Responsiveness was Similar between Direct and Mapped SF-6D in Colorectal Cancer Patients who declined

Submitted to Journal of Clinical Epidemiology

Original Article

Correspondence Author: Dr. Carlos King Ho WONG, PhD, MPhil, BSc

Institution: Department of Family Medicine and Primary Care, The University of Hong Kong

Address: 3/F, Ap Lei Chau Clinic, 161 Ap Lei Chau Main Street, Ap Lei Chau, Hong Kong

Contact: +852-25185688 (tel); +852-28147475 (fax) carlosho@hku.hk (email)

Order of Author: Carlos K.H. Wong, PhD, MPhil, BSc*[1], Brendan Mulhern, MRes[2], Y.F. Wan, MSc, BSc[1], Cindy L.K. Lam, MD, MBBS[1]

[*] First and correspondence Author

[1] Department of Family Medicine and Primary Care, The University of Hong Kong

[2] Health Economics and Decision Science, ScHARR, The University of Sheffield

Abstract

Objective: To evaluate the responsiveness of generic and mapped preference-based measures based on the anchor of global change in health condition of colorectal cancer (CRC) patients.

Study Design and Setting: A baseline sample of 333 Chinese CRC patients was recruited between 09/2009 and 07/2010, and was surveyed prospectively at 6-month follow-up. Preference-based indices were derived from the generic SF-6D measure ($SF\text{-}6D_{Direct}$), from the Short Form-12 ($SF\text{-}6D_{SF\text{-}12}$) and mapped from the condition-specific Functional Assessment of Cancer Therapy-Colorectal ($SF\text{-}6D_{FACT\text{-}C}$). Responsiveness of three measures was assessed using standardized effect size, standardized response mean, and responsiveness statistic, and receiver operating characteristic (ROC) curve analysis.

Results: The $SF\text{-}6D_{SF\text{-}12}$ and $SF\text{-}6D_{FACT\text{-}C}$ indices were significantly more responsive to detect positive changes than the $SF\text{-}6D_{Direct}$ index in improved groups. In worsened group, the $SF\text{-}6D_{Direct}$ and $SF\text{-}6D_{FACT\text{-}C}$ indices showed significant decline from baseline to 6-month. The areas under the ROC curve for $SF\text{-}6D_{Direct}$ and $SF\text{-}6D_{FACT\text{-}C}$ indices were not statistically different from 0.7. The $SF\text{-}6D_{FACT\text{-}C}$ index was more responsive to changes in health status compared with other indices.

Conclusion: Direct SF-6D measure was more responsive than mapped preference-based measures in improved group but the direction was reversed in worsened group. Use of a preference-based index mapped from a condition-specific measure captures both negative and positive important changes among CRC.

Running Title: Responsiveness of SF-6D measures for CRC

Keywords: Colorectal cancer; Responsiveness; Anchor; SF-6D; Preference-based; Mapping

Words Count: 202

What is new?

Key finding:

- The preference-based measure mapped from condition-specific FACT-C was responsive in both improved and worsened groups for patients with colorectal cancer.
- Direct preference-based measure was more responsive than mapped preference-based measures in improved group but the direction was reversed in worsened group.

What this adds to what was known:

- Responsiveness of generic and mapped preference-based measures was compared in a cancer population for the first time.

What is the implication, what should change now:

- Use of algorithm that maps condition-specific measure onto preference-based measure improved the responsiveness in patients with colorectal cancer.

Manuscript Text

Introduction

Colorectal cancer (CRC) is highly prevalent around the world [1], and the costs of treatment for CRC for health services are projected to increase [2]. Continuing advances in pharmaceutical and other treatments for CRC has meant that research has been directed towards developing and testing interventions with evidence of both clinical and cost effectiveness. In economic evaluations, the quality adjusted life year (QALY), that combines the quantity and quality of life into a single metric, can be used as the outcome measure [3]. The quality weight can be gained from a preference-based measure, which provides a single figure utility score (anchored on the 1 to 0; full health – dead scale) that is derived from the preferences of the general population for certain health states described by the measures over others. A range of generic preference-based measures are available (including the EQ-5D, SF-6D and HUI-3), and it is also possible to map condition specific measures onto the utility scale[4]. In part, the choice of measure to use can be informed by the psychometric properties of the instrument, including responsiveness, a subcategory of validity[5]. Responsiveness refers to the ability of a measure to detect the clinically important changes in health over time[6]. It is an important measurement property for preference-based indices because health policy implications and reimbursement decisions are impacted by the level at which preference-based indices capture actual changes in health linked to an intervention.

