Supplementary Table 1. Signaling questions in the original and modified QUADAS-2 tool

Domain		Original QUADAS-2 tool	Modified QUADAS-2 tool
1. Patient selection	Risk of bias	Was a consecutive or random sample of patients enrolled? Was a case-control design avoided? Did the study avoid inappropriate exclusions? Could the selection of patients have introduced bias?	Was a consecutive or random sample of patients enrolled? Was a case-control design avoided? Did the study avoid inappropriate exclusions? Could the selection of patients have introduced bias?
	Concerns regarding applicability	Is there concern that the included patients do not match the review question?	Is there concern that the included patients do not match the review question?
2. Index test	Risk of bias	Were the index test results interpreted without knowledge of the results of the reference standard? Could the conduct or interpretation of the index test have introduced bias?	Were investigators, calculating the FRAX score, blinded to fracture status? Was the execution of the FRAX tool described in sufficient detail to permit its replication? Could the conduct or interpretation of the index test have introduced bias?
	Concerns regarding applicability	Is there concern that the index test, its conduct, or interpretation differ from the review question?	Is there concern that the index test, its conduct, or interpretation differ from the review question?
3. Reference standard[a]	Risk of bias	Is the reference standard likely to correctly classify the target condition? Could the reference standard, its conduct, or its interpretation have introduced bias?	Was fracture status verified (yes) or only self-reported (no)? Or was osteoporosis defined according to the standard T-scores (yes)? Could the reference standard, its conduct, or its interpretation have introduced bias?
	Concerns regarding applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Is there concern that the target condition as defined by the reference standard does not match the review question?
4. Flow and timing	Risk of bias	Did all patients receive a reference standard? Did patients receive the same reference standard? Were all patients included in the analysis? Could the patient flow have introduced bias?	Were at least 90% of eligible participants included in the analysis?
Additional questions	Were the data on risk factors obtained by clinical interview (as opposed to self-reported)? 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]?) Was 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?) Is the study sample over 1000 subjects? Did the tool validation study include over 100 events of interest? 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)		

[a]One question from domain 2: index test ("If a threshold was used, was it pre-specified?"), 1 question from domain 3: reference standard ("Were the reference standard results interpreted without knowledge of the results of the index test?"), and 2 questions from domain 4: flow and timing ("Did all patients receive a reference standard?" and "Did patients receive the same reference standard?") were excluded from the modified QUADAS-2 tool due to irrelevance.

QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies 2; FRAX, fracture risk assessment tool; BMD, bone mineral density.