View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Sydney eScholarship

Journal of Clinical Epidemiology, in press, accepted 26/02/2013.

The Quality of Life Questionnaire Core 30 (QLQ-C30) and Functional Assessment of Cancer - General (FACT-G) differ in responsiveness, relative efficiency, and therefore required sample size

Madeleine T King,[1,2] Melanie L Bell,[1,2] Daniel Costa,[1,2] Phyllis Butow,[1,2] Byeongsang Oh[3]

1. Psycho-oncology Co-operative Research Group (PoCoG), University of Sydney, NSW, Australia
2. School of Psychology, University of Sydney, Sydney, NSW, Australia
3. Byeongsang Oh, Dept Medicine, Concord Repatriation General Hospital, University of Sydney, Concord, Australia

Abstract

Objective: QLQ-C30 and FACT-G are widely-used cancer-specific health-related quality of life (HRQOL) questionnaires. We aimed to compare their responsiveness to clinically important effects and statistical efficiency to detect such effects.

Study design and setting: Secondary analysis of QLQ-C30 and FACT-G data from a randomised controlled trial of Medical Qigong (n=162 heterogeneous cancer patients). Difference in responsiveness (DR) and relative efficiency (RE) were calculated for five domains.

Results: FACT-G total score was more efficient than QLQ-C30 global scale for detecting change within the intervention arm (RE=0.31 (0.083, 0.69)) and comparing change between trials arms (RE=0.17 (0.009, 0.58)). In the social domain, the QLQ-C30 scale was more responsive (DR=0.28 (0.024, 0.54)) and more efficient within arm only (RE=5.25 (1.21, 232.26)). In the physical, functional/role and emotional domains, neither questionnaire was more responsive or efficient.

Conclusion: FACT-G would require about one third the sample of QLQ-C30 to detect a given change in overall HRQOL, while in the social domain it would require five times the sample size. FACT-G won advantage in overall HRQOL by reduced “noise” (smaller standard deviation achieved by summing across 27 items), while QLQ-C30 won advantage in the social domain via a larger “signal” (achieved through well-targeted item content).

What is new?

- As a measure of overall health-related quality of life (HRQOL), the FACT-G total score is more responsive to change over time than the QLQ-C30 global scale, and has greater statistical efficiency and hence power, for both change within a group and for comparing change between two groups.
- In the social domain, the QLQ-C30 scale is more responsive than the FACT-G scale, and has greater statistical efficiency and hence power for the within-group change but not for comparing change between two groups.
- A randomised trial which used the QLQ-C30 to assess overall HRQOL would require a sample size approximately five times greater than one which used the FACT-G to detect a given difference between trial arms as statistically significant. FACT-G would require about one third the sample of QLQ-C30 to detect a given change in overall HRQOL, while in the social domain it would require five times the sample size.
- In the physical, emotional and role/functional domains, the FACT-G and QLQ-C30 have similar responsiveness, statistical efficiency and hence power and sample size requirements.

- These results should be included along with other relevant considerations to determine the optimal patient-reported outcome measure when planning clinical research in cancer.

Introduction

Various criteria come into play when choosing among candidate health-related quality of life (HRQOL) questionnaires (1). When choosing a HRQOL questionnaire to evaluate the effectiveness of an intervention, *responsiveness* to a clinically important effect is a key criterion (2). The more responsive the measure, the smaller the sample size required to detect a given effect (3) or the greater the power for a fixed sample size (4). Little information is currently available on the relative responsiveness of different HRQOL measures to guide researchers in their choice of such measures.

