Benchmarking Health-Related Quality-of-Life Data From a Clinical Setting

Janel Hanmer, MD, PhD; Rachel Hess, MD, MS; Sarah Sullivan, BS; Lan Yu, PhD; Winifred Teuteberg, MD; Jeffrey Teuteberg, MD; and Dio Kavalieratos, PhD

Healthcare in the United States is evaluated by a wide variety of organizations and measures, which often include conventional health indicators (eg, mortality rates, complication rates, health service use) and measures of patient satisfaction. These measures are important, but do not include one of the ultimate goals of healthcare: health-related quality of life (HRQoL).[1] Conventional health indicators are frequently used instead of HRQoL measures because they are more easily quantified, and there are often comparative data available for interpretation of the results.[2] As the field of HRQoL measurement has improved, there is increasing interest in including patient-reported outcomes (PROs), such as HRQoL, as an outcome of clinical care.[3-5]

HRQoL measures are either disease-specific or generic. Disease-specific measures provide a high level of detail about disease-specific symptoms and experiences; conversely, generic measures allow for comparison across disease groups.[6] These generic HRQoL measures can broadly be categorized into 2 groups: health status measures and health preference measures. Health status measures (or profile measures) provide a description of multiple domains of health, such as physical functioning, mental health, and pain.[7] These measures provide multiple scores—1 for each domain of health measured. In contrast, health preference measures can combine multiple domains into a single score, commonly referred to as a "utility score." This score is constructed using preferences of different descriptions of health elicited from the general population.[8]

Generic HRQoL measures provide a unique opportunity to compare outcomes across clinical practices. Unfortunately, these measures are not collected and published routinely enough for comparisons across clinics, health systems, or regions. However, many of these generic measures have been included in large, nationally representative datasets, providing a rich resource for comparisons across smaller population and patient groups.[6,9-11] Catalogs of age and sex normative values for generic HRQoL measures[6,9,10] have also been published from these datasets, but the values from these catalogs do not address specific patient groups.

ABSTRACT

OBJECTIVES: Health-related quality of life (HRQoL) is an important clinical outcome, yet there is little guidance for its interpretation in clinical settings. One approach would use benchmarking to contextualize HRQoL results. Our objective was to construct a nationally representative HRQoL benchmark for use with a clinical sample.

STUDY DESIGN: Cross-sectional analysis of HRQoL scores from: 1) the 2011 Medical Expenditures Panel Survey (MEPS), a representative sample of the noninstitutionalized US population; and 2) outpatient academic and community cardiology clinics within a large health system in 2012 and 2013.

METHODS: The 2011 MEPS includes 21,959 adults who completed the HRQoL measures; 414 reported visiting a cardiologist. Of 1945 outpatient index visits during the study period that were not for outpatient cardiac catheterization, 1434 patients completed the HRQoL measures. The primary outcome was the Short Form 6-Dimension questionnaire (SF-6D). The secondary outcomes were the Mental Component Summary score and the Physical Component Summary score.

RESULTS: The local cardiology clinic sample was 42% female with a mean Charlson Comorbidity Index (CCI) score of 1.74. The MEPS subsample of cardiology patients more closely matched the local cardiology clinic sample (43% female; mean CCI score of 1.57) than the entire MEPS sample (52% female; mean CCI score of 0.62). SF-6D scores for the local cardiology clinic sample were significantly better, statistically and clinically, in 4 of 5 age strata than the MEPS subsample of cardiology patients.

CONCLUSIONS: HRQoL benchmarks can be created from current public datasets. Subgroups in national samples may provide more valid benchmarks for clinical populations.

Am J Manag Care. 2016;22(10):669-675

TAKE-AWAY POINTS

Health-related quality of life (HRQoL) data is often collected during routine clinical care; however, there is limited guidance in interpreting the HRQoL results for a clinical practice. This article presents a method to construct a benchmark for subspecialty care clinics from a nationally representative dataset.

‣ Creating HRQoL benchmarks for subspecialty clinics is possible.

‣ HRQoL benchmarks create a point-of-comparison for local subspecialty clinics.

‣ Differences from benchmarks may inform local quality improvement projects.

