

Imperfect tests

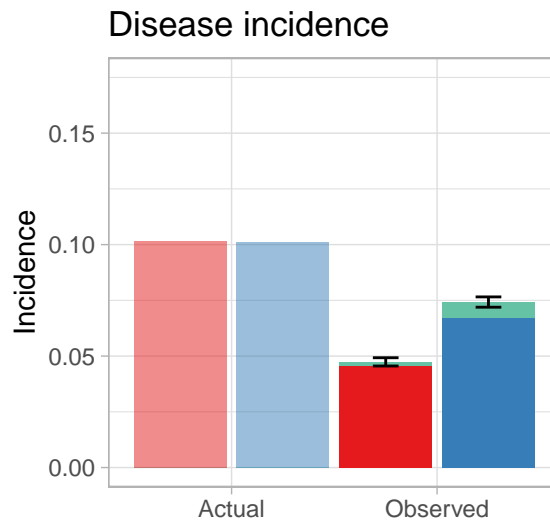
Rachael Caelie (Rocky) Aikens

11/4/2021

Here we consider a situation in which the diagnostic evaluation itself has some error rate. Suppose the test has a sensitivity and specificity of 80%, and the performance of the test itself does not depend on X or R .

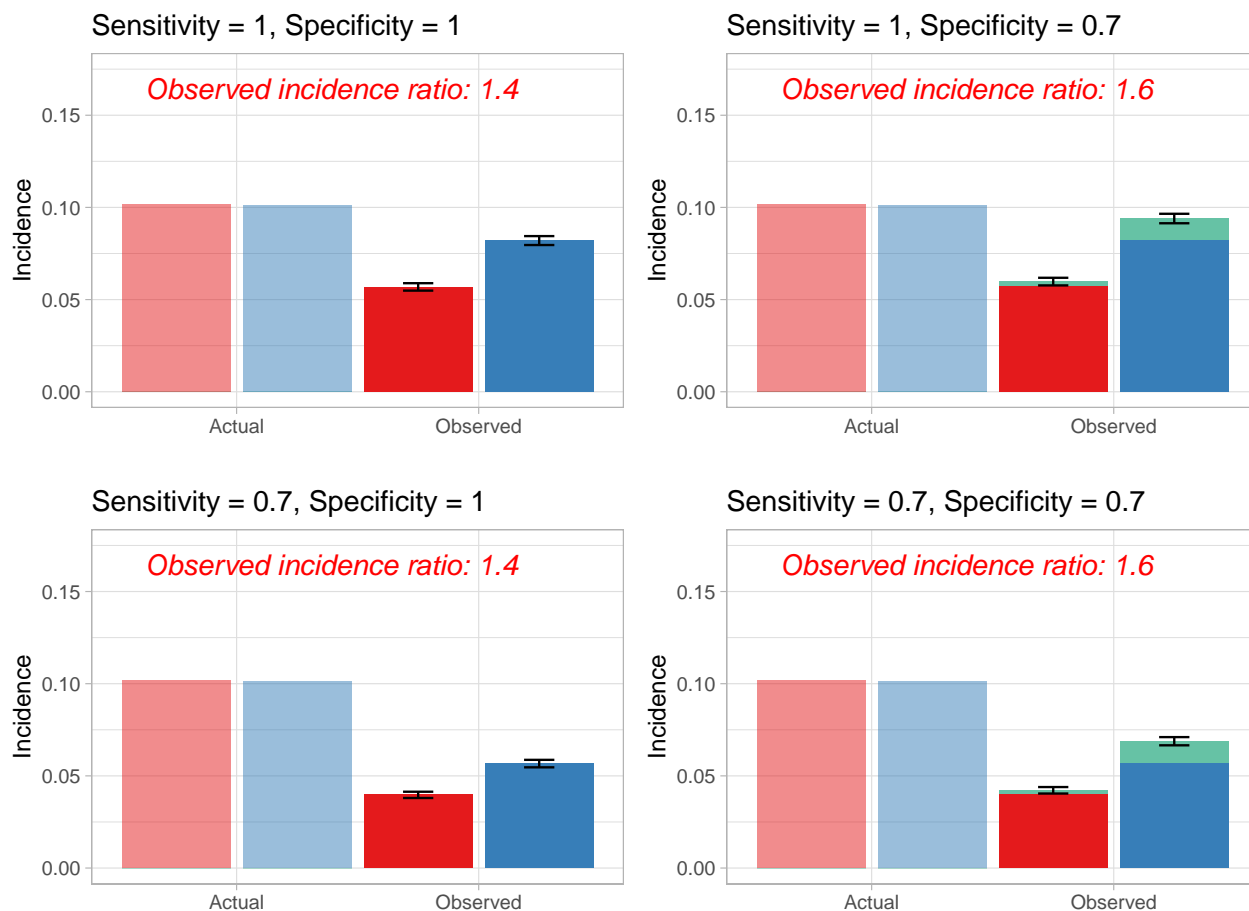
Case 2A with imperfect tests

For simplicity, I'll start with case 2A. The figure below shows the actual and observed incidences. The turquoise colored portion of the "Observed" bars represents the proportion of observed disease cases which are actually false positives. Because the $X = 1$ population is tested more, they have the greater burden of false positives *and* false negatives. However, because the $X = 0$ population is only tested for more representative cases, a greater proportion of their tests are positive overall.



x	Tests	True Positives	False Positives	False Negatives	Positive test rate
FALSE	3251	2308	86	550	0.7363888
TRUE	6016	3319	356	754	0.6108710

