Quality of Life Research (2006) 15: 527–536
DOI 10.1007/s11136-005-2096-4

The natural progression of health-related quality of life: Results of a five-year prospective study of SF-36 scores in a normative population

Wilma M. Hopman[1], Claudie Berger[2], Lawrence Joseph[3], Tanveer Towheed[4], Elizabeth VandenKerkhof[5], Tassos Anastassiades[6], Jonathan D. Adachi[7], George Ioannidis[7], Jacques P. Brown[8], David A. Hanley[9], Emmanuel A. Papadimitropoulos[10] & The CaMos Research Group[11]
[1]*Clinical Research Centre, Kingston General Hospital and Department of Community Health and Epidemiology, Queen's University, Kingston, Ontario, K7L 2V7, Canada (E-mail: hopman@kgh.kari.net);* [2]*CaMos Methods Centre, McGill University, Montreal, Quebec, Canada;* [3]*Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada;* [4]*Division of Rheumatology, Department of Medicine, Department of Community Health and Epidemiology, Queen's University, Kingston, Ontario, Canada;* [5]*Department of Anesthesiology; School of Nursing; Department of Community Health and Epidemiology, Queen's University, Kingston, Ontario, Canada;* [6]*Division of Rheumatology, Department of Medicine, Queen's University, Kingston, Ontario, Canada;* [7]*Department of Medicine, McMaster University, Hamilton, Ontario, Canada;* [8]*Laval University, Ste-Foy, Quebec, Canada;* [9]*University of Calgary, Calgary, Alberta, Canada;* [10]*Eli Lilly Canada Inc, Toronto, Canada and University of Toronto, Faculty of Pharmacy, Toronto, Canada;* [11]*See Acknowledgements for complete list*

Accepted in revised form 14 August 2005

Abstract

Background: Limited information exists regarding the natural progression of health-related quality of life (HRQOL) in the general population, as most research has been cross-sectional or has followed populations with specific medical conditions. Such norms are important to establish, because the effect of any intervention may be confounded by changes due to the natural progression of HRQOL over time. *Methods*: Participants were randomly selected from 9 Canadian cities and surrounding rural areas. Changes in the eight domains and 2 summary component scores of the Medical Outcomes Study 36-item short form (SF-36) were examined over a 5 year period (1996/1997–2001/2002). Mean changes were calculated for men and women within 10 year age categories. Multiple imputation was used to adjust for potential selection bias due to missing data. *Results*: The baseline sample included 6539 women and 2884 men. Loss to follow-up was 17% for women and 23% for men. Mean changes tended to be small, but there was an overall trend towards decreasing HRQOL over time. Changes were more pronounced in the older age groups and in the physically oriented domains. Younger age groups tended towards small mean improvements, particularly in the mentally oriented domains. Large standard errors suggest that on an individual level, large improvements in some participants are balanced by large declines in others. *Conclusion*: In general, the HRQOL of Canadians appears relatively stable over a 5 year period. However, care should be taken when assessing HRQOL longitudinally in certain age or gender groups, as changes associated with an intervention can potentially be confounded by the natural progression of HRQOL.

Key words: Longitudinal, Normative, Prospective, Quality of life, SF-36

Introduction

The widespread use of the Medical Outcomes Study 36-item short form (SF-36) [1, 2] for the assessment of health-related quality of life (HRQOL) has resulted in a large body of literature estimating the burden of a variety of diseases and assessing the benefits of a wide range of treatments and interventions [1–6]. The content of the SF-36 is based on earlier work of the Health Insurance Experiment (HIE), which had aimed to construct the best possible scales for measuring a broad array of functional status and well being concepts. The HIE clearly demonstrated that such tools were valid and reliable for assessing changes in health status in the general population [1]. However, little information is available regarding the natural progression of HRQOL in the general population, as most research using the SF-36 has been cross-sectional or has followed populations with specific medical conditions [3, 7, 8].

One longitudinal, population-based study examined British residents between 35 and 55 years of age at baseline, with a mean follow-up of 36 months [7]. They noted that cross-sectional data underestimated the within-person declines associated with increasing age, and emphasized the need for repeated measures of the SF-36 in population-based studies. A secondary analysis of a subset of civil servants between the ages of 54 and 59 at baseline noted small declines in physical functioning in both men and women, although mental functioning remained stable [9].

A second population-based study assessed participants in the Canadian Multicentre Osteoporosis Study (CaMos) [8] who were between the ages of 40 and 59 years at baseline, also over a 3 year period. In general, it appeared that while the mean HRQOL of the participants remained fairly stable over 3 years, declines were more evident in the physical than the mental domains, as in the British study. Although mean HRQOL changes were small, the standard deviations of the observations were generally high, suggesting that a number of participants had large improvements while others had large declines, essentially canceling each other out. This suggests that changes in HRQOL following an intervention are probably not confounded by declines due to population aging, as long as the follow-up period is of relatively short duration, but whether this holds over longer periods of time is unknown. Moreover, little is known about the stability of SF-36 scores in younger (under 35 years) and older (over 60 years) men and women.

