Fictional scenario

In a previous article, the basics about diagnostic studies were discussed through a fictional scenario on the prevalence of gingival recession (Papageorgiou, 2020). In that article, the terms sensitivity of a test (number of true cases picked from a test divided by the true cases), specificity (number of true negative test picked by a test divided by true negatives), positive predictive value (PPV; number of true cases picked from the test divided by all test positives) and negative predictive value (number of true negatives picked from the test divided by all test negatives) were introduced. Finally, I also discussed how our decision about the most appropriate diagnostic test might be affected by the following: (1) the severity of a potential disease which we test for; (2) the impact of somebody learning she/he has this particular disease on everyday life and prognosis; and (3) the efficacy/tolerability of existing treatments for a disease.

In the present article, we stay on the subject of diagnostic accuracy, but the focus will be shifted to additional factors that need to be taken into account when evaluating the performance of a diagnostic test. Sensitivity and specificity are terms widely used in the scientific literature, but represent concepts not intuitive to most patients, who like to know ‘If I get a positive (or a negative) test result, what does this mean to me’ or ‘If X people get a positive (or a negative) test result, what percentage of these people will have the disease?’ This is given by the PPV, which we will try to calculate and interpret for two different variations for clinical scenarios.

Suppose we want to check the diagnostic performance of a hypothetical novel clinical test to see if a tooth is ankylosed by hitting the tooth with the back side of a specially designed dental probe at an angle of 48° to the tooth’s long axis. Assuming we know that compared to the gold standard of histological analysis, this clinical percussion test normally has a sensitivity of 93% and a specificity of 55%.

In the first scenario variant (scenario 1), we try to use this ankylosis test in every single patient that enters our practice and has a deciduous tooth in the mouth during the patient’s first appointment. In the second scenario variant (scenario 2), we use this percussion test only on lower second deciduous molars that are still in place two years after their average expected exfoliation time and are in infraocclusion relative to the adjacent teeth. How would we expect the diagnostic test to perform in these scenarios?

Which of the following statements are true, if any?

(a) The test’s performance in terms of positive predictive value is the same in all clinical scenarios.

(b) The test’s performance and its interpretation are dependent on the prior expected event rate (for the disease) of the sample that it is used on.

(c) The test’s performance and its interpretation are dependent on the disease symptoms of the patient sample it is used upon.

(d) Assuming a patient gets a positive test result, the probability that a patient does indeed have the disease is the same in scenario 1 and scenario 2.

In order to see which of the following statements are correct, we need to work out the expected performance of the test in each of the scenario variants. Let’s start from the end with scenario 2, where the percussion ankylosis test is administered on lower second deciduous molars that are still in place two years after their average expected exfoliation time and are in infraocclusion relative to the adjacent teeth. An experienced clinician might see the above-mentioned clinical symptoms as hints of a possible ankylosis (Ponduri et al., 2009) and in this patient sample one might expect ankylosis to be found very often in 70% of the cases—even before doing any test!.

Then our novel percussion test would perform as shown in Table 1.

We see that this novel percussion test is indeed very good diagnostically, as one might interpret a positive test result as always telling you that you have a fairly high (83%) chance of ankylosis in every scenario, which is not correct. This PPV of 83% rather tells you, in this specific testing scenario, what proportion of the percussion test’s positives were actually true disease cases.
Let’s move now to scenario 1, where the percussion test is used indiscriminately to all patients entering our practice, be it young or old, and in any phase of their dentition without any specific clinical symptoms. In such a case, an experienced clinician might expect a very low rate of ankylosed teeth for the general population (Mubeen and Seehra, 2018)—let’s say of 1%. Then our novel percussion test would perform as shown in Table 2.

Here we see that the same test gives a very different PPV, of only 2%. This means that if we get a positive test result, there is only a 2% chance that there is in truth an ankylosis. This comes completely at odds with scenario 2, where the test was very useful in picking up and ruling out ankylosis cases. We can therefore see that, even though the PPV is an intuitive interpretation of a test’s diagnostic performance, it is heavily dependent on the true disease event rate of the population we are testing. Statement (a) is wrong and statement (b) is correct.

