CLSA Webinar Series

Anticipating new weights in the CLSA: Unpacking sampling weights and their use

Lauren Griffith, PhD, CLSA, McMaster University

12 pm to 1 pm ET | October 27, 2020

It is standard practice in surveys to use sampling weights, however, when surveys involve complex sampling, individuals in selected populations might not have equal probabilities of participation. Participants in the Canadian Longitudinal Study on Aging (CLSA) are assigned sample weights based on their inclusion probability. Essentially, the inflation weight provided with CLSA data tells a researcher how many people the participant represents in the target population. The use of weights can be complex, and the method of calculation might seem opaque to researchers. This webinar will present an overview of weight calculations in anticipation of new weights in the CLSA and aim to unpack the complexities of sampling weights and how they are implemented in the CLSA Tracking and Comprehensive cohorts.

Dr. Lauren Griffith is an associate scientific director and Hamilton site lead of the Canadian Longitudinal Study on Aging. She is an associate professor in the Department of Health Research Methods, Evidence, and Impact at McMaster University. Her research interests include physical functioning, injury and aging as well as the harmonization of longitudinal data.

Webinars will be broadcast using WebEx.
Further instructions will be sent by email.

Register online at:
clsa-elcv.ca/webinars

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

www.clsa-elcv.ca

Land Acknowledgement

The National Coordinating Centre of the Canadian Longitudinal Study on Aging (CLSA) is located on the traditional territories of the Mississauga and Haudenosaunee Nations, and within the lands protected by the Dish With One Spoon wampum agreement.

The CLSA Data Curation Centre (DCC) located at the Research Institute of the McGill University Health Centre is situated on the traditional territory of the Kanien’kehà:ka known as the Mohawk people, and is a place which has long served as a site of meeting and exchange amongst nations.

As attendees of this webinar, we want to acknowledge the original inhabitants of the land where we currently have the privilege to research, live and work, wherever that may be.

Anticipating new weights in the CLSA: Unpacking sampling weights and their use

Dr. Lauren Griffith

CLSA Associate Scientific Director

Dept. of Health Research Methods,

Evidence and Impact

McMaster University

on behalf of the CLSA Research Team

CLSA Webinar Series
Oct. 27, 2020

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

Acknowledgement

- Dr. Mary Thompson, Dr. Changbao Wu, Dr. Harry Shannon, Dr. Nazmul Sohel, Dr. Urun Erbas Oz, Dr. Hon Yiu (Henry) So

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

Webinar Outline

- Why do we use sampling weights?
- CLSA sampling and use of sampling weights
- Why do we need **new** sampling weights?
- How do the original weights and the new weights differ?
- When will the new weights be available?
- What is coming next?

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

Webinar Outline

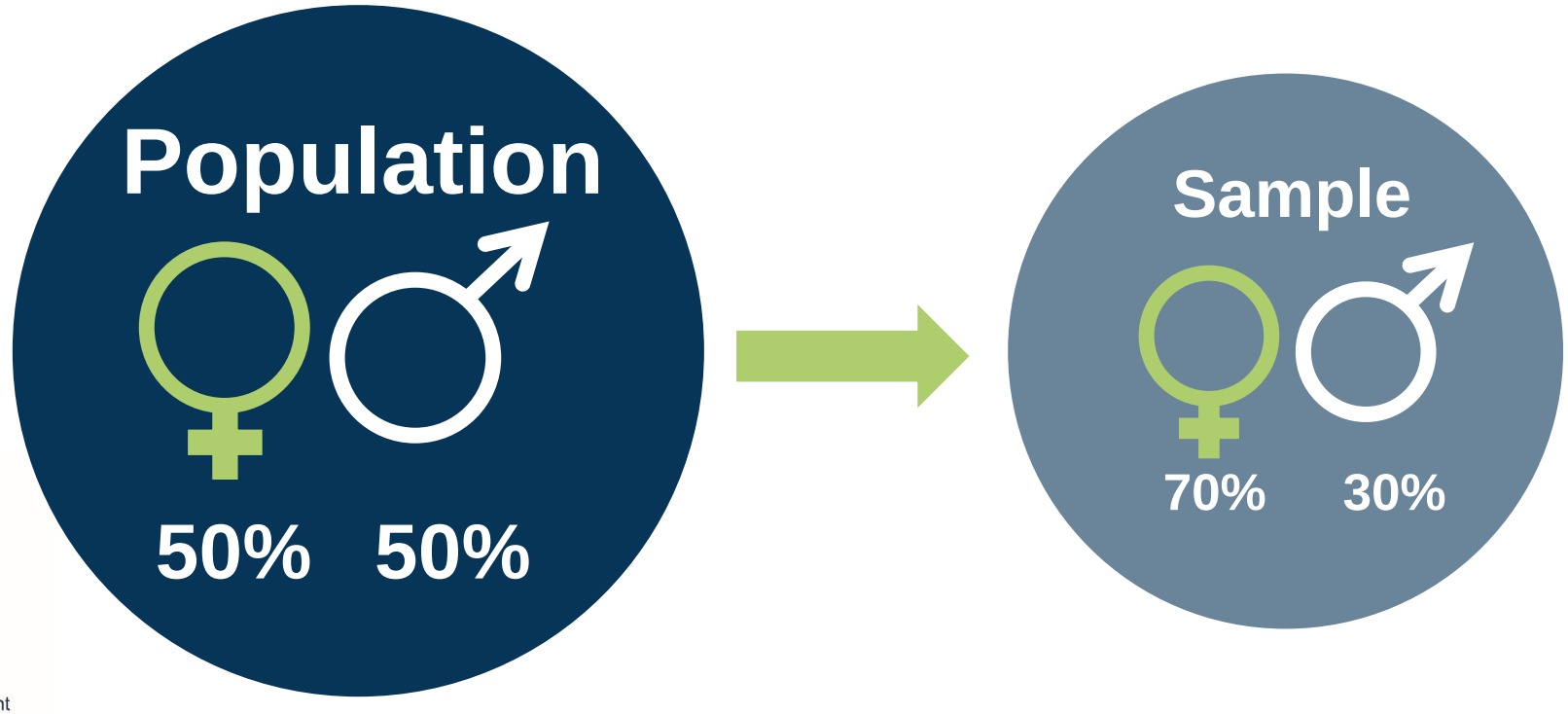
- What will not be included:
  - Technical guidance on the use of sampling weights

Why do we use sampling weights?