The initiation of CaMos in 1996 provided the opportunity to develop age- and sex-standardized norms for Canadians [10], assess regional differences [11], and more recently, examine changes over a 3 year follow-up for a subset of participants [8]. A follow-up of the full cohort took place 5 years after the baseline assessment, providing the opportunity to lengthen the follow-up period compared to the 3 year assessment, and expand the age range to include all adults ages 25 years and over. This information will provide the research community and health care professionals with some guidance as to how much change in HRQOL over time can be attributed to a natural evolution of HRQOL, and what can be attributed to an intervention or the progression of a chronic illness.

Methods

CaMos is an on-going, prospective cohort study of 9423 non-institutionalized, randomly selected men and women 25 years of age and older at baseline, drawn from a 50 km radius of nine Canadian cities (St John's, Halifax, Quebec City, Toronto, Hamilton, Kingston, Saskatoon, Calgary and Vancouver). A detailed description of the purpose, methodology and sampling framework for CaMos is available elsewhere [10, 12]. In brief, households within each region were selected by random draws of listed telephone numbers, and one randomly selected household member 25 years of age or older was asked to participate. Ethics approval was obtained through the Review Boards of each participating centre.

Baseline assessments took place between February 1996 and September 1997, and the 5 year follow-up took place between February 2001 and September 2002. The interviews were conducted in person at both time points, and included sociodemographic, family history, risk factors and lifestyle questions, as well as the Medical Outcomes Trust 36-item short form (SF-36) [1, 2]. The SF-36 yields 8 domain scores

including physical functioning (PF), role physical (RP), role emotional (RE), bodily pain (BP), general health perceptions (GH), vitality (VT), social function (SF) and mental health (MH) [1], as well as a Physical Component Summary (PCS) and a Mental Component Summary (MCS) [2]. Although a more recent SF-36 version is now available [3], version 1.0 was used to increase the validity of comparisons between the two time points. In order to assess change over time, baseline scores were subtracted from the year 5 scores, so that positive values indicate improvement in HRQOL.

Multiple imputation [13] was used to adjust for possible selection bias, as not all subjects who provided baseline data also participated at year 5. Participants who had died were omitted from the analysis, because change scores are not defined for this population. Imputed values were derived from regression models predicting SF-36 changes over 5 years, using data from respondents with complete data at both time points. We developed separate models for each age and gender category. Variables considered for the models are presented in Appendix A, and were selected on the basis of empirical evidence and/or demonstrated association with HRQOL in other CaMos analyses. Variables entered into each of the models differed, depending on the age and gender category.

The BIC model selection criterion [14] was used to select the best predictive model for each domain and summary score. We then entered baseline data from those missing year 5 data into the models to impute the changes in scores. Simultaneous regression parameter estimations and multiple imputations for all missing data were carried out via a Gibbs sampler algorithm as implemented in Bugs software (version 0.603, Cambridge, Institute of Public Health, MRC Biostatistics Unit, 1995). We ran up to 20,000 iterations of the Gibbs sampler to ensure accurate estimation. Multiple values were imputed for each missing item, and final results were averaged across all imputations. The multiple values account for the fact that the missing data predicted from the regression models are not known with certainty.

Results

Table 1 outlines the loss to follow-up over the 5-year period. More women were lost to follow-up, but the percentage of men was greater (1109 or 17.0% women; 662 or 23.0% men). However, a substantial number of these preferred to do a short questionnaire over the telephone (282 women; 147 men), and as a result they did not have SF-36 data at the time of the 5-year follow-up. The other reasons for loss to follow-up include death (401 women; 243 men); unable to contact (149 women, 128 men), not interested (92 women, 25 men), too sick (69 women, 29 men), moved away (64 women and 59 men) and other (52 women, 31 men) which included reasons such as cancelled, no time or no reason given.

Comparison of those with and without 5-year data

The baseline scores of those with and without 5-year follow-up data differed for a number of age-domain and age-component summary categories. For the women, those with a 5-year follow-up had higher scores in 40 of 48 (83.3%) age-domain cells and 11 of 12 (91.7%) age-component summary

Table 1. Sample size and loss to follow-up at year 5, by age group and gender

Age group (years)	Women			Men		
	Baseline total 1996/1997	5 year total 2001/2002	Lost to follow-up	Baseline total 1996/1997	5 year total 2001/2002	Lost to follow-up
25–34	200	170	30 (15.0%)	200	152	48 (24.0%)
35–44	286	261	25 (8.7%)	212	173	39 (18.4%)
45–54	1111	1024	87 (7.8%)	587	486	101 (17.2%)
55–64	1639	1461	178 (10.9%)	640	554	86 (13.4%)
65–74	2134	1771	363 (17.0%)	802	625	177 (22.1%)
$\geqslant 75$	1169	743	426 (36.4%)	443	232	211 (47.6%)
Total	6539	5430	1109 (17.0%)	2884	2222	662 (23.0%)

cells. For the men, those with follow-up data had higher scores in 41 of 47 (87.2%) age-domain cells (the mean score for one cell was identical) and 11 of 12 (91.7%) age-component summary cells. Higher scores represent better health status, so in general, those with follow-up had a better mean health status at baseline than those lost to follow-up. Results from those who provided data at both time points may be biased, and therefore the results in Tables 2–5 are based on the multiple imputation adjusted mean differences in scores from baseline to 5 years.

