①

History for Manuscript Number: QURE-D-12-05342

Close

Correspondence History

Correspondence Date ∧∨	Letter ∧∨	Recipient ∧∨	Revision ∧∨
08 Aug 2013	Editor Decision - Accept	Ron D. Hays®, Ph.D.	2
02 Aug 2013	Author Revision	Ron D. Hays®, Ph.D.	2
02 Aug 2013	AU - PDF Built	Ron D. Hays®, Ph.D.	2
26 Jul 2013	Editor Decision - Minor Revision	Ron D. Hays®, Ph.D.	1
30 Apr 2013	Author Revision	Ron D. Hays®, Ph.D.	1
30 Apr 2013	AU - PDF Built	Ron D. Hays®, Ph.D.	1
30 Apr 2013	AU - PDF Built	Ron D. Hays®, Ph.D.	1
30 Apr 2013	AU - PDF Built	Ron D. Hays®, Ph.D.	1
30 Apr 2013	AU - PDF Built	Ron D. Hays®, Ph.D.	1
24 Dec 2012	Editor decision - Major Revision	Ron D. Hays®, Ph.D.	0
03 Oct 2012	Author Submission	Ron D. Hays®, Ph.D.	0
03 Oct 2012	AU - PDF Built	Ron D. Hays®, Ph.D.	0
03 Oct 2012	AU - PDF Built	Ron D. Hays®, Ph.D.	0
03 Oct 2012	AU - PDF Built	Ron D. Hays®, Ph.D.	0
03 Oct 2012	AU - PDF Built	Ron D. Hays®, Ph.D.	0
03 Oct 2012	AU - PDF Built	Ron D. Hays®, Ph.D.	0

Close

Handwritten notes:

Month

① Initially submitted 10-03-12 — 0

② Editor decision 12-24-12 — ~3

③ Revision #1 submitted 04-30-13 — ~7

④ Editor decision 07-26-13 — ~10

⑤ Revision #2 submitted 08-02-13 — 10

⑥ Editor decision 08-08-13 — 10

View Letter

Close

Date: 24 Dec 2012
To: "Ron D. Hays®" hays@rand.org
From: "Carolien van der Gaag" Carolien.vanderGaag@springer.com
Subject: Major Revisions requested QURE-D-12-05342

Dear Dr. Hays®,

I have received the external reviews of the manuscript ("Associations of Cancer and Other Chronic Health Conditions with SF-6D Preference-Based Scores in Medicare Beneficiaries") you submitted to Quality of Life Research. You will find the comments and suggestions of the reviewers below. Based on the advice received and my own review, I have decided that your manuscript can be reconsidered for publication if you are prepared to make major revisions to the manuscript. Please note that I will be asking reviewers to evaluate the revised paper before I make a final decision regarding publication.

Please pay particular attention to the comments received from Reviewers 1 and 3. Reviewer 1 raises numerous important issues regarding the implications of this work for cancer patients. The authors need to address these issues when revising their introduction and discussion. Reviewer 3 suggests the use of more robust statistical approaches when analyzing the data, and I support the use of these methods. I ask that in preparing your revised manuscript you consider all comments carefully. Please check online for eventual reviewer attachments.

In submitting your revised manuscript, please include a cover letter giving specific details as to how you addressed each comment along with the page numbers where changes appear.

To submit your revision, please access the following web site: http://qure.edmgr.com/

Your username is: ********
Your password is: ********

I look forward to receiving your revised manuscript within eight weeks. If you are unable to complete it by this deadline, please contact me to request an extension. If you submit it after the deadline without prior approval from me, I will consider it a new manuscript.

Kind regards,

James W. Shaw, Ph.D., Pharm.D., M.P.H.
Associate Editor

==
REVIEWER COMMENTS FOR THE AUTHOR:

Reviewer #1: This paper reports the results of an analysis of the association between SF-6D utility scores of a large sample of Medicare recipients with a range of chronic disease conditions documented on the basis of self-report, with the exception of cancer conditions that were derived from a cancer registry. Not surprisingly, the results indicate that those with chronic health conditions have lower SF-6D scores than those without chronic health conditions. Most chronic health conditions had a statistically significant, negative association with the SF-6D scores. The strongest association was with self-reported depressive symptoms. Strong associations were also noted for arthritis of the hip, sciatica, COPD and stroke. Most cancer types also were associated significantly with SF-6D scores.

Comments:

Introduction:

1.   The introduction is very brief, and does little to build a case for the need for this study. Essentially the authors indicate that there are no data available on the association between preference-based health outcomes such as the SF-6D and chronic health conditions among Medicare recipients. That there is a significant association between both health profile scores and preference-based health outcomes and chronic disease status in the general population is not a mystery. This has been documented elsewhere It is unclear why the authors feel that it is important to confirm this among Medicare recipients. Do they suspect that there is something about Medicare status that would change the nature or strength of these associations?

2.   The authors pose two hypotheses: (1) that individuals with cancer will have significantly worse SF-6D scores than those without a chronic health condition (other than cancer); and (2) that individuals with cancer will have SF-6D scores similar to those with (other) chronic health conditions. The first hypothesis has been examined and confirmed in many population-based studies, although results vary as a function of type of cancer, stage of disease, treatment status (on or off treatment) and time since diagnosis. Again, the question arises as to why the authors feel that this question needs to be evaluated once again. In contrast, the authors do not provide any background information or arguments to support the plausibility of their second hypothesis, that having cancer has the same effect on preference-based health outcomes as having any other (chronic) health condition. This is such a broad hypothesis as to render it meaningless. Do the authors expect, for example, that a patient with advanced lung cancer would have the same utility score as a patient with mild hypertension? I hope not.

Methods:

3.   Information regarding cancer status was derived from the SEER database. This is in contrast to the other health conditions, all of which were documented on the basis of self-report. It would have been useful if the investigators had also asked respondents to self-report cancer. In this way, one could examine (at least partially) the validity of self-reported health conditions.

4.   Whereas for most health conditions respondents were asked to report whether a doctor had ever told them that they had that condition, for depression, respondents were asked whether they were depressed in the last year. As the authors themselves recognize, these are two very different types of questions. It is questionable if depressive symptoms should be included at all in the analysis, not only because of the difference in the type of question posed (diagnosis versus self-reported symptoms), but also because of the time frame used. For all of the other health conditions, the data essentially represent lifetime prevalence, whereas for depressive symptoms they represent a period prevalence (one year).

5.   On page 4 it is stated that the sample used in the analysis included those who "had completed enough questions to be included in the analysis." This is pretty vague and should be made more concrete.

6.   Much of the information provided in the methods section --- prevalence of various types of cancer, prevalence of other chronic health conditions - should actually be reported in the results section of the manuscript. Additionally, it would be helpful if these data were to be reported in tabular form. It should also be stated explicitly how many respondents reported no chronic health conditions (including cancer). I could not find this number anywhere in the body of the text or in the tables.

7.   In the paragraph on measures (page 5), the authors state that: "We used the scoring algorithm that is based on a consistent version of Model 10 in the published article (created by Donna Rown in accordance with changes agreed to by John Brazier, January 17, 2007). Is this something that the reader is supposed to understand? I fear not.

8.   Please explain briefly the method of "recycled predictions." I suspect that most readers will not be familiar with it (in any case, this reader is not).

Results:

9.   I had some difficulty in understanding the results of the regression analysis as reported in the text and in Table 2. The estimated SF-6D score for those without any chronic health conditions was 0.801. How then can, for example, colorectal cancer and female breast cancer have a statistically significant p value when the SF-6D adjusted score estimates for these conditions are also 0.80?

10.   Is it really prudent to test 52 interactions (cancer types x chronic health conditions)? Wouldn't it better (and perhaps more informative) to simply test the interaction between cancer and other chronic health conditions (none, 1, 2 or more)? Also, I miss results regarding the cumulative effects of having more than one chronic health condition (irrespective of which combination or combinations).

Discussion:

11.   Importantly, the authors provide some guidelines for interpreting the seemingly small effects of most of the cancer types and most of the chronic health conditions on SF-6D scores. On page 10 they explain that, according to several previous studies, differences in SF-6D scores as small as 0.1 may be minimally important, with the mean MID being about 0.03. I'm still trying to understand, however, how this translates into a difference in a quality-adjusted life year. Take the estimated SF-6D mean score of 0.80 for those without cancer or any other chronic health condition as a starting point. If I understand correctly, this means that, for this subpopulation of older adults, their average quality adjusted life year is .80 or 292 days. Take then the subgroup of respondents with arthritis; the subgroup with the largest regression coefficient (-0.044). This translates into an average quality adjusted life year of 0.762, or 278 days. If this math is correct, are we to believe that an average difference of 14 days in a quality adjusted life year is clinically or otherwise substantively significant? If the math is incorrect and is based on a misunderstanding of how the results should be interpreted, then the authors need to provide the correct math and provide the reader more explicitly with a correct interpretation of the results. In its current form, the discussion does not provide a clear interpretation of the results from a substantive point of view.

12.   I miss some discussion of the relevance of these data to either clinical practice or health policy decision-making. How will these results help allocate health resources, plan and delivery better and more effective health care, or improve the health-related quality of life of Medicare recipients?