We want to **generalize** from the sample to the population, but the sample is almost never fully representative

We want to generalize from the sample to the population, but the sample is almost never fully representative

Let's assume for example:



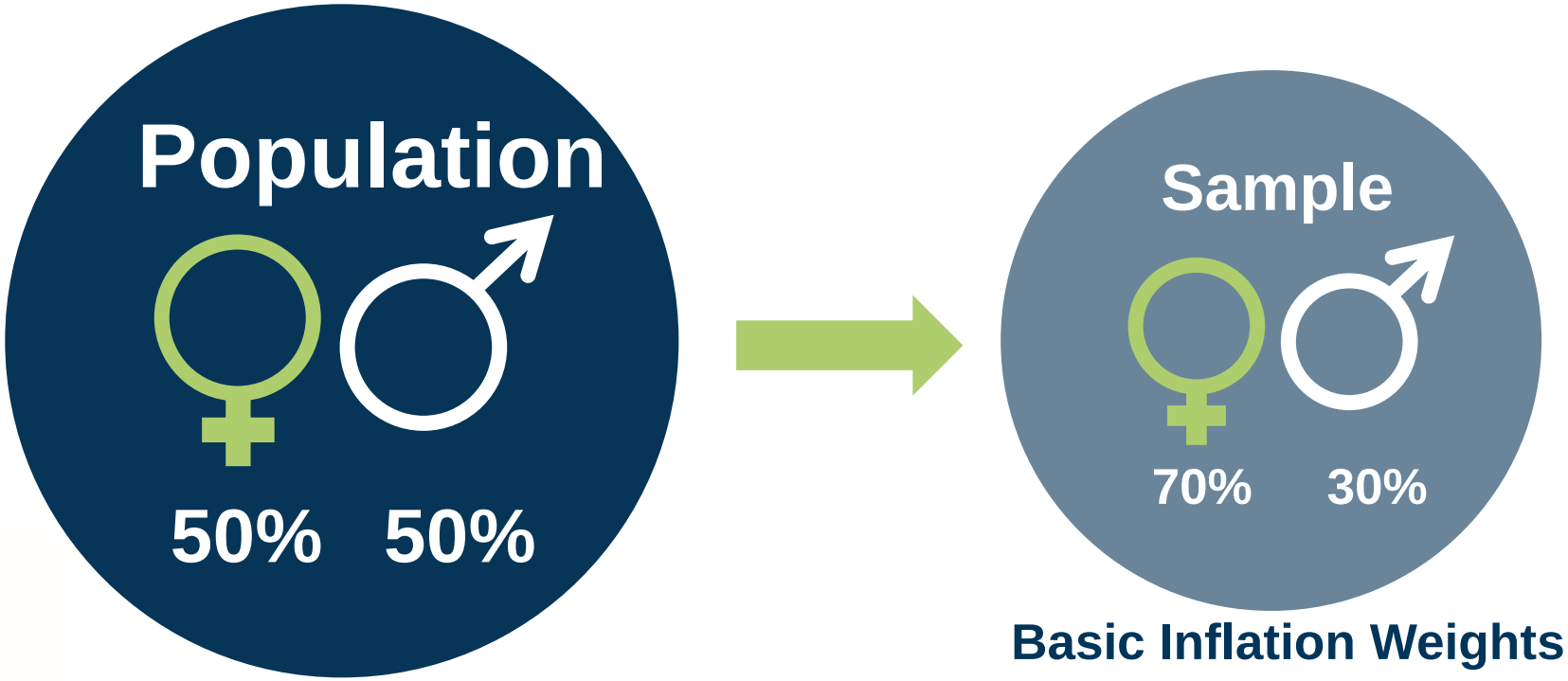
clsa élcv
Canadian Longitudinal Study on Aging
Étude longitudinale canadienne sur le vieillissement

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 to use sampling weights.
- **Each participant** in the study is assigned a sample weight constructed based on the inclusion probability.
- Sample weights are always **positive** and **non-zero**.