Comparing HRQoL results from a clinical sample to HRQoL outcomes for the general population is of limited value because we assume that a clinical sample has more health conditions and worse scores than the general population. Clinicians often worry about the interpretation of any outcome measure because of differing case mixes across clinicians and practices.[12] Those who collect clinical HRQoL data would benefit from a point of comparison or benchmark[13] for HRQoL results. Benchmarks exist for a wide variety of other clinical outcomes, ranging from physiological markers (eg, glycated hemoglobin and blood pressure control in patients with diabetes) to patient experience (eg, pain control measures, quietness of the hospital environment) to guideline adherence (eg, coronary artery disease, heart failure, atrial fibrillation).[14,15]

In this report, we construct a HRQoL benchmark for a clinical subspecialty using a nationally representative dataset called the Medical Expenditures Panel Survey (MEPS). MEPS includes HRQoL measures, which allow for the construction of overall population scores.[9] The MEPS dataset also includes information about which subspecialists the respondents have seen in the past year, allowing for the construction of subspecialty-specific HRQoL scores. Cardiology is the clinical subspecialty illustrated in this report; however, the technique we present herein can be used with other subspecialties included in MEPS. We compared the national cardiology benchmark to data collected in cardiology clinics from a large health system.

METHODS

Data: National Sample

MEPS is a nationally representative survey of healthcare utilization and expenditures for the US noninstitutionalized civilian population. It is a 2-year panel survey with an overlapping cohort design; each year, a new cohort is initiated and followed longitudinally. Cross-sectional analyses combine information from 2 MEPS cohorts. MEPS data, including the scores analyzed in this report, are freely available online through the Agency for Healthcare Research and Quality (AHRQ) website. We used the most recently released MEPS data for 2011.[16]

MEPS includes a self-administered questionnaire, which is distributed to all adults 18 years or older in eligible households par-

ticipating in MEPS. The 2011 MEPS included version 2 of the 12-Item Short Form Health Survey (SF-12v2), a generic HRQoL measure that is widely used in clinical practice and research.[17] We analyzed both the entire MEPS sample (referred to hereafter as "All MEPS") and the subset of patients who reported seeing a cardiologist in 2011 (referred to hereafter as "Cardiology MEPS").

Data: Clinical Sample

All patients presenting to any outpatient cardiology practice within our large academic and community health system were asked to complete a survey, including version 1 of the 36-Item Short Form Survey (SF-36v1), a generic HRQoL measure,[18] through the Patient-Reported Information Clinical Intake System (PRIcis). PRIcis is an institutional initiative to electronically collect PROs using a secure patient portal or tablet computers in the clinic. These PROs are securely transmitted to and stored within our electronic health record (EHR) system for use in clinical care. Data used in this analysis were collected from March 2012 to April 2013. We included information from a patient's first visit to a cardiology clinic within this time period; we excluded all other visits and all visits for outpatient heart catheterization. PRIcis was linked to the EHR for demographic and comorbidity information. These data are hereafter referred to as "Cardiology Local."

Outcome Variables

The primary outcome of interest was the Short Form 6-Dimension questionnaire (SF-6D) health preference score.[19] This score can be calculated from both the SF-12v2 in MEPS and the SF-36v1 from PRIcis. The questions in these measures were used to construct health scenarios that were evaluated using the standard gamble[20] technique in a representative sample of the UK population. Regression analysis was then used to model the preferences assigned to each health state. With the resulting scoring algorithm, a preference-based score can be assigned to each health state with "dead" anchored at 0 and "full health" anchored at 1.0. This scoring algorithm is published in the peer-reviewed literature.[19]

Secondary outcomes of interest were the Mental Component Summary (MCS) score and Physical Component Summary (PCS) score. The MCS and PCS were developed from a reduction of the 8 dimensions in the SF-36 (physical functioning, physical role limitations, emotional role limitations, pain, general health, vitality, social functioning, and mental health) to 2 dimensions by factor analysis. The SF-12 is an abridged version of the SF-36, which was constructed so that MCS and PCS scores from either form would be equivalent.[21,22]

The MCS and PCS scores calculated from the SF-36v1 were based on 1990 US normative data using the publicly available scoring algorithms for version 1.[23,24] Scores calculated from the SF-12v2 were based

on 1998 US normative data held by QualityMetric, Inc,[22] which is now part of Optum (Eden Prairie, Minnesota). The MCS and PCS scores were normalized so that both have averages of 50 and standard deviations of 10. We applied the standard correction to compare across the different versions,[22] and we used imputed scores in our analyses—which are freely available in the MEPS dataset—calculated by applying a proprietary algorithm of, what was then, QualityMetric, Inc.

