

Session 5

How to interpret the latent class

Matt Denwood

2022-06-09

Recap

- Adding more populations and more tests to a Hui-Walter model is technically easy
 - Particularly if using `template_huiwalter`
- Verifying that the assumptions you are making are correct is harder
 - The sensitivity and specificity must be consistent
 - Pairwise correlation between tests should be accounted for

How to interpret the latent class

What exactly is our latent class?

Think about what exactly the latent class is in these situations:

1. An antigen plus antibody test

What exactly is our latent class?

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?

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?

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'
 - And not 'diseased' !!!

What exactly is our latent class?

Think about what exactly the latent class is in these situations:

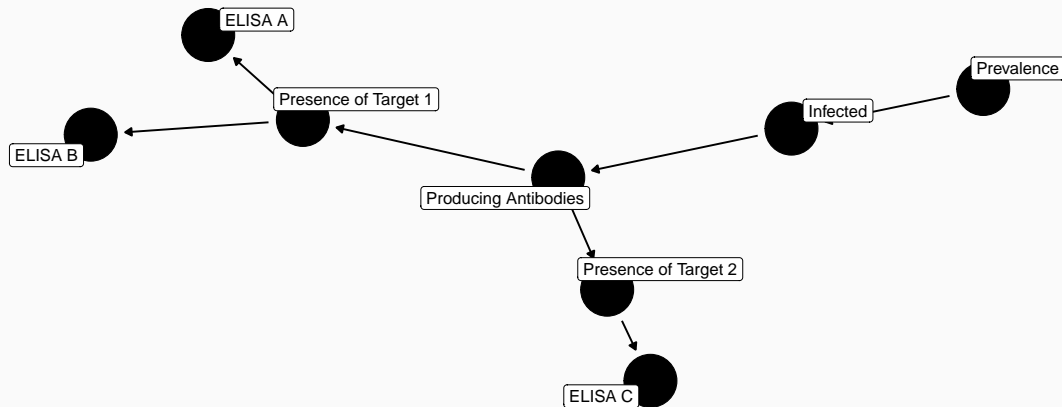
1. An antigen plus antibody test

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

2. Two antibody tests

- The latent status is actually ‘producing antibodies’
 - And not ‘diseased’ !!!
- What do we mean by “conditionally independent”?
 - Independent of each other conditional on the latent state
 - But the latent state is NOT always *disease*

A hierarchy of latent states



Branching of processes leading to test results

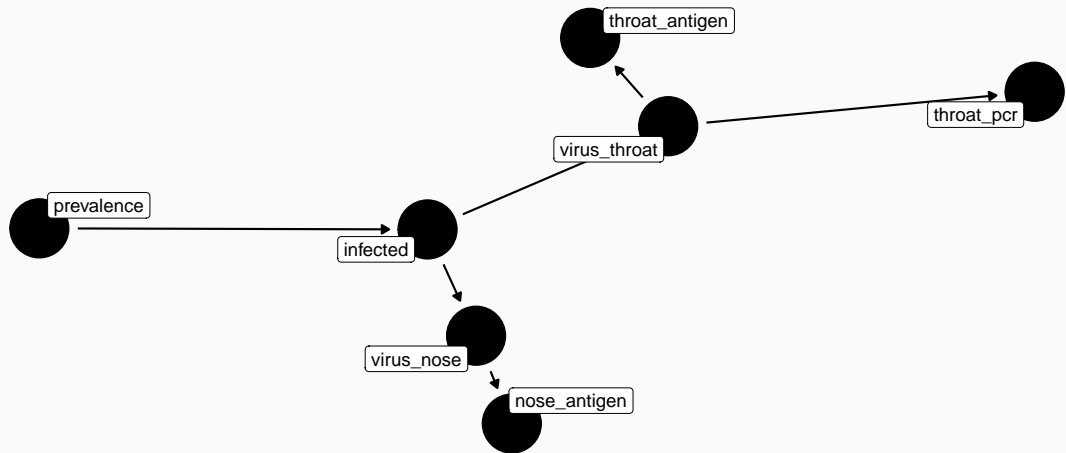
- Sometimes we have multiple tests detecting similar things
 - For example: two antibody tests and one antigen test
 - The antibody tests will be correlated

