

Evaluating the Utility of Existing Patient-Reported Outcome Scales in Novel Patient Populations with Pancreatic Cancer, Lung Cancer, and Myeloproliferative Neoplasms Using Medicare Current Beneficiary Survey Data

Jasmina I. Ivanova · Daniel S. Mytelka · Mei Sheng Duh ·
Howard G. Birnbaum · Alice Kate Cummings ·
Alexandra M. San Roman · Gregory L. Price · Ralph W. Swindle

Published online: 5 July 2013

© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract

Background While there are validated patient-reported outcomes (PRO) instruments for use in specific cancer populations, no validated general instruments exist for use in conditions common to multiple cancers, such as muscle wasting and consequent physical disability. The Medicare Current Beneficiary Survey (MCBS), a survey in a nationally representative sample of Medicare beneficiaries, includes items from three well known scales with general applicability to cancer patients: Katz activities of daily living (ADL), Rosow–Breslau instrumental ADL (IADL), and a subset of physical performance items from the Nagi scale.

Objective This study evaluated properties of the Katz ADL, Rosow–Breslau IADL, and a subset of the Nagi scale in patients with pancreatic cancer, lung cancer, and myeloproliferative neoplasms (MPN) using data from MCBS linked with Medicare claims in order to understand the potential utility of the three scales in these populations; understanding patient-perceived significance was not in scope.

Methods The study cohorts included Medicare beneficiaries aged ≥ 65 years as of 1 January of the year of their first cancer diagnosis with one or more health assessments

in a community setting in the MCBS Access to Care data from 1991 to 2009. Beneficiaries had at least two diagnoses in de-identified Medicare claims data linked to the MCBS for one of the following cancers: pancreatic, lung, or MPN. The Katz ADL, Rosow–Breslau IADL, and Nagi scales were calculated to assess physical functioning over time from cancer diagnosis. Psychometric properties for each scale in each cohort were evaluated by testing for internal consistency, test–retest reliability, and responsiveness by comparing differences in mean scale scores over time as cancer progresses, and differences in mean scale scores before and after hospitalization (for lung cancer cohort).

Results The study cohorts included 90 patients with pancreatic cancer, 863 with lung cancer, and 135 with MPN. Among each cancer cohort, the Katz ADL, Rosow–Breslau IADL, and Nagi scales had acceptable internal consistency (Cronbach’s alpha generally between 0.70 and 0.90) and test–retest reliability for consecutive surveys before diagnosis and consecutive surveys after diagnosis (when patients’ functioning was more stable). Compared with mean scale scores at the survey 1–2 years before cancer diagnosis (baseline), mean scale scores at the first survey after cancer diagnosis were significantly higher ($P < 0.05$), indicating worsening, for Katz ADL, Rosow–Breslau IADL, and Nagi scales (items scored 0–1) (0.54 vs. 1.45, 1.15 vs. 2.20, and 2.29 vs. 3.08, respectively, for pancreatic cancer; 0.73 vs. 1.24, 1.29 vs. 2.01, and 2.41 vs. 2.85 for lung cancer; and 0.44 vs. 0.86, 0.87 vs. 1.36, and 1.87 vs. 2.32 for MPN). Among lung cancer patients, scale scores increased significantly following a hospitalization, suggesting a worsening of functional status.

Conclusions The Katz ADL, Rosow–Breslau IADL, and Nagi scales collected in the MCBS demonstrate acceptable internal consistency and test–retest reliability among patients with pancreatic cancer, lung cancer, and MPN, and

J. I. Ivanova
Analysis Group, Inc., New York, NY, USA

D. S. Mytelka · G. L. Price · R. W. Swindle
Eli Lilly and Company, Indianapolis, IN, USA

M. S. Duh · H. G. Birnbaum (✉) · A. K. Cummings ·
A. M. San Roman
Analysis Group, Inc., 111 Huntington Avenue,
10th Floor, Boston, MA 02199, USA
e-mail: hbirnbaum@analysisgroup.com

are consistent with clinical worsening following diagnosis or hospitalization. These results suggest that using retrospective data may allow researchers to conduct preliminary assessments of existing PRO instruments in new populations of interest and generate useful exploratory disease information before embarking on de novo PRO development.

Key Points for Decision Makers

- The Katz activities of daily living (ADL), Rosow–Breslau instrumental ADL, and Nagi scales demonstrated acceptable reliability and responsiveness among patients with pancreatic cancer, lung cancer, or myeloproliferative neoplasms in the Medicare Current Beneficiary Survey
- Using retrospective survey data allows researchers to conduct preliminary assessment of existing patient-reported outcomes (PRO) scales (or select items from them) in populations of interest where de novo instrument development for each population may be impractical

1 Introduction

A number of instruments measuring patient-reported outcomes (PROs) are in common use in cancer populations. Some of these instruments are validated for use in specific individual cancer indications (e.g., Functional Assessment of Cancer Therapy-Lung [FACT-L] [1] and Functional Assessment of Cancer Therapy-Pancreatic cancer [FACT-PA] [2, 3]). Others are general instruments independent of cancer (e.g., Euro-QoL 5-Dimension [EQ-5D] [4]) or focus on a specific symptom that may present in similar ways in multiple cancers (e.g., Functional Assessment of Cancer Therapy: Fatigue [FACT-F] [5]). Developing PRO instruments that can be used to evaluate heterogeneous conditions like cachexia that can present in a variety of ways in different cancers can be very challenging. In this study, we describe an approach for the initial evaluation of the psychometric properties of PRO instruments in multiple cancers using retrospective data in order to understand whether the instruments could generate useful exploratory information in new populations.