Independent Variables

Both datasets included age in years, sex, and race. Race categories were collapsed into white, black, other, and unspecified to be consistent across the datasets. We calculated a Charlson Comorbidity Index (CCI) score constructed for administrative data,[25] which uses *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)* codes. We included *ICD-9-CM* codes entered in the EHR in the problem list or medical history before, or within 1 week, of the cardiology clinic visit. *ICD-9-CM* codes are included in MEPS if the individual had a condition linked to a 2011 event (eg, physician visit or taking medication) or disability day, or if the individual was, at that time, experiencing a condition as part of the MEPS condition enumeration assessment.

Statistical Analyses

We used Welch's 2-tailed *t* test with unequal variances to compare the mean for SF-6D, MCS, and PCS scores by age strata (18-44, 45-54, 55-64, 65-74, ≥75 years). MEPS results were weighted to be nationally representative in these comparisons. We used the unweighted number of respondents from MEPS for conservatism in the statistical testing.

As a sensitivity analysis to address different demographic compositions in the clinical and MEPS samples, we created a dataset that combined our clinical data with the MEPS cardiology subsample into a single dataset. We stratified by age group and used ordinary least squares regression with the outcome of interest (SF-6D, MCS, or PCS scores) on sex, number of comorbidities, and clinical versus national sample. These results were unweighted.

Statistical analyses were performed using SAS version 9.3 (SAS Institute, Cary, North Carolina). We considered $P < .05$ to be statistically significant. The clinically important difference, defined as "the smallest difference in score in the domain of interest which patients perceive as beneficial,"[26] for the SF-6D is 0.04,[27] and 5 for the MCS and PCS scores.

The University of Pittsburgh Institutional Review Board approved this study (#PRO 13060301).

RESULTS

Response Rates

Within the 2011 MEPS sample, the response rate of individuals invited to complete the self-administered questionnaire was 93%.

There were 21,959 respondents with SF-6D scores in All MEPS and 414 respondents with SF-6D scores in Cardiology MEPS. During the study period, 1945 patients were eligible to complete the PRIcis intake at their first visit. Of these, 1514 (78%) participated and 1434 (95%) completed the SF-6D. The response rate was consistent across age and sex groups, except for those 75 years or older, for whom the total response rate was 65%.

Demographics

Demographics from the samples are presented in **Table 1**. The demographics of the Cardiology Local and the All MEPS samples are quite different; the Cardiology Local sample has a substantially lower proportion of females (42% vs 52%), lower proportion of racial minorities (8% minorities vs 18% minorities), and substantially higher mean CCI score (1.74 vs 0.62). The Cardiology MEPS subsample (43% female, 13% minorities; mean CCI score of 1.57) is a closer fit to the Cardiology Local sample with these demographic parameters than the All MEPS subsample. However, the distribution of specific comorbidities is still different between the Cardiology Local and Cardiology MEPS samples, with the Cardiology Local sample having more than twice the reported rate of congestive heart failure and cerebrovascular disease and half the rate of acute myocardial infarction.

Primary Outcome: SF-6D

When the Cardiology Local sample is compared with the All MEPS sample, there are statistically significant differences between mean SF-6D scores in the 65-to-74-years and 75-years-or-older age groups; however, these do not reach a clinically important difference of 0.04[27] (**Table 2**). In contrast, when the Cardiology Local sample is compared with the Cardiology MEPS sample, there are statistically significant differences between mean SF-6D scores in 4 of the 5 age groups (45-54, 55-64, 65-74, ≥75). In 3 of these age groups (45-64, 55-64, and 65-75), the Cardiology Local sample has a clinically important improvement over Cardiology MEPS.

In the regression-based sensitivity analyses combining Cardiology Local and Cardiology MEPS data, SF-6D scores were statistically important and had a clinically important improvement in the Cardiology Local sample for the same 4 age groups (**Table 3**). For example, in those aged 55 to 64, Cardiology Local SF-6D scores were 0.059 higher than the Cardiology MEPS scores after adjusting for sex and number of comorbid conditions. As expected, as the number of comorbid conditions increased, the SF-6D scores decreased. SF-6D scores were also lower for females than males.

