



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/183866/>

Version: Published Version

Article:

Bansback, N., Trenaman, L., Mulhern, B.J. et al. (2022) Estimation of a Canadian preference-based scoring algorithm for the Veterans RAND 12-item Health Survey (VR-12): a population survey using a discrete choice experiment. *CMAJ open*, 10 (3). E589-E598. ISSN: 2291-0026

<https://doi.org/10.9778/cmajo.20210113>

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Estimation of a Canadian preference-based scoring algorithm for the Veterans RAND 12-Item Health Survey: a population survey using a discrete-choice experiment

Nick Bansback PhD, Logan Trenaman PhD, Brendan J. Mulhern PhD, Richard Norman PhD, Rebecca Metcalfe PhD, Richard Sawatzky PhD, John E. Brazier PhD, Donna Rowen PhD, David G.T. Whitehurst PhD

Abstract

Background: The Veterans RAND 12-Item Health Survey (VR-12) is a generic patient-reported outcome measure derived from the widely used 36- and 12-item Short Form Health Surveys. We aimed to estimate a Canadian preference-based scoring algorithm for the VR-12, enabling the derivation of health utility values for generating quality-adjusted life years (QALYs).

Methods: We conducted a discrete-choice experiment in a sample of the Canadian population in January and February 2019. Participants — recruited from a consumer research panel — completed an online survey, in English or French, that included 11 discrete-choice questions, each comprising 2 health profiles. We defined the health profiles using 8 VR-12 items and a duration attribute. Using conditional logit regressions, where each level of the respective VR-12 items was interacted with duration, we applied the coefficients to estimate health utility values interpretable on a scale of 0 (dead) to 1 (full health). Negative values reflect states considered worse than dead.

Results: A total of 3380 individuals completed the survey. Of these, 1688 (49.9%) were females, and 3101 (91.7%) completed the English version of the survey. Across all models, “feel downhearted and blue all of the time” and “pain interferes with your normal work extremely” were associated with the largest decrements in health utility. Excluding the 685 respondents (20.3%) who provided inconsistent responses had a negligible effect on the results. The recommended model, weighted to match population demographics, had health utility values ranging from −0.589 to 1.000.

Interpretation: Health utility values that reflect the preferences of the Canadian population can now be derived from responses to the VR-12. These values can be used to generate QALYs in future analyses.

Patient-reported outcome measures are used to collect information about aspects of respondents' health status at a particular point in time. Generic patient-reported outcome measures (as opposed to condition-specific patient-reported outcome measures) can be applied across different populations. They comprise questions about health-related aspects of quality of life, each with defined levels of impact or severity, resulting in the definition of a multitude of health states.¹ Preference-based patient-reported outcome measures include scoring algorithms that provide a “value set” — a set of index scores (often referred to as health utility values or preference weights) that provides a value for each unique instrument-defined health state.¹ Health utility values are interpreted on a scale anchored at full health = 1 and dead = 0, with negative scores reflecting health states worse than dead.²⁻⁴ Health utility values can be used to calculate quality-adjusted life years (QALYs) to inform analyses such as economic evaluations and population health comparisons.

Health utility values can be obtained by eliciting preferences that capture willingness to trade extra years of life for improvements in health (i.e., trading quantity for quality). In several jurisdictions, including Canada, guidelines for the conduct of economic evaluation recommend that health utility values reflect the preferences of a representative sample of the general population.⁵⁻⁷ Currently, Canadian value sets are available for a number of generic, preference-based, patient-reported outcome measures, including the Health Utilities Index (HUI) Mark 2 (HUI2)⁸ and Mark 3 (HUI3)⁹ and the 3-level (EQ-5D-3L)¹⁰ and 5-level (EQ-5D-5L)¹¹ versions of

Competing interests: None declared.

This article has been peer reviewed.

Correspondence to: Nick Bansback, nick.bansback@ubc.ca

CMAJ Open 2022 July 5. DOI:10.9778/cmajo.20210113

the EQ-5D. The availability of these value sets has the dual benefit of providing analysts with options when selecting outcome measures and ensuring that the preferences of Canadians are reflected in health care decision-making (e.g., in health technology appraisal processes).

Increasingly, the Veterans RAND 12-Item Health Survey (VR-12) is being used to measure health-related quality of life in routine data collection initiatives.^{12,13} The VR-12 is a shortened version of the Veterans RAND 36-item Health Survey,¹⁴ which was itself an adaptation of the RAND 36-item Health Survey (also known as the RAND SF-36).¹⁵ The VR-12 includes 12 questions that cover 8 domains, plus 2 questions that ask about health issues “compared with 1 year ago.”¹² The VR-12 was not designed to describe health states for the purposes of deriving health utility values. Scoring of the VR-12 is based on psychometric measurement methods for deriving scales that correspond with the 8 domains, as well as overarching physical and mental component scores.¹²

Given increasing use of the VR-12 in routine population measurement (in part because it is free for the user, although permission from the developers is required),¹⁶ considerable VR-12 data are available to inform real-world analyses such as economic evaluations. To facilitate this utilization, in a manner aligned with national guidelines, a value set that reflects the preferences of Canadians is required. Accordingly, our objective was to develop a scoring algorithm, based on the preferences of the Canadian general public, that estimates health utility values for health states described by the VR-12.

Methods

Study design

We used an online survey to administer a discrete-choice experiment in January and February 2019. The experiment required participants to choose their preferred option when presented with 2 health profiles (called a “choice set”). Each health profile described living in a health state defined using attributes and levels from 8 VR-12 items, for a certain duration. Respondents completed 11 choice sets. With aggregation across the sample, such data can be used to estimate population-level trade-offs between different health states (representing aspects of quality of life) and length of life, resulting in a societal value set — an approach that has been used in previous studies.^{17–19} We piloted the survey to get feedback on participants’ understanding of the questions.

In the following sections (supplemented by further details in Appendix 1, available at www.cmajopen.ca/content/10/3/E589/suppl/DC1, including the Checklist for Reporting Results of Internet E-Surveys [CHERRIES]²⁰), we describe development of the discrete-choice task and the survey, as well as the methods of analysis.