When selecting a HRQOL questionnaire for a cancer clinical trial, the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) (5) and Functional Assessment of Cancer Therapy - General (FACT-G) (6) are often obvious candidates (7). There is a large body of evidence supporting the validity and utility of each across a wide range of clinical research contexts, and both are available in many languages. Lucket *et al* (2011) provide a thorough comparison of the two measures, highlighting important differences in scale structure, social domains and tone that may inform choice for any particular study (7). Luckett *et al* concluded that available psychometric evidence does not recommend one questionnaire over the other. Their review highlights the limitations of available evidence about responsiveness, in particular that there are no head-to-head comparisons. The value of a head-to-head comparison is that it allows estimation of both the difference in responsiveness (4) and the relative statistical efficiency (3) of two measures. The latter is particularly useful as it estimates the factor by which sample size may be reduced if the more responsive measure was used.

Our aim was to compare the responsiveness, statistical efficiency and power of comparable scales from the EORTC QLQ-C30 and the FACT-G, and to quantify the sample size implications.

Methods

Data set

We conducted secondary analysis of data from a randomised trial of Medical Qigong (breathing and movement exercises). Details of the trial methods and results are reported elsewhere (8). In summary, the primary hypothesis of the trial was that patients randomised to Medical Qigong (intervention) would experience significant improvements in HRQOL compared with patients randomised to usual medical care (control). The Medical Qigong program ran for 10 weeks with two group-based supervised 90-minute sessions per week. Participants assigned to the usual care arm received usual medical care. Participants were 162 patients with mixed cancer diagnoses at variable stages of disease and treatment; n=79 Medical Qigong, n=83 usual care. The FACT-G (Version 4) total score was used as the *a priori* primary outcome measure; QLQ-C30 (Version 3) was included to assess responsiveness relative to FACT-G (note: the decision to include the QLQ was made mid-study). HRQOL was assessed prior to randomization ('pre') and 10 weeks later ('post'). The original trial analysis showed that participants in the Medical Qigong arm reported larger improvements in HRQOL as measured by the FACT-G than those in the usual care arm at 10-week follow-up (P < 0.001). Statistically significant benefits were also observed for all domains of HRQOL [physical, social, emotional and functional well-being, p < 0.001 for each].

HRQOL domains and scoring

In the current analysis, we focus on domains that are common to both the QLQ-C30 and FACT-G. Both questionnaires cover the four core domains of HRQOL: physical, functional/role, emotional, social. The wording and response scales of the items in these domains are shown in Table 1, as these may have a bearing on responsiveness and statistical efficiency. The QLQ-C30 has a global HRQOL scale (component items shown in Table 1), while the total score of the FACT-G is used as a composite measure of HRQOL. As per standard EORTC scoring algorithms, QLQ-C30 domain scores were calculated as the average response across component items, transformed linearly to a scale ranging from 0 to 100 (5). The standard FACT-G scoring algorithm is simply to sum responses of component items, after reversing selected items such that higher score represents better HRQOL for all (6). Further to this, we transformed these linearly to a 0-100 scale to facilitate comparison of mean changes between comparable domains of the QLQ-C30 and FACT-G. For both questionnaires, higher values represent better functioning, health and quality of life.

Statistical Methods

Responsiveness, power and relative efficiency

Responsiveness, difference in responsiveness and power were calculated following the methods of Tuley (4), and relative statistical efficiency was calculated following the method of Liang *et al* (3), in order to compare the QLQ-C30 with the FACT-G in each of four domains (physical, functional/role, emotional, social) and the QLQ-C30 global health /QoL scale with the FACT-G total scale.

A responsiveness index (RI) was calculated as the mean change in the intervention arm divided by the standard deviation of change in the control arm. The mean difference in RI between comparable QLQ-C30 and FACT-G scales was then calculated, with 95% confidence intervals. The null value for difference in responsiveness is 0.

Relative efficiency was calculated as the squared ratio of t-statistics, $(t_{QLQ}/t_{FACT})^2$. Here, t_{QLQ} is the t-statistic computed for a t-test using QLQ-C30 data, and t_{FACT} is the corresponding t-statistic for the FACT-G data. Two sets of relative efficiency (RE) statistics were calculated. The first assessed RE for change within trial arm; RE_{within} was based on a paired t-test (post-pre), using data from the intervention arm only. The second assessed RE for between group difference in change within group; $RE_{between}$ was based in a two sample t-test of the comparison of the pre-post difference in intervention versus usual care. For each RE estimate, bias-corrected accelerated 95% confidence intervals were generated with 1,000 bootstraps using SAS macros %boot and %bootci (9). The null value for RE is 1.