Secondary Outcomes: MCS and PCS

The MCS mean scores in the Cardiology Local sample are statistically, although not clinically, higher than the All MEPS sample in 4 age groups (18-44, 55-64, 65-74, ≥75) (**Table 4**). There are no statistically significant differences in the MCS values of the Cardiology Local sample compared with the Cardiology MEPS sample. In

METHODS

TABLE 1. Demographics From the Cardiology Local Sample and the 2011 Medical Expenditures Panel Survey Sample[a]

	Cardiology Local[b]	All MEPS[c]	Cardiology MEPS[d]
Number of respondents			
All ages	1434	21,959	414
18-44 years	100	11,076	34
45-54 years	181	4068	61
55-64 years	399	3337	98
65-74 years	393	2002	125
≥75 years	363	1476	112
Female, %			
All ages	42	52	43
18-44 years	58	51	27
45-54 years	35	51	43
55-64 years	38	52	32
65-74 years	38	53	35
≥75 years	49	59	46
Race, %			
White	92	81	87
Black	6	12	9
Other	1	7	4
Unspecified	2	0	0
Charlson Comorbidity Index			
Mean score	1.74	0.62	1.57
Acute myocardial infarction, %	13.3	2.8	26.5
Congestive heart failure, %	11.3	0.6	2.6
Peripheral vascular disease, %	12.7	1.2	6.8
Cerebrovascular disease, %	6.4	0.2	1.4
Dementia, %	0.6	0.4	0.9
Chronic obstructive pulmonary disease, %	8.0	7.9	9.2
Rheumatoid disease, %	7.5	2.9	6.9
Peptic ulcer disease, %	3.3	0.4	1.2
Liver disease (mild), %	4.6	0.4	0.6
Diabetes, %	23.7	10.0	25.9
Hemiplegia or paraplegia, %	8.1	1.7	6.7
Renal disease, %	7.6	0.4	1.2
Cancer, %	10.8	5.8	14.3
Liver disease (moderate or severe), %	8.0	5.6	7.8
Metastatic cancer, %	1.0	0.4	1.2
AIDS, %	0	0	0

MEPS indicates Medical Expenditures Panel Survey.
[a] All MEPS values are weighted to be nationally representative, except N, which is the unweighted number of respondents.
[b] "Cardiology Local" refers to patients presenting to any outpatient cardiology clinic within our healthcare system from March 2012 to April 2013.
[c] "All MEPS" refers to the entire 2011 MEPS sample.
[d] "Cardiology MEPS" refers to the subset of patients from the 2011 MEPS sample who reported seeing a cardiologist.

the sensitivity analyses combining Cardiology Local and Cardiology MEPS data, MCS scores are statistically different in 3 age strata (55-64, 65-74, ≥75), with the Cardiology Local sample reaching clinically important differences (of at least 5) in 2 of the age groups (55-64 and 65-74) (**Table 5**).

For PCS mean scores stratified by age, the Cardiology Local sample is statistically different from the All MEPS sample in 2 age groups (18-44, 45-54), and the Cardiology Local sample has clinically importantly lower scores than All MEPS in these groups (Table 4). There are 3 statistically different values when comparing the Cardiology Local sample with the Cardiology MEPS sample (age groups 45-54, 65-74, ≥75), and the differences reach clinical significance (improvement) in the 45-to-54 age group. In the sensitivity analyses, combining Cardiology Local and Cardiology MEPS data, PCS scores are statistically different in 4 age strata (45-54, 55-64, 65-74, ≥75), with the Cardiology Local sample reaching improved clinically important differences in these 4 age groups (Table 5).

DISCUSSION

This report illustrates the use of MEPS to generate nationally representative HRQoL benchmark scores for patients receiving care from a cardiologist. We used these benchmark scores to add an important dimension of understanding to the routine HRQoL data collected in cardiology clinics of a large healthcare system. Compared with the national benchmarks, the local cardiology clinics had both statistically significant and clinically important improvements in health preference scores for those aged 45 to 74. Our system did not have any scores that were statistically or clinically lower than the national benchmarks.

To better understand why overall health preference scores for local cardiology clinics were different than the national benchmarks, we also calculated MCS and PCS scores from the PROs. These analyses showed fewer differences than the overall health preference score. Compared with the national benchmarks, the improvements in the local cardiology clinics were generally larger for PCS than MCS scores.

