Appendix

The measured sensitivity is the probability that the interpreted test is positive given that the interpreted standard is positive:

\[ s' = (a'|a_0') = \frac{(a' & a_0')}{(a_0')} = \frac{\sum_X (X) \sum_{x,x_0}(x & x_0 | X)(a' & a_0'|x & x_0)}{\sum_X (X) \sum_{x_0}(x_0 | X)(a_0'|x_0)}. \] (S.1)

We use the assumptions that (A1) the outputs are independent when conditionalized on the patient state, \((x & x_0 | X) = (x | X)(x_0 | X)\), (A2) that the interpreters are not influenced by each other \((x' & x'_0 | x & x_0) = (x' | x & x_0)(x'_0 | x & x_0)\) and (A3) they are not influenced by the other device, \((x'|x) = (x'|x & x_0)\) and \((x'|x_0) = (x'_0 | x & x_0)\). Then the measured sensitivity can be written:

\[ s' = \frac{\sum_X (X) \sum_{x,x_0}(x|X)(a'|x)(x_0)(a_0'|x_0)}{\sum_X (X) \sum_{x_0}(x_0 | X)(a_0'|x_0)}. \] (S.2)

For the measured specificity a similar equation holds where events \(a\) are replaced by events \(n\). By substituting the definitions of the sensitivity and specificity of the test and standard and of the interpreters given in the text, Eq. 1 in the main article is obtained. Define the vector \(s' = [s',p']^T\). Eq. 1 shows that the relationship of \(s'\) and the test’s true properties, \(s\), has the form

\[ s' = Ms + b \] (S.3)

where each element of the 2x2 matrix \(M\) and vector \(b\) is an explicit function of the set of parameters \((A), s_0, p_0, r_0, q_0, r, q\). The true accuracy is then obtained as \(s = M^{-1}(s' - b)\) where a standard formula expresses \(M^{-1}\) in terms of the elements of \(M\), for any non-singular \(M\).

Eq. 1 involves the following definitions:

\[ \begin{align*}
\tilde{s}_0 &= s_0r_0 + (1 - s_0)(1 - q_0), \quad \tilde{p}_0 = (1 - p_0)r_0 + p_0(1 - q_0), \\
\bar{s}_0 &= s_0(1 - r_0) + (1 - s_0)q_0, \quad \bar{p}_0 = (1 - p_0)(1 - r_0) + p_0q_0.
\end{align*} \] (S.4)

It is straightforward to show that Eq. 1 reduces to the special case in Eq. 4 when \(s_0 = p_0 = r_0 = q_0 = 1\).

By considering a series of repeated pairs of interpretations, \(x'_{1}\) and \(x'_{2}\), of the test, the observed proportion of agreements can be estimated as

\[ P = (a'_1 & a'_2) + (n'_1 & n'_2) = \sum_X (X) \sum_x (x|X) \sum_{x'} (x'_1|x)(x'_2|x). \] (S.5)

where the second equality followed from the assumption (A2). By inserting the definitions of device and interpreter accuracy, this leads to Eq. 2. When interpreter sensitivity and specificity are the same, \(r\) can be replaced by a function of \(P\) and therefore by a function of \(\kappa\) and \((A)\). This leads by algebra to Eq. 3.

Given \((A), s_0, p_0, \kappa_0, \kappa\), the true test accuracy, \(s, p\), are determined as follows. 1) First we use the version of Eq. 2 for the standard interpreter that involves \(s_0, p_0, r_0, q_0\) and the definition of standard kappa \(\kappa_0 = (P_0 - P)/(1 - P)\). By choosing \(r_0\), we solve for \(q_0\) from these two equations. 2) This results in 4 unknowns \(s, p, r, q\), and 3 equations Eq. 1 and 2. Hence by choosing \(r\), we can solve for \(s, p, q\). For example we can solve for \(q\) from Eq. 2 in terms of \(s, p, r\), and substitute this into Eq. 1, and solve the resulting nonlinear equation by an iterative technique. Our actual implementation however is to choose pairs of values on a fine grid of values of \(r\) and \(q\), for each pair calculate \(s\) and \(p\) from Eq. 1, then determine kappa using 2, and retain the pair only if the resulting kappa is consistent with its given value to within a small tolerance. This solution yields a range of values for \(s\) and \(p\), since \(\kappa_0, \kappa\) do not uniquely determine the accuracy of the interpreters.