

Supplementary Figures S1 and S2– *Methodological quality of the studies with QUADAS-2.*

According to our assessment with QUADAS-2 (Figure S1), the average quality of 45 studies was higher in item 12 - similarities between data available during study interpretation and clinical practice; item 2 - description of the selection criteria; and items 8 and 9 - provision of sufficient details to allow replication. However, many studies did not report enough data to analyse the accuracy of the tools at the end of study (item 17) or at interim/intermediate analysis (item 13). The reasons for withdrawal are also lacking in many articles (item 14). The number of participants lost during the follow up due to death is conspicuously missing in most studies. Adherence to the recommended time of follow-up for the used tool (item 20) was only present in 16 studies.

Figure S1 – Quality assessment of studies testing fracture risk prediction tools (n=45). Green= Yes; Yellow=unclear; Red= No.



We also compared the quality of reports dealing with the 3 most developed tools (Figure S2). Articles on FRAX[®] performed better than average on items 11 and 13, while GARVAN's articles performed better on items 2, 8, 9, 14, 16, 17 and 20 and QFracture[®] studies on items 1, 2, 15 and 19.

Figure S2 – Percentage of articles complying with quality criteria, according to risk prediction tool under evaluation. * QFracture[®] has only been validated for the 10-years prediction interval.

