**Appendix 1**: Supplementary information about the development of the discrete choice experiment (DCE) task, pilot testing and methods of analysis

### *Development of the DCE task*

Different forms of instruction and presentation were explored with the aim of increasing respondent understanding and engagement in the DCE choice sets. An initial survey was developed, in both paper and online formats, that included survey instructions, a warm-up exercise and mock DCE choice sets.

With respect to the DCE choice sets, we explored three presentation issues: (i) comparing a description column with the alternative of including the full description for each attribute in both columns; (ii) comparing highlighting attributes that are different between options ('highlight differences') with highlighting all attribute levels with different intensities ('highlight intensity'); and (iii) comparing a 'two-stage' approach with a 'triplet' best-worst approach. A breakdown of the different versions tested is provided in **Appendix Table 1**.

**Appendix Table 1**: Survey iterations used in the think-aloud interviews

| Version | Mode of Delivery | | Highlighting | | Approach | |
|---|---|---|---|---|---|---|
| | Paper-based | Online | Differences | Intensity | Two-stage | Triplet |
| Version 1 | ✓ | | ✓ | | ✓ | ✓ |
| Version 2 | ✓ | ✓ | ✓ | | ✓ | ✓ |
| Version 3 | | ✓ | | ✓ | ✓ | ✓ |
| Version 4 | | ✓ | | ✓ | ✓ | |
| Version 5 | | ✓ | ✓ | ✓ | ✓ | |

Testing was conducted using think-aloud interviews. This process involved asking participants to verbalize their thoughts and decision-making process as they completed the respective survey version. All interviews were conducted by a single researcher using pre-defined instructions and probes. There were 16 think-aloud participants, with interviews lasting between 45 and 60 minutes. **Appendix Table 2** shows the demographic characteristics of the participants. The first seven interviews were conducted with the paper-based version of the survey. Changes were made iteratively, based on feedback from completed interviews. The final nine interviews were conducted with a computerized version of the survey.

Although there were benefits associated with a paper-based survey for some participants, we determined that these benefits did not outweigh the lower response rate that would be achieved with a paper-based survey compared with an online survey. This decision also took into account the project resources available.

Appendix 1, as supplied by the authors. Appendix to: Bansback N, Trenaman L, Mulhern BJ, et al. Estimation of a Canadian preference-based scoring algorithm for the Veterans RAND 12-Item Health Survey: a population survey using a discrete choice experiment. *CMAJ Open* 2022. doi:10.9778/cmajo.20210113. Copyright © 2022 The Author(s) or their employer(s). To receive this resource in an accessible format, please contact us at cmajgroup@cmaj.ca

**Appendix Table 2**: Think-aloud interview sample characteristics (n=16). Values are numbers (percentages) unless stated otherwise

| Characteristics | |
|---|---|
| Gender | |
| Female | 14 (87.5) |
| Male | 2 (12.5) |
| Other | 0 (0) |
| Age, years: mean (range) | 39.3 (20-72) |
| Education | |
| Some high school, but did not graduate | 1 (6.3) |
| High School or high school equivalency certificate | 2 (12.5) |
| College, CEGEP or other non-university certificate or diploma | 3 (18.8) |
| Some university or undergraduate degree | 7 (43.8) |
| Post-graduate or professional degree | 2 (12.5) |
| Prefer not to say | 1 (6.3) |
| Ethnicity* | |
| Caucasian or White | 12 (75.0) |
| Chinese | 3 (18.8) |
| Middle Eastern | 1 (6.3) |
| Prefer not to say | 1 (6.3) |
| Self-Reported Mental Health | |
| Excellent | 2 (12.5) |
| Very Good | 5 (31.3) |
| Good | 7 (43.8) |
| Fair | 2 (12.5) |
| Poor | 0 (0.0) |
| Self-Reported Physical Health | |
| Excellent | 2 (12.5) |
| Very Good | 5 (31.3) |
| Good | 4 (25.0) |
| Fair | 5 (31.3) |
| Poor | 0 (0.0) |

*Percentages do not add to 100% because participants could select more than one response option.