Missing data and imputation

Approximately 50% of the data were complete, with about 25% of patients having pre-intervention scores only. About 15% of patients had FACT-G scores for both time points but QLQ-C30 for the second time point only. We chose to impute the missing data rather than using the complete cases only, because the latter: 1) would reduce the effective sample size thereby reducing precision; 2) assumes the data are missing completely at random, which was not supported by the patterns of missingness (those who were missing the second assessment had poorer HRQOL scores in the first assessment, data not shown); 3) our estimates are likely to be less biased than if we had used the complete cases only (10). Multiple imputation was used. Four imputation models were implemented: QLQ-C30 measures by trial arm, and FACT-G measures by trial arm (11). Age and gender were also included in each of the imputation models. For each of the four imputation models, 100 multiple imputations were generated using the MCMC method in SAS Proc MI (12), then averaged by patient identification number to create one complete data set. The imputed dataset was analysed to address our study aims, as described above.

Descriptive statistics

The following descriptive statistics were calculated as these relate to scale precision and hence statistical efficiency and power: the standard deviation (SD) of domain scores, and the average inter-item correlation among items within a domain and Cronbach's alpha. The latter is a measure of scale reliability, or "internal consistency", and is a function of the number of items and the average correlation among those items; alpha increases with the number of items and with the degree of correlation among them (13). The degree of correlation between instruments within domains was also assessed. All but two of these statistics were calculated for both original data and imputed data to assess comparability; Cronbach's alpha and inter-item correlation could not be calculated for the imputed data as imputation was implemented for summated scales, not individual items. Histograms were plotted for the imputed dataset to assess the degree of ceiling effects (preponderance of scores at maximum value on scale), as these could limit the ability of a scale to register improvement, and hence its responsiveness to improvement in HRQOL. All descriptive statistics were calculated for baseline and post-intervention data.

Results

Descriptive statistics of the HRQOL scales

Descriptive statistics for baseline data are shown in Table 2; results for the post-intervention data were generally similar. In the physical domain, the FACT-G scale had higher inter-item correlation, more items and hence higher Cronbach alpha. In contrast, in the emotional domain, the FACT-G scale had lower inter-item correlation, and despite more items, lower Cronbach alpha. In the remaining domains, the effect of lower inter-item correlation in the FACT-G was countered by more items, such that Cronbach alpha values were similar between the two questionnaires. Estimates of SD and correlations tended to be slightly higher in the original data than in the imputed data. The lowest inter-scale correlation was for the social domain (<0.3) and the highest correlation was for the global domain (>0.6).

Histograms of the baseline imputed data (Figure A, online only) show ceiling effects in the social domain for both questionnaires. In the physical, functional and emotional domains, there was a slight degree of ceiling effect for both questionnaires. Neither the FACT-G total score or QLQ-C30 global score had a ceiling effect.

Responsiveness

Responsiveness index estimates ranged from 0.22 (FACT-G Social Well-being) to 0.84 (FACT-G total score) (Table 3). In the social domain, the QLQ-C30 responsiveness index was significantly larger than that of the FACT-G (Figure 1). For the remaining domains, responsiveness did not differ significantly.

Relative efficiency

The FACT-G total score was more efficient than the QLQ-C30 global scale, with a relative efficiency (RE) of 0.31 (95% CI: 0.083, 0.69) for the paired t-test in the intervention arm (Table 3, Figure 2A) and 0.17 (95% CI: 0.009, 0.58 for the two sample t-test (Table 4, Figure 2B). The QLQ-C30 was more efficient for the social domain for the paired t-test but not for the two-sample t-test.