Here are examples of a few runs of the simulation with different sensitivity and specificity levels. In particular, it seems that a low sensitivity test impacts both groups equally in this set-up, but a low specificity test results in a proportionally larger share of false positives for the more-tested group. Thus, the lower specificity testing scenarios also lead to a greater difference in the observed incidence rates between groups.



Here are the results over 1,000 simulations of case 2A in each of the four scenarios above.

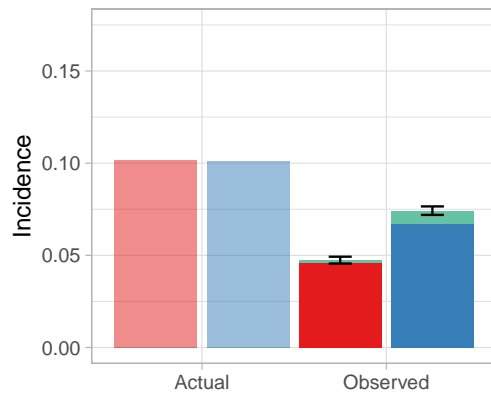
% cases diagnosed (X = 0)	% cases diagnosed (X = 1)	est. incidence (X = 0)	est. incidence (X = 1)	est. relative incidence	scenario
0.558	0.812	0.056	0.081	1.452	Perfect test
0.583	0.926	0.058	0.093	1.586	Poor specificity
0.391	0.568	0.039	0.057	1.456	Poor sensitivity
0.416	0.682	0.042	0.068	1.641	Poor sensitivity and specificity

SD % cases diagnosed (X = 0)	SD % cases diagnosed (X = 1)	SD est. incidence (X = 0)	est. incidence (X = 1)	SD est. relative incidence	scenario
0.007	0.006	0.001	0.001	0.035	Perfect test
0.008	0.008	0.001	0.001	0.036	Poor specificity
0.007	0.007	0.001	0.001	0.042	Poor sensitivity
0.007	0.009	0.001	0.001	0.044	Poor sensitivity and specificity

