The Centers for Medicare & Medicaid Services' Center for Medicare (CM) strives to make information available to all. Nevertheless, portions of our files including charts, tables, and graphics may be difficult to read using assistive technology. Persons with disabilities experiencing problems accessing portions of any file should contact CM through e-mail at 508_Compliance@cms.hhs.gov.

Medicare Health Outcomes Survey

Final Report

on

Analysis of Key Drivers of Improving or Maintaining Medicare Health Outcomes Survey (HOS) Scores

Health Services Advisory Group
May 2013

CMS
CENTERS FOR MEDICARE & MEDICAID SERVICES
CENTER FOR MEDICARE

HSAG HEALTH SERVICES ADVISORY GROUP

Table of Contents

List of Tables

List of Figures

Executive Summary

Using data from the HOS *2009 Cohort 12 Baseline* and *2011 Cohort 12 Follow-Up*, we describe how two-year mortality and two-year changes in the Veterans RAND 12-Item Health Survey (VR-12) items relate to key Medicare Health Outcomes Survey (HOS) measures used in the Medicare Star Ratings. The HOS measures relate to maintaining and improving health and are derived from changes in the physical component summary (PCS) and mental component summary (MCS) scores. The results from this study clarify the properties of several Centers for Medicare & Medicaid Services' (CMS) quality measures used for the Medicare Star Ratings (also known as Medicare Advantage Plan Ratings).

Background

Since Spring 2006, CMS has administered the VR-12 to Medicare enrollees as part of the Medicare HOS to monitor the quality of care in Medicare Advantage Organizations (MAOs) (Kazis et al., 2011). The VR-12 provides a reliable and valid measure of health status and is used by CMS to assess physical and mental health functioning of beneficiaries and to guide payment adjustments (Iqbal et al., 2007). The VR-12 items are summarized into two widely-used measures: the PCS and MCS scores. Since 2009, the PCS and MCS scores have been publicly reported in the Medicare Star Ratings as part of the CMS quality improvement (QI) efforts. Since 2012, the Medicare Star Ratings are used to link quality of care to payment for MAOs with the first Quality Bonus Payments to begin in 2013, thus incentivizing improvement of quality indicators such as the PCS and MCS scores.

In this report, we describe the relationships among two-year mortality, two-year changes in VR-12 items, and two-year changes in PCS and MCS scores. We undertake beneficiary and contract-level analyses to (a) identify which items most influence contract-level PCS and MCS scores, and (b) identify MAOs that do particularly poorly on mortality and individual VR-12 items given their performance on the other remaining 12 items to identify specific QI opportunities for the identified MAOs.

Instruments and Data Source

The HOS, a longitudinal survey that assesses the physical and mental health functioning of beneficiaries, was first fielded nationally in 1998, and is the first patient-based outcomes measure in Medicare managed care.

We used data from respondents to the 2009 HOS 2.0, which obtained a sample from all MAOs with a minimum enrollment of 500 members. MAOs with Medicare contracts in effect on or before January 1, 2008, including local and regional preferred provider organizations (PPOs) and continuing cost contracts, participated in the HOS Baseline Survey in 2009. MAOs composed exclusively of special needs plan (SNP) benefit packages, regardless of institutionalized, chronically ill or dual eligible enrollment, were also included in the requirement. Private fee-for-service (PFFS) plans could voluntarily report HOS in 2009 (NCQA, 2009). The data also included the follow up respondents to the 2011 HOS 2.0. MAO respondents from the HOS Baseline Survey in 2009 were remeasured using the HOS Follow-Up Survey in 2011 (NCQA, 2011).

We limited individual-level analyses to the 105,477 respondents aged 65 and over (from 381 MAOs) who had baseline PCS and MCS scores and had follow-up (2011 remeasurement) PCS and MCS scores.

Contract-level analyses included an additional 17,936 respondents who had died before remeasurement, and were part of the original sample of 424 MAOs at baseline.

Methods

We present five sets of analyses. First, we examined specific transitions between categories of baseline and follow-up responses for each VR-12 item. These included (a) the prevalence of each possible transition, (b) the conditional prevalence of each follow-up response given the baseline response, and (c) the relationship between each transition and individual risk-adjusted and categorized PCS and MCS scores.

Second, we used individual-level regression models to model how each VR-12 item contributes to the change in PCS or MCS scores. We modeled both the scoring distinguishing "better than expected" from "same as expected" and "worse than expected," and the scoring that distinguished "better than expected" and "same as expected" from "worse than expected." Our key predictor variables in individual-level regression models are follow-up minus baseline scores for each of the VR-12 measures. We used linear scoring of ordinal categories, and in a secondary scoring approach, we transformed these linear scores into Z-scores.

Third, we examined contract-level variation in mortality and VR-12 item change scores in order to identify those factors with the largest contract-level variation. In combination with the individual-level analyses from the first two sets of models, this provided insight into which items have the most influence on contract-level scores. Items with greater contract-level variation in change scores may also have greater potential for contract-level influence.

The fourth set of analyses predicted change in PCS and MCS scores at the contract level. These analyses complement the first three sets of analyses by including mortality information in the PCS (but not MCS) score, and by considering covariation in item changes at the contract level. In contract-level models, the summary scores are expressed as the percentage of Medicare Advantage (MA) members whose two-year changes in PCS and MCS scores are "better than expected" (or "better than or same as expected").

The fifth set of analyses differs from the first four in that it focused on identifying specific contracts that are outliers on specific VR-12 item changes or mortality. We classified regression residuals in order to identify specific contracts that performed particularly well or poorly on one or more of the VR-12 item change scores or mortality relative to their overall performance on the remaining items in the models. These analyses identified specific QI opportunities for individual contracts and may aid CMS oversight of MAOs.

Key Findings

The five transitions that most influence the MCS change from baseline to follow-up involve three items (*Moderate activities*, *Health interferes with social activities,* and *Felt downhearted and blue*). These five transitions were: (1) the transition from "not limited at all" to "limited a little" in *Moderate activities*, (2) the transition from "none of the time" to "a little of the time" in *Health interferes with social activities,* (3) the transition from "none of the time" to "some of the time" in *Health interferes with social activities*, (4) the transition from "none of the time to "a little of the time" in *Felt downhearted and blue,* and (5) the

transition from "a little of the time" to "none of the time" in *Felt downhearted and blue*. The 12 most influential transitions for PCS change involve four VR-12 items (*Moderate activities, Climbing several flights of stairs, Limited in the kind of work or activities due to physical problems*, and *How much pain interferes with normal work*). These 12 transitions were: (1) the transition from "not limited at all" at baseline to "limited a little" at follow-up in *Moderate activities*, (2) the transition from "limited a little" to "limited a lot" in *Moderate activities,* (3) the transition from "limited a little" to "not limited at all" in *Moderate activities,* (4) the transition from "not limited at all" to "limited a little" in *Climbing several flights of stairs,* (5) the transition from "limited a little" to "limited a lot" in *Climbing several flights of stairs*, (6) the transition from "limited a little" to "not limited at all" in *Climbing several flights of stairs,* (7) the transition from "limited a lot" to "limited a little" in *Climbing several flights of stairs,* (8) the transition from "none of the time" to "a little of the time" in *Limited in the kind of work or activities due to physical problems,* (9) the transition from "none of the time" to "some of the time" in *Limited in the kind of work or activities due to physical problems,* (10) the transition from "not at all" to "a little bit" in *How much pain interferes with normal work,* (11) the transition from "not at all" to "moderately" in *How much pain interferes with normal work,* and (12) the transition from "a little bit" to "not at all" in *How much pain interferes with normal work.*

The VR-12 items that most influence MCS per one-unit change are *Accomplished less than you would like to due to emotional problems* and *Felt downhearted and blue*. Because a one-unit change may be more difficult or less common for some VR-12 items than for others, we also modeled the effect of moving one standard deviation (SD) for each VR-12 item. The most influential VR-12 items per SD are *Accomplished less than like to due to emotional problems,* followed by *Felt calm and peaceful* and *Felt downhearted and blue.*

The VR-12 items that most influence PCS per one-unit change are *Moderate activities, Climbing several flights of stairs* and *How much pain interferes with normal work*. Moving just one level of improvement on each of these measures is associated with a 10% increase in PCS scores. When considering the effects of one SD VR-12 item change, *How much pain interferes with normal work* was the most influential item, followed by *Climbing several flights of stairs, Accomplished less than you would like to due to physical problems,* and *Moderate activities*. In the secondary contract-level models that allowed us to compare the effect of mortality with the effects of the VR-12 change scores, we find that the effect of a contract's preventing one death is equivalent to preventing a 1.5 unit decline or causing a 1.5 linear unit improvement on the most important VR-12 measure for one beneficiary. The standardized model similarly indicates that the effect of one contract-level SD in mortality rate is just slightly greater than the effect of one contract-level SD on the most influential VR-12 item, *How much pain interferes with normal work*.

Contract-level intra-class correlations (ICCs) in change scores measure the proportion of variance in change scores associated with contracts as opposed to individual beneficiaries. A statistically significant ICC indicates evidence for at least some contract-level variation. A statistically significant ICC of at least 0.10 indicates substantial contract-level variation. Twelve (of 13) ICCs associated with the VR-12 item change scores and mortality are statistically significant, with the one exception being *Felt downhearted and blue*. This may indicate that MAOs have some influence over all VR-12 items with the possible exception of *Felt downhearted and blue* change. Eight of the 13 items have ICCs exceeding 0.10: *2-year*

Mortality rate, both Role-Emotional items (*Didn't do work or other activities as carefully*, *Accomplished less than you would like to due to emotional problems*), *Vitality*, *General health*, *Climbing several flights of stairs*, and both role limitations due to physical problems measures (*Accomplished less than you would like to due to physical problems* and *Limited in the kind of work or activities*).

The VR-12 item *Accomplished less than you would like to due to emotional problems* is particularly influential in change in MCS scores and the extent to which change in this item varies substantially across contracts may indicate notable MAO influence on the item. Three other VR-12 items that both strongly influence change in PCS scores and vary notably across contracts in their degree of change over time are *Moderate activities*, *Climbing several flights of stairs*, and *How much pain interferes with normal work*. We identified specific transitions associated with these same three VR-12 items (*Moderate activities*, *Climbing several flights of stairs*, and *How much pain interferes with normal work*) that may merit particular focus by MAOs and CMS.

Finally, we examined contract-level outliers on individual mortality and individual VR-12 items given overall contract performance. The two Role-Physical items (*Accomplished less than you would like to due to physical problems* and *Limited in the kind of work or activities*) and the two Role-Emotional items (*Didn't do work or other activities as carefully* and *Accomplished less than you would like to due to emotional problems*) were the items best predicted by the other items. *General health*, *Felt downhearted and blue*, and *2-year Mortality rate* were least well predicted. For all 13 items, there are many more instances of negative outliers (a dimension far below what was expected from a contract's performance on other dimensions) than positive outliers. The number of negative outliers ranges from 3 contracts (for *Climbing several flights of stairs*) to 20 contracts (*Felt calm and peaceful*) and the range of positive outliers is 0 (*Moderate activities*, *Accomplished less than you would like to due to physical problems*, and *Felt downhearted and blue*) to 3 (*How much pain interferes with normal work*). A listing of outliers by contracts suggests that the outlier patterns are unrelated to a contract's parent organization. Efforts to improve performance on VR-12 item change scores and mortality are likely to be more effective at the contract rather than sponsor level.