Country specific scoring algorithms to produce preference-based indices are available. In China, there is currently no EQ-5D value set but an SF-6D preference-based index is available and has been validated[7]. The responsiveness of the SF-6D alongside other generic

preference-based measures has been examined in patients with cataract and heart failure[8], stroke[9], psychosomatic disorder[10, 11], rheumatic diseases[12-16], spine problems[17], knee pain[18], and HIV [19]. Two longitudinal studies have compared the responsiveness of HRQOL instruments in patients with CRC[20, 21]. Condition-specific HRQOL instruments were found to be more responsive than generic instruments except for the domains linked to social functioning[20]. However, neither study tested the responsiveness of preference-based measures in CRC.

Related to this patient group, preference-based indices can be estimated by three approaches: 1) direct elicitation techniques for CRC health states using time-trade off or/and standard gamble[22-26]; 2) indirect elicitation though the use of preference-weighted scoring valued by health state classification system such as SF-6D[27, 28] and EQ-5D[29-33]; and 3) indirect elicitation though the use of mapping algorithms that map condition-specific measures (EORTC[34] and FACT-C[35]) onto the preference-based utility scale using regression. This means that preference-based indices can be derived in trials and studies where a generic measure such as EQ-5D or SF-6D has not been used. The responsiveness of SF-6D preference-based index derived from SF-36 and SF-12 appeared similar across seven patient groups[36], providing evidence on the psychometric property of the SF-6D index derived from SF-12 in Western populations. There is limited work comparing the psychometric properties of the indices generated using the different methods. Thus, the responsiveness of the SF-6D index has not been tested in a Chinese population with CRC.

The aims of this study were to examine the internal and external responsiveness of three preference-based SF-6D indices: 1. Derived from asking respondents the directly

complete the SF-6D questionnaire; 2. Derived from the SF-12v2; and 3. Mapped from the FACT-C to the SF-6D based on a self-reported health change anchor. This enabled us to compare the responsiveness of the indices among groups of patients who self-reported both increases and decreases in health status, and those whose health did not change.

Methods

Subject and Study Design

A total of 698 patients attending colorectal specialist outpatient clinics of an academic teaching hospital in Hong Kong, China were contacted from September 2009 to July 2010. The inclusion criteria were aged 18 years or above with a documented diagnosis of colorectal polyp or cancer for at least 6 months. The exclusion criteria were: 1) their life expectancy to be less than 6 months, 2) unable to communicate in Chinese/Cantonese, 3) too ill to carry out an interview, and 4) refused to give consent. Of relevance to the current study, patients with colorectal polyp were excluded. The remaining 386 (55.3%) patients participated in the baseline visit by completing a set of questionnaires including the SF-6D and FACT-C (condition-specific) HRQOL instruments. At 6-month follow-up, 333 (86.3%) subjects also completed the questionnaires by telephone. The interviewers were instructed to go through each item stating from the beginning to the end of the questionnaire and to standardize how each item and its response options should be read out during the interviews at baseline and 6-month follow-up. Subject recruitment and data collection procedures were described in detail

elsewhere[20]. The protocol was approved by the local Ethics Committee (IRB Ref# UW 09-391) and the trial was registered with HK Clinical Trial Register (#HKCTR-973).

Health Anchor

All subjects also answered an one item Global Rating on Change Scale (GRS) to evaluate their subjective changes in global health condition by a retrospective question "Compared to the first visit (six months ago), how would you rate your overall health now?"[37]. The response was rated on a 7-point ordinal scale ranging from -3 to 3 anchored from the much worse to the much better options, with 0 indicating no change. GRS anchors have been commonly used as the external criterion of change on the estimation of minimally clinically importance difference of HRQOL measures[38-40], and applied in longitudinal studies comparing the responsiveness of HRQOL measures in colorectal cancer[20, 21] and responsiveness of preference-based measures[10, 16, 19, 41].

HRQOL Instruments

The Chinese (HK) Short Form-6 Dimensions (SF-6D)

The SF-6D is a generic preference-based measure for the generation of a composite index value on a scale of 0 (death) to 1 (full health). It consists of six dimensions namely

physical functioning (PF), role limitation (RL), social functioning (SF), bodily pain (BP), mental health (MH) and vitality (VT). The SF-6D scoring algorithm has been validated and established to produce direct SF-6D preference-based index, denoted as $SF\text{-}6D_{Direct}$, for the adult Chinese population in Hong Kong[42].

