BMI881 Homework 2

PPV calculation

Grace Yi Chen

08 November, 2022

This is in relation to the paper on predicting nonmelanoma skin cancer, Wang et al. (2019).

Question:

1. If a test has sensitivity = 80% and specificity 80% and the prevalence of the disease is 9/100,000, what is the positive predictive value (aka "precision") of the test?

   Here is the 2 by 2 table calculating sensitivity, specificity, prevalence and precision.

	True Values	
	Yes	No
Yes	A	B
No	C	D

Sensitivity: $\frac{A}{A+C} = 0.8$

Specificity: $\frac{D}{B+D} = 0.8$

Prevalence: $\frac{A+C}{A+B+C+D} = \frac{9}{100000}$

If we want to find $\frac{A}{A+B}$, we could first get

$A = 0.8 \times (A + C)$,

$D = 0.8 \times (B + D)$, $B = \frac{0.2}{0.8} \times D$

Then we will have $A + D = 0.8 \times (A + B + C + D)$

Thus, $\frac{A+C}{A+B+C+D} = \frac{A}{A+D} = \frac{9}{100000}$ So $A = \frac{\frac{9}{100000}}{1-\frac{9}{100000}} \times D$

So $\frac{A}{A+B} = \frac{\frac{\frac{9}{100000}}{1-\frac{9}{100000}}}{\frac{\frac{9}{100000}}{1-\frac{9}{100000}} \times + \frac{1}{4}} = 0.00036$

The positive predictive value (precision) is 0.00036.

2. Suppose sensitivity = specificity. What would they have to be to achieve positive predictive value = 50% when prevalence is 9/100,000?

Suppose we have Sensitivity: $\frac{A}{A+C} = x$. So $A = \frac{x}{1-x} \times C$

Specificity: $\frac{D}{B+D} = x$. So $D = \frac{x}{1-x} \times B$

Prevalence: $\frac{A+C}{A+B+C+D} = \frac{9}{100000}$

Since the sensitivity and specificity are equal, we would have So $A + D = x \times (A + B + C + D)$

Thus, $\frac{A+C}{A+B+C+D} = \frac{A}{A+D} = \frac{9}{100000}$

So $A = \frac{\frac{9}{100000}}{1-\frac{9}{100000}} \times D$

For the positive predictive value, we have $\frac{A}{A+B} = 0.5$, so $A = B$.

Finally, we will get $\frac{\frac{9}{100000}}{1-\frac{9}{100000}} \times \frac{x}{1-x} \times B = B$

The sensitivity and specificity would be $x = 1 - (9/100000)/(1 - 9/100000) = 0.99991$.

3. Comment on these results in relation to the precision values provided in Table 2 of Wang et al. (2019).

Table 2 of Wang et al.(2019) shows high sensitivity, specificity and moderate precision. This table is based on the study sample they collected and the ratio of patients with Nonmelanoma skin cancer (NMSC) to controls without cancer was 1:4. This means the prevalence of NMSC is around 0.2 in the study sample. We should be aware about the calculation of precision since it is sensitive to prevalence and is a bad measure when the positive group has few samples. We could derive precision mathematically according to Bayes rule using sensitivity, specificity and prevalence. Also, from the above two questions, if the disease has a very small prevalence, we know that in order to get a high precision we need to have extremely high sensitivity and specificity. In the real world setting, the prevalence of NMSC is much lower than the study prevalence 0.2. So it is not helpful to report precision in this case. In addition, a precision-recall curve would be more sensitive for a low prevalence dataset than AUC.

When the prevalence of NMSC is low, the precision will also be low, even using a test with high sensitivity and specificity. For such rare diseases, a large proportion of those with positive screening tests will inevitably be found not to have the disease upon further diagnostic testing (high false positive).