ROYAL STATISTICAL SOCIETY

DATA | EVIDENCE | DECISIONS

Possible criteria for MHRA antibody sensitivity and specificity assurance

David Spiegelhalter, on behalf of RSS Covid-19 Task Force

Co-Chair, Royal Statistical Society Covid-19 Task Force
Chair, Winton Centre for Risk and Evidence Communication, University of Cambridge

13th April 2020

Executive Summary

Two issues are addressed concerning the specification for serology point of care tests for SARS-Cov-2 antibodies. First, a suggestion that the required sensitivity and specificity should be specified both in terms of the point estimate and a required threshold for the lower end of the 95% confidence interval. Second, that the current requirement for the specificity is likely to mean that, for the people who are said to have antibodies, a substantial proportion actually do not have antibodies. A more stringent requirement for the specificity, possibly with a less stringent sensitivity, may be appropriate.

****

1 Expanded criterion including threshold for the lower end of the 95% confidence intervals

Current guidance (*Specification criteria for serology point of care tests and self-tests)* states that both sensitivity and specificity should be *"Greater than 98% (within 95% confidence intervals)"*.

This is somewhat ambiguous –

- If it means that 98% should be within a 95% confidence interval, then for specificity this could be achieved by 1 false positive in 5 negative samples (95% interval 30% to 99%)[1].
- If it means that the whole of the 95% interval must be greater than 98%, this would require an observed specificity considerably greater than 98%, for example 1 false positive in 350 negative samples (estimated sensitivity 99.7%, 95% interval 98.1% to 99.9%)

An alternative type of criterion is to specify an additional threshold for the lower end of the 95% confidence interval, for example –

[1] Throughout this document, confidence intervals for proportions have been calculated using the prop.test function in R.

Observed specificity should be greater than 98%, with a 95% confidence interval lying wholly above 96%

In this way there is a 'guaranteed' minimum specificity.

We can explore the implications of such a criterion for a range of lower thresholds and sample sizes, as shown in the table below.

Observed specificity > 98%, lower end of 95% confidence interval above **95%**		Observed specificity > 98%, lower end of 95% confidence interval above **96%**		Observed specificity > 98%, lower end of 95% confidence interval above **97%**	
Number of negative-controls	Number of false positives permitted to achieve criterion	Number of negative-controls	Number of false positives permitted to achieve criterion	Number of negative-controls	Number of false positives permitted to achieve criterion
92 - 125	0	116 - 158	0	156 - 212	0
126 - 157	1	159 - 197	1	213 - 264	1
158 - 186	2	198 - 234	2	265 - 314	2
187 - 215	3	235 - 270	3	315 - 362	3
216+	4	271+	4	363+	4

Table: For different achievement criteria, the number of permissible false-positives in negative-control samples of different size.

For example, if the antibody testing is conducted on 200 negative cases, then achieving 2 false positives would achieve the 96% lower-bound criterion (estimated specificity 99%, interval 96.1% to 99.8%); but 3 false positives would not (estimated specificity 98.5%, interval 95.3% to 99.6%).

Two final points:

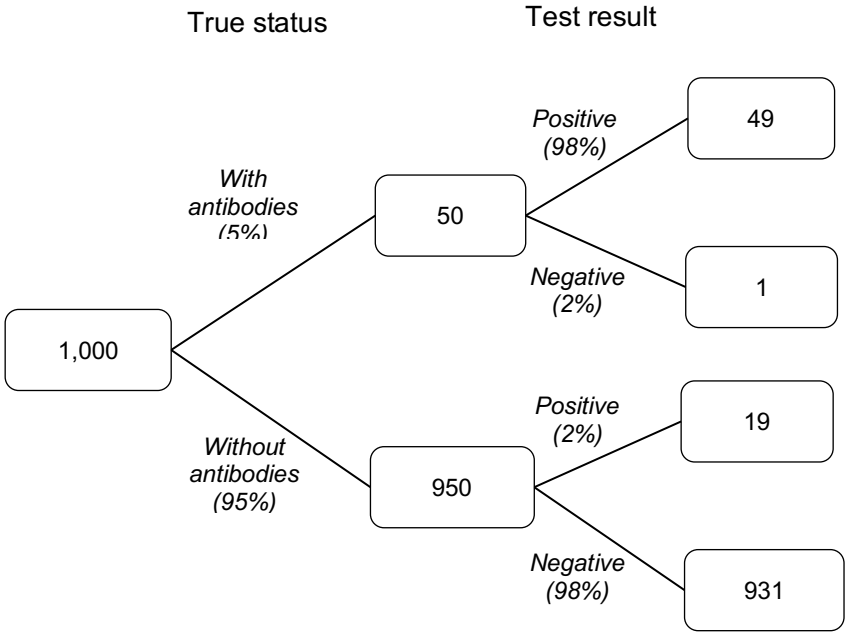
1. The composition of the negative control sample is clearly important, for example whether it is seeded with other coronaviruses.
2. If the aim is simply to find the proportion in a population with antibodies to coronavirus-2, far less stringent specifications may be adequate, as the prevalence can be estimated from the observed immune proportion, adjusted for the assumed sensitivity and specificity[2].

[2] If a test with sensitivity Se and specificity Sp gives a percentage p of positive results, an unbiased estimate of the true prevalence is $p^* = (p - 100 + Sp)/(Se + Sp - 100)$.

2 Choice of specificity threshold – should it be higher?

98% specificity means that 2% of those without antibodies will be falsely told they are immune (assuming lasting immunity). The impact of this depends on the prevalence of immunity. This is demonstrated in the diagram below for 1,000 people, assuming 98% sensitivity, 98% specificity, and an immunity prevalence of 5%.

Testing 1000 people, assuming 98% sensitivity, 98% specificity, and 5% prevalence of immunity



Out of 49 + 19 = 68 positive tests, 19 (28%) are false positives. Hence, of people being told they have antibodies, 28% are being falsely reassured and are in fact still susceptible.

The Table below shows how the magnitude of false claims, for 5%, 10% and 20% prevalence assumptions, and different values for sensitivity and specificity. The bold **28%** represents the calculation in the diagram above.

	Of those with a **negative** test result, the proportion who are actually immune			Of those with a **positive** test result, the proportion who are not immune		
Test specification	*Assuming **5%** of population have immunity*	*Assuming **10%** of population have immunity*	*Assuming **20%** of population have immunity*	*Assuming **5%** of population have immunity*	*Assuming **10%** of population have immunity*	*Assuming **20%** of population have immunity*
1. sens 98%, spec 98%	0.1%	0.2%	0.5%	**28%**	16%	8%
2. sens 90%, spec 98%	0.5%	1%	2.5%	30%	17%	8%
3. sens 98%, spec 99%	0.1%	0.2%	0.5%	16%	8%	4%
4. sens 90%, spec 99%	0.5%	1%	2.5%	17%	9%	4%

- Scenario 1 (sens 98%, spec 98%) represents the current assumptions, and shows a substantial proportion of claims of immunity will be false.
- Scenario 2 (sens 90%, spec 98%) shows a less sensitive test, which has little influence on the false claims of immunity.
- Scenario 3 (sens 98%, spec 99%) shows a more specific test, which roughly halves the proportion of claims of immunity that turn out to be false.
- Scenario 4 (sens 98%, spec 99%) shows a more specific but less sensitive test, which roughly halves the proportion of claims of immunity that turn out to be false, and has little influence on the proportion of people who are told they are not immune, but who actually are.

This suggests that if the aim is to provide reliable guidance for higher-risk people to return to work or to relax isolation, then 98% specificity may not be sufficient, and the sensitivity requirements might be relaxed.