Sensitivity analysis

Some of the RE distributions were quite skewed, as evidenced by their wide confidence intervals (note log scale for RE in Figure 2). To see if extreme differences in FACT and QLQ scores were driving results, we performed sensitivity analysis by setting to missing all values where the absolute value of the difference between the FACT change score and the QLQ change score was more than 50 (there were 11 such observations). We then performed the same multiple imputation, bootstrapping and computation of the RE as described above. We found that neither the estimate nor the confidence intervals changed substantially, except for the two sample t-test case for the physical domain, which changed from 1.63, 95% CI: (0.45,13.17) to 0.88 (0.23, 3.18). However, both confidence intervals contain 1, so the inference (no evidence of difference in efficiency) remains the same.

Discussion

We detected differences in responsiveness and statistical efficiency between FACT-G and QLQ-C30 for two of five pairs of scales. The FACT-G total score was more efficient than the QLQ-C30 global scale as a measure of overall HRQOL, for both change within the intervention arm (paired t-test) and for comparing change between trials arms (two sample t-test). The estimate of relative efficiency for the latter case was 0.17, meaning that a randomised trial which used the QLQ-C30 to assess global HRQOL would require a sample size approximately five times greater than one which used the FACT-G to detect a given difference as statistically significant. In the social domain, the QLQ-C30 scale was more responsive than the FACT-G scale and more efficient for the within-group t-test only. Since power and sample size are functions of one another, if we need fewer subjects for the same power, it is equivalent to saying that for a set sample size, one is more powerful than the other. We can therefore infer power from the relative efficiency results.

Given the conventional interpretation of the squared t ratio that we used to assess relative efficiency, our results suggest FACT total score would require about one third of the sample required by QLQ global QOL scale to detect a given change within a group. In contrast, FACT social wellbeing scale would require about five times the sample size required for QLQ social functioning scale. Wide confidence intervals on our estimates of relative efficiency reflect considerable uncertainty in the actual size of these differentials.

Statistical efficiency may not be the only consideration in choosing between these two measures. It is paramount that the content of the chosen measure's scales matches the specific QOL-related construct(s) of interest in any particular trial (14). So, for example, if a truly global assessment of QOL is required, then the QLQ-C30 global QOL scale would be more appropriate than the FACT-G total scale, regardless of the latter's superior statistical efficiency. This is because the latter is not a direct measure of global HRQOL but rather a composite of somewhat disparate items, each one either causing or reflecting QOL. Table 1 shows that only one of the 27 FACT-G items provides a direct measures of global quality of life: "I am content with my quality of life right now" (item FWB7), whereas the two items of the QLQ-C30 global scale directly assess "overall quality of life" and "overall health", allowing each respondent to implicitly define and weight the components of QOL and health. If the social domain of QOL was the primary outcome of a psychosocial intervention, then the choice between QLQ-C30 and FACT-G may be driven more by the difference in their content (Table 1) than by their relative statistical efficiency. Further, if multiple aspects of QOL are all important, then sample size needs to be adequate for the least efficient of the target scales within a measure. Thus statistical efficiency is only one of a number of considerations both in the choice of measure and the determination of sample size.

What might explain the observed differentials in responsiveness and statistical efficiency? Tuley's responsiveness index and the t-statistic (whether paired or two-sample) are 'signal-to-noise' ratios; in each case, the denominator ('signal') reflects the mean change observed on the HRQOL scale, and the numerator ('noise') reflects variability among individuals in change in HRQOL. What might amplify or attenuate the signal, and what might increase or decrease the noise? First, consider the content of the items in each scale; the extent to which these issues are likely to change, given the patient population, time period and intervention will have a direct bearing on signal. The social scale of the QLQ-C30 addresses interference with family and social activities, whereas the social scale of the FACT-G focuses on family and social support (Table 1). Patients in the Medical Qigong arm registered, on average, an 11% improvement on the QLQ-C30 scale, but less that 4% the FACT-G scale. It is plausible that interventions such as Medical Qigong can reduce the interference of cancer and its medical treatment on family and social activities, but have less impact on family and social support.

