External factors (e.g. financing, accreditation) were excluded as these are likely to be specific to the local health system.

Extent to which this was possible depended on the existence of agreed construct definitions in multiple included studies or, alternatively, in synthesised sources from the extant literature (i.e. recent or seminal review article).

**Stage 1 – Searching & screening**

- Abstract and full text screening for inclusion in Stage 2

**Stage 2 – Taxonomy development**

- Development of taxonomy for categorizing instrument content
  - Iterative process involving:
    - Comparison of instrument content with initial taxonomy to identify a comprehensive set of constructs/categories
    - Collapsing similar constructs and categories
    - Labelling constructs and categories to reflect prevailing conceptualisation

**Stage 3 – Categorization of instrument content**

- Categorization of instrument content using the taxonomy; tabulation to enable comparison of instruments and illustrate coverage of InQuIRe framework

**Stage 4 – Assessment of measurement properties**

- Summary and tabulation of:
  - The instrument’s purpose, content and theoretical basis
  - Methods of development
  - Methods and main findings of assessments of measurement properties (validity, reliability, responsiveness)
  - Appraisal of evidence supporting measurement properties, using COSMIN criteria

Sources and inclusion criteria by stage

- MEDLINE, PsycINFO, HAPI, reference lists of systematic reviews of measurement instruments
- Snowballing: reference lists, citation searching (main papers)

Papers meeting initial inclusion criteria:

1. Reported development or use of one or more instruments intended to measure a construct a) within the scope of the InQuIRe framework (team domains) OR b) identified by the study authors as a potential determinant of team effectiveness and judged as relevant to QI teams in primary care;
2. Instrument was quantitative, allowing for statistical analysis of its measurement properties;
3. Instrument was intended to be self-report;
4. Information about the instrument was published in English, in the peer-reviewed literature and sufficient detail was reported to enable assessment of its content

**Stage 3 – Categorization of instrument content**

- InQuIRe conceptual framework as initial basis for structure and content of taxonomy
- Data from content analysis of individual instruments
- Review and conceptual articles identified in search or subsequent purposive search
- Other relevant sources

Subset of instruments from Stage 2 that:

1. Based on content analysis, were confirmed as measuring a construct relevant to evaluating QI in primary care
2. a) Had item content and wording suitable for QI teams in primary care (minor changes to wording were acceptable); OR b) Measured a construct not adequately covered by more suitable instruments, and had potential to be adapted for QI teams in primary care.

Subset of instruments from Stage 3 that:

1. Based on content analysis, the instrument appears to be an adequate measure of a relevant construct;
2. The instrument a) was used in, or developed for, primary care OR b) has appropriate item content and wording for QI teams in primary care (minor changes to wording were acceptable) and a more appropriate equivalent was not identified;
3. Sufficient detail is reported in the paper, or linked additional files, to enable assessment of the development of the instrument and evidence of its measurement properties;
4. An English-language version is available; and
5. The instrument is not proprietary.