

## **SF-6D and EQ-5D result in widely divergent incremental cost effectiveness ratios in a clinical trial of older women: Implications for health policy decisions**

**Jennifer C Davis, PhD,**

Centre for Hip Health & Mobility, University of British Columbia & Vancouver Coastal Health Research Institute (VCHRI), 301-2647 Willow Street, Vancouver, British Columbia, Canada, V5Z 3P1

**Dr Teresa Liu-Ambrose [assistant professor],**

Centre for Hip Health & Mobility, University of British Columbia & Vancouver Coastal Health Research Institute (VCHRI), 301-2647 Willow Street, Vancouver, British Columbia, Canada, V5Z 3P1

**Associate Professor M Clare Robertson [research associate professor],**

Department of Medical & Surgical Sciences, Dunedin School of Medicine, University of Otago, P.O. Box 913, Dunedin, New Zealand, 9054

**Dr Karim M Khan [professor], and**

Centre for Hip Health & Mobility, University of British Columbia & Vancouver Coastal Health Research Institute (VCHRI), 301-2647 Willow Street, Vancouver, British Columbia, Canada, V5Z 3P1

**Associate Professor Carlo A Marra [director]**

Canada Research Chair in Pharmaceutical Outcomes, Collaboration for Outcomes Research and Evaluation, St Paul's Hospital, 620B 1081 Burrard Street, University of British Columbia, Vancouver, British Columbia, Canada, V6Z 1Y6

### **Abstract**

**Objectives**—Two major challenges facing decision makers within our health care system is the choice between alternative and new interventions for the same clinical indication or diagnosis. To guide such choices, economic evaluations that provide evidence of the costs and effects are increasingly prevalent. Very few studies have compared impact of different preference based utility instruments and the resultant incremental cost-effectiveness ratio on decision-making in population specific settings. We compared the incremental cost-effectiveness ratios, in a clinical trial of older women, estimated using the EuroQol-5D (EQ-5D) and the Short Form-6D (SF-6D) to discuss the implications on decision making.

**Methods**—Using both the EQ-5D and the SF-6D, we compared the incremental cost per quality-adjusted life year (QALY) gained in a clinical trial in older women. The trial consisted of two

parts, a 12-month randomized controlled trial with a 12-month following was conducted among 155 community-dwelling women aged 65 to 75 years. For the followup study, 123 of the 155 who originally were randomly allocated to once-weekly resistance training (n=54), twice-weekly resistance training (n=52), or to twice-weekly balance and tone exercises (i.e., control group) (n=49) participated in the 12-month follow-up study. For the intervention period, 155 individuals are included in the economic evaluation. For the followup study, 98 took part in the economic evaluation (twice-weekly balance and tone exercises, n=28, once-weekly resistance training, n=35; twice-weekly resistance training, n=35). Our primary outcome measure was incremental cost per quality adjusted life year (QALY) gained. Healthcare resource utilization was assessed over 21 months (2009 prices); health status was assessed using the EQ-5D and SF-6D to calculate QALYs using a 21 month time horizon.

**Results**—At cessation of the 12-month intervention, the incremental QALYs were 0.084 (EQ-5D) and 0.003 (SF-6D) for the once and 0.179 (EQ-5D) and 0.003 (SF-6D) twice weekly resistance training groups, compared with the twice weekly balance and tone classes. At cessation of the 12-month followup, the incremental QALY was -0.051 (EQ-5D) and -0.144 (SF-6D) for the once-weekly resistance training group and -0.081 (EQ-5D) and -0.127 (SF-6D) for the twice-weekly resistance training group compared with balance and tone classes

**Conclusions**—The choice of EQ-5D or SF-6D matters, when used to estimate QALYs for the incremental cost effectiveness ratio to inform decisions. We found the incremental QALYs estimated from the EQ-5D were threefold greater than those estimated from the SF-6D. Hence, the choice of preference based utility instrument holds the potential to substantially impact the decision to fund a new health care technology.

