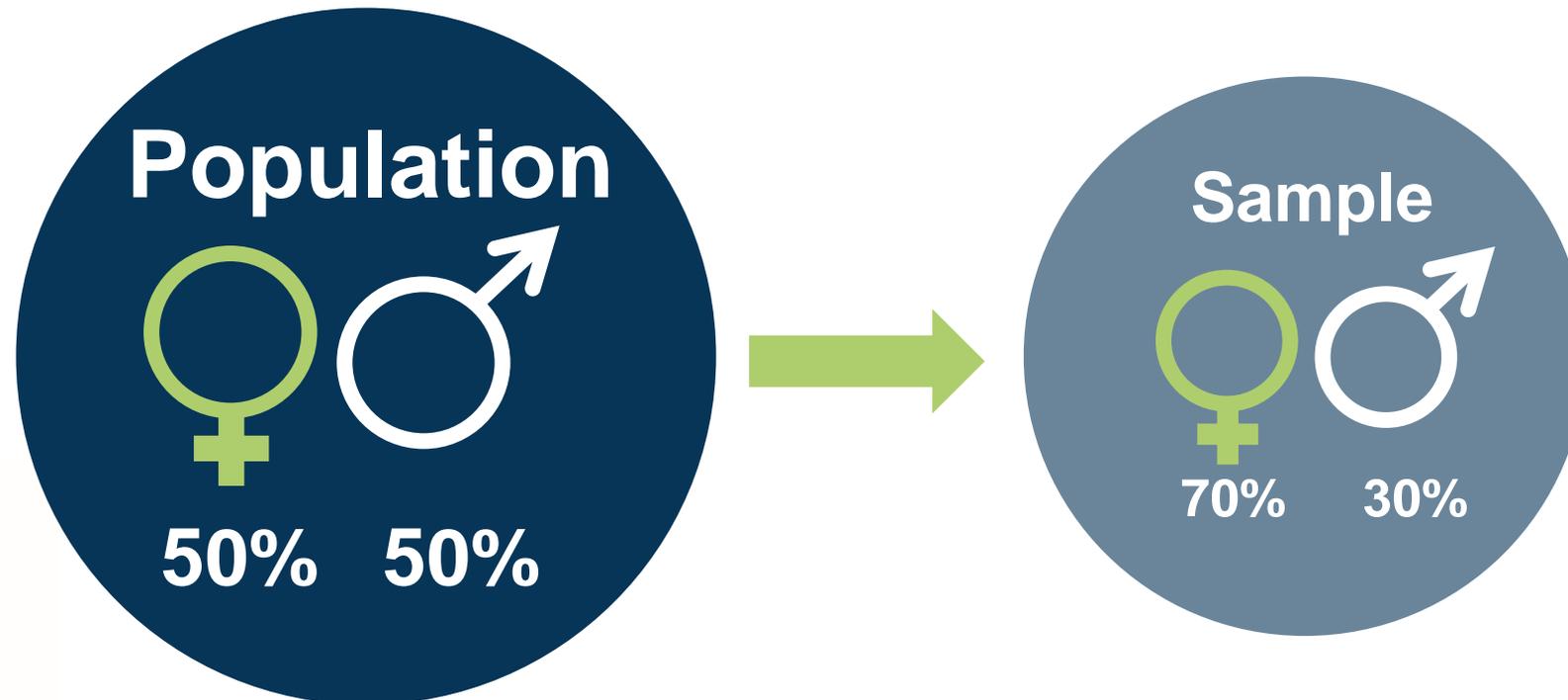
Why do we use sampling weights?