Cachexia is a multifactorial syndrome defined by an ongoing loss of skeletal muscle mass that cannot be fully reversed by conventional nutritional support and leads to progressive functional impairment [6]. As cachexia may differ by cancer type, it was of interest to identify attributes that could be used to evaluate the impact of cancer

cachexia across multiple cancers without de novo instrument development for each cancer type. An expected impact of cachexia is a decline in daily functioning, and the goal of this study was to evaluate changes in measures of patient daily function over time across a set of patients with three different cancer indications (pancreatic cancer, lung cancer, or myeloproliferative neoplasms [MPN]) using existing data. The MPN cohort included chronic myeloid leukemia (CML) and non-CML patients.

Three well known general scales potentially pertinent to the decline in health functioning in pancreatic cancer, lung cancer, or MPN are included in the Medicare Current Beneficiaries Survey (MCBS), a multipurpose survey of a nationally representative sample of the Medicare population linked to de-identified Medicare claims data [7]. These scales are the Katz [8] activities of daily living (ADL) items, the Rosow–Breslau [9] instrumental ADL (IADL) items, and the Nagi [10] physical performance items. These three scales have demonstrated reliability and consistent relationships to a number of objective tests of performance and health outcomes in a wide variety of prominent studies and populations taking place over decades of research [11–13]. For example, the consistency of these scales has been evaluated among 5,986 older adults from the Longitudinal Study on Aging [11], and the correlation of ADL scales with hemoglobin was evaluated among 586 elderly cancer patients undergoing chemotherapy [12]. The association between comorbidities and functional status measures using all three scales was evaluated among 4,162 older adults, including 376 self-reporting diagnoses of cancer enrolled in the Duke Established Populations for Epidemiologic Studies of the Elderly [13].

All of the items from the Katz ADL and Rosow–Breslau IADL scales, and a subset of the items from the Nagi scale, were selected for inclusion in the MCBS. The MCBS has been used to evaluate health functioning status in many populations—for example, in studies including older adults without initial ADL limitations [14], older women [15], patients diagnosed with Alzheimer’s and other dementias [16], patients diagnosed with diabetes [17], or studies evaluating the impact of cancer diagnosis on functional status [18].

This research evaluated the psychometric properties of the Katz, Rosow–Breslau, and Nagi instruments in the subpopulations of interest (pancreatic cancer, lung cancer, and MPN) using existing data from the MCBS and linked Medicare claims data. Specifically, this study evaluated the internal consistency, test–retest reliability, and mean scale scores of the scales over time as cancer progresses, and the association of the scales with a clinical outcome: hospitalization. This study evaluated the potential use of existing PRO instruments in future clinical trials with specific cancer subpopulations for which new treatments are under

development. The study did not aim to demonstrate full validity of the instruments analogous to US FDA guidelines but to demonstrate the usefulness of the instruments in understanding disease in additional populations. Understanding the usefulness of well known generic instruments in specific disease states could provide a less costly and more convincing alternative to developing and validating new disease-specific PRO scales for every disease state of interest, particularly when many characteristics are common across a number of related and relatively rare diseases.

2 Methods

2.1 Data Source

This study used an integrated database combining survey responses from the MCBS Access to Care files and the Centers for Medicare and Medicaid Services (CMS) administrative claims data for Medicare beneficiaries [19]. The data were available from 1991 to 2009 and included a national sample of approximately 16,000 participants each year. The MCBS Access to Care files contain de-identified information on socioeconomic and demographic characteristics, health status and functioning, health insurance, access to health care, satisfaction with care, and usual source of care for a representative sample of the Medicare population. De-identified administrative claims of survey respondents for medical services covered under Medicare Part A and Part B for survey respondents were linked to surveys of beneficiaries. Patient diagnoses, medical services, and pharmacy prescriptions covered under Medicare Part D were available for 2006–2009. A key feature of the survey is its longitudinal design. Each sample person is

interviewed up to three times a year for up to 4 years or until death or loss to follow-up [20]. Functional status is evaluated once a year in the last survey of each year.

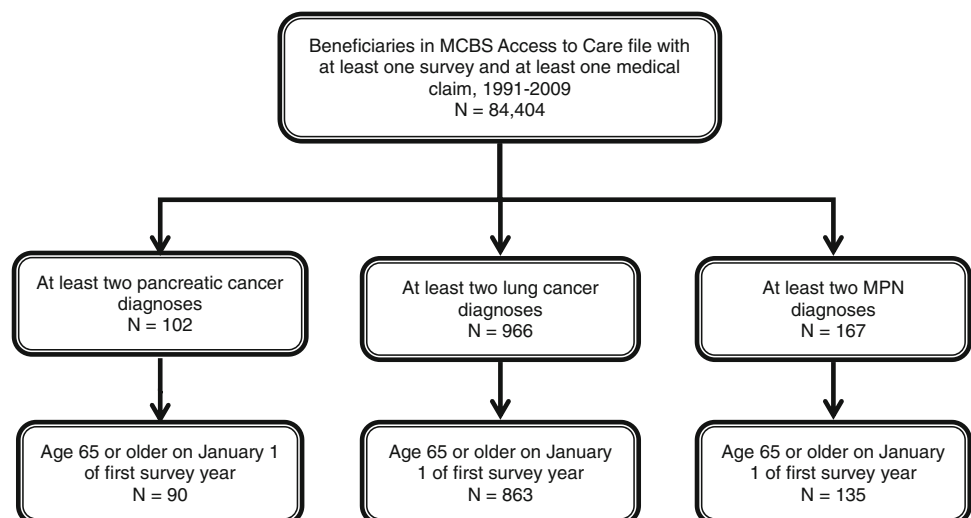
2.2 Patient Selection

Patients with at least two diagnoses on different dates for one of the following cancers were selected from the database: pancreatic (*International Classification of Diseases, Ninth Revision* [ICD-9-CM] code: 157.xx), lung (162.2x–162.9x), or MPN (205.1x, 205.9x, 238.4x, 238.71, 238.76, 289.83). At least two diagnoses were required to exclude patients with potential rule-out cancer diagnosis. Patients were required to be aged 65 years and older as of 1 January of the year of their first cancer diagnosis and have at least one health assessment in the MCBS Access to Care data for community-dwelling individuals from 1991 to 2009. After the sample selection criteria were applied, the study samples included 90 patients diagnosed with pancreatic cancer, 863 patients diagnosed with lung cancer, and 135 patients diagnosed with MPN. See Fig. 1 for sample selection.