Branching of processes leading to test results

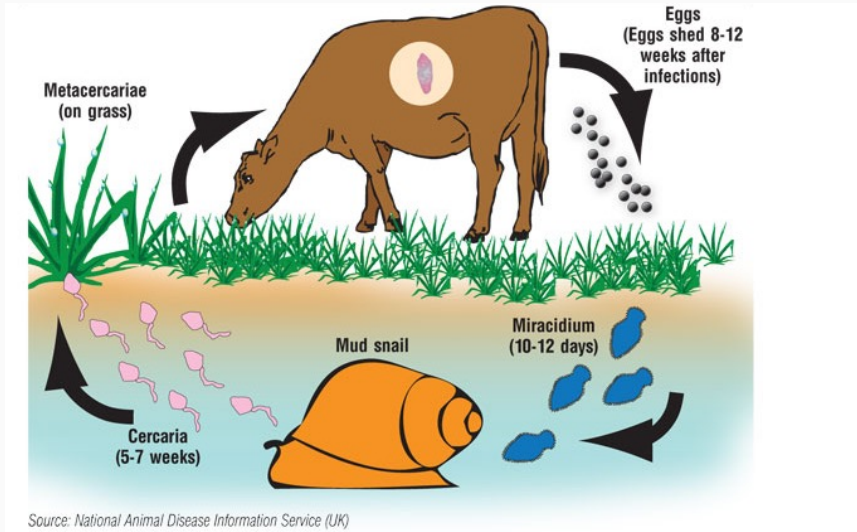
- Sometimes we have multiple tests detecting similar things
 - For example: two antibody tests and one antigen test
 - The antibody tests will be correlated
- Sometimes we have multiple tests on the same site / sample:
 - For example: two throat swab tests vs a nasal swab test
 - The throat swab tests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

Branching of processes leading to test results

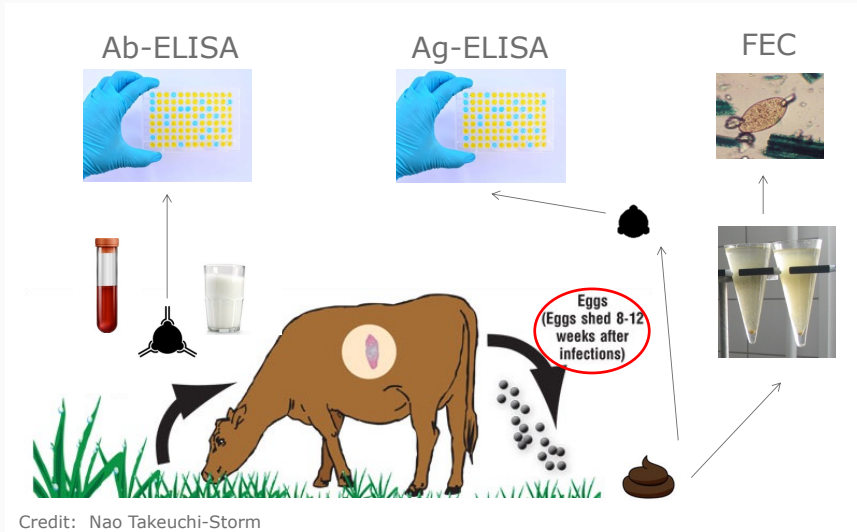
- Sometimes we have multiple tests detecting similar things
 - For example: two antibody tests and one antigen test
 - The antibody tests will be correlated
- Sometimes we have multiple tests on the same site / sample:
 - For example: two throat swab tests vs a nasal swab test
 - The throat swab tests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



Parasites generally have more complex life cycles



So diagnostic tests are more difficult to interpret!



What are the tests detecting?

- Faecal egg counts
 - Detect eggs from adult parasites
 - These are produced 8-12 weeks after infection
 - Eggs may persist in the gall bladder for some weeks after infection has been cleared
- Antigen ELISA
 - Detects presence of maturing/adult parasites in faeces
 - This occurs from 5-8 weeks after infection
 - Parasites only detectable during active infection
- Antibody ELISA
 - Triggered by migrating juveniles and adults
 - Persists (although declining) for several months after infection has been cleared

What is sensitivity and specificity?

- The probability of test status conditional on true disease status?
- 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?
- 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?
- 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”

- Nils Toft

Model complexity

How many parameters are in my latent class model?