Tables:

13.   Please add a better explanation of the meaning/interpretation of the unstandardized beta coefficient to the footnote of the table.
14.   Figure 2 is really unnecessary.

Other (minor):

15.   The text needs to be checked for spelling errors.
16.   The word 'data' is plural, not singular. Thus, for example, "the data were," rather then "the data was" gathered.
17.   The abstract states that the MHOAS covered the period 1998-2002. In the methods section (page 3) this is reported as being the period 1999-2003.

Reviewer #2: This is a very interesting study trying to estimate the incremental effect of different types of cancers on HRQOL on single index scale using SF-6D in Medicare Beneficiaries survey data set. These results has interest for both decision makers and clinicians who are making decisions and treating cancer patients. The results may also provide encouraging message for many cancer survivors that the (short/long term?) HRQOL outcomes seem to be relatively good.
Some comments:
-     Methods: in the sample section (end of first paragraph) you indicate that the cancer patients were incident cases. What does this mean considering the timing of the survey? Do you know how long time from the diagnosis and initial treatment they got the survey? Timing of the survey has significant implications for the results and their interpretation. Like if some patients are still getting their radiation or other therapy, it has a significant impact on their HRQOL.
-     Page 4; Cancer stage, could you shortly define the stages, since they may not be that obvious for all readers (distant=metastatic?).
-     The results show statistically significant negative HRQOL impact in all but two cancers as expected mainly because the sample size is so huge. However, many of the impacts are very small (magnitude of less or equal than 0.01) and as discussed in the discussion part they would not be considered clinically significant using many criteria in the literature. I would like to see a bit more conservative approach to the claim that SF-6D is sensitive to detect decrements in HRQOL in most cancers/stages in elderly general population. As shown by the authors, more severe cases had lower health status indicating that in those patient groups the measure is very sensitive and potentially also measure changes within the patient group. It would be good to discuss a bit about the difference between population health and clinical trial measurements and how it impacts on the level of the HRQOL differences and interpretations of them. Like if the measurement is done one or few years after initial treatment, those who did not survive are not included (value or 0 in the QALY scale) and those who have recovered well are already in relatively good condition.
-     Related to the above, did your comparison to the sensitivity of the EQ-5D related to the same type of surveys (population health). It would be good to concentrate mainly to population health application of the measure in this article. The PROMS application of the results is very interesting , especially if SF-6D data before cancer treatment and shortly after it would be available to give more perspective for the interesting study results.

Reviewer #4: General Comments

This compact and well-written paper makes a nice addition to the literature on the impact of chronic conditions (particularly cancer) on HRQoL, as well as to the growing body of work on the ability of the SF-6D instrument to capture such impact. Following are some more detailed comments and suggestions for revision.

Specific Comments

1. On page 4, it states that: "Individuals with more than one cancer primary diagnosis were excluded." How many such cases were there?

2. The regression analysis does not address the potential issue of clustering/nesting of patients within HMO and SEER site. I would imagine that patients under the same HMO and those who report to the same SEER site would tend to be more similar to each other than they would be to patients from other HMOs and SEER sites, for a variety of reasons (e.g., similar socio-contextual factors). Clustering induces correlations between cases on the study variables, making standard errors too small and significance tests too optimistic. Note that random sampling within clusters (e.g., within HMOs as on the MHOS) does not resolve this issue. One way to quantify the extent of the problem is to use intra-class correlations (ICCs), that is, to examine the magnitude of between-subjects correlations among cases for the variables of interest (using HMO and SEER site as the clustering variables):

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biological Reviews, 85(4), 935-956.

Thompson, D.M., Fernald, D.H., & Mold, J.W. (2012). Intraclass Correlation Coefficients Typical of Cluster-Randomized Studies: Estimates From the Robert Wood Johnson Prescription for Health Projects. Annals of Family Medicine, 10(3), 235-240.
http://www.annfammed.org/content/10/3/235.full#ref-11

Killip, S., Mahfoud, Z., & Pearce, K. (2004). What Is an Intracluster Correlation Coefficient? Crucial Concepts for Primary Care Researchers. Annals of Family Medicine, 2(3), 204-208.
http://www.annfammed.org/cgi/reprint/2/3/204

To address clustering, I would suggest using a design-based approach involving post-estimation adjustments to the conventionally computed standard errors, essentially treating the clustering as a nuisance parameter (in other words, you keep the same original hypothesized model, but just correct the significance tests). These techniques could include bootstrapping, a "sandwich" estimator of the variance (*highly recommended*), or more simple corrections using a design effect factor. Note in particular, however, that there are at least two levels of clustering in the MHOS-SEER data set: HMO and SEER. If these can be considered as "nested" levels of clustering (e.g., if cases from multiple SEER sites are nested within the same HMO), then one just creates clusters (using a new variable) at the highest level -- basically just look at whichever level would encompass the most cases within each cluster. This will ensure unbiased standard errors. However, it is possible that clustering is non-nested and then a multiway error components model is required -- although I would think that in the current case, just choosing the highest level will be appropriate. See the following papers:

Cameron, A.C., Gelbach, J.B., & Miller, D.L. (2011). Robust Inference With Multiway Clustering. Journal of Business & Economic Statistics, 29(2), 238-249.
http://www.tandfonline.com/doi/pdf/10.1198/jbes.2010.07136

Sainani, K. (2010). The Importance of Accounting for Correlated Observations. Physical Medicine and Rehabilitation, 2, 858-861.
http://www.stanford.edu/~kcobb/hrp259/correlateddata.pdf

McCoach, D.B., & Adelson, J.L. (2010). Dealing With Dependence (Part I): Understanding the Effects of Clustered Data. Gifted Child Quarterly, 54(2), 152-155
http://www.iapsych.com/articles/mccoach2010a.pdf

Galbraith, S., Daniel, J.A., & Vissel, B. (2010). A Study of Clustered Data and Approaches to Its Analysis. The Journal of Neuroscience, 30(32), 10601-10608.

http://www.jneurosci.org/content/30/32/10601.full.pdf

Hedges, L.V. (2007). Correcting a Significance Test for Clustering. Journal of Educational and Behavioral Statistics, 32(2), 151-179.

Hedges, L.V. (2009). Adjusting a Significance Test for Clustering in Designs With Two Levels of Nesting. Journal of Educational and Behavioral Statistics, 34(4), 464-490.

Binder, D.A., Roberts, G.R. (2003). Design-based and model-based methods for estimating model parameters. In R.L. Chambers & C. Skinner (eds.), Analysis of Survey Data (pp. 29-48). Chichester: Wiley.

Zhu, M., Chu, H., & Greenland, S. (2011). Biased standard errors from complex survey analysis: an example from applying ordinary least squares to the national hospital ambulatory medical care survey. Annals of Epidemiology, 21(11), 830-834.

3. I agree with the authors' use of the "recycled predictions" method, although a brief description of the logic and purpose of the technique, along with some more recent references, should be added:

Li, Z., & Mahendra, G. (2010). Using "Recycled Predictions" for Computing Marginal Effects. SAS Global Forum 2010, Statistics and Data Analysis, Paper 272-2010.

Basu, A., & Rathouz, P.J. (2005). Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. Biostatistics, 6(1), 93-109.
doi: 10.1093/biostatistics/kxh020
http://biostatistics.oxfordjournals.org/content/6/1/93.full.pdf

Kleinman, L.C., & Norton, E.C. (2009). What's the Risk? A Simple Approach for Estimating Adjusted Risk Measures from Nonlinear Models Including Logistic Regression. Health Services Research, 44(1), 288-302.

4. I understand why the SF-6D was used here, given that it is the only preference-based measure derivable from the MHOS-SEER. However, I would suggest noting in the limitations section that the results are potentially specific to the SF-6D system. A number of studies have demonstrated that preference scores from different instruments produce widely varying results when administered to the same patient groups, which has obvious implications for economic analysis and any recommendations based on them. Thus, it would be appropriate to note that the current work should be replicated in other large-scale surveys of cancer patients using different generic, preference-based HRQoL measures, such as the EQ-5D, HUI3 and possibly others. See the following:

Whitehurst, D.G.T., & Bryan, S. (2011). Another Study Showing that Two Preference-Based Measures of Health-Related Quality of Life (EQ-5D and SF-6D) Are Not Interchangeable. But Why Should We Expect Them to be? Value in Health, 14(4), 531-538.

Whitehurst, D.G.T., Bryan, S., & Lewis, M. (2011). Systematic Review and Empirical Comparison of Contemporaneous EQ-5D and SF-6D Group Mean Scores. Medical Decision Making, 31(6), E34-E44.
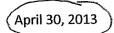
Grieve, R., Grishchenko, M., & Cairns, J. (2008). SF-6D versus EQ-5D: reasons for differences in utility scores and impact on reported cost-utility. The European Journal of Health Economics, 10(1), 15-23.
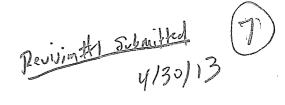
Sach, T.H., Barton, G.R., Jenkinson, C., Doherty, M., Avery, A.J., & Muir, K.R. (2009). Comparing cost-utility estimates: does the choice of EQ-5D or SF-6D matter? Medical Care, 47(8), 889-894.