2.3 Patient Characteristics

Patient characteristics were described using survey responses and medical claims data in the year of first cancer diagnosis for the cancer cohorts and in the first year with a survey in the overall MCBS population. Patient characteristics were also evaluated for subpopulations of pancreatic and lung cancer patients with only one survey after cancer diagnosis and with two or more surveys after cancer diagnosis. Patient characteristics included demographic characteristics (identified using responses to the survey), comorbidity profile, the Charlson Comorbidity Index (CCI; a weighted sum of 17 conditions predictive of

Fig. 1 Sample selection. *MCBS* Medicare Current Beneficiaries Survey, *MPN* myeloproliferative neoplasms



1-year mortality, with an index range of 0–33) and the individual conditions included in the index [21, 22], the proportions of patients actively using cancer treatments, and annual medical costs. Medical costs were inflated to \$US, year 2009 values, using the Consumer Price Index for Medical Care [23]. No statistical comparisons were made among the overall MCBS populations and the cancer cohorts.

2.4 Patient-Reported Outcomes Scales and Scoring

Using the MCBS questionnaires for the period of 1991–2009, the Katz ADL scale, Rosow–Breslau IADL scale, and a subset of the Nagi scale, were calculated to assess health status and physical functioning over time. The Katz ADL scale questions in the MCBS asked patients whether they had any difficulty doing the following everyday activities by themselves, without special equipment, because of a health or physical problem: (1) bathing or showering; (2) dressing; (3) eating; (4) getting in or out of bed or chairs; (5) walking; and (6) using the toilet. The Rosow–Breslau IADL scale questions asked patients whether they had any difficulty doing the following everyday activities by themselves because of a health or physical problem: (1) using the telephone; (2) doing light housework (like washing dishes, straightening up, or light cleaning); (3) doing heavy housework (like scrubbing floors or washing windows); (4) preparing (your/his/her) own meals; (5) shopping for personal items (such as toiletries or medicines); and (6) managing money (like keeping track of expenses or paying bills). The MCBS asked 5 of the 12 Nagi health functioning questions about how difficult it is on average for patients to do each of the following activities because of a health or physical problem: (1) stooping, crouching, or kneeling; (2) lifting or carrying objects as heavy as 10 pounds, like a sack of potatoes; (3) writing or handling and grasping small objects; (4) walking for a quarter mile—that is, about 2 or 3 blocks; and (5) reaching or extending your arms above shoulder level.

Following an approach by Wolinsky et al. [24], items were scored as dichotomous variables (0 = no; 1 = yes), and scale scores were calculated as the sum of item scores. For the Nagi scale, two item-scoring approaches were used—scoring each item 0–1 (0 = no difficulty at all; 1 = any level of difficulty); and scoring each item 1–5 (1 = no difficulty at all; 2 = a little difficulty; 3 = some difficulty; 4 = a lot of difficulty; 5 = not able to do it).

Scale scores with missing item responses were imputed only if half the items from the original scale were present by pro-rating the score (e.g., if scale range is 0–6, three items were answered and the sum of the answered items was three, then scale score = 6). Scale score ranges were

based on the number of questions asked in the MCBS. For example, the original Nagi health functioning scale has 12 questions but only five were asked in the MCBS, therefore the Nagi scale range in this study was 0–5, with 0–1 item scoring, or 5–25 with 1–5 item scoring. Moreover, because fewer than six (half of the original 12 questions) were asked in the MCBS, there was no imputation of the Nagi scale score if a patient did not respond to one of the Nagi items asked in the MCBS.

2.5 Psychometric Properties Evaluation

2.5.1 Test for Internal Consistency

Internal consistency measures the homogeneity of the scale, or the extent to which various items included in a scale measure a single concept. Internal consistency was evaluated using Cronbach's alpha, which reflects the average correlation among all the items in the scale. In general, alpha of 0.7 or greater indicates acceptable reliability [25]. The internal consistency of each scale was evaluated at each survey relative to cancer diagnosis, from up to 3 years before cancer diagnosis to 4 years after cancer diagnosis depending on sample size, and for each cancer cohort.

2.5.2 Test–Retest Reliability

Test–retest reliability was used to assess the concordance between scale scores obtained from the same patient at different points in time. For each scale and cancer cohort, test–retest reliability was evaluated by calculating the intraclass correlation coefficient (ICC) [26] and the concordance correlation coefficient (CCC) [27] for consecutive annual surveys: before first cancer diagnosis; for the first survey before and the first survey after the cancer diagnosis; and for the first two surveys after the cancer diagnosis. The sample size for test–retest reliability assessment was smaller for each cohort because included patients were required to have two consecutive surveys. Test–retest reliability was also evaluated for subpopulations of pancreatic and lung cancer patients with only one survey after cancer diagnosis and with two or more surveys after cancer diagnosis.

2.5.3 Responsiveness

2.5.3.1 Mean Scale Scores Pre and Post Cancer Diagnosis

Mean scale scores were reported for two surveys before the first cancer diagnosis and two surveys after the first cancer diagnosis for each of the cancer cohorts, as well as for subpopulations of pancreatic and lung cancer patients with only one survey after cancer diagnosis and

with two or more surveys after cancer diagnosis. Generalized estimating equation models taking into account repeated patient measures were used to compare mean scale scores approximately 2 years before diagnosis (baseline) and the first survey after diagnosis, as well as to compare mean scale scores at baseline and the second survey after diagnosis. P values < 0.05 were considered to indicate statistically significant differences.