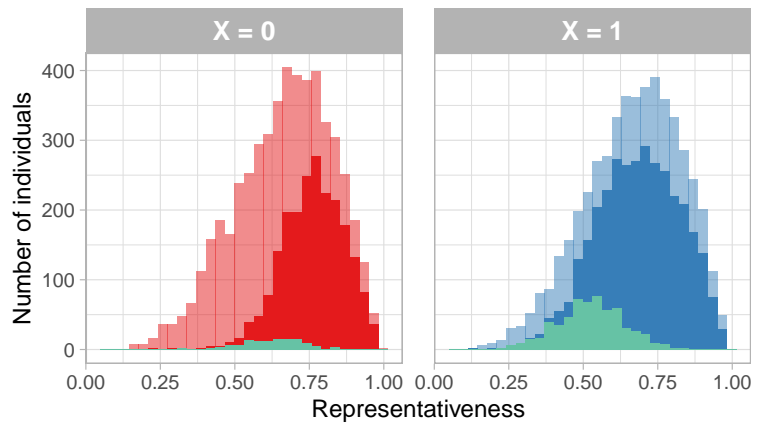
The distribution of representativeness

What does an imperfect test mean for the observed distribution of representativeness? False positives (turquoise) contribute to the observed distribution of representativeness even though they are not true disease cases. These false-positive individuals tend to fall lower on the distribution of representativeness than the true-positive individuals. In other words, they “fatten” the left tail of the observed distribution of representativeness, but this effect is more pronounced for the more-tested group ($X = 1$).

A Disease incidence

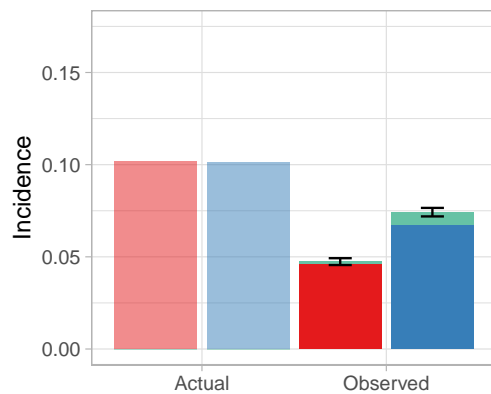


B Distribution of representativeness

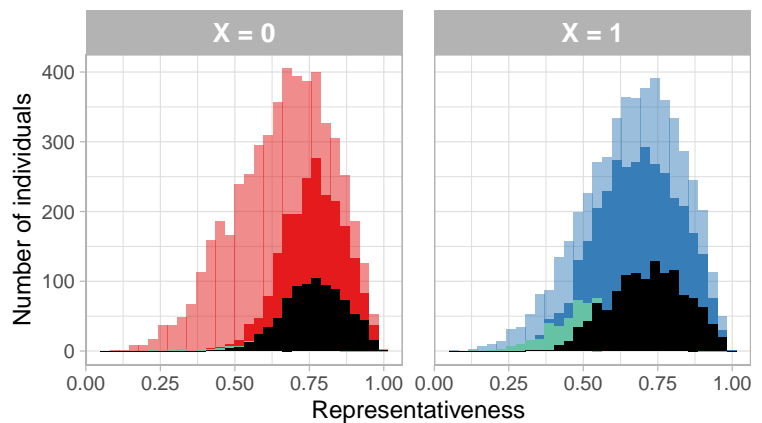


What about false negatives? The plot below adds false negatives in black. Since the sensitivity of the test does not depend on representativeness, the distribution of false negatives appears to mirror the distribution of true positives.