Differences in the expected event rate of a tested population might be due to pure chance, might be due to the way we collected our sample from the general population, or might be due to the specific characteristics of the patients. In this case, we explicitly defined scenario 1 to be based on a sampling of consecutively admitted patients, which given that a long time period was included, might be expected to result in a relatively random sample that might be generalisable to the average patient that visits a practice. On the other hand, scenario 2 was based on specific inclusion criteria that actually pertained to patient characteristics that are not random, but rather might be directly attributed to the disease in question. Persistence of a deciduous lower second molar for a prolonged period after its expected exfoliation period and infraocclusion might be clinical symptoms of a tooth ankylosis. A Bayesian thinking approach might aid in the interpretation of the diagnostic test here. A random patient in scenario 2, even before having the percussion test, has a much higher chance of having an ankylosed tooth (since there are more clinical symptoms associated with tooth ankylosis) than a random patient in scenario 1 (since no clinical symptoms are required). And we have seen that the prior knowledge of the patient’s characteristics before the test was performed had a direct influence on the test’s performance (through the expected ankylosis prevalence). Therefore, statement (c) is correct.

Moreover, using the Bayes’ theorem, we can calculate the posterior probability of a patient having indeed the disease (i.e. probability after the test is performed) using the prior probability of a patient having the disease (i.e. probability before the test is performed) and the test result (in a simplified form). In scenario 1, a patient before having the test has an expected prior probability of ~1% having ankylosis, and having a positive percussion test result gives a posterior ankylosis probability of ~0.02%. On the other hand, a patient of scenario 2 has a prior chance of ankylosis of 70% even before doing the test and a positive test result lifts this probability posteriorly to 92%, which is really helpful. This is a very simplified way to indicate that the way a positive test result is interpreted does not only rely on the diagnostic performance of a test, but also on the characteristics of the patient sample. Therefore, statement (d) is false.

This is one of the reasons why a diagnostic test that has been benchmarked and approved in terms of sensitivity/specificity among symptomatic patients would not necessarily be appropriate to be used for mass testing of asymptomatic patients, as was seen for rapid lateral flow tests for SARS-CoV-2 infection used in the United Kingdom (Deeks and Raffle, 2020).

### Table 1. Diagnostic performance of the novel percussion test (sensitivity of 93%, specificity of 55%) for scenario 2, with an expected event rate of 70%; here given for the first 1000 patients.

| In truth: ankylosis | In truth: no ankylosis | Total |
|--------------------|------------------------|-------|
| Percussion says ankylosis | 651 | 135 | 786 |
| Percussion says no ankylosis | 49 | 165 | 214 |
| Total | 700 | 300 | 1000 |

In this scenario, the positive predictive value is $\frac{651}{651+135}=\frac{651}{786} \approx 83\%$.

### Table 2. Diagnostic performance of the novel percussion test (sensitivity of 93%, specificity of 55%) for scenario 1, with an expected event rate of 1%; here given for the first 1000 patients.

| In truth: ankylosis | In truth: no ankylosis | Total |
|--------------------|------------------------|-------|
| Percussion says ankylosis | 9 | 445 | 454 |
| Percussion says no ankylosis | 1 | 545 | 546 |
| Total | 10 | 990 | 1000 |

In this scenario, the positive predictive value is $\frac{9}{9+445}=\frac{9}{454} \approx 2\%$. 
Declaration of Conflicting Interests
The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author received no financial support for the research, authorship, and/or publication of this article.

ORCID iD
Spyridon N Papageorgiou https://orcid.org/0000-0003-1968-3326

References
Deeks JJ and Raffle AE (2020) Lateral flow tests cannot rule out SARS-CoV-2 infection. British Medical Journal 371: m4787.
Mubeen S and Seehra J (2018) Failure of eruption of first permanent molar teeth: a diagnostic challenge. Journal of Orthodontics 45: 129–134.
Papageorgiou SN (2020) Prudens quaestio in diagnostic studies. Journal of Orthodontics 47: 363–365.
Ponduri S, Birnie DJ and Sandy JR (2009) Infraocclusion of secondary deciduous molars—an unusual outcome. Journal of Orthodontics 36: 186–189.