2.5.3.2 Comparison of Scale Scores Pre and Post Hospitalization Patient scale scores were compared before and after a hospitalization among lung cancer patients, the largest cohort, using paired t tests. Due to smaller sample sizes, similar analyses were not conducted in the pancreatic and MPN populations.

3 Results

3.1 Patient Characteristics

Patients with pancreatic cancer, lung cancer, or MPN had similar mean age and gender distribution as patients in the overall MCBS population, with the exception of a higher proportion of men among patients with lung cancer (likely due to higher historical smoking rates among men). Patients in cancer cohorts had a higher average CCI than patients in the MCBS population, and there was a higher proportion of patients with diagnoses for chronic conditions other than cancer (e.g., congestive heart failure, peripheral vascular disease, chronic pulmonary disease, diabetes for pancreatic and, to a lesser extent, lung cancer populations). Annual medical costs for pancreatic cancer, lung cancer, or MPN were also higher than those for the overall MCBS population (\$US56,023; \$US62,545; \$US20,734 vs. \$US9,088, respectively). Medical costs were mostly driven by outpatient/other costs and inpatient costs. See Table 1 for descriptive results. No statistical comparisons were conducted.

3.2 Test for Internal Consistency

The Katz ADL, Rosow–Breslau IADL, and Nagi scales had acceptable internal consistency (Cronbach’s alpha generally between 0.70 and 0.90) among patients with pancreatic cancer, lung cancer, or MPN (Table 2).

3.3 Test–Retest Reliability

Overall, the Katz ADL, Rosow–Breslau IADL, and Nagi scales had good test–retest reliability for consecutive surveys before diagnosis and consecutive surveys after diagnosis even though consecutive surveys were conducted a

year apart. As expected, test–retest reliability was higher for consecutive surveys before diagnosis and consecutive surveys after diagnosis (when patients’ functioning was more stable) than for the survey preceding cancer diagnosis and the first survey after cancer diagnosis, except for the MPN population. Test–retest reliability was the highest among pancreatic cancer patients for consecutive surveys before diagnosis and consecutive surveys after diagnosis (Table 3). The ICCs for first and second consecutive surveys after diagnosis for Katz ADL, Rosow–Breslau IADL, and Nagi scales scored 0–1 and 1–5 were 0.83, 0.77, 0.79, and 0.93, respectively, among pancreatic cancer patients; 0.58, 0.61, 0.54, and 0.67 among lung cancer patients; and 0.73, 0.72, 0.72, and 0.80 among MPN patients.

3.4 Responsiveness

3.4.1 Mean Scale Scores Over Time

Mean Katz ADL, Rosow–Breslau IADL, and Nagi scale scores were increasing (suggesting worsening of functional status) before cancer diagnosis and immediately after diagnosis among the three cancer cohorts. Changes in mean score were observed in multiple scale items. Compared with mean scale scores at the survey 1–2 years before cancer diagnosis (baseline), mean scale scores at the first survey after cancer diagnosis were significantly ($P < 0.05$) higher for Katz ADL, Rosow–Breslau IADL, and Nagi scales with items scored 0–1 (0.54 vs. 1.45, 1.15 vs. 2.20, and 2.29 vs. 3.08, respectively, for pancreatic cancer; 0.73 vs. 1.24, 1.29 vs. 2.01, and 2.41 vs. 2.85 for lung cancer; and 0.44 vs. 0.86, 0.87 vs. 1.36, and 1.87 vs. 2.32 for MPN). Mean scale score at the second survey after cancer diagnosis was also significantly higher compared with baseline except for the Katz ADL scale among pancreatic cancer patients (Table 4).

3.4.2 Comparison of Scale Scores Pre and Post Hospitalization

Among lung cancer patients with at least one hospitalization, Katz ADL, Rosow–Breslau IADL, and Nagi scale scores (items scored 0–1) increased significantly following a hospitalization (from 0.89 to 1.29; from 1.41 to 2.16; from 2.57 to 3.14; respectively), suggesting a worsening of functional status (Table 5).

3.5 Population Heterogeneity

Pancreatic and lung cancer patients with only one survey after cancer diagnosis had different characteristics, suggestive of more advanced cancer, than those with two or more surveys after diagnosis. Pancreatic and lung cancer