The Chinese (Hong Kong) Short Form-12 version 2 (SF-12v2)

The SF-12v2 Health Survey is a generic HRQOL measure with eight subscales (Physical functioning, PF; Role physical, RP; Bodily pain, BP; General health, GH; Vitality, VT; Social functioning, SF; Role emotional, RE; Mental health, MH), ranging from 0 to 100[36]. This generic instrument is shown to be valid and reliable in Chinese population [43, 44]. Seven out of twelve items from SF-12v2 has been selected to derive the SF-6D preference-based index, denoted as $SF\text{-}6D_{SF\text{-}12,}$ based on a preference-weighted scoring algorithm [42].

The Functional Assessment of Cancer Therapy-Colorectal (FACT-C)

The Functional Assessment of Cancer Therapy-Colorectal (FACT-C) is an extension of the FACT-G HRQOL instrument that emphasizes on a range of important aspects of quality of life specific to patients with CRC [45]. This condition-specific HRQOL instrument has been shown to have acceptable validity in Chinese patients with colorectal neoplasm using classical[46] and modern psychometric methods[47]. Four (Physical, emotional, functional well-being and colorectal cancer subscale) out of five subscales from FACT-C have been used to derive mapped SF-6D preference-based indices, denoted as $SF\text{-}6D_{FACT\text{-}C}$

[35]. Based on pilot data assessing the reproducibility [46], the intra-class correlation between in-person and telephone administration ranged from 0.66 to 0.82 for each subscale of FACT-C, resulting in acceptable reproducibility.

Statistical Analysis

The SF-6D preference-based measures ($SF\text{-}6D_{Direct}$, $SF\text{-}6D_{SF\text{-}12}$ and $SF\text{-}6D_{FACT\text{-}C}$) were derived from either preference-weighted scoring for the SF-6D or a mapping algorithm based on the source of SF-12v2[42] and FACT-C[35] instrument. The descriptive statistics (mean ± standard deviation and median) in $SF\text{-}6D_{Direct}$, $SF\text{-}6D_{SF\text{-}12}$ and $SF\text{-}6D_{FACT\text{-}C}$ were calculated. The ceiling and floor effects were compared and considered to be present if over 15% of subjects reported the minimum or maximum possible score [48]. The presence of a large ceiling effect may impact on responsiveness as the measure will not be able to detect an increase in health for patients reporting the highest possible score. The presence of a floor effect means that the measure may not be sensitive to decreases in health status amongst those reporting the lowest possible score.

Responsiveness was assessed using the self-reported health change anchor to define samples reflecting no change (rating of 0), improvement (rating of 1 to 3) and deterioration (rating of -3 to -1) in health status [49]. Mean and standard deviation of all HRQOL scores were calculated for each sample.

According to a literature review study [50], the assessment of responsiveness was characterized by two major aspects: “internal responsiveness” and “external responsiveness”.

For assessing the internal responsiveness, mean changes in HRQOL score over the past six months in patients with "Deterioration", "Unchanged" and "Improvement" were tested by paired t-test. Mean changes in utility were regarded as clinically important differences when the changes exceeded the absolute magnitude of 0.04, which was the mean minimally important difference of SF-6D[51]. The HRQOL score differences between baseline and follow-up assessments were also tested using the standardized effect size (SES)[37], standardized response mean (SRM)[52] and responsiveness statistics (RS)[6, 41, 53] separately for patients in each group. Three responsiveness statistics were reported because the method for calculating the most appropriate responsiveness statistic was still controversial [50]. The SES, SRM and RS were determined by dividing the observed differences between baseline and follow-up scores by the standard deviation of all subjects at baseline, the standard deviation of observed differences, and the standard deviation of observed differences among "Unchanged" group, respectively. The formula of these statistics is reported in the following:

$$\text{Standardized Effect Size (SES)} = (\mu_{Followup} - \mu_{Baseline})/\sigma_{Baseline}$$

$$\text{Standardized Response Mean (SRM)} = (\mu_{Followup} - \mu_{Baseline})/\sigma_{Followup\text{-}Baseline}$$

$$\text{Responsiveness statistic (RS)} = (\mu_{Followup} - \mu_{Baseline})/\sigma_{(Followup\text{-}Baseline)Unchanged}$$

The value of SES, SRM and RS were interpreted as trivial for <0.2, small for ≥0.2 and <0.5, moderate for ≥0.5 and <0.8 or large for ≥0.8, according to criteria defined by Cohen [54], Liang [52] and Norman [55] respectively. Internal responsiveness was supported if these changes are interpreted as small or above. 95% bootstrap bias-corrected and accelerated

confidence intervals [56] for SES, SRM and RS were obtained using the bootstrapping estimation method with 2000 replications.