Implications

Our analysis of the transitions influencing MCS has implications for monitoring asymptomatic beneficiaries (those in good mental health, with high MCS scores) in order to detect the emergence of mental health problems and suggests that MAOs should offer easy access to high-quality self-health tools for emergent symptoms. Our analysis has also identified four items that have particular potential for QI efforts, in that they may be both influential and subject to intervention. These QI efforts may involve initiatives that the literature suggests are effective at improving Role-Physical and Role-Emotional items. Our models predicting outliers indicate that MAOs that can improve any one of these items are likely to also positively affect other items as well.

Such QI initiatives might involve interventions to help identify members who could benefit from learning strategies to help them accomplish more given their health conditions and limitations rather than, for example, focusing on trying to change the underlying mental state. In other words, PCS and MCS scores appear more driven by the consequences of the underlying states than by the states themselves. Such interventions might focus on teaching coping strategies for living with chronic physical or mental health

conditions, for example, and screening for people who can benefit from such strategies. Several interventions have been shown to improve role functioning in sub-populations of older adults (Kazis et al., 2011). For example, a "behavior change-focused weight management program," that involved a bundled set of clinical interventions, improved *Role limitations due to physical problems* and *Physical functioning* among frail obese adults (Blissmer et al., 2006). An intervention that involves clinician counseling offered through "exercise prescriptions" in a sample of sedentary women found positive effects on *Role limitations due to physical problems* and *Mental health* after 12 months (Lawton et al., 2008). Such interventions, which involve cognitive-behavioral techniques such as motivational interviewing and goal setting, often positively affect other VR-12 measures.

Findings also suggest that efforts to improve performance on mortality and the VR-12 items are likely to be effective at the contract rather than sponsor level.

Limitations

The HOS is restricted to the MA population; we cannot be certain if the results would apply to a more general beneficiary population without conducting analyses in a different population. All associations described in this report are from observational (rather than experimental) data, which limits our ability to draw causal inferences.

While a large variation in VR-12 change across contracts is likely to reflect MAO influence on health, it is also possible that some of the observed variance in change scores at the contract level reflects differences in the mix of patients who are more likely to decline on different measures.

Future Work

We identified several VR-12 items that multiple strands of analyses identify as particularly worthy of targeted intervention: *Accomplished less than you would like to due to emotional problems, Moderate activities, Climbing several flights of stairs,* and *How much pain interferes with normal work*. Future qualitative work could probe MAOs that are positive (and negative) outliers along these specific domains to identify possible strategies that the MAOs may be using to move these items, and, indirectly, their PCS and MCS outcomes.

Since we were able to identify contracts with outliers on mortality and specific VR-12 items, future work can use these outliers to address substantive questions. Examples of such analyses include examining contract characteristics that are associated with outlier status to identify aggregate patterns. Future work can also investigate the year-to-year stability of outliers. If this stability is not high, this might indicate that change in a VR-12 item is unreliably measured or, alternatively, that MAOs are able to improve their lagging dimensions. Nearly one in three (30.6%) beneficiaries disenrolled from their MAO between baseline and follow-up. The current methodology for calculating PCS and MCS outcomes excludes disenrollees, so that their subsequent health is not attributed to the MAO from which they disenrolled. One research question is whether beneficiaries who receive poor care and are particularly likely to experience a decline in PCS or MCS scores are also more likely to disenroll. If so, then the current approach may fail to capture their poor care by excluding those who disenrolled because of poor care. Finally, future work may also focus on identifying specific strategies used by MAOs to improve Role-Physical and Role-Emotional measures for members, especially members living with chronic physical or

emotional conditions. Such strategies could be more widely shared with contracts that perform "worse than expected" based on their performance on mortality and other VR-12 measures to improve the overall performance of MAOs on change in PCS and MCS measures. These strategies could be supplemented with an inventory of evidence-based interventions and approaches that have been shown to improve these same outcomes in older populations through a review of the published literature.

Chapter 1. Introduction

Purpose

Using data from the HOS *2009 Cohort 12 Baseline* and *2011 Cohort 12 Follow-Up*, we describe how two-year mortality and two-year changes in the Veterans RAND 12-Item Health Survey (VR-12) items relate to key Medicare Health Outcomes Survey (HOS) measures used in the Medicare Star Ratings. The HOS measures relate to maintaining and improving health and are derived from changes in the physical component summary (PCS) and mental component summary (MCS) scores. Data collected from the HOS *2009 Cohort 12 Baseline* and *2011 Cohort 12 Follow-Up* were used. The results from this study will clarify the properties of the Centers for Medicare & Medicaid Services' (CMS) quality measures used for the Medicare Star Ratings. MAOs will benefit from identification of high leverage items that may help prioritize quality improvement (QI) efforts.

Background

The VR-12 was first administered in national VA surveys in 1997. Since Spring 2006, CMS has administered the VR-12 to Medicare enrollees as part of the Medicare HOS to monitor the quality of care in Medicare Advantage Organizations (MAOs) (Kazis et al., 2011). The VR-12 provides a reliable and valid measure of health status and is used by CMS to assess physical and mental health functioning of beneficiaries and to guide payment adjustments (Iqbal et al., 2007).

MAOs and other large health care systems can use VR-12 change scores as performance indicators to compare health outcomes among different beneficiary populations with one another and with beneficiaries in different systems. They can also be used to identify processes of care or management practices that can serve as best practices within a system and across systems (Iqbal et al., 2007).

The VR-12 is a self-administered 12-item instrument that primarily assesses quality of life. The 12 items correspond to eight physical and mental health domains: 1) Physical Functioning (PF02, PF04), 2) Role-Physical (RP2, RP3), 3) Role-Emotional (RE2, RE3), 4) Bodily Pain (BP2), 5) Social Functioning (SF2), 6) Mental Health (MH3, MH4), 7) Vitality (VT2), and 8) General Health (GH1). The 12 items are summarized into the PCS and MCS scores. Typically, the PCS and MCS measures are widely used.

Since 2009, the PCS and MCS measures have been publicly reported in the Medicare Star Ratings as part of the CMS QI efforts. Since 2012, the Medicare Star Ratings are used to link quality of care to payment for MAOs with the first Quality Bonus Payments beginning in 2013, thus incentivizing improvement of quality indicators such as the PCS and MCS outcomes measures.

In this report, we focus on two-year changes in the individual VR-12 items (and the mortality indicator, which in combination with PCS change is used to compute overall change in physical health). Specifically, we undertake beneficiary and contract-level analyses to (a) identify which VR-12 items (and mortality) have the most influence on contract-level PCS and MCS scores, and (b) identify MAOs that do particularly poorly on mortality or an individual VR-12 item given their performance on the 12 remaining items to identify specific QI opportunities (and potential strategies) for MAOs.

Description of HOS

The Medicare HOS is a longitudinal survey that has an annual baseline measurement of a sample of beneficiaries, including both aged and disabled, from each participating MAO, and a two-year follow-up of the baseline respondents (NCQA, 2009; NCQA, 2011). MAOs with 500 or more members at baseline are required to participate. For MAOs with greater than 1,200 members, a random sample of 1,200 is drawn, whereas all members are sampled from contracts with 500 to 1,200 members. The HOS is a mail survey with telephone follow-up of mail nonrespondents. If a member fails to respond after two mailings or returns an incomplete survey, survey vendors attempt at least six follow-up telephone calls.

Beneficiaries age 65 or older who (a) complete the baseline survey, (b) complete enough baseline VR-12 items for PCS or MCS scores to be calculated, and (c) are in contracts that remain in the HOS at follow-up, are considered the *Performance Measurement Analytic Sample*. The analytic sample is used to determine the number that died within each MAO between the baseline and follow-up measurements. Surviving members who remain in their same MAO are sent the follow-up survey. Members who complete the follow-up survey are the *Performance Measurement Respondent Sample* and those for whom changes in PCS and MCS scores can be calculated.

The HOS performance measurement analyses determine the percentages of an MAO's beneficiaries who are better, the same, or worse than expected in terms of physical and mental health at the two-year follow-up and compare these rates to national averages. The primary outcomes are death, change in physical health as measured by PCS scores, and change in mental health as measured by MCS scores. Death and PCS scores are combined into one overall measure of change in physical health (Rogers et al, 2004). The data are summarized at the beneficiary level to determine the actual and expected numbers who are alive with a PCS score that is better or the same at follow-up, who die or whose PCS score is worse at follow-up, as well as those whose MCS score is better, the same, or worse at follow-up. The data are aggregated at the contract-level to summarize the actual and expected mean death rates, the percentage of beneficiaries who were alive and had a PCS score that was the same or better at follow-up, and the percentage of beneficiaries who had an MCS score that was the same or better at follow-up.

HOS and the Medicare Star Ratings

CMS developed the Medicare Star Ratings in order to provide information about MAOs in the Medicare Plan Finder (MPF) tool on the www.medicare.gov website for consumers. CMS rates the relative quality of service of MAOs based upon a five-star rating scale that utilizes the contract level HOS scores combined with other measurement results. The Medicare Star Ratings helps Medicare beneficiaries compare MAOs, helps educate consumers on quality, and makes quality data more transparent and comparable among MAOs. Up to 49 unique quality measures are included in the ratings, including success in providing preventive services, managing chronic illness, access to care, Healthcare Effectiveness Data and Information Set (HEDIS®)[1] measures, the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) survey, and responsiveness.

Five HOS measures are included in the Medicare Star Ratings: two measures of functional health (the *Improving or Maintaining Physical Health* and the *Improving or Maintaining Mental Health* measures)

[1] HEDIS® is a registered trademark of the National Committee for Quality Assurance.

and three HEDIS Effectiveness of Care measures (*Improving Bladder Control*, *Monitoring Physical Activity*, and *Reducing the Risk of Falls* measures). The functional health measures are reported in each MAO's annual HOS Performance Measurement Report and on the Health Plan Management System (HPMS), which is the CMS data collection and maintenance system that houses MAO-related information, including summary level HOS results. These results are derived from a combination of case-mix adjusted PCS and MCS change scores and death status based on the baseline and two-year follow-up surveys. When the functional health measures are reported in the Medicare Star Ratings, scores for contracts with fewer than 30 respondents are not reported (CMS, 2012).