Table 1 Patient characteristics^a

	MCBS population (<i>N</i> = 83,985)			Pancreatic cancer pts (<i>N</i> = 90)			Lung cancer pts (<i>N</i> = 863)			MPN pts (<i>N</i> = 135)		
	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
Demographics												
Age on index date (years)	76.3	7.5	76.0	77.1	7.1	76.0	75.9	6.5	76.0	76.5	7.0	76.0
Male (%)	41.6			42.2			52.1			42.2		
Ever smoked tobacco (%)	54.8			57.8			87.6			62.4		
Weight (lbs)												
Female	146	32	142	151	38	146	139	32	136	145	29	145
Male	179	33	175	169	30	167	171	32	170	179	28	180
Comorbidities												
Charlson comorbidity index	1.3	1.7	1.0	6.3	2.8	7.0	5.9	2.8	6.0	2.3	2.4	2.0
Metastatic solid tumor (%)	1.2			50.0			41.7			2.2		
Malignancies (%)	8.0			48.9			57.2			17.0		
Congestive heart failure (%)	11.9			34.4			33.7			30.4		
Chronic pulmonary disease (%)	12.0			30.0			64.3			28.1		
Diabetes, mild to moderate (%)	10.6			27.8			15.9			10.4		
Diabetes, chronic complications (%)	4.2			22.2			9.3			5.9		
Peripheral vascular disease (%)	9.4			18.9			25.0			23.7		
Cerebrovascular disease (%)	9.3			14.4			20.3			22.2		
Renal disease (%)	2.5			14.4			7.1			13.3		
Peptic ulcer disease (%)	1.9			13.3			5.7			4.4		
Myocardial infarction (%)	2.6			5.6			11.7			5.9		
Rheumatologic disease (%)	2.4			4.4			4.1			2.2		
Hemiplegia or paraplegia (%)	0.8			3.3			2.3			1.5		
Mild liver disease (%)	0.4			2.2			1.9			2.2		
Dementia (%)	2.8			1.1			4.1			7.4		
Moderate or severe liver disease (%)	0.1			1.1			0.5			1.5		
HIV/AIDS (%)	0.0			0.0			0.1			0.0		
Secondary malignancies												
Liver and intrahepatic bile ducts (%)	0.2			27.8			7.1			0.0		
Rectum, rectosigmoid junction, anus, and colon (%)	1.1			10.0			7.0			0.7		
Prostate (%)	2.9			7.8			6.5			3.7		
Small intestine (%)	0.0			5.6			0.3			0.0		
Stomach (%)	0.1			4.4			0.7			0.0		
Lymphatic (%)	0.4			3.3			2.8			1.5		
Breast (%)	1.8			2.2			8.3			5.9		
Active treatments												
Chemotherapy (%)	2.2			40.0			32.3			10.4		
Radiotherapy (%)	1.0			17.8			25.5			1.5		
Surgery (%)	0.1			5.6			14.9			4.4		
Red blood cell growth factors (%)	0.5			5.6			9.8			4.4		
White blood cell growth factors (%)	0.2			2.2			7.2			1.5		
Targeted therapy (%)	0.9			1.1			8.5			0.7		
Annual medical costs (\$US, year 2009 values)												
Inpatient	2,518	9,753	0	16,859	21,621	8,764	15,451	21,465	8,964	6,067	14,702	0
Emergency room	218	1,913	0	760	1,291	233	751	1,692	103	1,021	4,944	0
Skilled nursing facility	278	2,205	0	850	3,234	0	1,141	4,804	0	526	2,654	0

Table 1 continued

	MCBS population (<i>N</i> = 83,985)			Pancreatic cancer pts (<i>N</i> = 90)			Lung cancer pts (<i>N</i> = 863)			MPN pts (<i>N</i> = 135)		
	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
Hospice	76	1,580	0	1,396	4,879	0	1,248	6,059	0	0	0	0
Home health agency	465	2,893	0	2,776	7,761	0	1,264	3,731	0	864	2,675	0
Durable medical equipment	447	4,385	0	966	2,751	0	2,188	7,589	0	780	3,048	0
Outpatient/other	5,086	17,338	1,195	32,417	45,009	17,850	40,502	68,387	17,504	11,476	17,532	5,305
Total medical costs	9,088	25,788	1,436	56,023	56,866	42,629	62,545	78,605	37,970	20,734	33,938	8,281

MCBS Medicare Current Beneficiary Survey, MPN myeloproliferative neoplasms, pts patients

^a Characteristics were evaluated using responses to the survey (for demographics) and medical claims data (for comorbidities, active treatments, and medical costs) in the year of first diagnosis for cancer cohorts and in the year of first survey for the MCBS population

patients with only one survey after diagnosis had a higher proportion of patients with metastatic solid tumor diagnosis (58.5 vs. 29.2 % among pancreatic cancer patients, 52.1 vs. 26.1 %, among lung cancer patients) a higher proportion of some secondary malignancies (e.g., liver and intrahepatic bile ducts, and rectum, rectosigmoid junction, anus and colon for pancreatic and lung cancer patients; brain and spinal cord for lung cancer patients), higher proportions of chemotherapy and radiation therapy, a lower proportion of surgery (possibly because fewer advanced cancer patients could benefit from surgery) and higher medical costs (\$US57,375 vs. 52,302 among pancreatic cancer patients with 1 vs. 2 or more surveys; \$US82,744 vs. 43,758 among lung cancer patients with 1 vs. 2 or more surveys). No statistical comparisons were conducted.

Subpopulations of pancreatic cancer and lung cancer patients with only one survey after diagnosis also had higher mean scale scores at the first survey after diagnosis (worse functional status) than those with two or more surveys after diagnosis (no statistical comparison conducted), and, for most scales, lower test–retest reliability assessed between the survey preceding cancer diagnosis and the first survey after cancer diagnosis, consistent with having more advanced cancer and greater declines in functioning than those with two or more surveys after cancer diagnosis (data available upon request).

4 Discussion

This study described an approach for the initial evaluation of the psychometric properties of existing PRO scales from existing data to demonstrate their usefulness for understanding disease in new populations of interest. Specifically, data from the MCBS linked to Medicare claims data were used to test the psychometric properties of items from Katz ADL, Rosow–Breslau IADL, and Nagi scales among patients with pancreatic, lung, and MPN cancers. The data

collected in the MCBS demonstrated acceptable internal consistency (Cronbach's alpha generally between 0.70 and 0.90) among the cancer cohorts and test–retest reliability for consecutive surveys before diagnosis and consecutive surveys after diagnosis (when patients' functioning was more stable), even though consecutive surveys were conducted a year apart. Compared with mean scale scores at the survey 1–2 years before cancer diagnosis (baseline), mean scale scores at the first survey after cancer diagnosis were significantly higher. Among lung cancer patients, scale scores increased significantly following a hospitalization, suggesting a worsening of functional status. The sample size for patients with pancreatic cancer and MPN was too small to compare the difference in scale scores before and after hospitalization. The psychometric findings are encouraging that inclusion of the Katz, Rosow–Breslau, and Nagi scales in confirmatory clinical trials of pancreatic, lung, and MPN cancers is appropriate, and could demonstrate changes in functional outcomes associated with efficacious treatment. Although the 1-year interval between functional assessments in MCBS precludes a fine-grained examination of the sensitivity of the scales over brief time-spans, the data do suggest the scales are sensitive, in cohort-level analyses, to functional changes associated with important events, such as cancer onset and hospitalization in the lung cancer cohort.