TABLE 2. Primary Outcome: SF-6D Scores Stratified by Age[a]

Age Group, Years	Cardiology Local,[b] Mean (SE)	All MEPS,[c] Mean (SE)	P Value Compared With Clinical Sample	Cardiology MEPS,[d] Mean (SE)	P Value Compared With Clinical Sample
18-44	0.79 (0.015)	0.81 (0.002)	.34	0.77 (0.028)	.51
45-54	0.77 (0.011)	0.78 (0.003)	.99	0.70 (0.026)	.001
55-64	0.77 (0.007)	0.76 (0.004)	.22	0.73 (0.018)	.040
65-74	0.78 (0.007)	0.76 (0.004)	.036	0.74 (0.016)	.009
≥75	0.72 (0.007)	0.71 (0.006)	<.0001	0.70 (0.016)	<.0001

MEPS indicates Medical Expenditures Panel Survey; SE, standard error; SF-6D, Short Form 6-Dimension questionnaire.
[a]The SF-6D is a health utility measure ranging from 0 to 1.0. Higher scores indicate better health-related quality of life.
[b]"Cardiology Local" refers to patients presenting to any outpatient cardiology clinic within our healthcare system from March 2012 to April 2013.
[c]"All MEPS" refers to the entire 2011 MEPS sample.
[d]"Cardiology MEPS" refers to the subset of patients from the 2011 MEPS sample who reported seeing a cardiologist.

TABLE 3. Primary Outcome: SF-6D Regression Coefficients[a]

Age Group, Years	Intercept (P value)	Number of Comorbid Conditions (P value)	Male vs Female (P value)	Cardiology Local[b] vs Cardiology MEPS[c] (P value)
18-44	0.755 (<.0001)	−0.019 (.03)	0.029 (.28)	0.047 (.11)
45-54	0.713 (<.0001)	−0.020 (<.0001)	0.019 (.36)	0.071 (.002)
55-64	0.740 (<.0001)	−0.020 (<.0001)	0.011 (.44)	0.058 (.001)
65-74	0.706 (<.0001)	−0.009 (.005)	0.056 (<.0001)	0.059 (.001)
≥75	0.645 (<.0001)	−0.008 (.008)	0.029 (.03)	0.034 (.033)

MEPS indicates Medical Expenditures Panel Survey; SF-6D, Short Form 6-Dimension questionnaire.
[a]The SF-6D is a health utility measure ranging from 0 to 1.0. Higher scores indicate better health-related quality of life.
[b]"Cardiology Local" refers to patients presenting to any outpatient cardiology clinic within our healthcare system from March 2012 to April 2013.
[c]"Cardiology MEPS" refers to the subset of patients from the 2011 MEPS sample who reported seeing a cardiologist.

These findings suggest that there is greater room for improvement in mental health outcomes than physical health outcomes; focusing quality improvement on mental health outcomes within our cardiology clinics may have a great impact on patient HRQoL. Findings such as these illustrate the power of using generic health measurement, even in subspecialty care.

This study also shows the limitations of comparing normative values in clinic populations to the general population. There are several generic HRQoL measures that have been included in nationally representative datasets, such as MEPS, the Joint Canada/United States Survey of Health, the National Health Measurement Study, and the US Valuation of the EuroQol 5-Dimension (EQ-5D) Health States Survey[28]; however, these datasets are primarily filled with healthy respondents. When the Cardiology Local sample was compared with the All MEPS sample in this study, there were no clinically important differences in health preference scores or MCS

scores. The Cardiology Local sample had clinically importantly worse PCS scores in 2 of the age groups (18-44 and 45-54). Compared with the individuals in MEPS who reported seeing a cardiologist, the Cardiology Local sample was the same as or better than, the Cardiology MEPS benchmarks.

