Catalogue no. 82-003-X
ISSN 1209-1367

Health Reports

Mapping the Washington Group on Disability Statistics disability measure to the Health Utilities Index Mark 3: Development and validation of a predictive multivariable model in a general population sample

by Thomas Charters, Dafna Kohen and Julie Bernier

Release date: January 15, 2025

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Mapping the Washington Group on Disability Statistics disability measure to the Health Utilities Index Mark 3: Development and validation of a predictive multivariable model in a general population sample

by Thomas Charters, Dafna Kohen and Julie Bernier

DOI: https://www.doi.org/10.25318/82-003-x202500100001-eng

ABSTRACT

Background

Statistics Canada routinely collects information on functional health and related concepts. Recently, the Washington Group on Disability Statistics (WG) measure of disability has been introduced to the Canadian Community Health Survey (CCHS). The WG measure is used as a tool for developing internationally comparable data on disability. In alternate cycles of the CCHS, it replaces the Health Utilities Index Mark 3 (HUI3), a generic preference-based measure of health-related quality of life. The HUI3 is used to derive evaluative health measures common in population health and economic evaluations. Since the WG measure is not preference-based, it is unable to derive these measures. To address resulting data gaps, this study empirically maps the health state utility values of the HUI3 score from the WG measure.

Data and methods

Empirical mapping used a "head-to-head" subsample of the 2017 CCHS where WG and HUI3 measures were collected from the same respondents aged 40 and over. Empirical mapping used regression models to estimate the statistical relationship between WG and HUI3 measures in addition to health and demographic variables. Out-of-sample predictive performance was assessed through descriptive statistics, mean absolute error, and other measures of predictive accuracy.

Results

The preferred estimation strategy resulted in reasonably precise estimates of the HUI3 score corresponding to trends across health and demographic characteristics and reflecting distributional properties of the HUI3 score. Inclusion of different components of the WG measure influenced predictive accuracy.

Interpretation

Empirical mapping offers a potential method to estimate health state utility scores from the WG measure and addresses data gaps in health-related quality of life measures in the CCHS when HUI3 is not collected.

Keywords

Adult; quality of life; data interpretation, statistical; models, theoretical; health surveys/statistics and numerical data; Canada

AUTHORS

Thomas Charters and Dafna Kohen are with the Health Analysis Division at Statistics Canada. Julie Bernier was formerly with the Health Analysis Division.

What is already known on this subject?

- The Health Utilities Index Mark 3 (HUI3) has been included in various health and social surveys from Statistics Canada for several decades. Health status information derived using the HUI3 has been used extensively in economic and population health analyses.
- The Washington Group on Disability Statistics (WG) measure of disability cannot be used to generate utility scores necessary for common evaluative health measures since it does not have preference-based scoring functions, yet it is useful as an internationally comparable measure.
- Empirical mapping provides a common method to predict health state utility values from measures that lack preference-based scoring functions but show sufficient conceptual overlap.
- Previous work has conceptually mapped the WG to HUI3 measure and found sufficient overlap in the attributes and measurement approach between the measures to justify empirical mapping.

What does this study add?

- Empirical mapping provides a feasible means to estimate health state utility scores of the HUI3 from the WG measure. This approach is verified through tests of predictive accuracy and validation across sample characteristics.
- A predictive formula to estimate health state utility scores from the WG measure in addition to health and demographic variables in the core content of the CCHS.
- Results from this study provide a means to derive and validate health-related quality of life measures in the general Canadian population aged 40 years and over such as Quality Adjusted Life Years (QALYs) and Health-Adjusted Life Expectancy (HALE).

Population health surveys commonly collect information on health status as represented by functional abilities. Questions assess ability levels of respondents carrying out various tasks or activities in addition to health states that may impede this functioning. Disability is a related concept, involving interactions between these elements of functional status and environmental factors that limit or restrict participation in society.[1,2] The Washington Group on Disability Statistics (WG) measure of disability was developed by an international consortium and sponsored by the United Nations Statistical Commission. The purpose of the WG was to develop an internationally comparable population-based measure of disability to be used in censuses or national surveys, through measurement of functional limitations across domains closely associated with social participation.[2] To facilitate comparability across different countries and cultural contexts, the WG measure assesses functional health through difficulties with universal basic activities. While the WG measures were developed within the International Classification of Functioning, Disability and Health framework,[3,4] they do not include social or environmental factors implicit to this framework for reasons of brevity and comparability. The WG measures are intended to be used in conjunction with other information sources to highlight inequalities between limitations in health and functioning and social inclusion, and thereby targets for intervention as per the United Nations 2030 Sustainable Development Goals.[5-7] The validity and reliability of the WG measure have been demonstrated in international contexts,[3,8] and the WG measure has been adopted in censuses or surveys in over 80 countries.[6]

Several surveys and measures have historically been used by Statistics Canada to estimate levels of disability through measures of impairment, functional health, or activity limitations. Among these, the Health Utilities Index® Mark 3[9] has been incorporated into several health and social surveys for several decades.[10] The HUI system was developed to provide a standardized measure to assess and compare health and health-related quality of life (HRQoL) in patient groups and the general population, and in evaluation of health interventions.[11] Further, the HUI3 has been used to derive common evaluative health measures, such as health-adjusted life expectancy (HALE)[12] and quality-adjusted life years (QALYs), commonly used measures in population health and economic evaluations.[13] The validity, reliability, and responsiveness of the HUI3 system are well established in clinical and population health settings.[9,14-16]

Since 2000, the Canadian Community Health Survey (CCHS) has been administered by Statistics Canada to provide comprehensive health information on the Canadian population.[17] The HUI3 instrument has been included in the CCHS since its inception. In 2015, the CCHS underwent a major redesign which saw updates to its content, sampling methods, and administration.[18] Following the redesign, the HUI3 and the questions from the Washington Group Short Set on Functioning (WG-SS) were included as part of a two-year theme content, collected in alternate cycles to optimize data

collection. Inclusion of the WG-SS measure meets commitments for collection of internationally integrated data on disability.[6] Both measures describe functional capacities (what can you do) intrinsic to the person (within or near the skin) rather than performance (what you do) to avoid influence by context-dependent environmental factors.[19,20] Unlike the HUI3, the WG measure is unable to generate health state utility values, since these are derived from a preference-based scoring function.[21,22] As such, the WG does not permit calculation of HALEs or QALYs.

Collection of the HUI3 in alternating years will lead to data gaps. While the WG and HUI3 measures play complementary roles in the measurement of functional health, the WG measure is not suitable for use in the calculation of important health measures used in population health and program evaluations. Mapping provides a potential solution to estimate health state utility values from the WG measure. Mapping involves the estimation of a relationship between a target measure (HUI3) and a source measure (WG).[21] The relationship may be estimated through use of a statistical model or algorithm[23] or through equating or linking equivalent values between instruments.[24] Importantly, the validity and feasibility of mapping rely on sufficient conceptual overlap between the measures.[21,25,26] Mapping studies are common[21,27] and include several examples of successfully mapping the HUI from other measures.[28-33] Estimation of the HUI3 health state utility score would alleviate data gaps in years when not collected and may

Table 1
Descriptive characteristics of head-to-head sample and general sample of adult noninstitutionalized population aged 40 years and over, Canada 2017