- $2 \times k + d \times (d - 1) + p$
- Where k =total tests, d =covarying tests, p =populations

Model complexity

How many parameters are in my latent class model?

- $2 \times k + d \times (d - 1) + p$
- Where k =total tests, d =covarying tests, p =populations

BUT remember the effect of priors:

- $Se \sim \text{dbeta}(1, 1)$ vs. $Se \sim \text{dbeta}(2, 1)$ vs. $Se \sim \text{dbeta}(2978, 234)$

Model complexity

How many parameters are in my latent class model?

- $2 \times k + d \times (d - 1) + p$
- Where k =total tests, d =covarying tests, p =populations

BUT remember the effect of priors:

- $Se \sim \text{dbeta}(1, 1)$ vs. $Se \sim \text{dbeta}(2, 1)$ vs. $Se \sim \text{dbeta}(2978, 234)$

Effective number of parameters:

- pD (and therefore DIC) is ill-defined for LCM
- p_waic (and $WAIC$) is a better bet - will be much easier in JAGS 5.0!

Discussion session 5 - when should we correct for correlation?

Discussion session 5 - when should we correct for correlation?

Aims:

- To get you to think about when and why we should try to correct for correlation
- To get you to think about what is NOT achieved by including correlation terms

Discussion session 5 - when should we correct for correlation?

Aims:

- To get you to think about when and why we should try to correct for correlation
- To get you to think about what is NOT achieved by including correlation terms

Main discussion points for each example:

- What is the latent class
- Which correlation terms should we include and why

Discussion session 5 - when should we correct for correlation?

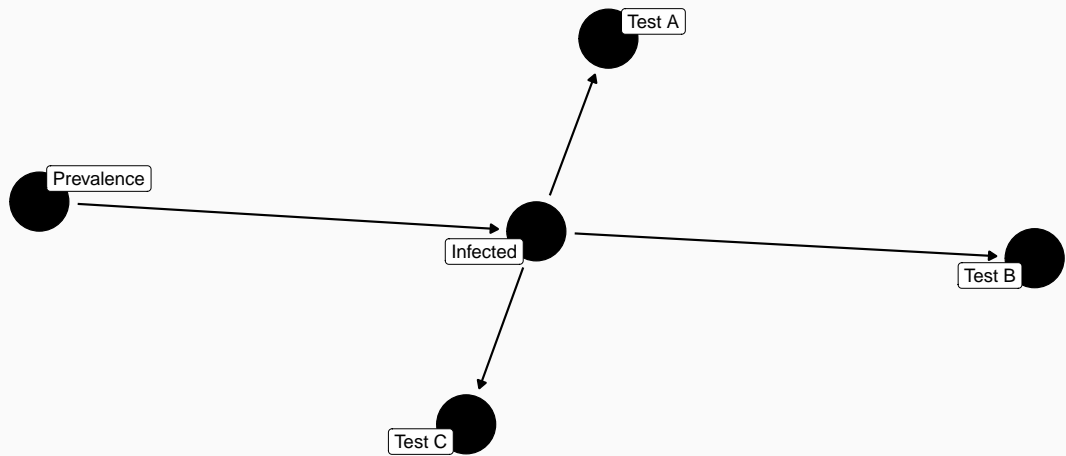
Aims:

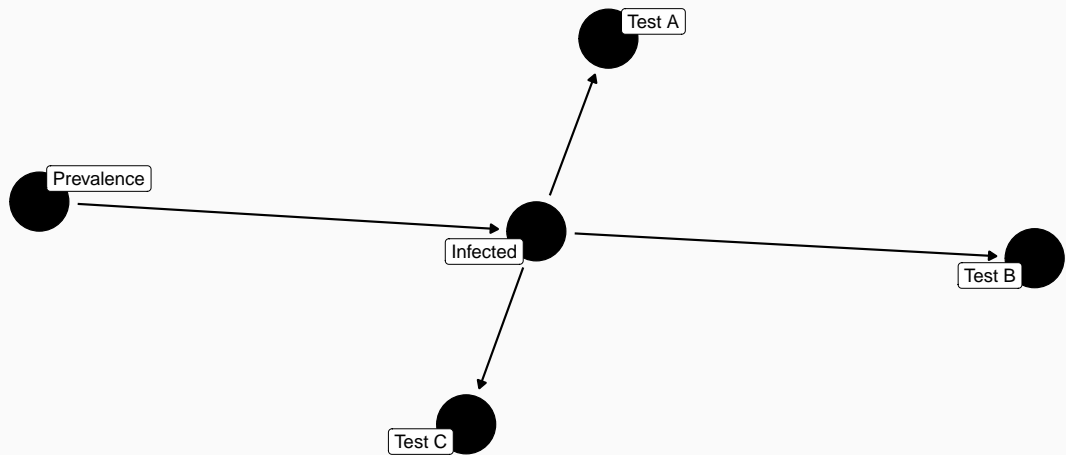
- To get you to think about when and why we should try to correct for correlation
- To get you to think about what is NOT achieved by including correlation terms