Appendix 1, as supplied by the authors. Appendix to: Bansback N, Trenaman L, Mulhern BJ, et al. Estimation of a Canadian preference-based scoring algorithm for the Veterans RAND 12-Item Health Survey: a population survey using a discrete choice experiment. *CMAJ Open* 2022. doi:10.9778/cmajo.20210113. Copyright © 2022 The Author(s) or their employer(s). To receive this resource in an accessible format, please contact us at cmajgroup@cmaj.ca

Feedback from the interviews also resulted in the following survey design decisions: (i) a maximum of 10 DCE questions per respondent; (ii) the inclusion of an attribute description column, with shortened phrasing to describe attribute levels; (iii) allowing levels for five attributes only to vary for the first eight DCE questions (all eight attributes vary for the remaining two questions to meet DCE requirements that all attributes are traded by each individual), and position these five attributes at the top of the choice set, under the duration attribute; (iv) highlighting the levels that differ between the two presented health states using different intensities (i.e., shading); (v) maintaining a single order of attributes for all ten DCE questions at the individual-level; and (vi) including an interactive warm-up task that begins with three attributes, moves on to five attributes, then all nine attributes, with pop-ups providing feedback on the selections and a description of the trade-offs that individuals had made. The impact of decisions (ii), (iii) and (iv) on the presentation of the task can be seen in Figure 1.

***Experimental Design***

Using these design decisions, we used experimental design theory to select the levels for each attribute that would be incorporated into each choice set. We considered that potentially 121 separate parameters might need to be estimated from the choices, with 30 main effects for health attributes (7 [attributes] x (5-1) [with 5 levels] + 1 [attribute] x (3-1) [with 3 levels] = 30), for 4 duration interactions ((30x4) = 120), and 1 continuous duration main effect. We decided to generate 200 choice sets in total, more than the number of parameters needed to estimate, but given restrictions imposed by our design decisions, this limited the inefficiency compared to an unrestricted design. A D-Optimal design was generated using a Modified Federov algorithm in NGene [1]. The choice sets were randomly 'blocked' into 20 blocks of 10, with respondents randomly allocated to a block. To overcome potential order effects, attributes within a block were randomly ordered (but remained in the same order for each question in the block).

***Pilot testing***

One hundred-three individuals were recruited using Amazon Mechanical Turk. **Appendix Table 3** shows the demographic characteristics of the participants. In the beta version of the survey (see Section 2.3 in the paper), all DCE questions required a choice. At the end of the survey, a free text question asked participants about their levels of ease with the navigation and understanding of the survey. Pilot testing did not identify any usability or navigation issues. However, the free-text feedback provided some suggestions for improving the wording of the instructions – for the warm-up exercise, in particular. A revised version of the survey was developed and translated into French. On the entry page to the survey, participants could select which version they wanted to complete. Official translations of the VR-12 were used; other components of the survey were forward and backward translated.

***Statistical analysis***

Inconsistent coefficients are defined as when an increase in severity (defined by the naturally monotonic levels of the VR-12 items) leads to an increase rather than a decrease in utility. We followed accepted practice by examining patterns of inconsistency, and imposed consistency on estimates by combining adjacent levels with disordered coefficients and re-estimating the model where differences were not statistically significant [2]. We chose which estimates to combine by choosing those that best improved model fit.

Appendix 1, as supplied by the authors. Appendix to: Bansback N, Trenaman L, Mulhern BJ, et al. Estimation of a Canadian preference-based scoring algorithm for the Veterans RAND 12-Item Health Survey: a population survey using a discrete choice experiment. *CMAJ Open* 2022. doi:10.9778/cmajo.20210113. Copyright © 2022 The Author(s) or their employer(s). To receive this resource in an accessible format, please contact us at cmajgroup@cmaj.ca