A Disease incidence



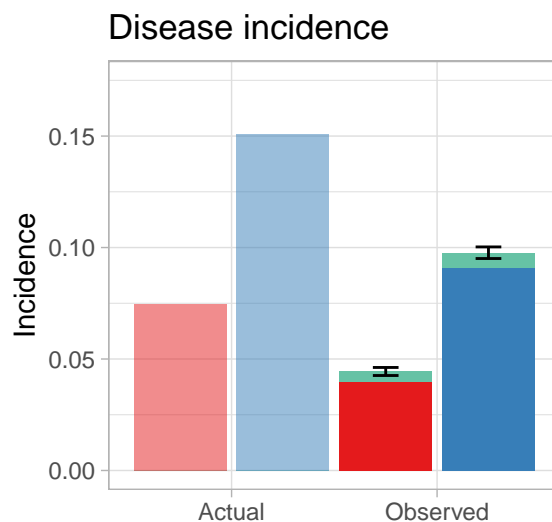
B Distribution of representativeness



Case 2B with imperfect tests

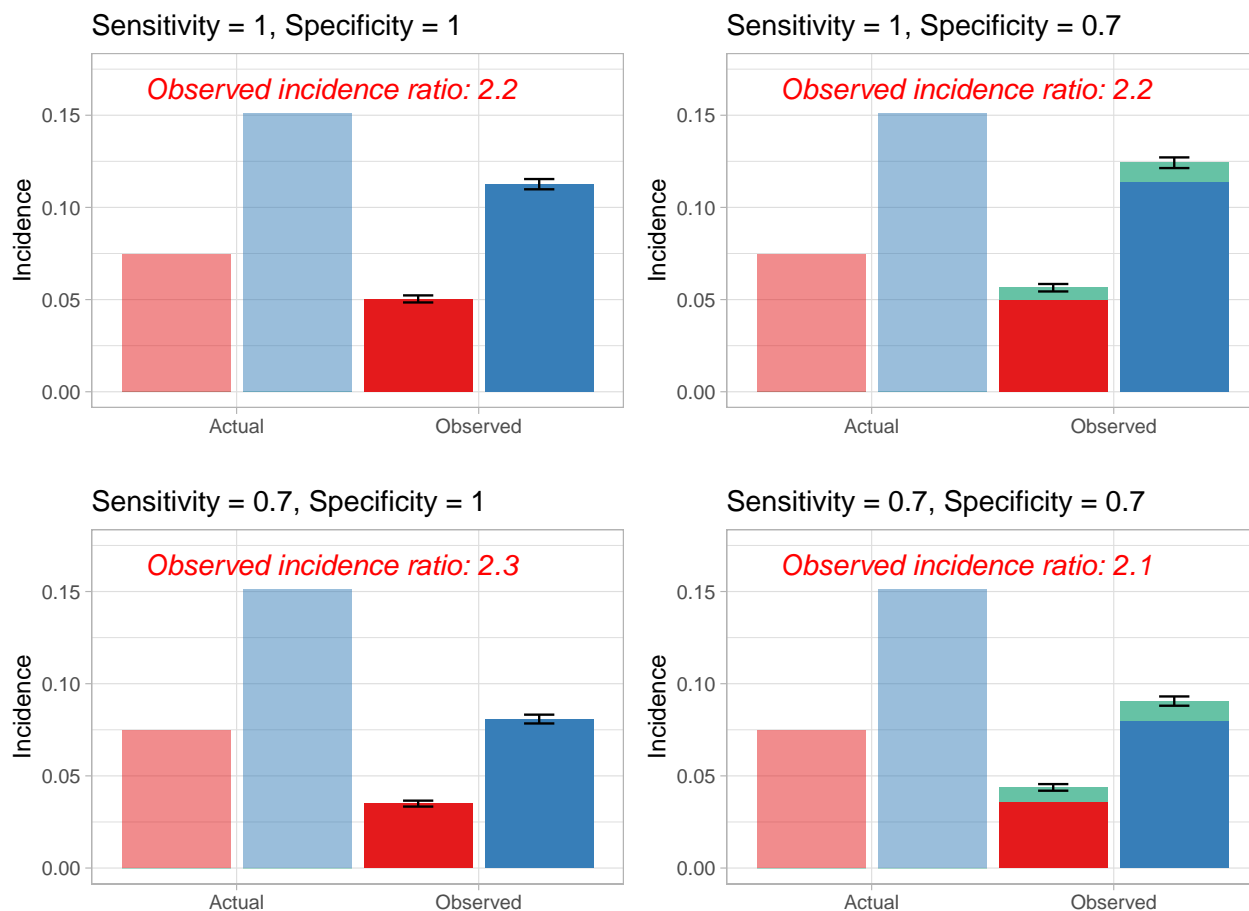
For completeness, here's the same set of simulations and visualizations for case 2B. These can be harder to interpret because the base rates of disease are not at parity between groups.

