Diagnostic and Interventional Imaging (2018) xxx, xxx—xxx

EDITORIAL / *Genitourinary imaging*

Pitfalls in interpreting positive and negative predictive values: Application to prostate multiparametric magnetic resonance imaging

''A certain elementary training in statistical method is becoming as necessary for everyone living in this world of today as reading and writing''.

H.G. Wells (1866—1946)

Multiparametric magnetic resonance imaging (mpMRI) has excellent sensitivity in detecting clinically-significant prostate cancer (csPCa) [1,2]. Recently, the multicentric PROMIS study that used template saturation biopsy as a reference test found a sensitivity ranging from 87% to 93%, depending on the definition used for csPCa [3]. For many clinicians, these excellent results are the ultimate proof that patients with negative mpMRI findings can safely avoid prostate biopsy. Indeed, it seems logical to think that a highly sensitive test necessarily has a high negative predictive value (NPV): if the test can detect most cancers, and if my results are negative, then surely I am safe, aren't I? Unfortunately, this is not always true.

NPV and positive predictive value (PPV) are useful indicators because they assess what doctors want to know: the probability that their patient has or does not have the suspected disease. However, interpreting NPV and PPV may be difficult because, unlike sensitivity and specificity, they depend on disease prevalence (Fig. 1). Failing to understand the implications of this dependence may lead to misinterpretation of published results and errors in individual patient management [4—6].

Studies have repeatedly shown that clinicians poorly evaluate the probability of disease based on sensitivity, specificity and prevalence estimates [7—11]. In a large survey performed in Switzerland, 1361 physicians from all clinical specialties were randomly separated into six groups and asked to estimate the probability of disease when a test with a sensitivity and a specificity of 99% was positive. Each group received a different prevalence value (1%, 2%, 10%, 25%, 95% and no information on prevalence for the last group). Possible answers for the probability of disease were < 60%, 60—79%, 80—94%, 95—99.9% and > 99.9%. In each group, the majority of respondents selected a post-test probability of 95—99.9% regardless of the prevalence of disease. Taking into account possible random answers, it was estimated that only 9% of respondents correctly knew how to assess the correct post-test probability that was 50%, 67%, 92%, 97% and 99.95% for prevalence values of 1%, 2%, 10%, 25% and 95% respectively. The proportion of correct answers did not vary according to sex, age, number of years after graduation, clinical specialty or practice setting [11].

There are several explanations for these disappointing results. Some doctors may confuse NPV with sensitivity and PPV with specificity and believe that if the sensitivity and specificity of the test are 99%, then its NPV and PPV must be 99%. Others may not comprehend the true meaning of prevalence. They may think that the 'population' from which a given prevalence is estimated refers to people in a given geographical area, or from a specified race. As a result, they assume that it is approximately the same for all patients in their clinic. In reality, the 'population' refers to a constellation of people with similar symptoms and/or signs [4]. In a given city, the prevalence of csPCa is higher in, say, 60-year-old patients with PSA levels of 20 ng/ml and suspicious digital rectal examination than in 60-year-old patients from the same city with PSA levels of 3 ng/ml and normal digital rectal examination. Thus, the prevalence must be regarded as the probability that a patient has the disease based on the information (personal and family

https://doi.org/10.1016/j.diii.2018.07.008

Please cite this article in press as: Rouvière O, et al. Pitfalls in interpreting positive and negative predictive values: Application to prostate multiparametric magnetic resonance imaging. Diagnostic and Interventional Imaging (2018), https://doi.org/10.1016/j.diii.2018.07.008

