Session 5

How to interpret the latent class

Matt Denwood

2021-06-30

Recap

- Adding more populations and more tests to a Hui-Walter model is technically easy
  - Particualrly if using template_huiwalter
- Verifying that the assumptions you are making are correct is harder
  - The sensitivity and specificity must be consistent across populations
  - Pairwise correlation between tests should be accounted for (with $>$2 tests)

How to interpret the latent class

What exactly is our latent class?

Homework (reminder): think about what exactly the latent class is in these situations:

1. An antigen plus antibody test

What exactly is our latent class?

Homework (reminder): think about what exactly the latent class is in these situations:

1. An antigen plus antibody test

- The latent status is probably close to the true disease status

What exactly is our latent class?

Homework (reminder): think about what exactly the latent class is in these situations:

1. An antigen plus antibody test
   - The latent status is probably close to the true disease status
2. Two antibody tests

What exactly is our latent class?

Homework (reminder): think about what exactly the latent class is in these situations:

1. An antigen plus antibody test

- The latent status is probably close to the true disease status

2. Two antibody tests

- The latent status is actually 'producing antibodies' not 'diseased'

What exactly is our latent class?

Homework (reminder): think about what exactly the latent class is in these situations:

1. An antigen plus antibody test

- The latent status is probably close to the true disease status

2. Two antibody tests

- The latent status is actually 'producing antibodies' not 'diseased'
- What do we mean by "conditionally independent" (revisited) ?
    - Independent of each other conditional on the latent state
    - But the latent state is NOT always *disease*

A hierarchy of latent states

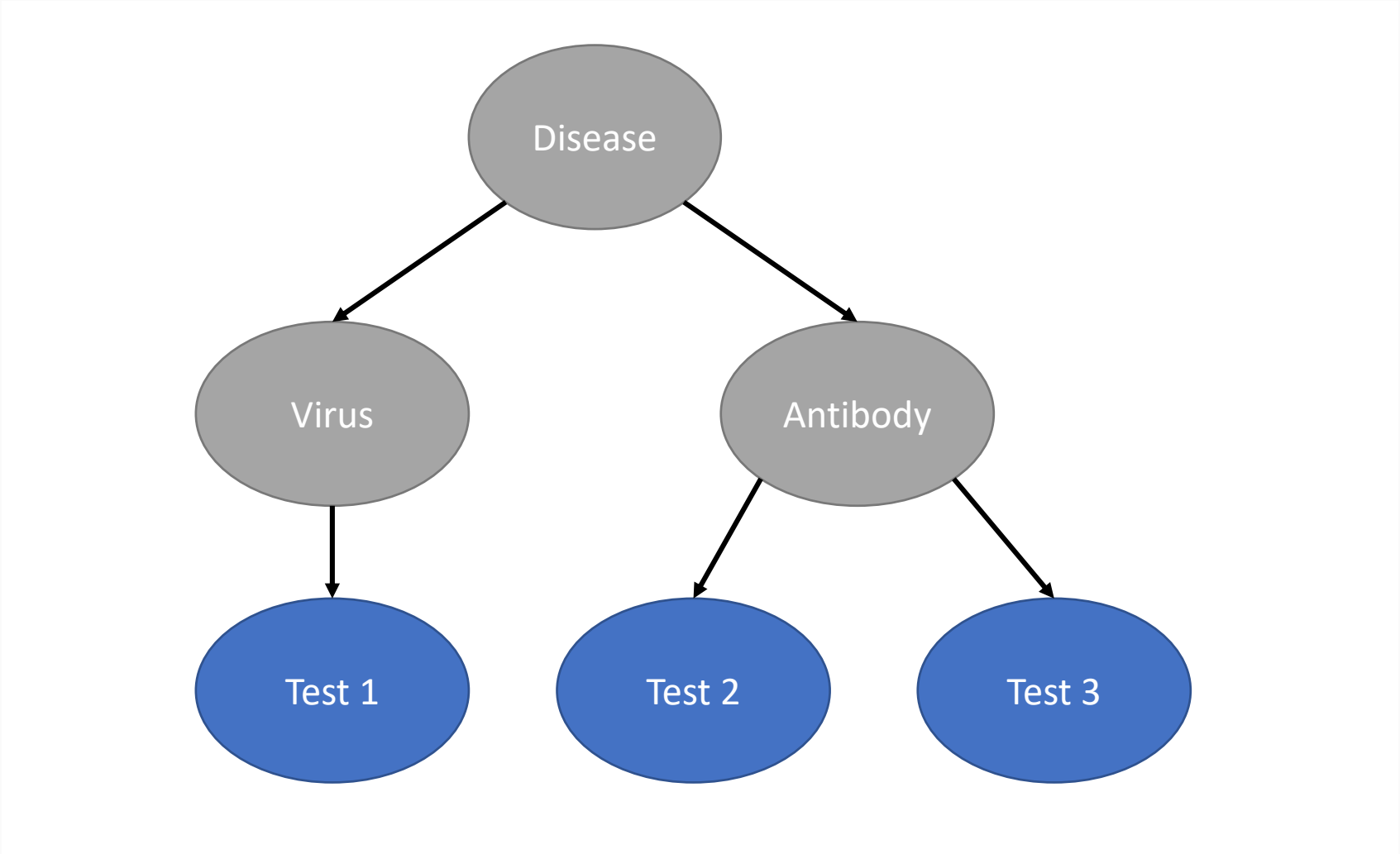

Figure 1: DAG with 3 tests and 2 intermediate states

Branching of processes leading to test results

- Sometimes we have multiple tests that are detecting a similar thing
  - For example: two antibody tests and one antigen test
    - The antibody tests will be correlated