The figure below shows the actual and observed incidences in this scenario. Because the $X = 1$ population is tested more, they have the greater raw numbers of false positives *and* false negatives. However, it's not clear to me if the relative abundances of false positives and false negatives are somehow “proportional” to the underlying rates of disease. In this case, the positive tests rates are closer to parity between groups (I'm not sure whether the difference is systematic or due to chance). Maybe this is because the $X = 0$ group has a higher underlying disease risk than the $X = 1$ group.



x	Tests	True Positives	False Positives	False Negatives	Positive test rate
FALSE	3686	1986	236	482	0.6028215
TRUE	7424	4530	350	1110	0.6573276

Here are examples of a few runs of the simulation with different sensitivity and specificity levels. It's harder to parse the relative contributions of sensitivity and specificity in this context. Since the apparent differences are small, it's less obvious whether any differences are due to something systematic, or just mostly noise.



Finally, 1000 simulations of the same results.

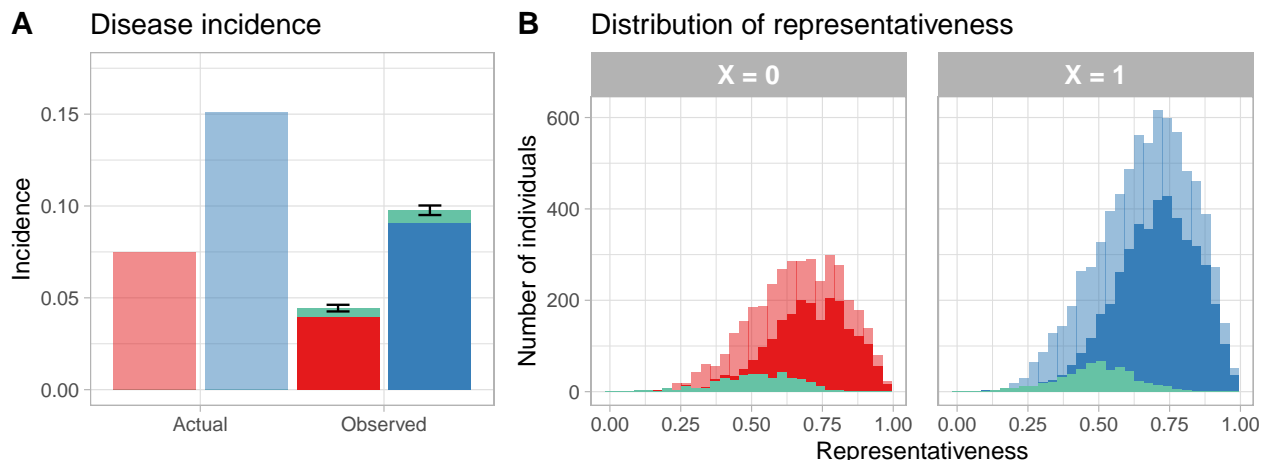
% cases diagnosed (X = 0)	% cases diagnosed (X = 1)	est. incidence (X = 0)	est. incidence (X = 1)	est. relative incidence	scenario
0.673	0.757	0.050	0.114	2.252	Perfect test
0.770	0.829	0.058	0.124	2.155	Poor specificity
0.471	0.530	0.035	0.080	2.256	Poor sensitivity
0.568	0.602	0.043	0.090	2.119	Poor sensitivity and specificity

SD % cases diagnosed (X = 0)	SD % cases diagnosed (X = 1)	SD est. incidence (X = 0)	SD est. incidence (X = 1)	SD est. relative incidence	scenario
0.008	0.006	0.001	0.001	0.054	Perfect test
0.010	0.007	0.001	0.002	0.047	Poor specificity
0.009	0.006	0.001	0.001	0.065	Poor sensitivity
0.010	0.007	0.001	0.001	0.055	Poor sensitivity and specificity