**Appendix Table 3**: Characteristics of participants in the pilot testing of the survey (n=103). Values are numbers (percentages) unless stated otherwise

| **Characteristics** | |
|---|---|
| Gender | |
| Female | 58 (56%) |
| Male | 45 (44%) |
| Other | 0 (0%) |
| Age, years: mean (range) | 35.46 (22-66) |
| Education | |
| Some high school, but did not graduate | 0 (0%) |
| High School or high school equivalency certificate | 19 (18%) |
| College, CEGEP or other non-university certificate or diploma | 23 (22%) |
| Some university or undergraduate degree | 35 (38%) |
| Post-graduate or professional degree | 10 (6%) |
| Prefer not to say | 0 (0%) |
| Ethnicity* | |
| Caucasian or White | 82 (80%) |
| Black | 15 (15%) |
| Chinese | 3 (3%) |
| Latin American | 3 (3%) |
| Indigenous | 2 (2%) |
| Korean | 1 (1%) |
| Southeast Asian | 1 (1%) |
| South Asian | 1 (1%) |
| Prefer not to say | 0 (0%) |
| Self-Reported Health Conditions* | |
| Back Pain | 23 (22%) |
| High blood pressure | 11 (11%) |
| Anemia | 4 (4%) |
| Depression | 7 (7%) |
| Osteoarthritis | 7 (7%) |
| Cancer | 5 (5%) |
| Rheumatoid Arthritis | 4 (4%) |
| Diabetes | 4 (4%) |
| Lung disease | 1 (1%) |

*Percentages do not add to 100% because participants could select more than one response option.

Appendix 1, as supplied by the authors. Appendix to: Bansback N, Trenaman L, Mulhern BJ, et al. Estimation of a Canadian preference-based scoring algorithm for the Veterans RAND 12-Item Health Survey: a population survey using a discrete choice experiment. *CMAJ Open* 2022. doi:10.9778/cmajo.20210113. Copyright © 2022 The Author(s) or their employer(s). To receive this resource in an accessible format, please contact us at cmajgroup@cmaj.ca

### *Weighting*

A combination of the 2016 Census and 2015/16 Canadian Community Health Survey (CCHS) was used as the population standard for weighting [3,4]. An interactive proportional fitting method, also known as raking, was implemented using the SAS Raking Macro – Generation IV [5]. The aim of this method is to match the distribution of each variable in the Internet sample to the known population distribution based on Statistics Canada data in terms of gender, region, education, age group, and health status. Age group and health status were nested and were the only two variables taken from the CCHS. The raking procedure was iterated until there was close agreement in the distribution of variables in the study sample and the known population distribution. The weighting was integrated into the final model using the pweights function in stata.

### *Statistical Code*

The key code for the analysis is below

'model 1 – conditional logit – each parameter interaction with lifeyear duration in profile

clogit choice ly lypf1 lypf2 lyrp1 lyrp2 lyrp3 lyrp4 lyre1 lyre2 lyre3 lyre4 lybp1 lybp2 lybp3 lybp4 lymp1 lymp2 lymp3 lymp4 lyvt1 lyvt2 lyvt3 lyvt4 lymd1 lymd2 lymd3 lymd4 lysf1 lysf2 lysf3 lysf4, group(pairm)

'model 2 – levels combined

clogit choice ly lypf1 lypf2 lyrp123 lyrp4 lyre123 lyre4 lybp12 lybp3 lybp4 lymp2 lymp3 lymp4 lyvt1 lyvt23 lyvt4 lymd12 lymd3 lymd4 lysf2 lysf3 lysf4, group(pairm)

'model 3 – model 2 but exclude same choice

clogit choice ly lypf1 lypf2 lyrp123 lyrp4 lyre123 lyre4 lybp12 lybp3 lybp4 lymp2 lymp3 lymp4 lyvt1 lyvt23 lyvt4 lymd12 lymd3 lymd4 lysf2 lysf3 lysf4 & same!=1, group(pairm)

'model 4 – model 2 but exclude inconsistent responses

clogit choice ly lypf1 lypf2 lyrp123 lyrp4 lyre123 lyre4 lybp12 lybp3 lybp4 lymp2 lymp3 lymp4 lyvt1 lyvt23 lyvt4 lymd12 lymd3 lymd4 lysf2 lysf3 lysf4 & inconsistent!=1, group(pairm)

'model 5 – model 2 but exclude those completing too quickly.

clogit choice ly lypf1 lypf2 lyrp123 lyrp4 lyre123 lyre4 lybp12 lybp3 lybp4 lymp2 lymp3 lymp4 lyvt1 lyvt23 lyvt4 lymd12 lymd3 lymd4 lysf2 lysf3 lysf4 & time!=1, group(pairm)