Second, consider the number of items in a scale. Assuming that the items are drawn from an infinitely large pool of items reflecting a common domain, increasing the number of items in a scale should increase the scale's reliability, hence precision, and thereby reduce 'noise' (15). However, the number of items is only half of the story; the underlying assumption is the other half. Cronbach's alpha is a function of the number of items and the average correlation among those items; alpha increases with the number of items in a scale and with the degree of correlation among those items (13). The FACT-G global scale (27 items) had the largest alpha (0.90), but only marginally greater than that of the 2-item QLQ-C30 global QOL scale (0.86). In the social domain, the 2-item QLQ-C30 scale and the 7-item FACT-G achieved the same alpha (0.84). Inspection of the content of Table 1 reveals that within any domain, the FACT-G items tend to be a more disparate collection than the QLQ-C30; Table 2 confirms that have lower inter-item correlations for all but the physical domain. Thus the advantage of more items in FACT-G scales is countered by their lower inter-item correlation. Luckett *et al* anticipated that FACT-G subscales should be more responsive than their QLQ-C30 counterparts due to their larger number of items (1). In the current analysis, this prediction was born out only for extreme case: the 27-item FACT-G total score versus the 2-item QLQ-C30 global QOL score, highlighting the importance of inter-item correlation. Interestingly, the scale with the lowest alpha (0.70) was the QLQ-C30 Physical Functioning scale, which is unique among the scales we studied in being a Guttman scale; as Table 1 shows, its items are ranked in order of difficulty so that an individual who agrees with a particular item is also likely to agree with items of lower rank-order. Despite this low alpha score, this scale was no less responsive or efficient than the FACT-G Physical Wellbeing scale in our analyses.

Perhaps a more directly relevant indicator of 'noise' in the current analysis is the standard deviation of change of each scale, as this plays directly into the calculation of both the responsiveness index and the t-tests on which relative efficiency is based. For the Global domain, the SD of the FACT-G scale was about half the size of the SD of the QLQ-C30 scale. As both detected about the same 'signal' , the greater statistical efficiency of the FACT-G scale must be attributed to its smaller 'noise'. What led to this smaller SD? As an aggregate of 27 items, the FACT-G total score has many possible values, and thus a far more finely graded (high resolution) scale than the relatively coarse QLQ-C30 Global QOL scale; this may be one explanation. Another is that for the QLQ-C30, respondents are required to make a global rating on two items (QOL and health), an approach which potentially introduces varying interpretations of these items as an additional source of noise.

The third potential factor in the observed differentials in responsiveness and statistical efficiency is ceiling effects, as these can limit the ability of a scale to register improvement, and hence its responsiveness to improvement in HRQOL. However, as the degree of ceiling effects was about the same in both the social scales, and as the neither of the global scales displayed a ceiling effect, this factor was clearly not at play in our results.

A strength of our analysis is that it is based on data from a randomised trial which demonstrated a benefit to HRQOL across all FACT-G domains. Responsiveness index estimates ranged from 0.22 (FACT-G Social Well-being) to 0.84 (FACT-G total score). Although Tuley's responsiveness index in not a standard effect size (since the denominator and numerator are from intervention and control samples, respectively), the form is similar enough to interpret their sizes in a similar way; as ranging from small to large with all but the Social Well-being results being moderate to large (16). This is not surprising, as it is well-established that exercise and yoga improve the HRQOL and psychological well-being of people with cancer (17, 18); note that many of the studies in these two meta-analyses used FACT-G or QLQ-C30. The Medical Qigong RCT (8) therefore provides a suitable dataset for the estimation of responsiveness (19). However, our conclusions may not generalise to medical interventions, which may affect other aspects of HRQOL, such as disease symptoms or treatment side-effects. A limitation of our dataset is the missing data, particularly for QLQ-C30, which was added to the assessment schedule after the trial had begun. We chose to impute missing data in order to maximise precision and minimise bias, and used the best available means to do so, multiple imputation. Further strengths of our analysis are the calculation of confidence intervals and sensitivity analysis. We note that the confidence intervals on our estimates of relative efficiency are very wide (note the log scale on Figure 2). In part, this is because relative efficiency is a squared ratio.