MEPS is a rich dataset with information about many medical expenditures, including subspecialty visits, and it also includes generic health preference measures, such as the EQ-5D from 2000 to 2003, the SF-12v1 from 2000 to 2002, and the SF-12v2 from 2003 to present. The analysis shown in this report could be used to create similar benchmarks for other clinic or patient populations for which the MEPS sampling would reflect the local sample of interest. For example, it would be possible to select respondents in MEPS who have been to an endocrinology clinic or who report expenditures for thyroid disorders. Selecting rare conditions would reduce the sample size, although analysts could consider combining multiple years of MEPS data for more respondents. MEPS is a representative sample of the US civilian noninstitutionalized population, explicitly excluding institutionalized individuals in prisons or nursing homes and US citizens in the military. We expected these exclusions to bias our results in the opposite direction that was found, as the clinics in our healthcare system serve individuals in nursing homes.

Comparing local cardiology clinic samples to MEPS is limited by differing methods of eliciting the comorbid conditions appearing in each dataset. Comorbid conditions in the local dataset were included in the problem list or medical history in the EHR; conditions are physician-coded and do not have to be active at the time. Comorbid conditions in MEPS are reported by the respondent for the Household Component field and then recorded by the interviewer as verbatim text. This text was then coded by professional coders to fully specified *ICD-9-CM* codes, including medical condition and V codes. Although codes are verified, and error-rate percentages for any coder are low, the Agency for Healthcare Research and Quality recommends against assuming that household respondents have high accuracy in reporting condition data.[29] More work is necessary to determine the comparability of these methods to elicit comorbid conditions.

Usual techniques for case-mix adjustment include age, sex, and comorbidity information.[30] Our primary analyses compared age

METHODS

TABLE 4. Secondary Outcomes: Mental Component Summary and Physical Component Summary Scores[a]

Age Group, Years	Cardiology Local,[b] Mean (SE)	All MEPS,[c] Mean (SE)	P Value Compared With Clinical Sample	Cardiology MEPS,[d] Mean (SE)	P Value Compared With Clinical Sample
MCS					
18-44	48.3 (1.12)	50.8 (0.13)	.017	48.4 (2.19)	.90
45-54	49.6 (0.82)	50.3 (0.22)	.29	47.0 (1.99)	.27
55-64	52.4 (0.47)	50.9 (0.22)	.012	50.5 (1.11)	.16
65-74	54.2 (0.45)	52.7 (0.24)	.011	52.3 (1.13)	.16
≥75	54.2 (0.46)	51.7 (0.37)	<.0001	52.0 (1.16)	.11
PCS					
18-44	48.7 (1.03)	53.0 (0.11)	.003	49.9 (1.62)	.96
45-54	46.3 (0.84)	49.4 (0.23)	.024	42.8 (2.06)	.041
55-64	44.3 (0.60)	46.6 (0.28)	.073	43.4 (1.70)	.27
65-74	44.1 (0.56)	44.7 (0.35)	.45	42.3 (1.36)	.051
≥75	38.7 (0.60)	38.8 (0.40)	.17	35.8 (1.33)	.007

MCS, Mental Component Summary; MEPS, Medical Expenditures Panel Survey; PCS, Physical Component Summary; SE, standard error.

[a]The MCS and PCS are health profile scores which are constructed so that 50 is the mean score in the US general population. Higher scores indicate better health-related quality of life. MCS and PCS scores from the clinical sample, which use version 1 of the 36-Item Short Form Survey, have been adjusted to be comparable to version 2 of the 12-Item Short Form Survey scores collected in the MEPS using the standard method.[17]

[b]"Cardiology Local" refers to patients presenting to any outpatient cardiology clinic within our healthcare system from March 2012 to April 2013.

[c]"All MEPS" refers to the entire 2011 MEPS sample.

[d]"Cardiology MEPS" refers to the subset of patients from the 2011 MEPS sample who reported seeing a cardiologist.

TABLE 5. Secondary Outcomes: Regression Coefficients for the Mental Component Summary and Physical Component Summary Scores[a]