Variables selected for the best models for imputation of missing data differed for men and women, across the age strata and for each domain and summary score, and therefore are not described in detail. Appendix A contains the list of variables that contributed to change in one or more domains of the SF-36. In general, the baseline values of the SF-36 domain scores or component summary scores were included in the models, and these were always the strongest predictors.

Mean changes

The estimated mean changes for women and men are presented in Tables 2 and 3, respectively, along with 95% credible intervals (CrIs). The CrI is the Bayesian analogue to frequentist confidence intervals, and provides an interval within which the true parameter value lies with 95% probability, given the data and any prior information included in the model. As we used very low information prior distributions for all analyses, the 95% CrIs are essentially based on information from the data alone. While the mean changes were not large, the standard deviations (SDs) of the observations were generally high, suggesting that large HRQOL improvements in some participants were cancelled out by large declines in others.

Table 2. Estimated mean SF-36 change for women

Age group at baseline	SF-36 Domains/Summary scores									
	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
25–34 (n = 198–200)*										
Mean difference	−0.6	2.8	0.8	1.3	2.0	2.9	5.0	2.5	−0.1	1.7
95% CrI of mean difference	−1.3;0.1	1.1;4.5	−0.3;1.8	0.5;2.1	1.1;2.9	1.9;3.9	3.4;6.6	1.7;3.1	−0.4;0.3	1.3;2.2
35–44 (n = 268–285)										
Mean difference	−1.7	1.7	−1.9	−0.6	−0.4	2.9	2.8	0.9	−0.7	1.1
95% CrI of mean difference	−2.2;−1.1	0.5;3.0	−2.7;−1.1	−1.1;−0.1	−1.0;0.2	2.1;3.6	1.6;4.0	0.4;1.4	−1.0;−0.4	0.7;1.4
45–54 (n = 766–1101)										
Mean difference	−3.0	0.0	−1.4	−0.5	0.7	1.5	2.4	1.6	1.4	1.3
95% CrI of mean difference	−3.3;−2.8	−0.6;0.6	−1.8;−1.1	−0.7;−0.2	0.5;1.0	1.2;1.8	2.0;2.9	1.3;1.8	−1.5;−1.2	1.2;1.5
55–64 (n = 1221–1604)										
Mean difference	−3.7	−2.0	−2.0	−0.9	−0.3	0.9	2.1	1.6	−1.6	1.3
95% CrI of mean difference	−3.9;−3.4	−2.5;−1.5	−2.3;−1.7	−1.2;−0.7	−0.5;−0.1	0.6;1.2	1.6;2.5	1.3;1.8	−1.7;−1.4	1.2;1.4
65–74 (n = 1865–2013)										
Mean difference	−7.1	−4.0	−3.2	−2.0	−4.1	−0.7	1.2	0.4	−2.5	0.7
95% CrI of mean difference	−7.4;−6.8	−4.6;−3.4	−3.6;−2.9	−2.2;−1.8	−4.4;−3.8	−1.0;−0.4	0.6;1.7	0.2;0.6	−2.6;−2.4	0.6;0.8
75+ (n = 555–999)										
Mean difference	−11.8	−2.9	−2.7	−4.0	−5.8	−3.6	1.7	−0.5	−3.5	0.6
95% CrI of mean difference	−12.5;−11.1	−4.7;−1.2	−3.4;−2.0	−4.5;−3.4	−6.4;−5.3	−4.5;−2.7	0.6;2.8	−1.0;0.1	−3.8;−3.1	0.3;0.9

* These values represent the range of the sample size for each age category. The sample size varies due to the differing amounts of missing data available for the imputation models.
CrI = Credible Interval.
Bias adjusted through multiple imputation.
PF = Physical function, RP = Role physical, BP = Bodily pain, GH = General health perceptions, VT = Vitality, SF = Social function, RE = Role emotional, MH = Mental health, PCS = Physical component summary, MCS = Mental component summary.