Main discussion points for each example:

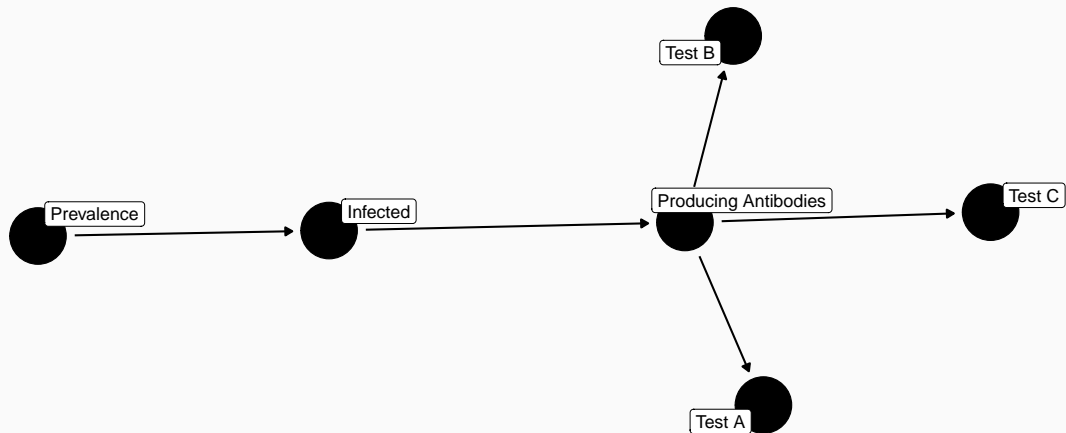
- What is the latent class
- Which correlation terms should we include and why

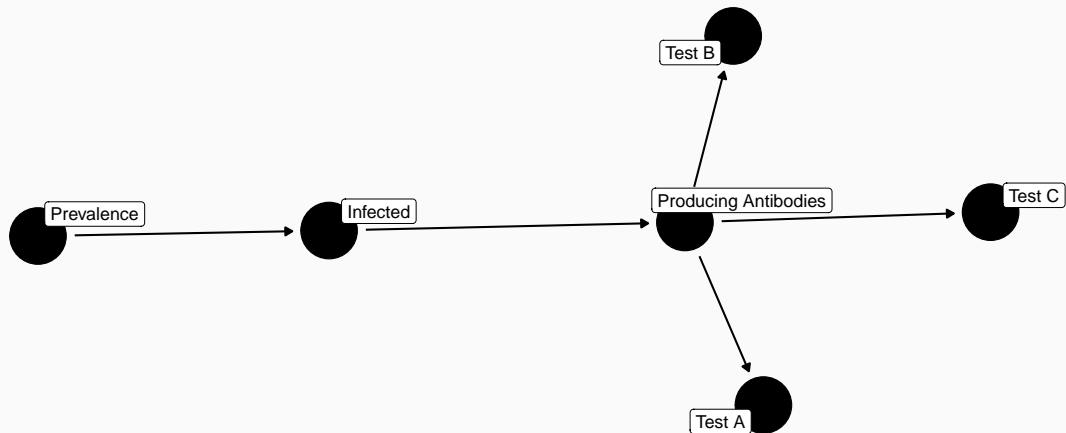
Please feel free to hijack the discussion at any point :)