Sample Weights

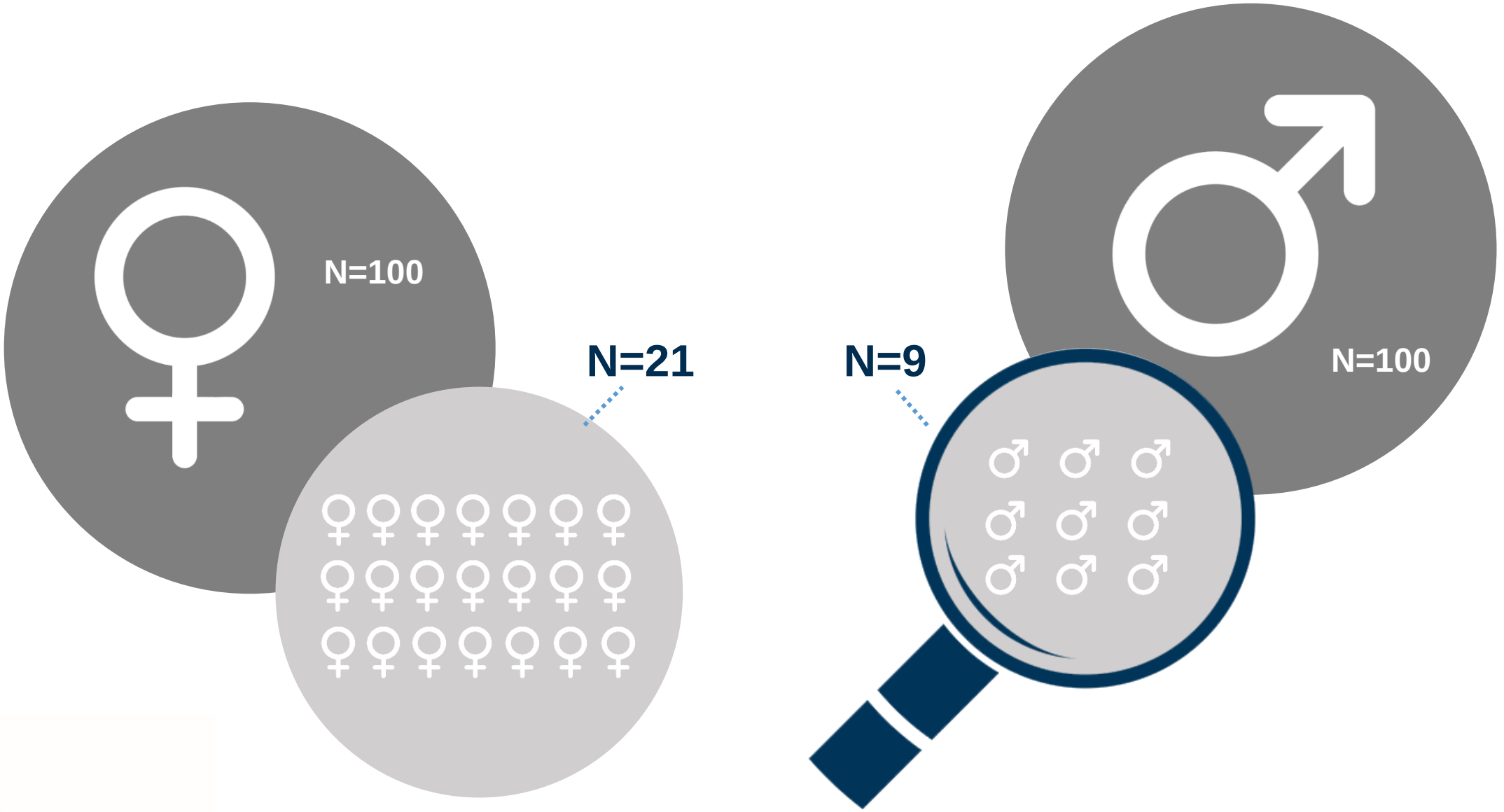
- Respondent in under-represented group gets higher weight; respondent in over-represented group gets lower weight.