Chapter 2. Methods

For each VR-12 item, we first calculated the point value and probability of all possible beneficiary-level transitions from each response level at baseline to each level at follow-up. Separately for PCS and MCS, we developed an impact index to quantify the importance of a given transition to PCS and MCS scores. For each possible transition of a VR-12 item from baseline to follow-up, this index was defined as the product of (a) the probability of the transition and (b) the absolute value of the change in PCS or MCS points associated with the change, simply using the PCS and MCS weights. In order to identify transitions that contribute most to change in PCS and MCS scores, we designated impact index values of 0.20 or greater as notable. We then used regression models to examine the relative contributions of VR-12 change scores on individual-level *dichotomous* PCS and MCS outcomes measures; the extent to which change in a given VR-12 item was associated with change of the individual across the better than expected/same as expected/worse than expected categories. Next, we examined contract-level variation in mortality and changes in VR-12 items in order to identify those factors with the largest contract-level variation. Large contract-level variation in change scores as calculated by the intra-class correlations (ICCs) may indicate greater potential for contract-level influence. In the first three sets of analyses, we excluded beneficiaries who died after baseline. In order to anchor the effects of mortality relative to the other VR-12 items, we estimated contract-level models predicting PCS and MCS outcomes measures from the 13 rolled-up VR-12 items and mortality.

Data Sources

The Medicare HOS 2.0 was used for the *2009 Cohort 12 Baseline* and *2011 Cohort 12 Follow*-Up. The merged *2009-2011 Cohort 12 Baseline and Follow-Up* file with death status information was used for all analyses. Death status at the two-year follow up was obtained from the CMS database by RTI International and merged to the HOS data using the Medicare Health Insurance Claim (HIC) Number. The *Cohort 12 Baseline* included a sample of 487,861 beneficiaries, including both aged and disabled, from 424 MAOs. Of these beneficiaries, 59.2% (288,794) completed the baseline survey. Of this group, 250,733 were seniors (age 65 or older). During the two years between the baseline and follow-up, a number of MAOs discontinued or consolidated with other MAOs. Consequently, 382 contracts with 230,555 baseline respondents remained in the HOS; this group comprised the *2009-2011 Cohort 12 Performance Measurement* analytic sample.

Of the 230,555 senior baseline respondents, 17,936 (7.8%) died between baseline and the two-year follow-up in 2011, 70,589 (30.6%) voluntarily disenrolled from their MAO, and 142,030 (61.6%) were still alive and in the same MAO at the time of follow-up in 2011. Of the 142,030 seniors who were eligible for follow-up sampling, 34,644 (24.4%) did not respond, 1,330 (0.9%) were determined to be ineligible members at follow-up,[2] and 106,056 (75.4%) were respondents who had a follow-up PCS or MCS score from the 2011 remeasurement. The average number of respondents per MAO was 278, with a range of 0 (one MAO) to 787 respondents. The middle half of MAOs had between 157 and 389 respondents, with 10% of MAOs having 78 or fewer respondents. We excluded 579 beneficiaries because they were missing a PCS or MCS score at baseline or follow-up, leaving 105,477 respondents for whom

[2] Ineligible members at follow-up met one of the following criteria: had an incorrect address and phone number; or had a language barrier.

VR-12 item transitions could be computed. Of this group, 79,670 respondents from 381 contracts were available with complete information on the VR-12 items for the individual-level ordinary least squares (OLS) regression analyses. Death status information for the 17,926 who died among all 230,555 senior baseline respondents was used to determine the percentage that died within each MAO for the contract-level analysis.

Survey Instruments

In 2006, CMS implemented the Medicare HOS 2.0 for MAOs (NCQA, 2011). The HOS 2.0 evaluates the Health-Related Quality of Life (HRQOL) of MA beneficiaries by measuring their physical and mental health status using the VR-12 (Iqbal et al, 2007), instead of the 36-item health survey used in the HOS 1.0. The HOS 2.0 also contains questions about socio-demographics, activities of daily living (ADLs), chronic medical conditions, depression risk, and height and weight used for calculation of Body Mass Index (BMI). Four HEDIS Effectiveness of Care measures are included to evaluate management of urinary incontinence, physical activity, osteoporosis testing and fall risk management.

The VR-12 was derived from the Veterans RAND 36-Item Health Survey (VR-36) and is a generic, multipurpose health instrument, which includes the VR-36 items that best predict VR-36 PCS and MCS scores (questions Q1-Q7 in the HOS 2.0). It also includes two items that assess change in physical and emotional health compared to one year ago (Q8 and Q9) that are not used in the calculation of the summary scores. Two additional modifications were made in the VR-12 compared to the 36-item survey used in the HOS 1.0. The first modification was an increase in the number of response choices for the items used for Role-Physical (Q3a, Q3b) and Role-Emotional (Q4a, Q4b) from a dichotomous "Yes" or "No" option to a five-point Likert scale ("No, none of the time," "Yes, a little of the time," "Yes, some of the time," "Yes, most of the time," and "Yes, all of the time"). The Role-Physical questions assess whether respondents' physical health limits them in the kind of work or other usual activities they perform, whereas the Role-Emotional questions assess whether emotional problems have caused respondents to accomplish less in their work or other usual activities. The second modification was the use of the two items to assess change in health compared to one year ago; one focusing on physical health (Q8) and one on emotional problems (Q9), in contrast to the one general change item in the 36-item survey (Kazis, Lee et al, 2004; Kazis, Miller et al, 2004).

The VR-12 measures the same eight health domains as the 36-item survey: 1) Physical Functioning (PF02, PF04), 2) Role-Physical (RP2, RP3), 3) Role-Emotional (RE2, RE3), 4) Bodily Pain (BP2), 5) Social Functioning (SF2), 6) Mental Health (MH3, MH4), 7) Vitality (VT2), and 8) General Health (GH1). Each domain aggregates one or two items; all eight domains are utilized in the calculation of the two summary measures.

Variable Descriptions (Tables 1-2)

Our key predictor variables in individual-level regression models were follow-up minus baseline scores for each of the VR-12 measures. We used linear scoring of ordinal categories, and in a secondary scoring approach, we transformed these linear scores into Z-scores. The key predictor variables at the contract level were the mean VR-12 change scores and the mortality rate. Our individual-level dependent variables were each member's risk-adjusted and categorized changes in PCS and MCS scores at the two-year

follow-up. We model both the scoring distinguishing "better than expected" from "same as expected"/"worse than expected" and the scoring that distinguished "better than expected"/"same as expected" from "worse than expected." In contract-level models, the summary scores were expressed as the percentage of MA members whose two-year changes in PCS or MCS scores were "better than expected" (or "better than or same as expected").

Predictor variables. Table 1 lists the VR-12 items and the scoring of beneficiary-level baseline and follow-up PCS and MCS scores (Spiro et al., 2004). Our key predictor variables at the individual level were follow-up minus baseline scores for each of the VR-12 measures. These individual-level variables involved linear scoring of ordinal categories and a secondary scoring that transformed these linear scores into Z-scores. The key predictor variables at the contract level were the mean VR-12 change scores and the mortality rate (derived as described below).

Outcome variables. A constant was included in the calculation to standardize the PCS and MCS scores to the general U.S. population (Spiro et al., 2004), with a mean of 50 and a standard deviation (SD) of 10. In order to retain as many cases as possible (including cases with missing items), a regression model was employed for imputing and scoring missing data so that summary scores can be calculated even if as many as 9 of the 12 items are missing from the VR-12 (Spiro et al., 2004; Iqbal et al., 2007). In Table 1, the shaded cell entries are coefficients associated with the largest influence on PCS and MCS scores (magnitudes greater than 5 units, or 0.5 SD). The range of possible PCS scores is 6.2 - 72.5 and the range of possible MCS scores is 4.0 - 76.4. The items with the largest coefficients for PCS scores (in terms of absolute magnitude) were responses to the Pain item, with items in the Role-Physical, Physical Functioning, and General Health domains also having large magnitudes. The items with the largest coefficients for the MCS score were the Mental Health items, with items from Social Functioning, Role-Emotional, and Vitality also having large coefficients.

As described in Table 2, after each member's PCS and MCS scores were calculated, they were risk adjusted and categorized as "better than expected," "same as expected" or "worse than expected," depending on the difference in follow-up and baseline summary scores. Results were expressed as the percentage of MA members whose two-year changes in PCS scores are "better than expected" or the "same as expected," compared with those whose physical health is "worse than expected" or who died, and members whose two-year change in MCS scores are "better than expected" or the "same as expected," compared with those whose mental health is "worse than expected." Table 2 summarizes the intermediate steps involved in producing the contract-level PCS and MCS performance measures. For the individual-level analyses, the four dependent variables are the four individual-level dichotomous outcomes produced in step 3 (Actual PCS better, Actual PCS same or better, Actual MCS better, and Actual MCS same or better).

Table 2 also describes how the beneficiary-level derived scores were converted into contract-level measures used in the assignment of Medicare Star Ratings. The PCS and MCS change scores are risk adjusted and categorized as "better than expected," "same as expected," and "worse than expected." At the contract-level, members who died within the two-year window following the baseline interview but before they completed the follow-up survey were included in the PCS analysis; they were excluded from the MCS analysis. Results are expressed as the percentage of beneficiaries whose two-year PCS change

scores are "better than expected" or the "same as expected" compared with those who died or whose PCS change scores are "worse than expected" and as the percentage of beneficiaries whose two-year MCS change scores are "better than expected" or the "same as expected" compared with those whose MCS change scores were "worse than expected."

Analyses

Person-level transition analyses. We developed a beneficiary-level transition matrix to describe for each VR-12 item the various possible transitions between each baseline response level and all other follow-up response levels. Beneficiaries who died before follow-up were excluded from this analysis and the one that follows.

Person-level models predicting beneficiary-level dichotomous HOS outcomes from VR-12 item change scores. To estimate the relative contribution of each VR-12 change score on the beneficiary-level HOS outcomes (i.e., classification as did not improve to improved or vice versa), we estimated four beneficiary-level models (Models 1 and 2 with two subscripts). These eight models correspond to all combinations of two outcome domains (PCS and MCS), two versions of the outcome change measures (*a* in the first subscript corresponding to "Better" versus "Same" or "Worse" outcome and *b* in the first subscript corresponding to the "Same or Better" versus "Worse" outcome), and two versions of the predictors (*a* in the second subscript indicating unstandardized and *b* in the second subscript indicating standardized). Each model predicts dichotomous beneficiary-level HOS outcomes from VR-12 item change scores. Beneficiaries who die before completing follow-up HOS are excluded from these models. Although the outcomes are dichotomous, we estimated each pair of models using linear OLS regression, because OLS coefficients are easier to interpret (as changes in scores) than logit coefficients or odds ratios in this context, and at the sample sizes in the HOS data, the p-values are valid (Chen 1995).

Model 1aa: Actual PCS Better = diffPF02 + diffPF04 + diffVRP2 + diffVRP3 + diffBP2 + diffGH1 + diffVT2 + diffSF2 + diffVRE2 + diffVRE3 + diffMH3 + diffMH4

Model 1ba: Actual PCS Same or Better = diffPF02 + diffPF04 + diffVRP2 + diffVRP3 + diffBP2 + diffGH1 + diffVT2 + diffSF2 + diffVRE2 + diffVRE3 + diffMH3 + diffMH4

Model 2aa: Actual MCS Better = diffPF02 + diffPF04 + diffVRP2 + diffVRP3 + diffBP2 + diffGH1 + diffVT2 + diffSF2 + diffVRE2 + diffVRE3 + diffMH3 + diffMH4

Model 2ba: Actual MCS Same or Better = diffPF02 + diffPF04 + diffVRP2 + diffVRP3 + diffBP2 + diffGH1 + diffVT2 + diffSF2 + diffVRE2 + diffVRE3 + diffMH3 + diffMH4

Next, we investigated the variation of mortality and the VR-12 item change scores from contract to contract.