	Rapid Response subsample[1]			Annual CCHS sample aged 40 years and over[2]		
		95% confidence interval			95% confidence interval	
	Percentage	from	to	Percentage	from	to
Sex						
Male	48.3	45.3	51.3	48.6	48.3	48.9
Female	51.7	48.7	54.7	51.4	51.1	51.7
Age						
40 to 49 years	25.9	22.9	28.9	25.8	25.4	26.2
50 to 59 years	28.1	25.2	31.1	28.1	27.5	28.7
60 to 69 years	23.3	20.9	25.8	24.7	24.0	25.4
70 to 79 years	16.7	14.6	18.8	14.5	14.1	14.9
80 years or over	5.9	4.8	7.0	6.9	6.6	7.3
Educational attainment						
Less than high school graduation	12.2	10.3	14.0	14.3	13.7	14.8
Secondary school graduation	21.9	19.5	24.3	22.4	21.7	23.1
Postsecondary certificate or degree	63.2	60.3	66.1	61.1	60.4	61.9
Missing	2.8	1.5	4.0 E	2.2	1.9	2.5
Marital status						
Married or common law	73.8	71.3	76.3	70.7	69.8	71.6
Widowed, separated, divorced or never married	26.1	23.6	28.5	29.2	28.2	30.1
Missing	F	...	...	0.1	0.1	0.2 E
Self-rated health						
Poor	2.7	1.9	3.4	4.6	4.3	5.0
Fair	9.2	7.6	10.8	10.4	9.9	10.9
Good	30.6	27.8	33.5	30.9	30.1	31.7
Very Good	36.9	34.1	39.8	33.7	32.9	34.4
Excellent	20.5	18.0	23.0	20.2	19.5	20.9
Not stated	F	...	...	0.2	0.1	0.2 E
Self-rated mental health						
Poor	0.8	0.4	1.2 E	1.4	1.2	1.6
Fair	4.1	3.0	5.2	4.9	4.6	5.3
Good	23.3	20.5	26.1	22.0	21.3	22.7
Very Good	38.0	35.0	41.0	36.3	35.4	37.2
Excellent	31.8	28.8	34.8	31.4	30.6	32.2
Not stated	2.0	1.3	2.7 E	4.0	3.7	4.3
Number of chronic conditions						
None	24.7	22.2	27.2	26.0	25.3	26.7
One	30.0	27.1	33.0	26.8	26.1	27.6
Two	16.8	14.6	19.1	18.6	18.0	19.3
Three or more	25.5	22.9	28.1	24.9	24.2	25.6
Missing	3.0	1.9	4.0 E	3.7	3.4	4.0

... not applicable
E use with caution
F too unreliable to be published
[1] N=2,597; Weighted N=1,247,600
[2] N=37,609; Weighted N=18,148,700
Note: CCHS = Canadian Community Health Survey
Source: 2017 Canadian Community Health Survey annual survey and Rapid Response subsample.

optimize resources used in data collection. The purpose of this study is to empirically map the health state utility values of the HUI3 score from the WG measure and to validate the results. This report builds off of previous research, which has established necessary levels of conceptual overlap between these two measures (available upon request).[34]

Materials and methods

Data

The 2017 CCHS annual component was used in this study. The CCHS is a cross-sectional representative survey covering a range of topics relevant to the health status, health behaviours, and demographic profiles of the Canadian population aged 12 and over living in private dwellings. Persons living on Indian reserves, on Crown lands, in institutions, in remote regions, or serving in the Canadian Forces are excluded from the sample. Additionally, individuals residing in the territories are excluded from one-year survey cycles. Approximately 98% of Canadians aged 12 and over are represented in the CCHS.[35] Health, demographic, and socioeconomic variables collected from the 2017 CCHS included the respondents' sex and age, highest educational attainment, marital status, self-rated general and mental health, chronic conditions, and the WG-SS.

Mapping used a unique "head-to-head" subsample of the 2017 CCHS containing both WG and HUI3 measures from the same respondents aged 40 years and over. The subsample contains three additional variables from the Washington Group Extended Set on Functioning (WG ES-F) in addition to the multi-attribute health status classification system questionnaire and derived attribute-specific and overall scores for the HUI3. The CCHS Rapid Response file included 2,837 respondents who had been provided modules for both the HUI3 and WG ES-F, with 2,597 having non-missing responses to the HUI3 target measure.

Measures of functional health

All domains from the Washington Group Short-Set (WG-SS) were included in the core content of the annual 2017 CCHS. The WG-SS consists of six attributes, including vision, hearing, mobility, cognition, self-care, and communication. Each attribute was assessed by a single question containing four response options: "No difficulty," "Yes-some difficulty," "Yes-a lot of difficulty," and "Cannot do at all."[4] The Rapid Response subsample of the 2017 CCHS contained three additional attributes from the WG ES-F (pain, anxiety, depression). Response categories for each attribute were derived from one question measuring the frequency of the attribute and a second on its intensity.[3] All WG measures included an additional response category of "Not stated," combining response categories of "Don't know" and "Refusal." Appendix I shows derivation of categories for the WG ES-F.

The HUI3 multi-attribute classification system consists of eight attributes: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain, with five or six response options per attribute.[9,11] Preference-based scoring functions convert this

Table 2
Measure of agreement in observed and predicted categories of Health Utilities Index Mark 3 score (1, 0.973, less than 0.973)

	Ordered logistic regression					Multinomial logistic regression				
Model	Percentage HUI3 <0.973	Percentage HUI3= 0.973	Percentage HUI3= 1.00	Kappa	Percentage observed agreement	Percentage HUI3 <0.973	Percentage HUI3= 0.973	Percentage HUI3= 1.00	Kappa	Percentage observed agreement
HUI3	59.5	24.5	16.1	...	...	59.5	24.5	16.1	...	...
Mapped										
1	99.8	0.2	0.0	-0.003	59.2	99.7	0.2	0.1	-0.005	59.1
2	64.8	35.1	0.1	0.323	64.2	62.7	37.0	0.3	0.322	63.6
3	63.5	36.4	0.1	0.305	62.9	63.2	36.8	0.0	0.308	63.0
4	82.2	12.6	5.2	0.124	58.7	65.2	25.5	9.2	0.253	60.0
5	63.9	29.7	6.5	0.353	65.2	62.8	29.0	8.2	0.361	65.4
6	62.8	30.0	7.2	0.332	63.9	62.2	28.8	9.0	0.354	64.8
7	82.6	11.9	5.5	0.137	59.4	64.0	27.4	8.7	0.252	59.7
8	64.1	29.3	6.6	0.352	65.2	62.8	28.6	8.5	0.374	66.1
9	62.8	30.3	6.9	0.332	63.9	61.8	29.0	9.2	0.354	64.7
10	62.9	30.6	6.5	0.327	63.6	63.3	28.2	8.5	0.328	63.6
11	63.4	28.9	7.7	0.318	63.2	61.8	29.7	8.5	0.353	64.7
12	64.8	27.1	8.1	0.316	63.4	61.9	30.1	8.0	0.327	63.3

... not applicable

Note: WG-SS = Washington Group Short-Set questions; WG ES-F = Washington Group Extended Set of Functioning; HUI3=Health Utilities Index Mark 3

Model 1: WG-SS

Model 2: WG-SS, WG ES-F (pain, anxiety)

Model 3: WG-SS, WG ES-F (pain, depression)

Model 4: WG-SS, sex, age, age^2, age^3

Model 5: WG-SS, WG ES-F (pain, anxiety), sex, age, age^2, age^3

Model 6: WG-SS, WG ES-F (pain, depression), sex, age, age^2, age^3

Model 7: WG-SS, sex, age, age^2, age^3, marital status

Model 8: WG-SS, WG ES-F (pain, anxiety), sex, age, age^2, age^3, marital status

Model 9: WG-SS, WG ES-F (pain, depression), sex, age, age^2, age^3, marital status

Model 10: WG-SS, sex, age, age2, age3, marital status, general health, mental health

Model 11: WG-SS, WG ES-F (pain, anxiety), sex, age, age^2, age^3, marital status, general health, mental health

Model 12: WG-SS, WG ES-F (pain, depression), sex, age, age^2, age^3, marital status, general health, mental health

Source: 2017 Canadian Community Health Survey Rapid Response subsample

Table 3
Model performance of two-step empirical mapping to HUI3 with arcsine transformation and imputation of discrete Health Utilities Index Mark 3 scores using Washington Group Extended Set on Functioning

Model	Mean: Score	Mean: Lower 95% confidence interval	Mean: Upper 95% confidence interval	Minimum	25th percentile	50th percentile	75th percentile	Maximum	Mean absolute error	Root mean squared error	Kendall's rank coefficient	Percentage difference +/- 0.03 units or more	Model R^2
HUI3	0.848	0.828	0.869	-0.160	0.744	0.919	0.973	1.000	...	...	...	...	...
Mapped													
1	0.878	0.862	0.894	-0.223	0.816	0.922	0.973	1.000	0.095	0.172	0.459	60.8	0.423
2	0.875	0.857	0.893	-0.202	0.837	0.925	0.973	1.000	0.087	0.163	0.476	60.6	0.526
3	0.873	0.854	0.892	-0.291	0.839	0.934	0.973	1.000	0.086	0.161	0.484	57.9	0.539
4	0.878	0.863	0.894	-0.208	0.823	0.919	0.973	1.000	0.095	0.170	0.452	61.5	0.425
5	0.876	0.858	0.893	-0.194	0.825	0.930	0.973	1.000	0.086	0.163	0.479	59.4	0.533
6	0.874	0.855	0.892	-0.296	0.834	0.929	0.973	1.000	0.086	0.160	0.486	57.6	0.543
7	0.878	0.862	0.894	-0.208	0.823	0.919	0.973	1.000	0.095	0.170	0.452	61.4	0.425
8	0.876	0.859	0.894	-0.198	0.825	0.931	0.973	1.000	0.086	0.163	0.479	59.6	0.533
9	0.874	0.856	0.892	-0.298	0.835	0.929	0.973	1.000	0.086	0.160	0.486	57.8	0.543
10	0.879	0.863	0.895	-0.267	0.811	0.936	0.973	1.000	0.092	0.163	0.473	59.5	0.493
11	0.878	0.860	0.895	-0.215	0.824	0.943	0.973	1.000	0.087	0.160	0.487	57.2	0.577
12	0.876	0.859	0.894	-0.255	0.829	0.940	0.973	1.000	0.086	0.158	0.486	57.8	0.581