Researchers who wish to use or adapt existing PRO scales for a population of interest, in which the scales have not been previously validated, could incorporate a similar approach to create a PRO instrument that would provide useful information about the target population, even though the instrument should not be implied to be substitutable for one created by ground-up development. Review of existing data could initially test the psychometric properties and the sensitivity of the PRO scales in the population of interest, before conducting costly prospective development and validation work. Existing survey data that capture health status and functioning over time, especially if it is linked to

Table 2 Internal consistency testing: Cronbach's alpha coefficients^a using scales scores^b evaluated in surveys before and after initial cancer diagnosis

	Pancreatic cancer patients (N = 90)				Lung cancer patients (N = 863)				MPN patients (N = 135)								
	Survey pre diagnosis (years)		Survey post diagnosis (years)		Survey pre diagnosis (years)		Survey post diagnosis (years)		Survey pre diagnosis (years)		Survey post diagnosis (years)						
	-2	-1	1	2	-3	-2	-1	1	2	3	4	-2	-1	1	2	3	4
Katz ADL scale																	
Sample size	41	59	65	24	116	246	455	747	344	211	111	39	66	130	79	51	32
Cronbach's alpha	0.62	0.83	0.87	0.85	0.71	0.83	0.81	0.86	0.87	0.84	0.83	0.74	0.62	0.78	0.81	0.78	0.82
Rosow-Breslau IADL scale																	
Sample size	41	58	63	24	116	244	449	737	338	207	109	39	65	127	78	50	32
Cronbach's alpha	0.73	0.78	0.83	0.85	0.79	0.76	0.77	0.82	0.82	0.80	0.83	0.66	0.85	0.80	0.81	0.83	0.86
Nagi scale																	
Sample size	41	59	65	24	116	246	454	743	344	211	111	39	66	130	76	51	32
Cronbach's alpha (scored 0-1)	0.74	0.81	0.81	0.80	0.71	0.70	0.75	0.72	0.71	0.72	0.67	0.56	0.67	0.78	0.78	0.66	0.79
Cronbach's alpha (scored 1-5)	0.84	0.87	0.90	0.88	0.79	0.78	0.80	0.82	0.81	0.82	0.76	0.67	0.76	0.87	0.86	0.80	0.81

ADL activities of daily living, IADL instrumental activities of daily living, MPN myeloproliferative neoplasms

^a In general, Cronbach's alpha between 0.7 and 0.9 is considered acceptable [25]

^b Cronbach's alpha for each scale was calculated using only surveys with non-missing responses to all items in the scale

Table 3 Test–retest reliability using intraclass correlation coefficients (ICC) and concordance correlation coefficients (CCC) using scales scores^a for pairs of consecutive surveys before, before and after, and after diagnosis for cancer patients

	Pancreatic cancer patients (<i>N</i> = 90)			Lung cancer patients (<i>N</i> = 863)						MPN patients (<i>N</i> = 135)		
	–2 vs. –1	–1 vs. 1	1 vs. 2	–3 vs. –2	–2 vs. –1	–1 vs. 1	1 vs. 2	2 vs. 3	3 vs. 4	–2 vs. –1	–1 vs. 1	1 vs. 2
Katz ADL scale ^b												
Sample size	41	34	24	116	246	339	344	211	111	39	61	79
ICC	0.75	0.51	0.83	0.50	0.59	0.44	0.58	0.62	0.65	0.61	0.47	0.73
CCC	0.75	0.58	0.84	0.47	0.65	0.49	0.56	0.57	0.64	0.59	0.55	0.71
Rosow–Breslau IADL scale ^b												
Sample size	41	34	24	116	246	339	344	211	111	39	61	79
ICC	0.80	0.21	0.77	0.67	0.64	0.44	0.61	0.64	0.67	0.46	0.64	0.72
CCC	0.81	0.43	0.81	0.70	0.71	0.52	0.62	0.63	0.71	0.41	0.74	0.71
Nagi scale ^c												
Sample size	41	34	24	116	246	336	343	211	111	39	61	76
Scored 0–1												
ICC	0.68	0.39	0.79	0.51	0.58	0.39	0.54	0.61	0.48	0.64	0.59	0.72
CCC	0.71	0.54	0.85	0.60	0.64	0.52	0.62	0.64	0.55	0.65	0.65	0.76
Scored 1–5												
ICC	0.76	0.47	0.93	0.59	0.73	0.47	0.67	0.72	0.67	0.74	0.62	0.80
CCC	0.78	0.63	0.95	0.65	0.78	0.58	0.70	0.73	0.70	0.74	0.68	0.83

ADL activities of daily living, *IADL* instrumental activities of daily living, *MCBS* Medicare Current Beneficiaries Survey, *MPN* myeloproliferative neoplasms

^a For each scale, ICC and CCC were calculated using pairs of consecutive surveys with non-missing scale scores

^b Patients were required to have responses to at least half of the items included in the original scale in order to calculate values for a scale

^c Patients were required to have responses to all scale items asked in the MCBS in order to calculate values

claims so that patient diagnoses could be confirmed, are a valuable resource for researchers. This approach can help researchers evaluate the psychometric properties of scales, identify trends in health status/functioning changes in specific populations, and conduct hypothesis-generating analyses for future prospective studies. Advantages of using longitudinal retrospective survey data, and specifically MCBS linked to Medicare claims data, include the low cost (as it does not require recruitment and prospective data collection), a diverse nationally representative population of Medicare beneficiaries, the ability to select cohorts of interest using diagnoses and procedures recorded in claims data, bigger samples that may allow testing for statistical significance, PRO assessments across multiple years for many patients, and the availability of patient resource use and cost data from medical claims.