## Keywords

health related quality of life; older adults; cost utility analysis; incremental cost-effectiveness ratio

---

## Introduction

A primary goal of the health care system is to maximize health benefit when resources are scarce. Two major challenges facing decision makers are the choice between already existing and new interventions for the same clinical manifestation (1, 2). To guide health policy decision, economic evaluations that provide the costs and effects are increasingly prevalent (3, 4). A widely accepted strength of cost-utility analyses is that it provides a common metric with which to compare existing and new health care interventions with standard of care. QALYs are meant to permit comparison across conditions and populations. Despite QALYs providing a ‘common metric’, the methodology and valuations used to estimate QALYs differ (5, 6). Despite the number of studies comparing utility score, few evaluate the impact of different preference based utility instruments and the incremental cost-effectiveness ratio on decision-making (7). Of these few, researchers have demonstrated that although the ICERs differ based on QALY estimate method used, the implications on decision making are minimal (2, 7, 8).

However, this broad spanning claim needs to be established at a condition specific level. Most recently, investigators have reviewed the relative merits of the various methods used to

elicit health state utility values for estimating QALYs. Two widely used indirect techniques are two questionnaires: Euro-QoL5D (EQ-5D) (9, 10) and the Short Form 6D (SF-6D) (11, 12). In a key paper, Joore and colleagues (13) found that that using the EQ-5D in patient groups with mild health condition, the likelihood of accepting the incremental cost-utility ratio was substantially greater. In contrast, using the SF-6D in patient groups with worse health, the likelihood of accepting the incremental cost-utility ratio was substantially greater. These findings could pose an interesting debate because the goal is not to use an instrument that will result in increased probability of funding a program/intervention. Hence, further research is needed to determine the construct validity of these instruments in population specific settings as Marra and colleagues have done in rheumatoid arthritis (14). Applying Joore's findings, we would expect that a cost utility analysis of relatively healthy older women would be expected to be more likely favorable if the EQ-5D was used to estimate utility.

There are many challenging decisions needed to be made to contribute to healthy aging and mobility in aging so we used both the EQ-5D and the SF-6D in a cost-utility analysis. The objective of our study was to compare the incremental cost-effectiveness ratios estimated from QALYs calculated using the EuroQoL-5D (EQ-5D) and the Short Form-6D (SF-6D) and the resultant implications on decision making. We designed a concurrent, prospective economic analysis using individual level data on cost and effectiveness outcomes as part of the Brain Power study, a three arm randomized controlled trial (15). The main outcomes of the clinical trial and the economic evaluations are reported elsewhere (16, 17). Hence, our purpose was to add a novel setting to the few studies that have made these condition specific comparisons. In turn this will provide a basis for recommending condition specific methods for QALY estimation that may be more appropriate for decision makers.

## Methods

### Sample

We previously reported details of the Brain Power randomized controlled trial and followup study (15, 16). Briefly, the study sample included 155 community dwelling women aged 65 to 75 years. Participants enrolled in Brain Power had a Mini Mental State Examination (MMSE) score  $\geq 24$  (i.e., were cognitively intact) and visual acuity 20/40 or better with or without corrective lenses. Participants excluded: were unable to write and speak English, were partaking in resistance training in the last six months, had a current medical condition for which exercise is contraindicated, had a neurodegenerative disease, were taking cholinesterase inhibitors, being treated currently for depression or on hormone replacement therapy during the previous 12 months. The interventions for the Brain Power study included three participant groups: once weekly resistance training, twice weekly resistance training and the control group, twice weekly balance and tone classes (comparator). The resistance training program used a progressive, high intensity protocol. The balance and tone program consisted of stretching exercises, range-of-motion exercises, basic core-strength exercises including kegels (pelvic floor exercises), balance exercises, and relaxation techniques. All classes were 60 minutes long, with a 10-minute warm-up, 40 minutes of core content, and a 10-minute cool down period.

The 12-month followup study sample included 135 community dwelling women aged 65 to 75 years who completed the 12-month intervention. Of these, 89 participants completed the EQ-5D at all three time points and 127 participants completed the SF-36 at all three time points. We use these two complete case sets for all data analyses.