Table 3. Estimated mean SF-36 change for men

Age group at baseline	SF-36 Domains/Summary scores									
	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
25–34 (n = 53–200)*										
Mean difference	−2.7	−1.0	2.4	−1.8	−3.7	1.3	0.7	0.4	−0.7	0.0
95% CrI of mean difference	−5.0;−0.4	−3.3;1.1	0.9;3.9	−2.9;−0.6	−4.8;−2.5	−0.1;2.8	−1.5;3.0	−0.9;1.6	−1.2;−0.2	−0.6;0.7
35–44 (n = 117–212)										
Mean difference	−0.3	2.3	0.6	−0.3	−0.7	1.3	8.0	0.3	0.0	0.5
95% CrI of mean difference	−1.2;0.5	0.6;4.0	−0.8;2.0	−1.2;0.6	−1.6;0.2	0.3;2.4	6.3;9.8	−0.6;1.2	−0.4;0.5	−0.0;1.0
45–54 (n = 544–575)										
Mean difference	−1.3	0.4	−1.0	−1.1	0.5	3.4	5.4	2.6	−1.1	2.0
95% CrI of mean difference	−1.8;−0.8	−0.5;1.3	−1.6;−0.3	−1.5;−0.6	−0.0;1.0	2.9;4.0	4.7;6.2	2.1;3.0	−1.3;−0.9	1.8;2.2
55–64 (n = 454–627)										
Mean difference	−1.6	0.1	0.6	−1.0	−0.5	1.1	−0.2	1.5	−0.4	0.5
95% CrI of mean difference	−2.0;−1.2	−0.7;1.0	0.1;1.2	−1.4;−0.6	−0.8;−0.1	0.7;1.6	−0.8;0.4	1.2;1.9	−0.6;−0.2	0.3;0.7
65–74 (n = 703–735)										
Mean difference	−7.8	−4.4	−2.4	−3.2	−4.4	−2.7	−1.6	−0.8	−2.6	−0.2
95% CrI of mean difference	−8.4;−7.2	−5.4;−3.2	−3.1;−1.8	−3.7;−2.7	−4.8;−3.9	−3.3;−2.1	−2.5;−0.7	−1.2;−0.4	−2.9;−2.3	−0.4;0.1
75+ (n = 319–350)										
Mean difference	−12.5	−7.2	−2.8	−4.4	−8.6	−4.3	2.9	−0.5	−4.4	0.5
95% CrI of mean difference	−14.3;−10.7	−10.2;−4.2	−4.5;−1.1	−5.6;−3.2	−9.9;−7.2	−6.0;−2.7	0.6;5.2	−1.6;0.6	−5.2;−3.7	−0.1;1.1

* These values represent the range of the sample size for each age category. The sample size varies due to the differing amounts of missing data available for the imputation models.
CrI = Credible interval.
Bias adjusted through multiple imputation.
PF = Physical function, RP = Role physical, BP = Bodily pain, GH = General health perceptions, VT = Vitality, SF = Social function, RE = Role emotional, MH = Mental health, PCS = Physical component summary, MCS = Mental Component summary.

Overall, for the 48 age-domain cells, the number of mean scores that declined outnumbered the number of mean scores that improved, for both men and women. Women reported HRQOL improvements in 21 of 48 (43.8%) age-domain cells; declines were noted in 27 of 48 (56.3%) age-domain cells. Men reported improvements in 19 of 48 (39.6%) age-domain cells, and declines in 29 of 48 (60.4%) age-domain cells.

For the component summary scores, improvements and declines were close to equal. Women improved in 7 of 12 (58.3%) and declined in 5 of 12 (41.7%) of the age-component summary cells. Men improved in 5 of 12 (41.7%), declined in 6 of 12 (50.0%) and stayed the same in 1 of 12 (8.3%) of the age-component summary cells. Declines were much more apparent in the physically oriented domains of PF, RP, BP and GH and the PCS, for both men and women. By contrast, the mentally oriented domains of VT, SF, RE and MH were more likely to improve, which was more pronounced in women.

Size of mean changes

There is debate in the literature about whether five or ten point changes are required for clinical relevance. A minimum change of five points is considered clinically and socially relevant by some [1], while others consider a larger change of ten points as important [15, 16]. For both men and women, older age groupings showed much more decline than the younger age groupings, particularly in the physically oriented domains. For women, the PF for those 65–74 and 75+ years had a mean decline that exceeded five points. The mean change in VT also exceeded five points for the age group of 75+ years.

For men, mean changes greater than five points were seen for the PF domain in those 65–74 and

Table 4. Percent of women who improved by ⩾10 points, stayed within 10 points of their baseline score, or declined by ⩾10 points over the 5 year follow-up (5-point changes for the PCS and MCS)