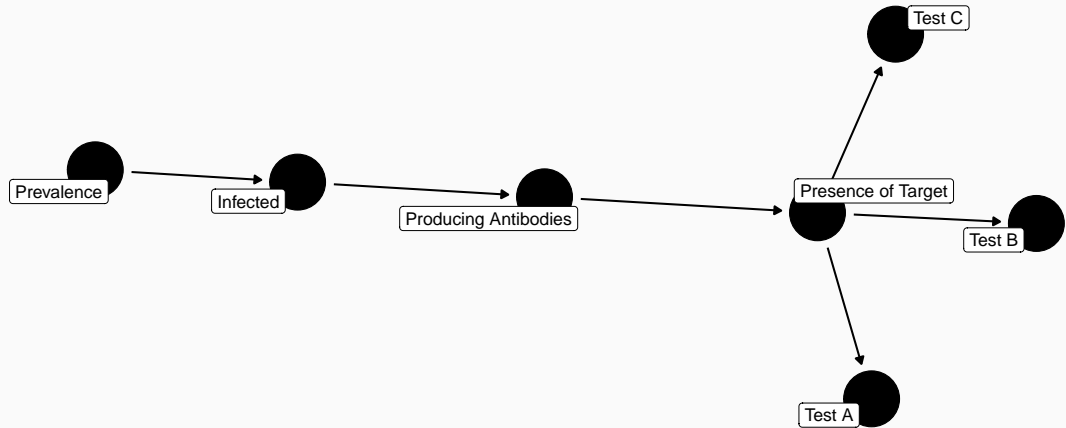


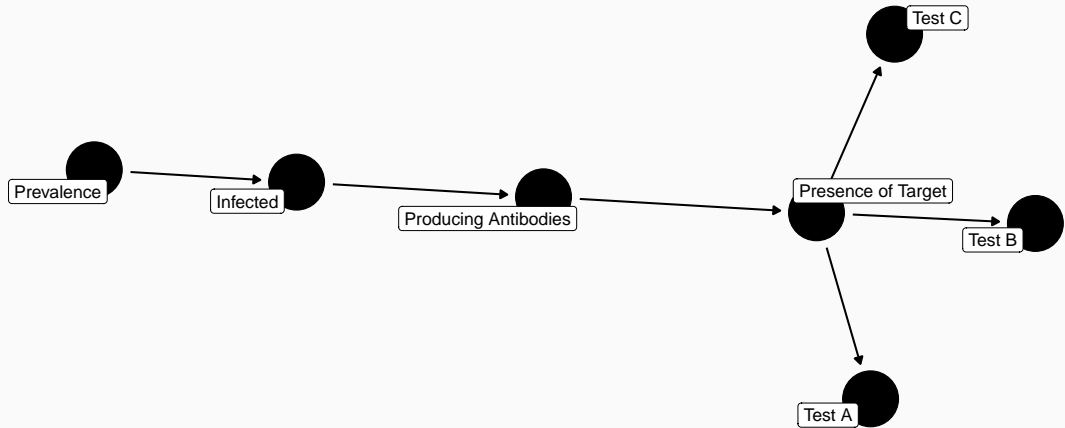
- No correlation to model!