## Instruments

**Health care costs – Health resource utilization questionnaire**—We have previously described our collection of health care costs for the Brain Power economic evaluation and followup study (16, 18). Briefly, the health resource utilization questionnaire asked participants to report the following visits over a specified time period: 1) health care professionals, 2) admissions or visits to hospital and 3) laboratory work. The health resource questionnaire is described and supported in previous studies (19, 20). In total, we collected nine months of data from the Brain Power study on associated health care resource use using a 9-month time horizon and a 21-month time horizon for the followup study. Participants recalled their health care resource use every three months during the 12-month followup study. A detailed list of cost items collected was previously reported (18). We estimated total health care related costs over the nine and 21 months from a Canadian health care system perspective and presented all costs in 2008 Canadian dollars. We costed all items using the BC Ministry of Health Fee Payment Schedule and the Vancouver General Hospital Fully Allocated Cost Model developed by Marra and colleagues.

**EuroQol-5D**—The EuroQol EQ-5D (EQ-5D) is a generic preference based utility instrument developed by the EuroQol Group and is one of the most commonly used instruments (9, 10). This five item questionnaire includes the following domains: mobility, self-care, usually activity, pain/discomfort, and anxiety/depression with each attribute having three possible options: 1) no problems, 2) some problems and 3) major problems. These options combine for a total possible 243 health states identified. Individuals preferences for the scoring of the EQ-5D were measured using the TTO technique on a random sample adult population living in the UK (n=3000) (21). Zero is defined as a health state equivalent to death, 1.0 is defined as a state of “full health” and less than zero is a health state worse than death. Health states less than zero are possible for the EQ-5D (range: -0.594 to 1.00) (21). Participants rated their health on the ‘day’ the questionnaire was administered.

**Short Form-6D**—The SF-6D is also a generic preference based utility instrument that is based on a widely used health related quality of life questionnaire, the Short Form 36.(11) The SF-36 can be used to calculate a utility score for the SF-6D. This six-item questionnaire contained six domains that include: physical functioning, role limitations, social functioning, pain, mental health and vitality. Each attribute contains four to six levels that account for the 18 000 unique health states captured by the SF-6D. Unlike the EQ-5D, the scoring model for the SF-6D is based on the standard gamble utility measurements. A random sample (n=836) of a general adult population from the UK was used to estimate the utilities for 249 different health states. Each participant was required to provide utilities for six states. The range in possible health states for the SF-6D is narrower than the EQ-5D at 0.30 to 1.00. Participants rated their health status using a 4-week recall from the day the survey was administered.

**Incremental cost-effective ratio**—The main outcome for our economic evaluation was the incremental cost per QALY gained. QALYs are defined as the benefit of a health intervention in terms of time in a series of quality-weighted health states, in which the quality weights reflect the desirability of living in the particular health state, typically anchored at “perfect” health (weighted 1.0) to dead (weighted 0.0) (22). The quality weights spent in each state are multiplied by the time spent in each state. The total number of QALYs for that time period is the sum of all these products.

### Data analysis

We analysed all data using STATA version 10.0. We report descriptive data as mean (standard deviation) and/or median (interquartile range) for all baseline characteristics and primary outcomes measures. To compare the EQ-5D and the SF-6D, we used both an available and a complete case analysis approach (23, 24). For the complete case analysis, participants who had QALY estimates for both measures were included in all analyses. We also calculated summary statistics including the mean change in cost and mean QALY gain for the once weekly and twice weekly resistance training groups compared with the balance and tone comparator.

## Results

### Sample

Characteristics of the 155 participants who were randomized at baseline, the 135 who completed the 12-month intervention study, and the 123 who consented to the 12-month follow-up study have been reported previously (16, 18). There were no differences in the baseline characteristics for the 123 participants who took part in the follow-up study compared with the 32 of the original 155 participants who declined to participate in the follow-up study. Further, of the 123 who consented to participate in the follow-up study, 98 (80%) participants had complete data for health care resource utilization and were included in the economic evaluation.

### Incremental QALYs and cost-effectiveness ratios using the EQ-5D for the Brain Power intervention and followup study

After controlling for baseline EQ-5D levels, the incremental QALY after 12 months calculated using the EQ-5D was  $-0.051$  for the once-weekly resistance training group and  $-0.081$  for the twice-weekly resistance training group compared with balance and tone classes (Table 1).

Based on the point estimates for total healthcare resource use and QALYs calculated from the EQ-5D for our base case analysis, we found that the incremental cost-utility ratio for once-weekly resistance training per QALY gained was less costly and equally effective compared with BAT. The incremental cost-utility ratio for twice-weekly resistance training per QALY gained was less costly and less effective compared with BAT. The twice weekly resistance training group showed a higher incremental QALY loss.