Participants

For the pilot phase of the project, we sought about 100 Canadians (recruited from the Amazon Mechanical Turk marketplace) to complete a survey hosted on the PrefApp platform

located at the Centre for Health Evaluation and Outcome Sciences (British Columbia, Canada). We performed recruitment for the subsequent main survey (aiming for a sample of about 3000 respondents) through a larger study, in which half the participants were randomly assigned to complete the survey described here (again, on the PrefApp platform); the other half completed a different survey on an unrelated research question concerning the VR-12. A market research company (Ipsos) invited members of their national consumer research panel, primarily by email, to take part in “a study about a measure of quality of life.” Participants who clicked on the link were given further information about both surveys, including the type of questions and approximate completion time. Those who continued were given further study information and asked if they wanted to consent to proceed.

After providing consent and answering demographic questions, participants were randomly assigned to 1 of the 2 surveys. Participants who completed the survey were awarded points by the company, which could be used toward financial rewards. We used quota sampling to encourage representation of the Canadian general population in terms of age, biological sex and location (based on provinces). We checked Internet Protocol addresses for duplicates, to reduce the likelihood that individuals completed the survey more than once.

Data sources

We obtained data for the main analysis by asking participants to answer discrete-choice questions based on health states described using the VR-12. Not all items in the VR-12 are relevant for the purposes of defining and valuing a health state (e.g., questions asking “compared with 1 year ago”), and multiple items measure similar constructs; therefore, 3 of the authors (B.J.M., J.E.B., D.R.) used an established process to construct a health state classification system.^{21,22} The process comprises factor analysis to assess dimensionality, Rasch analysis to understand item performance and tests of differential item function. This resulted in the selection of 8 health attributes, each of which corresponded to a VR-12 item.

We used the following naming conventions for the VR-12-related attributes: physical functioning, role physical, role emotional, bodily pain, mental health — anxiety, mental health — depression, vitality and social functioning. We constructed the levels for each attribute on the basis of response options for the respective VR-12 item, with the exception of items that included “a good bit of the time.” This response option was deemed too similar to “a little of the time” and was removed from the classification system. The final classification system, presented in Appendix 2 (available at www.cmajopen.ca/content/10/3/E589/suppl/DC1), defines 234 375 (3⁵7) health states.

Following previous studies,^{17–19} we used a discrete-choice experiment to elicit trade-offs between health states described by the 8 VR-12 items and life duration. To do this, we added a ninth attribute, the duration attribute, termed “life years.” We described this attribute using 4 levels (“1 year,” “4 years,” “7 years” and “10 years”), an approach used in previous studies.^{17,23} We chose an upper limit of 10 years for comparability with the time horizon in most national valuation studies.^{24–26}

We developed a pilot version of the survey using feedback from 16 cognitive interviews (conducted by R.M.; details provided in Appendix 1, including Appendix Table 1). Feedback from the pilot survey ($n = 103$) informed some revisions to the wording and layout of the final survey (Appendix 3, available at www.cmajopen.ca/content/10/3/E589/suppl/DC1). The survey was then forward- and back-translated into French. The demographic characteristics of those who participated in the cognitive interviews and in the survey pilot testing are available in Appendix 1 (Appendix Tables 2 and 3, respectively).

The final survey began with details about the study, followed by a consent form and questions to collect information about the characteristics and health history of participants.

Participants were then asked to complete the VR-12 for themselves (with instruction that later questions would be based on this questionnaire), followed by a warm-up discrete-choice exercise and, finally, the discrete-choice questions.

Feedback from the cognitive interviews informed the development of the discrete-choice task in the final survey. Figure 1 provides an example of a choice set that was presented to participants in the final survey. One of the authors (R.N.) used experimental design theory (NGene, modified Federov algorithm) to select 200 choice sets (from the millions of possible choice sets) that would provide the most information for analysis purposes, i.e., a d-efficient design. These 200 choice sets were “blocked” into 20 groups of 10 questions to ensure that each block had representation of

	Health state A	Health state B
Live for ___ with the following and then you die ...	7 years	4 years
Feel downhearted and blue ...	none of the time	some of the time
Have a lot of energy ...	a little of the time	all of the time
Feel calm and peaceful ...	some of the time	none of the time
Accomplish less than you would like due to your physical health ...	all of the time	most of the time
Health interferes with social activities ...	most of the time	most of the time
Moderate activities are ...	limited a lot	limited a lot
Accomplish less than you would like due to your emotional problems ...	most of the time	most of the time
Pain interferes with your normal work ...	moderately	moderately
Which do you prefer?	<input type="checkbox"/>	<input type="checkbox"/>

Figure 1: Example of a choice set presented to participants. Details about the use of different intensities of highlighting (i.e., the various shades of yellow) for 5 of the attributes are provided in Appendix 1.

attributes and levels. We randomly assigned each participant to 1 of the blocks. To assess the consistency of participants' responses, we repeated the warm-up discrete-choice question along with the set of 10. The repeated question (i.e., an 11th question) appeared fourth in the sequence, and we did not use the responses to this question in the modelling. We explored engagement with the task by timing how long participants took to answer all 11 discrete-choice questions.

Statistical analysis

Our analysis followed the methods reported in peer-reviewed studies that used discrete-choice experiments to elicit population value sets.^{17–19,23} We (N.B., B.J.M., R.N., J.E.B., D.R., D.G.T.W.) used conditional logit regression to model the data. This method estimates coefficients for each level of the VR-12-related attributes (dummy-coded, with the best level used as the reference category) interacted with the corresponding number of life years, along with a coefficient for the duration attribute as a main effect. The coefficient for the life years main effect is the value of living in full health for 1 year (expected to be positive). The coefficients for each level of the VR-12-related attributes reflect the disutility of living with the respective impairment for a duration of 1 year (expected to be negative). A random error term captured unobservable factors, as well as potential inconsistencies in decision-making.²⁷ We tested interactions between dummy-coded levels of duration and, if satisfied that they were sufficiently linear, coded them as continuous, both in the main effect and the set of interaction terms.

The estimated regression coefficients are on a latent scale: they do not have cardinal properties and, therefore, their magnitudes are not directly comparable (“unanchored” coefficients). We estimated coefficients anchored on the full health – dead scale as the marginal willingness to trade an extra year of life for each VR-12-related attribute level, calculated by dividing each interaction coefficient by the duration coefficient. Additive combinations of anchored coefficients provide the health utility values for the health states defined by the classification system.