Branching of processes leading to test results

- Sometimes we have multiple tests that are detecting a similar thing
    - For example: two antibody tests and one antigen test
        - The antibody tests will be correlated
- Sometimes we have multiple tests on the same site / sample:
    - For example: two throat swab tests vs a nasal swab test
        - The throat swab tests will be correlated

Branching of processes leading to test results

- Sometimes we have multiple tests that are detecting a similar thing
    - For example: two antibody tests and one antigen test
        - The antibody tests will be correlated
- Sometimes we have multiple tests on the same site / sample:
    - For example: two throat swab tests vs a nasal swab test
        - The throat swab tests will be correlated
- Or even three antibody tests where two are primed to detect the same thing, and one has a different target!
    - In this case all three tests are correlated, but two are more strongly correlated

What is sensitivity and specificity?

The probability of test status conditional on true disease status?

What is sensitivity and specificity?

The probability of test status conditional on true disease status?

Or the probability of test status conditional on the latent state?

What is sensitivity and specificity?

The probability of test status conditional on true disease status?

Or the probability of test status conditional on the latent state?

So is the latent state the same as the true disease state?

What is sensitivity and specificity?

The probability of test status conditional on true disease status?

Or the probability of test status conditional on the latent state?

So is the latent state the same as the true disease state?

Important quote:

"Latent class models involve pulling **something** out of a hat, and deciding to call it a rabbit"

- Some Danish guy

Publication of your results

STARD-BLCM: A helpful structure to ensure that papers contain all necessary information

- You should follow this and refer to it in your articles!

Publication of your results

STARD-BLCM: A helpful structure to ensure that papers contain all necessary information

- You should follow this and refer to it in your articles!

If you use the software, please cite JAGS:

- Plummer, M. (2003). JAGS : A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling JAGS : Just Another Gibbs Sampler. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20–22,Vienna, Austria. ISSN 1609-395X. https://doi.org/10.1.1.13.3406

And R:

```
citation()
##
## To cite R in publications use:
##
##   R Core Team (2021). R: A language and environment
##   for statistical computing. R Foundation for
##   Statistical Computing, Vienna, Austria. URL
##   https://www.R-project.org/.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {R: A Language and Environment for Statistical
↪  Computing},
##     author = {{R Core Team}},
##     organization = {R Foundation for Statistical Computing},
##     address = {Vienna, Austria},
##     year = {2021},
##     url = {https://www.R-project.org/},
##   }
##
## We have invested a lot of time and effort in creating
## R, please cite it when using it for data analysis.
## See also 'citation("pkgname")' for citing R packages.
```

And runjags:

```
citation("runjags")
##
## To cite runjags in publications use:
##
##   Matthew J. Denwood (2016). runjags: An R Package
##   Providing Interface Utilities, Model Templates,
##   Parallel Computing Methods and Additional
##   Distributions for MCMC Models in JAGS. Journal of
##   Statistical Software, 71(9), 1-25.
##   doi:10.18637/jss.v071.i09
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {{runjags}: An {R} Package Providing Interface Utilities,
↪  Model Templates, Parallel Computing Methods and Additional
↪  Distributions for {MCMC} Models in {JAGS}},
##     author = {Matthew J. Denwood},
##     journal = {Journal of Statistical Software},
##     year = {2016},
##     volume = {71},
##     number = {9},
##     pages = {1--25},
##     doi = {10.18637/jss.v071.i09},
##   }
```

Discussion session 5

Points to consider

1. Interpreting the results of latent class models is much more difficult than running them
2. How can we be sure that e.g. `probability of a positive test result conditional on the latent state` is the same thing as `sensitivity`?
3. How can we make sure that our publications contain all of the necessary information to allow others to interpret our findings?

Exercise

1. Read the STARD-BLCM guidelines, checklist, and examples documents. Make sure you understand what the documents ask for.
2. Read the *Diagnosing diagnostic tests* paper provided for Day 3, and try to understand how the issues discussed in this paper relate to what we have discussed yesterday and today.
3. Be ready with questions for the group discussion! You get several bonus points if you can ask a question that the paper authors are unable to answer.

Summary

- Latent class models are MUCH more complex to interpret than traditional models
    - Take time to think about what the latent class means
- Think about which tests might be correlated and if you should include covariance terms
- Think about the biology of where your data comes from, particularly if populations are fundamentally different
- Follow the STARD checklist!