This secondary analysis of data provides the first head-to-head comparison of the responsiveness and relative statistical efficiency of the QLQ-C30 and FACT-G. This is valuable information for people considering these as candidate measures for assessing HRQOL in a clinical trial. More generally, this paper demonstrates the utility of estimating relative efficiency. Results for difference in responsiveness (after Tuley) and relative efficiency (after Liang et al) were consistent, and given its practical interpretation in terms of sample size implications, the latter is probably the most useful. Further, the two-sample version of relative efficiency, based on data from a RCT, is likely to be the most informative for future RCTs. However, it relies on the use of two candidate instruments which measure the same thing (head-to-head comparison), which is rarely done in an RCT, due to cost and patient-burden. Thus, while replication of the analyses in this paper for a range of interventions and patient populations would be interesting and valuable, it may not be feasible.

References

1. Luckett T, King MT. Choosing patient-reported outcome measures for cancer clinical research--practical principles and an algorithm to assist non-specialist researchers. European Journal of Cancer. 2010;46(18):3149-57.
2. Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. Clinical Therapeutics. 1996;18(5):979-92.
3. Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. Arthritis & Rheumatism. 1985;28(5):542-7.
4. Tuley MR, Mulrow CD, McMahan CA. Estimating and testing an index of responsiveness and the relationship of the index to power. Journal of Clinical Epidemiology. 1991;44(4-5):417-21.
5. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. Journal of the National Cancer Institute. 1993;85:365-76.
6. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. Journal of Clinical Oncology. 1993;11(570-579).
7. Luckett T, King MT, Butow PN, Oguchi M, Rankin N, Price MA, et al. Choosing between the EORTC QLQ-C30 and FACT-G for measuring health-related quality of life in cancer clinical research: issues, evidence and recommendations. Annals of Oncology. 2011;22(10):2179-90.
8. Oh B, Butow P, Mullan B, Clarke S, Beale P, Pavlakis N, et al. Impact of medical Qigong on quality of life, fatigue, mood and inflammation in cancer patients: a randomized controlled trial. Annals of Oncology. 2010;21(3):608-14.
9. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Statistics in Medicine. 2000;19(9):1141-64.
10. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient reported outcomes. WHAT? 2012;in press.

11. Little RJA, Rubin DB. Statistical analysis with missing data. 2nd ed. Chichester: Wiley; 2002.

12. SAS 9.2 Online Documentation. Cary, NC: SAS Institute Inc; 2008.

13. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;6:297-334.

14. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage M, et al. Reporting of Patient Reported Outcomes in Randomised Trials: the CONSORT PRO Extension. JAMA. in press.

15. Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. New York: McGraw-Hill; 1994.

16. King MT, Stockler MR, Cella DF, Osoba D, Eton DT, Thompson J, et al. Meta-analysis provides evidence-based effect sizes for a cancer-specific quality-of-life questionnaire, the FACT-G. Journal of Clinical Epidemiology. 2010;63(3):270-81.

17. Ferrer RA, Huedo-Medina TB, Johnson BT, Ryan S, Pescatello LS. Exercise interventions for cancer survivors: a meta-analysis of quality of life outcomes. Annals of Behavioral Medicine. 2011;41(1):32-47.

18. Lin K-Y, Hu Y-T, Chang K-J, Lin H-F, Tsauo J-Y. Effects of yoga on psychological health, quality of life, and physical health of patients with cancer: a meta-analysis. Evidence-Based Complementary and Alternative Medicine. 2011;Article ID 659876:12 pages.