We want to <u>generalize</u> from the sample to the population, but the sample is almost never fully representative

Let's assume for example:
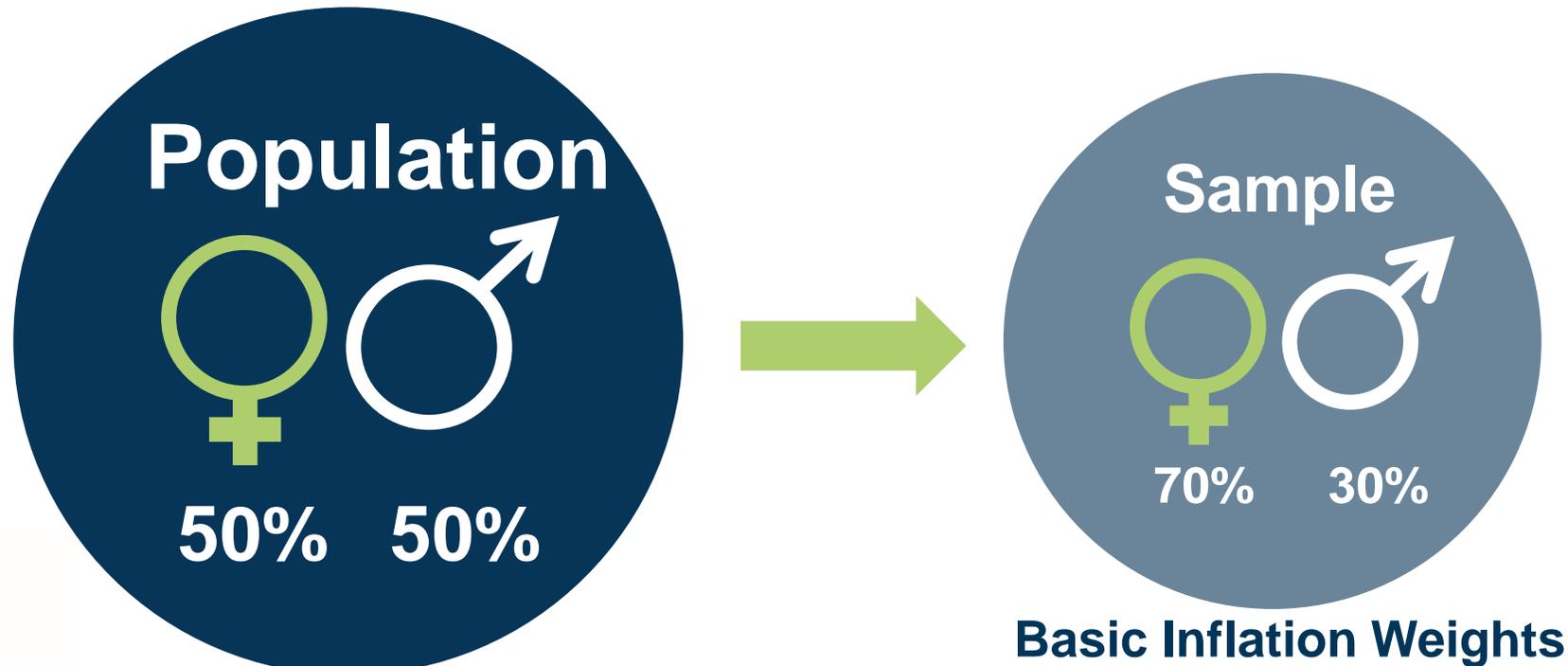

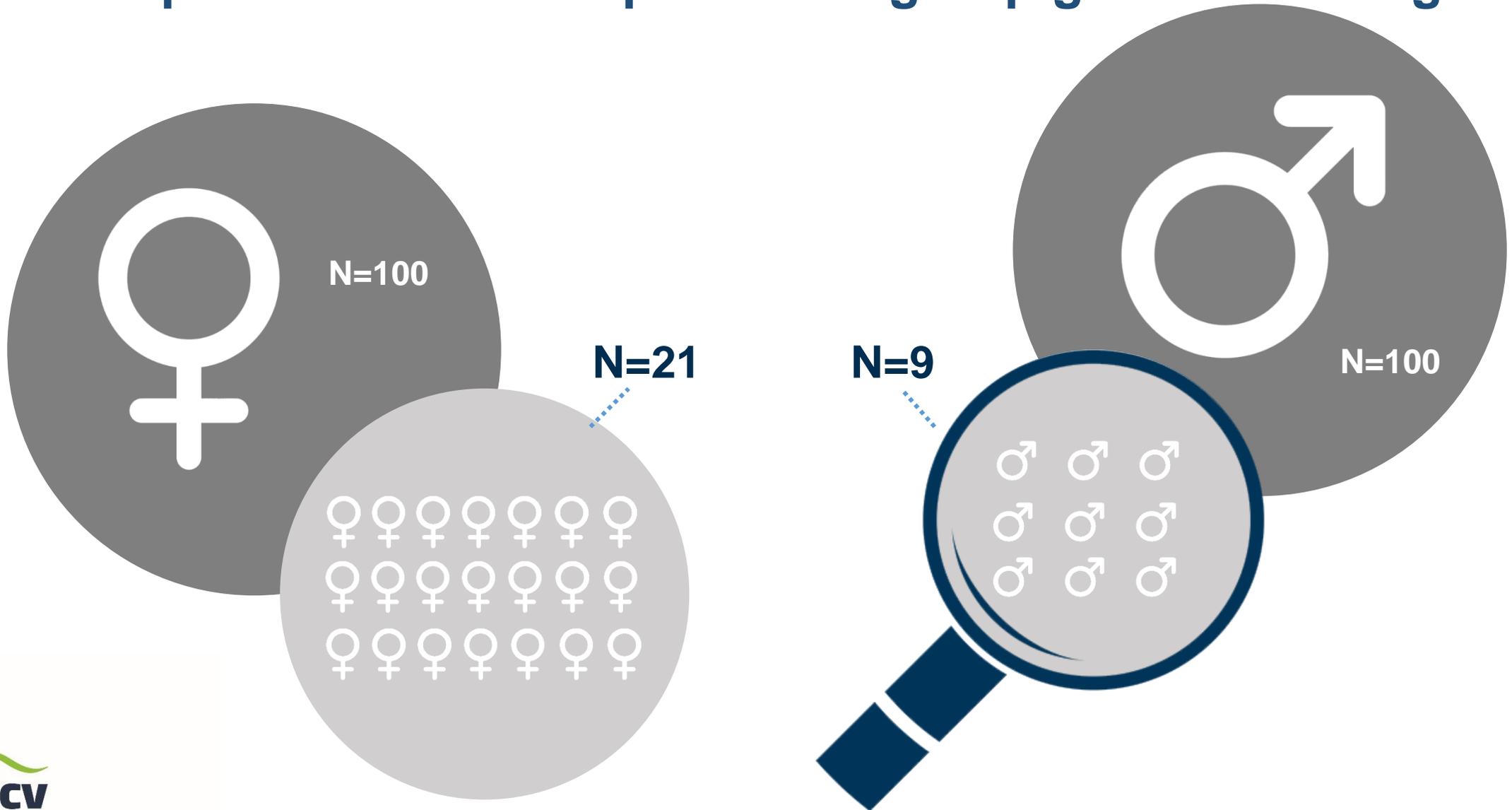
Population
50% 50%

Sample
70% 30%

Sample Weights

- Sample weights are used to make statistics computed from the data more **representative** of the population.
- It is a **standard practice** in surveys and many cohorts have sampling weights.
- **Each participant** in the study is assigned a sample weight constructed based on the inclusion probability

clsa élcv
Canadian Longitudinal Study on Aging
Étude longitudinale canadienne sur le vieillissement

Sample Weights

- Respondent in under-represented group gets higher weight; respondent in over-represented group gets lower weight.