For assessing the external responsiveness, the "Worsened" group was compared with "Unchanged" and "Improved" groups. Independent t-tests were performed to compare the HRQOL score in patients among two groups. External responsiveness was determined by not only the score change between groups but also ROC curve analysis [50] that assesses the ability of the instrument to detect HRQOL score change with health condition changes, or to discriminate between groups. The ROC curve is a plot of the true-positive rate (sensitivity) against the false-positive rate (1-specificity). Conceptually, the area under ROC curves (AUC) represents the probability of a random patient with unchanged/improved health status to have a higher HRQOL score than a random patient with worsened health status. Perfect discriminatory power is defined as a value of 1 but a value of 0.5 is considered no discriminatory power. The adequate threshold of AUC was considered as 0.7 [48] and its 95% confidence intervals were reported to assess if the confidence interval contained the adequate threshold of 0.7. Moreover, the Pearson correlation between GRS and score change from baseline to follow-up was presented to assess the external responsiveness. The correlation was considered small for the coefficient between 0.1 and 0.3, moderate for the coefficient between 0.3 and 0.5 and large for coefficient larger than 0.5 [54]. External responsiveness was supported if the 95% confidence interval of AUC contained the adequate threshold of 0.7 and the Pearson correlation was statistically significant.

The preference-based indices tested were regarded as more responsive overall when it was found to be both internally and externally responsive. All statistical analyses were

conducted by the SPSS version 20.0 for Windows (SPSS, IBM Inc., Chicago, Illinois, USA). P-value < 0.05 was considered as statistically significant.

Results

Baseline Characteristics

The baseline characteristics of all respondents and those who did and did not complete follow up are shown in Table 1. The majority of patients were male, married, not currently working, low income, non-smoker and non-drinker at baseline. There was no significant difference in the characteristics of those who did and did not complete follow up, except in the proportion of people on active treatment and distribution of cancer stage. Subjects with advanced cancer stage or palliative treatment were more likely to default than subjects without.

Baseline and 6-month follow-up on the HRQOL scores are shown in Table 2. In baseline and 6-month follow-up, the score of $SF\text{-}6D_{Direct}$, $SF\text{-}6D_{SF\text{-}12}$ and $SF\text{-}6D_{FACT\text{-}C}$ were greater than 0.8. $SF\text{-}6D_{Direct}$ and $SF\text{-}6D_{SF\text{-}12}$ detected statistically significant mean changes between the two time points. There were no floor and ceiling effects for the $SF\text{-}6D_{Direct}$ and $SF\text{-}6D_{FACT\text{-}C}$ at either baseline or 6-month follow-up. Table 3 shows that most subjects reported no change (52.0%) in global health, and 15.1% and 32.9% of patients rated better and worse health conditions respectively.

Internal Responsiveness

A summary of mean change and responsiveness statistics in each health state change group (i.e. worsened, unchanged and improved) is shown in Table 4 and is illustrated in Figure 1. For the group where health worsened, the mean of SF-$6D_{Direct}$ and SF-$6D_{FACT-C}$ between baseline and 6 months follow-up was significantly different and also exceeded the minimum clinically important changes of 0.04. SF-$6D_{Direct}$ and SF-$6D_{FACT-C}$ had moderate sensitivity to change in SES, SRM and RS and illustrated a greater sensitivity to change than SF-$6D_{SF-12}$. For the unchanged group, there was no significant difference in the mean change of SF-$6D_{Direct}$, SF-$6D_{SF-12}$ and SF-$6D_{FACT-C}$ as expected. For the improved group, significant differences were found in the mean of SF-$6D_{SF-12}$ and SF-$6D_{FACT-C}$ between baseline and follow up but the mean changes in SF-$6D_{Direct}$ and SF-$6D_{FACT-C}$ did not exceed the clinically important changes of 0.04. Despite the small changes with SES, SRM and RS in the SF-$6D_{SF-12}$ and SF-$6D_{FACT-C}$, these were more sensitive to change than SF-$6D_{Direct}$.

As far as internal responsiveness was concerned, SF-$6D_{Direct}$ had the best sensitivity to change in the health decrease group, but did not perform as well in the health improvement group. SF-$6D_{SF-12}$ performed well in the health improvement group but unsatisfactorily in the health decrease group. In comparison, SF-$6D_{FACT-C}$ had satisfactory performance in all three groups - deteriorated, unchanged and improved groups.

