Appendix 1: Analytic plan

Analysis of variance tests were calculated to test (i) if domain quality ratings of the appraised guideline using the AGREE with the 7-point scale varied as a function of User Type and (ii) if usefulness ratings of the items and domains varied as a function of User Type.

Multiple regressions were calculated to test if overall outcomes (endorsements of, intentions to use, and overall quality) could be predicted from domain quality ratings, User Type of the interaction between the two factors.

Descriptive statistics were calculated to report on proportion of respondents who recommended modifications, edits or deletions to items.

Cronbach's α was used to calculate the internal consistency of each domain.

Intraclass-correlations coefficients were calculated to evaluate the inter-rater reliability of the domains for the average of 1, 2, 3, and 4 raters.

The Spearman Brown Prophecy coefficient was used to calculate the number of raters required to reach a minimum inter-rater reliability threshold of r = 0.70 .