Population
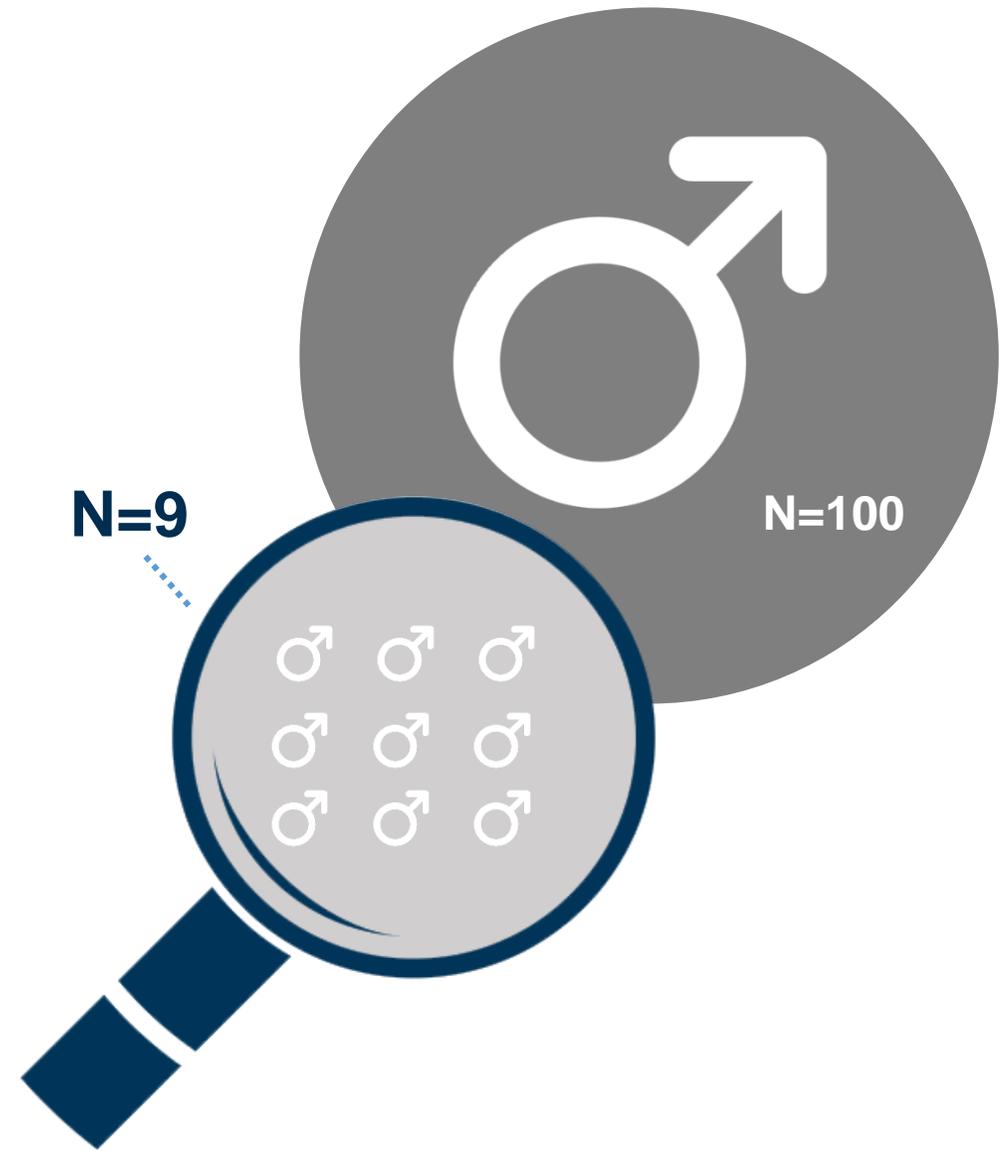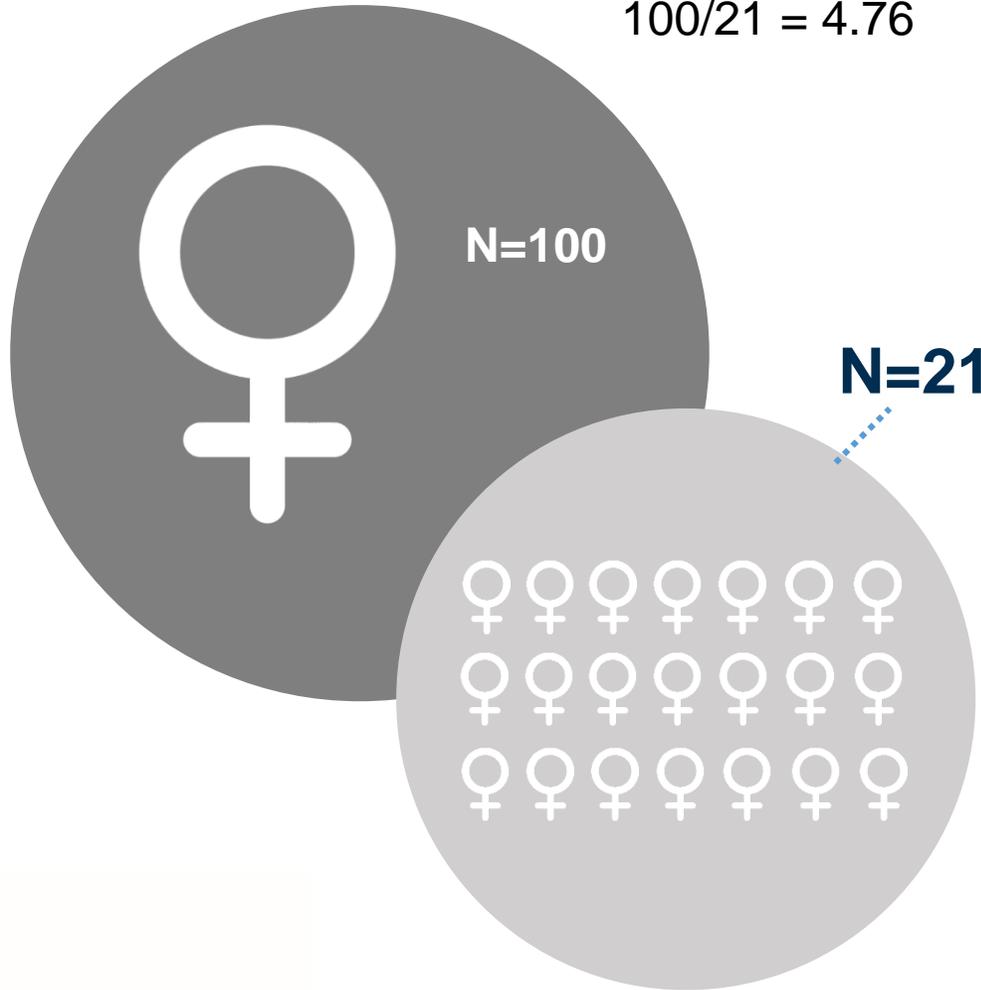50%   50%