External responsiveness

Table 5 and Figure 2 show the similar result of mean change and AUC of the ROC between "Worsened" and "unchanged" group as well as that between "Worsened" and "improved" group. Mean differences for $SF\text{-}6D_{Direct}$ and $SF\text{-}6D_{FACT\text{-}C}$ were statistically significant. The correlation coefficients for $SF\text{-}6D_{Direct}$, $SF\text{-}6D_{SF\text{-}12}$ and $SF\text{-}6D_{FACT\text{-}C}$ were small. However, the 95% confidence interval of AUC for $SF\text{-}6D_{Direct}$ and $SF\text{-}6D_{FACT\text{-}C}$ overlapped with the adequate threshold of 0.70 so they could be able to distinguish between patients who had worsened from those who had not.

Discussion

For Chinese patients with CRC, there are a number of methods for estimating SF-6D preference-based indices, including directly administering by the SF-6D instrument[42], derived from the generic SF-12v2 instrument[42], and mapping from the FACT-C[35] instrument This study has found evidence of differences in the level of responsiveness of preference-based indices derived using these different methods. This is the first study, to our knowledge, assessing the comparative responsiveness of preference-based indices using direct and mapped SF-6D preference-based measures in cancer populations using indicators of both "internal" and "external" [50] responsiveness that have been utilized in previous studies[12, 20].

Mapping functions are principally designed to provide preference-based indices for economic evaluation when a generic measure has not been used, and is characterized as an alternative and second-best way of obtaining utility values[4]. It is unclear whether mapped

SF-6D scores are as responsive as generic SF-6D scores. In those whose health was unchanged, the measures did not show change over time. This is as expected, and provides evidence of the reliability of the instruments.

In the group where health got worse, the direct preference-based index was more responsive to change on the basis of three responsiveness statistics than preference-based measures derived from SF-12v2 and FACT-C instruments. Between indices derived from SF-6D and SF-12, possible explanation of the worsened subjects favoring direct preference-based measure were the discrepancy between descriptive systems and tariff set for the convention of directly administered SF-6D and SF-12v2 data. Direct measure required the complete responses of six items whereas the seven out of twelve items in SF-12 were administered to produce the index. According to the scoring algorithm converting raw SF-6D responses to SF-6D preference-based index, several health states have been shown to result in a significantly larger decrement. In cases of the worse possible health states in physical functioning and bodily pain was the strong reflection of reasonably lower valuations of SF-6D scores in worsened group, while direct preference-based index enabled to capture the more severe levels of quality of life compared to preference-based measures from SF-12. In contrary, the direct preference-based measure was not responsive to change in improved group whereas the mapped preference-based measures were significantly responsive to change in improved group. It might be due to the greater relevance of the FACIT instrument to the health condition, leading to the increasingly responsive to change over time. However mapping functions are not without concerns, including the inherent level of error between the actual and predicted utility values [57]. It was expected that mapping algorithm may lead to the slight over-estimation of preference-based measure when the subjects rated at less severe end of preference-based index. As a whole, mapped preference-based indices sufficiently

captured improvement and deterioration in patients with CRC which provides some support for the use of mapped utility values in colorectal neoplasm patients.

Studies comparing the responsiveness of the SF-6D and EQ-5D[11, 13, 14, 16, 18, 19] have concluded that the SF-6D was more responsive than the EQ-5D in the health improvement group but the opposite phenomenon was apparent in the sample whose health worsened. Few studies [12-14] reported that relatively small variance in SF-6D derived from SF-36 provided a conservative estimation in QALYs for the cost-effectiveness analysis of interventions. , Since the direct preference-based index was responsive to change for those whose health deteriorated from healthy to severe health states, the preference-based indices tended to result in a higher QALY gained and lower cost per QALYs gained calculated with preference-based indices administered directly by SF-6D compared with scores derived from other generic or condition-specific instruments. Appraised interventions adopting the direct preference-based index may have a higher likelihood of being cost-effective strategy in theoretical sense. Conversely, appraised interventions undertaken in patients following pathway of disease trajectories starting from treatment to pro-treatment recovery deemed to vary in both positive and negative direction with considerable magnitude.