Previous valuation studies have reported inconsistencies in the directionality of coefficients when valuing large descriptive systems (e.g., when an increase in severity leads to an increase rather than a decrease in utility). We used models involving combinations of attribute levels (merged on the basis of best fit) to overcome any inconsistencies, while also exploring issues of participant engagement and understanding regarding the survey. Because preferences may vary in relation to different sociodemographic characteristics, we weighted the “preferred” final model to ensure a representative population estimate and to account for non-response and noncoverage. We applied a raking approach using data from the 2016 Census²⁸ and the 2015/16 Canadian Community Health Survey (<https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=238854>) to develop weights, and included these weights in the conditional logit models. We performed the analyses with Stata and SAS software.

Ethics approval

The study was approved by the University of British Columbia Behavioural Research Ethics Board (study no. H18-00969).

Results

A total of 8110 individuals clicked on the initial link to find out further information about the larger study (i.e., the 2 surveys), of whom 6731 (83.0%) provided consent to take part. The discrete-choice survey was randomly assigned to 3445 (51.2% of the 6731) potential participants; 3380 (98.1% of the 3445) completed all survey questions, which yielded 33 800 choice-set responses for analysis.

Table 1 reports the demographic characteristics of the survey respondents. Relative to the Canadian general population, differences included a higher proportion of study participants who identified as white, with lower reported annual income, with higher levels of education and having never legally married. A lower proportion of respondents chose to complete the survey in French, relative to the proportion of Canadians reporting French as their first official language spoken.

Of the 10 379 choice sets where the level of life duration varied between health profiles, the profile with the lower duration was selected on 2433 (23.4%) occasions.

Unanchored values

Results from the conditional logit models are provided in Table 2. Model 1 includes all respondents. Most coefficients (26 of 30) were negative, with the magnitude of the disutility tending to increase with more severe levels of impairment. Based on the coefficients of the worst levels, the most important attributes were mental health — depression (–0.145) and bodily pain (–0.127), whereas the least important were role physical (–0.047) and role emotional (–0.048). As expected, there were some directional inconsistencies; however, vitality level 4 was the only statistically significant inconsistency at the 5% level.

In models 2 to 5, attribute levels were combined to overcome directional inconsistencies (Table 2). Model 2 included all participants; model 3 excluded the 37 respondents (1.1%) who provided the same answer to all 10 discrete-choice questions (i.e., always selected health state A or always selected health state B); model 4 excluded the 685 respondents (20.3%) whose responses failed the consistency check; and model 5 excluded the 315 respondents (9.3%) who completed all questions in less than 2 minutes. Across all 4 consistent models, a similar pattern was observed when attributes were ordered according to the magnitude of the coefficients of the worst levels. Given the similarities across models 1 to 5, model 2 was considered the most defensible because of the directional consistency and inclusivity of participants. Model 2 was then weighted to adjust for the differences between the study sample and the Canadian general population (i.e., model 6 is a weighted version of model 2).

Table 1 (part 1 of 2): Characteristics of survey respondents, with comparable statistics for the Canadian general population, drawn from Statistics Canada 2016 Census²⁸

Characteristic	No. (%) of sample n = 3380	% of Canadian general population
Age group, yr		
18–29	663 (19.6)	19.3
30–39	599 (17.7)	17.5
40–49	615 (18.2)	17.0
50–59	670 (19.8)	18.4
60–69	367 (10.9)	15.3
≥ 70	466 (13.8)	12.6
Gender		
Male	1686 (49.9)	48.6
Female	1688 (49.9)	51.4
Other	6 (0.2)	0
Survey language		
English	3101 (91.7)	–
French*	279 (8.3)	–
Province/territory		
Alberta	373 (11.0)	11.2
British Columbia	504 (14.9)	13.5
Manitoba	154 (4.6)	3.5
New Brunswick	77 (2.3)	2.2
Newfoundland and Labrador	48 (1.4)	1.5
Northwest Territories	3 (0.1)	0.1
Nova Scotia	125 (3.7)	2.7
Nunavut	1 (< 0.1)	0.1
Ontario	1283 (38.0)	38.3
Prince Edward Island	16 (0.5)	0.4
Quebec	695 (20.6)	23.4
Saskatchewan	99 (2.9)	3.0
Yukon	2 (0.1)	0.1
Education		
No certificate, diploma or degree	92 (2.7)	18.3
Secondary (high) school diploma or equivalency certificate	806 (23.8)	26.5
Apprenticeship or trades certificate or diploma	215 (6.4)	9.8
College, CEGEP or other non-university certificate or diploma	816 (24.1)	19.4
University certificate or diploma below bachelor level	298 (8.8)	2.8
University certificate, diploma or degree at bachelor level or above	1153 (34.1)	23.3

Table 1 (part 2 of 2): Characteristics of survey respondents, with comparable statistics for the Canadian general population, drawn from Statistics Canada 2016 Census²⁸

Characteristic	No. (%) of sample n = 3380	% of Canadian general population
Ethnicity†		
White	2707 (80.1)	74.5
South Asian	116 (3.4)	5.3
Chinese	247 (7.3)	4.7
Black	71 (2.1)	3.1
Filipino	29 (0.9)	2.2
Latin American	34 (1.0)	1.3
Arab	27 (0.8)	1.3
Southeast Asian	33 (1.0)	0.9
West Asian	21 (0.6)	0.8
Korean	12 (0.4)	0.6
Japanese	16 (0.5)	0.3
First Nation (North American Indian)	48 (1.4)	2.4
Métis	13 (0.4)	1.6
Inuit	1 (< 0.1)	0.2
Indigenous/Aboriginal, not included elsewhere	22 (0.7)	0.1
Other	88 (2.6)	0.4
Marital status		
Never legally married	1136 (33.6)	22.9
Legally married (and not separated)	1434 (42.4)	49.0
Separated, but still legally married	95 (2.8)	2.6
Divorced	227 (6.7)	6.7
Widowed	124 (3.7)	6.0
Living with a common-law partner	363 (10.7)	12.8
Missing	1 (< 0.1)	0
Annual income, \$		
0–19 999	363 (10.7)	4.4
20 000–49 999	911 (27.0)	20.0
50 000–99 999	1256 (37.2)	35.5
100 000–149 999	557 (16.5)	21.8
≥ 150 000	274 (8.1)	18.4
Missing	19 (0.6)	0

*About one-fifth (20.6%) of Canadians report French as their first official language spoken.
†Ethnicity was self-reported using predetermined categories. Participants could select as many categories as were applicable.