Brazier, J., Roberts, J., Tsuchiya, A., & Busschbach, J. (2004) A comparison of the EQ-5D and SF-6D across seven patient groups. Health Economics, 13, 873-884.
http://eprints.whiterose.ac.uk/279/1/brazierje7.pdf

Hatoum, H.T., Brazier, J.E., & Akhras, K.S. (2004). Comparison of the HUI3 with the SF-36 preference based SF-6D in a clinical trial setting.Value in Health, 7(5), 602-609.
http://onlinelibrary.wiley.com/doi/10.1111/j.1524-4733.2004.75011.x/pdf

O'Brien, B.J., Spath, M., Blackhouse, G., Severens, J.L., Dorian, P., & Brazier, J. (2003). A view from the bridge: agreement between the SF-6D utility algorithm and the Health Utilities Index. Health Economics, 12(11), 975-981

April 30, 2013

Dear Dr. Shaw and reviewers:

Enclosed is our revised manuscript ("Associations of Cancer and Other Chronic Health Conditions with SF-6D Preference-Based Scores in Medicare Beneficiaries"). We found the comments of the reviewers to be very insightful and helpful and we have responded to each one in the attached response to reviewer comments document. In some cases we felt that there was not a consensus about making a specific change and have tried to provide rationale for why we did not implement the suggestion in the manuscript. In these cases we noted our willingness to make the change if the editor decides that we should do so.

We honestly found the reviews to be thorough, constructive and extremely helpful. The manuscript has been improved substantially as a result.

If you have any questions, do not hesitate to contact me by phone (310-794-2294) or email (drhays@ucla.edu).

Thank you,

Ron D. Hays, Ph.D.

Reviewer #1:

1. The introduction is very brief, and does little to build a case for the need for this study. Essentially the authors indicate that there are no data available on the association between preference-based health outcomes such as the SF-6D and chronic health conditions among Medicare recipients. That there is a significant association between both health profile scores and preference-based health outcomes and chronic disease status in the general population is not a mystery. This has been documented elsewhere It is unclear why the authors feel that it is important to confirm this among Medicare recipients. Do they suspect that there is something about Medicare status that would change the nature or strength of these associations?

Response: **We revised the introduction to show the importance of this study. "While it is apparent from the work to date that there are significant associations of cancer with HRQOL, the relative impact of cancer and other conditions on HRQOL overall is unknown. Although individual unique associations of conditions on HRQOL may seem small, a difference of about 0.03 on a 0-1 preference-based score may be important; interventions that produce that level of difference are non-trivial. In addition, the cumulative effect of multiple conditions could be substantial. Because the likelihood of chronic conditions increase with age, it is especially important to examine the impacts of these conditions on HRQOL among older individuals" (pp. 1-2).**

2. The authors pose two hypotheses: (1) that individuals with cancer will have significantly worse SF-6D scores than those without a chronic health condition (other than cancer); and (2) that individuals with cancer will have SF-6D scores similar to those with (other) chronic health conditions. The first hypothesis has been examined and confirmed in many population-based studies, although results vary as a function of type of cancer, stage of disease, treatment status (on or off treatment) and time since diagnosis. Again, the question arises as to why the authors feel that this question needs to be evaluated once again. In contrast, the authors do not provide any background information or arguments to support the plausibility of their second hypothesis, that having cancer has the same effect on preference-based health outcomes as having any other (chronic) health condition. This is such a broad hypothesis as to render it meaningless. Do the authors expect, for example, that a patient with advanced lung cancer would have the same utility score as a patient with mild hypertension? I hope not.

Response: **We modified the wording of the hypothesis and include a hypothesis about stage of disease, as noted by the reviewer. "We hypothesize that individuals with cancer will have significantly worse SF-6D scores than those without a chronic health condition and of comparable or greater magnitude than those with chronic medical conditions other than cancer. But the differences may vary by cancer type. For example, a previous analysis found that lung cancer was more strongly related to decrements in physical and mental health than other cancers [3]. We also hypothesize that more advanced stage of cancer will be associated with worse HRQOL" (p. 2).**

1

3. Information regarding cancer status was derived from the SEER database. This is in contrast to the other health conditions, all of which were documented on the basis of self-report. It would have been useful if the investigators had also asked respondents to self-report cancer. In this way, one could examine (at least partially) the validity of self-reported health conditions.

Response: We revised what we previously noted in the limitations section of the paper as follows: "Aside from cancer, we had to rely on self-reports of chronic conditions. However, data comparing comorbidities self-reported on the same survey we used (MHOS) versus abstraction of medical records suggests reasonably good correspondence [20]" (p. 11).

4. Whereas for most health conditions respondents were asked to report whether a doctor had ever told them that they had that condition, for depression, respondents were asked whether they were depressed in the last year. As the authors themselves recognize, these are two very different types of questions. It is questionable if depressive symptoms should be included at all in the analysis, not only because of the difference in the type of question posed (diagnosis versus self-reported symptoms), but also because of the time frame used. For all of the other health conditions, the data essentially represent lifetime prevalence, whereas for depressive symptoms they represent a period prevalence (one year).

Response: While we agree with the reviewer's comments and included a paragraph noting this issue in the discussion section of the original submission, we also found that the unique associations for the other 22 conditions with the SF-6D score were robust. We believe it is better to have looked at the association of depressive symptoms and compare it with a model that excluded it than to have not looked at all. If the editors feel that we should move this in accordance with the reviewer's suggestion we are happy to do that.

5. On page 4 it is stated that the sample used in the analysis included those who "had completed enough questions to be included in the analysis." This is pretty vague and should be made more concrete.

Response: We agree and revised the phrase to say "responded to the survey questions included in the analysis" (p. 3).

6. Much of the information provided in the methods section --- prevalence of various types of cancer, prevalence of other chronic health conditions - should actually be reported in the results section of the manuscript. Additionally, it would be helpful if these data were to be reported in tabular form. It should also be stated explicitly how many respondents reported no chronic health conditions (including cancer). I could not find this number anywhere in the body of the text or in the tables.

Response: Although the descriptive information about the chronic conditions could be moved to the Results section, we believe this is a matter of style. The advantage of leaving it in the Methods section is that it provides coherence to the reader when we describe the nature of the groups we compare. The intent of this study was not to estimate the prevalence of different chronic conditions. Rather, we are estimating the associations of the conditions (as given in the dataset) to SF-6D scores. In addition, we report the sample sizes for each condition in Table 2 already. The other reviewers did not raise an issue with where we presented this descriptive information. We would be happy to move this descriptive information to the results if the editor feels this is important or if it better reflects the journal's general formatting.

7. In the paragraph on measures (page 5), the authors state that: "We used the scoring algorithm that is based on a consistent version of Model 10 in the published article (created by Donna Rown in accordance with changes agreed to by John Brazier, January 17, 2007). Is this something that the reader is supposed to understand? I fear not.

Response: We revised the text to: "We used the revised SF-6D scoring algorithm described by Brazier, Rowen and Hanmer [10]" (pp. 4-5).

8. Please explain briefly the method of "recycled predictions." I suspect that most readers will not be familiar with it (in any case, this reader is not).

Response: We revised the sentence to say, "Recycled predictions [12], or averaging of individual marginal effects, are used to understand the incremental effect of an independent variable on a dependent variable. We used recycled predictions to obtain adjusted SF-6D means for each cancer and chronic condition group" (p. 6).

9. I had some difficulty in understanding the results of the regression analysis as reported in the text and in Table 2. The estimated SF-6D score for those without any chronic health conditions was 0.801. How then can, for example, colorectal cancer and female breast cancer have a statistically significant p value when the SF-6D adjusted score estimates for these conditions are also 0.80?

Response: We understand this confusion and have replaced the information about the intercept from the regression model with the recycled mean for those without any of the chronic conditions (p. 6 and in footnote to Table 2 on p. 17).

10. Is it really prudent to test 52 interactions (cancer types x chronic health conditions)? Wouldn't it better (and perhaps more informative) to simply test the interaction between cancer and other chronic health conditions (none, 1, 2 or more)? Also, I miss results regarding the cumulative effects of having more than one chronic health condition (irrespective of which combination or combinations).

Response: We believe it is important to assess whether or not the specific health conditions included in the model interact with the four prevalent cancers. The fact that we found little evidence of interactions provides further comfort in focusing on the main effect model. We can drop this text if the editor prefers for us to exclude it. We also revised a sentence that notes: "the adjusted score for those reporting two or more conditions was on average 0.11 lower (results not reported earlier). " (p. 10).

11. Importantly, the authors provide some guidelines for interpreting the seemingly small effects of most of the cancer types and most of the chronic health conditions on SF-6D scores. On page 10 they explain that, according to several previous studies, differences in SF-6D scores as small as 0.01 may be minimally important, with the mean MID being about 0.03. I'm still trying to understand, however, how this translates into a difference in a quality-adjusted life year. Take the estimated SF-6D mean score of 0.80 for those without cancer or any other chronic health condition as a starting point. If I understand correctly, this means that, for this subpopulation of older adults, their average quality adjusted life year is .80 or 292 days. Take then the subgroup of respondents with arthritis; the subgroup with the largest regression coefficient (-0.044). This translates into an average quality adjusted life year of 0.762, or 278 days. If this math is correct, are we to believe that an average difference of 14 days in a quality adjusted life year is clinically or otherwise substantively significant? If the math is incorrect and is based on a misunderstanding of how the results should be interpreted, then the authors need to provide the correct math and provide the reader more explicitly with a correct interpretation of the results. In its current form, the discussion does not provide a clear interpretation of the results from a substantive point of view.