Contract-level variation in VR-12 change scores and mortality. We computed contract-level ICCs for each of the VR-12 change scores in four different ways (see below) and mortality (in a single manner). First, we used the unstandardized change scores for VR-12 items. Second, we used the standardized change scores (change of items on Z-score scales) for VR-12 items. Third, we used the change in the PCS

coefficient associated with each VR-12 item. Fourth, we used the change in the MCS coefficient associated with each VR-12 item.

Contract-level models predicting dichotomous HOS outcomes from VR-12 item mean change scores and mortality. To estimate the contribution of mortality relative to each VR-12 contract-level mean PCS change score, we estimated four contract-level models. We used the same naming convention as described in the beneficiary-level models above. These four models corresponded to all combinations of the PCS (Model 1), two versions of the outcome change measures (*a* in the first subscript corresponding to change in proportion of beneficiaries in the plan in "Better" versus "Same" or "Worse" outcome and *b* in the first subscript corresponding to change in the proportion reporting "Same or Better" versus "Worse" outcome), and two versions of the predictors (*a* in the second subscript indicating unstandardized and *b* in the second subscript indicating standardized). Each model predicts a continuous contract-level HOS outcome from VR-12 mean change score and mortality. We estimated each pair of models using linear OLS regression.

Contract-level Model 1aa: Adjusted Percent PCS Better= mdiffPF02 + mdiffPF04 + mdiffVRP2 + mdiffVRP3 + mdiffBP2 + mdiffGH1 + mdiffVT2 + mdiffSF2 + mdiffVRE2 + mdiffVRE3+ mdiffMH3 + mdiffMH4 + mdeath

Contract-level Model 1ba: Adjusted Percent PCS Same or Better= mdiffPF02 + mdiffPF04 + mdiffVRP2 + mdiffVRP3 + mdiffBP2 + mdiffGH1 + mdiffVT2 + mdiffSF2 + mdiffVRE2 + mdiffVRE3+ mdiffMH3 + mdiffMH4 + mdeath

Identification of minimum contract sample size for outlier detection. The determination of a minimum contract sample size (number of longitudinal observations) for outlier detection is a two-stage process. In the first stage, we tested if the inclusion of small contracts in the outlier model might have biased the overall model coefficients used to detect outliers in all contracts. This stage led to the development of a preliminary minimum sample size (MSS). The second stage asked if the outlier estimates for the small contracts themselves were unreliable, leading to a final recommended MSS for outlier detection. In both stages, we considered three thresholds: 0 (no minimum), 30 (the official threshold for public reporting of longitudinal measures from HOS), and 100.

In the first stage, we compared the R-squares resulting from our models used to identify outliers using each of the three candidate MSS thresholds. R-square would be expected to increase with the candidate MSS if smaller contracts (of a size between the MSSs being compared) have sufficiently low reliability (because of their sample sizes) to downwardly bias R-square. Thus, a markedly higher R-square (0.1 units or more) for the larger of two MSSs being compared may indicate that the smaller MSS is inadequate. In the second stage, we compared the number of positive and negative outliers identified in models using the three candidate MSS thresholds (0, 30, 100). Here we wanted to explore whether there was evidence that small contracts were more likely to have outlying dimensions than larger contracts, which would also indicate poor reliability for measuring outlier dimensions in the smaller contracts. Our criterion was that lower thresholds are inadequate if they have a markedly higher percentage of outliers than is observed at higher thresholds.

Contract-level modeling of outlying components. The analyses described in this section were based on the sample of MAOs that met the MSS identified through the steps above. We used contract-level regression analyses to identify specific contracts that performed particularly well or poorly on mortality or one of the individual VR-12 change scores relative to their overall performance on all 13 change scores. We ran 13 linear regression (OLS) models using the mortality indicator and individual VR-12 change scores, with each model predicting one of the 13 variables from the other 12. For example, one model predicted linear change in *How much pain interferes with normal work* score from the contract-level mortality rate and each of the other 11 VR-12 linearly scored change scores. From the output of each of 13 models, we obtained the residual score (actual score minus predicted score). We created 13 separate box-and-whisker plots of the residuals from each of the 13 models. Using Tukey's definition, we defined as an outlier any residual at least 1.5 interquartile ranges (IQRs) beyond the 25th or 75th percentile of the distribution of residuals (i.e., the residual exceeds inner fence using Tukey's box-and-whisker plots). We defined extreme outliers as those where the observed score was greater than 3 IQRs beyond the 25th or 75th percentile (i.e., the residual exceeds the outer fence using Tukey's box-and-whisker plots). These analyses allow us to identify specific contracts that are high or low outliers on each of the 13 items.

Chapter 3. Results of Analyses to Identify Key Drivers of the Physical Component Scores and Mental Component Scores

This chapter describes the analyses used to identify which VR-12 items (and mortality) have the most influence on contract-level PCS and MCS scores.

Findings

The beneficiary-level analyses of transitions found that the most influential transitions on MCS change involved *Moderate activities*, *Health interferes with social activities*, and *Felt downhearted and blue*. The most influential transitions influencing PCS change involved *Moderate activities*, *Climbing several flights of stairs*, *Limited in the kind of work or activities due to physical problems*, and *How much pain interferes with normal work.*

Individual-level regression models of VR-12 linear change scores on MCS change identified *Accomplished less than you would like to due to emotional problems* and *Felt downhearted and blue* as the VR-12 items for which a one-unit change was most likely to change a beneficiary's classified status (e.g., from *better/same* to *worse*). The most influential VR-12 items per standard deviation were *Accomplish less than you would like to due to emotional problems*, followed by *Felt calm and peaceful* and *Felt downhearted and blue.*

Individual-level regression models of VR-12 linear change scores on PCS change identified *Moderate activities*, *Climbing several flights of stairs*, and *How much pain interferes with normal work* as the VR-12 items for which a one-unit change was most likely to change a beneficiary's classified status. The most influential VR-12 items per standard deviation were *How much pain interferes with normal work* as the most influential VR-12 change score, followed by *Climbing several flights of stairs*, *Accomplished less than you would like to due to physical problems*, and *Moderate activities.* The contract-level models predicting PCS indicate that the effect of a one SD difference in mortality is approximately equivalent to the effect of a one SD difference in the amount of change in the most influential VR-12 item. Thus, per contract-level SD, mortality was about as influential to PCS as was a similar change in the most influential VR-12 item, *How much pain interferes with normal work*. The linear change contract-level models indicated that preventing one death is equivalent to preventing a 1.5 unit decline on *How much pain interferes with normal work*, the most influential item per unit change.

Twelve (of 13) ICCs associated with mortality and the VR-12 item change scores were statistically significant, with the one exception being *Felt downhearted and blue*.

Transitions that Contribute Most to Changes in PCS and MCS Scores (Table 3)

Table 3 presents the results of the individual-level transition analysis. Each row corresponds to each possible transition within each VR-12 item. Underlined cell entries correspond to transitions that involve no change in response level from baseline to follow-up.

Columns 3-4 show changes in PCS and MCS scores associated with each transition (calculated by subtraction from the values in Table 1). PCS and MCS coefficient differences that are ≤-5.0 points are shaded in pale red and the coefficients ≥5.0 points are shaded in pale green.

The impact index (columns 8 and 9) quantifies the importance of a given transition to PCS and MCS scores. We designated (bolded) values of ±0.20 or greater as notable. The most influential transitions on MCS change from baseline to follow-up involved three VR-12 items. These five transitions were: (1) the transition from "not limited at all" to "limited a little" in *Moderate activities*, (2) the transition from "none of the time" to "a little of the time" in *Health interferes with social activities*, (3) the transition from "none of the time" to "some of the time" in *Health interferes with social activities*, (4) the transition from "none of the time to "a little of the time" in *Felt downhearted and blue*, and (5) the transition from "a little of the time" to "none of the time" in *Felt downhearted and blue*.

There were 12 influential transitions on PCS change involving four VR-12 items: (1) the transition from "not limited at all" at baseline to "limited a little" at follow-up in *Moderate activities*, (2) the transition from "limited a little" to "limited a lot" in *Moderate activities*, (3) the transition from "limited a little" to "not limited at all" in *Moderate activities*, (4) from "not limited at all" to "limited a little" in *Climbing several flights of stairs*, (5) the transition from "limited a little" to "limited a lot" in *Climbing several flights of stairs*, (6) the transition from "limited a little" to "not limited at all" in *Climbing several flights of stairs*, (7) the transition from "limited a lot" to "limited a little" in *Climbing several flights of stairs*, (8) the transition from "none of the time" to "a little of the time" in *Limited in the kind of work or activities due to physical problems*, (9) the transition from "none of the time" to "some of the time" in *Limited in the kind of work or activities due to physical problems*, (10) the transition from "not at all" to "a little bit" in *How much pain interferes with normal work*, (11) the transition from "not at all" to "moderately" in *How much pain interferes with normal work*, and (12) the transition from "a little bit" to "not at all" in *How much pain interferes with normal work.*

Relative Contributions of VR-12 Item Change Scores to Dichotomous PCS and MCS Outcomes (Tables 4a-6)

The VR-12 item change scores that most influence MCS per one-unit change were *Accomplished less than you would like to due to emotional problems* and *Felt downhearted and blue*. Because a one-unit change may be more difficult/less common for some items than others, we also modeled the effect of moving one SD for each VR-12 item. The most influential VR-12 items per SD of change were: *Accomplish less than you would like to due to emotional problems*, followed by *Felt calm and peaceful* and *Felt downhearted and blue.*

The linearly scored VR-12 change scores that have the largest effects on PCS were associated with *Moderate activities*, *Climbing several flights of stairs* and *How much pain interferes with normal work;* moving just one level of improvement on these measures is associated with a 10% increase in PCS scores. When considering the effects of one SD change, *How much pain interferes with normal work* was the most influential VR-12 change score, followed by *Climbing several flights of stairs*, *Accomplished less than you would like to due to physical problems*, and *Moderate activities*. In the secondary contract-level models that allowed us to compare the effect of mortality with the effects of the VR-12 change scores, we found that the effect of an MAO's preventing one death was equivalent to preventing a 1.5 unit decline or causing a 1.5 linear unit improvement on the most important VR-12 measure for one beneficiary. The standardized model similarly indicates that the effect of one contract-level SD in mortality rate was just

slightly greater than the effect of one contract-level SD on the most influential VR-12 item, *How much pain interferes with normal work*.