'model 6 – model 2 weighted results

clogit choice ly lypf1 lypf2 lyrp123 lyrp4 lyre123 lyre4 lybp12 lybp3 lybp4 lymp2 lymp3 lymp4 lyvt1 lyvt23 lyvt4 lymd12 lymd3 lymd4 lysf2 lysf3 lysf4 [pweight=weight], group(pairm)

'calculate utility values for models

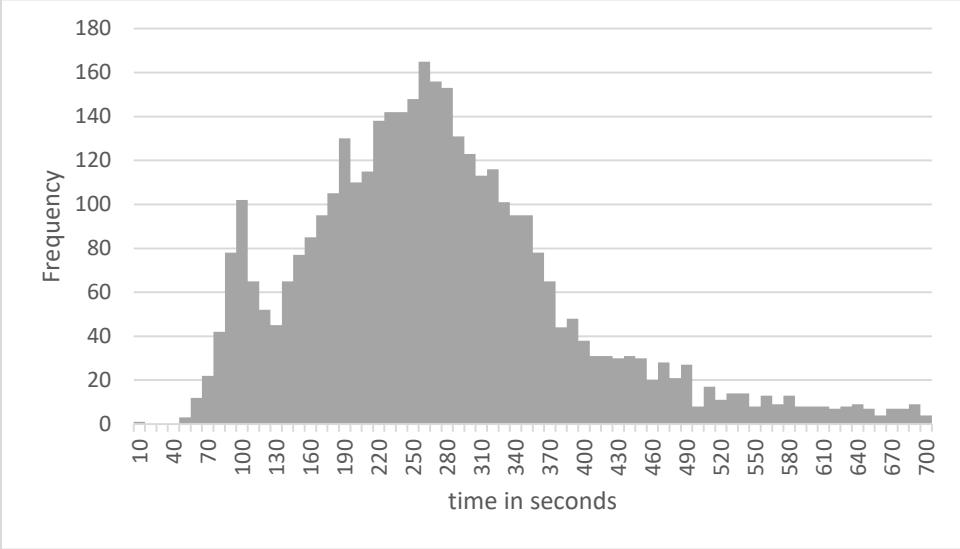
wtp ly lypf1 lypf2 lyrp123 lyrp4 lyre123 lyre4 lybp12 lybp3 lybp4 lymp2 lymp3 lymp4 lyvt1 lyvt23 lyvt4 lymd12 lymd3 lymd4 lysf2 lysf3 lysf4

Appendix 1, as supplied by the authors. Appendix to: Bansback N, Trenaman L, Mulhern BJ, et al. Estimation of a Canadian preference-based scoring algorithm for the Veterans RAND 12-Item Health Survey: a population survey using a discrete choice experiment. *CMAJ Open* 2022. doi:10.9778/cmajo.20210113. Copyright © 2022 The Author(s) or their employer(s). To receive this resource in an accessible format, please contact us at cmajgroup@cmaj.ca

***Selected additional Results***

The average time taken to complete the 10 DCE questions was 4 minutes 36 seconds. The distribution of times is shown below.

**Appendix Figure 1:** Time taken to complete DCE questions




The anchored coefficients for each of the models described in Table 2 in the main paper are shown in Appendix Table 4.

An analysis of the data using a mixed logit model (model 7) which enables certain coefficients to be randomly distributed across individuals, enabling exploration of preference heterogeneity [6]. As can be seen in Appendix Figure 2, This introduces more variation around the values, but distributed around the conditional logit model, and so does not alter the main values.