Response: The math is correct for one year of life. Of course, the expected life expectancy is longer than one year. We appreciate this suggested way of expressing the MID and have added the following sentence to the discussion section: the decrements in HRQOL cumulate over time so the full impact relative to those without chronic conditions is the observed decrement multiplied by the number of years with the condition" (p. 10). Finally, the decrements in HRQOL accumulate over time so the full impact relative to those without chronic conditions is the observed decrement multiplied by the number of years with the condition.

4

12. I miss some discussion of the relevance of these data to either clinical practice or health policy decision-making. How will these results help allocate health resources, plan and delivery better and more effective health care, or improve the health-related quality of life of Medicare recipients?

Response: We revised the end of the paper as follows: "As SEER-MHOS adds new cohorts and larger sample sizes by stage of disease become available, further investigation of the impact of cancer and stage of disease will be possible. Tracking the association of chronic conditions and stage of disease on HRQOL periodically can give providers and federal agencies such as the Center for Medicare and Medicaid Services important information about the extent to which the health needs of older Americans are being addressed" (pp. 11-12).

13. Please add a better explanation of the meaning/interpretation of the unstandardized beta coefficient to the footnote of the table.

Response: We added the following to the Table 2 (p. 17) footnote. "The unstandardized beta coefficients column indicate the direction and magnitude of difference in the adjusted SF-6D score for each chronic condition."

14. Figure 2 is really unnecessary.

Response: We think the reviewer may be referring to Figure 1. We are happy to omit it but are unsure if this opinion is shared by the other reviewers and the editor. We have retained the figure, but we made the text more specific on page 8 in case the consensus is to drop Figure 1.

15. The text needs to be checked for spelling errors.

Response: We checked and corrected all spelling errors we could find.

16. The word 'data' is plural, not singular. Thus, for example, "the data were," rather then "the data was" gathered.

Response: Very true and duly corrected: "In addition, the data were collected using both mail surveys and telephone interviews..." (p. 11).

17. The abstract states that the MHOS covered the period 1998-2002. In the methods section (page 3) this is reported as being the period 1999-2003.

Response: **The latter was corrected on page 3 to be consistent with the abstract.**

Reviewer #2:

- Methods: in the sample section (end of first paragraph) you indicate that the cancer patients were incident cases. What does this mean considering the timing of the survey? Do you know how long time from the diagnosis and initial treatment they got the survey? Timing of the survey has significant implications for the results and their interpretation. Like if some patients are still getting their radiation or other therapy, it has a significant impact on their HRQOL.

Response: **This paper includes all cancer patients regardless of time since diagnosis and treatment. We note this in the discussion section of the paper. "A previous study found similar SF-36 scores by time since diagnosis but there was some indication of a healthy survivor effect in colorectal and lung cancer patients [3]" (p. 10)**

- Page 4; Cancer stage, could you shortly define the stages, since they may not be that obvious for all readers (distant=metastatic?).

Response: **We added metastatic within parentheses following "distant" to the text (p. 6).**

- The results show statistically significant negative HRQOL impact in all but two cancers as expected mainly because the sample size is so huge. However, many of the impacts are very small (magnitude of less or equal than 0.01) and as discussed in the discussion part they would not be considered clinically significant using many criteria in the literature. I would like to see a bit more conservative approach to the claim that SF-6D is sensitive to detect decrements in HRQOL in most cancers/stages in elderly general population. As shown by the authors, more severe cases had lower health status indicating that in those patient groups the measure is very sensitive and potentially also measure changes within the patient group. It would be good to discuss a bit about the difference between population health and clinical trial measurements and how it impacts on the level of the HRQOL differences and interpretations of them. Like if the measurement is done one or few years after initial treatment, those who did not survive are not included (value or 0 in the QALY scale) and those who have recovered well are already in relatively good condition.

Response: **We added further argument about the importance of the observed differences. "While the individual impact of conditions including cancer was typically not large, the**

6

differences could matter in group comparisons. For example, a difference of 0.02 on the utility scale would be considered cost-effective if it cost about $1,000 to produce (i.e., $50,000/QALY). Moreover, the cumulative effect of multiple conditions is substantial as evidenced by the plethora of significant unique associations in the regression model. In fact, the adjusted score for those reporting any one condition was on average 0.03 lower than those reporting no conditions while the adjusted score for those reporting two or more conditions was on average 0.11 lower (results not reported earlier). Finally, the decrements in HRQOL cumulate over time so the full impact relative to those without chronic conditions is the observed decrement multiplied by the number of years with the condition" (pp. 9-10).

- Related to the above, did your comparison to the sensitivity of the EQ-5D related to the same type of surveys (population health). It would be good to concentrate mainly to population health application of the measure in this article. The PROMS application of the results is very interesting, especially if SF-6D data before cancer treatment and shortly after it would be available to give more perspective for the interesting study results.

Response: We clarified in the text that "While some preference-based measures (e.g., the EQ-5D) have been criticized for lack of sensitivity [21], the results of this study suggest that the SF-6D is sensitive to the impact of cancer on HRQOL" (p. 11).

Reviewer #4:
1. On page 4, it states that: "Individuals with more than one cancer primary diagnosis were excluded." How many such cases were there?

Response: Clauser et al. [8] reported that 13% of the incident cases had multiple primary cancer diagnoses. We added this information to the manuscript (pp. 3- 4).

2. The regression analysis does not address the potential issue of clustering/nesting of patients within HMO and SEER site. I would imagine that patients under the same HMO and those who report to the same SEER site would tend to be more similar to each other than they would be to patients from other HMOs and SEER sites, for a variety of reasons (e.g., similar socio-contextual factors). Clustering induces correlations between cases on the study variables, making standard errors too small and significance tests too optimistic. Note that random sampling within clusters (e.g., within HMOs as on the MHOS) does not resolve this issue. One way to quantify the extent of the problem is to use intra-class correlations (ICCs), that is, to examine the magnitude of between-subjects correlations among cases for the variables of interest (using HMO and SEER

site as the clustering variables):

To address clustering, I would suggest using a design-based approach involving post-estimation adjustments to the conventionally computed standard errors, essentially treating the clustering as a nuisance parameter (in other words, you keep the same original hypothesized model, but just correct the significance tests). These techniques could include bootstrapping, a "sandwich" estimator of the variance (*highly recommended*), or more simple corrections using a design effect factor. Note in particular, however, that there are at least two levels of clustering in the MHOS-SEER data set: HMO and SEER. If these can be considered as "nested" levels of clustering (e.g., if cases from multiple SEER sites are nested within the same HMO), then one just creates clusters (using a new variable) at the highest level -- basically just look at whichever level would encompass the most cases within each cluster. This will ensure unbiased standard errors. However, it is possible that clustering is non-nested and then a multiway error components model is required -- although I would think that in the current case, just choosing the highest level will be appropriate.

Response: **We reran the regression models and modified the text to say that "Regression parameter standard errors were adjusted for clustering at the health plan level using the sandwich estimator of variance [13]" (p. 6).**

3. I agree with the authors' use of the "recycled predictions" method, although a brief description of the logic and purpose of the technique, along with some more recent references, should be added:

Response: **We substituted the Basu and Rathouz citation for the Korn and Graubard reference originally include in the paper (p. 14). The problem with the Li and Mahendra reference is that it is a macro for binary dependent variables. We revised the text in the manuscript as follows. "Recycled predictions [12], or averaging of individual marginal effects, are used to understand the incremental effect of an independent variable on a dependent variable. We used recycled predictions to obtain adjusted SF-6D means for each cancer and chronic condition group" (p. 6).**

4. I understand why the SF-6D was used here, given that it is the only preference-based measure

8

derivable from the MHOS-SEER. However, I would suggest noting in the limitations section that the results are potentially specific to the SF-6D system. A number of studies have demonstrated that preference scores from different instruments produce widely varying results when administered to the same patient groups, which has obvious implications for economic analysis and any recommendations based on them. Thus, it would be appropriate to note that the current work should be replicated in other large-scale surveys of cancer patients using different generic, preference-based HRQoL measures, such as the EQ-5D, HUI3 and possibly others.

Response: We agree and added the following text to the limitations paragraph in the manuscript. "However, previous research indicates that different preference measures may not yield the same results [22]. The current work needs to be replicated in other large-scale surveys of cancer patients using different preference-based measures" (p. 11)

Close |

Date:	26 Jul 2013
To:	"Ron D. Hays®" hays@rand.org
cc:	carolyn.schwartz@deltaquest.org
From:	"Carolien van der Gaag" Carolien.vanderGaag@springer.com
Subject:	Your Submission QURE-D-12-05342R1

Dear Dr. Hays®,

I have received the external reviews of the manuscript ("Associations of Cancer and Other Chronic Health Conditions with SF-6D Preference-Based Scores in Medicare Beneficiaries") you submitted to Quality of Life Research. You will find the comments and suggestions of the reviewers below. Based on the advice received and my own review, I have decided that your manuscript can be reconsidered for publication if you are prepared to incorporate minor revisions.