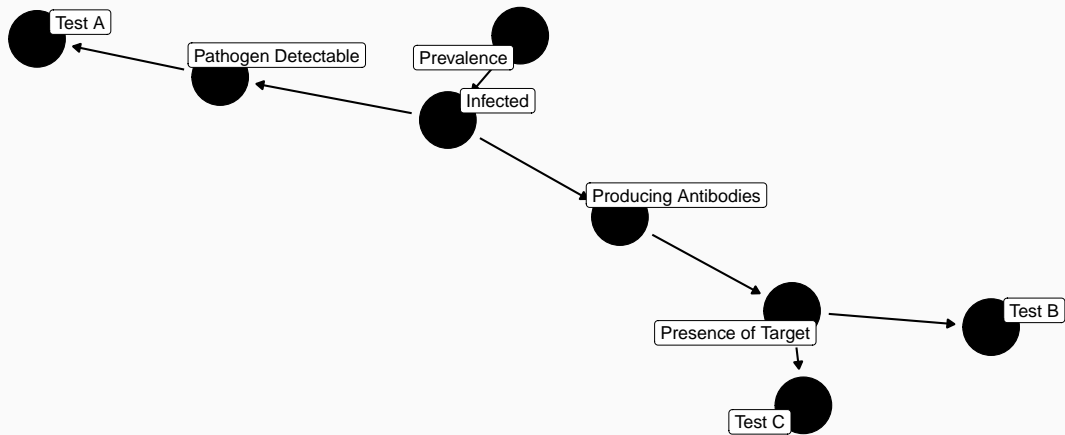


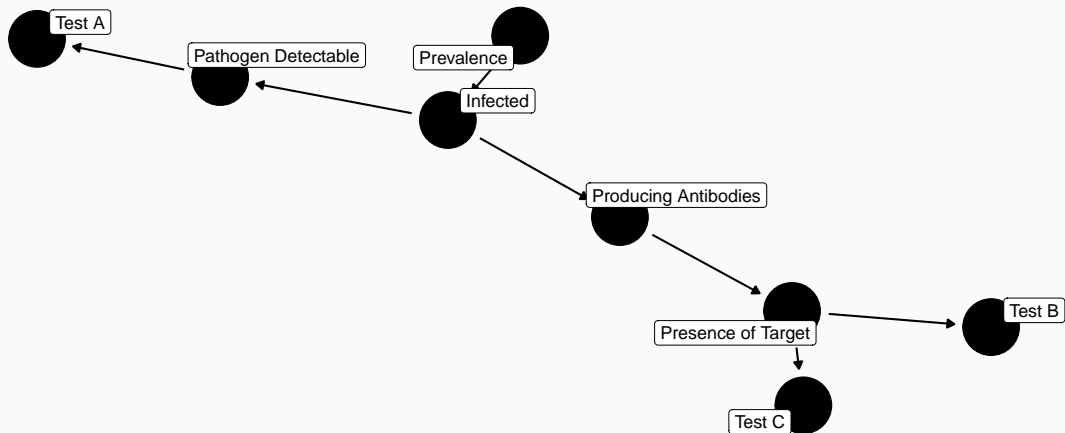
- No correlation to model ... but “infected” is not the latent class