Basic Inflation Weights

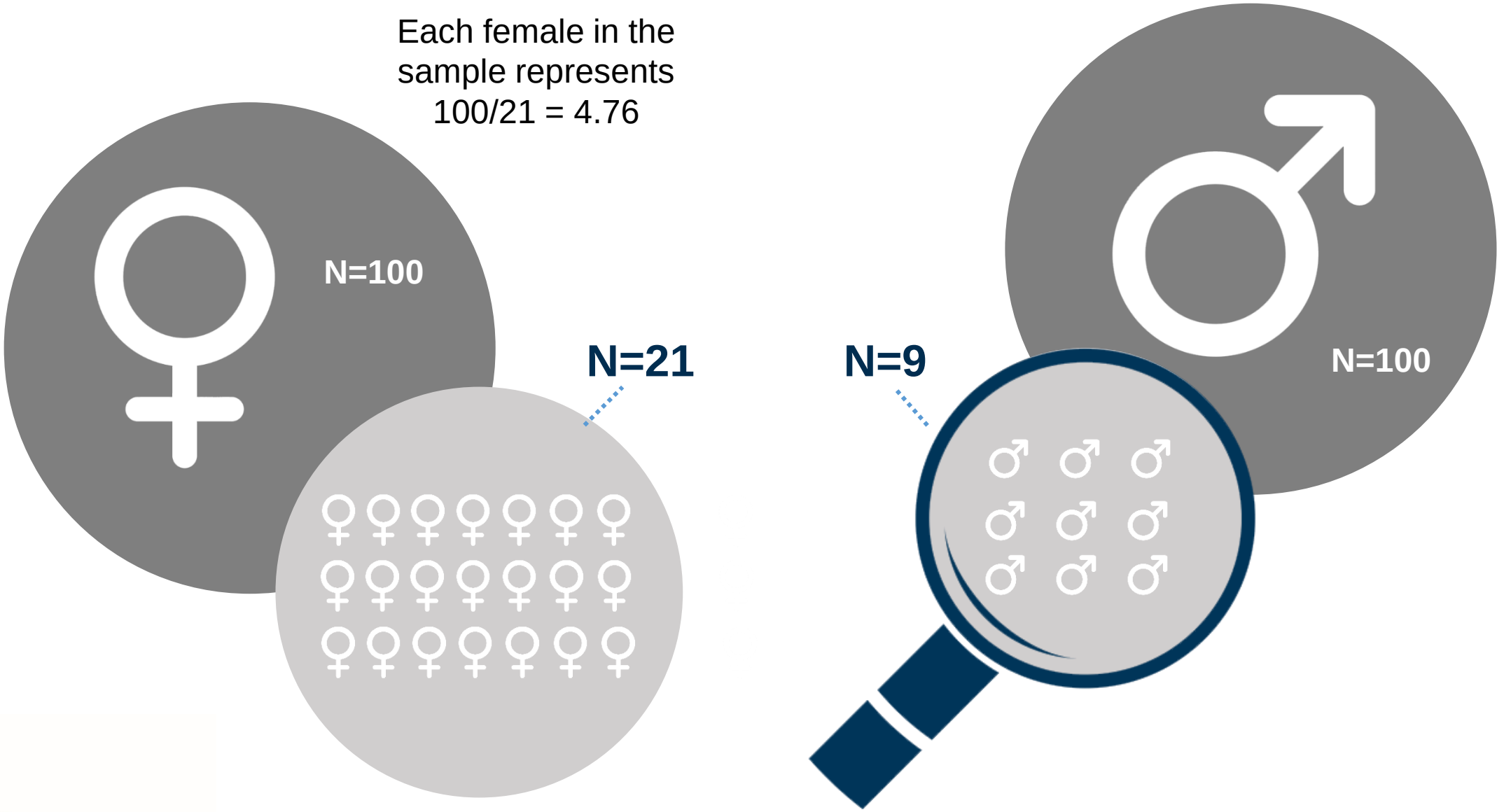
(0.5/0.7)x(N/n) **(0.5/0.3)x(N/n)**

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



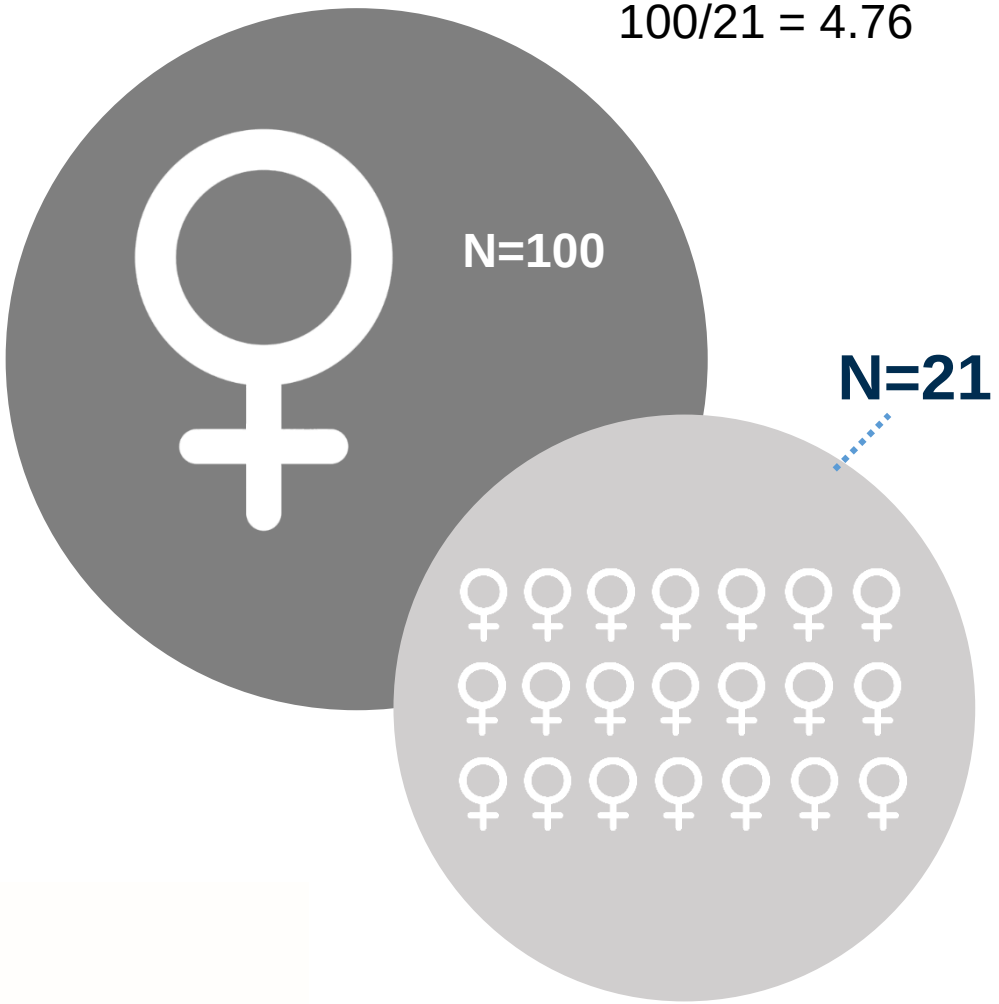
clsa élcv
Canadian Longitudinal Study on Aging
Étude longitudinale canadienne sur le vieillissement