I ask that in preparing your revised manuscript you consider all comments carefully. Please check online for eventual reviewer attachments.

In submitting your revised manuscript, please include a cover letter giving specific details as to how you addressed each comment along with the page numbers where changes appear.

To submit your revision, please access the following web site: http://qure.edmgr.com/

Your username is: ********
Your password is: ********

I look forward to receiving your revised manuscript within 8 weeks. If you are unable to complete it by this deadline, please contact me to request an extension. If you submit it after the deadline without prior approval from me, I will consider it a new manuscript.

Kind regards,

Carolyn Emily Schwartz, Sc.D.
Editor in Chief

==
REVIEWER COMMENTS FOR THE AUTHOR:

Reviewer #1: 1. The authors have rephrased their second hypothesis in response to my concerns about the plausibility of their original formulation. Unfortunately, the revised text is not entirely clear: "We hypothesize that individuals with cancer will have significantly worse SF-6D scores than those without a chronic health condition and of comparable or greater magnitude those with chronic medical conditions other than cancer." (italics added). First, are the authors making a distinction between health conditions and medical conditions? Second, do they really mean that the SF-6D scores of cancer patients will be similar to or better than those with other medical conditions other than cancer? That is the way the sentence currently reads. They probably intend to say that: "...and SF-6D scores that are comparable or worse than those with chronic medical conditions other than cancer." It might make sense to split this into two sentences/hypotheses.
2. The authors have revised the text in the limitations section of the paper, indicating "reasonably good correspondence" between information derived from self-report versus medical record audits. What is reasonably good? Please be more specific, and leave it to the reader to decide if your qualification as being "reasonably good" is reasonable.
3. I still think that Figure 1 is unnecessary, particularly given the results presented in the figure could and have now been summarized in a single sentence in the results section.
4. There is one additional minor point not raised in my earlier review. The authors use the term

"distant stage of disease." This term is not one that is typically used to classify stage of cancer. "Metastatic disease" or "advanced disease stage" or "metastatic disease stage" are more appropriate terms.

Reviewer #4: General Comments:

The authors have been extremely thorough in addressing my concerns, and I appreciate their efforts. I only have one minor suggestion. Please consider replacing:

13. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 1980; 48: 817-830.

with one or more of:

Rogers, W.H. (1993). Regression standard errors in clustered samples. Stata Technical Bulletin Reprints, 3, 88-94.

Williams, R.L. (2000). A note on robust variance estimation for cluster-correlated data. Biometrics, 56, 645-646.

Hardin, J.W. (2003). The sandwich estimate of variance. In T.B. Fomby & R.C. Hill (Eds.), Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later (Advances in Econometrics, vol. 17, pp. 45-73). Emerald Group Publishing LTD.

Ying, G.-S., & Liu, C.-C. (2006). Statistical Analysis of Clustered Data using SAS® System. NESUG 2006, Data Manipulation and Analysis. Cary, NC: SAS Institute.

Close

August 2, 2013

Carolyn Schwartz, Sc.D.

Editor in Chief, *Quality of Life Research*

Dear Dr. Schwartz:

Thank you for the July 26 minor revisions decision letter. We have revised the manuscript based on the reviewers' comments. Below we indicate how we responded to each suggestion. In addition, we decided to add an additional column to Table 2 with alternative recycled prediction values. The reason we did this is because the ones we provided before were derived by fixing the values for all other chronic conditions to zero, but some people prefer fixing these other values to the mean for the sample. The advantage of fixing at the mean is that the adjusted scores correspond to the sample. (Fixing the conditions to zero leads to higher adjusted scores.) The disadvantage of using the mean is that it can yield counterintuitive results. In our data, for example, the unstandardized beta for sciatica (-0.037) is more negative than for GI (-0.031) yet the recycled prediction based on the means for the other conditions is slightly larger for sciatica (0.701) than GI (0.700). This occurs because sciatica is more prevalent than GI--the mean for sciatica (0.21) is larger than that for GI (0.05). Hence, when we predict the GI score the larger sciatica mean drives the estimated GI score down more than the GI mean drives down the estimated sciatica score.

Sincerely,

Ron D. Hays

Reviewer #1

Comment: The authors have rephrased their second hypothesis in response to my concerns about the plausibility of their original formulation. Unfortunately, the revised text is not entirely clear: "We hypothesize that individuals with cancer will have significantly worse SF-6D scores than those without a chronic health condition and of comparable or greater magnitude those with chronic medical conditions other than cancer." (italics added). First, are the authors making a distinction between health conditions and medical conditions? Second, do they really mean that the SF-6D scores of cancer patients will be similar to or better than those with other medical conditions other than cancer? That is the way the sentence currently reads. They probably intend to say that: "...and SF-6D scores that are comparable or worse than those with chronic

1

medical conditions other than cancer." It might make sense to split this into two sentences/hypotheses.

Response: We modified the text to read as follows: "We hypothesize that individuals with cancer will have significantly worse SF-6D scores than those without a chronic medical condition. We also hypothesize that those with cancer will have SF-6D scores that are comparable or worse than those with chronic medical conditions other than cancer" (p. 2).

Comment: The authors have revised the text in the limitations section of the paper, indicating "reasonably good correspondence" between information derived from self-report versus medical record audits. What is reasonably good? Please be more specific, and leave it to the reader to decide if your qualification as being "reasonably good" is reasonable.

Response: We revised the sentence to say the following. "However, data comparing comorbidities self-reported on the same survey we used (MHOS) versus abstraction of medical records suggests reasonably good correspondence, with median specificity (% of time condition is not self-reported when it is not in the medical record) and sensitivity (% of time condition is self-reported when it is in the medical record) of 69% and 91%, respectively" (p. 11).

Comment: I still think that Figure 1 is unnecessary, particularly given the results presented in the figure could and have now been summarized in a single sentence in the results section.

Response: We deleted the figure.

Comment: There is one additional minor point not raised in my earlier review. The authors use the term "distant stage of disease." This term is not one that is typically used to classify stage of cancer. "Metastatic disease" or "advanced disease stage" or "metastatic disease stage" are more appropriate terms.

Response: As noted previously, when we describe "distant" stage of disease in the Methods section we also include "metastatic" within parentheses (p. 6). This terminology is consistent with classification used by the National Cancer Institute. Please see Item #4 in: http://www.cancer.gov/cancertopics/factsheet/detection/staging. Distant stage of disease is one of the "summary stage" categories used by the SEER program to classify all types of cancer. Our data come directly from SEER registries; thus, our terminology is the most accurate fit to the data, and is well understood among researchers studying cancer. Although the reviewer suggests language that would also be understood, we want to be consistent with classification by the SEER program, and therefore prefer to retain this terminology.

Reviewer #4

Comment: The authors have been extremely thorough in addressing my concerns, and I appreciate their efforts. I only have one minor suggestion. Please consider replacing:

13. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 1980; 48: 817-830. with one or more of:

Rogers, W.H. (1993). Regression standard errors in clustered samples. Stata Technical Bulletin Reprints, 3, 88-94.

Williams, R.L. (2000). A note on robust variance estimation for cluster-correlated data. Biometrics, 56, 645-646.

Hardin, J.W. (2003). The sandwich estimate of variance. In T.B. Fomby & R.C. Hill (Eds.), Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later (Advances in Econometrics, vol. 17, pp. 45-73). Emerald Group Publishing LTD.

Ying, G.-S., & Liu, C.-C. (2006). Statistical Analysis of Clustered Data using SAS® System. NESUG 2006, Data Manipulation and Analysis. Cary, NC: SAS Institute.

Response: We replaced the White citation with the Williams (2000) reference (p. 14).

Associations of cancer and other chronic medical conditions with SF-6D preference-based scores in Medicare beneficiaries

Ron D. Hays · Bryce B. Reeve · Ashley Wilder Smith ·
Steven B. Clauser

Abstract

Purpose Documenting the impact of different types of cancer on daily functioning and well-being is important for understanding burden relative to other chronic medical conditions. This study examined the impact of 10 different cancers and 13 other chronic medical conditions on health-related quality of life.

Methods Health-related quality of life data were gathered on the Medicare Health Outcomes Survey (MHOS) between 1998 and 2002. Cancer information was ascertained using the National Cancer Institute's surveillance, epidemiology, and end results program and linked to MHOS data.

Results The average SF-6D score was 0.73 (SD = 0.14). Depressive symptoms had the largest unique association with the SF-6D, followed by arthritis of the hip, chronic obstructive pulmonary disease/asthma, stroke, and sciatica. In addition, the majority of cancer types were significantly associated with the SF-6D score, with significant negative weights ranging from -0.01 to -0.02 on the 0–1 health utility scale. Distant stage of cancer was associated with large decrements in the SF-6D ranging from -0.04 (prostate) to -0.08 (female breast).

Conclusion A large number of chronic conditions, including cancer, are associated uniquely with decrements in health utility. The cumulative effects of comorbid conditions have substantial impact on daily functioning and well-being of Medicare beneficiaries.

