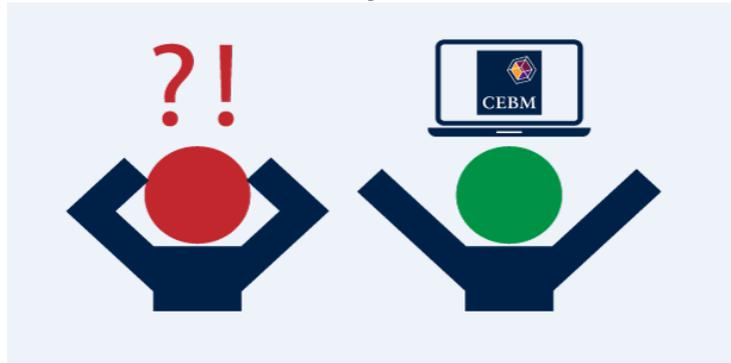# Tip for data extraction in meta-analysis - 1

What can you do if the 2x2 classification table in a diagnostic accuracy study is not reported?

Kathy Taylor

At the Centre for Evidence Based Medicine (CEBM), we often conduct systematic reviews of diagnostic accuracy, a topic with great relevance to primary care and other fields. In this blog post, I'll look at a common problem with extracting data from diagnostic accuracy studies.

First some background. To pool diagnostic accuracy data in a meta-analysis, for each study and every reported combination of index test, reference test, measure and threshold reported, the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) of the diagnostic test are required (Table).

Table. Diagnostic accuracy 2x2 classification

	Disease	No disease	
Test Positive	TP	FP	TP+FP
Test Negative	FN	TN	TN+FN
	TP+FN	FP+TN	Total

TP is the number with the disease who are correctly identified

TN is the number without the disease who are correctly identified

FP is the number without the disease who are incorrectly identified (test positive)

FN is the number with the disease who are incorrectly identified (test negative)

Although most studies don't report this table, fortunately it may be derived from the reported statistics. Let's suppose that the sensitivity, specificity and prevalence have been reported, and it is also clear how many were in the study. The sensitivity is the proportion of diseased patients (1st column in Table) that correctly test disease-positive. The specificity is the proportion of 'healthy' patients (2nd column) that correctly test disease-negative. Prevalence is the proportion of all those in the study who have the disease.

A bit of maths (see below if you're interested) shows us that

$$TP = Sensitivity \ \times Prevalence \ \times Total$$
$$FN = Prevalence \ \times Total \ \times (1 - Sensitivity)$$

$$TN = Specificity \ \times Total \ \times (1 - Prevalence)$$
$$FP = Total \ \times (1 - Prevalence) \times (1 - Specificity)$$

The first calculation (TP) is to multiply the reported sensitivity by the reported prevalence and then by the total number of people. Be careful – statistics such as prevalence can be reported as percentages or as decimal fractions. We're assuming that sensitivity, specificity and prevalence are all reported as fractions.

Let me now show you an example from a review that I worked on. It's a study of the detection of heart failure in a population with prevalence 34% using brain natriuretic peptide (BNP). Measuring BNP with the Biosite Triage device at a threshold of 69 pg/mL, compared to clinical assessment, the sensitivity is reported at 97% and specificity at 44%.

This study reports percentage inputs, so we first need to convert these to decimal fraction inputs by dividing the percentages by 100 (e.g. sensitivity of 97% becomes 0.97). Then applying the equations
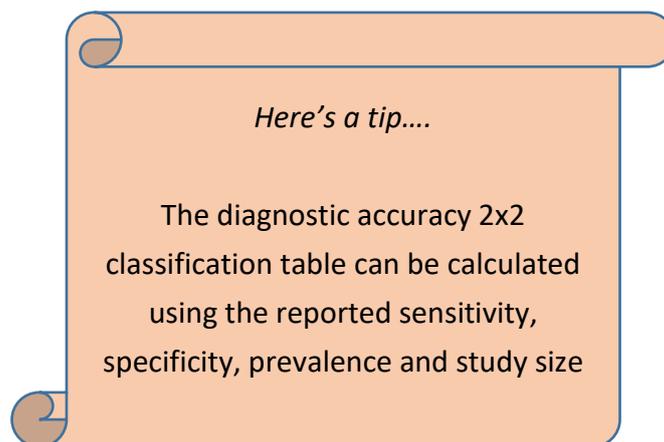
TP = 0.97 x 0.34 x 205 = 67.6
FN = 0.34 x 205 x (1 – 0.97) = 2.1
TN = 0.44 x 205 x (1 – 0.34) = 59.5
FP = 205 x (1 – 0.34) x (1 – 0.44) = 75.8

Note that the equations will not necessarily produce whole numbers because reported statistics are rounded up or down. Rounding up or down to the nearest integer and checking that the total is still 205 we see TP+FN+TN+FP = 68+2+59+76 = 205.