Sample
70%   30%

Basic Inflation Weights

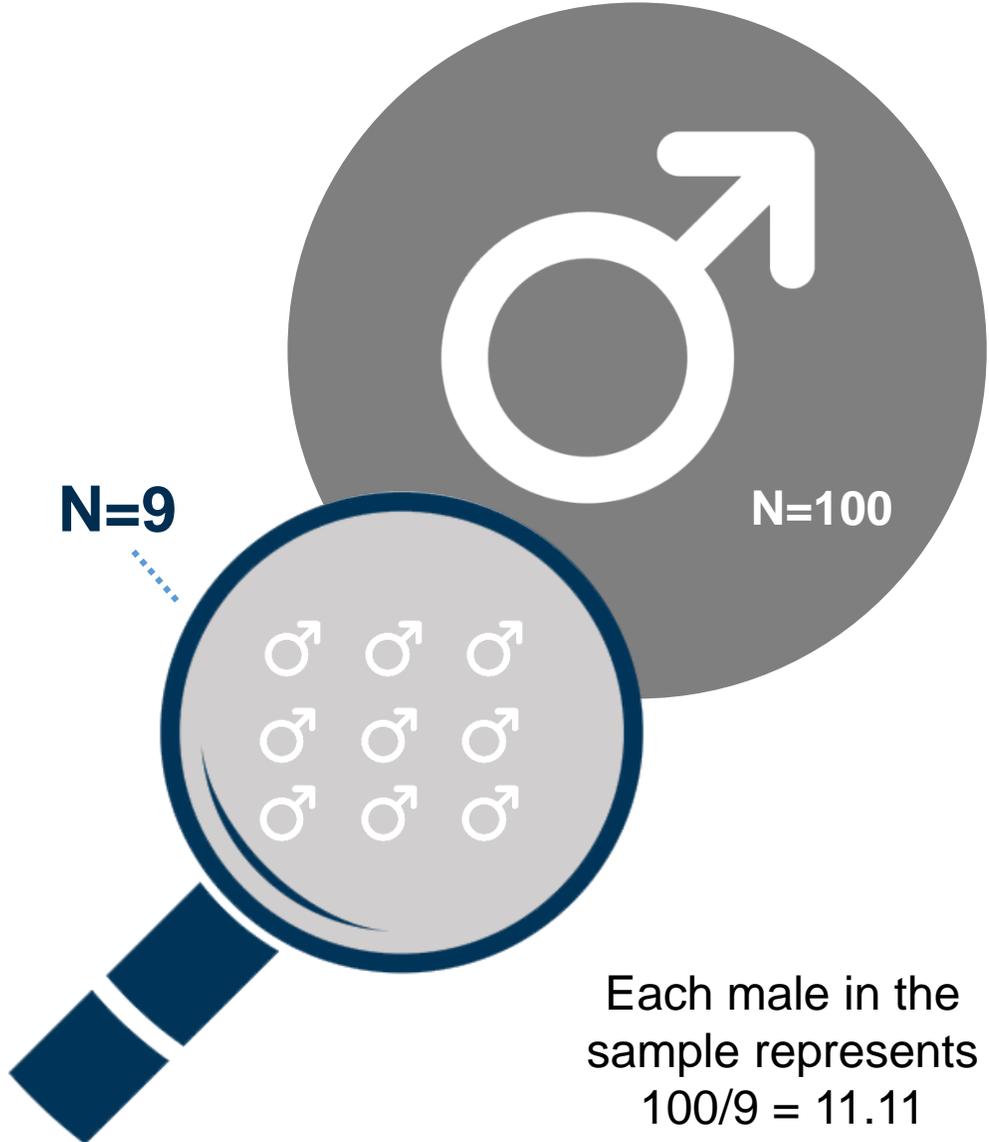# Respondent in under-represented group gets higher weight; respondent in over-represented group gets lower weight.