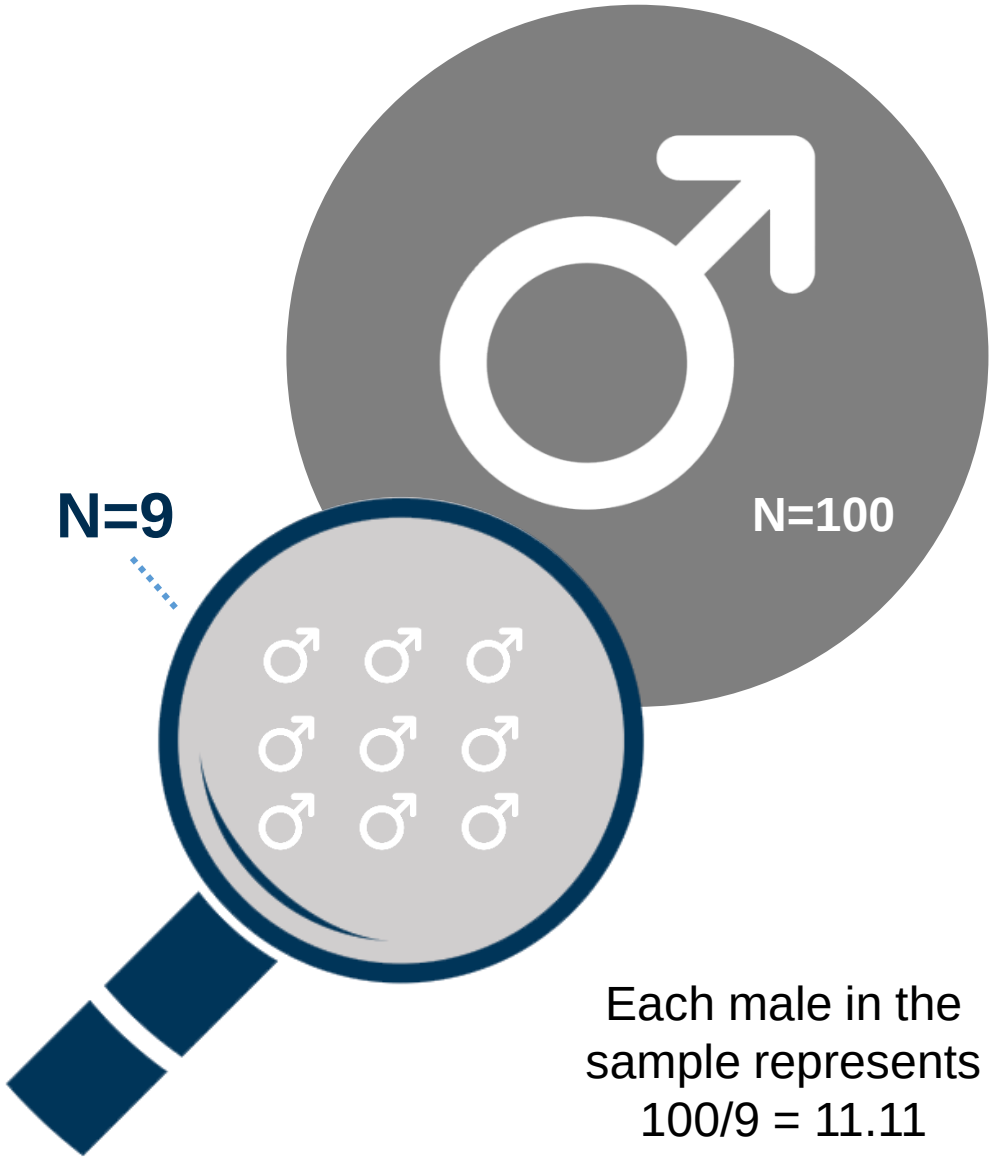
Each female in the sample represents 100/21 = 4.76



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

Each female in the sample represents 100/21 = 4.76





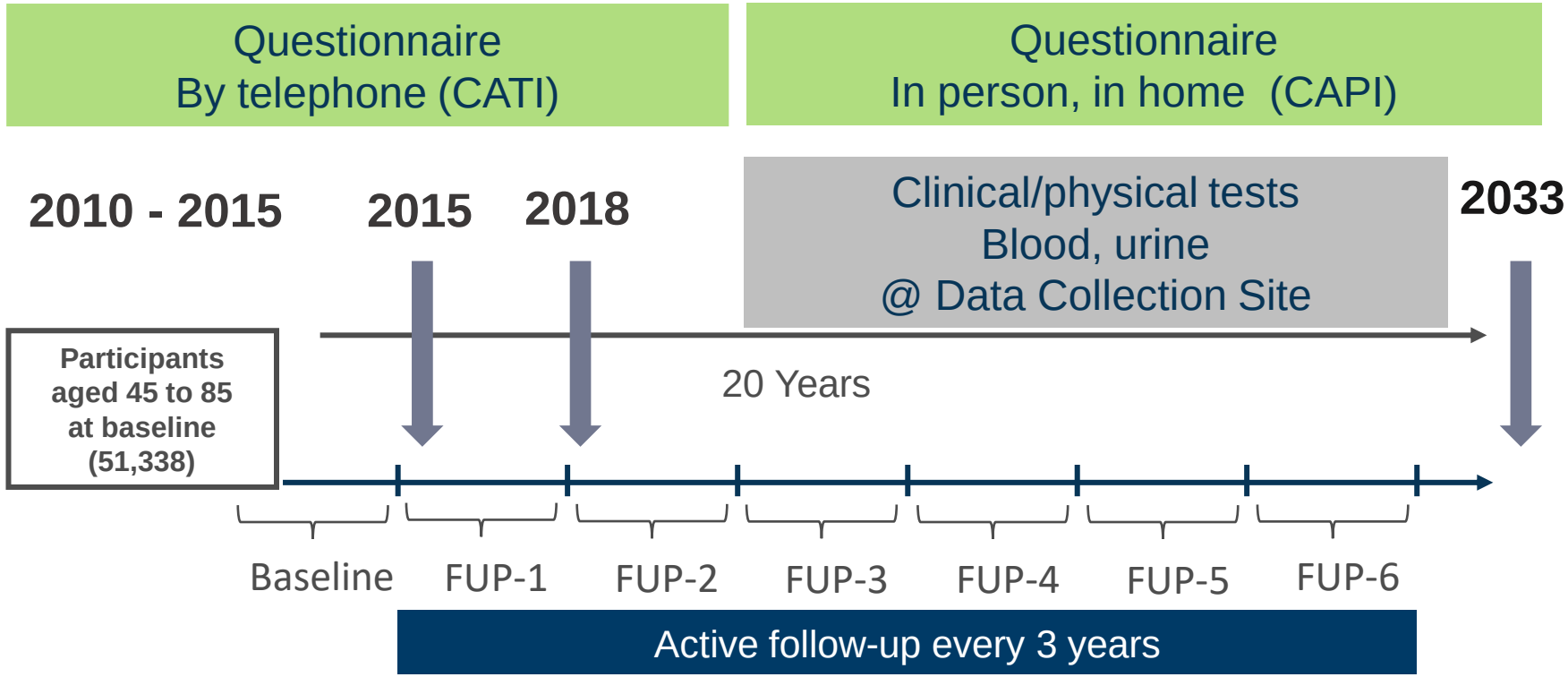
Each male in the sample represents 100/9 = 11.11