... not applicable

Note: WG-SS = Washington Group Short-Set on Functioning; WG ES-F = Washington Group Extended Set on Functioning; HUI3=Health Utilities Index Mark 3; CI= confidence interval; MAE=mean absolute error; Two-step method (1) imputes discrete HUI3 scores (1.00, 0.973) using predictions from multinomial model regressing on WG-SS, WG ES-F (pain and anxiety), age, age^2, age^3, sex, marital status (2) linear regression on arcsine transformed HUI3 score

Model 1: WG-SS

Model 2: WG-SS, WG ES-F (pain, anxiety)

Model 3: WG-SS, WG ES-F (pain, depression)

Model 4: WG-SS, sex, age, age^2, age^3

Model 5: WG-SS, WG ES-F (pain, anxiety), sex, age, age^2, age^3

Model 6: WG-SS, WG ES-F (pain, depression), sex, age, age^2, age^3

Model 7: WG-SS, sex, age, age^2, age^3, marital status

Model 8: WG-SS, WG ES-F (pain, anxiety), sex, age, age^2, age^3, marital status

Model 9: WG-SS, WG ES-F (pain, depression), sex, age, age^2, age^3, marital status

Model 10: WG-SS, sex, age, age2, age3, marital status, general health, mental health

Model 11: WG-SS, WG ES-F (pain, anxiety), sex, age, age^2, age^3, marital status, general health, mental health

Model 12: WG-SS, WG ES-F (pain, depression), sex, age, age^2, age^3, marital status, general health, mental health

Source: 2017 Canadian Community Health Survey Rapid Response subsample

descriptive information in utilities, cardinal scores describing preferences over various health states. The multi-attribute scoring function generates overall HUI3 utility scores through multiplicative models and describes up to 972,000 different health states. These scores range from -0.36 (implying a state worse than death) to 1.00 (perfect functional health) with 0.00 representing death.

Analytical techniques

Descriptive statistics of the CCHS Rapid Response subsample and overall CCHS population aged 40 years and over were estimated for key health, sociodemographic, and socioeconomic variables. Sampling weights were used to estimate sample means and standard errors were obtained through bootstrap repeated replications.

Regression techniques were used to estimate the statistical relationship between WG and HUI3 measures enabling prediction of the overall HUI3 score in the head-to-head subsample. The continuous HUI3 overall score (target measure) was regressed on categorical WG variables in addition to demographic, socioeconomic, and health variables of interest (source measures). Respondent health and demographic characteristics were included since these were anticipated to have associations to the HUI3 independent from that of the WG measures, and that their inclusion had potential to improve overall predictive accuracy. Characteristics were selected given their inclusion in the annual CCHS core content, having independent associations to the HUI3 score, and use in other mapping studies.[28-33] Missing data in categorical independent variables were permitted,[36] being treated as distinct item scores.

Responses to the WG questions were entered into the model as discrete dummy variables with the response "no difficulty" as the reference category. Use of item scores in regression models was chosen given the potential to improve model flexibility.[21] Three different sets of WG items were tested together: the WG-SS only (25 potential coefficients), and the two sets with both the WG-SS and WG ES-F, including measures of pain in both and, alternately, anxiety or depression for affect (33 potential coefficients each). Age, sex, marital status (married or common law; single, windowed, separated, or divorced; missing), and self-rated general health and mental health (for each: poor, fair, good, very good, excellent, not stated) were included in succession in 12 sets of prediction models. Self-rated health has been demonstrated to be a reliable measure,[37] positively correlated to physicians' health ratings,[38] chronic disease incidence,[39] and mortality.[40] Inclusion of self-rated health was selected given common use of additional clinical or health state measures in mapping studies.[21,27] Age was centred at 62, the unweighted mean age, to improve interpretability of coefficients. Consistent with previous studies,[28,41,42] age was

Table 4
Model performance of two-step empirical mapping to HUI3 with arcsine transformation and imputation of discrete HUI3 scores not using Washington Group Extended Set on Functioning

Model	Mean: score	Mean: Lower 95% confidence interval	Mean: Upper 95% confidence interval	Minimum	25th percentile	50th percentile	75th percentile	Maximum	Mean absolute error	Root mean squared error	Kendall's rank coefficient	Percentage difference +/- 0.03 units or more	Model R^2
HUI3	0.848	0.828	0.869	-0.160	0.744	0.919	0.973	1.000	…	…	…	…	…
Mapped													
1	0.882	0.866	0.897	-0.220	0.829	0.933	0.973	1.000	0.095	0.176	0.443	59.8	0.438
4	0.882	0.867	0.897	-0.208	0.834	0.928	0.973	1.000	0.094	0.174	0.434	60.1	0.441
7	0.883	0.867	0.898	-0.214	0.836	0.927	0.973	1.000	0.094	0.174	0.435	60.5	0.442
10	0.881	0.866	0.897	-0.274	0.825	0.936	0.973	1.000	0.096	0.169	0.439	61.9	0.500

… not applicable

Note: WG-SS = Washington Group Short-Set questions; WG ES-F = Washington Group Extended Set on Functioning; HUI3=Health Utilities Index Mark 3; CI= confidence interval; MAE= mean absolute error; Two-step method (1) imputes discrete HUI3 scores (1.00, 0.973) using predictions from multinomial model regressing on WG-SS, age, age^2, age^3, sex, marital status, self-rated general health, self-rated mental health (2) linear regression on arcsine transformed HUI3 score

Model 1: WG-SS

Model 4: WG-SS, sex, age, age^2, age^3

Model 7: WG-SS, sex, age, age^2, age^3, marital status

Model 10: WG-SS, sex, age, age^2, age^3, marital status, general health, mental health

Source: 2017 Canadian Community Health Survey Rapid Response subsample.

entered in linear, quadratic, and cubic forms, accounting for declines in functional health at advanced ages, assuming a nonlinear functional relationship. All other variables were entered into the model as categorical dummies.

Models fit on one set of data may predict characteristics randomly unique to that dataset and not necessarily show comparable predictive validity if replicated in other data sources. The head-to-head dataset was randomly partitioned into an "analytical" dataset making up two-thirds of the subsample (N=1,731) used to statistical predict HUI3 scores and a "hold-out" dataset with one-third of the data (N=866) used to assess the predictive accuracy. Models that accurately predict target measures in out-samples were expected to perform well in routine use and avoid problems intrinsic to overfitting.[36] In the "hold-out" dataset predictive accuracy was assessed through descriptive statistics of the predicted health state utility scores (mean, median, interquartile range, minimum, maximum) in addition to several forecast statistics, including mean absolute error (MAE), root mean squared error (RMSE), Kendall's rank correlation coefficient, and the model adjusted R^2 (from the analytical dataset). Proportions of predicted scores that differed from observed values by +/-0.03 units or more (the smallest change in HUI3 considered clinically important)[11] were calculated.

Characteristics of the HUI3 score presented challenges to empirical mapping. The population distribution of the overall HUI3 score is known to be highly skewed, with most respondents having perfect or near-perfect functional health.[15,42,43] Because of the skewed nature of the HUI3 distribution, the normality of residuals assumption may be violated through use of regression techniques during empirical mapping. Further, since the HUI3 score is calibrated between -0.36 and 1.00,[9,11] modelling must ensure predicted results fall within these theoretical bounds to be interpretable. Several different regression modelling methods and outcome transformations were explored to improve predictive accuracy, attain a more normal distribution of residuals, and better replicate the distributional properties of the HUI3 score (see[44] for the complete record of mapping procedures tested).