19. Revicki D, Hays RD, Cella D, Sloan J, Revicki D, Hays RD, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. Journal of Clinical Epidemiology. 2008;61(2):102-9.

Figure 1 Difference in responsiveness index (95% confidence intervals) of comparable scales of FACT-G and QLQ-C30

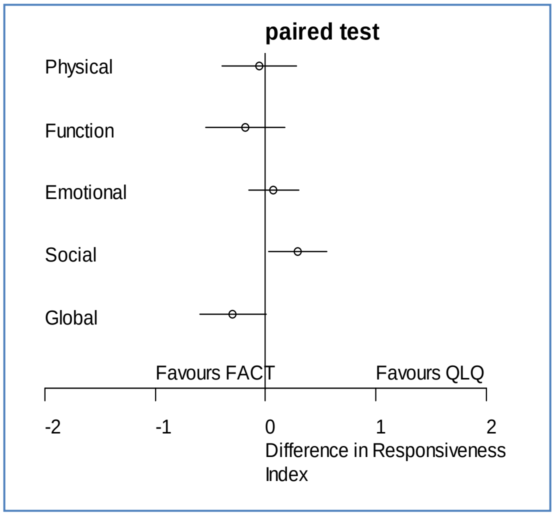
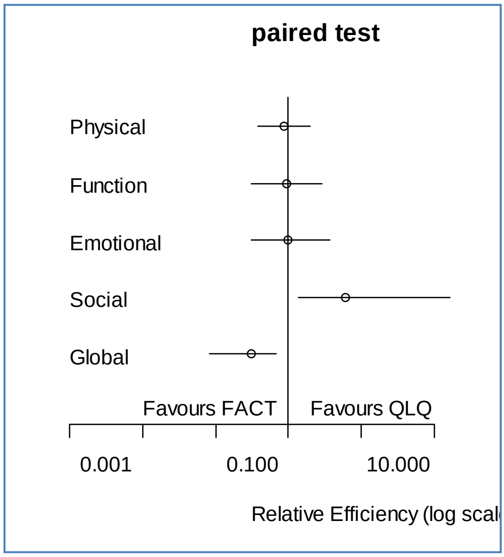
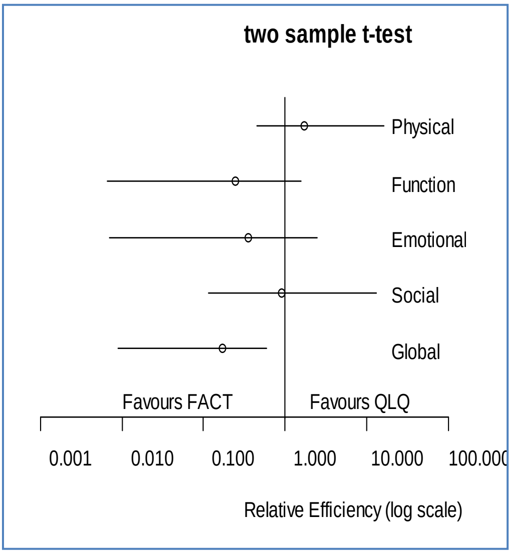


Figure 2 Relative efficiency (95% confidence intervals) of comparable scales of FACT-G and QLQ-C30, based on the paired t-test (assessing change within the Medical Qigong arm, Panel A) and the two-sample t-test (comparing change in the intervention arm with change in the control arm, Panel B)

Panel A **Panel B**


paired test
Physical
Function
Emotional
Social
Global
Favours FACT
Favours QLQ
0.001
0.100
10.000
Relative Efficiency (log scal


two sample t-test
Physical
Function
Emotional
Social
Global
Favours FACT
Favours QLQ
0.001
0.010
0.100
1.000
10.000
100.000
Relative Efficiency (log scale)