Age group at baseline		SF-36 Domains/Summary scores									
		PF	RP	BP	GH	VT	SF	RE	MH	PCS*	MCS*
25–34	Improved by ⩾10 points	13.3	20.5	31.6	23.3	31.9	32.3	24.7	28.1	9.1	19.4
	Change of <10 points	66.5	60.1	35.6	56.4	40.8	45.2	59.1	52.6	80.8	68.3
	Declined by ⩾10 points	20.2	19.4	32.8	20.3	27.3	22.5	16.2	19.3	10.1	12.3
35–44	Improved by ⩾10 points	16.9	22.6	**32.7**	25.8	31.5	34.2	21.4	20.9	11.1	14.2
	Change of <10 points	59.6	58.1	**29.8**	49.6	38.5	45.7	64.1	61.6	77.9	74.4
	Declined by ⩾10 points	23.5	19.3	**37.5**	24.6	30.0	20.1	14.5	17.5	11.0	11.4
45–54	Improved by ⩾10 points	19.2	18.0	**31.1**	24.0	30.6	29.8	18.1	23.0	8.5	13.7
	Change of <10 points	52.4	63.1	**31.3**	51.0	42.2	46.6	67.5	60.0	78.4	77.0
	Declined by ⩾10 points	28.4	18.9	**37.6**	25.0	27.2	23.6	14.3	17.0	13.1	9.3
55–64	Improved by ⩾10 points	18.8	20.8	**31.1**	22.2	27.8	26.7	17.3	23.8	8.5	12.0
	Change of <10 points	49.3	56.7	**32.1**	51.6	43.8	49.4	68.4	59.0	76.9	79.7
	Declined by ⩾10 points	31.9	22.5	**36.8**	26.2	28.4	23.9	14.3	17.2	14.6	8.3
65–74	Improved by ⩾10 points	**16.6**	21.8	**29.1**	21.9	20.6	23.7	19.2	21.7	7.4	11.7
	Change of <10 points	**40.6**	49.0	**31.8**	47.4	40.9	48.8	63.0	58.4	74.6	79.4
	Declined by ⩾10 points	**42.8**	29.2	**39.1**	30.7	38.5	27.5	17.8	19.9	18.0	8.9
75+	Improved by ⩾10 points	**14.9**	29.5	**29.6**	19.4	**21.1**	**25.7**	24.3	22.7	7.8	13.2
	Change of <10 points	**29.5**	37.2	**31.2**	44.4	**36.2**	**37.0**	53.8	52.2	70.1	75.5
	Declined by ⩾10 points	**55.6**	33.3	**39.2**	36.2	**42.7**	**37.3**	21.9	25.1	22.1	11.4

PF = Physical function, RP = Role physical, BP = Bodily pain, GH = General health perceptions, VT = Vitality, SF = Social function, RE = Role emotional, MH = Mental health, PCS = Physical component summary, MCS = Mental component summary.

* PCS and MCS changes are based on 5 points rather than 10 points,

Bold items are the 9 cells where percent that declined exceeds the percent stable and percent improved.

75+ years of age. For those 75+ years, the domains of RP and VT also showed mean declines of over five points. However, for the RE domain, women 25–34 years had a mean improvement of 5.0 points, while men 35–44 and 45–54 years had a mean improvement of 8.0 and 5.4 points, respectively.

In addition, for women the credible intervals were both negative for 24 of 27 (88.9%) age-domain groups that had a mean decline, as well as 4 of 5 (80.0%) of the age-component summary cells that showed a mean decline, suggesting real decline. Likewise, the CrIs were both positive for 20 of 21 (95.2%) of the age-domain cells that improved, and 6 of 7 (85.7) of the age-component summary cells that improved, suggesting real improvement. For men, the CrIs were both negative for 23 of 29 (79.3%) age-domain groups that had a mean decline, as well as 5 of 6 (83.3%) of the age-component summary cells that showed a mean decline, and both positive for 11 of 19 (57.9%) of the age-domain cells that improved, and 2 of 6 (33.3%) of the age-component summary cells that improved.

Percent changing by 10 points or more

In order to attain a better understanding of the distribution of moderate changes, Tables 4 (women) and 5 (men) contain the adjusted percentage of those who improved by ⩾10 points, those who stayed within 10 points of their baseline score, and those who declined by ⩾10 points. However, a 5-point change was used for the PCS and MCS, as the distribution of the summary scores is narrower (8–72 for the PCS and 9–74 for the MCS [2]).

Only one domain, BP for men 25–34 years, had a larger percentage who improved by ⩾ 10 points as compared to the percentage who declined by ⩾ 10 points or stayed within 10 points of their baseline score. Fourteen age-domain cells (9 for women, 5 for men) had more who declined by ⩾ 10 points than the percent who improved ⩾ 10 points or stayed within 10 points of their original score. These 14 are bolded in the two tables. Seven of these involved the BP domain (all but the youngest age group for women, and the two oldest age groups for men). The other 7 included PF for

Table 5. Percent of men who improved by $\geqslant 10$ points, stayed within 10 points of their baseline score, or declined by $\geqslant 10$ points over the 5 year follow-up (5-point changes for the PCS and MCS)

