Supplementary Table S2 – Modified version of QUADAS-2. The checklist was used to assess the study quality. All items were scored with "yes", "no" or "unclear". Items 3-7 and 10 were excluded as they were not considered relevant in the current context. We added 6 new items to the checklist (items 15 to 20) as relevant for our review.

Item	Yes	No	Unclear
1. Was the spectrum of patients representative of the patients who will receive the test in practice? *(Unselected patients recruited from the general population?)*	( )	( )	( )
2. Were selection criteria clearly described? *(Clear definition of the criteria used in- and exclusion criteria for entry into the study)*	( )	( )	( )
3. ~~Is the reference standard likely to correctly classify the target condition?~~			
4. ~~Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?~~			
5. ~~Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?~~			
6. ~~Did patients receive the same reference standard regardless of the index test result?~~			
7. ~~Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?~~			
8. Was the execution of the index test described in sufficient detail to permit its replication? *(Was the tool/tools described in sufficient detail to permit its replication (a final algorithm)?)*	( )	( )	( )
9. Was the execution of the reference standard described in sufficient detail to permit its replication? (*Was the fracture collection verified and not only self-reported?)*	( )	( )	( )
10. ~~Were the index test results interpreted without knowledge of the results of the reference standard?~~			
11. Were the reference standard results interpreted without knowledge of the results of the index test? (*Was the risk of fracture calculated without the knowledge of the outcome (fracture)?)*	( )	( )	( )
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice? *(Is it possible to collect the risk factors included the tool in clinical practice?)*	( )	( )	( )
13. Were uninterpretable, indeterminate or intermediate test results reported? *(Were the any uninterpretable, indeterminate or intermediate results and were the results reported for all patients who were described as having been entered into the study?)*	( )	( )	( )
14. Were withdrawals from the study explained? *(A patient flow diagram or results available for all patients who were reported to have been entered into the study)*	( )	( )	( )
15. Were the data on risk factors obtained by clinical interview (as opposed to self-reported)?	( )	( )	( )
16. Were the baseline demographic and clinical features of study participants adequately described? *(Age, (BMD if measured) and risk factors for fracture included in the tool/tools used in the study (no more than 2 risk factors not reported in baseline description)?)*	( )	( )	( )
17. Were all the data needed to calculate the score of the tool/tools available on all subjects? ***(****No missing data on the risk factors included in the tool/tools?)*	( )	( )	( )
18. Is the study sample over 1.000 subjects?	( )	( )	( )
19. Did the tool validation study include over 100 events of interest?	( )	( )	( )
20. Was the follow-up period equal to the "recommended" by the tools included in the study? *(5 or 10 years for all subjects included in the study, depending on the outcome period of the tools)*	( )	( )	( )