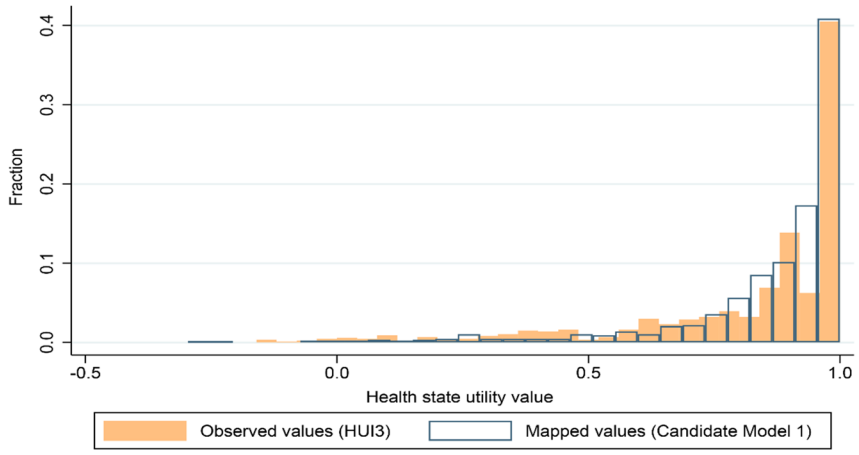
The validity of selected models was assessed by comparing the distributions of mean observed HUI3 scores and predicted scores across key respondent characteristics in the full CCHS Rapid Response subsample. Comparisons were made across age groups (40 to 49 years, 50 to 59 years, 60 to 69 years, 70 to 79 years, 80 years and over), sex, and presence of chronic conditions (no chronic conditions, one chronic condition, two chronic conditions, three or more chronic conditions, not stated). Accurate empirical mapping would imply little difference between observed and predicted variables in respect to key demographic and health characteristics.

Results

Sample characteristics from the head-to-head subsample (N=2,597) are compared with those from the main survey (N=37,609) for respondents aged 40 and over (Table 1). The subsample had a mean age of 59, was 52% female, was mostly educated at a postsecondary level (63%), and mostly married or in a common-law relationship (74%). Respondents were most likely to report "very good" self-perceived health (37%), "very good" self-perceived mental health (38%), and to have one chronic condition (30%). Overall, demographic, socioeconomic, and health characteristics of the CCHS Rapid Response subsample corresponded closely to those in the full sample.

In total, nine modelling strategies differing by use of regression model, transformation, and estimation of highly prevalent discrete scores were tested on 12 sets of covariates. No prespecified criteria determined model success. Two broad observations were drawn from model testing, which led to the selection of the final set of candidate models. First, regression methods led to mapped estimates of reduced variability, which inadequately corresponded to the properties of the HUI3 distribution. Second, regression of the untransformed HUI3 score led to mapped scores, which were somewhat less accurate

Figure 1
Distribution of observed and predicted health state utility values



Note: HUI3 = Health Utilities Index Mark 3

Source: Statistics Canada, 2017 Canadian Community Health Survey Rapid Response subsample

pand tended to exceed the theoretical upper bound of the HUI3 score (for more details on the modelling procedures and results from model testing, see[44]).

The candidate models addressed these limitations by applying a two-step procedure to empirically map the HUI3 score. First, ordinal or multinomial logistic regression was used to predict highly prevalent discrete scores of the HUI3 variable of 1.00, 0.973, and all scores less than0.973 on source measures through regressing on a three-category variable representing these scores. After deriving predicted probabilities for each category of the three-level variable, mapped categories were assigned based on each respondent's highest predicted probability. Table 2 shows that 16.1% of the sample had a HUI3 score of 1.00, 24.5% had a score of 0.973, and 59.5% had a score less than 0.973. Greater accuracy in predicting these categories was attained, with multinomial logistic regression in models controlling for the WG-SS, the WG ES-F (pain and anxiety), age, age^2, age^3, sex, and marital status (Model 8) having the highest agreement between categories (66%, kappa=0.374). The highest agreement in models not using the WG ES-F was found in Model 10 (64%, kappa=0.328), which additionally included self-rated general and mental health.

The next step regressed an arcsine transformed HUI3 score on source measures on the 63% of the "hold-out" sample projected to have HUI3 scores less than 0.973. The arcsine transformation takes the form: $arcsine\left[2*\left(\frac{HUI3+0.36}{1+0.36}\right)-1\right]$. which first binds the HUI3 score to the [-1, 1] interval necessary to facilitate transformation by the arcsine function. Mapped scores were derived from reverse-transforming predicted scores and otherwise imputing discrete scores of 1.00 or 0.973 based on projected categories from the first estimation step. Table 3 shows descriptive and forecast statistics for this two-step approach, whereby discrete scores of 1.00 and 0.973 were derived from Model 8 (Table 2) of the first estimation step. The HUI3 score in the hold-out sample had a mean of 0.848 (95% confidence interval=0.828, 0.869), a median and interquartile range of 0.919, 0.744 to 0.973, respectively, and a range of -0.16 to 1.00. Models including the WG ES-F in the second estimation step routinely performed better than those without, but indicated little difference based on selection of anxiety or depression for affect or from inclusion of other non-WG predictors. The greatest predictive accuracy based on MAE estimates was found to be 0.086, with slight improvements in forecast statistics favouring inclusion of depression. Mean predicted scores were generally higher than observed scores, although by less than the clinically important difference of 0.03. Predicted scores were constrained to the bounds of the HUI3 and aligned with the median and 75th percentiles while routinely overestimating the 25th percentile. To investigate the predictive accuracy of models not using WG ES-F measures, derivation of HUI3 categories used Model 10 covariates (Table 2). The lowest MAE was found to be 0.094, with little variation observed in predictive performance across models (Table 4).

To validate empirical mapping, scores of the predicted health state utility values in candidate models were compared to observed values of the HUI3 across key demographic and health factors. Selected models included Model 6 (Table 3), which included covariates for WG-SS, WG ES-F (pain and depression), age, age^2, age^3, and sex (Candidate Model 1); and Model 4 (Table 4), which included covariates for WG-SS, age, age^2, age^3, and sex (Candidate Model 2). Candidates were selected based on high predictive performance and exclusion of variables that did not improve predictive accuracy. Figure 1 shows the distribution of health state utility values from the HUI3 in comparison to those from Candidate Model 1. Table 5

shows the overall HUI3 score generally declined in older ages, a pattern replicated in predicted health state utility scores from both candidate models. Candidate models consistently overestimated the mean HUI3 score by age by amounts proximate to or less than the threshold for clinical significance (0.03). The distributions of predicted utility scores also followed those of sex, showing higher mean scores in men than women and overestimating scores to similar degrees. Both observed and mapped scores were highest among those with no chronic conditions and declined in those with three or more conditions or those missing this covariate.

Figure 2 shows a calibration plot between observed and mapped health state utility values from Candidate Model 1 in the hold-out dataset. Results show a calibration coefficient (slope=0.938), showing reasonably strong agreement between observed and predicted utility scores.[45,46] The statistic for calibration-in-the-large (CITL) of -0.039 reflects minor overestimation. The 95% confidence interval around this slope reflects less accurate model calibration at lower utility scores.

Discussion

Information on health status and quality of life obtained from the HUI3 system plays important roles in economic, clinical, and population health analysis in Canada. While the adoption of the WG measure permits collection on a validated and internationally comparable measure of functional capacity, the lack of preference-based scoring functions makes it unsuited for estimation of HRQoL, HALE, QALY, and other common measures. This data gap was addressed through empirical mapping of the overall HUI3 on the WG measure and select demographic and health characteristics in a CCHS subsample of the Canadian general population aged 40 years and over.

The head-to-head subsample provided a comparatively large and detailed dataset representative of the non-institutionalized Canadian population. The preferred model included two estimation steps: first, multinomial logistic regression to predict the HUI3 score falling in categories defined as 1, 0.973, or less than 0.973, and a second step predicting the HUI3 score through linear regression of the arcsine transformed score on respondents projected to have scores less than 0.973 or imputing discrete values of 1.00 or 0.973 based on the first estimation step. Arcsine transformation has been shown to improve the distribution of residuals in regression modelling and to maintain prediction within the theoretical bounds of the HUI3 score.[42] This estimation strategy was able to predict reasonably precise health state utility scores corresponded to trends across demographic and health characteristics, reflected the distributional properties of the skewed HUI3 score, and retained prediction to its theoretical bounds. Mapping using the WG-SS and WG ES-F resulted in a MAE in the candidate model of 0.086, about 6.3% of the total range of the HUI3 health state utility score (1.36), and exceeds the predictive accuracy of many mapping studies.[21] While demonstrating group-level predictive accuracy, about 60% of the sample had mapped scores exceeding the minimum clinically important difference of 0.03, implying difficulties in mapping at the individual level.