Age group at baseline		SF-36 Domains/Summary scores								
		PF	RP	BP	GH	VT	SF	RE	MH	PCS*
25–34	Improved by $\geqslant 10$ points	13.0	17.3	37.4	31.4	19.6	28.6	20.5	23.3	5.5
	Change of < 10 points	64.1	63.1	35.7	43.7	46.1	49.7	61.0	56.9	85.6
	Declined by $\geqslant 10$ points	22.9	19.6	26.9	24.9	34.3	21.7	18.5	19.8	8.9
35–44	Improved by $\geqslant 10$ points	16.0	20.7	31.9	25.0	24.5	26.7	22.5	20.3	7.5
	Change of < 10 points	68.9	64.2	37.3	51.8	51.2	53.3	68.5	61.3	83.0
	Declined by $\geqslant 10$ points	15.1	15.1	30.8	23.2	24.3	20.0	9.0	18.4	9.5
45–54	Improved by $\geqslant 10$ points	16.3	18.0	30.9	20.7	27.9	28.9	18.4	21.8	5.9
	Change of < 10 points	64.8	64.7	35.7	55.0	47.7	52.4	71.3	65.9	83.0
	Declined by $\geqslant 10$ points	18.9	17.3	33.4	24.3	24.4	18.7	10.3	12.3	11.1
55–64	Improved by $\geqslant 10$ points	17.7	19.7	33.9	23.8	25.1	23.0	13.7	21.3	8.1
	Change of < 10 points	60.0	61.4	34.5	48.4	46.7	55.8	72.1	64.4	83.3
	Declined by $\geqslant 10$ points	22.3	18.9	31.6	27.8	28.2	21.2	14.2	14.3	8.6
65–74	Improved by $\geqslant 10$ points	15.8	20.9	**28.5**	21.0	18.5	19.4	15.7	19.1	6.8
	Change of < 10 points	44.5	49.5	**34.0**	45.6	43.0	52.2	65.2	59.8	74.0
	Declined by $\geqslant 10$ points	39.7	29.6	**37.5**	33.4	38.5	28.4	19.1	21.1	19.2
75+	Improved by $\geqslant 10$ points	**16.1**	**25.3**	**29.8**	18.1	**16.7**	21.7	25.0	23.2	5.8
	Change of < 10 points	**29.4**	**36.2**	**33.1**	46.8	**35.7**	42.5	52.8	53.1	69.8
	Declined by $\geqslant 10$ points	**54.5**	**38.5**	**37.1**	35.1	**47.5**	35.8	22.2	23.7	24.4

PF = Physical Function, RP = Role Physical, BP = Bodily Pain, GH = General Health Perceptions, VT = Vitality, SF = Social Function, RE = Role Emotional, MH = Mental Health, PCS = Physical Component Summary, MCS = Mental Component Summary.

* PCS and MCS changes are based on 5 points rather than 10 points.

Bolded items are the 5 cells where the percent that declined exceeds the percent stable and percent improved.

women 65 – 74 years and 75+ years, as well as VT and SF in women 75+ years. The remaining 3 involved the men 75+ years, and included the PF, RP and VT domains.

For most of the cells, the largest percentage demonstrated a change of < 10 points (< 5 for the PCS and MCS). For women, this included 39 of 48 (81.3%) of the age-domain scores cells and all 12 of the age-component summary cells (5 points for the PCS and MCS). For men, this included 42 of 48 (87.5%) of the age-domain scores cells and all 12 of the age-component summary cells.

Discussion

There is a consistent trend across all age groups and in both men and women towards lower quality of life for the physically oriented domains (PF, RP, BP and GH) as age increases. However, most of the mentally oriented domains (SF, MH and particularly RE) showed small mean improvements despite increasing age. The Vitality domain was the exception; despite being considered one of the mentally oriented domains [2], it showed more mean declines than improvements with advanced age.

While the overall trend was one of decline, the changes tended to be small. The majority of respondents exhibited changes that were less than 10 points in magnitude (5 points for the PCS and MCS), suggesting that in general, SF-36 scores display reasonable stability over time.

This finding is consistent with previous work. Hemingway et al. noted mean declines in all domain scores over a 3 year period, ranging from 0.3 (MH) to 3.0 (RP) in men and 0.3 (MH) to 2.4 (RP) in women with no physical or psychiatric disease [7]. Declines were larger for those with physical disease, psychiatric disease or both [7]. Previous work with a subset of the CaMos population (40–59 years) identified declines in 68.8% of the domain-age groups for women and 56.3% of the domain-age groups for men over a 3 year period [8]. Our results indicate that when examined over a 5 year period and including a wider age group (25–75+ years), 56.2% of the age-domain groups for women and 60.4% of the age-domain groups for men exhibit decline.

The pattern of change remains similar. Previous data identified the greatest variability in the BP, VT and DF domains [8], which was replicated in the current study. Direct comparison of the slopes of the changes is difficult as the age groupings are not the same. However, in general, domains that exhibited declines or improvements over the 3 year study exhibited similar declines or improvements over the 5 year study. Our results, as well as those of the two previous longitudinal studies [7, 8], suggest that changes accumulated over time could become quite large.

This study is the first to assess change in younger (< 35 years) and older (> 55 years) age groups. HRQOL may *improve* quite substantially over time, particularly in the mentally oriented domains and in the younger age groups. For women, the mean change remains positive across all age groups for RE, and for MH, only those in the 75+ year group show a mean decline. For men, declines in the mentally oriented domains only become apparent after age 65 years. For the physically oriented domains, mean declines are more common for both men and women, particularly in those over 65 years of age.

The assessment of SF-36 change over time provides additional information for establishing meaningful change over time [17]. In general, changes in HRQOL following an intervention are probably not confounded with changes due to aging provided that the follow-up period is of relatively short duration. However, in some groups, and in particular the oldest age groups, it is possible that the effects of a beneficial treatment or intervention may not be evident, if the effect of the intervention is of the same magnitude (but in the opposite direction) as the HRQOL decline due to aging. It should be noted, however, that these results are based on a random sample of the population, which may have different HRQOL than patients selected for an intervention or a clinical trial.