Table 2: Results from the unanchored models

Parameters	Model; coefficient (SE)					
	Model 1: All levels, all respondents	Model 2: Combined levels, all respondents	Model 3: Combined levels, excluding nonswitchers	Model 4: Combined levels, excluding inconsistent respondents	Model 5: Combined levels, excluding completions < 2 min	Model 6: Weighted model 2
PF1xLY	–	–	–	–	–	–
PF2xLY	–0.028 (0.005)*	–0.025 (0.005)*	–0.026 (0.005)*	–0.035 (0.006)*	–0.028 (0.005)*	–0.023 (0.007)*
PF3xLY	–0.080 (0.005)*	–0.080 (0.005)*	–0.080 (0.005)*	–0.093 (0.006)*	–0.086 (0.005)*	–0.073 (0.007)*
RP1xLY	–	–	–	–	–	–
RP2xLY	0.008 (0.005)	–0.011 (0.004)*	–0.012 (0.004)*	–0.014 (0.005)*	–0.011 (0.005)*	–0.004 (0.006)
RP3xLY	0.009 (0.006)	†	†	†	†	†
RP4xLY	–0.035 (0.005)*	†	†	†	†	†
RP5xLY	–0.047 (0.006)*	–0.053 (0.005)*	–0.054 (0.005)*	–0.059 (0.006)*	–0.057 (0.006)*	–0.050 (0.008)*
RE1xLY	–	–	–	–	–	–
RE2xLY	–0.016 (0.006)*	–0.006 (0.005)	–0.006 (0.005)	–0.008 (0.005)	–0.006 (0.005)	–0.009 (0.007)
RE3xLY	0.013 (0.006)	†	†	†	†	†
RE4xLY	–0.013 (0.006)	†	†	†	†	†
RE5xLY	–0.048 (0.005)*	–0.052 (0.005)*	–0.051 (0.005)*	–0.052 (0.006)*	–0.058 (0.006)*	–0.051 (0.008)*
BP1xLY	–	–	–	–	–	–
BP2xLY	–0.028 (0.006)*	–0.026 (0.005)*	–0.027 (0.005)*	–0.034 (0.006)*	–0.030 (0.005)*	–0.018 (0.007)*
BP3xLY	–0.027 (0.006)	†	†	†	†	†
BP4xLY	–0.096 (0.005)*	–0.092 (0.005)*	–0.091 (0.005)*	–0.106 (0.006)*	–0.101 (0.005)*	–0.085 (0.007)*
BP5xLY	–0.127 (0.006)*	–0.129 (0.006)*	–0.128 (0.006)*	–0.149 (0.006)*	–0.138 (0.006)*	–0.123 (0.008)*
MA1xLY	–	–	–	–	–	–
MA2xLY	–0.002 (0.005)	–0.019 (0.005)*	–0.019 (0.005)*	–0.016 (0.006)*	–0.019 (0.005)*	–0.018 (0.007)*
MA3xLY	–0.020 (0.006)*	†	†	†	†	†
MA4xLY	–0.058 (0.005)*	–0.058 (0.005)*	–0.059 (0.005)*	–0.063 (0.005)*	–0.058 (0.005)*	–0.059 (0.007)*
MA5xLY	–0.112 (0.006)*	–0.112 (0.005)*	–0.114 (0.005)*	–0.125 (0.006)*	–0.120 (0.005)*	–0.107 (0.007)*
MD1xLY	–	–	–	–	–	–
MD2xLY	–0.028 (0.006)*	–0.024 (0.005)*	–0.024 (0.005)*	–0.030 (0.006)*	–0.028 (0.006)*	–0.027 (0.008)*
MD3xLY	–0.015 (0.006)*	†	†	†	†	†
MD4xLY	–0.112 (0.006)*	–0.113 (0.005)*	–0.112 (0.005)*	–0.128 (0.006)*	–0.125 (0.006)*	–0.113 (0.008)*
MD5xLY	–0.145 (0.006)*	–0.148 (0.006)*	–0.150 (0.006)*	–0.168 (0.007)*	–0.160 (0.006)*	–0.148 (0.009)*
VT1xLY	–	–	–	–	–	–
VT2xLY	–0.020 (0.006)*	–0.019 (0.006)*	–0.018 (0.006)*	–0.017 (0.007)*	–0.021 (0.006)*	–0.022 (0.008)*
VT3xLY	–0.044 (0.006)*	–0.034 (0.005)*	–0.035 (0.005)*	–0.037 (0.006)*	–0.036 (0.005)*	–0.037 (0.007)*
VT4xLY	–0.023 (0.006)*	†	†	†	†	†
VT5xLY	–0.082 (0.006)*	–0.082 (0.006)*	–0.083 (0.006)*	–0.091 (0.006)*	–0.088 (0.006)*	–0.081 (0.008)*
SF1xLY	–	–	–	–	–	–
SF2xLY	0.005 (0.005)	–0.021 (0.004)*	–0.021 (0.004)*	–0.021 (0.005)*	–0.021 (0.005)*	–0.024 (0.006)*
SF3xLY	–0.017 (0.005)*	†	†	†	†	†
SF4xLY	–0.068 (0.005)*	–0.072 (0.005)*	–0.073 (0.005)*	–0.082 (0.005)*	–0.079 (0.005)*	–0.065 (0.007)*
SF5xLY	–0.083 (0.006)*	–0.094 (0.005)*	–0.093 (0.005)*	–0.104 (0.006)*	–0.100 (0.005)*	–0.086 (0.007)*
LY	0.464 (0.012)*	0.473 (0.011)*	0.472 (0.011)*	0.537 (0.012)*	0.507 (0.011)*	0.452 (0.016)*
R ²	0.167	0.164	0.164	0.196	0.179	0.154
No. of choice set responses	33 800	33 800	33 430	26 950	30 650	33 800

Note: BP = bodily pain, LY = life years, MA = mental health — anxiety, MD = mental health — depression, PF = physical functioning, RE = role emotional, RP = role physical, SE = standard error, SF = social functioning, VT = vitality.
 *Coefficient for the VR-12-related attribute is statistically different from the coefficient of the attribute level above, at the 5% level of significance. For the life-years attribute, the coefficient is statistically different from 0 at the 5% level of significance.
 †Attribute levels are combined to overcome the directional inconsistencies observed in model 1. For example, in model 3, SF2 and SF3 have the same disutility (–0.021).