N=100

N=21

N=9

N=100

Each female in the sample represents 100/21 = 4.76

N=100

N=21

N=9

N=100

Each female in the sample represents 100/21 = 4.76

N=100

N=21

N=9

N=100
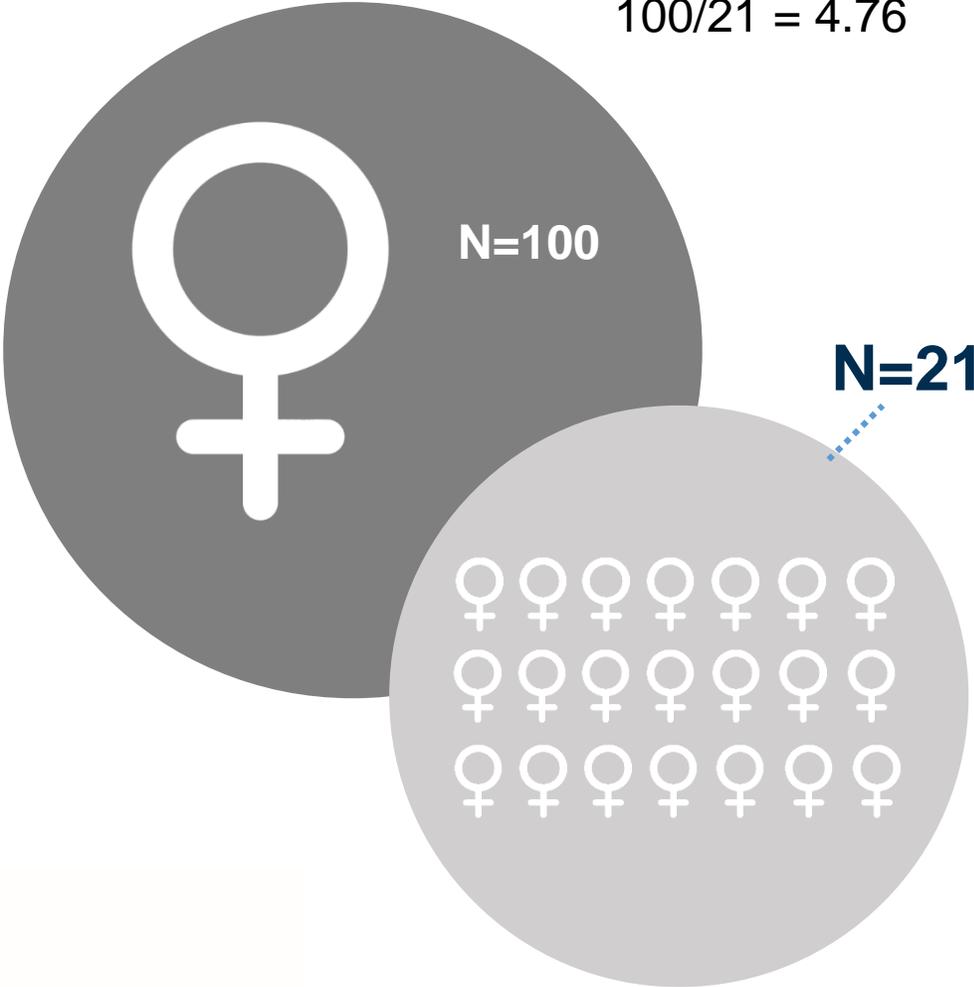
Each male in the sample represents 100/9 = 11.11