The limitations of current study should be noted. First, a general self-reported retrospective health anchor was used but it is debatable whether this is the most appropriate anchor. This is because the anchor may not strongly relate to change across the other generic and condition specific measures, and also due to potential recall bias [49]. Taking this into consideration, several previous studies [9, 10] collected multiple independent anchors in terms of clinical endpoint-based, patient-based or physician-based anchors. Further studies

exploring different aspects of anchors can be warranted. Second, there was a high frequency of patients in one category of the anchor with over half of subjects (52.0%) self-reporting "unchanged" health over time.. Thirdly, patients with colorectal polyp were excluded from this analysis although polyp patients were also recruited in our sampling plan. The focus on colorectal cancer only was due to the fact that colorectal cancer and polyp patients differed with respect to their treatment, the severity of disease, and level of HRQOL. Fourth, the differences in administration modes between the two time assessments may contribute to potential differences in HRQOL scores and impact on responsiveness. According to a randomized study on administration modes[58], cancer patients were found to inflate their HRQOL and report more socially favorable desirable responses when the HRQOL instrument was interviewer-administered. Given that the baseline and follow-up assessments were interviewer-administered, both the in-person and telephone administrations may be subject to social desirability bias. Finally, patients were recruited at a single clinical site (Queen Mary Hospital) rather than across multiple centers, and this limits the generalizability of the results to the wider Chinese population.

Conclusion

The direct SF-6D measure was more responsive than mapped preference-based measures in worsened group but the direct SF-6D was not responsive in improved group. The mapped SF-6D had satisfactory performance in all three health change groups. The use of a preference-based index mapped from a condition-specific measure captures both negative and

positive change in HRQOL among CRC patients. Furthermore, the external responsiveness from the mapped SF-6D measure was greater compared with the direct SF-6D measure.

(Words Count: 3425)

Acknowledgements

This work was supported by Small Project Funding (grant number 200907176135) from the committee on research and conference grants of The University of Hong Kong, and Health and Health Services Research Fund (project number 08090851) from the Food and Health Bureau of the Hong Kong SAR. The authors would like to thank Mandy Tai, Joyce Sing, Winnie Chan, Eileen Yeung, Wincy Wong and Deki Pun for data collection.

Abbreviations: HRQOL, Health-related Quality of Life; CRC, Colorectal Cancer; QALY, Quality-adjusted life year; FACT-C, Functional Assessment of Cancer Therapy-Colorectal; EORTC, European Organization for Research and Treatment of Cancer; PWB, Physical well-being; SWB, Social well-being; EWB, Emotional well-being; FWB, Functional well-being; CCS, Colorectal cancer subscale; TOI, Trial Outcome Index; PF, Physical functioning; RP, Role physical; BP, Bodily pain; GH, General health; VT, Vitality; SF, Social functioning; RE, Role emotional; MH, Mental health; PCS, Physical composite score; MCS, Mental composite score; SES, Standardized effect size; SRM, Standardized response mean ; RS, Responsiveness statistic; GRS, Global Rating on Change Scale

Reference

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. International Journal of Cancer. 2010;127(12):2893-917.
2. Yabroff KR, Mariotto AB, Feuer E, Brown ML. Projections of the costs associated with colorectal cancer care in the United States, 2000–2020. Health Economics. 2008;17(8):947-59.
3. National Institute for Clinical Excellence. Guide to the Methods of Technology Appraisal (reference N1618). London: NICE; 2008.
4. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. The European Journal of Health Economics. 2010;11(2):215-25.
5. Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. Quality of Life Research. 1992;1(1):73-5.
6. Guyatt G, Walter S, Norman G. Measuring change over time: Assessing the usefulness of evaluative instruments. Journal of Chronic Diseases. 1987;40(2):171-8.
7. Lam CLK, Brazier J, McGhee SM. Valuation of the SF-6D Health States Is Feasible, Acceptable, Reliable, and Valid in a Chinese Population. Value in Health. 2008;11(2):295-303.
8. Kaplan RM, Tally S, Hays RD, Feeny D, Ganiats TG, Palta M, et al. Five preference-based indexes in cataract and heart failure patients were not equally responsive to change. Journal of Clinical Epidemiology. 2011;64(5):497-506.
9. Pickard AS, Johnson JA, Feeny DH. Responsiveness of generic health-related quality of life measures in stroke. Quality of Life Research. 2005;14(1):207-19.
10. Gerhards SA, Huibers MJ, Theunissen KA, de Graaf LE, Widdershoven GA, Evers SM. The responsiveness of quality of life utilities to change in depression: a comparison of instruments (SF-6D, EQ-5D, and DFD). Value in Health. 2011;14(5):732-9.
11. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. Quality of Life Research. 2008;17(3):485-95.
12. Buitinga L, Braakman-Jansen LMA, Taal E, Kievit W, Visser H, van Riel PLCM, et al. Comparative responsiveness of the EuroQol-5D and Short Form 6D to improvement in patients with rheumatoid arthritis treated with tumor necrosis factor blockers: Results of the Dutch Rheumatoid Arthritis Monitoring registry. Arthritis Care & Research. 2012;64(6):826-32.
13. Gaujoux-Viala C, Rat AC, Guillemin F, Flipo RM, Fardellone P, Bourgeois P, et al. Responsiveness of EQ-5D and SF-6D in patients with early arthritis: results from the ESPOIR cohort. Annals of the Rheumatic Diseases. 2012;71(9):1478-83.
14. Harrison MJ, Davies LM, Bansback NJ, McCoy MJ, Verstappen SM, Watson K, et al. The comparative responsiveness of the EQ-5D and SF-6D to change in patients with inflammatory arthritis. Quality of life Research. 2009;18(9):1195-205.
15. Russell AS, Conner-Spady B, Mintz A, Maksymowych WP. The responsiveness of generic health status measures as assessed in patients with rheumatoid arthritis receiving infliximab. J Rheumatol. 2003;30(5):941-7.