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

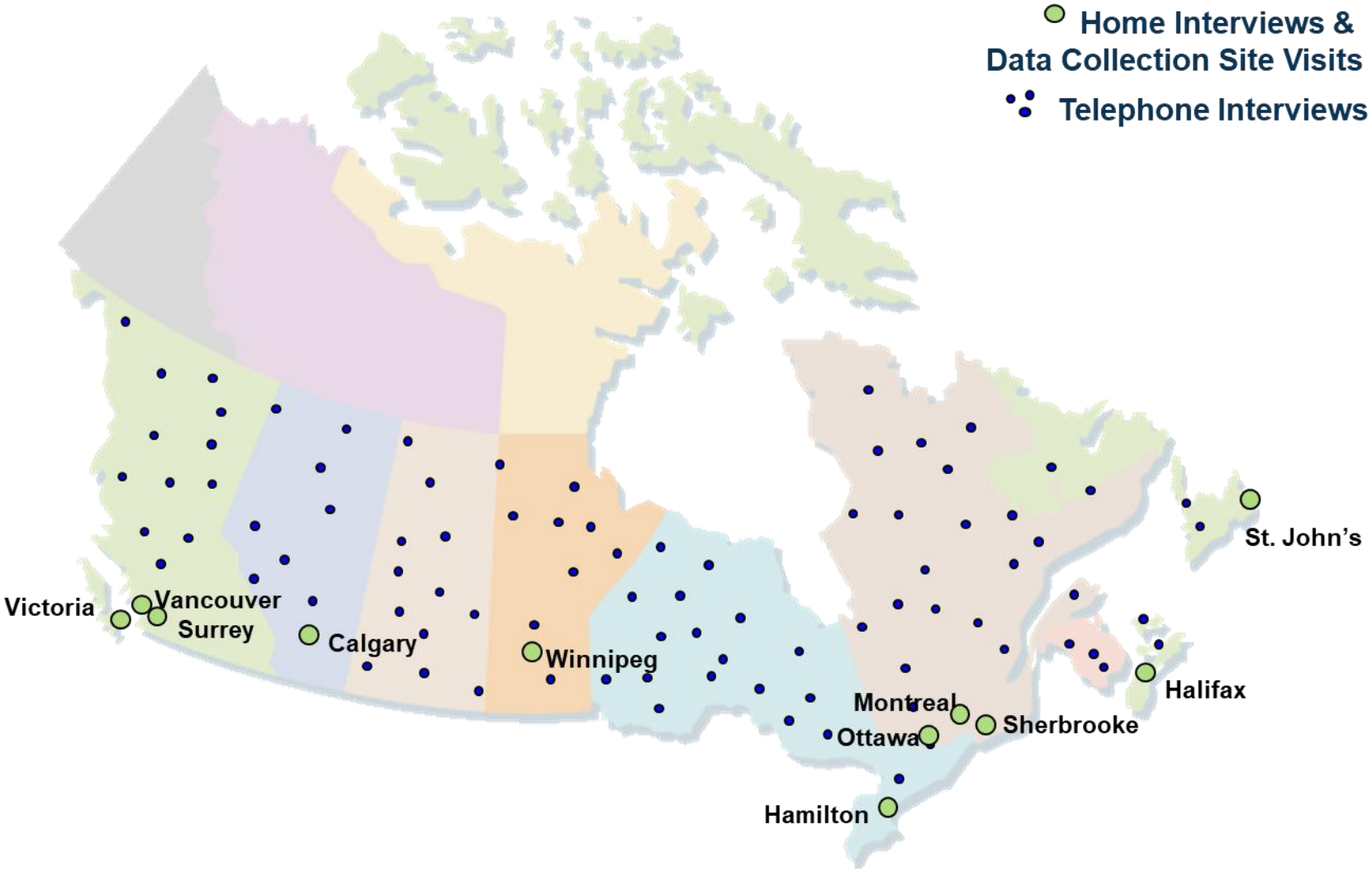
CLSA Research Platform

50,000 participants aged 45 - 85 at baseline

Target: 20,000 Actual: 21,241 Randomly selected within provinces	Target: 30,000 Actual: 30,097 Randomly selected within 25-50 km of 11 sites



National Scope



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 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

CLSA Strata

- **10 Provinces**
  - 10 (Tracking)
  - 7 (Comprehensive)
- **Age groups**
  - 45-54
  - 55-64
  - 65-74
  - 75-85
- **Sex**
  - Female
  - Male
- **Geographic areas**
  - DCS
  - Non-DCS