Factors with the Largest Contract-Level Variation (Table 7)

Contract-level ICCs in change scores measure the proportion of variance in change scores associated with contracts, as opposed to individuals. A statistically significant ICC indicates evidence for at least some contract-level variation. An ICC of at least 0.10 indicates substantial contract-level variation. Twelve (of 13) ICCs associated with the VR-12 item change scores and mortality were statistically significant, with the one exception being *Felt downhearted and blue*. This may indicate that MAOs have some influence over nearly all VR-12 items. Eight change scores have ICCs exceeding 0.10, indicating particularly strong contract-level variation: *2-year Mortality rate*, both Role-Emotional items (*Didn't do work or other activities as carefully*, *Accomplished less than you would like to due to emotional problems*), Vitality, General Health, Physical Functioning item of *Climbing several flights of stairs*, and both Role-Physical items (*Accomplished less than you would like to due to physical problems, Limited in the kind of work or activities*).

Chapter 4. Results of Analyses to Identify Contract Outliers on VR-12 Items and Mortality

In this chapter, we describe analyses used to identify MAOs that do particularly poorly on mortality and the individual VR-12 item change scores given their performance on the remaining 12 items in order to identify specific QI opportunities (and potential strategies) for specific MAOs. We began by identifying a recommended MSS for a contract's inclusion in the outlier models.

Findings

We found that an MSS of 30 completed longitudinal surveys per MAO, the current CMS requirement for public reporting for longitudinal HOS measures, was adequate for outlier detection models (Tables 8a-9). Using this MSS, the two Role-Physical items (*Accomplished less than you would like to due to physical problems* and *Limited in the kind of work or activities*) and the two Role-Emotional items (*Didn't do work or other activities as carefully* and *Accomplished less than you would like to due to emotional problems*) were the items best predicted by other VR-12 items and *General health*, *Felt downhearted and blue*, and *2-year Mortality rate* were the items least well predicted by other VR-12 items.

For all 13 items, more contracts were negative outliers than positive outliers. The number of negative outliers ranged from 3 contracts (for *Climbing several flights of stairs*) to 20 contracts (*Felt calm and peaceful*), whereas the range of positive outliers was from 0 (for *Moderate activities*, *Accomplished less than you would like to due to physical problems*, and *Felt downhearted and blue*) to 3 (*How much pain interferes with normal work*).

Finally, we list contracts with outliers by contract and by the item (mortality or specific VR-12 item) on which they were an outlier.

Identifying Recommended MSS for Outlier Detection (Tables 8a-9)

We first compared the R-squares resulting from our models used to identify outliers. We conclude that the inclusion of small contracts (<30 beneficiaries) biases the R-squares: we found that change in five of 13 R-squares exceeded 0.10 when we compared candidate MSSs of 30 (Table 8b) to 0 (Table 8a).[3] When we compared an MSS of 100 (Table 8c) to 30, there was a change of 0.10 in just one R-square.

Second, as shown in Table 9, we found a marked decline in the number of positive *and* negative outliers between models that use an MSS of 0 and models that use an MSS of 30, but little difference between models that use 30 and those that use 100. This suggests that sample sizes below 30 are insufficiently reliable for outlier detection but that sample sizes of 30 are adequate. We recommend that for purposes of identifying contract-level outliers, users employ the current CMS-recommended MSS of 30 longitudinal completed cases per MAO. All subsequent analyses in this chapter use this recommended MSS.

[3] Among the contracts with <30 beneficiaries, one contract has no scoreable data.

Predicting Mortality and Individual VR-12 Item Changes from the Others (Table 8b; Figure 1)

Table 8b predicts Medicare Star Ratings for the item listed in each column from the items listed in each row for contracts using the MSS threshold of 30 (identified through analyses described above). The correlations among the mortality rate and individual VR-12 item change scores are mainly positive, with the strongest correlations between the two Role-Physical items (*Accomplished less than you would like to due to physical problems* and *Limited in the kind of work or activities)* and the two Role-Emotional items (*Accomplished less than you would like to due to emotional problems* and *Didn't do work or other activities as carefully*).

The R-square rows summarize how well each item is predicted from all other items. Role-Physical and Role Emotional items were the items best predicted by the other items (R-squares ranging from 0.50 to 0.64 for the four items). These four role limitation items correspond to changes in how well beneficiaries perform their work; i.e., how much they accomplish and how productively they do so. Change in one of these items tended to be associated with changes in several other items, as evidenced by the correlations, making them promising items for MAOs to target in QI efforts. The least well-predicted items were two global items, *General health* and *Felt downhearted and blue*, and the *2-year Mortality rate* (R-squares of 0.14 - 0.16 for these three measures). Consequently, change in these items tend to be unrelated to changes in other items, so that even if MAOs can influence these items, improving them is less likely to carry over to improvement in other items.

The box-and-whisker plots in Figure 1 illustrate the outlier patterns for the mortality rate and VR-12 change scores. In these plots, the rectangular box contains the middle 50% of contracts, and the line within the box corresponds to the median. The length of the middle box is the IQR, the difference between the 75^{th} and 25^{th} percentiles. The "whiskers" (the endpoints of the lines extending from the box) are the minimum and maximum mean contract change scores without outliers. The unlabeled hollow squares beyond the whiskers are outliers that are at least 1.5 times the IQR above or below the $75^{th}/25^{th}$ percentiles. The labeled hollow squares represent extreme outliers (those that are 3 times the IQR above or below the 75^{th}/25th percentiles). The units on the vertical scale are the original units for the corresponding items. Thus, a positive outlier of 0.50 units indicates a contract whose mean change score or mortality was 0.5 units higher than expected on that item, given their changes on the other items.

Contracts with Outliers (Tables 10-11)

Note that Tables 10 and 11 are removed in this report. Table 10 lists all contracts with an outlier for at least one individual VR-12 item change score or mortality among all MAOs. The contracts are listed alphabetically by name. On this table, green shading indicates a positive outlier and red shading indicates a negative outlier. There is not much evidence that contracts within the same parent organization (as inferred by the plan's name) show similar patterns of outliers. For example, of the 14 contracts for one parent organization, two are positive outliers and four are negative outliers on *How much time health interferes with social activities (SF2).* Thus, the VR-12 item change scores and mortality appear to be more within the control of the specific contract than of the parent organization.

Table 11 displays all contracts with an outlier, arranged by the item on which the contract was an outlier. These tables are useful for identifying contracts that may otherwise have average PCS or MCS scores, but which do unexpectedly well on a key measure, such as on any of the four Role- Physical (RP2, RP3) or Role-Emotional (RE2, RE3) items, for improving overall score. They are also useful for identifying contracts that could benefit from strategies to improve specific measures.

Chapter 5. Discussion

Findings and Implications

The most influential transitions on MCS change involved three VR-12 items: *Moderate activities, Health interferes with social activities,* and *Felt downhearted and blue,* and generally involved respondents exhibiting the least impaired categories. These patterns have implications for monitoring asymptomatic beneficiaries (those in good mental health, with high MCS scores) in order to detect the emergence of mental health problems and suggest that MAOs should offer easy access to high-quality self-health tools for emergent symptoms. The most influential transitions influencing PCS change involved four VR-12 items: *Moderate activities, Climbing several flights of stairs, Limited in the kind of work or activities due to physical problems,* and *How much pain interferes with normal work.*

The VR-12 items that most influenced MCS per one-unit change were *Accomplished less than you would like to due to emotional problems* and *Felt downhearted and blue*. Because a one-unit change may be more difficult/less common for some items than others, we also modeled the effect of moving one standard deviation (SD) for each VR-12 item. The most influential VR-12 items per SD were *Accomplish less than you would like to due to emotional problems,* followed by *Felt calm and peaceful* and *Felt downhearted and blue.*

The linearly scored VR-12 change scores that have the largest effects on PCS were associated with *Moderate activities, Climbing several flights of stairs* and *How much pain interferes with normal work;* moving just one level of improvement on these measures is associated with a 10% increase in PCS. When considering the effects of one SD change, *How much pain interferes with normal work* was the most influential VR-12 change score, followed by *Climbing several flights of stairs, Accomplished less than you would like to due to physical problems,* and *Moderate activities*. In the secondary contract-level models that allowed us to compare the effect of mortality with the effects of the VR-12 change scores, we found that the effect of an MAO's preventing one death was equivalent to preventing a 1.5 linear decline or causing a 1.5 linear unit improvement on the most important VR-12 measure for one beneficiary. The standardized model similarly indicates that the effect of one contract-level SD in mortality rate was just slightly greater than the effect of one contract-level SD on the most influential VR-12 item, *How much pain interferes with normal work*.

Contract-level ICCs in change scores measure the proportion of variance in change scores associated with MAOs as opposed to individual beneficiaries. A statistically significant ICC indicates evidence for at least some contract-level variation, and a value of 0.10 or greater indicates substantial contract-level variation. Twelve (of 13) ICCs associated with the VR-12 item change scores and mortality were statistically significant (all but *Felt downhearted and blue*).

Eight of the 13 items had ICCs exceeding 0.10: *2-year Mortality rate*, both Role-Emotional items (*Didn't do work or other activities as carefully, Accomplished less than you would like to due to emotional problems*), *Vitality, General health, Climbing several flights of stairs,* and both Role-Physical measures (*Accomplished less than you would like to due to physical problems* and *Limited in the kind of work or activities*). This may indicate MAOs have some influence over these items.

Finally, we examined contract-level outliers on mortality and individual VR-12 items given overall contract performance. The two Role-Physical items (*Accomplished less than you would like to due to physical problems* and *Limited in the kind of work or activities due to physical problems*) and the two Role-Emotional items (*Didn't do work or other activities as carefully* and *Accomplished less than you would like to due to emotional problems*) were the items that are best predicted by other VR-12 items, and *General health*, *Felt downhearted and blue*, and *2-year Mortality rate* were the items that were least well predicted. For all items, there are many more instances of negative outliers (a dimension far below what was expected from a contract's performance on other dimensions) than positive outliers. The number of negative outliers ranges from 3 contracts (for *Climbing several flights of stairs)* to 20 contracts (*Felt calm and peaceful*), and the range of positive outliers is 0 contracts (for *Moderate activities, Accomplished less than you would like to due to physical problems,* and *Felt downhearted and blue*) to 3 contracts *(How much pain interferes with normal work)*. Findings also suggest that efforts to improve performance on mortality and VR-12 items are likely to be more effective at the contract rather than sponsor level.

Our analysis has identified four items with particular potential for QI efforts, in that they may be both influential and subject to intervention. These QI efforts may involve initiatives that the literature suggests are effective at improving Role-Physical and Role-Emotional items. Our models predicting outliers indicate that MAOs that can improve any one of these items are likely to also positively affect other items as well.