Anchored values

Figure 2 shows the anchored value sets produced by all models; the corresponding coefficients are presented in Appendix 1, Appendix Table 4. The anchored values for each attribute level of model 6 are provided in Table 3. These results provide the means to calculate health utility values for all health states defined by the classification system. The health utility values for model 6 ranged from 1.000 (no impairment on any attribute) to -0.589 (the worst level of impairment on all attributes); 5.9% of the health states had values less than or equal to 0. The distribution of health utility values for model 6 is shown in Figure 3.

Interpretation

In this study, we developed a value set for the VR-12 based on the preferences of the Canadian general public. We suggest use of model 6 (Table 3) because it has consistently ordered levels, is based on the preferences of all respondents and is weighted to the demographic characteristics of the Canadian population. Further information and tools to facilitate the scoring of health utility values from VR-12 responses is available on the study website (www.bcpcm.ca/resources/VR-12/).

Many preference-based, patient-reported outcome measures are available for use,¹ and 1 previous study has estimated

health utility values for the VR-12 (based on a mapping exercise using values from the United Kingdom).²⁹ Although there are myriad reasons why value sets differ across different preference-based instruments, recent studies in Canada and the UK have been consistent in identifying pain and mental health items as being associated with the largest utility decrements.^{11,23,25} In the UK value set for the SF-6Dv2, the most severe level of pain had a disutility of 0.620,²³ compared with 0.272 in the current study. The difference in magnitude is likely related to the items used in the respective studies, with the SF-6Dv2 using the SF-36v2 pain severity item,²¹ whereas the single pain item in the VR-12 is framed around interference “with your normal work.”

We recognize that there is no “best” preference-based patient-reported outcome measure, and health utility values across instruments are expected to differ,^{30,31} something that is reflected in Canadian guidelines for the conduct of economic evaluation.⁵ Instead, researchers are required to consider the relative merits of the available options for the evaluative purpose being pursued. For example, anxiety and depression, which are combined in a single item in the EQ-5D instruments, have been shown to be valued differently.³² The Canadian value set for the VR-12 separates anxiety from depression, meaning that it may better reflect changes in health outcome attributable to interventions that improve one but not the other. This is a context-specific empirical question

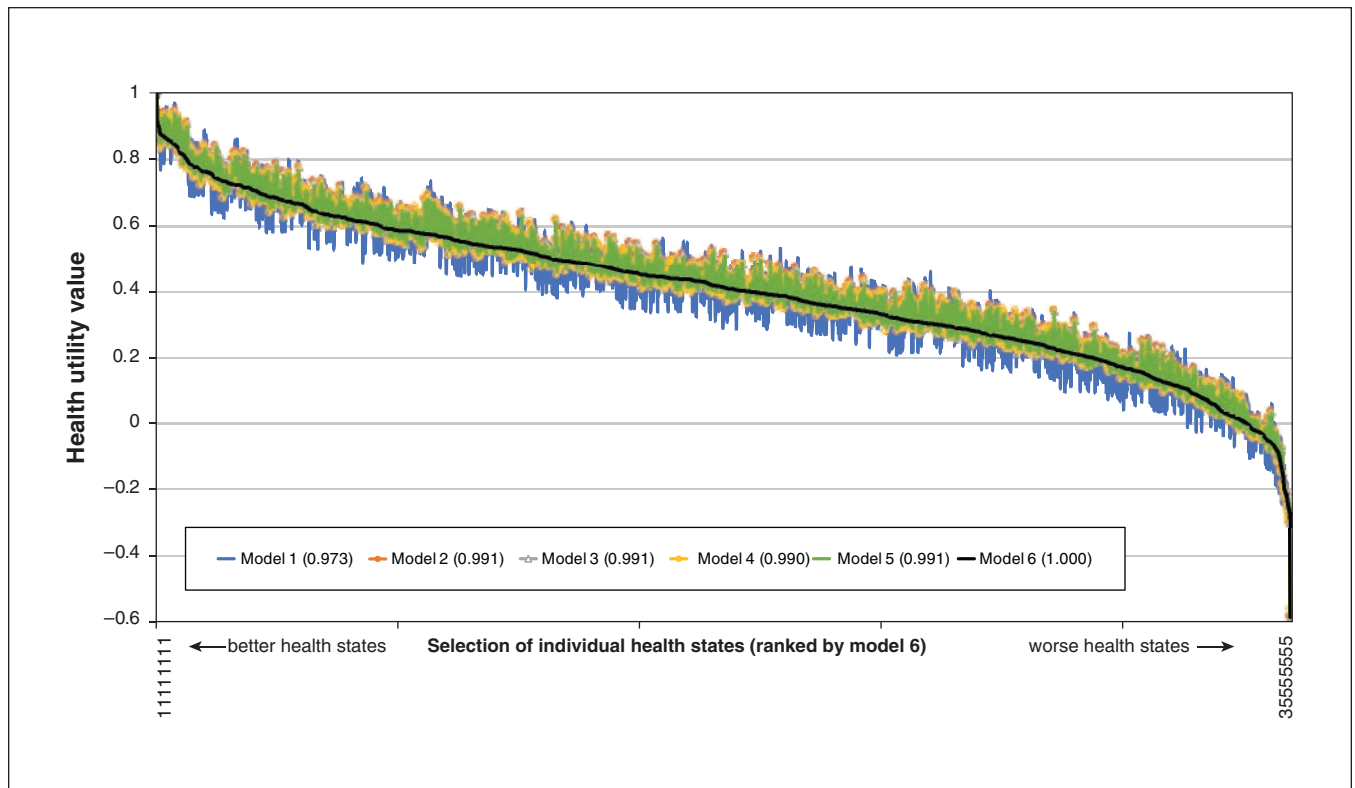


Figure 2: Anchored health utility values from the 6 models, presented as the health utility values for each model for each health state. The “best” health state (i.e., level 1 in all attributes, denoted as health state 11111111) is on the far left, with a health state equal to 1. The “worst” health state (i.e., the lowest level on each attribute, denoted as 35555555) is on the far right. The Pearson correlation with model 6 is shown in parentheses for each of the other models.

