

Supplementary Text S1: Quality indicator assessment criteria

Conflict of interest in the development or implementation processes was assessed from any statements of conflict of interest in the article; if no statement made, the reviewers' judgment was used to assess whether it was likely any conflict of interest would have affected the results.

Source of evidence/evidence synthesis: indicators should ideally have been based upon evidence derived from a systematic review.[23]

Consensus method used in deciding on the indicator (including selection and prioritisation for implementation studies): a RAND Appropriateness Method was rated most highly, followed by the Delphi technique.[24]

Method of updating: if the article was published since the EULAR, NICE or OARSI guidelines, it was regarded as having access to a broadly similar evidence base as the guidelines themselves. Ideally, indicators should have an established mechanism for an update cycle to integrate new evidence and review the continued validity of the indicators.

Reproducibility of an indicator was assessed by identification of similar indicators independently created from more than one study.

Reliability was assessed through implementation studies which used multiple data extractions of quality indicator results from the same dataset to determine whether the same result was obtained.

External validity was assessed by examination of implementation studies for evidence of quality improvement in populations in which the indicator had been used.[25]

Sensitivity to change was assessed through implementation studies to determine whether quality improvement over time was demonstrated where quality improvement was known independently to have occurred.

Feasibility, including identification of the applicable population, method of measurement: this was assessed through any recommendation of method for implementation in the development studies, or through implementation studies, and whether such methods were judged compatible with routine general practice in the UK.

Bias was not considered in the implementation studies as no comparison was made between groups.