Estimating the force of infection and incidence using routine data affected by outcome dependent sampling

Sereina A. Herzog, Steven Abrams, Niel Hens

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, KU Leuven, Hasselt University, Belgium.

contact: sereina.herzog@uantwerpen.be

Introduction

In routine data, we are often faced with outcome dependent sampling (ODS), i.e. the probability to have a further observation from the same person depends on the outcome of the current one.

For example, Austrian prenatal care includes screening for the early detection and treatment of toxoplasmosis infection:

- first test negative → up to two further tests
- first test positive → no further tests

The ODS in this example leads to an overrepresentation of negative test results which needs to be accounted for when estimating the force of infection (FOI) from (serial) cross-sectional data.

Objective

To determine an estimator for the force of infection (FOI) - for a Susceptible-Infected (SI) course of infection - using routine data and thereby dealing with outcome dependent sampling (ODS).

Methods

- We derived our proposed estimator theoretically while relying on conditional independence assumptions, i.e. the probability to have a new observation depends only on the outcome (negative or positive test) of the previous observation.
- We compare four FOI (λ) estimators (Table 1) differing in considering data structure and ODS. With the estimator ‘our’, we estimate also the probabilities to have a 2nd observation given the outcome of the first observation is negative or positive (ν1|0, ν1|1).
- In contrast, the estimator ‘ws’ assumes that the probability to have a 2nd observation is independent of the current outcome (ν|1).

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
<th>Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>our</td>
<td>Considers ODS &amp; data structure</td>
<td>( \hat{\lambda}^{our} )</td>
</tr>
<tr>
<td>ws</td>
<td>Considers data structure but no ODS</td>
<td>( \hat{\lambda}^{ws} )</td>
</tr>
<tr>
<td>ns</td>
<td>Treats all observations as being independent (no-stratification)</td>
<td>( \hat{\lambda}^{ns} )</td>
</tr>
<tr>
<td>y2</td>
<td>Uses only the first observation from each person</td>
<td>( \hat{\lambda}^{y2} )</td>
</tr>
</tbody>
</table>

We simulated 5000 datasets for each of the 144 simulation settings (Table 2) assuming that all individuals have their 1st and potential 2nd observation at the same time points, i.e. \( t_1, t_2 \).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Number of individuals</td>
<td>1000</td>
</tr>
<tr>
<td>( \tilde{A} )</td>
<td>FOI per time unit</td>
<td>0.01, 0.05, 0.223, 0.511</td>
</tr>
<tr>
<td>( \tilde{A} )</td>
<td>FOI per time unit</td>
<td>1%, 5%, 20%, 40%</td>
</tr>
<tr>
<td>( \nu_{1</td>
<td>0} )</td>
<td>( P[y_2 = 0 \mid \text{1st obs.}] )</td>
</tr>
<tr>
<td>( \nu_{1</td>
<td>1} )</td>
<td>( P[y_2 = 0 \mid \text{1st obs.}] )</td>
</tr>
</tbody>
</table>

Results

- Figure 1A illustrates that the ‘ws’ estimator over- or underestimates the FOI unless there is no ODS (i.e. \( \nu_{1|0} = \nu_{1|1} \)).
- The other estimators (‘our’, ‘ws’, ‘y2’) have a similar magnitude of bias (Figure 1B).
- In general, \( y_2 \) has the highest variance (Figure 1C) as a consequence of information contained in the observations at \( t_2 \).

Often, the MSE(\( \lambda \)) is a bit higher in ‘our’ than in ‘ws’ (not shown) due to the slightly higher variance in ‘our’ (Figure 1C). However, our proposed estimator provides additional estimates for the ODS probabilities (Figure 2).

We estimated for toxoplasmosis \( \lambda^{our} = 0.01154 \) [0.000157], this implies an estimated prevalence of 20.6% in 20-year-old women. We observed similar estimates for \( \lambda^{ws} \) and \( \lambda^{ns} \) but \( \lambda^{y2} = 0.09938 \) was lower. We estimated \( \lambda^{ws} = 0.765 \) [0.00287] and \( \lambda^{ns} = 0.378 \) [0.00473], despite the fact that guidelines imply that these values should be equal to 1 and 0, respectively.

Conclusion

- Our proposed estimator for the FOI performs well in terms of bias and MSE.
- Our approach yields additional estimates with regard to the ODS probabilities.
- Not considering the structure in routine data leads to bias in both directions, either under- or overestimating the FOI.

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 642540 – TransMED).

References