One limitation of the study is the loss to follow-up, at 17% of the women and 23% of the men. In addition, a small proportion of those lost to follow-up indicated that they were too sick, which is selected loss to follow-up and may lead to a slight underestimation of the real natural progression of HRQOL. While multiple imputation was used to adjust for bias due to missing data, this technique also has assumptions and limitations. For example, we assumed that the baseline data were sufficiently detailed to predict 5 year changes, and that patterns of change were similar in both respondents and non-respondents after adjusting for baseline differences. Therefore, despite adjusting for baseline characteristics, we conclude that non-respondents at 5 years may still differ from those with complete data.

A second limitation is that the SF-36 domains have a pronounced ceiling effect in the general population for certain domains, in particular RP and RE. This is somewhat more pronounced in men and in younger age groups. An evaluation of change may therefore be underestimated by our inability to detect changes in those at the ceiling of the domain scores.

In conclusion, it appears that overall, the mean HRQOL of Canadian residents remains fairly stable over 5 years, as the majority of the mean differences were relatively small for most domain and age categories. However, large standard errors suggest that on an individual level, large improvements in some participants are balanced by large declines in others, resulting in small mean differences. Notable improvements are evident in the younger age groups in the mentally oriented domains, as are large declines in the older age groups in the physically oriented domains.

Care should therefore be taken when assessing HRQOL changes over time, as changes associated with a treatment or intervention can potentially be confounded by the natural progression of HRQOL over time. Changes in HRQOL may need to be examined within the context of the age, gender and SF-36 domain or component summary score under investigation.

Acknowledgements

CaMos was funded by the Canadian Institutes of Health Research (CIHR), Merck Frosst Canada Ltd., Eli Lilly Canada Inc., Novartis Pharmaceuicals Inc., The Alliance: Safoni-Aventis & Procter and Gamble Pharmaceuticals Canada Inc., The Dairy Farmers of Canada, The Arthritis Society. Lawrence Joseph is funded by a Senior Scientist Award from the CIHR. The authors thank all the participants in the Canadian Multicentre

Osteoporosis Study. Investigators in the CaMos Research Group are: McGill University, Royal Victoria Hospital, Montreal, Quebec (Co-ordinating centre): David Goltzman (Co-principal investigator), Alan Tenenhouse (Senior researcher), Suzette Poliquin (National co-ordinator), Suzanne Godmaire (Research assistant), Lawrence Joseph (Study statistician), Claudie Berger (Statistician) Memorial University, St. John's Newfoundland: Carol Joyce (Centre director), Emma Sheppard (Centre co-ordinator). Dalhousie University, Halifax, Nova Scotia: Susan Kirkland, Stephanie Kaiser (Centre co-directors), Barbara Stanfield (Centre co-ordinator). Laval University, Quebec City, Quebec: Jacques P. Brown (Centre director), Louis Bessette (Co-director), Evelyne Lejeune (Centre co-ordinator), Marc Gendreau (Imaging co-ordinator) Queen's University, Kingston, Ontario: Tassos Anastassiades (Centre director), Tanveer Towheed (Co-director), Barbara Matthews (Centre co-ordinator). University of Toronto, Toronto, Ontario: Nancy Kreiger (Co-principal investigator), Bob Josse (Centre director), Tim Murray (co-director), Barbara Gardner-Bray (Centre co-ordinator). McMaster University, Hamilton, Ontario: Jonathan D. Adachi (Centre director), Alexandra Papaioannou (Co-director), Laura Pickard (Centre co-ordinator). University of Saskatchewan, Saskatoon, Saskatchewan: Wojciech P. Olszynski (Centre director), Jola Thingvold (Centre co-ordinator). University of Calgary, Calgary, Alberta: David A. Hanley (Centre director), Jane Allan (Centre co-ordinator). University Alberta, Edmonton, Alberta: Kerry Siminoski (Quality Assurance) University British Columbia, Vancouver, British Columbia: Jerilynn C. Prior (Centre director), Yvette Vigna (Centre co-ordinator).

Appendix

Appendix A. Variables considered for entry into the imputation models

Baseline SF-36
All baseline SF-36 domain and component summary scores
Administrative information centre
Age
Employment status
Education level
Physical characteristics
Height
Body mass index
Bone mineral density spine (L1–4)
Bone mineral density, femoral neck
Comorbid conditions (Self-reported)
Osteoporosis
Osteoarthritis
Rheumatoid arthritis
Chronic obstructive pulmonary disease
Thyroid disease
Scoliosis
Eating disorder
Liver disease
Breast cancer
Uterine cancer
Prostate cancer
Irritable bowel syndrome
Kidney stones
Kidney disease
Hypertension
Heart disease
Cerebrovascular accident or transient ischemic attack
Neuromuscular disease
Diabetes type I
Diabetes type II
Phlebitis
Number of comorbid conditions
Previous surgery
Stomach
Intestine
Gall bladder
Thyroid
Fracture and immobility
Immobilized for more than one month
Deformity
History of fracture
Fracture of forearm or wrist, minimal trauma
Fracture of ribs, minimal trauma
Fracture of pelvis, minimal trauma
Fracture of back, minimal trauma
Lifestyle and activity
Smoked cigarettes for at least 6 months
Lifetime packs of cigarettes
Activity level at work
Time spent walking, typical week, past 6 months
Current participation in regular activity
Number of sedentary hours per day
Daily calcium intake, past 12 months
Caffeine intake, past 12 months
Number of alcoholic beverages
Weekly kilocalories spent in vigorous or moderate activity
Reproductive history
Number of children fathered
Pregnant in the past 12 months