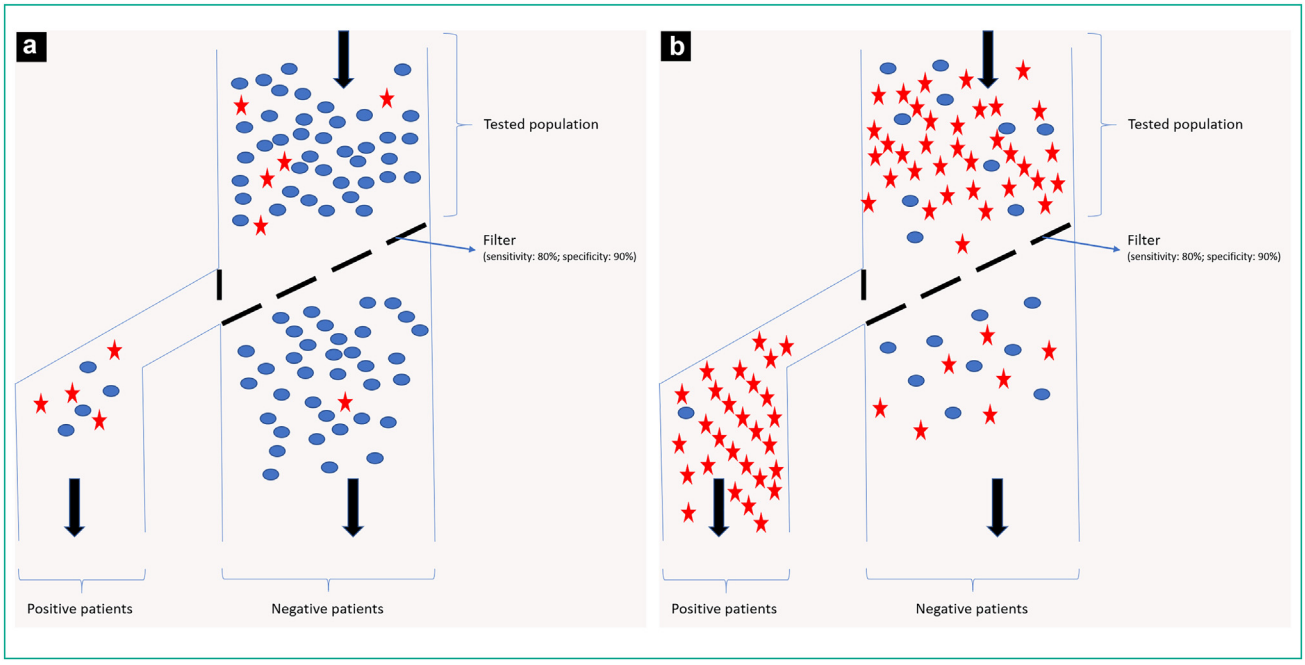

Figure 1. Illustration of the impact of prevalence on negative and positive predictive values. A diagnostic test can be illustrated as a filter located in a pipe that filters out patients with the disease (red stars) and lets patients without the disease (blue ovals) pass through. Let us suppose that the test filters out 80% of the patients with the disease (sensitivity: 80%) and 10% of the patients without the disease (specificity: 90%). In Fig. 1a, the disease prevalence in the tested population is 5/45 = 11.1%. Four of the five patients with the disease (80%) and four of the 40 patients without the disease (10%) are filtered out (positive patients). This gives a positive predictive value of 4/8 = 50%. Similarly, one patient with the disease and 36 patients without the disease pass through the filter (negative patients); the negative predictive value is thus 36/37 = 97.3%. In Fig. 1b, the disease prevalence in the tested population is 35/45 = 77.8%. Twenty-eight of the 35 patients with the disease (80%) and one of the 10 patients without the disease (10%) are filtered out (positive patients). This gives a positive predictive value of 28/29 = 96.6%. Similarly, 7 patients with the disease and 9 patients without the disease pass through the filter (negative patients); the negative predictive value is thus 9/16 = 56.3%. The filter has the same performances (sensitivity and specificity) in both figures. Yet, the proportion of patients with and without the disease in the group that was filtered out (positive patients) and in the group that passed through the filter (negative patients) are deeply influenced by the composition of the initial population (prevalence).

history, physical examination, laboratory results) available before the diagnostic test is performed, e.g. the pre-test probability of disease. The 'prevalence' of csPCa will therefore be different in two consecutive patients seen in consultation.

Even when clinicians do take disease prevalence into account, they usually fail to understand how deeply it can affect NPV and PPV. Again, this is probably explained by the unconscious belief that a test with an excellent sensitivity (or specificity) necessarily has an excellent NPV (or PPV). As shown in Fig. 2, NPV decreases as prevalence increases, but the relationship is not linear and is mostly influenced by sensitivity. Highly sensitive tests show high NPV over a large range of prevalence values. However, if the prevalence is too high, the NPV falls, even for these highly sensitive tests. Similarly, PPV increases as prevalence increases, and their relationship is mostly influenced by specificity. While highly specific tests show excellent PPV over a large range of prevalence values, PPV falls dramatically when the prevalence is too low.

Can mpMRI therefore be used to define who should have prostate biopsy? As discussed above, negative mpMRI findings must be regarded with care in high-risk patients. Similarly, because mpMRI only has a low-to-moderate specificity, a substantial proportion of low-risk patients with positive mpMRI findings will undergo unnecessary biopsy. To rationally assess the need for prostate biopsy, one must take into account the pre-mpMRI risk profile of the patient. Risk models integrating mpMRI findings and clinical parameters (e.g., history, age, digital rectal examination findings, PSA density) have yielded good results in predicting biopsy results in preliminary studies [12—14]. We believe this is the only way to reduce unnecessary biopsies while optimizing csPCa detection.

Please cite this article in press as: Rouvière O, et al. Pitfalls in interpreting positive and negative predictive values: Application to prostate multiparametric magnetic resonance imaging. Diagnostic and Interventional Imaging (2018), https://doi.org/10.1016/j.diii.2018.07.008