Appendix 1, as supplied by the authors. Appendix to: Bansback N, Trenaman L, Mulhern BJ, et al. Estimation of a Canadian preference-based scoring algorithm for the Veterans RAND 12-Item Health Survey: a population survey using a discrete choice experiment. *CMAJ Open* 2022. doi:10.9778/cmajo.20210113. Copyright © 2022 The Author(s) or their employer(s). To receive this resource in an accessible format, please contact us at cmajgroup@cmaj.ca

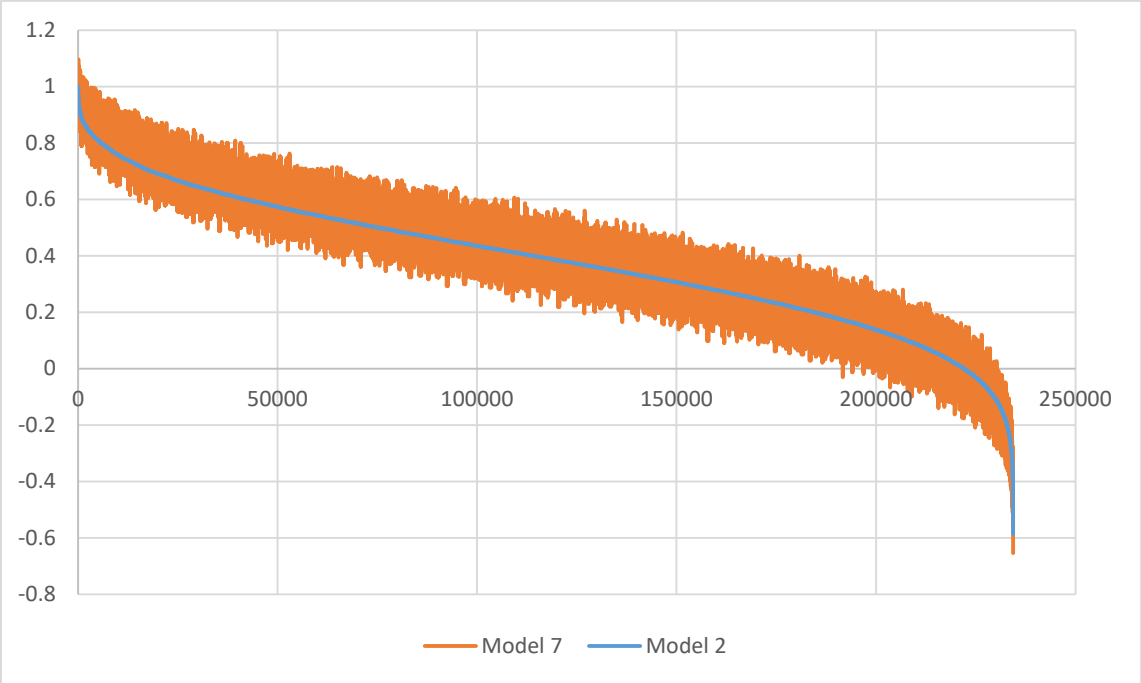
**Appendix Table 4:** Anchored disutility values for each model

