On the independent evaluation of Abbott and Roche SARS-CoV-2 antibody tests

Sheila M. Bird, formerly Programme Leader at MRC Biostatistics Unit, CAMBRIDGE CB2 OSR, Honorary Professor, Edinburgh University College of Medicine and Veterinary Medicine & member of RSS Covid-19 Taskforce.

22 May 2020

Public health and statistical science require robust experimental design and reporting standards.

Independent evaluations have been conducted under the auspices of Public Health England of two serological tests, developed respectively by Abbott Laboratories and by Roche, for detection of Immunoglobulin G (IgG) antibodies against SARS-CoV-2[1] and for use in population surveillance. The test developers had, of course, conducted their own evaluations. For the criteria set by UK's Medicines and Healthcare products Regulatory Authority for accreditation as point-of-care test, please see https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/883897/Target_Product_Profile_antibody_tests_to_help_determine_if_people_have_immunity_to_SARS-CoV-2_Version_2.pdf.

Specificity [Sp, true negative rate or percentage of those truly negative who actually test negative] and sensitivity [Sn, true positive rate or percentage of those truly positive who actually test positive] of at least 98% each are considered necessary to determine an individual's antibody status; but lower values can be acceptable for population surveillance because appropriate adjustment can be made[2 3].

Background considerations: Demographic factors (age-group; being male) are strong prognostic factors for severe versus mild manifestation of COVID-19 disease. Hence, were the tests to be deployed in population surveillance for IgG antibodies against SARS-CoV-2 (*have I had it?*), we need to know **in advance** whether the serological tests' performances in terms of Sp and Sn vary by age-group and/or gender. If so, adjustments to population surveillance estimates would need to be made for age and/or gender.

In addition, robust evaluation of IgG antibody tests needs to source sufficient sera from people who tested PCR-positive in swab-test (*have I got it?)* but experienced only mild symptoms as their prevalence (or persistence) of IgG antibodies could be different from patients whose symptoms led to their hospitalization.

Serological testing for IgG antibodies to SARS-CoV-2 may be confounded by the presence of other viruses and so a key aspect of any independent evaluation is to amass sera from patients with a range of confounding conditions (but not coronavirus-2) to check if the COVID-19 IgG antibody-test signals positive.

Finally, persistence of IgG antibodies - measured alternatively from the date of symptom onset (subject to recall bias) or from swab-sample-date (subject to access to testing) – matters and so any rigorous independent evaluation needs to access repeat sera (pairs, triples, quartets) from recovered COVID-19-swab-positive patients and, for masking purposes, from patients who are not known to have encountered SARS-CoV-2.

Acquisition and banking of sera: The above background considerations mean that sera must be acquired as shown in **Table 1,** which includes dating alternatively from symptom-onset-date and

swab-positive-date. Ideally, each patient will have contributed sufficient blood that their research-donation or residual blood-samples can be used in at least 5 evaluations. If each cell in **Table 1** were represented by at least 100 sera, then 1400 negative sera (700 from males; 700 from females; 100 per age-group per gender) would have to be acquired together with 800 confounder sera (400 from males; 400 from females). The goal in **Table 1** would be to acquire sera from six times as many swab-positive patients (8400 in all; 1680 sera from patients who were hospitalized with COVID-19 disease).

A different acquisition scheme is needed to test robustly for persistence of IgG antibodies within an individual.

Weaknesses in independent evaluation under the auspices of Public Health England (PHE): Ten reservations are listed.