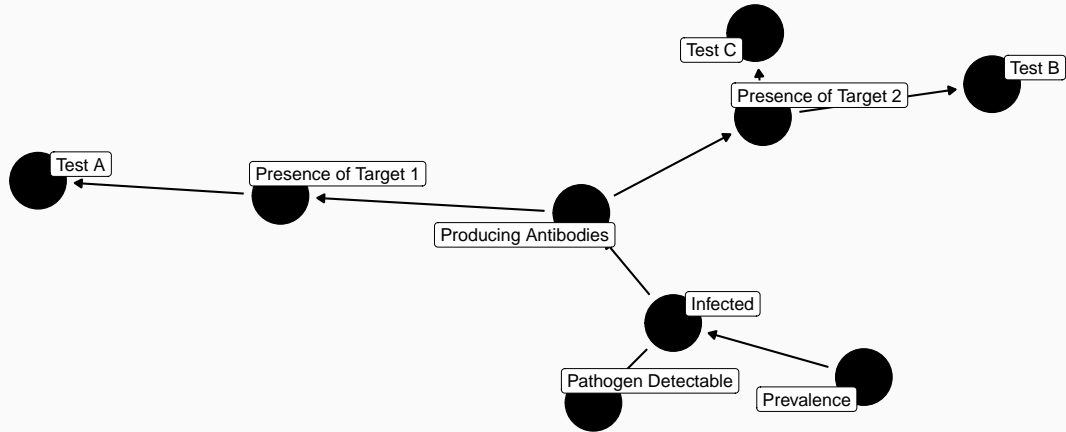


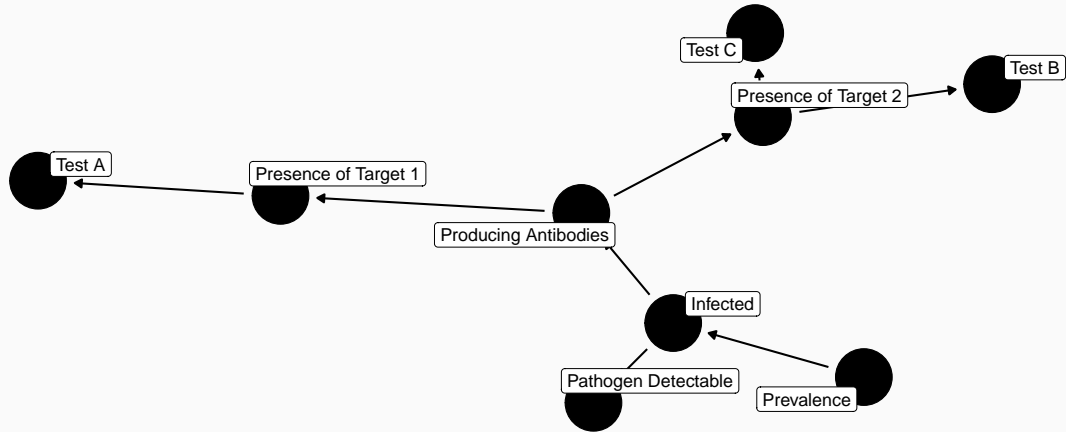
- Same as above!