16. Marra CA, Rashidi AA, Guh D, Kopec JA, Abrahamowicz M, Esdaile JM, et al. Are indirect utility measures reliable and responsive in rheumatoid arthritis patients? Quality of Life Research. 2005;14(5):1333-44.
17. McDonough CM, Tosteson TD, Tosteson AN, Jette AM, Grove MR, Weinstein JN. A longitudinal comparison of 5 preference-weighted health state classification systems in persons with intervertebral disk herniation. Medical Decision Making. 2011;31(2):270-80.
18. Barton G, Sach T, Avery A, Doherty M, Jenkinson C, Muir K. Comparing the performance of the EQ-5D and SF-6D when measuring the benefits of alleviating knee pain. Cost Effectiveness and Resource Allocation. 2009;7(1):12.
19. Stavem K, Frøland SS, Hellum KB. Comparison of preference-based utilities of the 15D, EQ-5D and SF-6D in patients with HIV/AIDS. Quality of Life Research. 2005;14(4):971-80.
20. Wong CKH, Lam CLK, Law WL, Poon JTC, Kwong DLW, Tsang J, et al. Condition-specific Measure was more responsive than Generic Measure in Colorectal Cancer: All but Social Domains. Journal of Clinical Epidemiology. 2013;66(5):557-65.
21. Uwer L, Rotonda C, Guillemin F, Miny J, Kaminsky MC, Mercier M, et al. Responsiveness of EORTC QLQ-C30, QLQ-CR38 and FACT-C quality of life questionnaires in patients with colorectal cancer. Health and Quality of Life Outcomes. 2011;9:70.
22. Best JH, Garrison LP, Hollingworth W, Ramsey SD, Veenstra DL. Preference values associated with stage III colon cancer and adjuvant chemotherapy. Quality of Life Research. 2010;19(3):391-400.
23. Shiroiwa T, Fukuda T, Tsutani K. Health utility scores of colorectal cancer based on societal preference in Japan. Quality of Life Research. 2009;18(8):1095-103.
24. Dominitz JA, Provenzale D. Patient Preferences and Quality of Life Associated with Colorectal Cancer Screening. American Journal of Gastroenterology. 1997;92(12):2171-8.
25. Dobrez D, Cella D, Pickard AS, Lai JS, Nickolov A. Estimation of Patient Preference-Based Utility Weights from the Functional Assessment of Cancer Therapy - General. Value in Health. 2007;10(4):266-72.
26. Ness RM, Holmes AM, Klein R, Dittus R. Utility valuations for outcome states of colorectal cancer. American Journal of Gastroenterology. 1999;94(6):1650-7.
27. Hornbrook MC, Wendel CS, Coons SJ, Grant M, Herrinton LJ, Mohler MJ, et al. Complications Among Colorectal Cancer Survivors: SF-6D Preference-Weighted Quality of Life Scores. Medical Care. 2011;49(3):321-6.
28. Wong CKH, Lam CLK, Poon JTC, Kwong DLW. Clinical Correlates of Health Preference and Generic Health-related Quality of Life in Patients with Colorectal Neoplasms. PLoS One. 2013;8(3):e58341.
29. Odom D, Barber B, Bennett L, Peeters M, Zhao Z, Kaye J, et al. Health-related quality of life and colorectal cancer-specific symptoms in patients with chemotherapy-refractory metastatic disease treated with panitumumab. International Journal of Colorectal Disease. 2011;26(2):173-81.
30. Bennett L, Zhao Z, Barber B, Zhou X, Peeters M, Zhang J, et al. Health-related quality of life in patients with metastatic colorectal cancer treated with panitumumab in first- or second-line treatment. British Journal of Cancer. 2011;105(10):1495-502.
31. Kim SH, Hwang JS, Kim TW, Hong YS, Jo MW. Validity and reliability of the EQ-5D for cancer patients in Korea. Supportive Care in Cancer. 2012;20(12):3155-60.
32. Färkkilä N, Sintonen H, Saarto T, Järvinen H, Hänninen J, Taari K, et al. Health-related quality of life in colorectal cancer. Colorectal Disease. 2013;15(5):e215–e22.
33. Hamashima C. Long-term quality of life of postoperative rectal cancer patients. Journal of Gastroenterology & Hepatology. 2002;17(5):571-6.