Table 3: Anchored disutility values for each attribute level of the preferred model (model 6)*

Attribute	Level and level description†	Value (95% CI)‡
PF	PF1: Moderate activities are <i>not limited at all</i>	0 (0)
	PF2: Moderate activities are <i>limited a little</i>	0.052 (0.021 to 0.083)
	PF3: Moderate activities are <i>limited a lot</i>	0.161 (0.130 to 0.192)
RP	RP1: Accomplish less than you would like as a result of your physical health <i>none of the time</i>	0.000 (0.000 to 0.000)
	RP2: Accomplish less than you would like as a result of your physical health <i>a little of the time</i>	0.010 (−0.018 to 0.038)
	RP3: Accomplish less than you would like as a result of your physical health <i>some of the time</i>	0.010 (−0.018 to 0.038)
	RP4: Accomplish less than you would like as a result of your physical health <i>most of the time</i>	0.010 (−0.018 to 0.038)
	RP5: Accomplish less than you would like as a result of your physical health <i>all of the time</i>	0.111 (0.078 to 0.143)
RE	RE1: Accomplish less than you would like as a result of your emotional problems <i>none of the time</i>	0 (0)
	RE2: Accomplish less than you would like as a result of your emotional problems <i>a little of the time</i>	0.019 (−0.009 to 0.048)
	RE3: Accomplish less than you would like as a result of your emotional problems <i>some of the time</i>	0.019 (−0.009 to 0.048)
	RE4: Accomplish less than you would like as a result of your emotional problems <i>most of the time</i>	0.019 (−0.009 to 0.048)
	RE5: Accomplish less than you would like as a result of your emotional problems <i>all of the time</i>	0.113 (0.082 to 0.145)
BP	BP1: Pain <i>does not</i> interfere with your normal work	0 (0)
	BP2: Pain interferes with your normal work <i>a little bit</i>	0.040 (0.009 to 0.070)
	BP3: Pain interferes with your normal work <i>moderately</i>	0.040 (0.009 to 0.070)
	BP4: Pain interferes with your normal work <i>quite a bit</i>	0.187 (0.157 to 0.218)
	BP5: Pain interferes with your normal work <i>extremely</i>	0.272 (0.237 to 0.307)
MA	MA1: Feel calm and peaceful <i>all of the time</i>	0 (0)
	MA2: Feel calm and peaceful <i>most of the time</i>	0.040 (0.009 to 0.070)
	MA3: Feel calm and peaceful <i>some of the time</i>	0.040 (0.009 to 0.070)
	MA4: Feel calm and peaceful <i>a little of the time</i>	0.130 (0.102 to 0.159)
	MA5: Feel calm and peaceful <i>none of the time</i>	0.237 (0.206 to 0.267)
MD	MD1: Feel downhearted and blue <i>none of the time</i>	0 (0)
	MD2: Feel downhearted and blue <i>a little of the time</i>	0.061 (0.027 to 0.094)
	MD3: Feel downhearted and blue <i>some of the time</i>	0.061 (0.027 to 0.094)
	MD4: Feel downhearted and blue <i>most of the time</i>	0.249 (0.217 to 0.281)
	MD5: Feel downhearted and blue <i>all of the time</i>	0.326 (0.291 to 0.362)
VT	VT1: Have a lot of energy <i>all of the time</i>	0 (0)
	VT2: Have a lot of energy <i>most of the time</i>	0.048 (0.013 to 0.083)
	VT3: Have a lot of energy <i>some of the time</i>	0.081 (0.051 to 0.112)
	VT4: Have a lot of energy <i>a little of the time</i>	0.081 (0.051 to 0.112)
	VT5: Have a lot of energy <i>none of the time</i>	0.179 (0.145 to 0.213)
SF	SF1: Health interferes with social activities <i>none of the time</i>	0 (0)
	SF2: Health interferes with social activities <i>a little of the time</i>	0.053 (0.026 to 0.080)
	SF3: Health interferes with social activities <i>some of the time</i>	0.053 (0.026 to 0.080)
	SF4: Health interferes with social activities <i>most of the time</i>	0.143 (0.115 to 0.172)
	SF5: Health interferes with social activities <i>all of the time</i>	0.191 (0.162 to 0.219)

Note: BP = bodily pain, CI = confidence interval, MA = mental health — anxiety, MD = mental health — depression, PF = physical functioning, RE = role emotional, RP = role physical, SF = social functioning, VT = vitality.

*For illustration, the estimated health utility value for health state 34454212 = 0.294 (i.e., 1 [full health] − 0.161 [PF3] − 0.010 [RP4] − 0.019 [RE4] − 0.272 [BP5] − 0.130 [MA4] − 0.061 [MD2] − 0.000 [VT1] − 0.053 [SF2]). Details for handling “a good bit of the time” VR-12 responses are provided in Appendix 2 (see footnote).

†Text in *italics* indicates the information that was varied in the description of health profiles (Figure 1).

‡Reported to 3 decimal places.