### Incremental QALYs and cost-effectiveness ratios using the SF-6D for the Brain Power intervention and followup study

After controlling for baseline SF-6D levels, the incremental QALY after 12 months calculated using the SF-6D was  $-0.144$  for the once-weekly resistance training group and  $-0.127$  for the twice-weekly resistance training group compared with balance and tone classes (Table 1).

Based on the point estimates for total healthcare resource use and QALYs calculated from the SF-6D for our base case analysis, we found that the incremental cost-utility ratio for once- or twice-weekly resistance training per QALY gained was less costly and less effective compared with BAT. In contrast to the QALYs estimated from the EQ-5D, the once weekly resistance training group showed a higher incremental QALYs loss.

## Discussion

This indirect approach of preference elicitation enables researchers to use societal health state utility values for a minimal cost compared with direct elicitation techniques (25). Generic preference based utility instruments are often an essential component of clinical research because they offer a feasible method to assess health status, specifically to QALYs – a measure of health gains or losses for economic evaluations. Each of these preference-based utility measures provides weightings for QALYs. The SF-6D describes 18,000 discrete health states and will likely capture small changes in health status (11) compared to the EQ-5D that captures the fewest (i.e., 243 health states), but has a broader range of possible health state utility values (HSUVs) (14). However, the EQ-5D and the SF-6D are based on different domains (i.e., health state descriptions) and valuation techniques. Resultantly, a number of studies have demonstrated that these two instruments produced different utility scores (26–28).

Falls are the leading cause of chronic disability (29, 30) and chronic disability is associated with a decline in health related quality of life. Individuals who experience the greatest declines in health related quality of life have the greatest mortality.(31) To our knowledge, this study is the first to 1) examine the incremental cost effectiveness ratios generated from two generic preference based utility instruments, 2) provide a descriptive comparison of two generic preference based utility instruments in terms of their estimated QALYs measured prospectively over one year among high functioning community-dwelling senior women and 3) to ascertain significant and independent predictors of QALYs and health resource utilization.

Of particular interest, this independent associations for health resource utilization, physiological profile assessment and Timed Up and Go on QALYs estimated from the EQ-5D or SF-6D were found in this cohort of senior women after accounting for age, group and functional comorbidities. One notable discrepancy in thesis findings was that for QALYs estimated from the SF-6D, number of falls in the past 3 months was also a significant and independent predictor. The QALYs estimated from the EQ-5D did not demonstrate this finding. One potential reason for this is that the EQ-5D and the SF-6D are comprised of different domains. Specifically, the SF-6D has a physical function and vitality

domain whereas the EQ-5D has a mobility domain. Therefore, it is possible that individuals who sustained falls in the past three months may report deficits in the physical function or vitality domain that would not be picked up by the EQ-5D. Specific to older adults, few peer-reviewed studies have compared the EQ-5D and SF-6D. One recent study found that the EQ-5D and SF-6D score were highly correlated in adults aged 45 years and older.(32) Further, individuals who were healthier, had higher mean scores on the EQ-5D ( $p<0.001$ ), whereas individuals who were less healthy had higher mean scores on the SF-6D ( $p<0.001$ ). (32) This is not surprising given that the range of possible values from the SF-6D is 0.30–1.00 compared with the EQ-5D's range of –0.54–1.00.(33) Thus our study builds upon previous findings emphasizing that although there are strong similarities, the differences need to be explored further.

To our knowledge, no other study has examined the relationship between health related quality of life and health resource utilization specifically among older adults. Our novel finding that health related quality of life is inversely, independently and significantly associated with health resource utilization does build upon previous findings in other fields. (34)

Our findings should be interpreted with caution due to the following study limitations. Our study sample consisted only of community dwelling women who were cognitively intact. Therefore, we cannot say with certainty these findings are generalizable to all older women or older adults in general. Further, the primary outcome of the Brain POWER study was to ascertain a change in cognitive performance. For this reason, we limited our comparisons of the EQ-5D and the SF-6D to a descriptive analysis. Thus, we recommend this study be used to highlight ideas for future prospective studies to ascertain whether 1) our present finding applied to other clinical populations, 2) the direction of the causal relationship between health related quality of life and health resource utilization and 3) subgroup analysis for the SF-6D and EQ-5D based on levels of mobility in older adults. Our findings indicate that EQ-5D or SF-6D HSUVs over time can be largely explained by baseline measures of age, weight, functional comorbidities index, health resource utilization, Physiological Profile Assessment, number of falls in the past 3 months and Timed Up and Go.

