Multi-test|Multi-pop - Adjusting for conditional dependencies between tests

CA18208 HARMONY Zurich Training School - https://harmony-net.eu/

Julio Alvarez Eleftherios Meletis

2022-09-01

Recap Hui-Walter

- Model assumptions
  - The population is divided into two or more populations with different prevalences in which two or more tests are evaluated
  - *Se* and *Sp* are the same in all populations
  - The tests are *conditionally independent* given the disease status.

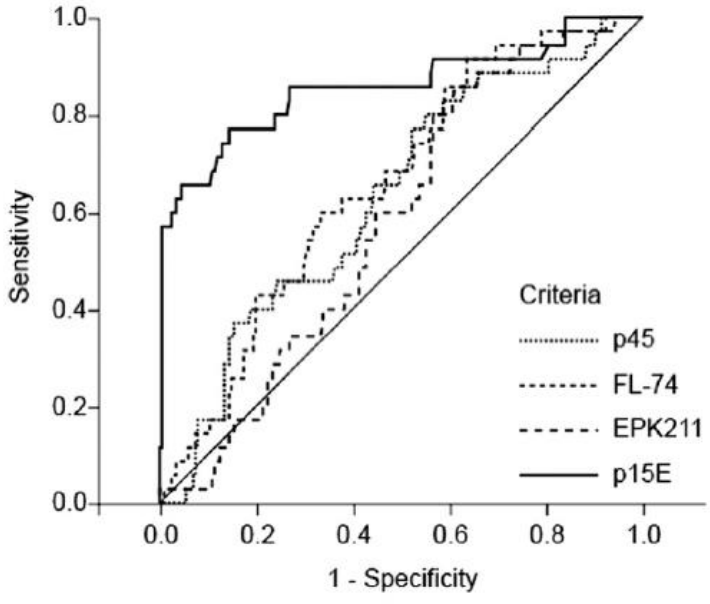
Criticism of the Hui-Walter paradigm assumption (1)

- The assumption (1) of distinct prevalences is necessary for the Hui-Walter model because otherwise, the data can be collapsed into a single 2x2 table with only three degrees of freedom for estimation.
- The smaller the difference between disease prevalences, the larger are the posterior credible intervals, indicating a loss in precision.
- The smallest difference in prevalence assessed by simulation was 10%. In case of rare diseases, it might be difficult to find populations with prevalences higher than 10%.

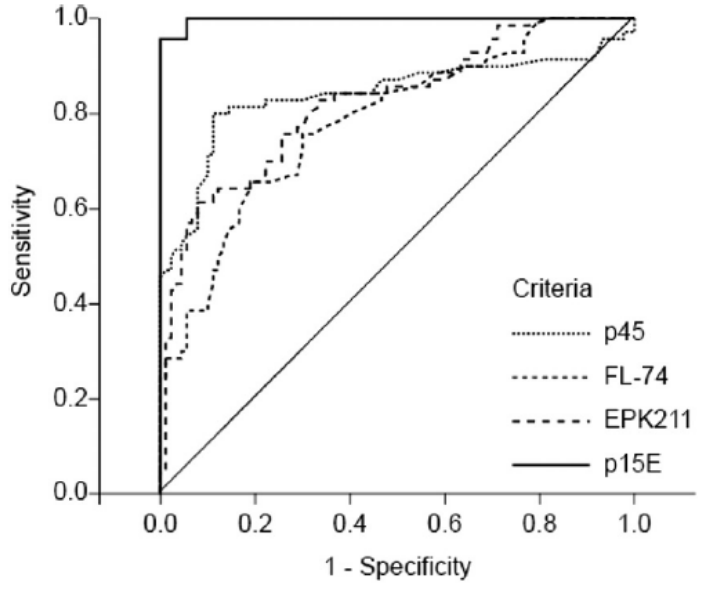
Criticism of the Hui-Walter paradigm assumption (2)

- In reality, the assumption of constant *Se* and *Sp* is questionable as already illustrated in the figure. Here the same diagnostic tests have been applied in the same laboratory to experimentally and naturally infected cats.

DOI: https://doi.org/10.1128/JCM.02584-13



Naturally infected cats



Experimentally infected cats

Criticism of the Hui-Walter paradigm assumption (2)

- If assumption (2) is not satisfied, the accuracies would differ between two populations, this would add four additional parameters to be estimated, while there are only three additional degrees of freedom.
- *Se* and *Sp* are assumed to vary with external factors.
- *Se*, for example, especially when detecting an infectious agent, may depend on the prevalence and the stage of disease.
- The occurrence of a so-called *spectrum bias* contradicts this assumption.

