Sample size calculation for estimating key epidemiological parameters using serological data and mathematical modelling

Stéphanie Blaizot [1], **Sereina A. Herzog** [1], **Steven Abrams** [2,3], **Heidi Theeten** [4], **Amber Litzroth** [5], **Niel Hens** [1,2,3]

1. Centre for Health Economics Research and Modelling Infectious Diseases (CHERMID), Vaccine and Infectious Disease Institute, University of Antwerp, Belgium; 2. Global Health Institute, Department of Epidemiology and Social Medicine, University of Antwerp, Belgium; 3. Center for Statistics, I-BioStat, Hasselt University, Belgium; 4. Center for the Evaluation of Vaccination (CEV), Vaccine and Infectious Disease Institute, University of Antwerp, Belgium; 5. Service of Epidemiology of infectious diseases, Scientific Directorate Epidemiology and Public Health, Sciensano, Brussels, Belgium

contact: sblaizot@yahoo.fr

publication: BMC Med Res Methodol 2019; 19:51

Introduction

The epidemiology of infectious diseases is commonly studied by cross-sectional serological surveys. Several key epidemiological parameters, e.g. seroprevalence or force of infection (FOI), can be computed through mathematical models using data from these surveys. The restriction on serum availability and/or financial resources raises several questions:

- Which samples in a serum bank should be tested, e.g. how should the age-based sampling be structured?
- Does this selection depend on the epidemiological parameter of interest and are there differences between pathogens?

Objective

To determine how to allocate a given number of samples over various age groups in order to estimate key epidemiological parameters with an acceptable precision level in a cross-sectional serological survey.

Methods

- We performed simulation-based sample size calculations.[1] Our calculations are based on Belgian serological survey data collected in 2001-2003 where testing was done for the presence of IgG antibodies against measles, mumps, rubella, varicella-zoster virus (VZV), and parvovirus B19 (PB19).[2]
- Different statistical and mathematical models (**Table 1**) were fitted to get age group specific "true" seroprevalence and FOI for the simulation study.

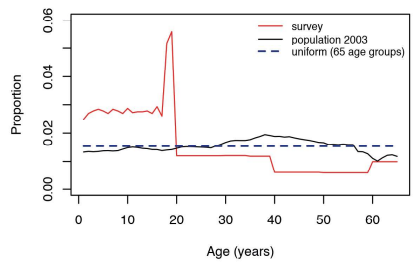
Table 1: Description of models.

Infection	Models	Setting
Measles Mumps Rubella	- Logistic model with piecewise constant prevalence	Non-endemic
VZV	- MSIR piecewise constant force of infection - Exponentially damped model	Endemic
PB19	- MSIR piecewise constant force of infection - Exponentially damped model - MSIR model with boosting and waning	Endemic

MSIR = Maternally-derived immunity - Susceptible - Infectious - Recovered

- We investigated whether the same or different age-based sampling structure (**Fig.1**) would be chosen for different epidemiological parameters using precision as a criterion.

Figure 1: Comparison of three age structures.

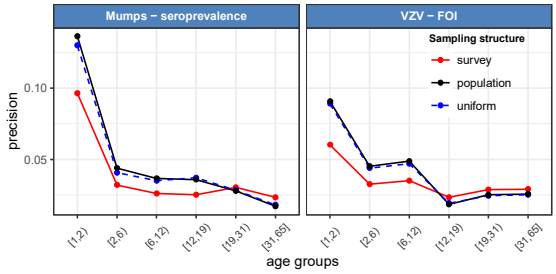


- Precision defined to be half the length of the 95% percentile-based confidence interval (CI) calculated over 500 simulations. Optimal allocation determined by calculating the precisions obtained using different age structures.

Results

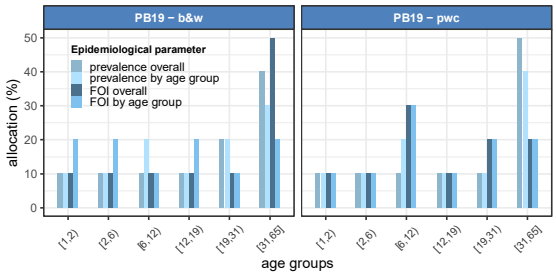
- The best age structure / optimal allocation to use in the sampling of a serological study varied with the epidemiological parameter of interest and with the infection.[1]
- ***Best age-based sampling structure*** for seroprevalence and FOI by age group: survey-based better for young and uniform/population-based better for oldest age groups.
  Figure 2 exemplifies this for mumps (seroprevalence) and VZV (FOI, exponentially damped model) for a sample size N=3300. Note, lower value = better precision.

Figure 2: Best age-based sampling structure for seroprevalence and FOI by age group - mumps & VZV.



- **Figure 3** illustrates the variability in ***optimal allocation*** of a fixed sample size among age groups for 4 key epidemiological parameters: PB19 with MSIR model boosting & waning (b&w) and piecewise constant FOI (pwc); sample size N=3300.

Figure 3: Optimal allocation of fixed sample size among age groups - PB19.



Conclusion

- Simulation-based calculations in combination with mathematical modelling can be used for choosing the optimal allocation of a given number of samples.
- The choice of sampling design should be adapted to prior knowledge about the infection.
- Attention should be given to the age-based sampling structure when estimating epidemiological parameters with acceptable levels of precision within the context of a single cross-sectional serological survey

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 682540 – TransMID).

References

[1] Blaizot S. BMC Med Res Methodol. 2019; 19:51. [2] Nardone A. Euro Surveill. 2004; 5-7.

Interuniversity Institute for Biostatistics and statistical Bioinformatics

Health Economics & Modelling Infectious Diseases
Vaccine & Infectious Disease Institute
University of Antwerp