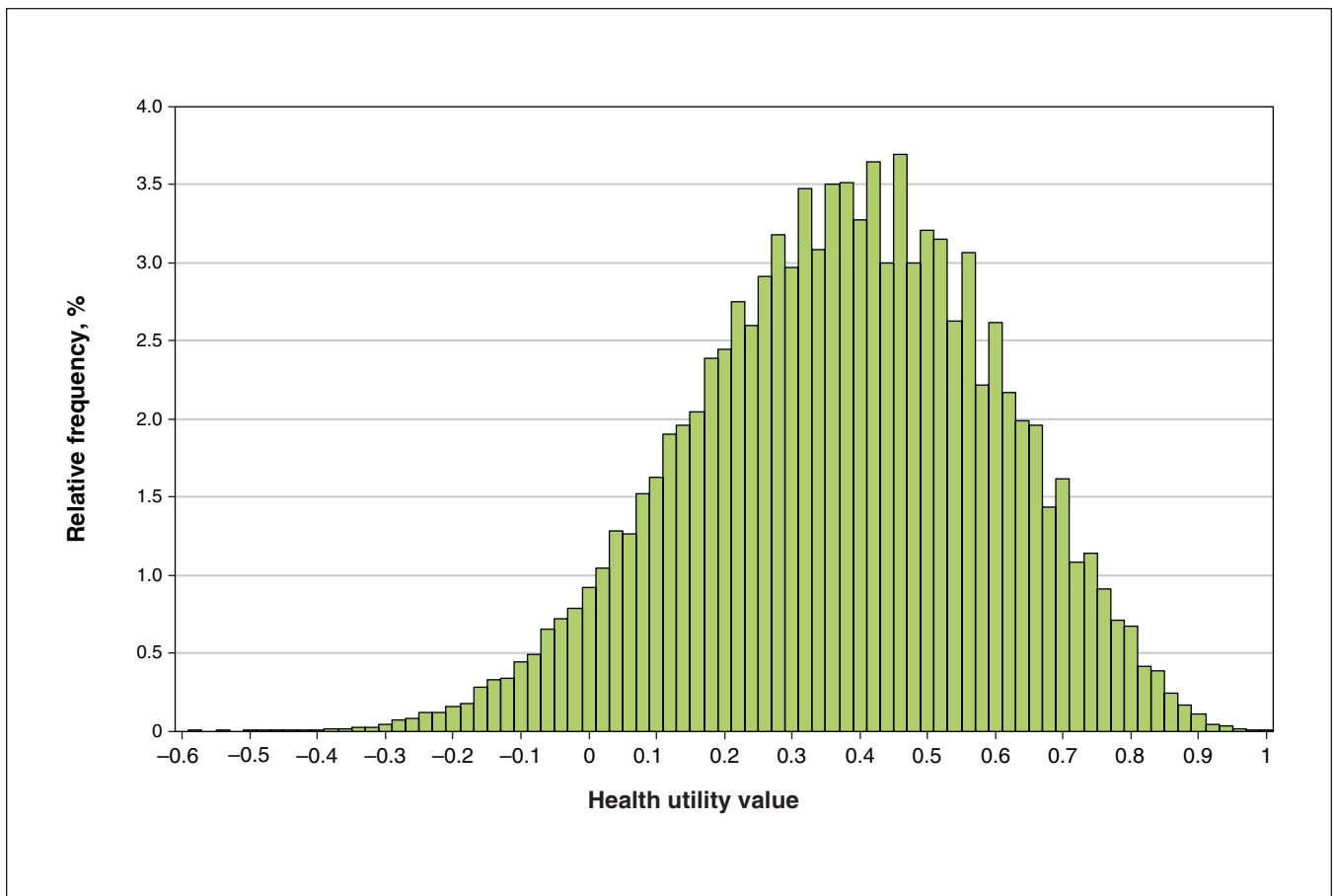


Figure 3: Distribution of health utility values for model 6 (the preferred model) across all possible health states.

and highlights the need for comprehensive comparative psychometric evaluation of the new classification system and value set (i.e., compared with other preference-based, patient-reported outcome measures).

Another empirical question is whether health state values from patient populations would differ from those elicited from the general public,³³ and whether the use of societal preferences undervalues the quality of life of people with health conditions and disabilities.^{34,35} The approach taken in this study — elicitation of societal preferences — reflects the explicit expectations detailed in national guidelines for the economic evaluation of health technologies.⁵⁻⁷

Limitations

Preference elicitation tasks, such as discrete-choice experiments, are cognitively challenging,^{36,37} which can introduce bias into the sample and the subsequent responses and results. To mitigate the potential of such biases, we used cognitive interviews and pilot testing in designing our discrete-choice experiment, and we employed novel methods to simplify the task (Appendix 1). Such methods have been used successfully in other discrete-choice studies.^{19,37}

Although our results showed some inconsistencies, these were resolved by combining adjacent levels. We do not know

if these inconsistencies were a result of the choice sets presented (i.e., a “real” inconsistency), participants’ lack of comprehension of or attention to the task, or merely random. Regardless of the reason, a methodologic priority is to advance the design of preference elicitation tasks to address issues of inclusivity. Such advances must extend beyond recognition of individuals with lower cognitive abilities.³⁶

Our study included the preferences of more Canadians than any other Canadian value set, yet it cannot be said to have explored preferences that reflect the full diversity of the population. For example, we did not include French speakers in the cognitive interview phase, and only 8.3% of the sample chose to complete the French version of the survey.

Our sample, recruited from a consumer research panel to complete an online survey, likely had higher literacy skill levels than the general population.

Conclusion

Health utility values reflecting Canadian preferences can now be generated from VR-12 responses. The new scoring algorithm provides a means to make further use of existing data sets, as well as an alternative to the HUI and EQ-5D instruments for the estimation of QALYs in economic evaluations involving primary data collection.