Empirical mapping using only the short-set attributes of the WG measure was also able to generate reasonable measures in this population group, although with less predictive accuracy (MAE=0.094, or 6.9% of the overall HUI3 range). This finding may highlight the importance of conceptual overlap between the measures,[21,25,26] since the WG-SS alone may not adequately match the attributes of pain and emotion contained in the HUI3 system[34] and may not be adequately represented by routinely collected health data, such as self-rated general and mental health. This highlights potential benefits to extending the existing WG-SS portion, included biennially in the CCHS two-

Table 5
Comparison of HUI3 score and mapped values from candidate models over demographic and health characteristics

	HUI3			Candidate Model 1[1]						Candidate Model 2[2]					
		95% confidence interval			95% confidence interval			95% confidence interval			95% confidence interval			95% confidence interval	
	Mean	from	to	Mean	from	to	Difference	from	to	Mean	from	to	Difference	from	to
a) Age															
40 to 49 years	0.889	0.866	0.911	0.924	0.907	0.941	0.035	0.019	0.052	0.933	0.919	0.947	0.044	0.027	0.061
50 to 59 years	0.860	0.838	0.881	0.890	0.871	0.909	0.030	0.015	0.046	0.896	0.881	0.912	0.037	0.021	0.053
60 to 69 years	0.842	0.817	0.866	0.871	0.851	0.890	0.029	0.014	0.044	0.873	0.851	0.895	0.031	0.015	0.048
70 to 79 years	0.852	0.828	0.877	0.858	0.835	0.880	0.005	-0.010	0.020	0.867	0.847	0.888	0.015	-0.002	0.032
80 years or over	0.773	0.728	0.818	0.799	0.762	0.837	0.027	0.007	0.047	0.796	0.761	0.832	0.024	-0.006	0.053
b) Sex															
Male	0.871	0.855	0.887	0.892	0.878	0.906	0.021	0.010	0.031	0.901	0.887	0.915	0.030	0.018	0.041
Female	0.843	0.827	0.859	0.875	0.862	0.889	0.033	0.022	0.043	0.879	0.867	0.891	0.036	0.025	0.047
c) Chronic conditions															
None	0.933	0.922	0.944	0.954	0.947	0.961	0.021	0.012	0.030	0.952	0.945	0.959	0.019	0.008	0.030
One	0.890	0.870	0.910	0.928	0.917	0.939	0.038	0.022	0.053	0.920	0.907	0.934	0.030	0.016	0.045
Two	0.858	0.835	0.880	0.880	0.862	0.897	0.022	0.005	0.040	0.893	0.878	0.908	0.035	0.017	0.054
Three or more	0.755	0.729	0.781	0.780	0.755	0.806	0.025	0.008	0.043	0.805	0.783	0.826	0.050	0.031	0.068
Not stated	0.748	0.652	0.844	0.754	0.648	0.861	0.006	-0.029	0.041	0.769	0.648	0.890	0.021	-0.050	0.093

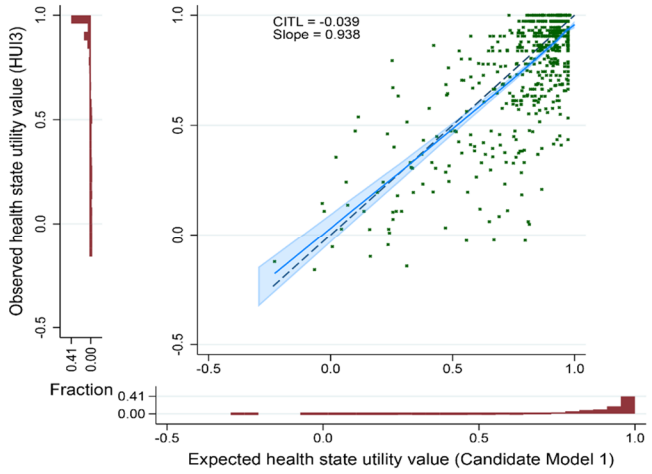
[1] Two-step method (1) imputes discrete HUI3 scores (1.00, 0.973) using predictions from multinomial model regressing on WG-SS, WG ES-F (pain and anxiety), age, age^2, age^3, sex, marital status (2) linear regression on arcsine transformed HUI3 score

[2] Two-step method (1) imputes discrete HUI3 scores (1.00, 0.973) using predictions from multinomial model regressing on WG-SS, age, age^2, age^3, sex, marital status, self-rated general health, self-rated mental health (2) linear regression on arcsine transformed HUI3 score

Note: N=2,597; Weighted N=1,247,600; HUI3=Health Utilities Index Mark 3; Difference represents difference in means from mapped - HUI3 score

Source: 2017 Canadian Community Health Survey Rapid Response subsample

Figure 2
Calibration of observed and predicted health state utility values



Note: Distribution of health state utility values by fraction (histogram). CITL = calibration in the large; HUI3 = Health Utilities Index Mark 3

Source: Statistics Canada, 2017 Canadian Community Health Survey Rapid Response subsample

year theme content, to include WG ES-F measures for pain and affect. Interestingly, inclusion of demographic and health covariates appeared more informative in modelling discrete categories of the HUI3 representative of perfect or near-perfect functional health than modelling lower functional health. Appendix II outlines regression coefficients and methods to map the HUI3 score for both candidate models.

Limitations of this study should be noted. Empirical mapping was tested and validated on a head-to-head sample of the Canadian population aged 40 years and over and is not generalizable to younger ages. The household population under 40 years generally has higher levels of functional health as measured by the HUI3,[12] and additional methods may be required to map the score to these groups. Further, some applicable categories of the WG group were absent in the head-to-head sample and may reduce reliability and replicability. Third, mapped health state utility scores were overestimated in the "hold-out" sample and across demographic categories. Generally, levels of error were greater at lower HUI3 health state utility scores, a similar finding to other mapping studies.[47] Next, conceptual limitations may arise if mapped scores are used in research comparisons across demographic characteristics included in the prediction equation. Assessment and prediction coefficients excluding sociodemographic factors for these uses will be presented elsewhere.[44] Finally, mapping functions were generated on a non-institutionalized general population sample and may not be appropriate for use in other population groups, such as patient data or respondents from institutional settings. Further work may incorporate methods[12] to adjust population-level mapped health state utility scores for institutionalized populations.

This study offers a potential method through empirical mapping to estimate health state utility scores from the WG measure and, as such, addresses data gaps in HRQoL measurement in the CCHS. Mapping was validated through comparisons of the distribution of the overall HUI3 score and selected mapped scores over key demographic and health characteristics. Mapped health state utility values may be used in future population studies of health-adjusted life expectancy (HALE) and quality-adjusted life years (QALYs), although further validation specific to these uses is required. Future work may further expand on mapping to the population aged less than 40 years.

Appendix I

Washington Group Extended Set on Functioning: Pain

How much pain you had last time you had pain	Frequency of pain in past three months				
	Never	Some days	Most days	Every day	Don't know
Not asked	(1)	…	…	…	(5)
A little	…	(1)	(2)	(2)	(5)
In between	…	(2)	(2)	(3)	(5)
A lot	…	(2)	(3)	(4)	(5)
Don't know	…	(5)	(5)	(5)	(5)

… not applicable

(1) Never had pain OR had a little pain some days

(2) Had pain every day (a little) OR had pain most days (a little OR in between) OR had pain some days (in between OR a lot)

(3) Had pain every day (in between) OR had pain most days (a lot)

(4) Had pain every day (a lot)

(5) Not stated

Washington Group Extended Set on Functioning: Anxiety

Levels of feelings last time felt worried, nervous, or anxious	How often feel worried, nervous, or anxious?					
	Daily	Weekly	Monthly	A few times a year	Never	Don't know
Not asked	…	…	…	…	(1)	(5)
A little	(2)	(2)	(2)	(1)	(1)	(5)
In between	(3)	(2)	(2)	(1)	(1)	(5)
A lot	(4)	(3)	(2)	(1)	(1)	(5)
Don't know	(5)	(5)	(5)	(5)	(5)	(5)

… not applicable

(1) Never feel worried, nervous or anxious OR feel worried, nervous or anxious a few times a year

(2) Feel worried, nervous or anxious monthly OR feel worried, nervous or anxious weekly (a little OR in between) OR feel worried, nervous or anxious daily (a little)

(3) Feel worried, nervous or anxious weekly (a lot) OR feel worried, nervous or anxious daily (in between)

