OBJECTIVES: The EQ-5D, a measure of self-reported health status, has been operationalized in ways that depart from the original format. This study examines the equivalence of the original paper-based vertical format with a touch-screen-based horizontal format. METHODS: Non-probability sampling was employed to recruit 314 subjects intended to reflect the primary socio-demographic characteristics of the general adult population. A two part questionnaire was administered in a randomized crossover design. One part was the original paper-based 20cm vertical EQ-VAS; the other part was touch screen computer-based (designed by Assist Technologies) and included, among other items/scales, a horizontal EQ-VAS, the SF-36, and socio-demographic items. The two EQ-VAS formats were completed roughly ten minutes apart. To test for minimally important differences (MID) between EQ-VAS scores, a difference of half a standard deviation (~8 points on the 100 point scale) was used as the equivalence threshold. RESULTS: The mean (SD) EQ-VAS score was 81.0 (15.4) on the paper and 79.6 (15.2) on the touch-screen. The mean (CI) difference between scores on the two formats was 1.4 (0.19 to 2.58) points and the mean absolute difference was 5.3 (4.22 to 6.44) points. The intraclass correlation coefficient (ICC) was 0.75, indicating good agreement between the two scores. Almost a third (30.1%) of the respondents reported identical scores on both formats and 80.1% of the respondents had difference scores within ± eight points. Using nonparametric bootstrap techniques, both the mean difference and the mean absolute difference between scores on the two formats were significantly less (p < 0.001) than the equivalence threshold. In addition, data collected via touch screen may be more reliable since 22.2% of subjects did not complete the EQ-VAS paper format as instructed. CONCLUSION: These results provide evidence for the measurement equivalence of this EQ-VAS touch screen format with the original paper format.

METHODS: Two main linking approaches were assessed; linking the two scales by the nine additional (common) items identified and use of an independent anchor or test (the PGWB). Here, all items in the PGWB are combined with all items in each of the scales. RESULTS: Adding the nine common items identified to the RAQoL led to a scale with excellent fit to the Rasch model; Item Fit (mean = -0.19, SD = 1.22), Person Fit (mean = -0.21, SD = 0.89) and Person Separation Index (0.94). Adding the nine items to the QoL-AGHDA also led to excellent fit to the model; Item Fit (mean = -0.14, SD = 1.46), Person Fit (mean = 0.015, SD = 0.84) and Person Separation Index (0.96). Comparison of scores on the nine common items suggested that the AGH group had worse quality of life than RA patients. Use of the PGWB as a linking test led to considerable item misfit in both scales. CONCLUSION: Use of the PGWB as an anchor test was unsuccessful (probably as it assesses well-being (impairment) rather than QoL). For the purposes of constructing an item bank common item equating appears to be feasible. Such co-calibration provides an opportunity for valid and accurate comparisons of the impact of different diseases on patient groups. It must be noted that co-calibration requires that the scales to be linked adopt the same measurement model.