Here's a tip....

The diagnostic accuracy 2x2 classification table can be calculated using the reported sensitivity, specificity, prevalence and study size

In the next post I'll explain what you might do if the prevalence is not reported.

Where did the equations come from?

(You can skip this if you are only interested in carrying out the calculations)

$$Sensitivity = \frac{TP}{TP + FN} = Proportion\ with\ disease\ who\ are\ correctly\ identified. \quad \textbf{equation 1}$$

$$Specificity = \frac{TN}{TN + FP} = Proporion\ with\ no\ disease\ who\ are\ correctly\ identified. \quad \textbf{equation 2}$$

$$Prevalence = \frac{TP + FN}{Total} = Proportion\ with\ the\ disease \quad\quad\quad\quad \textbf{equation 3}$$

$$TOTAL = TP + FN + FP + TN = All\ patients \quad\quad\quad\quad\quad \textbf{equation 4}$$

To calculate TP

Rearrange equation 1:	TP=Sensitivity x (TP+FN)
Rearrange equation 3:	TP+FN=Prevalence x Total **equation 5**
Substitute for TP+FN:	TP=Sensitivity x Prevalence x Total

To calculate FN

Rearrange equation 3:	FN=Prevalence x Total – TP
Substitute for TP:	FN=Prevalence x Total – Sensitivity x Prevalence x Total
Tidy up with some brackets:	FN=Prevalence x Total x (1-Sensitivity)

To calculate TN

Rearrange equation 2:	TN=Specificity x (TN+FP) **equation 6**
Rearrange equation 4:	TN+FP=Total – (TP+FN)
Substitute for TP+FN (equation 5):	TN+FP=Total – Prevalence x Total **equation 7**
Substitute for TN+FP in equation 6 :	TN=Specificity x (Total – Prevalence x Total)
Tidy up with some brackets:	TN=Specificity x Total x (1 – Prevalence)

To calculate FP

Rearrange equation 7:	FP=Total – Prevalence x Total – TN
Substitute for TN:	FP=Total x (1 – Prevalence) – Specificity x Total x (1 – Prevalence)
Tidy up with some brackets:	FP=Total x (1 – Prevalence) x (1– Specificity)

To calculate TP	Rearrange equation 1	*TP=Sensitivity x (TP+FN)*	
	Rearrange equation 3	*TP+FN=Prevalence x Total*	***equation 5***
	Substitute for TP+FN	*TP=Sensitivity x Prevalence x Total*	
To calculate FN	Rearrange equation 3	*FN=Prevalence x Total – TP*	
	Substitute for TP	*FN=Prevalence x Total – Sensitivity x Prevalence x Total*	
	Tidy up with brackets	*FN=Prevalence x Total x (1-Sensitivity)*	
To calculate TN	Rearrange equation 2	*TN=Specificity x (TN+FP)*	***equation 6***
	Rearrange equation 4	*TN+FP=Total –(TP+FN)*	
	Substitute for TP+FN (equation 5)	*TN+FP=Total–Prevalence x Total*	***equation 7***
	Substitute for TN+FP in equation 6	*TN=Specificity x (Total –Prevalence x Total)*	
	Tidy up with brackets	*TN=Specificity x Total x (1–Prevalence)*	
To calculate FP	Rearrange equation 7	*FP=Total –Prevalence x Total – TN*	
	Substitute for TN	*FP=Total x (1– Prevalence) – Specificity x Total x (1–Prevalence)*	
	Tidy up with brackets	*FP=Total x (1–Prevalence) x (1– Specificity)*	

Dr Kathy Taylor teaches data extraction in [Meta-analysis](). This is a short course that is also available as part of our [MSc in Evidence-Based Health Care](), [MSc in EBHC Medical Statistics](), and [MSc in EBHC Systematic Reviews.]()

Follow updates on this blog, related news, and to find out about other examples of statistics being made more broadly accessible on Twitter @dataextips