| Attribute | Level | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---|---|---|---|---|---|---|---|
| Physical Functioning (PF) | PF1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | PF2 | -0.059 | -0.052 | -0.052 | -0.063 | -0.056 | -0.052 |
| | PF3 | -0.173 | -0.168 | -0.168 | -0.175 | -0.169 | -0.161 |
| Role Physical (RP) | RP1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | RP2 | 0.016 | -0.024 | -0.024 | -0.026 | -0.021 | -0.010 |
| | RP3 | 0.019 | -0.024 | -0.024 | -0.026 | -0.021 | -0.010 |
| | RP4 | -0.076 | -0.024 | -0.024 | -0.026 | -0.021 | -0.010 |
| | RP5 | -0.102 | -0.112 | -0.112 | -0.110 | -0.113 | -0.111 |
| Role Emotional (RE) | RE1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | RE2 | -0.034 | -0.013 | -0.013 | -0.014 | -0.011 | -0.019 |
| | RE3 | 0.029 | -0.013 | -0.013 | -0.014 | -0.011 | -0.019 |
| | RE4 | -0.026 | -0.013 | -0.013 | -0.014 | -0.011 | -0.019 |
| | RE5 | -0.103 | -0.111 | -0.111 | -0.097 | -0.114 | -0.113 |
| Bodily Pain (BP) | BP1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | BP2 | -0.061 | -0.057 | -0.057 | -0.065 | -0.060 | -0.040 |
| | BP3 | -0.060 | -0.057 | -0.057 | -0.065 | -0.060 | -0.040 |
| | BP4 | -0.209 | -0.194 | -0.194 | -0.198 | -0.199 | -0.187 |
| | BP5 | -0.275 | -0.275 | -0.275 | -0.278 | -0.273 | -0.272 |
| Mental Health - Anxiety (MA) | MA1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | MA2 | -0.003 | -0.039 | -0.039 | -0.030 | -0.038 | -0.040 |
| | MA3 | -0.042 | -0.039 | -0.039 | -0.030 | -0.038 | -0.040 |
| | MA4 | -0.123 | -0.121 | -0.121 | -0.117 | -0.114 | -0.130 |
| | MA5 | -0.242 | -0.237 | -0.237 | -0.234 | -0.237 | -0.237 |
| Mental Health - Depression (MD) | MD1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | MD2 | -0.040 | -0.038 | -0.038 | -0.030 | -0.041 | -0.048 |
| | MD3 | -0.093 | -0.071 | -0.071 | -0.066 | -0.071 | -0.081 |
| | MD4 | -0.047 | -0.071 | -0.071 | -0.066 | -0.071 | -0.081 |
| | MD5 | -0.175 | -0.170 | -0.170 | -0.167 | -0.174 | -0.179 |
| Vitality (VT) | VT1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | VT2 | -0.042 | -0.052 | -0.052 | -0.057 | -0.056 | -0.061 |
| | VT3 | -0.035 | -0.052 | -0.052 | -0.057 | -0.056 | -0.061 |
| | VT4 | -0.244 | -0.240 | -0.240 | -0.240 | -0.247 | -0.249 |
| | VT5 | -0.313 | -0.314 | -0.314 | -0.315 | -0.316 | -0.326 |
| Social Functioning (SF) | SF1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | SF2 | 0.012 | -0.045 | -0.045 | -0.040 | -0.042 | -0.053 |
| | SF3 | -0.038 | -0.045 | -0.045 | -0.040 | -0.042 | -0.053 |
| | SF4 | -0.148 | -0.153 | -0.153 | -0.153 | -0.155 | -0.143 |
| | SF5 | -0.181 | -0.199 | -0.199 | -0.194 | -0.198 | -0.191 |

*Reported to three decimal places.*

Appendix 1, as supplied by the authors. Appendix to: Bansback N, Trenaman L, Mulhern BJ, et al. Estimation of a Canadian preference-based scoring algorithm for the Veterans RAND 12-Item Health Survey: a population survey using a discrete choice experiment. *CMAJ Open* 2022. doi:10.9778/cmajo.20210113. Copyright © 2022 The Author(s) or their employer(s). To receive this resource in an accessible format, please contact us at cmajgroup@cmaj.ca

**Appendix Figure 2:** Mixed Logit Model (model 7) compared to conditional logit model (model 2)




---

**REFERENCES**


1. ChoiceMetrics. Ngene User Manual and Reference Guide. Sydney, New South Wales, Australia: ChoiceMetrics; 2012.
2. Brazier J, Roberts J, Rowan D. Methods for developing preference-based measures of health. In: The Elgar Companion to Health Economics, Second Edition. Edward Elgar Publishing; 2012.
3. Statistics Canada. Language Highlight Tables: 2016 Census of Population [Internet]. 2017. Available from: http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/lang/index-eng.cfm (accessed April 06, 2021)
4. Statistics Canada. Canadian Community Health Survey (2015/16). Available from: https://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=assembleInstr&Item_Id=260675 (accessed April 06,2021)
5. Izrael D, Battaglia M, Battaglia A, Ball S. SAS Raking Macro – Generation IV [Internet]. 2017. Available from: www.abtassociates.com/raking-survey-data-aka-sample-balancing (accessed April 06, 2021)
6. Hensher DA, Greene WH. The mixed logit model: the state of practice. Transportation. 2003 May;30(2):133-76.



Appendix 1, as supplied by the authors. Appendix to: Bansback N, Trenaman L, Mulhern BJ, et al. Estimation of a Canadian preference-based scoring algorithm for the Veterans RAND 12-Item Health Survey: a population survey using a discrete choice experiment. *CMAJ Open* 2022. doi:10.9778/cmajo.20210113. Copyright © 2022 The Author(s) or their employer(s). To receive this resource in an accessible format, please contact us at cmajgroup@cmaj.ca