(4) Feel worried, nervous or anxious daily (a lot)

(5) Not stated

Washington Group Extended Set on Functioning: Depression

Level of feelings last time felt depressed	How often do you feel depressed?					
	Daily	Weekly	Monthly	A few times a year	Never	Don't know
Not asked	…	…	…	…	(1)	(5)
A little	(2)	(2)	(2)	(1)	(1)	(5)
In between	(3)	(2)	(2)	(1)	(1)	(5)
A lot	(4)	(3)	(2)	(1)	(1)	(5)
Don't know	(5)	(5)	(5)	(5)	(5)	(5)

… not applicable

(1) Never feel depressed OR feel depressed a few times a year

(2) Feel depressed monthly OR feel depressed weekly (a little OR in between) OR feel depressed daily (a little)

(3) Feel depressed weekly (a lot) OR feel depressed daily (in between)

(4) Feel depressed daily (a lot)

(5) Not stated

Appendix II
Candidate Model 1

	First step coefficients		Second step coefficients
Variable description	**β**	**γ**	**δ**
WG-SS Vision			
No difficulty	0	0	0
Some difficulty	-0.5138	-0.69896	-0.0463124
A lot of difficulty	-0.58369	-16.7024	-0.0806459
Unable to do	0.851381	-16.5612	-0.0700441
Not stated	18.54257	18.99978	-0.6048409
WG-SS Hearing			
No difficulty	0	0	0
Some difficulty	-0.35549	-0.40425	-0.0253059
A lot of difficulty	-1.88281	-17.5026	-0.3140145
Unable to do	-11.2348	18.89366	-0.3513188
Not stated	-33.2204	-34.8649	0
WG-SS Mobility			
No difficulty	0	0	0
Some difficulty	-0.5592	-0.73422	-0.2015701
A lot of difficulty	-2.89708	-16.5066	-0.4700025
Unable to do	-17.1235	-16.742	-0.5949685
Not stated	-19.3796	-18.8958	-0.5706593
WG-SS Cognition			
No difficulty	0	0	0
Some difficulty	-1.07527	-1.90556	-0.1667633
A lot of difficulty	-16.7047	-16.0525	-0.4495562
Unable to do	2.448472	14.38663	-0.2082227
Not stated	18.69593	0.866978	0.0657736
WG-SS Self-care			
No difficulty	0	0	0
Some difficulty	-0.73333	-15.6698	-0.1676797
A lot of difficulty	-15.6507	-14.1999	-0.5187314
Unable to do	-11.1156	3.405978	-0.5250122
Not stated	-0.53158	-0.62957	0.1093124
WG-SS Communication			
No difficulty	0	0	0
Some difficulty	-1.09505	-0.80345	-0.1651021
A lot of difficulty	3.739904	-10.3529	-0.0016886
Unable to do	…	…	…
Not stated	-30.9669	-16.7047	-0.2340906
WG ES-F Pain			
Never had pain OR had a little pain some days	0	0	0
Had pain every day (a little) OR most days (a little or in between) OR some days (in between or a lot)	-0.96752	-0.96959	-0.1372369
Had pain every day (in between) OR most days (a lot)	-2.27801	-2.09324	-0.3591698
Had pain every day (a lot)	-2.30331	-2.08809	-0.5353398
Not stated	…	…	…
WG ES-F Anxiety			
Never feel worried, nervous or anxious OR feel worried, nervous or anxious a few times a year	0	0	0
Feel worried, nervous or anxious monthly OR feel worried, nervous or anxious weekly (a little OR in between) OR feel worried, nervous or anxious daily (a little)	-0.53158	-0.62957	…
Feel worried, nervous or anxious weekly (a lot) OR feel worried, nervous or anxious daily (between a little and a lot)	-1.13668	-2.04243	…
Feel worried, nervous or anxious daily (a lot)	-1.7661	-1.18641	…
Not stated	-1.36895	0.016042	…
WG ES-F Depression			
Never feel depressed OR feel depressed a few times a year	…	…	0
Feel depressed monthly OR feel depressed weekly (a little or between a little and a lot) OR feel depressed daily (a little)	…	…	-0.1741433
Feel depressed weekly (a lot) OR feel depressed daily (between a little and a lot)	…	…	-0.2838989
Feel depressed daily (a lot)	…	…	-0.4179881
Not stated	…	…	-0.2063329
Age (years, centre at 62)			
Age	-0.01504	-0.02237	-0.0015583
Age^2	-0.00125	0.001793	4.43E-06
Age^3	5.85E-05	-3.70E-05	-2.95E-06
Sex			
Male	0	0	0
Female	0.436339	-0.00818	0.0151686
Marital status			
Married or common law	0	0	…
Not stated	-0.89605	-17.6181	…
Constant	0.625385	-0.06105	1.2501338

… not applicable

Note: WG-SS = Washington Group Short Set on Functioning. WG ES-F = Washington Group Extended Set on Functioning.

Source: 2017 Canadian Community Health Survey Rapid Response subsample.

Candidate Model 2

	First step coefficients		Second step coefficients
Variable description	β	γ	δ
WG-SS Vision			
No difficulty	0	0	0
Some difficulty	-0.4427192	-0.622193	-0.0342789
A lot of difficulty	-0.58153122	-16.993984	-0.15771655
Unable to do	1.2729682	-16.476993	-0.06623129
Not stated	18.700756	19.271366	-1.1416135
WG-SS Hearing			
No difficulty	0	0	0
Some difficulty	-0.38875093	-0.44841755	-0.03267161
A lot of difficulty	-1.897773	-17.463538	-0.29629429
Unable to do	-13.133998	18.051263	-0.54203977
Not stated	-34.77817	-35.736802	0
WG-SS Mobility			
No difficulty	0	0	0
Some difficulty	-0.81222301	-0.96174524	-0.27593801
A lot of difficulty	-3.2810702	-17.149309	-0.69897206
Unable to do	-16.847342	-16.271138	-0.73437133
Not stated	-18.989008	-18.591085	-0.49377781
WG-SS Cognition			
No difficulty	0	0	0
Some difficulty	-0.83183092	-1.6617851	-0.19371209
A lot of difficulty	-17.038335	-16.768265	-0.61849676
Unable to do	1.0537305	12.539575	-0.30501034
Not stated	19.115092	1.3470731	-0.37191609
WG-SS Self-care			
No difficulty	0	0	0
Some difficulty	-0.7143783	-15.766142	-0.23042272
A lot of difficulty	-14.405838	-12.887276	-0.68072406
Unable to do	-11.89192	2.17125	-0.67758229
Not stated	0.90863264	0.59832036	0.10931236
WG-SS Communication			
No difficulty	0	0	0
Some difficulty	-1.016783	-0.63671982	-0.19149598
A lot of difficulty	4.5300097	-10.420647	0.00564958
Unable to do	...	...	...
Not stated	-33.052687	-17.591905	-0.18038516
Age (years, centre at 62)			
Age	-0.00242904	-0.01161276	0.0037914
Age^2	-0.00092281	0.00195602	0.00008082
Age^3	0.00005345	-0.00003192	-0.000009276
Sex			
Male	0	0	0
Female	0.18365673	-0.26527423	-0.02518461
Marital status			
Married or common law	0	0	...
Widowed, separated, divorced, single never married	-0.19089891	0.03987409	...
Not stated	-0.52903053	-17.207523	...
Self-rated health			
Poor	-1.0378036	-0.55357327	...
Fair	-0.59380978	-0.89205454	...
Good	-0.8547589	-0.76134274	...
Very good	-0.0067599	-0.12386246	...
Excellent	0	0	...
Not stated	-16.716375	-2.7547615	...
Self-rated mental health			
Poor	-17.455517	-17.043811	...
Fair	-1.6257774	-1.8077786	...
Good	-1.1140278	-0.99820412	...
Very good	-0.20539518	-0.39988375	...
Excellent	0	0	...
Not stated	-1.027783	-0.90652266	...
Constant	0.6743065	0.09994616	1.1266818

... not applicable

Note: WG-SS = Washington Group Short Set on Functioning. WG ES-F = Washington Group Extended Set on Functioning.

Source: 2017 Canadian Community Health Survey Rapid Response subsample.

