# Incidence Estimation for Uncertain Events

**Sarah Perkins, 2021, Ecology and Evolutionary Biology**
Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam
Funded by CHW under the Internships in Global Health program

## Introduction

•Diagnostic tools often operate with imperfect sensitivity and specificity.

•Thus, two consecutive test results are sometimes required to confirm diagnoses.

•Most studies which require two tests to confirm diagnoses censor data incorrectly, potentially biasing results.

## Objective of the Study

Investigate the bias of different censoring strategies for disease events.

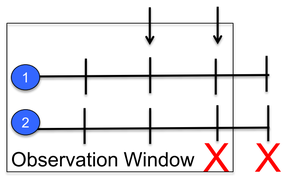


*Figure 1. Data from two-test diagnostic approaches is correctly censored at the penultimate observation, but commonly censored at the ultimate observation.*

## Methods

• Use R to simulate and analyze existing data for uncertain events such as HCV infection.

•Compare bias in incidence rates estimated for this data via different censoring strategies.

•Investigate errors underlying bias.

## Results

100% sensitivity and specificity

| | True | II.e | II | I |
|---|---|---|---|---|
| 6 mos. | 5 | 3.0 | 5.1 | 5.0 |
| 12 mos. | 5 | 3.0 | 5.0 | 5.0 |
| 6 mos. | 10 | 6.0 | 10.1 | 10.0 |
| 12 mos. | 10 | 6.2 | 10.1 | 10.1 |

90% sensitivity and specificity

| | True | II.e | II | I |
|---|---|---|---|---|
| 6 mos. | 5 | 3.7 | 6.2 | 22.3 |
| 12 mos. | 5 | 3.3 | 5.3 | 13.5 |
| 6 mos. | 10 | 6.5 | 10.7 | 27.0 |
| 12 mos. | 10 | 6.1 | 9.8 | 18.1 |

*Figure 2. Incidence rates per 1000 person-months estimated for simulated diagnostic data. Rates are respectively estimated via correct censoring (II) and incorrect censoring (II.e) with two-test diagnostic analysis and one-test diagnostic analysis (I), and compared with the true rate, where tests occur every 6 or 12 months.*

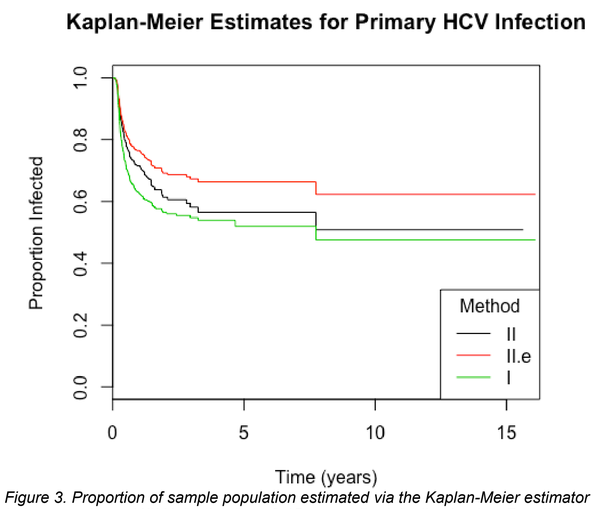


*Figure 3. Proportion of sample population estimated via the Kaplan-Meier estimator to have a primary HCV infection from 0-15 years after infection begins. Results are included for the three main methods of analysis described above. Data provided by the InC3 study.*

- One-test diagnostic analysis demonstrates significant upward bias at suboptimal sens./spec.
- Two-test diagnostic analysis with correct censoring yields only slight bias at suboptimal sens./spec.
- Incorrect censoring with two-test analysis produces downward bias in unidirectional and bidirectional simulations
- Visit rate significantly affects bias in bidirectional simulations

100% sensitivity and specificity

| | True | II.e | II | I |
|---|---|---|---|---|
| Infection | 10 | 9.3 | 9.5 | 9.9 |
| Clearance | 5 | 4.6 | 4.7 | 4.9 |

*Figure 4. Incidence rates per 1000 person-months estimated for simulated bidirectional diagnostic data, where tests occur every 2 months.*

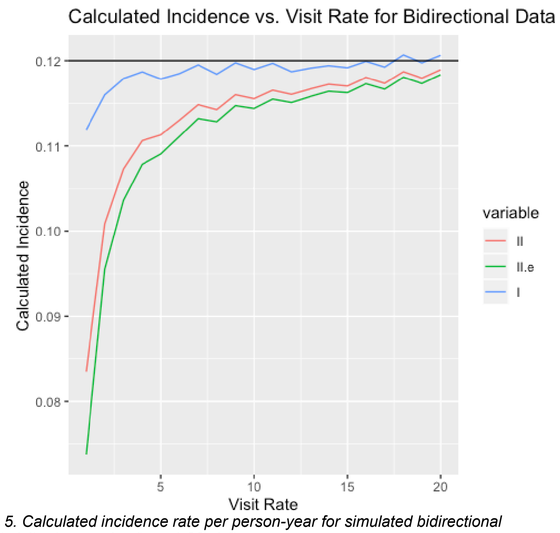


*Figure 5. Calculated incidence rate per person-year for simulated bidirectional data at visit rates varying from 1-20 visits per year.*

## Discussion

• In the two-test analysis of simulated data, incorrectly censoring at the ultimate event contributed to significant downward bias in incidence estimates.

•This downward bias is apparent in real HCV primary infection data and further persists in simulated bidirectional data.

## Questions

- How much bias is observed across censoring methods for diverse data types, such as body temperature data?
- Is the two-test diagnostic standard appropriate for bidirectional data when sens./spec are suboptimal?

## Conclusion

When two tests are used for diagnosis, the popular but incorrect data censoring method is likely to yield significantly biased results.

## Acknowledgements

I gratefully acknowledge the mentorship and guidance of Ronald Geskus and the Biostatistics Group at the Oxford University Clinical Research Unit. I furthermore acknowledge the support of the Global Health Program and the Center for Health and Wellbeing at Princeton University, and the data provided by the InC3 study.