References

- Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annu Rev Public Health* 2000;21:587-611.
- Torrance GW. Utility approach to measuring health-related quality of life. *J Chronic Dis* 1987;40:593-603.
- Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J Technol Assess Health Care* 1989;5:559-75.
- Macran S, Kind P. "Death" and the valuation of health-related quality of life. *Med Care* 2001;39:217-27.
- CADTH methods and guidelines: guidelines for the economic evaluation of health technologies: Canada. 4th ed. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2017. Available: https://www.cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed.pdf (accessed 2021 Apr. 6).
- Guide to the methods of technology appraisal 2013. London (UK): National Institute for Health and Care Excellence; 2013. Available: <https://www.nice.org.uk/article/pmg9/chapter/Foreword> (accessed 2021 Apr. 6).
- Sanders GD, Neumann PJ, Basu A, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *JAMA* 2016;316:1093-103.
- Torrance GW, Feeny DH, Furlong WJ, et al. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Med Care* 1996;34:702-22.
- Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 system. *Med Care* 2002;40:1113-28.
- Bansback N, Tsuchiya A, Brazier J, et al. Canadian valuation of EQ-5D health states: preliminary value set and considerations for future valuation studies. *PLoS One* 2012;7:e31115.
- Xie F, Pullenayegum E, Gaebel K, et al.; Canadian EQ-5D-5L Valuation Study Group. A time trade-off-derived value set of the EQ-5D-5L for Canada. *Med Care* 2016;54:98-105.
- Selim AJ, Rogers W, Fleishman JA, et al. Updated U.S. population standard for the Veterans RAND 12-item Health Survey (VR-12). *Qual Life Res* 2009;18:43-52.
- Health Outcomes Survey (HOS). Centres for Medicare and Medicaid Services; 2021. Available: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/HOS> (accessed 2022 June 20).
- Kazis LE. The Veterans SF-36 health status questionnaire: development and application in the Veterans Health Administration. *Med Outcomes Trust Monit.* 2000;5:1-14.
- Stewart A, Ware JE. Measuring functioning and well-being: the Medical Outcomes Study approach. Santa Monica (CA): RAND Corporation; 1992.
- About the VR-36®, VR-12 and VR-6D®. Boston: Boston University, School of Public Health; 2021. Available: <http://www.bu.edu/sph/research/research-landing-page/vr-36-vr-12-and-vr-6d/about-the-vr-36-vr-12-and-vr-6d/> (accessed 2015 Oct. 12).
- Bansback N, Brazier J, Tsuchiya A, et al. Using a discrete choice experiment to estimate health state utility values. *J Health Econ* 2012;31:306-18.
- Viney R, Norman R, Brazier J, et al. An Australian discrete choice experiment to value EQ-5D health states. *Health Econ* 2014;23:729-42.
- Norman R, Viney R, Aaronson NK, et al. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format [published erratum in *Qual Life Res* 2016;25:2401]. *Qual Life Res* 2016;25:637-49.
- Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *J Med Internet Res* 2004;6:e34.
- Brazier JE, Mulhern BJ, Björner JB, et al.; SF-6Dv2 International Project Group. Developing a new version of the SF-6D health state classification system from the SF-36v2: SF-6Dv2. *Med Care* 2020;58:557-65.
- Brazier JE, Rowen D, Mavranzeouli I, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess* 2012;16:1-114.
- Mulhern BJ, Bansback N, Norman R, et al.; SF-6Dv2 International Project Group. Valuing the SF-6Dv2 classification system in the United Kingdom using a discrete-choice experiment with duration. *Med Care* 2020;58:566-73.
- Dolan P, Jones-Lee M. The time trade-off: a note on the effect of lifetime reallocation of consumption and discounting. *J Health Econ* 1997;16:731-9.
- Devlin NJ, Shah KK, Feng Y, et al. Valuing health-related quality of life: an EQ-5D-5L value set for England. *Health Econ* 2018;27:7-22.
- Szende A, Oppe M, Devlin N, editors. *EQ-5D value sets: inventory, comparative review and user guide*. Dordrecht (Netherlands): Springer; 2007:1-96.
- Lancsar E, Fiebig DG, Hole AR. Discrete choice experiments: a guide to model specification, estimation and software. *PharmacoEconomics* 2017;35:697-716.
- 2016 Census of population. Cat no 98-402-X2016005. Ottawa: Statistics Canada; 2017. Available: <https://www12.statcan.gc.ca/census-recensement/2016/index-eng.cfm> (accessed 2022 June 16).
- Selim AJ, Rogers W, Qian SX, et al. A preference-based measure of health: the VR-6D derived from the Veterans RAND 12-Item Health Survey. *Qual Life Res* 2011;20:1337-47.
- Whitehurst DGT, Bryan S. Another study showing that two preference-based measures of health-related quality of life (EQ-5D and SF-6D) are not interchangeable. But why should we expect them to be? *Value Health* 2011;14:531-8.
- Whitehurst DGT, Bryan S, Lewis M. Systematic review and empirical comparison of contemporaneous EQ-5D and SF-6D group mean scores. *Med Decis Making* 2011;31:E34-44.
- McDonald R, Mullett TL, Tsuchiya A. Understanding the composite dimensions of the EQ-5D: an experimental approach. *Soc Sci Med* 2020;265:113323.
- Versteegh MM, Brouwer WBF. Patient and general public preferences for health states: a call to reconsider current guidelines. *Soc Sci Med* 2016;165:66-74.
- Whitehurst DGT, Engel L. Disability discrimination and misdirected criticism of the quality-adjusted life year framework. *J Med Ethics* 2018;44:793-5.
- Helgesson G, Ernstsson O, Åström M, et al. Whom should we ask? A systematic literature review of the arguments regarding the most accurate source of information for valuation of health states. *Qual Life Res* 2020;29:1465-82.
- Engel L, Bansback N, Bryan S, et al. Exclusion criteria in national health state valuation studies: a systematic review. *Med Decis Making* 2016;36:798-810.
- Jonker MF, Donkers B, de Bekker-Grob E, et al. Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. *Health Econ* 2019;28:350-63.

Affiliations: School of Population and Public Health (Bansback), University of British Columbia; Centre for Health Evaluation and Outcome Sciences (Bansback, Trenaman, Metcalfe, Sawatzky), Vancouver, BC; Centre for Health Economics Research and Evaluation (Mulhern), University of Technology Sydney, Sydney, New South Wales, Australia; School of Health and Related Research (Mulhern, Brazier, Rowen), University of Sheffield, Sheffield, UK; School of Public Health (Norman), Curtin University, Perth, Western Australia, Australia; School of Nursing (Sawatzky), Trinity Western University, Langley, BC; Faculty of Health Sciences (Whitehurst), Simon Fraser University, Burnaby, BC

Contributors: Nick Bansback, Logan Trenaman, Richard Sawatzky and David Whitehurst conceived and designed the study; Nick Bansback, Logan Trenaman, Rebecca Metcalfe and David Whitehurst acquired the data; and Nick Bansback, Logan Trenaman, Brendan Mulhern, Richard Norman, Richard Sawatzky, John Brazier, Donna Rowen and David Whitehurst analyzed and interpreted the data. Nick Bansback and David Whitehurst drafted the manuscript. All of the authors revised the manuscript critically for important intellectual content, approved the final version for publication and agreed to be accountable for the work.

Funding: This study was funded by the British Columbia Patient-Centred Measurement Steering Committee.

Content licence: This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY-NC-ND 4.0) licence, which permits use, distribution and reproduction in any medium, provided that the original publication is properly cited, the use is noncommercial (i.e., research or educational use), and no modifications or adaptations are made. See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Data sharing: Aggregated data presented in this manuscript are available to other investigators upon request to the corresponding author. Further analysis of individual-level data is also available upon request.

Acknowledgements: The authors acknowledge Judy Chiu for support in manuscript preparation. They also thank the respondents for taking part in the study and attendees of the 41st Annual North American Meeting of the Society for Medical Decision Making (Portland, Oregon), who provided valuable feedback.

Supplemental information: For reviewer comments and the original submission of this manuscript, please see www.cmajopen.ca/content/10/3/E589/suppl/DC1.