Age Group, Years	Intercept (P value)	Number of Comorbid Conditions (P value)	Male vs Female (P value)	Cardiology Local[b] vs Cardiology MEPS[c] (P value)
MCS				
18-44	48.1 (<.0001)	−0.8 (.23)	1.2 (.54)	0.7 (.76)
45-54	48.5 (<.0001)	−1.0 (.004)	−1.2 (.44)	3.2 (.063)
55-64	50.2 (<.0001)	−0.5 (.037)	0.1 (.91)	3.1 (.004)
65-74	49.8 (<.0001)	0.2 (.38)	1.7 (.046)	3.0 (.002)
≥75	52.0 (<.0001)	−0.5 (.023)	1.2 (.17)	2.5 (.013)
PCS				
18-44	46.5 (<.0001)	−0.7 (.24)	3.1 (.088)	1.7 (.39)
45-54	42.8 (<.0001)	−1.2 (.0005)	1.9 (.21)	3.7 (.032)
55-64	43.9 (<.0001)	−1.9 (<.0001)	0.9 (.40)	3.1 (.022)
65-74	39.5 (<.0001)	−1.1 (<.0001)	5.2 (<.0001)	3.4 (.003)
≥75	35.4 (<.0001)	−0.7 (.009)	2.5 (.022)	3.4 (.007)

MCS, Mental Component Summary; MEPS, Medical Expenditures Panel Survey; PCS, Physical Component Summary.

[a]The MCS and PCS are health profile scores constructed so that 50 is the mean score in the US general population. Higher scores indicate better health-related quality of life. MCS and PCS scores from the clinical sample, which use version 1 of the 36-Item Short Form Survey (SF-36v1), have been adjusted to be comparable to version 2 of the 12-Item Short Form Survey scores collected in MEPS using the standard method.[22]

[b]"Cardiology Local" refers to patients presenting to any outpatient cardiology clinic within our healthcare system from March 2012 to April 2013.

[c]"Cardiology MEPS" refers to the subset of patients from the 2011 MEPS sample who reported seeing a cardiologist.

strata between the Cardiology Local sample and the Cardiology MEPS subsample because the number of comorbid conditions and the proportion of females were similar in each data source. Given the small strata sample sizes, we did not further stratify by sex. Given the small number of nonwhite respondents, we were unable to look for differences by race, although there are well-known differences in HRQoL reports by race in the general population.[31]

The stratified analysis allowed us to incorporate MEPS weights so the estimates are representative of the US noninstitutionalized civilian population. For a sensitivity analysis, we combined the Cardiology Local and Cardiology MEPS samples, which adjusted for sex and number of comorbid conditions within the age strata. These analyses gave similar results to the stratified analyses but did not allow the use of MEPS weights. Techniques should be developed to include case adjustment and weighting when comparing clinic samples to national normative values.

CONCLUSIONS

Our work demonstrates that collection and interpretation of PROs in a clinical setting is feasible. Furthermore, incorporating PROs provides information that complements traditional clinical measures by emphasizing patient-centered care. We illustrate a technique to create benchmarks to allow interpretation of patient-reported outcomes in a clinic setting. Our findings suggest the clinical practices in our sample are achieving or exceeding benchmarks created from nationally representative data. They also provide clues about areas of improvement that could be high yield within these practices. If used more broadly, PROs could add an important dimension to our understanding of high-value care.

Author Affiliations: Division of General Internal Medicine (JH, SS, LY, WT, DK) and Division of Cardiology (JT), University of Pittsburgh Medical Center, Pittsburgh, PA; Health System Innovation and Research, University of Utah Schools of the Health Sciences (RH), Salt Lake City, UT.

Address Correspondence to: Janel Hanmer, MD, PhD, Department of General Internal Medicine, University of Pittsburgh Medical Center, 230 McKee Pl, Ste 600, Pittsburgh, PA 15213. E-mail: hanmerjz@upmc.edu.

REFERENCES

1. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life: a conceptual model of patient outcomes. *JAMA*. 1995;273(1):59-65.

2. Lohr KN. Quality measures: patient-reported outcomes for quality improvement of clinical Practice. Agency for Healthcare Research and Quality website. http://www.qualitymeasures.ahrq.gov/expert/expert-commentary.aspx?id=36851. Published May 14, 2012. Accessed January 14, 2015.

3. Osoba D. Translating the science of patient-reported outcomes assessment into clinical practice. *J Natl Cancer Inst Monogr.* 2007;(37):5-11.

4. Snyder CF, Aaronson NK, Choucair AK, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. *Qual Life Res.* 2012;21(8):1305-1314. doi: 10.1007/s11136-011-0054-x.

5. Spertus JA. Evolving applications for patient-centered health status measures. *Circulation.* 2008;118(20):2103-2110. doi: 10.1161/CIRCULATIONAHA.107.747568.

6. Fryback DG, Dunham NC, Palta M, et al. US norms for six generic health-related quality-of-life indexes from the National Health Measurement study. *Med Care.* 2007;45(12):1162-1170.

