How to assess representativeness when using residual specimens?

Example scenario: An investigator wants to know if provincial measles seroprevalence estimates are lower than 90%, in order to evaluate the need for a vaccination campaign. She identified a laboratory at a health facility as a potential source of residual specimens.

Types of information you can use assess representativeness

Setting the Scene

Considerations

Actions / Assessments

What are the sociodemographic variables of interest (e.g., age, place of residence)?

Variables of interest (depends on research question):

- **Measles seropositivity:** outcome of interest
- **Province**: grouping variable
- **Age**: effect modifier, related to measles seropositivity (i.e., opportunities for vaccination and risk of infection) and differs by province

Consider the following:

- ❑ Do the specimens include the full range across the variables of interest? (i.e., all provinces and all ages)
- ❑ Are the variables of interest available and can be matched to the collected specimens?
- ❑ What is the "truth" in the population of interest based on other data sources?

Using the province and age information about the residual specimens, and the provincial age profiles from the census (i.e. "truth"), the investigator can make an initial assessment about the representativeness of the residual specimen source.

What adjustment could be made in the sampling design and/or analysis to minimize biases?

Within the sampling design.

- Sample ART clinic specimens proportional to the "true" HIV prevalence of each province.
- Sample age-specific specimens from each province proportional to the "true" age distribution of each province.

Within the data analysis.

- For example, use inverse probability weights to correct for over or under-sampling of HIV prevalence and age distribution of each province

What other factors could be biasing representativeness of the data?

The identified laboratory processes specimens from the antiretroviral therapy (ART) clinic at the facility

Consider the following:

- ❑ What percent of the specimens collected each month are from the HIV-positive patients at the ART clinic?
- ❑ What is the "true" provincial HIV seroprevalence based on other data sources?

Since HIV positivity is associated with measles seroprevalence and knowing the provincial HIV positivity is considerably lower than the percent of specimens from the ART clinic, the investigator determines collecting all possible specimens will result in a biased estimate.