Keywords Cancer and comorbidity · Health-related quality of life · Preference-based measures · Utilities

Introduction

The majority of US adults (133 million) have at least one chronic medical condition [1], and 12 million Americans are living with cancer [2]. Rothrock et al. [1] documented that most medical conditions have a negative impact on daily functioning and well-being, or health-related quality of life (HRQOL) measured by the Patient-Reported Outcomes Measurement Information System® (PROMIS®). Having a single condition had a negative impact on the PROMIS® HRQOL domain scores of about 0.1–0.4 standard deviations (SDs), depending on the condition and the specific HRQOL domain. Having multiple conditions compared to a single condition impacted negatively in the range of 0.2–0.7 SDs.

Smith et al. [3] found that Medicare managed care beneficiaries with cancer had significantly worse physical health (as measured by the SF-36 v.1 physical component summary score) than those without cancer. Beneficiaries with non-small cell lung, non-Hodgkin lymphoma, female breast, colorectal, or bladder cancer reported worse mental health (on SF-36 v. 1 mental component summary score) than did those without cancer. While the SF-36's two
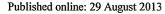
R. D. Hays (✉)
Division of General Internal Medicine & Health Services
Research, David Geffen School of Medicine at UCLA,
Los Angeles, CA, USA
e-mail: drhays@ucla.edu

B. B. Reeve
Lineberger Comprehensive Cancer Center, Department of
Health Policy and Management, Gillings School of Global
Public Health, University of North Carolina, Chapel Hill
NC, USA

A. W. Smith · S. B. Clauser
National Cancer Institute, Bethesda, MD, USA

summary scores provide useful information about physical and mental health, a single preference-based score is very helpful when "bottom-line" comparisons of different therapies are needed such as in comparative effectiveness research. Preference-based measures are designed to integrate across domains of health to provide a summary measure anchored relative to "dead" (score of 0) and "perfect health" (score of 1). A preference-based score is essential when a decision about overall health impact is required.

While it is apparent from the work to date that there are significant associations of cancer with HRQOL, the relative impact of cancer and other health conditions on HRQOL overall is unknown. Although individual unique associations of conditions on HRQOL may seem small, a difference of about 0.03 on a 0–1 preference-based measure may be important; interventions that produce that level of difference are non-trivial. In addition, the cumulative effect of multiple conditions could be substantial. Because the likelihood of chronic conditions increases with age, it is especially important to examine the impacts of these conditions on HRQOL among older individuals.

This study uses the SF-6D [4, 5] to estimate the unique impact of different health conditions on HRQOL for Medicare managed care beneficiaries aged 65 years or older. We also investigate the impact of several types of cancers including both highly prevalent (prostate, female breast, colorectal, non-small cell lung) and less common (endometrial, bladder, melanoma, non-Hodgkin lymphoma, and kidney) cancers. We hypothesize that individuals with cancer will have significantly worse SF-6D scores than those without a chronic medical condition. We also hypothesize that those with cancer will have SF-6D scores that are comparable or worse than those with chronic medical conditions other than cancer. But the differences may vary by cancer type. For example, a previous analysis found that lung cancer was more strongly related to decrements in physical and mental health than other cancers [3]. We also hypothesize that more advanced stage of cancer will be associated with worse HRQOL.

Methods

Sample

Our sample is derived from a dataset that links surveillance, epidemiology, and end results (SEER) cancer registry information with the Medicare Health Outcomes Survey (MHOS). A detailed description of the SEER-MHOS data is provided by Ambs et al. [6]. Briefly, the MHOS evaluated outcomes of care provided by health maintenance organizations to Medicare beneficiaries.

During the 1998–2003 study period, the MHOS was a yearly survey administered to a random sample of 1,000 Medicare beneficiaries from each managed care plan under contract with the Centers for Medicare and Medicaid Services (CMS). The SEER program includes population-based cancer registry sites throughout the USA that collect standardized clinical and demographic information for persons with incident cancer [7].

The linked SEER-MHOS dataset includes four MHOS cohorts (baseline and follow-up year): 1998 and 2000; 1999 and 2001; 2000 and 2002; and 2001 and 2003. Response rates to the MHOS baseline surveys ranged from 63 to 72 %. The majority of the MHOSs were completed by mail (88 %), and the rest were administered by phone (12 %). In addition, 12 % of the surveys were completed by a proxy. Across the four cohorts, we identified a total sample of 126,366 persons, both with and without cancer, age 65 or older who had completed at least one survey and responded to the survey questions included in the analysis.

Participants with cancer ($n = 22,740$; 18 % of the sample) were identified through SEER, and the first survey completed after their cancer diagnosis was used. We restricted this cancer subgroup to those with a first diagnosis of one of nine prevalent cancers: (1) prostate ($n = 5,593$; 4 % of the sample), (2) female breast ($n = 4,311$; 3 % of the sample), (3) colorectal ($n = 3,012$; 2 % of the sample), (4) non-small cell lung ($n = 1,792$; 1 % of the sample), (5) bladder ($n = 1,299$; 1 % of the sample), (6) melanoma ($n = 1,135$; 1 % of the sample), (7) endometrial ($n = 902$; 1 % of the sample), (8) non-Hodgkin lymphoma ($n = 668$; 1 % of the sample), and (9) kidney cancer ($n = 488$; 0.4 % of the sample). The remaining cancer diagnoses were classified as "other" ($n = 3,540$, 3 % of the sample). Cancer stage data came from SEER. Individuals with more than one cancer diagnosis, or who self-reported cancer, but were not identified in SEER, were excluded. Clauser et al. [8] reported that 13 % of the incident cases had multiple primary cancer diagnoses. For the 103,626 people without a cancer diagnosis, we included the first survey they completed.

Respondents were asked on the MHOS whether they had ever been told by a doctor that they had any of 12 chronic medical conditions: (1) hypertension or high blood pressure ($n = 66,968$; 53 % of the sample), (2) arthritis of the hip ($n = 44,524$; 35 % of the sample), (3) arthritis of the hand ($n = 40,402$; 32 % of the sample), (4) sciatica ($n = 26,878$; 21 % of the sample), (5) angina or coronary artery disease ($n = 18,017$; 14 % of the sample), (6) myocardial infarction or heart attack ($n = 11,982$; 9 % of the sample), (7) stroke ($n = 9,479$; 8 % of the sample), (8) congestive heart failure ($n = 7,893$; 6 % of the sample), (9) other heart disease ($n = 25,455$; 20 % of the sample), (10) diabetes ($n = 20,089$; 16 % of the sample), (11)

chronic obstructive pulmonary disease ($n = 15,445$; 12 % of the sample), and (12) inflammatory bowel disease ($n = 5,882$; 5 % of the sample). In addition, they were asked whether they were depressed in the last year ($n = 14,815$; 12 % of the sample). The percentage of the sample who reported no conditions was 13 % ($n = 15,833$), 21 % ($n = 26,126$) reported one condition, 21 % ($n = 26,653$) reported two conditions, and the remainder reported 3–14 conditions (including cancer).

Measures

The MHOS includes the SF-36 health survey, version 1 [9]. We used the SF-6D preference-based score as the dependent variable in this study. The SF-6D is computed from a subset (11 of the 36 questions) of the SF-36 questionnaire [5]. The SF-6D reduced the SF-36 to six domains (physical functioning, role limitations, social function, pain, emotional well-being, vitality), each comprised of four to six levels, and jointly defined 18,000 health states. Scoring was derived from standard gamble assessments by a population sample from the United Kingdom [5]. We used the revised SF-6D scoring algorithm described by Brazier, Rowen and Hanmer [10]. The algorithm produces scores ranging from 0.30 to 1.00 for those alive [5].

Participants' self-reported age, gender, race/ethnicity, marital status, education, and income came from the MHOS.

Analysis plan

The analyses were performed using SAS 9.3 and STATA 12 software. We provide descriptive statistics for the sample, followed by least square regression adjusted means on the SF-6D. The SF-6D mean was slightly lower than the median (0.73 vs. 0.75), indicating minor negative skewness (-0.37) and an approximately normal distribution of standard errors for regression coefficients [11]. We estimated the unique associations of each chronic condition with the SF-6D, controlling for the other conditions, education (8th grade or less; some high school; high school graduate; some college; 4 year college graduate; >4 year college degree), gender, marital status (married; widowed; separated; divorced; never married), age, race/ethnicity (Hispanic; non-Hispanic white; non-Hispanic black; Asian; American Indian; other race; race/ethnicity missing), income (<$10,000/year; $10,000–19,999/year; $20,000–29,999/year; $30,000–39,999/year; $40,000–49,999/year; $50,000–79,999/year; $80,000 and above; do not know or missing income), whether a proxy completed the survey, and mode of administration (mail vs. telephone).

We also evaluated whether there were interactions between the four most prevalent cancers (female breast, prostate, colorectal, lung) and the 13 non-cancer chronic conditions in the model. We examined whether stage of disease was related to SF-6D scores for the four most prevalent cancers (female breast, prostate, colorectal, non-small cell lung). For uniformity across conditions, we coded stage of disease into localized cancer, distant (metastatic) cancer, and onstage. Regression parameter standard errors were adjusted for clustering at the health plan level using the sandwich estimator of variance [12].

