Day 1 Hui Walter II

Sonja Hartnack Valerie Hungerbühler

2021-07-13

With some inspiration from Matt Denwood & Giles Innocent from a previous training

Recap Hui-Walter

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

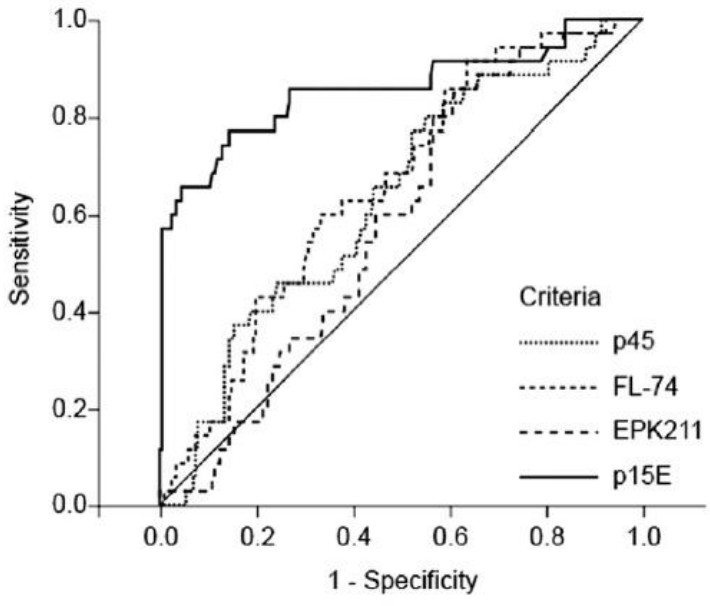
Criticism of the Hui-Walter paradigm assumption (1)

- 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%.

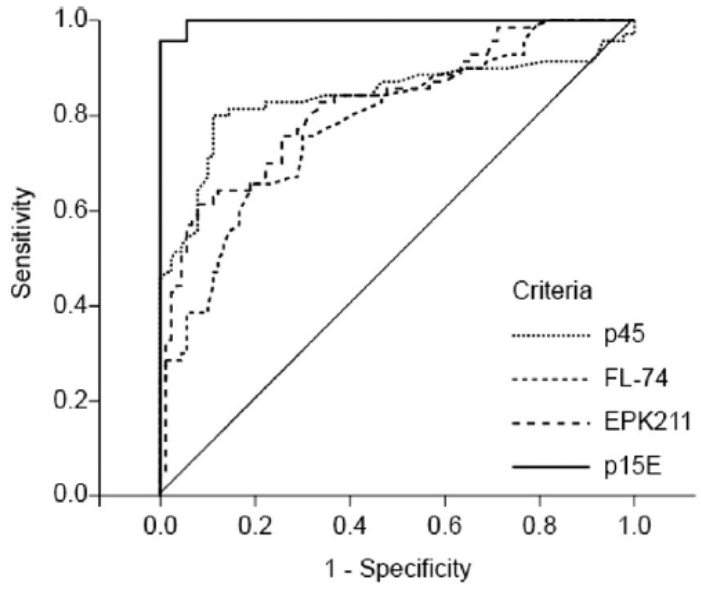
Criticism of the Hui-Walter paradigm assumption (2)

- In reality, the assumption of constant sensitivity and specificity 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.
- Sensitivity and specificity are assumed to vary with external factors.
- Sensitivity, 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 in 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.