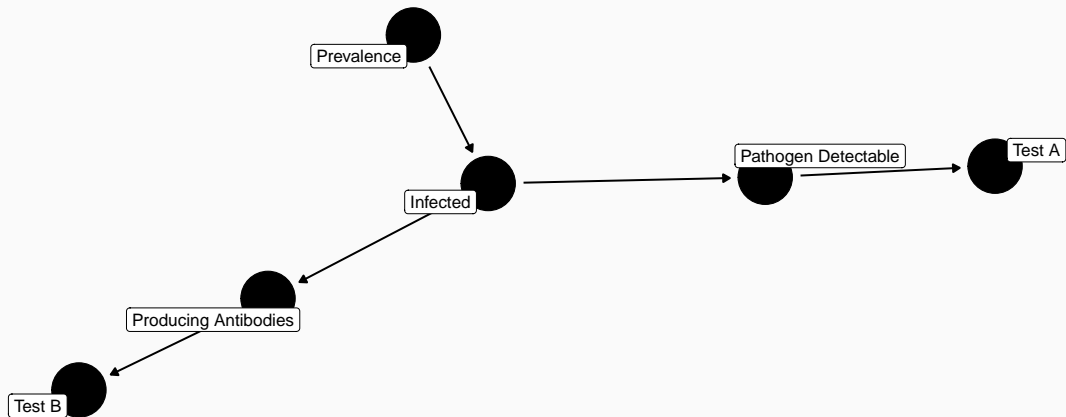


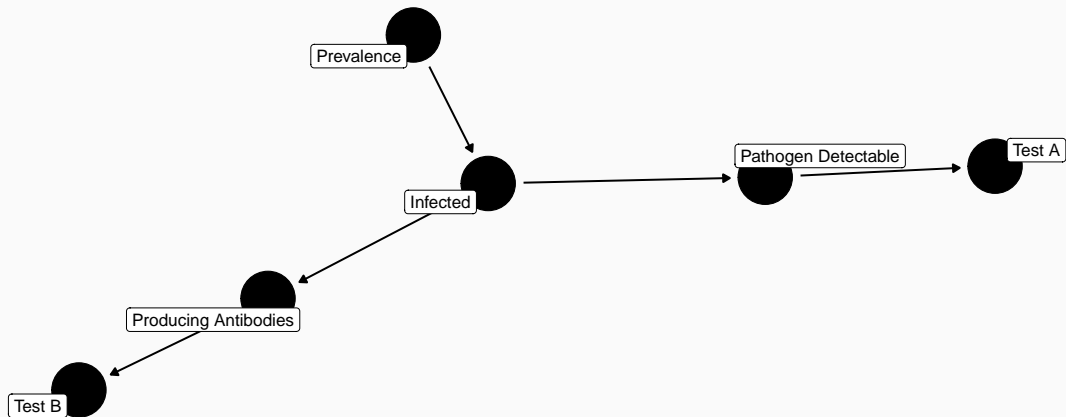
- Tests B and C are correlated



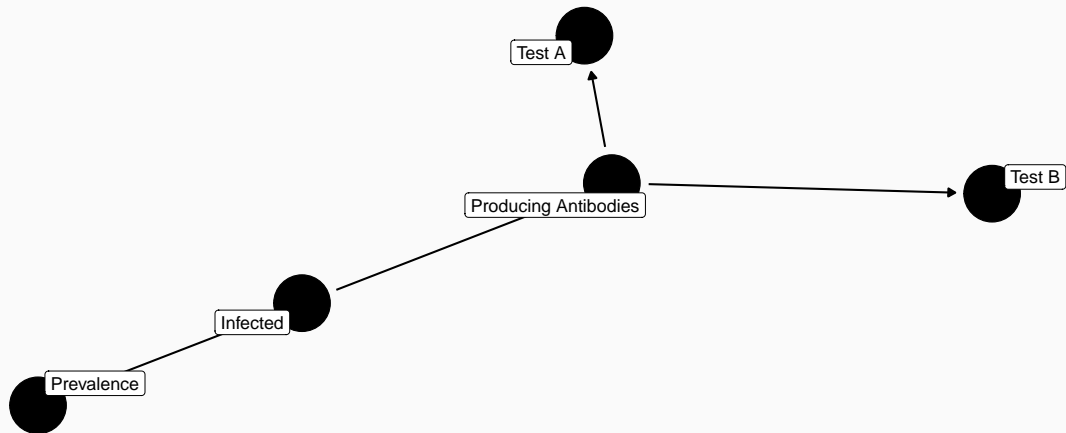


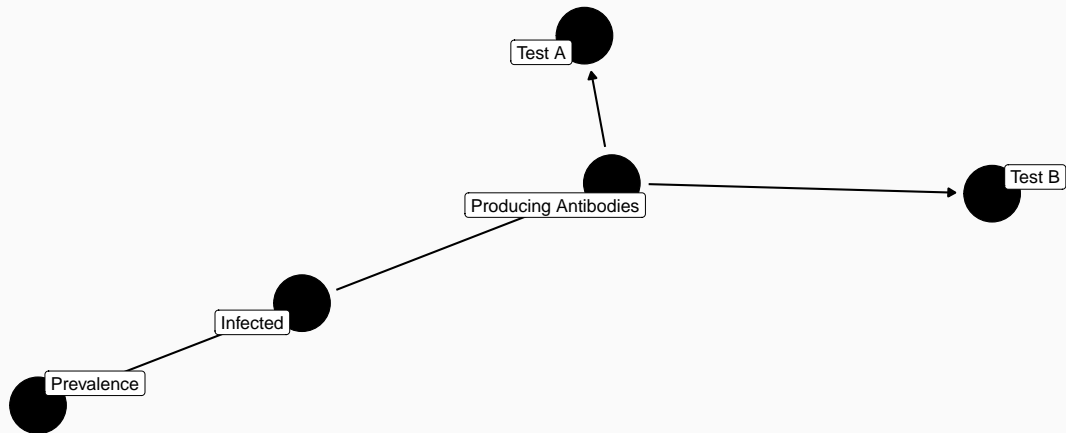
- All tests are correlated with respect to infected BUT infected is not the latent class
- Tests B and C are correlated with respect to antibodies - but maybe not substantially?





- No correlation to model





- No correlation to model - but “infected” is not the latent class

Other examples?

[Insert discussion here. . .]

Summary: between-test correlation structure

My approach:

- If you have < 3 tests then forget correlation BUT you must (always) consider what the latent class really means

Summary: between-test correlation structure

My approach:

- If you have <3 tests then forget correlation BUT you must (always) consider what the latent class really means
- Otherwise, start with the biology. Which pairwise correlation terms are plausible?

Summary: between-test correlation structure

My approach:

- If you have <3 tests then forget correlation BUT you must (always) consider what the latent class really means
- Otherwise, start with the biology. Which pairwise correlation terms are plausible?
- If you have a LOT of terms then consider eliminating some based on the posterior being close to zero . . . but check that other estimates do not change substantially between “full” and “reduced” models

Summary: between-test correlation structure

My approach:

- If you have <3 tests then forget correlation BUT you must (always) consider what the latent class really means
- Otherwise, start with the biology. Which pairwise correlation terms are plausible?
- If you have a LOT of terms then consider eliminating some based on the posterior being close to zero ... but check that other estimates do not change substantially between “full” and “reduced” models
- I dislike DIC ... maybe WAIC is better?

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

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

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 (2022). 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 = {2022},  
##   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},
```