

Appendix 2 – Indicators for the assessment of quality using the QUADAS-2.

Domain	Signaling questions and indicators
Patient selection	<p data-bbox="363 365 1372 443"><i>Was the spectrum of patients representative of the patients who will receive the test in practice?</i></p> <p data-bbox="363 454 1372 533">Patients with an unclassified headache will receive the tests in practice, therefore this should be the sample used in the studies.</p> <ul data-bbox="411 544 1372 1462" style="list-style-type: none"> <li data-bbox="411 544 1372 813">• Classify as ‘yes’ if (1) patient enrollment was consecutive or random, (2) a case-control design was avoided and (3) inappropriate exclusions were avoided. Information should be given about clinical setting, recruitment of patients and the in- and exclusion criteria. <ul data-bbox="507 734 1372 813" style="list-style-type: none"> <li data-bbox="507 734 1372 813">○ When all three items are classified as yes, there was considered to be a low risk of bias. <li data-bbox="411 824 1372 1137">• Classify as ‘no’ if (1) patient enrollment was not consecutive or random, or (2) a case-control design was applied or (3) inappropriate exclusions were present. Examples of inappropriate exclusions are: (1) excluding secondary headaches when the test was aimed at diagnosing primary headaches or (2) excluding patients with multiple headache types (<i>patients in practice may have a second(ary) headache, the test should be able to discriminate between the headaches properly to be used in practice</i>) <ul data-bbox="507 1149 1372 1227" style="list-style-type: none"> <li data-bbox="507 1149 1372 1227">○ When one or more items were classified as no, there was considered to be a high risk of bias. <li data-bbox="411 1238 1372 1462">• Classify as ‘unclear’ if the information stated above was not available or unclear from the article. <ul data-bbox="507 1339 1372 1462" style="list-style-type: none"> <li data-bbox="507 1339 1372 1462">○ When one or more items were classified as unclear and no items were classified as no, the risk of bias was considered to be unclear.
Index test	<p data-bbox="363 1485 1372 1585"><i>Is the index test likely to classify the target condition correctly? Could the conduct or interpretation of the index test have introduced bias?</i></p> <p data-bbox="363 1597 1372 1765">Blinding of the results of other tests or specific patient population will decrease the risk of bias. Using pre-specified thresholds when applicable will also decrease the chance that bias is introduced after the index test is applied. As different index tests were included, no definition can be given for the index test.</p> <ul data-bbox="411 1776 1372 2002" style="list-style-type: none"> <li data-bbox="411 1776 1372 2002">• Classify as ‘yes’ if (1) the index tests results were interpreted without knowledge of the results of the reference standard and (2) if the threshold used was pre-specified (when applicable). The index test should be clearly described as well as how the diagnosis was determined (cut-off points, thresholds or criteria).

	<ul style="list-style-type: none"> ○ When both items are classified as yes, there was considered to be a low risk of bias. ● Classify as ‘no’ if (1) the index tests results were interpreted with knowledge of the results of the reference standard or (2) if the threshold used was determined after the test was applied (when applicable). <ul style="list-style-type: none"> ○ When one or more items were classified as no, there was considered to be a high risk of bias. ● Classify as ‘unclear’ if the information stated above was not available or unclear from the article. <ul style="list-style-type: none"> ○ When one or more items were classified as unclear and no items were classified as no, the risk of bias was considered to be unclear.
Reference standard	<p><i>Is the reference standard likely to classify the target condition correctly? Could the conduct or interpretation of the reference standard have introduced bias?</i></p> <p>Blinding of the results of other tests or specific patient population will decrease the risk of bias. The reference test should be the gold standard: ICHD, ICHD-II or ICHD-3 or a headache specialist trained in using these criteria.</p> <ul style="list-style-type: none"> ● Classify as ‘yes’ if (1) the reference standard is the gold standard (ICHD, ICHD-II or ICHD-3) or otherwise likely to correctly classify the target condition and (2) the reference standard results were interpreted without knowledge of the results of the index test. <ul style="list-style-type: none"> ○ When both items are classified as yes, there was considered to be a low risk of bias. ● Classify as ‘no’ if (1) the reference standard is not the gold standard and not likely to correctly classify the target condition (<i>e.g. interview with someone other than a trained headache specialist</i>) and (2) the reference standard results were interpreted with knowledge of the results of the index test. <ul style="list-style-type: none"> ○ When one or more items were classified as no, there was considered to be a high risk of bias. ● Classify as ‘unclear’ if the information stated above was not available or unclear from the article. For example: when the diagnosis from a neurologist was the reference standard, without explicitly stating the neurologist is trained in, or used the ICHD criteria. <ul style="list-style-type: none"> ○ When one or more items were classified as unclear and no items were classified as no, the risk of bias was considered to be unclear.

<p>Flow and timing</p>	<p><i>Was the application of the index test and reference standard adequate in terms of interval between the two tests and inclusion of all participants?</i></p> <p>Ideally the index test and reference standard should be carried out on the same day. However, it is unlikely that headache complaints will change within 0 – 4 weeks, so this time interval was determined to be adequate. By including all participants in the analyses, the results will be most accurate and true to reality.</p> <ul style="list-style-type: none"> • Classify as ‘yes’ if (1) an appropriate time interval between the index test and reference standard was present, (2) all patients received (the same) reference standard and (3) all patients were included in the analysis. <ul style="list-style-type: none"> ○ When all three items are classified as yes, there was considered to be a low risk of bias. • Classify as ‘no’ if (1) there was not an appropriate interval between the index test and reference standard, or (2) not all patients received (the same) reference standard or (3) not all patients were included in the analysis. <ul style="list-style-type: none"> ○ When one or more items were classified as no, there was considered to be a high risk of bias. • Classify as ‘unclear’ if the information stated above was not available or unclear from the article. <ul style="list-style-type: none"> ○ When one or more items were classified as unclear and no items were classified as no, the risk of bias was considered to be unclear.