## Conclusion

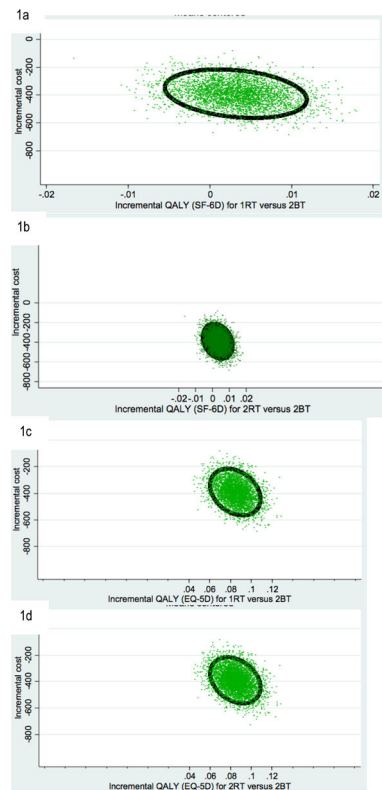
## References

1. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). *Lancet*. 2002; 360:711–5. [PubMed: 12241891]
2. Fenwick E, Marshall DA, Blackhouse G, et al. Assessing the impact of censoring of costs and effects on health-care decision-making: an example using the Atrial Fibrillation Follow-up Investigation of Rhythm Management (AFFIRM) study. *Value Health*. 2008; 11:365–75. [PubMed: 17854433]
3. Neumann PJ, Greenberg D, Olchanski NV, et al. Growth and quality of the cost-utility literature, 1976–2001. *Value Health*. 2005; 8:3–9. [PubMed: 15841889]
4. Siegel JE, Weinstein MC, Russell LB, et al. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. *JAMA*. 1996; 276:1339–41. [PubMed: 8861994]