34. Wong CKH, Lam CLK, Wan YF, Rowen D. Predicting SF-6D from the European Organization for Treatment and Research of Cancer Quality of Life Questionnaire Scores in Patients with Colorectal Cancer. Value in Health. 2013;16(2):373-84.
35. Wong CKH, Lam CLK, Rowen D, McGhee SM, Ma KP, Law WL, et al. Mapping the Functional Assessment of Cancer Therapy-General or -Colorectal to SF-6D in Chinese Patients with Colorectal Neoplasm. Value in Health. 2012;15(3):495-503.
36. Brazier JE, Roberts J. The Estimation of a Preference-Based Measure of Health From the SF-12. Medical Care. 2004;42(9):851-9.
37. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. Journal of Clinical Epidemiology. 1997;50(1):79-93.
38. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407-15.
39. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of- life scores. Journal of Clinical Oncology. 1998;16(1):139-44.
40. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. Journal of Clinical Epidemiology. 1994;47(1):81-7.
41. Wright JG, Young NL. A Comparison of Different Indices of Responsiveness. Journal of Clinical Epidemiology. 1997;50(3):239-46.
42. McGhee SM, Brazier J, Lam CLK, Wong LC, Chau J, Cheung A, et al. Quality-adjusted life years: population-specific measurement of the quality component. Hong Kong Med J. 2011;17(Suppl 6):17-21.
43. Fong DYT, Lam CLK, Mak KK, Lo WS, Lai YK, Ho SY, et al. The Short Form-12 Health Survey was a valid instrument in Chinese adolescents. Journal of Clinical Epidemiology. 2010;63(9):1020-9.
44. Lam CLK, Tse EYY, Gandek B. Is the standard SF-12 Health Survey valid and equivalent for a Chinese population? Quality of Life Research. 2005;14(2):539-47.
45. Ward WL, Hahn EA, Mo F, Hernandez L, Tulsky DS, Cella D. Reliability and validity of the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) quality of life instrument. Quality of Life Research. 1999;8(3):181-95.
46. Wong CKH, Lam CLK, Law WL, Poon JTC, Chan P, Kwong DLW, et al. Validity and Reliability Study on Traditional Chinese FACT-C in Chinese Patients with Colorectal Neoplasm. Journal of Evaluation in Clinical Practice. 2012;18(6):1186-95.
47. Wong CKH, Lam CLK, Mulhern B, Law WL, Poon JTC, Kwong DLW, et al. Measurement Invariance of the Functional Assessment of Cancer Therapy-Colorectal Quality-of-life Instrument among Modes of Administration. Quality of Life Research. 2013;22(6):1415-26.
48. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology. 2007;60(1):34-42.
49. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. Journal of Clinical Epidemiology. 2008;61(2):102-9.
50. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. Journal of Clinical Epidemiology. 2000;53(5):459-68.
51. Walters S, Brazier J. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Quality of Life Research. 2005;14:1523-32.

52. Liang MH, Fossel AH, Larson MG. Comparisons of Five Health Status Instruments for Orthopedic Evaluation. Medical Care. 1990;28(7):632-42.
53. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PMM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Quality of Life Research. 2003;12(4):349-62.
54. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
55. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. Journal of Clinical Epidemiology. 1997;50(8):869-79.
56. Efron B. Better Bootstrap Confidence Intervals. Journal of the American Statistical Association. 1987;82(397):171-85.
57. Longworth L, Rowen D. NICE DSU Technical Support Document 10: The use of mapping methods to estimate health state utility values. 2011 [updated April 2011February 2013]; Available from: http://www.nicedsu.org.uk/TSD%2010%20mapping%20FINAL.pdf.
58. Buskirk TD, Stein KD. Telephone vs. mail survey gives different SF-36 quality-of-life scores among cancer survivors. Journal of Clinical Epidemiology. 2008;61(10):1049-55.