

Supplementary Material. Sources of Bias and Variations in Studies Evaluating Diagnostic Test Accuracy (DTA)

1. Population-related bias/variations

- 1) Spectrum effect (sometimes called spectrum bias): The phenomenon that the diagnostic test accuracy may vary in terms of demographic features, disease severity, disease prevalence, and clinical setting. If a study is carried out in a certain spectrum of patients that is not similar to the population in which the test will be used in clinical practice, the results of the study may have limited applicability.
- 2) Sample selection bias: The selection process determines the composition of the study sample. For example, if an investigator excludes patients whose MRI image is of poor quality, the study would report an overestimated diagnostic accuracy of the MRI.

2. Bias related to reference standards

- 1) Bias due to inappropriate reference standard: The use of inaccurate reference standards leads to misclassification of the true disease status and makes the index test results inaccurate. This leads to biased (usually underestimated) results of the test accuracy.
- 2) Differential verification bias: This occurs when part of the index test results is verified by a different reference standard. It is common in radiological studies. For example, if the index test result is positive for disease (e.g., probable liver metastasis seen on MRI), the reference standard of surgery will be performed. In contrast, if the index test is negative for disease (e.g., probable hemangioma rather than liver metastasis), the reference standard would be follow-up.
- 3) Partial verification bias: This occurs when only a selected sample of patients is verified by the reference standard. For example, bias arises if patients with a positive result on the index test undergo the reference standard test.
- 4) Disease progression bias: This occurs when there is a sufficient time delay between the application of the index test and of the reference standard, leading to a significant change in the target disease state.

3. Interpretation-related bias

- 1) Diagnostic review bias: This occurs when the results of the index test are known to reviewers who interpret the reference standard.
- 2) Test review bias: This occurs when the results of the reference standard are known to reviewers who interpret the index test.
- 3) Clinical review bias: This occurs when the reviewers use certain clinical information (e.g., symptoms, co-morbidities) to interpret the index test or reference standard. Clinical information may affect the estimates of DTA.
- 4) Incorporation bias: This occurs when the result of the index test is used in establishing the final diagnosis (i.e., it forms part of the reference standard). It usually overestimates the test accuracy of index tests.

4. Analysis-related bias

- 1) Managing indeterminate results: A diagnostic test may produce indeterminate results (e.g., lesions that are too small to characterize). If these indeterminate results are removed from the analysis, biased assessment of the test accuracy occurs (generally overestimation of test accuracy).
 - 2) Arbitrary choice of the threshold value: If the choice of the threshold (i.e., cutoff) value for the index test maximizes its sensitivity and specificity, the test accuracy may be overestimated. The test may perform less well at the selected cutoff when evaluated in a new set of patients or population.
-