Such QI initiatives might involve interventions to help identify members who could benefit from learning to accomplish more given their health conditions and limitations rather than, for example, focusing on changes in mental state. In other words, HOS outcomes based on PCS and MCS measures appear more driven by members' experiences of what they are doing and how well they are able to do what they do rather than on how well they are feeling. Such interventions might focus, for example, on teaching coping strategies for living with chronic physical or mental health conditions, and screening for people who can benefit from such strategies. Several interventions have been shown to improve role functioning in sub-populations of older adults (Kazis et al., 2011). For example, a "behavior change-focused weight management program" that involved a bundled set of clinical interventions improved *Role limitations due to physical problems* and *Physical functioning* among frail obese adults (Blissmer et al., 2006). An intervention that involves clinician counseling offered through "exercise prescriptions" in a sample of sedentary women found positive effects on *Role limitations due to physical problems* and *Mental health* after 12 months (Lawton et al., 2008). Such interventions, which involve cognitive-behavioral techniques such as motivational interviewing and goal setting, often positively affect other VR-12 measures.

Conclusions

The VR-12 item *Accomplished less than you would like to due to emotional problems* stands out as a major driver of MCS scores and one with evidence of possible MAO influence; as such, it may be an especially good target for QI efforts. Three VR-12 items met similar criteria with respect to improving PCS scores: *Moderate activities*, *Climbing several flights of stairs*, and *How much pain interferes with normal work*. We identified specific transitions associated with two VR-12 items (*Climbing several flights of stairs* and *Limited in kind of work because of physical problems*) that may merit particular focus.

Limitations

The HOS is restricted to the MA population; we cannot be certain if the results would apply to a more general beneficiary population without conducting analyses in a different population. All associations described in this report are from observational (rather than experimental) data, which limits our ability to draw causal inferences.

While a large variation in VR-12 change across contracts is likely to reflect MAO influence on health, it is also possible that some of the observed variance in change scores at the contract level reflects differences in the mix of patients who are more likely to decline on different measures.

Future Work

We identified several VR-12 items that multiple strands of analyses identify as particularly worthy of targeted intervention: *Accomplished less than you would like to due to emotional problems, Moderate activities, Climbing several flights of stairs,* and *How much pain interferes with normal work*. Future qualitative work could probe MAOs that are positive (and negative) outliers along these specific domains to identify possible strategies that these MAOs may be using to move these items, and, indirectly, their PCS and MCS outcomes.

Since we were able to identify contracts with outliers on mortality and specific VR-12 items, future work could use these outliers to address substantive questions. Examples of such analyses include examining MAO characteristics that are associated with outlier status to identify aggregate patterns. Future work could also investigate the year-to-year stability of outliers. If this stability is not high, this might indicate that change in a VR-12 item is unreliably measured or, alternatively, that MAOs are able to improve their lagging dimensions. Nearly one in three (30.6%) beneficiaries disenrolled from their MAO between baseline and follow-up. The current methodology for calculating PCS and MCS outcomes excludes disenrollees, so that their subsequent health is not attributed to the MAO from which they disenrolled. One research question is whether beneficiaries who receive poor care and are particularly likely to experience a decline in PCS or MCS outcomes measures are also more likely to disenroll. If so, then the current approach may fail to capture their poor care by excluding those who disenrolled because of poor care.

Finally, future work may also focus on identifying specific strategies used by MAOs to improve Role-Physical and Role-Emotional measures for members, especially members living with chronic physical or emotional conditions. Such strategies could be more widely shared with contracts that perform worse than expected based on their performance on other measures to improve overall PCS and MCS performance of MAOs. These strategies could be supplemented with an inventory of evidence-based interventions and approaches that have been shown to improve these same outcomes in older populations through a review of the published literature.

References

Chen L. Testing the mean of skewed distributions. *J American Statistical Assoc.* 1995;90:767-772.

Blissmer B, Riebe D, Dye G, Ruggiero L, Greene G, Caldwell M. Health-related quality of life following a clinical weight loss intervention among overweight and obese adults: intervention and 24 month follow-up effects. *Health Qual Life Outcomes*. 2006 Jul 17;4:43.

Centers for Medicare & Medicaid Services (CMS). October 10, 2012. Medicare Health & Drug Plan Quality and Performance Ratings 2013 Part C & Part D Technical Notes. Accessed on March 5, 2013 at www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData.html

Iqbal SU, Rogers W, Selim A, Qian S, Lee A, Ren XS, Rothendler J, Miller D, Kazis L. 2007. The Veterans RAND 12 Item Health Survey (VR-12): What it is and How it is used. Accessed March 5, 2013 at www.chqoer.research.va.gov/docs/VR12.pdf

Kazis LE, Lee A, Spiro III A, Rogers W, Ren XS, Miller DR, Selim A, Hamed A, Haffer SC. Measurement comparisons of the Medical Outcomes Study and the Veterans SF-36 Health Survey. *Health Care Financing Review*. 2004; 25(4):43-58.

Kazis LE, Miller DR, Clark JA, Skinner KM, Lee A, Ren XS, Spiro III A, Rogers WH, Ware Jr JE. Improving the response choices on the Veterans SF-36 Health Survey role functioning scales: results from the Veterans Health Study. *Journal of Ambulatory Care Management*. 2004; 27(3):263-280.

Kazis LE, Selim AJ, Qian SX, Rogers W, Tobin A, Xu P, Arya D, Rothendler J, Shapiro G. 2011. Functional Status in Older Adults: Intervention Strategies for Impacting Patient Outcomes. Technical report, Center for the Assessment of Pharmaceutical Practices, Department of Health Policy and Management, Boston University School of Public Health. Accessed March 5, 2013 at www.hosonline.org/surveys/hos/download/Functional_Status_in_Older_Adults_2011.pdf

National Committee for Quality Assurance (NCQA). *HEDIS® 2009, Volume 6: Specifications for the Medicare Health Outcomes Survey.* Washington, DC: NCQA Publication, 2009.

National Committee for Quality Assurance (NCQA). *HEDIS® 2011, Volume 6: Specifications for the Medicare Health Outcomes Survey.* Washington, DC: NCQA Publication, 2011.

Rogers WH, B Gandek, SJ Sinclair. 2004. Calculating Medical Health Outcomes Survey Performance Measurement Results. Centers for Medicare & Medicaid Services' Office of Research, Development, and Information. Accessed March 5, 2013 at www.hosonline.org/surveys/hos/download/HOS_Calculating_PM_Results.pdf

Spiro A, Rogers WR, Qian S, Kazis LE. 2004. Imputing physical and mental summary scores (PCS and MCS) for the Veterans SF-12 Health Survey in the Context of Missing Data. Technical report prepared by The Health Outcomes Technologies Program, Health Services Department, Boston University School of Public Health and The Institute for Health Outcomes and Policy, Center for Health Quality, Outcomes and Economic Research, Veterans Affairs Medical Center, Bedford, MA. Accessed March 5, 2013 at www.hosonline.org/surveys/hos/download/HOS_Veterans_12_Imputation.pdf

Table 1. Scoring Veterans VR-12 PCS and MCS

Table 1. Scoring Veterans VR-12 PCS and MCS						
VR-12 Item	**Response Choice**	**Linear Score**	**Proportion**	**Z-Score[1]**	**PCS Coefficient[2]**	**MCS Coefficient[2]**
Intercept					47.2	44.9
Moderate activities (Physical functioning)	Limited a lot	0	22.3	-1.2	0.0	0.0
	Limited a little	1	35.1	-0.1	3.2	-1.7
	Not limited at all	2	42.7	1.3	6.4	-3.4
Climbing several flights of stairs (Physical functioning)	Limited a lot	0	29.2	-1.2	0.0	0.0
	Limited a little	1	36.3	0.0	3.8	-1.9
	Not limited at all	2	34.5	1.2	6.9	-3.4
Accomplished less than you would like to (Role limitations because of physical problems)	None of the time	4	32.1	1.2	0.0	0.0
	A little of the time	3	21.6	0.8	-2.3	0.8
	Some of the time	2	24.4	0.2	-4.2	1.3
	Most of the time	1	14.2	-0.6	-5.9	1.8
	All of the time	0	7.7	-1.6	-6.5	2.1
Limited in the kind of work or activities (Role limitations because of physical problems)	None of the time	4	34.2	1.2	0.0	0.0
	A little of the time	3	20.3	0.8	-2.9	0.9
	Some of the time	2	23.2	0.2	-4.8	1.5
	Most of the time	1	13.8	-0.6	-6.3	1.9
	All of the time	0	8.6	-1.6	-6.8	2.1
How much pain interferes with normal work (Pain)	Not at all	4	31.4	1.2	0.0	0.0
	A little bit	3	27.4	0.8	-3.8	0.7
	Moderately	2	18.6	0.2	-6.9	1.3
	Quite a bit	1	17.1	-0.6	-9.7	1.8

Table 1. Scoring Veterans VR-12 PCS and MCS						
VR-12 Item	**Response Choice**	**Linear Score**	**Proportion**	**Z-Score[1]**	**PCS Coefficient[2]**	**MCS Coefficient[2]**
	Extremely	0	5.5	-1.6	-12.6	2.3
In general, would you say your health is (General health)	Excellent	4	5.8	1.3	0.0	0.0
	Very good	3	24.4	0.9	-1.4	<0.1
	Good	2	38.6	0.0	-3.2	>-0.1
	Fair	1	24.6	-0.9	-5.7	-0.2
	Poor	0	6.6	-1.3	-7.6	-0.4
Have a lot of energy (Vitality)	All of the time	5	7.5	1.3	0.0	0.0
	Most of the time	4	28.3	1.0	-0.5	-0.9
	A good bit of the time	3	16.9	0.4	-1.1	-2.0
	Some of the time	2	23.6	-0.3	-1.6	-3.3
	A little of the time	1	15.1	-0.9	-2.0	-4.7
	None of the time	0	8.5	-1.5	-2.6	-6.0
How much time health interferes with social activities (Social functioning)	All of the time	0	4.3	-1.0	0.0	0.0
	Most of the time	1	8.1	-0.8	0.2	2.1
	Some of the time	2	18.1	-0.3	0.3	5.0
	A little of the time	3	16.5	0.4	0.5	7.6
	None of the time	4	53.0	1.7	0.8	10.3
Accomplished less than you would like to (Role limitations because of emotional problems)	All of the time	0	3.5	-1.8	0.0	0.0
	Most of the time	1	7.1	-0.3	1.9	-3.9
	Some of the time	2	14.6	0.3	3.5	-7.7
	A little of the time	3	17.0	0.8	4.6	-10.3
	None of the time	4	57.9	1.0	4.5	-10.0
Didn't do work or other activities as	All of the time	0	3.6	-1.8	0.0	0.0

Table 1. Scoring Veterans VR-12 PCS and MCS						
VR-12 Item	**Response Choice**	**Linear Score**	**Proportion**	**Z-Score[1]**	**PCS Coefficient[2]**	**MCS Coefficient[2]**
carefully (Role Limitations because of emotional problems)	Most of the time	1	6.4	-0.3	1.2	-3.1
	Some of the time	2	13.6	0.4	2.2	-5.7
	A little of the time	3	16.9	0.8	2.8	-7.6
	None of the time	4	59.4	1.0	2.3	-6.7
Felt calm and peaceful (Mental Health)	All of the time	5	16.8	1.1	0.0	0.0
	Most of the time	4	43.4	0.9	0.5	-1.9
	A good bit of the time	3	13.9	0.6	1.3	-3.9
	Some of the time	2	16.6	0.1	2.1	-6.1
	A little of the time	1	6.9	-0.8	3.1	-8.2
	None of the time	0	2.4	-1.8	3.8	-9.8
Felt downhearted and blue (Mental Health)	All of the time	0	3.0	-0.9	0.0	0.0
	Most of the time	1	3.7	-0.8	-0.7	2.8
	A good bit of the time	2	4.8	-0.6	-1.8	6.2
	Some of the time	3	20.0	-0.2	-3.0	9.5
	A little of the time	4	27.0	0.7	-3.9	12.1
	None of the time	5	42.1	1.9	-4.9	14.7

Scoring: Each beneficiary starts with the baseline score and then additions or subtractions are made based on response to each VR-12 item.
Source: Appendix A of Spiro et al. 2004, at www.hosonline.org/surveys/hos/download/HOS_Veterans_12_Imputation.pdf.
[1]Z-score is the inverse standard normal transformation of the cumulative probability distribution derived from the Proportion cell entry.
[2] The shaded cells are coefficients associated with the largest influence on PCS and MCS scores, magnitude of ± 5 or more points, with red shading for items with a large negative influence and green shading for items with a large positive influence.