Correlated test results

- If there are two tests for detecting a disease, the results can interpreted in one of two ways:
  - *series*, only animals that test positive to both tests are considered test positive.
  - *parallel*, animal that test positive to at least one test are considered positive.
- Series interpretation increases Sp, but decreases Se; whereas parallel testing increases Se and decreases Sp

Example from Veterinary Epidemiologic research, 2nd ed., Dohoo, p. 111-113

- Test are considered conditionally independent if the probability of getting a given test results on one test does not depend on the result from the other test, given the disease status of the individual.
- If tests are conditionally independent, the formulae for Se and Sp under parallel (Se_p, Sp_p) and series (Se_s, Sp_s) interpretation are:
- $Se_p = Se_1 + Se_2 - (Se_1 * Se_2)$
- $Sp_p = Sp_1 * Sp_2$
- $Se_s = Se_1 * Se_2$
- $Sp_s = Sp_1 + Sp_2 - (Sp_1 * Sp_2)$

Example from Veterinary Epidemiologic research, 2nd ed., Dohoo, p. 111-113

Example 5.8 Multiple tests—series versus parallel interpretation
data = ISA_test

The data in this example are from the ISA_test dataset. The tests we are using are the indirect fluorescent antibody test (IFAT) and the polymerase chain reaction (PCR) test, with clinical disease status (see dataset description Chapter 31) as the gold standard. The observed joint distributions of test results and virus presence are shown below along with the 4 possible test interpretation criteria.

	Number of fish by test-result category				Totals
IFAT result	+	+	0	0	
PCR result	+	0	+	0	
Diseased fish	134	4	29	9	176
Non-diseased fish	0	28	12	534	574
Series interpretation	+	0	0	0	
Parallel interpretation	+	+	+	0	

Se of IFAT only = 138/176 = 0.784

Sp of IFAT only = 546/574 = 0.951

Se of PCR only = 163/176 = 0.926

Sp of PCR only = 562/574 = 0.979

Se of series interpretation = 134/176 = 0.761

Se of parallel interpretation = (134+4+29)/176 = 0.949

Sp of series interpretation = (28+12+534)/574 = 1.000

Sp of parallel interpretation = 534/574 = 0.930

Example from Veterinary Epidemiologic research, 2nd ed., Dohoo, p. 111-113

	Number of fish by test-result category				Totals
IFAT result	+	+	0	0	
PCR result	+	0	+	0	
Diseased fish	134	4	29	9	176
Non-diseased fish	0	28	12	534	574
Series interpretation	+	0	0	0	
Parallel interpretation	+	+	+	0	

Se of IFAT only = 138/176 = 0.784

Sp of IFAT only = 546/574 = 0.951

Se of PCR only = 163/176 = 0.926

Sp of PCR only = 562/574 = 0.979

Se of series interpretation = 134/176 = 0.761

Se of parallel interpretation = (134+4+29)/176 = 0.949

Sp of series interpretation = (28+12+534)/574 = 1.000

Sp of parallel interpretation = 534/574 = 0.930

Table 5.3 Expected Se and Sp levels with combined tests for ISA assuming conditional independence (data from Example 5.8)

	Sensitivity		Specificity	
Interpretation	**Expected**	**Observed**	**Expected**	**Observed**
Parallel	0.784+0.926 - 0.784*0.926=0.984	0.949	0.951*0.979=0.931	0.930
Series	0.784*0.926=0.726	0.761	0.951+0.979 - 0.979*0.951=0.999	1.000

Estimating covariances between test results

- Using the Se and Sp estimates from the ISA example, the covariances in the $D+$ and the $D-$ groups are:
  - $covar(+) = Se_s(obs) - (Se_1 * Se_2) = 0.761 - 0.726 = 0.035$
  - $covar(-) = Sp_p(obs) - (Sp_1 * Sp_2) = 0.930 - 0.931 = -0.001$

Conditional dependencies

- Conditional independence implies that given 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 dependence, in contrast, implies that

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

and / or

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

- 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 specificities 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.

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

- 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

- 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

JOURNAL OF
MEDICAL VIROLOGY

LETTER TO THE EDITOR | Open Access

Bayesian latent class models to estimate diagnostic test accuracies of COVID-19 tests

Sonja Hartnack ✉, Paolo Eusebi, Polychronis Kostoulas

First published: 08 August 2020 | https://doi.org/10.1002/jmv.26405

University of Zurich

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?