Appendix A. Continued

Appendix A. Continued

Number of live births
Breastfeeding history
Menopausal status
Uterus removed
Ovaries removed
Medication use
Current hormone replacement therapy
Ever taken cortisone daily
Current corticosteroid use
Current bisphosphonate use

References

1. Ware JE Jr, Snow KK, Kosinski M, Gandek B. SF-36 Health Survey Manual and Interpretation Guide. Boston: The Health Institute, New England Medical Center, 1993.
2. Ware JE Jr, Kosinski M, Keller SD. SF-36 Physical and Mental Summary Scales: A User's Manual. Boston: The Health Institute, New England Medical Center, 1994.
3. Ware JE Jr.. SF-36 health survey update. Spine 2000; 25: 3130–3139.
4. Andresen EM, Meyers AR. Health-related quality of life outcomes measures. Arch Phys Med Rehab 2000; 81(12 Suppl 2) S30–45.
5. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: Bibliographic study of patient assessed health outcomes measures. Br Med J 2002; 324: 1417–1421.
6. Bayliss EA, Bayliss MS, Ware JE Jr, Steiner JF. Predicting declines in physical function in persons with multiple chronic medical conditions: What we can learn from the medical problem list. Health and Quality of Life Outcomes 2004; 2: 47–55.
7. Hemingway H, Stafford M, Stansfield S, Shipley M, Marmot M. Is the SF-36 a valid measure of change in population health? Results from the Whitehall II study. Br Med J 1997; 315: 1273–1279.
8. Hopman WM, Berger C, Joseph L, Towheed T, vanden Kerkhof E, Anastassiades T, Cranney A, Adachi JD, Ioannidis G, Poliquin S, Brown JP, Murray TM, Hanley DA, Papadimitropoulos M, Tenenhouse A the CaMos Research Group. Stability of Normative Data for the SF-36 Health Survey: Results of a three-year prospective study in middle-aged Canadians. Can J Public Health 2004; 95: 387–391.
9. Mein G, Martikainen P, Hemingway H, Stansfeld S, Marmot M. Is retirement good or bad for mental and physical functioning? Whitehall II longitudinal study of civil servants. J Epidemiol Commun Health 2003; 57: 46–49.
10. Hopman WM, Towheed T, Anastassiades T, Tenenhouse A, Poliquin S, Berger C, Joseph L, Brown J, Murray T, Adachi JD, Hanley D, Papadimitropoulos M the CaMos Research Group. Canadian Normative Data for the SF-36 Health Survey. Can Med Assoc J 2000; 163: 265–271.
11. Hopman WM, Towheed T, Anastassiades T, Tenenhouse A, Poliquin S, Berger C, Joseph L, Brown J, Murray T, Adachi JD, Hanley D, Papadimitropoulos M the CaMos Research Group. Is there regional variation in the SF-36 scores of Canadian adults? Can J Public Health 2002; 93: 233–236.
12. Kreiger N, Tenenhouse A, Joseph L, MacKenzie T, Poliquin S, Brown J, Prior JC, Rittmaster R. Research Notes: The Canadian Multicentre Osteoporosis Study (CaMos): Background, Rationale, Methods. Can J Aging 1999; 18: 376–387.
13. Rubin D. Multiple Imputation for Non-response in Surveys. New York: Wiley, 1987.
14. Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc 1995; 90: 773–795.
15. Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. Health and Quality of Life Outcomes 2003; 11: 4.
16. Panopalis P, Petri M, Manzi S, Isenberg A, Gordon C, Senecal J-L, Penrod JR, Joseph L, St. Pierre Y, Pineau C, Fortin PR, Sutcliffe N, Goulet J-R, Choquette D, Grodzicky T, Esdaile JM, Clarke AEthe Tri-Nation Study Group. The systemic lupus erythematosus tri-nation study: longitudinal changes in physical and mental well-being. Rheumatology 2005; 44: 751–55.
17. Testa MA. Interpretation of quality of life outcomes: Issues that affect magnitude and meaning. Medical Care 2000; 38(Supplement II) 166–174.

Address for correspondence: Wilma M. Hopman, Research Facilitator, Clinical Research Centre, Kngston General Hospital, Kingston, Ontario K7L 2V7, Canada
Tel: 613 549 6666; Fax: 613 548 2428
E-mail: Hopmanw@kgh.kari.net