7. McDowell I. *Measuring Health: A Guide to Rating Scales and Questionnaires*. 3rd ed. New York, NY: Oxford University Press; 2006.

8. Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ*. 1986;5(1):1-30.

9. Hanmer J, Lawrence WF, Anderson JP, Kaplan RM, Fryback DG. Report of nationally representative values for the noninstitutionalized US adult population for 7 health-related quality-of-life scores. *Med Decis Making.* 2006;26(4):391-400.

10. Luo N, Johnson JA, Shaw JW, Feeny D, Coons SJ. Self-reported health status of the general adult US population as assessed by the EQ-5D and Health Utilities Index. *Med Care.* 2005;43(11):1078-1086.

11. Sullivan PW, Ghushchyan V. Preference-based EQ-5D index scores for chronic conditions in the United States. *Med Decis Making.* 2006;26(4):410-420.

12. Iezzoni LI. The risks of risk adjustment. *JAMA. 1997;*278(19):1600-1607.

13. Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: a method for continuous quality improvement in health. *Healthc Policy.* 2012;7(4):e101-e119.

14. Module 7. measuring and benchmarking clinical performance. Agency for Healthcare Research and Quality website. http://www.ahrq.gov/professionals/prevention-chronic-care/improve/system/pfhandbook/mod7.html. Accessed January 14, 2015.

15. Chan PS, Oetgen WJ, Buchanan D, et al. Cardiac performance measure compliance in outpatients: the American College of Cardiology and National Cardiovascular Data Registry's PINNACLE (Practice Innovation And Clinical Excellence) program. *J Am Coll Cardiol.* 2010;56(1):8-14. doi: 10.1016/j.jacc.2010.03.043.

16. Medical Expenditure Plan Survey (MEPS). Agency for Healthcare Research and Quality website. http://www.ahrq.gov/research/data/meps/index.html. Accessed January 14, 2014.

17. Ware JE, Kosinski M, Keller SD. *How to Score the SF-12 Physical and Mental Health Summary Scales*. 3rd ed. Boston, MA: The Health Institute, New England Medical Center, 1998.

18. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-item health survey 1.0. *Health Econ.* 1993;2(3):217-227.

19. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care.* 2004;42(9):851-859.

20. Torrance GW. Measurement of health state utilities for economic appraisal: a review. *J Health Econ.* 1986;5(1):1-30.

21. Gandek B, Ware JE, Aaronson NK, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol*. 1998;51(11):1171–1178.

22. Ware JE Jr, Kosinski M, Keller SD. A 12-item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996;34(3):220-233.

23. 36-Item Short Form Survey scoring instructions. RAND Health website. http://www.rand.org/health/surveys_tools/mos/36-item-short-form/scoring.html. Accessed April 2016.

24. 12-Item Short Form Survey. RAND Health website. http://www.rand.org/health/surveys_tools/mos/mos_core_12item.html. Accessed April 2016.

25. D'Hoore W, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson comorbidity index with administrative data bases. *J Clin Epidemiol.* 1996;49(12):1429-1433.

26. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10(4):407-415.

27. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res.* 2005;14(6):1523-1532.

28. Hanmer J, Hays RD, Fryback DG. Mode of administration is important in US national estimates of health-related quality of life. *Med Care.* 2007;45(12):1171-1179.

29. Medical Expenditure Panel Survey—MEPS HC-146: 2011 medical conditions. Agency for Healthcare Research and Quality website. http://meps.ahrq.gov/mepsweb/data_stats/download_data/pufs/h146/h146doc.shtml#2524. Published September 2013. Accessed January 14, 2015.

30. Pope GC, Kautter J, Ingber MJ, Freeman S, Sekar R, Newhart C. Evaluation of the CMS-HCC Risk Adjustment Model [final report]. CMS website. https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/Evaluation_Risk_Adj_Model_2011.pdf. Accessed January 14, 2015.

31. Robert SA, Cherepanov D, Palta M, Dunham NC, Feeny D, Fryback DG. Socioeconomic status and age variations in health-related quality of life: results from the National Health Measurement Study. *J Gerontol B Psychol Sci Soc Sci.* 2009;64(3):378-389. doi: 10.1093/geronb/gbp012. ■