Recycled predictions [13], or averaging of individual marginal effects, are used to understand the incremental effect of an independent variable on a dependent variable. We used recycled predictions to obtain adjusted SF-6D means for each cancer and non-cancer condition group. We created two variants of these predicted score. In the first approach, we fixed all other independent variables other than the condition being predicted at their means. In the second approach, we fixed the 22 other conditions at zero and the remaining independent variables at their means. The advantage of fixing the other conditions at the mean (first approach) is that the adjusted scores correspond to the overall sample mean on the dependent variable. However, this approach can yield counterintuitive results. In our data, for example, the unstandardized beta for sciatica (−0.037) is more negative than for GI (−0.031) yet the recycled prediction based on the means for the other conditions is slightly larger for sciatica (0.701) than GI (0.700). This occurs because sciatica is more prevalent than GI–the mean for sciatica (0.21) is larger than that for GI (0.05). Hence, when we predict the GI score, the larger sciatica mean drives the estimated GI score down more than the GI mean drives down the estimated sciatica score. The second approach (fixing the other conditions at zero) avoids these sorts of differences between the betas and the recycled predictions but the predicted scores are higher and do not correspond to the sample mean on the dependent variable.

Results

As shown in Table 1, the sample of 126,366 respondents had an average age of 74 years (range 65–106). Forty-five percent were male, 79 % non-Hispanic white, 60 % married, and 27 % had less than a high school degree. The median income was less than $30,000. The average number of reported chronic medical conditions other than cancer was 2.44 (range = 0–13). Sample characteristics for those with and without cancer were similar, but those with cancer were more likely to be male, white, and married.

The average SF-6D score in the entire sample was 0.73 (SD = 0.14), ranging from 0.30 to 1.00. Only 0.16 and 1 % of the sample had the lowest and highest observed

Table 1 Sample characteristics

	Overall (n = 126,366)	Non-cancer (n = 103,626)	Cancer (n = 22,740)
Mean age (range)	74 (65–106)	74 (65–106)	75 (65–101)
Mean number of conditions over than cancer (range)	2.44 (0–13)	2.42 (0–13)	2.51 (0–13)
Male (%)	45	43	53
Race/ethnicity			
Hispanic (%)	7	8	5
Non-hispanic white (%)	79	79	82
Non-hispanic black (%)	5	5	5
Asian (%)	5	5	4
American Indian (%)	1	1	1
Other race (%)	2	2	2
Missing race (%)	1	1	1
Married (%)	60	60	62
Education			
8th grade or less (%)	11	11	10
Some high school (%)	14	14	14
High school graduate (%)	33	33	32
Some college (%)	24	24	25
4-year college graduate (%)	8	8	9
More than 4-year degree (%)	9	9	10
Income			
< $10,000 (%)	11	11	10
$10,000–19,999 (%)	22	22	22
$20,000–29,999 (%)	18	18	18
$30,000–39,999 (%)	12	12	12
$40,000–49,999 (%)	7	7	8
$50,000–79,999 (%)	8	8	8
$80,000 or more (%)	4	4	5
Do not know/missing (%)	18	18	17

scores, respectively. SF-6D scores were consistent across the 4 MHOS cohorts, with the same mean and SD observed for each cohort as for the entire sample.

The regression model with 43 degrees of freedom in the numerator accounted for 39 % of the (adjusted) variance in the SF-6D (see Table 2). All except two conditions (melanoma, endometrial cancer) had significant unique negative associations with the SF-6D score. As a sensitivity analysis to address the concern that depressive symptoms overlap with the dependent variable, we reran the regression model without it and found little impact on the coefficients for the other conditions.

Adjusted mean scores (recycled predictions) using the first approach (fixing other conditions to their mean values) rounded to two decimal places were 0.73 (melanoma), 0.72 (endometrial cancer, colorectal cancer, female breast cancer, prostate cancer, bladder cancer, non-Hodgkin lymphoma, kidney cancer, myocardial infarction/heart attack, hypertension, angina/coronary artery disease, other heart disease), 0.71 (other cancer, non-small cell lung cancer, diabetes, arthritis of the hand), 0.70 (congestive heart failure, arthritis of the hip, inflammatory bowel disease, sciatica, chronic obstructive pulmonary disease), 0.69 (stroke), and 0.61 (depressive symptoms). Adjusted mean scores using the second approach are given in the last column of Table 2.

Only 6 of the 52 two-way interactions between the four most prevalent cancers and the 13 non-cancer conditions were statistically significant ($p < 0.05$). Small negative coefficients were found for the interactions between colorectal cancer and diabetes, and lung cancer and chronic obstructive pulmonary disease/asthma, while small positive coefficients were found between sciatica and lung cancer, hypertension and prostate cancer, hypertension and colorectal cancer, and other heart disease and female breast cancer. We do not interpret these interactions given the inconsistent directions and because they could have occurred by chance.

Table 2 Unique associations of chronic health conditions with SF-6D scores from ordinary least squares regression model

Condition	Unstandardized beta	Standard error	t-statistic	p value	Adjusted score*	Adjusted score**
Cancers						
Melanoma ($n = 1,135$)	−0.002	0.003	−0.48	0.630	0.728	0.799
Endometrial cancer ($n = 902$)	−0.006	0.004	−1.54	0.124	0.724	0.795
Colorectal cancer ($n = 3,012$)	−0.006	0.002	−2.87	0.004	0.724	0.795
Female breast cancer ($n = 4,311$)	−0.006	0.002	−3.85	0.000	0.724	0.795
Prostate cancer ($n = 5,593$)	−0.008	0.002	−4.78	0.000	0.722	0.793
Bladder cancer ($n = 1,299$)	−0.008	0.003	−2.98	0.003	0.722	0.793
Non-Hodgkin lymphoma ($n = 668$)	−0.012	0.005	−2.60	0.010	0.718	0.789
Kidney cancer ($n = 488$)	−0.014	0.004	−3.22	0.001	0.716	0.787
Other cancer ($n = 3,540$)	−0.016	0.002	−7.43	0.000	0.714	0.785
Non-small cell lung cancer ($n = 1,792$)	−0.024	0.002	−9.81	0.000	0.706	0.777
Non-cancer chronic conditions						
Myocardial infarction/heart attack ($n = 11,982$)	−0.006	0.001	−4.79	0.000	0.725	0.795
Hypertension ($n = 66, 968$)	−0.013	0.001	−18.77	0.000	0.724	0.788
Other heart disease ($n = 25,455$)	−0.017	0.001	−21.18	0.000	0.716	0.794
Angina/coronary artery disease ($n = 18,017$)	−0.017	0.001	−18.93	0.000	0.715	0.784
Diabetes ($n = 20,089$)	−0.022	0.001	−23.22	0.000	0.712	0.779
Arthritis of the hand ($n = 40,402$)	−0.023	0.001	−25.71	0.000	0.714	0.778
Congestive heart failure ($n = 7,893$)	−0.029	0.001	−20.87	0.000	0.703	0.772
Inflammatory bowel disease ($n = 5,882$)	−0.031	0.001	−24.11	0.000	0.700	0.770
Sciatica ($n = 26,878$)	−0.037	0.001	−52.64	0.000	0.701	0.764
Stroke ($n = 9,479$)	−0.039	0.001	−30.42	0.000	0.694	0.762
Chronic obstructive pulmonary disease/asthma ($n = 5,445$)	−0.040	0.001	−43.77	0.000	0.695	0.761
Arthritis of the hip ($n = 44,524$)	−0.044	0.001	−64.29	0.000	0.702	0.758
Depressive symptoms ($n = 14,815$)	−0.131	0.001	−131.29	0.000	0.614	0.670

The model adjusted for each chronic condition, education, age, and marital status, mode of data collection, race, and income. The regression model had 43 degrees of freedom in the numerator and accounted for 39 % of the adjusted variance. The unstandardized beta coefficients column indicates the direction and magnitude of difference in the adjusted SF-6D score for each chronic condition

* Adjusted score is from recycled predictions with other independent variables fixed at their means. The adjusted score for those without a condition was 0.749

** Adjusted score is from recycled predictions with the other 22 conditions fixed at zero and other independent variables fixed at their means. The adjusted score for those without a condition was 0.806

There were small numbers of distant stage cancer (26 female breast, 61 prostate, 48 colorectal, and 47 lung) but distant stage of disease was significantly associated with worse health utility scores, with distant stage of disease being worse than localized disease by 0.044 for prostate cancer, 0.046 for colorectal cancer, 0.058 for lung cancer, and 0.077 for female breast cancer.

Discussion

The average SF-6D score in this sample (0.73) is similar to the mean of 0.77 reported for participants 70–79 years of age in the National Health Measurement Study during 2005–2006 [14]. The lower mean in this sample is understandable given that the majority of respondents (88 %) completed the survey by mail, whereas telephone mode of data collection was used for the National Health Measurement Study. Telephone administration tends to yield more positive HRQOL scores [15]. The average SF-36 v. 1 physical component and mental component summary scores (T-score metric with 50 mean and SD of 10 in US general population) were 42 and 52, respectively. Hence, this sample of older individuals had substantially worse physical health (large effect size) but slightly better mental health (small effect size) than the US adult general population. The means for the MHOS sample are very similar to those observed for persons ages 65–74 in the US general

population where the PCS and MCS were 43 and 53, respectively [16].