clsa élcv
Canadian Longitudinal Study on Aging
Étude longitudinale canadienne sur le vieillissement

CLSA Sample

Sample was obtained via four sources:
- Canadian Community Health Survey-Healthy Aging (CCHS-HA) [Tracking only]
- Provincial Health Registries (HR)
- Telephone Sampling (TS)
- Quebec Longitudinal Study on Nutrition and Aging (NuAge) [Comprehensive only]

CLSA Sample

Stratified Random Sampling:

- A population is subdivided into mutually exclusive subpopulations
- A simple random sample is drawn from each subpopulation

clsa élcv
Canadian Longitudinal Study on Aging
Étude longitudinale canadienne sur le vieillissement

CLSA Sample

Why Stratified Random Sampling?

- Can be done for convenience
- To obtain more precise estimates (under many circumstances)
- To obtain an estimate for the subpopulations

CLSA Sample

Why Stratified Random Sampling?

- Can be done for convenience
- To obtain more precise estimates (under many circumstances)
- To obtain an estimate for the subpopulations

Types of Weights

- **<u>Basic design weights</u>:** proportional to the reciprocals of the individual inclusion probabilities

- **<u>Inflation Weight</u>:** re-calibrated to the sum of the targeted (eligible) Canadian population
- **<u>Analytic Weights</u>:** rescaled to sum to the sample size within each province, so that their mean value is 1 within each province

Types of Weights

Inflation Weights

- For the estimation of a descriptive parameter
- Reflect the estimated parameters in the target population, e.g.
  - Mean grip strength
  - Prevalence of CHD

Analytic Weights

- For use in modeling, e.g. regression analyses, where the weighting variables are included in the models.

Why Analytic vs. Inflation Weights?

- Provinces with larger populations will tend to have much higher inflation weights compared to smaller provinces
- The observations from those strata would tend to dominate the statistical analysis
- With analytic weights point estimators will remain the same, but they are more efficient if the model is correctly specified

Analysis of Cohort Data

Needs to take into account sampling strategy as well as sampling weights

CLSA Strata

- **10 Provinces**
  - 10 (Tracking)
  - 7 (Comprehensive)

- **Age groups**
  - 45-48
  - 49-54
  - 55-64
  - 65-74
  - 75-85

- **Sex**
  - Female
  - Male

- **Geographic areas**
  - DCS (10)
  - Non-DCS (10)

- **Education**
  - Low-Ed
  - Med-Ed
  - Higher-Ed Lower
  - Higher-Ed Upper

clsa élcv
Canadian Longitudinal Study on Aging
Étude longitudinale canadienne sur le vieillissement

Analysis of Cohort Data

- Primary Sampling Unit for CLSA is the individual

- Other studies, such as the Irish Longitudinal Study on Ageing (Tilda), identified households and invited all adults aged 50 or over to participate

clsa élcv
Canadian Longitudinal Study on Aging
Étude longitudinale canadienne sur le vieillissement

Special issues in longitudinal cohort studies – Replenishment Sampling

- Used in longitudinal studies to replace younger participants who move into older age strata over time
- Appropriate when interest in cross-sectional estimates across age groups in addition to studying longitudinal trajectories
- Can also be used to compensate for disproportionate attrition

clsa élcv
Canadian Longitudinal Study on Aging
Étude longitudinale canadienne sur le vieillissement

QUESTIONS?