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

CLSA Strata

- Early analyses showed **under-representation of people with lower SES (education, income)**
- This could potentially lead to low statistical power
- Thus, to increase heterogeneity in SES, we chose to over-sample people from dissemination areas with higher percent of people with lower levels of education

CLSA Strata

- **10 Provinces**
  - 10 (Tracking)
  - 7 (Comprehensive)
- **Age groups**
  - 45-54
  - 55-64
  - 65-74
  - 75-85
- **Sex**
  - Female
  - Male
- **Geographic areas**
  - DCS
  - Non-DCS
- **Education**
  - Low-Ed
  - High-Ed

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

Types of Weights: Inflation Weights

- The CLSA Tracking and Comprehensive Cohort inflation weights were constructed

- First, **basic design weights**,which are proportional to the reciprocals of the individual inclusion probabilities, were computed; they were then **re-calibrated** to the sum of the targeted (eligible) Canadian population

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

Types of Weights: Inflation Weights

- The CLSA Tracking and Comprehensive Cohort inflation weights were constructed

- First, **basic design weights**,which are proportional to the reciprocals of the individual inclusion probabilities, were computed; they were then **re-calibrated** to the sum of the targeted (eligible) Canadian population → Using CCHS-HA

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

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

Types of Weights: Inflation Weights

- Prevalence of CHD

$$\frac{\sum_{i=1}^{N} w_i \, y_i}{\sum_{i=1}^{N} w_i}$$

Where:

N = Number in the sample

w_i = weight for the ith participant

y_i = 1 if participant has CHD and 0 otherwise

Types of Weights: Inflation Weights

- Prevalence of CHD

$$\frac{\sum_{i=1}^{N} w_i \, y_i}{\sum_{i=1}^{N} w_i}$$

Number of people in the **target population** with CHD

Where:

N = Number in the sample

w_i = weight for the ith participant

y_i = 1 if participant has CHD and 0 otherwise

Types of Weights: Inflation Weights

- Prevalence of CHD

$$\frac{\sum_{i=1}^{N} w_i \, y_i}{\sum_{i=1}^{N} w_i}$$

Number of people in the **target population** with CHD

Number of people in the **target population**

Where:

N = Number in the sample

w_i = weight for the ith participant

y_i = 1 if participant has CHD and 0 otherwise

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

Types of Weights: Inflation Weights

- Prevalence of CHD

$$\frac{\sum_{i=1}^{N} w_i \, y_i}{\sum_{i=1}^{N} w_i}$$

Number of people in the **CLSA** with CHD

Number of people in the **CLSA**

If all $w_i = 1$

Where:

N = Number in the sample

w_i = weight for the ith participant

y_i = 1 if participant has CHD and 0 otherwise

Types of Weights: Analytic Weights

- Analytic weights are proportional to the inflation weights but **rescaled** to sum to the sample size within each province, so that their mean value is **1** within each province.
- They are intended 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

Sample Weights for Pooled Data

- Inflation weights were also provided for the pooled sample from two cohorts based on:
  - Combined Tracking and Comprehensive inclusion probability for participants in the DCS areas
  - Tracking inclusion probability for participants in the non-DCS areas.

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

Sample Weights in CLSA Data (previous)

Inflation Weights

WGHTS_TRIMMED_TRM

WGHTS_TRIMMED_COM

WGHTS_TRIMMED_CLSA

Analytic Weights

WGHTS_ANALYTIC_TRM

WGHTS_ANALYTIC_COM

WGHTS_ANALYTIC_CLSA

Why new weights?

Originally

- We anticipated that most analyses would be conducted at the Province-level
- CCHS-HA was used to calibrate both Tracking and Comprehensive weights
- CCHS-HA weights could be used to estimate parameters at the level of a Health Region

Why new weights?

Our thinking evolved

- While CCHS-HA worked well to calibrate the Tracking cohort it worked less well at the DCS-level (Area represented 25-50 km around the DCSs)
- There was interest in using the CLSA data at sub-Province level

How do the old and new weights differ?

- Calibration Source
  - CCHS-HA used 2006 Census
  - CLSA Recruitment began in 2011; we now use 2011 National Household Survey
- Additional refinement
  - Use of individual rather than geographic education level for weight calibration