5. Brazier J, Roberts J, Tsuchiya A, et al. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ.* 2004; 13:873–84. [PubMed: 15362179]
6. Tsuchiya A, Brazier J, Roberts J. Comparison of valuation methods used to generate the EQ-5D and the SF-6D value sets. *J Health Econ.* 2006; 25:334–46. [PubMed: 16271783]
7. McDonough CM, Tosteson AN. Measuring preferences for cost-utility analysis: how choice of method may influence decision-making. *Pharmacoeconomics.* 2007; 25:93–106. [PubMed: 17249853]
8. Sach TH, Barton GR, Jenkinson C, et al. Comparing cost-utility estimates: does the choice of EQ-5D or SF-6D matter? *Med Care.* 2009; 47:889–94. [PubMed: 19584759]
9. Brooks R. EuroQol: the current state of play. *Health Policy.* 1996; 37:53–72. [PubMed: 10158943]
10. Dolan P, Roberts J. Modelling valuations for Eq-5d health states: an alternative model using differences in valuations. *Med Care.* 2002; 40:442–6. [PubMed: 11961478]
11. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ.* 2002; 21:271–92. [PubMed: 11939242]
12. Brazier J, Usherwood T, Harper R, et al. Deriving a preference-based single index from the UK SF-36 Health Survey. *J Clin Epidemiol.* 1998; 51:1115–28. [PubMed: 9817129]
13. Joore M, Brunenberg D, Nelemans P, et al. The impact of differences in EQ-5D and SF-6D utility scores on the acceptability of cost-utility ratios: results across five trial-based cost-utility studies. *Value Health.* 2011; 13:222–9.
14. Marra CA, Woolcott JC, Kopec JA, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc Sci Med.* 2005; 60:1571–82. [PubMed: 15652688]
15. Liu-Ambrose T, Nagamatsu LS, Graf P, et al. Resistance training and executive functions: a 12-month randomized controlled trial. *Arch Intern Med.* 2010; 170:170–8. [PubMed: 20101012]
16. Davis JC, Marra CA, Beattie BL, et al. Sustained cognitive and economic benefits of resistance training among community-dwelling senior women: a 1-year follow-up study of the Brain Power study. *Arch Intern Med.* 170:2036–8.
17. Davis JC, Marra CA, Robertson MC, et al. Economic evaluation of dose-response resistance training in older women: a cost-effectiveness and cost-utility analysis. *Osteoporos Int.*
18. Davis JC, Marra CA, Robertson MC, et al. Economic evaluation of dose-response resistance training in older women: a cost-effectiveness and cost-utility analysis. *Osteoporos Int.* 2010
19. Liu-Ambrose T, Ashe MC, Marra C, et al. Among Older Adults with Multiple Chronic Conditions, Physical Activity is Independently and Inversely Associated with Health Care Utilization. *Br J Sports Med.* 2008
20. Maetzel A, Li LC, Pencharz J, et al. The economic burden associated with osteoarthritis, rheumatoid arthritis, and hypertension: a comparative study. *Ann Rheum Dis.* 2004; 63:395–401. [PubMed: 15020333]
21. Dolan P. Modeling valuations for EuroQol health states. *Med Care.* 1997; 35:1095–108. [PubMed: 9366889]
22. Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annu Rev Public Health.* 2000; 21:587–611. [PubMed: 10884966]
23. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ.* 2005; 14:487–96. [PubMed: 15497198]
24. Manca A, Palmer S. Handling missing data in patient-level cost-effectiveness analysis alongside randomised clinical trials. *Appl Health Econ Health Policy.* 2005; 4:65–75. [PubMed: 16162026]
25. Drummond, MF., Sculpher, MJ., Torrance, GW., et al. *Methods for the economic evaluation for health care programmes.* 3. New York. United States of America: Oxford University Press; 2005.
26. Conner-Spady B, Suarez-Almazor ME. Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Med Care.* 2003; 41:791–801. [PubMed: 12835603]
27. Marra CA, Esdaile JM, Guh D, et al. A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. *Med Care.* 2004; 42:1125–31. [PubMed: 15586840]