1. Equally-powered evaluation of the two tests was not provided: PHE's evaluation of the Roche test accommodated 85 confounder samples versus 364 for the Abbott test; each was evaluated on fewer than 100 convalescent sera.
2. Uncertain differentiation between repeat samples from the same patient and single samples from distinct patients. Throughout, there is a lack of transparency about the use of repeat sera per patient. Initial evaluation should be straightforwardly on the basis of one-sample per patient.
3. Hence, readers of the PHE reports cannot discern how many positive patients (versus sera) contributed to the evaluation of sensitivity (true positive rate). Neither evaluation included more than 100 convalescent patients (both genders, all ages, hospitalized or mildly symptomatic).
4. Age-group and gender of the patients whose sera were being analysed was not heeded. Demography matters for whether patients develop severe COVID-19 disease. Also, since immune responses differ intrinsically with age, the proportion of patients who develop IgG antibodies may differ by age; as may their persistence.
5. Level playing-field between the Roche and Abbott evaluations was not apparent in terms of whether the tested sera came from patients who had been hospitalized for COVID-19 disease or had been only mildly symptomatic. Both matter, especially if IgG antibody tests are to be used for population surveillance, and test-performance may be different by symptomatology.
6. Confounder samples, amongst which no positive was found in PHE's evaluation of the Roche test, did double-duty in the evaluation of the Roche test by being counted together with the "negative" samples, thereby increasing precision in an unprincipled manner.
7. Even playing field for both tests was lacking in PHE's evaluation of confounders. Confounder samples came from a range of patient-conditions, totalling 85 sera for the Roche evaluation (35/85 were from Lyme disease patients) versus 364 for the Abbott evaluation (11/364 were seasonal coronavirus positives).
8. Lack of clarity about analysis plan in respect of whether test-evaluation was primarily in respect of a) time-since-swab-positive-sample-date; or b) time-since-symptom-onset-date. Both matter in public health terms. Test-developers' focus may be preferential.
9. Lack of clarity about the analysis plan for exclusion of outliers when fitting half-normal distribution to the $\log_{10}$(test read-outs).
10. Lack of clarity about PHE's evaluation having not been designed to test the persistence of IgG antibodies within-person over time.

The above list is not exhaustive[4].

Public Health England's evaluation does not meet the standards that would be expected for confirmation that these tests meet the criteria set by UK's Medicines and Healthcare products Regulatory Authority for accreditation as point-of-care tests. When evaluating tests suitable for population surveillance, similar precision in estimating Sn and Sp is appropriate. Moreover, tests' suitability for age- and gender-specific population surveillance has not been addressed.

1. COVID-19: laboratory evaluations of serological assays, see https://www.gov.uk/government/publications/covid-19-laboratory-evaluations-of-serological-assays
2. Diggle, P.J. (2011). Estimating prevalence using an imperfect test. *Epidemiology Research International,* Article ID 608719. doi:10.1155/2011/608719.
3. Royal Statistical Society COVID-19 Taskforce Statement on COVID-19 Antibody Testing, issued on 14 April 2020. https://rss.org.uk/RSS/media/Policy-and-campaigns/Policy/Statement-Antibody-testing-14-04-2020.pdf.
4. Deeks J. Independent Evaluation of Immunoglobulin G antibodies against SARS-CoV-2: Statements of "100% accuracy" are not supported by the data now available. It is important that the public are properly informed of the accuracy of these tests. @deeksj.

Table 1. Data per single blood sample from an individual who tested PCR-positive for COVID-19 include: *birth-year***,** *gender,* symptom-onset date, swab-date, swab-PCR-result, whether hospitalized with COVID-19 disease, *blood-sample-date*. *Italicised data* apply also per single blood sample from individual who could not have been infected by SARS-CoV-2, for example because blood-sample-date was in the first semester of 2019.

Demography: Gender & age-group	**COVID-negative**	**Swab-PCR-positives: 20% hospitalized**			**Swab-PCR-positives: 20% hospitalized**			**Confounder samples**
		14-20 days since swab-date	21-69 days since swab-date	70-139 days since swab-date	14-20 days since onset-date	21-69 days since onset-date	70-139 days since onset-date	
Males								
5-12 years								
13-17 years								
18-44 years								
45-64 years								
65-74 years								
75-84 years								
85+ years								
Females								
5-12 years								
13-17 years								
18-44 years								
45-64 years								
65-74 years								
75-84 years								
85+ years								