Before running the first mapping step, the user may derive a three-category variable to represent prominent categories of the HUI3 variable. The new variable, which will be named *H3*, should take the form: H3=1 if HUI3<0.973, H3=2 if HUI3=0.973, H3=3 if HUI3=1.00. The predicted probability of each level of HUI3 may be ascertained for each individual respondent based on their observed characteristics through the following equations:

$$\Pr\left(\mathrm{H3}=3|\mathrm{X}=\mathrm{x}\right)=\frac{\exp\left(\gamma_0+\gamma_1X_1+\gamma_2X_2+\ldots\gamma_nX_n\right)}{1+\exp(\beta_0+\beta_1X_1+\beta_2X_2+\ldots\beta_nX_n)+\exp\left(\gamma_0+\gamma_1X_1+\gamma_2X_2+\ldots\gamma_nX_n\right)} \quad (1)$$

$$\Pr\left(\mathrm{H3}=2|\mathrm{X}=\mathrm{x}\right)=\frac{\exp\left(\beta_0+\beta_1X_1+\beta_2X_2+\ldots\beta_nX_n\right)}{1+\exp(\beta_0+\beta_1X_1+\beta_2X_2+\ldots\beta_nX_n)+\exp\left(\gamma_0+\gamma_1X_1+\gamma_2X_2+\ldots\gamma_nX_n\right)} \quad (2)$$

$$\Pr\left(\mathrm{H3}=1|\mathrm{X}=\mathrm{x}\right)=1-\left[\Pr\left(\mathrm{H3}=3|\mathrm{X}=\mathrm{x}\right)+\Pr\left(\mathrm{H3}=2|\mathrm{X}=\mathrm{x}\right)\right] \quad (3)$$

Work-through using Candidate Model 1: There is a vector 36 of coefficients $\beta_{n=1\ldots36}$ plus the constant β_0 are used to predict for each individual respondent the probability that H3=2 (HUI3=0.973) and a vector of 36 coefficients $\gamma_{n=1\ldots36}$ plus the constant γ_0 are used to predict the probability that H3=3 (HUI3=1). The probability that the HUI3 score is less than 1 (H3=1) can be derived from equation (3). The 36 coefficients relate to item scores for the Washington Group Short Set on Functioning (WG-SS) and Washington Group Extended Set on Functioning (WG ES-F) (with pain and anxiety), age (centred at 62 and entered as linear, quadratic, and cubic forms), sex, and marital status. In some instances, applicable categories of the Washington Group on Disability Statistics (WG) measure are missing coefficients given these were not represented in the analytical dataset.

Each respondent in the dataset will now have a predicted probability of each value of H3. Based on the highest predicted probability for each value of H3, the values corresponding to H3=3 (HUI3=1) and H3=2 (HUI3=0.973) may be imputed directly onto a new mapped health state utility score that will be named *HUI3map*. Individual records where the value of H3=1 shows the highest predicted probability undergo an additional step. Since the predictive coefficients used to estimate this step are derived from arcsine transformation of a linearly transformed HUI3 score of the form $arcsine\left[2*\left(HUI3+0.36/1+0.36\right)\right]-1$ (Materials and methods: Empirical mapping) the equation must reverse-transform the predicted values.

$$HUI3map=\left[\frac{\left(\sin\left(\delta_0+\delta_1X_1+\delta_2X_2+\ldots\delta_nX_n\right)+1\right)*\left(1+0.36\right)}{2}\right]-0.36 \quad (4)$$

Where a vector of 36 coefficients coefficients $\delta_{n=1\ldots36}$ plus the constant δ_0 are used to predict, for each respondent, the arcsine and linear transformed HUI3 score. Since this value is not interpretable, additional steps of reverse transformation as outlined in equation (4) are necessary.

Scenario 1:

A respondent has the following response characteristics: male sex, age 71, widowed, separated, divorced, or single, never married. Responses to the WG-SS and WG ES-F show no difficulties in any functional domain. Estimation Step 1 first takes the predicted sum of β and γ coefficients including the intercept term. For brevity, response categories that fall within the reference category are not shown or those for which the responses are 0. The age term is centred at 62, so it is entered in the prediction equation as 9 (linear), 81 (quadratic), and 729 (cubic). $\sum\beta_{1-n}X_{1-n}$: 0.6253845 (intercept) + -0.01504*9 (age) + -0.001247*81 (Age^2) + 0.0000585*729 (age^3) + -0.237704*1 (marital status)= 0.19396. $\sum\gamma_{1-n}X_{1-n}$: -0.061049 (intercept) + -0.022366* 9(age) + 0.0017927*81 (age^2) + -0.0000372*729 (age^3) + 0.0303471*1 (marital status)= -0.11391. From this, probabilities of falling into prominent discrete categories of 0.973, 1.0, or less than 0.973 (H3 indicator) can be derived. Pr(H3=2): exp(0.19396)/(1 + exp(0.19396) + exp(-0.11391))=0.391 while Pr(H3=3): exp(-0.11391)/(1 + exp(0.19396) + exp(-0.11391))=0.287 and Pr(H3=1) = 0.322. The probability of H3=2 is estimated by the model to be the most likely, and the value of 0.973 is imputed for this respondent and the second estimation step is not needed.

Scenario 2:

A respondent has the following response characteristics: female sex, aged 44, and married or common law. Responses to the WG-SS indicate "No difficulty" in vision, hearing, cognition, self-care, and communication and responses of "Some difficulty" to mobility. Responses to the WG ES-F show pain was experienced "every day (in between a little and a lot) OR most days (a lot)," that the person, "felt worried, nervous or anxious monthly (a little or in between a little and a lot), OR feel worried, nervous or anxious daily (a little)," and that the individual "never feels depressed OR feels depressed a few times a year."

Estimation Step 1 first takes the predicted sum of β and γ coefficients, including the intercept term. Note that the age term is centred at 62, so is included as 44-62= -18 (linear), 324 (quadratic) and -5832 (cubic) forms. $\sum \beta_{1-n} X_{1-n}$: 0.6253845 (intercept) + -0.5592043*1(mobility) + -2.278011*1 (pain) + -0.531578*1(anxiety) + -0.01504*-18 (age) + -0.001247*324 (Age^2) + 0.0000585*-5832 (age^3) + 0.436339*1 (sex) = -2.78162. $\sum \gamma_{1-n} X_{1-n}$: -0.061049 (intercept) + -0.73422*1 (mobility) + -2.093236*1 (pain) + -0.629575*1 (anxiety) + -0.022366*-18 (age) + 0.0017927*324 (age^2) + -0.0000372*-5832 (age^3) + -0.008179*1 (sex) = -2.32589. From this, probabilities of falling into prominent discrete categories of 0.973, 1.0, or less than 0.973 (H3 indicator) can be derived. Pr(H3=2): exp(-2.7816)/(1 + exp(-2.7816) + exp(-2.3259)) = 0.0534 while Pr(H3=3): exp(-2.3259)/(1 + exp(-2.7816) + exp(-2.3259))=0.0842, leaving the difference as Pr(H3=1)=0.862. In this case, the probability of health state utility scores of 1.0 (0.0842) or 0.973 (0.0534) are lower than the probability of scores being less than 0.973, so direct imputation of one of these scores is not possible (in which case the estimation would be completed) and the second estimation step follows.

Estimation Step 2 uses the same variables except marital status and anxiety, which are not included in the predictive model in this step. The sum of the second-step δ coefficients can be expressed as $\sum \delta_{1-n} X_{1-n}$: 1.2501338 (intercept) + -0.20157005*1 (mobility) + -0.35916977*1 (pain) + -0.0015583*-18 (age) + 0.000004428*324 (age^2) + -0.000002948*-5832 (age^3) + 0.01516857*1 (sex) =0.75124. To reverse transform the estimation coefficients to express the health state utility score, ((SIN(0.75124)+1)*(1 + 0.36) / 2) - 0.36 = 0.78413 is used, the mapped HUI3 health state utility score for this respondent.