Retrospective data studies could be subject to limitations. Limitations of this study included relying on the accuracy of diagnosis coding in claims data to identify cancer patients and a lack of disease staging or cachexia-specific disease information. To ensure that patients had

pancreatic cancer, lung cancer, or MPN, two claims on different dates with cancer diagnosis codes were required. No washout period was required prior to first cancer diagnosis. To the extent that the first cancer diagnosis in the available claims data was not the first cancer diagnosis, the survey timing relative to cancer diagnosis may not be accurate, and changes in scale scores as cancer progresses may appear smaller. Preliminary exploratory analyses requiring washout of different durations before first cancer diagnosis suggested that findings were not sensitive to washout requirements. This study was limited to community-dwelling cancer patients aged 65 years and over at first diagnosis and therefore may not be generalizable to all patients with pancreatic cancer, lung cancer, or MPN. In addition, these populations may not be representative of other cancer or non-cancer populations.

Another limitation of this study is that no work was carried out to establish content validity, the extent to which the scales measure all the dimensions of the disease state. However, the purpose of the study was to initially evaluate the utility of using PRO scales in populations in which they

Table 4 Mean scales scores^a evaluated in surveys before and after first observed cancer diagnosis

	Pancreatic cancer patients (N = 90)				Lung cancer patients (N = 863)				MPN patients (N = 135)				
	Survey pre diagnosis (years)		Survey post diagnosis (years)		Survey pre diagnosis (years)		Survey post diagnosis (years)		Survey pre diagnosis (years)		Survey post diagnosis (years)		P value
	-2	-1	1	2	-2	-1	1	2	-2	-1	1	2	
Time from survey to index date (months), mean	16.8	5.1	4.7	18.1	18.0	5.9	6.0	18.8	18.5	6.3	6.3	18.8	
Katz ADL scale^b													
Sample size	41	59	65	24	246	455	747	344	39	66	130	79	
Mean	0.54	0.81	1.45	1.04	0.73	0.85	1.24	0.97	<0.0001*	<0.0001*	0.44	0.61	0.0043*
SD	1.00	1.50	1.90	1.70	1.30	1.40	1.80	1.70	1.00	1.00	1.40	1.30	0.0145*
Rosow-Breslau IADL scale^b													
Sample size	41	59	65	24	246	455	747	344	39	66	130	79	
Mean	1.15	1.47	2.20	1.83	1.29	1.34	2.01	1.73	<0.0001*	<0.0001*	0.87	1.17	0.0197*
SD	1.40	1.70	2.00	2.10	1.60	1.60	2.00	1.90	1.20	1.60	1.70	1.60	0.0213*
Nagi scale^c													
Sample size	41	59	65	24	246	454	743	344	39	66	130	76	
Scored 0-1													
Mean	2.29	2.24	3.08	2.63	2.41	2.42	2.85	2.81	<0.0001*	<0.0001*	1.87	2.23	0.0018*
SD	1.70	1.80	1.70	1.90	1.60	1.70	1.60	1.60	1.40	1.50	1.70	1.70	0.0008*
Scored 1-5													
Mean	10.37	10.53	13.17	11.13	10.56	10.90	12.33	11.78	<0.0001*	<0.0001*	8.92	10.15	0.0004*
SD	5.20	5.50	6.50	5.80	4.80	5.10	5.40	5.10	3.60	4.60	5.70	5.50	0.0003*

ADL activities of daily living, IADL instrumental activities of daily living, MCBS Medicare Current Beneficiaries Survey, MPN myeloproliferative neoplasms

* P < 0.05. P values are for comparisons of mean scores versus baseline mean scores (measured at the survey 1-2 years before diagnosis). Generalized estimating equations models were used for comparisons adjusting for correlation within the survey scores of the same patient, with exchangeable covariance structure between surveys within patient

^a For each scale, maximum score per patient survey is equivalent to the number of questions asked in the MCBS

^b Patients were required to have responses to at least half of the items included in the original scale in order to calculate values for the scale

^c Patients were required to have responses to all scale items asked in the MCBS in order to calculate values for the scale

Table 5 Impact of hospitalization on scale scores^a—lung cancer (*N* = 395)

	Surveys pre-hospitalization	Surveys post-hospitalization	<i>P</i> value
Katz ADL scale, mean (SD)	0.89 (1.41)	1.29 (1.79)	<0.0001*
Rosow–Breslau IADL scale, mean (SD)	1.41 (1.65)	2.16 (1.96)	<0.0001*
Nagi scale (scored 0–1), mean (SD)	2.57 (1.60)	3.14 (1.52)	<0.0001*
Nagi scale (scored 1–5), mean (SD)	11.27 (4.99)	13.17 (5.35)	<0.0001*

ADL activities of daily living, *IADL* instrumental activities of daily living

* *P* < 0.05

^a Scale scores for two consecutive surveys before and after hospitalization were compared using matched pairs *t* tests

have not been validated without costly de novo PRO instrument development. The Nagi scales evaluated in this study included only 5 of the 12 questions in the original Nagi questionnaire, but even the scales constructed from the five questions asked in the MCBS had good psychometric properties among the three cancer cohorts. Moreover, while patients participating in the MCBS were surveyed four times a year, the PRO scales were collected only once a year. More frequent assessments would have allowed for more precise evaluation of the test–retest reliability (typically evaluated within 2 weeks), the relationship between cancer diagnosis and PRO scales, as well as the association of PRO scales and clinical outcomes. While a finding of test–retest instability over consecutive annual periods would be ambiguous because it would be unclear whether it is due to measurement variability or changes over time, acceptable test–retest reliability for consecutive surveys before cancer diagnosis and following cancer diagnosis still suggests test–retest reliability. Direct linkage of the changes observed in daily function to cachexia was outside of the scope of this study. In addition, the small sample size, especially among pancreatic cancer patients, limited the ability to test for the association of PRO scales with clinical outcomes.