Table 2. Calculation of Four Beneficiary-Level and Four Contract-Level PCS and MCS Outcomes for Analyses

Table 2. Calculation of Four Beneficiary-Level and Four Contract-Level PCS and MCS Outcomes for Analyses

Beneficiary-Level Outcomes[1]

1. Derive variables used to calculate expected outcomes based on beneficiary baseline characteristics (age, age75 spline, age85 spline, female, age*female, black, Asian, Hispanic, Medicaid, telephone mode, poverty status, high school education, currently married, home ownership, 13 chronic conditions, 4 variables based on averages of number of conditions a beneficiary has within condition grouping, and a functional status scale). Include variables derived from CMS administrative data and HOS.
2. Calculate expected values for each beneficiary for five outcomes: Death, PCS Same or Better, PCS Better, MCS Same or Better, and MCS Better. The models for death are calculated separately from the models for PCS Better and PCS Same or Better.
3. Calculate actual outcomes for each beneficiary. If beneficiary died before completing follow-up survey, code Actual death=1 (else 0). If beneficiary Actual death=0 and valid baseline and follow-up PCS and MCS scores, apply mode adjustments to baseline and follow-up score. Then create the following 0/1 flags based on difference between baseline and follow-up scores.
   1a. Actual PCS Better=1 if difference in PCS scores $\geq$ 5.66, else=0.
   1b. Actual PCS Same or Better=1 if difference in PCS scores $\geq$ -5.66; else=0.
   2a. Actual MCS Better=1 if difference in MCS scores $\geq$ 6.72; else=0
   2b. Actual MCS Same or Better=1 if difference in MCS scores $\geq$ -6.72; else=0

Plan-Level Outcomes[1]

4. Calculate national averages for percentage of beneficiaries with following outcomes: Actual Death, Actual PCS Better, Actual PCS Same or Better, Actual MCS Better, and Actual MCS Same or Better.
5. Within health plans, calculate the mean expected and actual rates for Death, Actual PCS Same or Better, Actual PCS Better, Actual MCS Same or Better, and Actual MCS Better

Table 2. Calculation of Four Beneficiary-Level and Four Contract-Level PCS and MCS Outcomes for Analyses

6. Derive adjusted plan-level results and statistics:
    a. Combine plan-level death and PCS results (both actual and expected). Death and change in PCS measures are combined to create two measures of change in physical health: Alive and PCS Same or Better, and Alive and PCS Better.
    b. Calculate the following plan-level deviation scores (actual – expected):
        1a. Alive and PCS Better
        1b. Alive and PCS Same or Better
        2a. MCS Better
        2b. MCS Same or Better
    c. Calculate the adjusted percent Better, Same, and Worse for each plan
    d. Calculate the standard error for plan-level outcome variables
    e. Calculate t statistics for plan-level estimates.

7. Calculate overall F statistics for plan comparisons and identify outliers (based on comparison of observed and expected percentages)

[1]Adapted from Rogers, Gandek, and Sinclair 2004.

Table 3. Results of Beneficiary-Level Transition Analysis

Table 3. Results of Beneficiary-Level Transition Analysis								
VR-12 Item	**Baseline Response Level**	**Follow-Up Response Level**	**Score Change Associated with Transition**		**Prevalence of Transition (%)**	**Probability of Follow-up Level Conditional on Baseline Level (%)**	**PCS Impact Index**	**MCS Impact Index**
			PCS[1]	**MCS[1]**				
Moderate activities (Physical Functioning)	2	2	0.0	0.0	31.9	68.7	0.00	0.00
	1	1	0.0	0.0	20.3	57.1	0.00	0.00
	2	1	-3.2	1.7	**11.7**	**25.2**	**-0.37**	**0.20**
	0	0	0.0	0.0	11.5	64.1	0.00	0.00
	1	0	-3.2	1.7	7.8	**21.8**	**-0.25**	0.13
	1	2	3.2	-1.7	7.5	**21.0**	**0.24**	-0.13
	0	1	3.2	-1.7	4.6	**25.8**	0.15	-0.08
	2	0	-6.4	3.4	2.9	6.2	-0.18	0.10
	0	2	6.4	-3.4	1.8	10.1	0.12	-0.06
Climbing several flights of stairs (Physical Functioning)	2	2	0.0	0.0	24.2	65.0	0.00	0.00
	1	1	0.0	0.0	21.8	58.1	0.00	0.00
	0	0	0.0	0.0	17.6	69.6	0.00	0.00
	2	1	-3.1	1.5	**10.3**	**27.7**	**-0.32**	0.15
	1	0	-3.8	1.9	9.1	**24.2**	**-0.34**	0.17
	1	2	3.1	-1.5	6.6	17.7	**0.21**	-0.10
	0	1	3.8	-1.9	5.7	**22.6**	**0.22**	-0.11
	2	0	-6/9	3.4	2.7	7.3	-0.19	0.09
	0	2	6.9	-3.4	2.0	7.8	0.14	-0.07
Accomplished less than you would like to	4	4	0.0	0.0	19.2	57.1	0.00	0.00
	2	2	0.0	0.0	10.6	41.9	0.00	0.00
	3	3	0.0	0.0	8.2	34.8	0.00	0.00

Table 3. Results of Beneficiary-Level Transition Analysis								
VR-12 Item	**Baseline Response Level**	**Follow-Up Response Level**	**Score Change Associated with Transition**		**Prevalence of Transition (%)**	**Probability of Follow-up Level Conditional on Baseline Level (%)**	**PCS Impact Index**	**MCS Impact Index**
			PCS[1]	**MCS[1]**				
(Role Limitations - Physical Problems)	4	3	-2.3	0.8	7.7	**22.9**	-0.18	0.06
	3	2	-1.9	0.5	7.1	**30.0**	-0.13	0.04
	3	4	2.3	-0.8	5.5	**23.5**	0.13	-0.04
	2	3	1.9	-0.5	5.5	**21.8**	0.10	-0.03
	1	1	0.0	0.0	4.9	39.1	0.00	0.00
	2	1	-1.7	0.5	4.7	18.6	-0.08	0.02
	4	2	-4.2	1.3	4.5	13.5	-0.19	0.06
	2	4	4.2	-1.3	3.2	12.5	0.13	-0.04
	1	2	1.7	-0.5	3.1	**25.2**	0.05	-0.02
	0	0	0.0	0.0	2.3	43.5	0.00	0.00
	1	0	-0.6	0.3	2.1	17.0	-0.01	0.01
	3	1	-3.6	1.0	2.1	8.7	-0.07	0.02
	4	1	-5.9	1.8	1.5	4.3	-0.09	0.03
	0	1	0.6	-0.3	1.4	**26.4**	0.01	0.00
	1	3	3.6	-1.0	1.3	10.7	0.05	-0.01
	2	0	-2.3	0.8	1.3	5.1	-0.03	0.01
	1	4	5.9	-1.8	1.0	8.0	0.06	-0.02
	4	0	-6.5	2.1	0.7	2.1	-0.05	0.01
	3	0	-4.2	1.3	0.7	3.0	-0.03	0.01
	0	2	2.3	-0.8	0.7	12.9	0.02	-0.01
	0	4	6.5	-2.1	0.5	10.3	0.03	-0.01
	0	3	4.2	-1.3	0.4	6.9	0.02	0.00
Limited	4	4	0.0	0.0	21.4	59.8	0.00	0.00

Table 3. Results of Beneficiary-Level Transition Analysis								
VR-12 Item	**Baseline Response Level**	**Follow-Up Response Level**	**Score Change Associated with Transition**		**Prevalence of Transition (%)**	**Probability of Follow-up Level Conditional on Baseline Level (%)**	**PCS Impact Index**	**MCS Impact Index**
			PCS[1]	**MCS[1]**				
in the kind of work or activities (Role Limitations - Physical Problems)	2	2	0.0	0.0	9.9	41.1	0.00	0.00
	4	3	-2.9	0.9	7.7	**21.6**	**-0.22**	0.07
	3	3	0.0	0.0	7.6	34.4	0.00	0.00
	3	2	-1.9	0.6	6.5	**29.4**	-0.12	0.04
	3	4	2.9	-0.9	5.5	**24.7**	0.16	-0.05
	2	3	1.9	-0.6	5.1	**21.3**	0.10	-0.03
	1	1	0.0	0.0	4.7	39.1	0.00	0.00
	2	1	-1.5	0.4	4.6	19.2	-0.07	0.02
	4	2	-4.8	1.5	4.5	12.6	**-0.22**	0.07
	2	4	4.8	-1.5	3.1	12.7	0.15	-0.05
	1	2	1.5	-0.4	3.0	**24.5**	0.04	-0.01
	0	0	0.0	0.0	2.8	46.8	0.00	0.00
	1	0	-0.5	0.2	2.4	19.6	-0.01	0.00
	3	1	-3.4	1.0	1.9	8.5	-0.06	0.02
	0	1	0.5	-0.2	1.6	**26.9**	0.01	0.00
	4	1	-6.3	1.9	1.4	4.0	-0.09	0.03
	2	0	-2.0	0.6	1.4	5.7	-0.03	0.01
	1	3	3.4	-1.0	1.1	9.4	0.04	-0.01
	1	4	6.3	-1.9	0.9	7.4	0.06	-0.02
	4	0	-6.8	2.1	0.7	2.1	-0.05	0.02
	0	2	2.0	-0.6	0.7	12.0	0.01	0.00
	3	0	-3.9	1.2	0.7	3.1	-0.03	0.01
	0	4	6.8	-2.1	0.5	8.3	0.03	-0.01
	0	3	3.9	-1.2	0.4	6.0	0.01	0.00