References

1. Cieza A, Stucki G. The International Classification of Functioning Disability and Health: its development process and content validity. *European journal of physical and rehabilitation medicine*. 2008;44(3):303-13.
2. Madans JH, Loeb ME, Altman BM, editors. Measuring disability and monitoring the UN Convention on the Rights of Persons with Disabilities: the work of the Washington Group on Disability Statistics. *BMC public health*. 2011 Jun 3;11(Suppl 4):S4.
3. Washington Group on Disability Statistics (WG), Budapest Initiative (BI), United Nations Economic & Social Commission for Asia & the Pacific (UNESCAP) [Internet]. *Development of Disability Measures for Surveys: The Extended Set on Functioning 2016*. [cited 2024 Oct 11]. Available from: https://www.cdc.gov/nchs/data/washington_group/development_of_disability_measures_for_surveys_the_extended_set_on_functioning.pdf
4. Washington Group on Disability Statistics (WG) [Internet]. *Understanding and Interpreting Disability as Measured using the WG Short Set of Questions*. Hyattsville, MD: Washington Group on Disability Statistics; 2009 [cited 2024 Oct 11]. Available from: https://www.cdc.gov/nchs/data/washington_group/meeting8/interpreting_disability.pdf
5. Mont D. [Internet*] How Are The Washington Group Questions Consistent With The Social Model Of Disability?* Hyattsville, MD: Washington Group on Disability Statistics; 2019 [cited 2024 Oct 11]. Available from: https://www.washingtongroup-disability.com/wg-blog/how-are-the-washington-group-questions-consistent-with-the-social-model-of-disability-65/
6. Washington Group on Disability Statistics (WG) [Internet]. *An Introduction to the Washington Group on Disability Statistics Question Sets*. Hyattsville, MD: Washington Group on Disability Statistics; 2020 [cited 2024 Oct 11]. Available from: https://www.washingtongroup-disability.com/resources/an-introduction-to-the-washington-group-on-disability-statistics-question-sets-459/
7. United Nations. *United Nations Sustainable Development Goals*. United Nations; 2015 [cited 2024 Oct 11]. Available from: https://www.un.org/sustainabledevelopment/
8. Miller K, Mont D, Maitland A, Altman B, Madans J. Results of a cross-national structured cognitive interviewing protocol to test measures of disability. *Quality & quantity*. 2011;45(4):801-15.
9. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical care*. 2002;40(2):113-28.
10. Health Utilities Inc. [Internet] *Sources of Information on the Health Utilities Index (HUI®)*. Hamilton, ON; 2015 [cited 2024 Oct 11]. Available from: http://healthutilities.com/
11. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI®): concepts, measurement properties and applications. *Health and quality of life outcomes*. 2003;1(1):1-13.
12. Bushnik T, Tjepkema M, Martel L. Health-adjusted life expectancy in Canada. *Health Reports*. 2018;99(4):14-22.
13. Heintz E, Wiréhn A-B, Peebo BB, Rosenqvist U, Levin L-Å. QALY weights for diabetic retinopathy—a comparison of health state valuations with HUI-3, EQ-5D, EQ-VAS, and TTO. *Value in health*. 2012;15(3):475-84.
14. Boyle M, Furlong W, Feeny D, Torrance GW, Hatcher J. Reliability of the health utilities index—mark III used in the 1991 cycle 6 Canadian general social survey health questionnaire. *Quality of life research*. 1995;4:249-57.
15. Feng Y, Bernier J, McIntosh C, Orpana H. Validation of disability categories derived from Health Utilities Index Mark 3 scores. *Health Reports* (Statistics Canada Catalogue 83-003) 2009;20(2):43-50.
16. Kopec JA, Willison KD. A comparative review of four preference-weighted measures of health-related quality of life. *Journal of clinical epidemiology*. 2003;56(4):317-25.
17. Béland Y, Dale V, Dufour J, Hamel M, editors. The Canadian Community Health Survey: Building on the success from the past. In *Proceedings of the American Statistical Association Joint Statistical Meetings, Section on Survey Research Methods; 2005 Aug*. Minneapolis: American Statistical Association, 2005.
18. Statistics Canada. [Internet] *Canadian Community Health Survey - Annual Component (CCHS): Detailed Information for 2015*. Ottawa: Statistics Canada; 2016 Jun [cited 2024 Oct 11]. Available from https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=238854
19. Madans JH, Loeb M. Methods to improve international comparability of census and survey measures of disability. *Disability and rehabilitation*. 2013;35(13):1070-3.
20. Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI®) system for assessing health-related quality of life in clinical studies. *Annals of medicine*. 2001;33(5):375-84.
21. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *The European journal of health economics*. 2010;11(2):215-25.
22. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford university press; 2015.
23. Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. *Value in health*. 2013;16(1):202-10.
24. Fayers PM, Hays RD. Should linking replace regression when mapping from profile-based measures to preference-based measures? *Value in health*. 2014;17(2):261-5.

25. Round J, Hawton A. Statistical alchemy: conceptual validity and mapping to generate health state utility values. *PharmacoEconomics-open*. 2017;1(4):233-9.

26. Wailoo AJ, Hernandez-Alava M, Manca A, Mejia A, Ray J, Crawford B, et al. Mapping to estimate health-state utility from non–preference-based outcome measures: an ISPOR good practices for outcomes research task force report. *Value in health*. 2017;20(1):18-27.

27. Mukuria C, Rowen D, Harnan S, Rawdin A, Wong R, Ara R, et al. An updated systematic review of studies mapping (or cross-walking) measures of health-related quality of life to generic preference-based measures to generate utility values. *Applied health economics and health policy*. 2019;17(3):295-313.

28. Grootendorst P, Marshall D, Pericak D, Bellamy N, Feeny D, Torrance GW. A model to estimate health utilities index mark 3 utility scores from WOMAC index scores in patients with osteoarthritis of the knee. *The Journal of rheumatology*. 2007;34(3):534-42.

29. Franks P, Lubetkin EI, Gold MR, Tancredi DJ. Mapping the SF-12 to preference-based instruments: convergent validity in a low-income, minority population. *Medical care*. 2003;41(11):1277-83.

30. Marshall D, Pericak D, Grootendorst P, Gooch K, Faris P, Frank C, et al. Validation of a prediction model to estimate health utilities index Mark 3 utility scores from WOMAC index scores in patients with osteoarthritis of the hip. *Value in health*. 2008;11(3):470-7.

31. Bartman BA, Rosen MJ, Bradham DD, Weissman J, Hochberg M, Revicki DA. Relationship between health status and utility measures in older claudicants. *Quality of life research*. 1997;7(1):67-73.

32. Nichol MB, Sengupta N, Globe DR. Evaluating quality-adjusted life years: estimation of the health utility index (HUI2) from the SF-36. *Medical decision making*. 2001;21(2):105-12.

33. Sengupta N, Nichol MB, Wu J, Globe D. Mapping the SF-12 to the HUI3 and VAS in a managed care population. *Medical care*. 2004:927-37.

34. Asakawa K, Bernier J, Kohen D. *Conceptual mapping of the Washington Group (WG) measure into the Health Utilities Mark 3 (HUI3)* [Internal Working Document]. Ottawa: Statistics Canada; 2017.

35. Statistics Canada. *Canadian Community Health Survey (CCHS) Annual component: User guide 2017 Microdata file.* Ottawa: Statistics Canada; 2018.

36. Shmueli G. To explain or to predict? *Statistical science*. 2010;25(3):289-310.

37. Lundberg O, Manderbacka K. Assessing reliability of a measure of self-rated health. *Scandinavian journal of social medicine*. 1996;24(3):218-24.

38. Maddox GL, Douglass EB. Self-assessment of health: A longitudinal study of elderly subjects. *Journal of health and social behavior*. 1973:87-93.

39. Kaplan GA, Goldberg DE, Everson SA, Cohen RD, Salonen R, Tuomilehto J, et al. Perceived health status and morbidity and mortality: evidence from the Kuopio ischaemic heart disease risk factor study. *International journal of epidemiology*. 1996;25(2):259-65.

40. Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of health and social behavior*. 1997;38(1):21-37.

41. Orpana HM, Ross N, Feeny D, McFarland B, Bernier J, Kaplan M. The natural history of health-related quality of life: a 10-year cohort study. *Health Reports*. 2009;20(1):29-35.

42. Bernier J, Feng Y, Asakawa K. Strategies for handling normality assumptions in multi-level modeling: a case study estimating trajectories of Health Utilities Index Mark 3 scores. *Health Reports*. 2011;22(4):45-51.

43. Van Doorslaer E, Jones AM. Inequalities in self-reported health: validation of a new approach to measurement. *Journal of health economics*. 2003;22(1):61-87.

44. Charters TJ, Bernier J, Kohen D. Mapping the Washington Group on Disability Statistics disability measure to the Health Utilities Index Mark 3: Development and testing of qualitative and predictive multivariable models in a general population sample. *Methods and References*. Ottawa: Statistics Canada; Forthcoming.

45. Stevens RJ, Poppe KK. Validation of clinical prediction models: what does the "calibration slope" really measure? *Journal of clinical epidemiology*. 2020;118:93-9.

46. Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC medicine*. 2019;17(1):1-7.

47. Gray AM, Rivero-Arias O, Clarke PM. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. *Medical decision making*. 2006;26(1):18-29.