5 Conclusions

Results of the psychometric examination of the Katz ADL, Rosow–Breslau IADL, and Nagi scales collected in the MCBS using Medicare claims data demonstrate acceptable internal consistency and test–retest reliability among patients with pancreatic cancer, lung cancer, and MPN. The psychometric findings suggest that inclusion

of the Katz, Rosow–Breslau, and Nagi scales in confirmatory clinical trials of pancreatic, lung, and MPN cancers is appropriate, and could demonstrate changes in functional outcomes associated with efficacious treatment. More generally, these results suggest that retrospective survey data may be useful for the initial assessment of the psychometric properties of existing PRO scales in other populations of interest. In some cases, analyses of retrospective survey data may also be useful for preliminary exploration of disease hypotheses in those populations. Further research in this area could greatly facilitate the ability to understand small populations of interest before costly and lengthy de novo PRO instrument development.

Acknowledgments Funding for this study was provided by Eli Lilly and Company, Indianapolis, IN, USA, to Analysis Group, Inc., Boston, MA, USA. Dr. Mytelka, Mr. Price, and Dr. Swindle, employees of Eli Lilly and Company, provided input on study design and interpretation of data, reviewed and approved the manuscript. Ms. Ivanova, Dr. Duh, Dr. Birnbaum, Ms. Cummings, and Ms. San Roman are employees of Analysis Group, Inc., which received research funds from Eli Lilly and Company for the study. Ms. Ivanova, Dr. Duh, Dr. Birnbaum, Ms. Cummings, and Ms. San Roman participated in study design and interpretation of data, revised and approved the manuscript. Ms. Cummings and Ms. San Roman were responsible for the data analysis. Ms. San Roman was responsible for creating the first draft of the manuscript. Ms. Ivanova is the guarantor for the overall content.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Cella DF, Bonomi AE, Lloyd SR, et al. Reliability and validity of the Functional Assessment of Cancer Therapy–Lung (FACT–L) quality of life instrument. *Lung Cancer*. 1995;12(3):199–220.
2. Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Qual Life Res*. 2002;11(3):207–21.
3. Wong GY, Schroeder DR, Cams PE, et al. Effect of neurolytic celiac plexus block on pain relief, quality of life, and survival in patients with unresectable pancreatic cancer: a randomized controlled trial. *JAMA*. 2004;291(9):1092–9.
4. EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16:199–208.
5. Yellen SB, Cella DF, Webster K, et al. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. *J Pain Symptom Manag*. 1997;13(2):63–74.
6. Fearon K, Strasser F, Anker SD, et al. Definition and classification of cancer cachexia: an international consensus. *Lancet Oncol*. 2011;12(5):489–95.
7. Centers for Medicare and Medicaid Services. Medicare Current Beneficiary Survey (MCBS) (online). <http://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/index.html?redirect=/mcbs/>. Accessed 9 June 2011.

8. Katz S, Ford AB, Moskowitz RW, et al. Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychosocial function. *JAMA*. 1963;21(185):914–9.
9. Rosow I, Breslau N. A Guttman health scale for the aged. *J Gerontol*. 1966;21(4):556–9.
10. Nagi SZ. An epidemiology of disability among adults in the United States. *Milbank Mem Fund Q Health Soc*. 1976 Fall;54(4):439–67.
11. Fitzgerald JF, Smith DM, Martin DK, et al. Replication of the multidimensionality of activities of daily living. *J Gerontol*. 1993;48(1):S28–31.
12. Doni L, Perin A, Manzione L, et al. The impact of anemia on quality of life and hospitalisation in elderly cancer patients undergoing chemotherapy. *Crit Rev Oncol Hematol*. 2011;77(1):70–7.
13. Garman KS, Pieper CF, Seo P, et al. Function in elderly cancer survivors depends on comorbidities. *J Gerontol A Biol Sci Med Sci*. 2003;58(12):M1119–24.
14. Federman AD, Penrod JD, Livote E, et al. Development of and recovery from difficulty with activities of daily living: an analysis of national data. *J Aging Health*. 2010;22(8):1081–98.
15. Blustein J, Weiss LJ. The use of mammography by women aged 75 and older: factors related to health, functioning, and age. *J Am Geriatr Soc*. 1998;46(8):941–6.
16. Hill J, Fillit H, Thomas SK, et al. Functional impairment, healthcare costs and the prevalence of institutionalization in patients with Alzheimer's disease and other dementias. *Pharmacoeconomics*. 2006;24(3):265–80.
17. Salas M, Bubolz T, Caro JJ. Impact of physical functioning of health status on hospitalizations, physician visits, and costs in diabetic patients. *Arch Med Res*. 2000;31(2):223–7.
18. Stafford RS, Cyr PL. The impact of cancer on the physical function of the elderly and their utilization of health care. *Cancer*. 1997;80(10):1973–80.
19. Centers for Medicare and Medicaid Services. Medicare Current Beneficiary Survey (MCBS) (online). http://www.cms.gov/LimitedDataSets/11_MCBS.asp#TopOfPage. Accessed 9 June 2011.
20. Centers for Medicare and Medicaid Services. 2003 Appendix A. Technical Documentation for the Medicare Current Beneficiary Survey (online). http://www.cms.gov/MCBS/Downloads/CNP_2003_appendixA.pdf. Accessed 9 June 2011.
21. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373–83.
22. Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *J Clin Epidemiol*. 1993;46(10):1075–9.
23. Bureau of Labor Statistics. Databases, tables and calculators by subject: Consumer Price Index—all urban consumers, CUUR0000SAM, U.S. city average, medical care (online). <http://data.bls.gov/cgi-bin/surveymost>. Accessed 13 Aug 2012.
24. Wolinsky FD, Stump TE, Callahan CM, et al. Consistency and change in functional status among older adults over time. *J Aging Health*. 1996;8(2):155–82.
25. Nunnally JC. *Psychometric theory*. 2nd ed. New York: McGraw-Hill; 1978.
26. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–8.
27. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45(1):255–68.