Criticism of the Hui-Walter paradigm assumption (3)

- Assumption (3) demanding conditional independence was the first to be questioned by Vacek (1985).
- Not accounting for potential conditional dependence may lead to misleading, biased estimates with a positive correlation leading to an over-estimation of the test accuracies and a negative of an under-estimation.

Condtional independencies

- Test are considered conditionally independent if the probability of getting a given test result on one test does not depend on the result from the other test, given the disease status of the individual.
- Conditional independence implies that given that an animal is diseased (or not) the probability P of positive (or negative) outcomes for T_1, the test results of the first test, is the same - regardless of the known outcome for the second test, T_2.

Conditional independence

$$P(T_1{}^+, T_2{}^+|D^+) = P(T_1{}^+|D^+) \times P(T_2{}^+|D^+)]$$

Conditional dependence

$$P(T_1{}^+, T_2{}^+|D^+) \neq P(T_1{}^+|D^+) \times P(T_2{}^+|D^+)$$

Conditional dependencies

- Conditional dependence, in contrast, implies that

$$P({T_1}^+|{T_2}^+) \neq P({T_1}^+|{T_2}^-)$$

and / or

$$P({T_1}^-|{T_2}^-) \neq P({T_1}^-|{T_2}^+)$$

Conditional (in)dependencies Interpretation

- Seen from a biological perspective, conditional dependency between two diagnostic tests could occur if both tests are based on the same biological principle.
- For example, the *Sp*s of two ELISAs might be conditionally dependent because they are both affected by the same cross-reacting agent. Another example would be two PCRs utilising fecal material which might contain substances potentially inhibiting the PCR reaction.

Conditional dependencies

- Obviously, conditional dependencies or covariances are additional parameters to be estimated, which in the frequentist situation (without any constraints put on the parameters) would lead to a non-identifiable problem (over-parameterisation).
- Whereas under the assumption of conditional independence at least three tests per sample allowing to estimate seven parameters are needed, under the assumption of conditional dependence 15 parameters need to estimated thus leading to non-identifiability.
- It is of course vital, that the parameters of a latent class model are identifiable to obtain meaningful estimates.

Conditional dependencies

TABLE 2. Maximum Number of Estimable Parameters and Number of Parameters to Be Estimated in the Absence of Conditional Independence and Under Conditional Independence as a Function of the Number of Tests per Subject

Number of Tests	Maximum Number of Estimable Parameters	Parameters to be Estimated Under Conditional Dependence	Parameters to Be Estimated Under Conditional Independence
1	1	3	3
2	3	7	5
3	7	15	7
4	15	31	9
5	31	63	11
h	$2^h - 1$	$2^{h+1} - 1$	$2h + 1$

Berkvens D et al. (2006) Estimating Disease Prevalence in a Bayesian Framework Using Probabilistic Constraints. doi: 10.1097/01.ede.0000198422.64801.8d

Conditional dependencies

- For example for two diagnostic tests named T_1 and T_2 the probabilities of the four different options of binary test results $(++, +-, -+, --)$ including also two conditional dependencies
  - *covs*12, the covariance between the sensitivities of test 1 and 2
  - *covc*12, the covariance between specificities of test 1 and 2)

  could be modelled as follows:

$$P(T_1^+, T_2^+) = [Pr \cdot (Se_1 \cdot Se_2 + covs12) + (1 - Pr) \cdot ((1 - Sp_1) \cdot (1 - Sp_2) + covc12)]^{a_i}$$
$$P(T_1^+, T_2^-) = [Pr \cdot (Se_1 \cdot (1 - Se_2) - covs12) + (1 - Pr) \cdot ((1 - Sp_1) \cdot Sp_2 - covc12)]^{b_i}$$
$$P(T_1^-, T_2^+) = [Pr \cdot ((1 - Se_1) \cdot Se_2 - covs12) + (1 - Pr) \cdot (Sp_1 \cdot (1 - Sp_2) - covc12)]^{c_i}$$
$$P(T_1^-, T_2^-) = [Pr \cdot ((1 - Se_1) \cdot (1 - Se_2) + covs12) + (1 - Pr) \cdot (Sp_1 \cdot Sp_2 + covc12)]^{d_i}.$$

Example of a COVID-19 data set

```
#![](C:\Users\shartn\Documents\Cost Harmony\Cost WS 2021\Trial files\jmv.pdf)
```

Example of a COVID-19 data set

1. how to prepare the data set in the correct format
   - create_data_cassaniti.R
2. how to describe the model
   - model_final.bug
3. how to run the model in JAGS with runjags
   - runjags.version.R
4. how to check convergence
5. how to analyse the data

Exercises

- Ex.3
  - Can you re-run the exercises?
  - Assess what happens if you add other covariances?
  - How many could you add?
  - Try different priors
- Ex. 4 (Bonus)
  - Could you expand the model with a fourth test with simulated data?