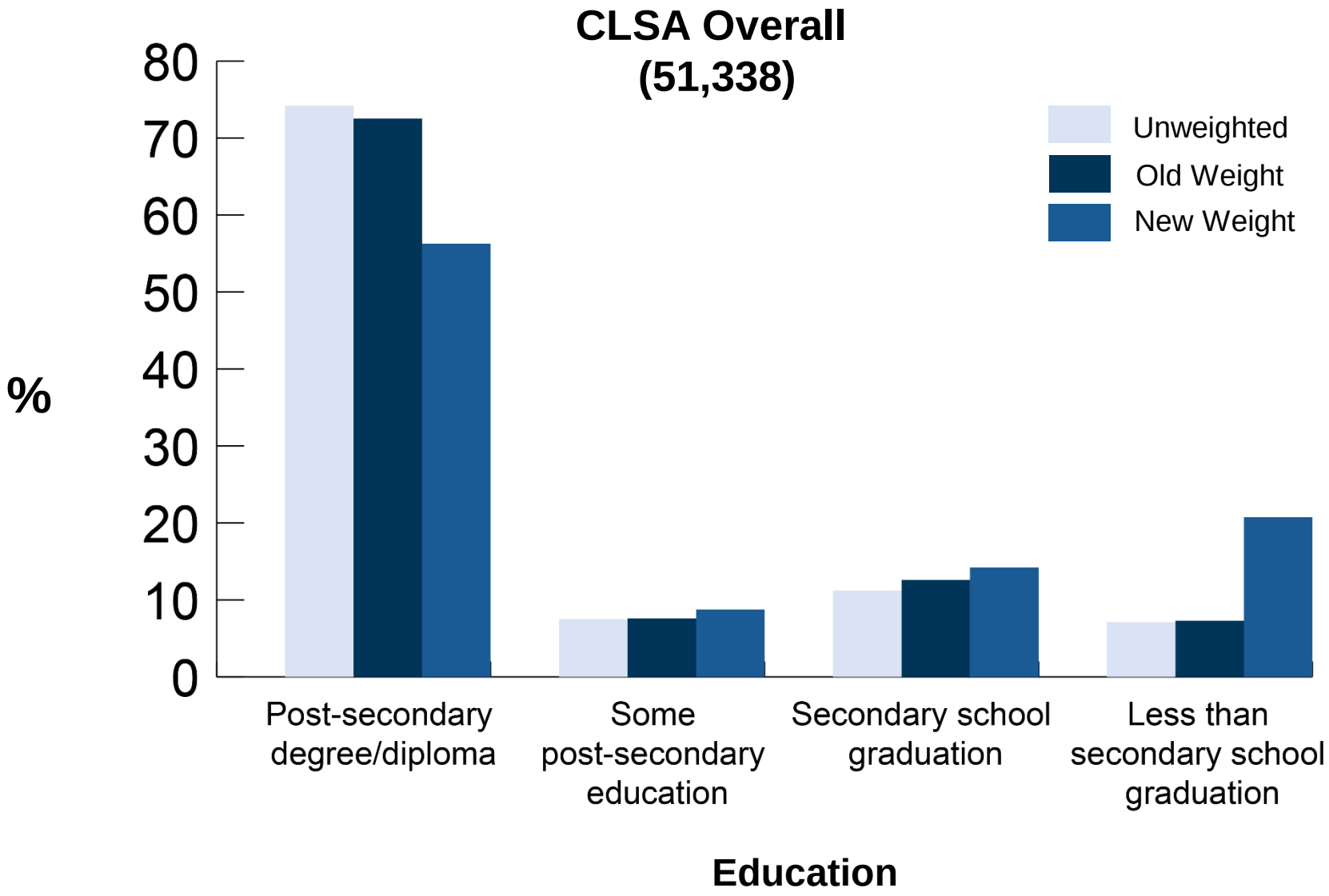
How do the old and new weights differ?

Age categories	CLSA Tracking (21,241)			CLSA Comprehensive (30,097)			CLSA Overall (51,338)		
	Un-weighted %	Old Weight %	New Weight %	Un-weighted %	Old Weight %	New Weight %	Un-weighted %	Old Weight %	New Weight %
45-54	27.46	36.67	38.09	25.24	41.97	39.02	26.15	37.56	38.09
55-64	30.90	30.91	31.40	32.75	29.76	31.02	31.98	30.88	31.40
65-74	21.82	19.64	19.02	24.46	17.16	18.31	23.37	19.17	19.02
75-85	19.82	12.77	11.50	17.56	11.11	11.65	18.50	12.39	11.50
%Female	51.00	51.50	51.80	50.90	50.40	52.40	50.90	51.50	51.83

How do the old and new weights differ?

	CLSA Tracking (21,241)			CLSA Comprehensive (30,097)			CLSA Overall (51,338)		
	Un-weighted %	Old Weight %	New Weight %	Un-weighted %	Old Weight %	New Weight %	Un-weighted %	Old Weight %	New Weight %
Education									
Post-secondary degree/diploma	69.30	72.60	56.50	77.60	79.50	62.10	74.20	72.50	56.27
Some post-secondary education	7.70	7.50	8.50	7.40	6.70	9.10	7.50	7.60	8.76
Secondary school graduation	13.60	12.80	14.60	9.40	9.00	11.50	11.20	12.60	14.22
Less than secondary school graduation	9.40	7.20	20.40	5.50	4.90	17.40	7.10	7.30	20.75

How do the old and new weights differ?



What should a researcher expect?

- In most cases, points estimates of prevalence or association will be similar
- Better reflect the target population, especially in the DCS catchment areas

What should a researcher expect?

- Underestimates of low socioeconomic status will be lessened
- Parameter estimates for variables strongly associated with SES are likely to be more affected
- Overall estimates will better reflect the target population

What should a researcher not expect?

- While data will better reflect the target population in the DCS area, they will not provide estimates at the “city-level”
- DCSs are in only 11 locations across Canada
- DCS catchment areas include a geographic region 25-50 km around the DCS

When will new weights be available?

- New weights will be provided as a separate data release this Fall
- Old weights will no longer be provided except by special request

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

What is coming next?

- Baseline Analytic weights can be used for longitudinal analyses using Baseline and F1
- F1 Inflation weights are being calculated to estimate descriptive parameters at F1
  - Reflect CLSA population attrition
  - Reflect changes in target population from BL to F1

A New CLSA Technical Report

Modelling Complex Survey Data Using R, SAS, SPSS and Stata: A Comparison Using CLSA Datasets

- Available soon

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

CIHR IRSC
Canadian Institutes of Health Research
Instituts de recherche en santé du Canada

INNOVATION.CA
FONDATION CANADIENNE POUR L'INNOVATION
CANADA FOUNDATION FOR INNOVATION

Contact:
Data inquiries: access@clsa-elcv.ca
General inquiries: info@clsa-elcv.ca

CLSA is funded by the Government of Canada through CIHR and CFI, and provincial governments and universities

www.clsa-elcv.ca

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

QUESTIONS?

Upcoming CLSA Webinar

CLSA COVID-19 Research Update

November 2020

Register: clsa-elcv.ca/webinars