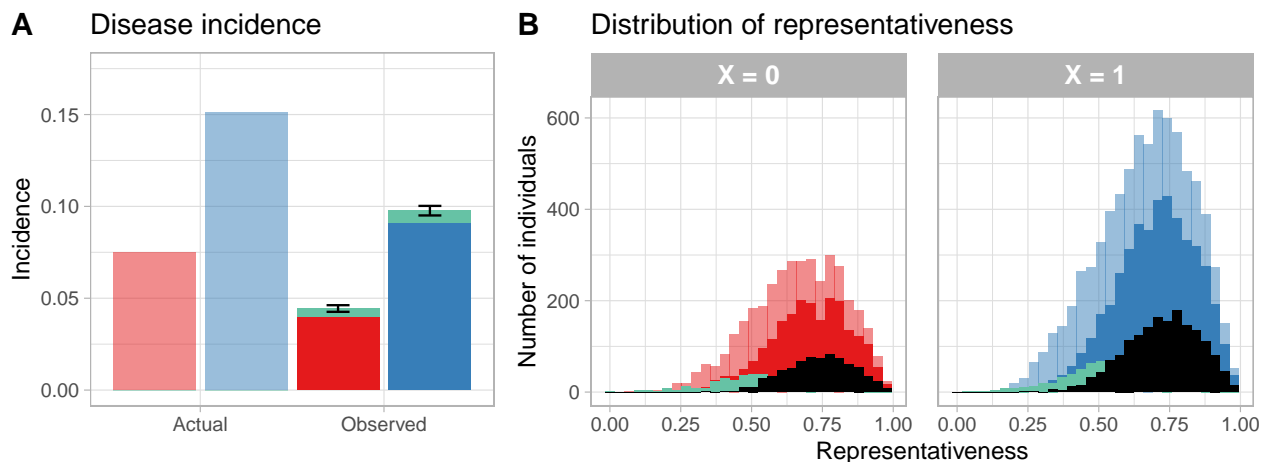
This is surprising! It seems the worse tests actually *decrease* the error in estimation for Case 2B. But why? Does this stem from the fact that the testing rates are roughly proportional to true disease risk?

The distribution of representativeness

Unsurprisingly, false-positive individuals tend to fall lower on the distribution of representativeness than the true-positive individuals. In other words, they “fatten” the left tail of the observed distribution of representativeness. It’s hard to tell if this effect is more pronounced for either group.



As before, the distribution of false negatives appears to mirror the distribution of true positives.



Case 2C: True risk factor, but overestimated impact

I made up this case in the hopes of better understanding the results from Case 2b. Here, X is a risk factor (as in Case 2B), but the association between X and disease risk is overemphasized in the diagnostic process (as in Case 2B). Specifically, here we’ll use the disease model of Case 2B but the biased clinical diagnostic process of Case 2A.

% cases diagnosed ($X = 0$)	% cases diagnosed ($X = 1$)	est. incidence ($X = 0$)	est. incidence ($X = 1$)	est. relative incidence	scenario
0.558	0.812	0.042	0.122	2.908	Perfect test
0.593	0.883	0.044	0.132	2.981	Poor specificity
0.391	0.568	0.029	0.085	2.910	Poor sensitivity
0.425	0.640	0.032	0.096	3.013	Poor sensitivity and specificity

SD % cases diagnosed (X = 0)	SD % cases diagnosed (X = 1)	SD est. incidence (X = 0)	SD est. incidence (X = 1)	SD est. relative incidence	scenario
0.008	0.004	0.001	0.001	0.075	Perfect test
0.008	0.005	0.001	0.002	0.069	Poor specificity
0.008	0.006	0.001	0.001	0.087	Poor sensitivity
0.008	0.006	0.001	0.001	0.084	Poor sensitivity and specificity