The strongest association with the SF-6D preference-based score was observed for depressive symptoms (−0.131). This is not surprising because the depression question was the only one that had a 1-year reference period. Other health conditions captured in the MHOS could have occurred recently or several years ago because people were asked whether they had ever been told by a doctor that they had the condition. Further, a strong negative association of depression with the SF-6D was expected because the measure includes mental health items. Thus, depressive symptoms are represented to some extent on both sides of the equation. Dropping depressive symptoms from the model had no impact on the interpretation of the associations for the other 20 comorbid conditions that had significant unique associations with the SF-6D score.

The largest decrements in HRQOL for the remaining conditions were observed for arthritis of the hip, sciatica, chronic obstructive pulmonary disease/asthma, stroke, inflammatory bowel disease, and congestive heart failure (betas ranging from −0.029 to −0.044). In contrast, the smallest significant associations were −0.006.

The four conditions with the strongest significant unique associations with the SF-6D in this study (arthritis of the hip, sciatica, chronic obstructive pulmonary disease/ asthma, stroke) were found to have relatively large associations with the SF-36 physical component summary score by Smith et al. [3]. Three of the conditions also had strong associations with the SF-36 mental health summary score, while arthritis of the hip was not as strongly related [3].

The majority of cancer types were significantly associated with the SF-6D score, with beta coefficients (rounded to two decimal places) ranging from −0.01 to −0.02 on the 0–1 health utility scale. These results are similar to what was observed in a sample of 38,678 individuals from the 2000–2002 in the medical expenditure panel survey by Sullivan et al. [17]. Specifically, Sullivan et al. [17] reported "disabilities" of −0.02 for prostate cancer, −0.01 for breast cancer, and −0.01 for other cancer.

Walters and Brazier [18] reviewed 7 studies and found that estimates of the minimally important difference (MID) for the SF-6D ranged from 0.01 to 0.05 with a weighted mean of 0.03. Similarly, Khanna et al. [19] reported MID estimates of about 0.03 in a sample of persons with systemic sclerosis. Hence, the magnitudes of the unique associations of types of cancer with health utility score are not trivial. In comparison, the largest unique association of chronic conditions (other than depressive symptoms) with the SF-6D in the current study was −0.04.

While the individual impact of conditions including cancer was typically not large, the differences could matter in group comparisons. For example, a difference of 0.02 on the utility scale would be considered cost-effective if it cost about $1,000 to produce (i.e., $50,000/QALY). Moreover, the cumulative effect of multiple conditions is substantial as evidenced by the plethora of significant unique associations in the regression model. In fact, the adjusted score for those reporting any one condition was on average 0.03 lower than those reporting no conditions, while the adjusted score for those reporting two or more conditions was on average 0.11 lower (results not reported earlier). Finally, the decrements in HRQOL cumulate over time, so the full impact relative to those without chronic conditions is the observed decrement multiplied by the number of years with the condition.

We found that stage of disease had a profound impact on HRQOL. Those with distant (metastatic) stage of disease had health utility scores that were 0.044–0.077 worse than those with localized cancer. Because the number of people with distant stage of disease was so small in the dataset, the overall relationship of cancer with health utility scores was determined entirely by those with less advanced disease when stage was excluded from the model. The large negative decrement in HRQOL in late stage of disease highlights the importance of clinical interventions to ameliorate these negative effects on functioning and well-being.

Although this study has a number of strengths, it also has limitations. Due to the cross-sectional design, we are unable to make definite conclusions about directionality. In addition, other than stage of cancer the study does not have information about the severity of the chronic conditions examined, nor the time when the condition was diagnosed. A previous study found similar SF-36 scores by time since diagnosis but there was some indication of a healthy survivor effect in colorectal and lung cancer patients [3]. We also have not captured information about some common conditions among older individuals such as osteoporosis, benign prostatic hypertrophy, or dementia. Aside from cancer, we had to rely on self-reports of chronic conditions. However, data comparing comorbidities self-reported on the same survey we used (MHOS) versus abstraction of medical records suggest reasonably good correspondence, with median specificity (% of time condition is not self-reported when it is not in the medical record) and sensitivity (% of time condition is self-reported when it is in the medical record) of 69 and 91 %, respectively [20]. In addition, the data were collected using both mail surveys and telephone interviews and included some proxy responses. But we adjusted for these variables in the regression model. Moreover, the SEER-MHOS dataset includes four MHOS cohorts and data collected back in 1998. However, average SF-6D scores for the four cohorts were exactly the same, indicating no temporal shifts. While some preference-based measures (e.g., the EQ-5D) have

been criticized for lack of sensitivity [21], the results of this study suggest that the SF-6D is sensitive to the impact of cancer on HRQOL. However, previous research indicates that different preference measures may not yield the same results [22]. The current work needs to be replicated in other large-scale surveys of cancer patients using different preference-based measures.

This study provides important information about the relative burden of different chronic conditions on HRQOL in Medicare beneficiaries. It indicates that a large number of conditions are associated uniquely with decrements in health utility and that the cumulative effects are substantial. In addition, distant stage of disease for the four big cancers (female breast, prostate, colorectal, lung) is associated with large, negative impact on utility among older individuals in the United States. The findings reported here are particularly important given the aging US population and increasing number of persons 65 years and older. As SEER-MHOS adds new cohorts and larger sample sizes by stage of disease become available, further investigation of the impact of cancer and stage of disease will be possible. Tracking the association of chronic conditions and stage of disease on HRQOL periodically can give providers and federal agencies such as the Center for Medicare and Medicaid Services important information about the extent to which the health needs of older Americans are being addressed.

References

1. Rothrock, N. E., Hays, R. D., Spritzer, K., et al. (2010). Relative to the general US population chronic diseases are associated with poorer health-related quality of life as measured by the patient-reported outcomes measurement information system (PROMIS). *Journal of Clinical Epidemiology, 63*, 1195–1204.
2. Altekruse SF, Kosary CL, Krapcho M, et al (eds) SEER Cancer Statistics Review 1975–2007 National Cancer Institute: Bethesda MD. Available from URL: http://seer.cancer.gov/csr/1975_2007 Accessed 27 Aug 2012.
3. Smith, A. W., Reeve, B. B., Bellizzi, K. M., et al. (2008). Cancer comorbidities and health-related quality of life of older adults. *Health Care Financing Review, 29*, 41–56.
4. Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of health economics, 21*, 271–292.
5. Brazier, J., & Roberts, J. (2004). The estimation of a preference-based measure of health from the SF-12. *Medical Care, 42*, 851–859.
6. Ambs, A., Warren, J. L., Bellizzi, K. M., et al. (2008). Overview of the SEER-medical health outcomes survey (MHOS) linked dataset. *Health Care Financing Review, 29*, 5–22.
7. Ries LAG, Melbert D, Krapcho M, et al (eds.) SEER Cancer Statistics Review 1975–2004 Available from URL: http://seer.cancer.gov/csr/1975_2004/ Based on November 2006 SEER data submission. Accessed 27 Aug 2012.
8. Clauser, S. B., Arora, N. K., Bellizzi, K. M., et al. (2008). Disparities in HRQOL of cancer survivors and non-cancer managed care enrollees. *Health Care Financing Review, 29*, 23–40.
9. Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36): Conceptual framework and item selection. *Medical Care, 30*, 473–483.
10. Brazier, J., Rowen, D., & Hanmer, J. (2008). Revised SF-6D scoring programmes: a summary of improvements. *Patient Reported Outcomes Newsletter, 48*, 16–17. Fall.
11. Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review Publish Health, 23*, 151–169.
12. Williams, R. L. (2000). A note on robust variance estimation for cluster-correlated data. *Biometrics, 56*, 645–646.
13. Basu, A., & Rathouz, P. J. (2005). Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. *Biostatistics, 6*, 93–109.
14. Hanmer, J. (2009). Predicting an SF-6D preference-based score using MCS and PCS scores from the SF-12 or SF-36. *Value Health, 12*, 958–966.
15. Hays, R. D., Kim, S., Spritzer, K. L., et al. (2009). Effects of mode and order of administration on generic health-related quality of life scores. *Value Health, 12*, 1035–1039.
16. Ware JE, Kosinski M, Keller SD. (1994). SF-36 physical and mental health summary scales: A user's manual boston massachusetts: New England medical center.
17. Sullivan, P. W., Lawrence, W. F., & Ghushchyan, V. (2005). A national study of preference-based scores for chronic conditions in the United States. *Medical Care, 43*, 736–749.
18. Walters, S. J., & Brazier, J. E. (2003). What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health Quality Life Outcomes, 11*, 1–4.
19. Khanna, D., Furst, D. E., Wong, W. K., et al. (2007). Reliability validity and minimally important differences of the SF-6D in systemic sclerosis. *Quality of Life Research, 16*, 1083–1092.
20. Miller, D. R., Rogers, W. H., Kazis, L. E., et al. (2008). Patients' self-report of diseases in the medicare health outcomes Survey based on comparisons of linked survey and medical data from the veterans administration. *The Journal of ambulatory care management, 31*, 161–177.
21. Garau, M., Shah, K., Mason, A. R., et al. (2011). Using QALYs in cancer: A review of the methodological limitations. *Pharmacoeconomics, 29*, 673–685.
22. Feeny, D., Spritzer, K., Hays, R. D., et al. (2011). Agreement about identifying patients who change over time: Cautionary results in cataract and heart failure patients. *Medical Decision Making, 32*, 273–286.