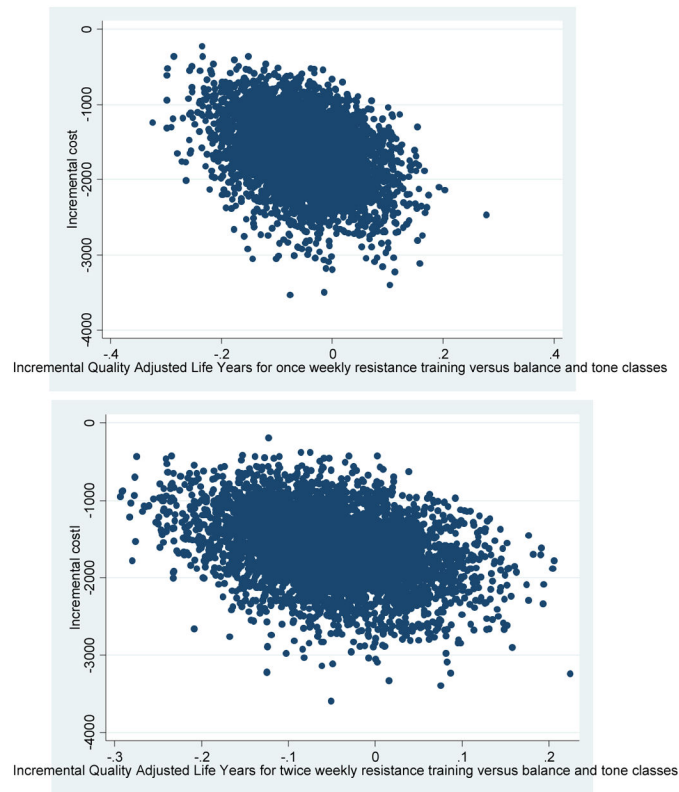


28. Pickard AS, Johnson JA, Feeny DH. Responsiveness of generic health-related quality of life measures in stroke. *Qual Life Res.* 2005; 14:207–19. [PubMed: 15789955]
29. Murray, CWS., Lopez, A. Global and regional descriptive epidemiology of disability: Incidence, prevalence, health expectancies, and year lived with disability. Boston: The Harvard School of Public Health; 1996.
30. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. *N Engl J Med.* 1988; 319:1701–7. [PubMed: 3205267]
31. Kroenke CH, Kubzansky LD, Adler N, et al. Prospective change in health-related quality of life and subsequent mortality among middle-aged and older women. *Am J Public Health.* 2008; 98:2085–91. [PubMed: 18511734]
32. Barton GR, Sach TH, Avery AJ, et al. A comparison of the performance of the EQ-5D and SF-6D for individuals aged  $\geq 45$  years. *Health Econ.* 2008; 17:815–32. [PubMed: 17893863]
33. Bharmal M, Thomas J 3rd. Comparing the EQ-5D and the SF-6D descriptive systems to assess their ceiling effects in the US general population. *Value Health.* 2006; 9:262–71. [PubMed: 16903996]
34. Pare P, Gray J, Lam S, et al. Health-related quality of life, work productivity, and health care resource utilization of subjects with irritable bowel syndrome: baseline results from LOGIC (Longitudinal Outcomes Study of Gastrointestinal Symptoms in Canada), a naturalistic study. *Clin Ther.* 2006; 28:1726–35. discussion 10-1. [PubMed: 17157129]



**Figure 1a.**

**1a)** Cost effective plane depicting the 95% confidence ellipses of incremental cost and effectiveness for comparison between twice weekly resistance training (2RT) and twice weekly balance and tone (2BT, comparator); **1b)** Cost effective plane depicting the 95% confidence ellipses of incremental cost and effectiveness for comparison between once weekly resistance training (1RT) and twice weekly balance and tone (2BT, comparator). QALY estimates are based on utility scores from the EQ-5D. QALY estimates are based on utility score from the SF-6D for figures 1a and 1b. **1c)** Cost effective plane depicting the 95% confidence ellipses of incremental cost and effectiveness for comparison between twice weekly resistance training (2RT) and twice weekly balance and tone (2BT, comparator); **1d)** Cost effective plane depicting the 95% confidence ellipses of incremental cost and effectiveness for comparison between once weekly resistance training (1RT) and twice weekly balance and tone (2BT, comparator). QALY estimates are based on utility scores from the EQ-5D for figures 1c and 1d.



**Figure .**

**1a.** Cost effective plane depicting the 95% confidence ellipses of incremental cost and effectiveness for comparison between once-weekly resistance training and balance and tone (comparator) with Quality Adjusted Life Years (QALYs) estimated from the EuroQol (EQ-5D); **1b.** Cost effective plane depicting the 95% confidence ellipses of incremental cost and effectiveness for comparison between twice-weekly resistance training and balance and tone (comparator) with QALYs estimated from the EQ-5D.

**Table 1**

Estimates of the Incremental QALY, Incremental Cost and Incremental cost effectiveness ratio estimated using the EQ-5D or the SF-6D

	Option					
	BAT (n=28)		1x RT (n=35)		2x RT (n=35)	
	EQ-5D	SF-6D	EQ-5D	SF-6D	EQ-5D	SF-6D
Mean Baseline utility score (SD)	0.834 (0.183)	0.711 (0.055)	0.848 (0.193)	0.702 (0.081)	0.866 (0.161)	0.699 (0.072)
Mean 9-month utility score (SD)	0.792 (0.212)	0.690 (0.066)	0.806 (0.245)	0.701 (0.093)	0.862 (0.125)	0.715 (0.059)
Mean 21-month utility score (SD)	0.859 (0.136)	0.769 (0.093)	0.829 (0.216)	0.751 (0.089)	0.824 (0.135)	0.749 (0.081)
Adjusted incremental QALY gain at 9 months	Reference	Reference	0.084	0.003	0.179	0.003
Adjusted incremental QALY gain at 21 months	Reference	Reference	-0.051	-0.144	-0.081	-0.127
Incremental cost at 9 months	Reference	Reference	Reference	\$-393	Reference	\$-88
Incremental cost at 21 months	Reference	Reference	Reference	Reference	Reference	\$-1077
Incremental cost per QALY (ICER) at 9 months	Reference	Reference	D	D	D	D
Incremental cost per QALY (ICER) at 21 months	Reference	Reference	-	-	-	-