Table 3. Results of Beneficiary-Level Transition Analysis								
VR-12 Item	**Baseline Response Level**	**Follow-Up Response Level**	**Score Change Associated with Transition**		**Prevalence of Transition (%)**	**Probability of Follow-up Level Conditional on Baseline Level (%)**	**PCS Impact Index**	**MCS Impact Index**
			PCS[1]	**MCS[1]**				
How much pain interferes with normal work (Pain)	4	4	0.0	0.0	19.8	59.8	0.00	0.00
	3	3	0.0	0.0	12.4	42.7	0.00	0.00
	4	3	-3.8	0.7	8.7	**26.1**	**-0.33**	0.06
	1	1	0.0	0.0	7.1	46.2	0.00	0.00
	3	4	3.8	-0.7	7.1	**24.3**	**0.27**	-0.05
	2	2	0.0	0.0	6.5	35.1	0.00	0.00
	3	2	-3.1	0.6	6.2	**21.2**	-0.19	0.04
	2	3	3.1	-0.6	5.0	**26.7**	0.15	-0.03
	2	1	-2.8	0.5	4.4	**23.9**	-0.12	0.02
	1	2	2.8	-0.5	3.3	**21.8**	0.09	-0.02
	3	1	-5.9	1.1	3.0	10.3	-0.18	0.03
	4	2	-6.9	1.3	3.0	9.1	**-0.21**	0.04
	2	4	6.9	-1.3	2.0	10.9	0.14	-0.03
	1	0	-2.9	0.5	2.0	13.1	-0.06	0.01
	1	3	5.9	-1.1	2.0	13.1	0.12	-0.02
	0	0	0.0	0.0	1.6	40.8	0.00	0.00
	0	1	2.9	-0.5	1.4	**36.2**	0.04	-0.01
	4	1	-9.7	1.8	1.3	4.0	-0.13	0.02
	1	4	9.7	-1.8	0.9	5.8	0.09	-0.02
	2	0	-5.7	1.0	0.6	3.5	-0.04	0.01
	0	2	5.7	-1.0	0.4	11.2	0.02	0.00
	3	0	-8.8	1.6	0.4	1.5	-0.04	0.01
	4	0	-12.6	2.3	0.3	1.0	-0.04	0.01
	0	3	8.8	-1.6	0.3	6.6	0.02	0.00

Table 3. Results of Beneficiary-Level Transition Analysis								
VR-12 Item	**Baseline Response Level**	**Follow-Up Response Level**	**Score Change Associated with Transition**		**Prevalence of Transition (%)**	**Probability of Follow-up Level Conditional on Baseline Level (%)**	**PCS Impact Index**	**MCS Impact Index**
			PCS[1]	**MCS[1]**				
	0	4	12.6	-2.3	0.2	5.3	0.03	0.00
In general, would you say your health is (General Health)	2	2	0.0	0.0	24.8	61.4	0.00	0.00
	3	3	0.0	0.0	15.3	55.6	0.00	0.00
	1	1	0.0	0.0	12.8	58.8	0.00	0.00
	3	2	-1.8	-0.2	8.7	**31.6**	-0.16	-0.02
	2	1	-2.5	-0.1	7.8	19.2	-0.19	-0.01
	2	3	1.8	0.2	6.5	16.1	0.12	0.01
	1	2	2.5	0.1	5.5	**25.2**	0.14	0.01
	4	4	0.0	0.0	2.8	44.6	0.00	0.00
	1	0	-1.9	-0.2	2.6	12.0	-0.05	-0.01
	4	3	-1.4	0.1	2.5	**40.3**	-0.04	0.00
	0	0	0.0	0.0	2.1	51.0	0.00	0.00
	3	4	1.4	-0.1	1.9	7.0	0.03	0.00
	0	1	1.9	0.2	1.6	**37.5**	0.03	0.00
	3	1	-4.3	-0.3	1.3	4.9	-0.06	0.00
	2	0	-4.4	-0.3	0.9	2.1	-0.04	0.00
	1	3	4.3	0.3	0.8	3.5	0.03	0.00
	4	2	-3.2	-0.1	0.7	11.0	-0.02	0.00
	2	4	3.2	0.1	0.5	1.1	0.01	0.00
	0	2	4.4	0.3	0.4	8.6	0.02	0.00
	3	0	-6.2	-0.5	0.2	0.9	-0.02	0.00
	4	1	-5.7	-0.2	0.2	3.2	-0.01	0.00
	1	4	5.7	0.2	0.1	0.5	0.01	0.00
	0	3	6.2	0.5	0.1	2.3	0.01	0.00

Table 3. Results of Beneficiary-Level Transition Analysis								
VR-12 Item	**Baseline Response Level**	**Follow-Up Response Level**	**Score Change Associated with Transition**		**Prevalence of Transition (%)**	**Probability of Follow-up Level Conditional on Baseline Level (%)**	**PCS Impact Index**	**MCS Impact Index**
			PCS[1]	**MCS[1]**				
	4	0	-7.6	-0.4	0.1	0.8	0.00	0.00
	0	4	7.6	0.4	0.0	0.7	0.00	0.00
Have a lot of energy (Vitality)	4	4	0.0	0.0	16.0	51.8	0.00	0.00
	2	2	0.0	0.0	9.4	40.9	0.00	0.00
	4	3	-0.6	-1.1	6.3	**20.5**	-0.04	-0.07
	3	3	0.0	0.0	5.8	31.4	0.00	0.00
	2	1	-0.4	-1.4	5.2	**22.7**	-0.02	-0.07
	3	2	-0.5	-1.3	5.2	**27.9**	-0.03	-0.07
	1	1	0.0	0.0	5.0	37.1	0.00	0.00
	3	4	0.6	1.1	4.9	**26.5**	0.03	0.05
	4	2	-1.1	-2.4	4.1	13.3	-0.05	-0.10
	1	2	0.4	1.4	3.8	**28.4**	0.02	0.05
	2	3	0.5	1.3	3.8	16.3	0.02	0.05
	5	4	-0.5	-0.9	3.3	**42.7**	-0.02	-0.03
	5	5	0.0	0.0	2.8	36.6	0.00	0.00
	0	0	0.0	0.0	2.8	44.1	0.00	0.00
	2	4	1.1	2.4	2.8	11.9	0.03	0.07
	1	0	-0.6	-1.3	2.5	19.0	-0.02	-0.03
	4	5	0.5	0.9	2.5	7.9	0.01	0.02
	0	1	0.6	1.3	1.8	**27.9**	0.01	0.02
	3	1	-0.9	-2.7	1.8	9.4	-0.02	-0.05
	2	0	-1.0	-2.7	1.5	6.7	-0.02	-0.04
	4	1	-1.5	-3.8	1.5	4.7	-0.02	-0.06
	1	3	0.9	2.7	1.0	7.6	0.01	0.03

Table 3. Results of Beneficiary-Level Transition Analysis								
VR-12 Item	**Baseline Response Level**	**Follow-Up Response Level**	**Score Change Associated with Transition**		**Prevalence of Transition (%)**	**Probability of Follow-up Level Conditional on Baseline Level (%)**	**PCS Impact Index**	**MCS Impact Index**
			PCS[1]	**MCS[1]**				
Have a lot of energy (Vitality) (continued)	0	2	1.0	2.7	1.0	15.6	0.01	0.03
	1	4	1.5	3.8	0.9	6.6	0.01	0.03
	5	3	-1.1	-2.0	0.6	7.9	-0.01	-0.01
	4	0	-2.1	-5.1	0.5	1.7	-0.01	-0.03
	5	2	-1.6	-3.3	0.5	6.8	-0.01	-0.02
	3	0	-1.5	-4.0	0.5	2.7	-0.01	-0.02
	3	5	1.1	2.0	0.4	2.1	0.00	0.01
	0	4	2.1	5.1	0.3	5.5	0.01	0.02
	2	5	1.6	3.3	0.3	1.4	0.01	0.01
	0	3	1.5	4.0	0.3	4.2	0.00	0.01
	5	1	-2.0	-4.7	0.3	3.3	-0.01	-0.01
	5	0	-2.6	-6.0	0.2	2.6	-0.01	-0.01
	0	5	2.6	6.0	0.2	2.7	0.00	0.01
	1	5	2.0	4.7	0.2	1.2	0.00	0.01
How much time health interferes with social activities (Social Functioning)	4	4	0.0	0.0	42.0	72.7	0.00	0.00
	4	3	-0.3	-2.7	7.9	13.7	-0.02	**-0.21**
	2	2	0.0	0.0	6.3	38.0	0.00	0.00
	3	4	0.3	2.7	5.8	**35.0**	0.02	0.16
	4	2	-0.5	-5.3	5.5	9.6	-0.03	**-0.29**
	3	3	0.0	0.0	4.8	28.6	0.00	0.00
	3	2	-0.2	-2.6	4.3	**25.9**	-0.01	-0.11
	2	3	0.2	2.6	3.5	**21.2**	0.01	0.09
	2	4	0.5	5.3	3.5	**20.9**	0.02	0.18
	2	1	-0.1	-2.9	2.5	14.9	0.00	-0.07

Table 3. Results of Beneficiary-Level Transition Analysis								
VR-12 Item	**Baseline Response Level**	**Follow-Up Response Level**	**Score Change Associated with Transition**		**Prevalence of Transition (%)**	**Probability of Follow-up Level Conditional on Baseline Level (%)**	**PCS Impact Index**	**MCS Impact Index**
			PCS[1]	**MCS[1]**				
	1	1	0.0	0.0	1.9	30.0	0.00	0.00
	1	2	0.1	2.9	1.8	**27.6**	0.00	0.05
	4	1	-0.6	-8.2	1.5	2.7	-0.01	-0.13
	3	1	-0.3	-5.5	1.3	7.6	0.00	-0.07
	1	4	0.6	8.2	1.0	14.9	0.01	0.08
	1	0	-0.2	-2.1	1.0	14.9	0.00	-0.02
	0	0	0.0	0.0	0.9	33.5	0.00	0.00
	2	0	-0.3	-5.0	0.8	5.0	0.00	-0.04
	1	3	0.3	5.5	0.8	12.7	0.00	0.04
	4	0	-0.8	-10.3	0.8	1.4	-0.01	-0.08
	0	1	0.2	2.1	0.6	**23.4**	0.00	0.01
	3	0	-0.5	-7.6	0.5	2.8	0.00	-0.04
	0	2	0.3	5.0	0.4	17.2	0.00	0.02
	0	4	0.8	10.3	0.4	17.0	0.00	0.05
	0	3	0.5	7.6	0.2	8.9	0.00	0.02
Accomplished less than you would like to (Role Limitations - Emotional Problems)	4	4	0.0	0.0	45.3	73.7	0.00	0.00
	4	3	0.1	-0.3	8.6	14.0	0.01	-0.03
	3	4	-0.1	0.3	7.0	**40.3**	-0.01	0.02
	3	3	0.0	0.0	5.0	29.1	0.00	0.00
	4	2	-1.0	2.3	5.0	8.1	-0.05	0.11
	2	2	0.0	0.0	4.4	32.5	0.00	0.00
	2	4	1.0	-2.3	3.7	**27.3**	0.04	-0.09
	3	2	-1.1	2.6	3.6	**21.0**	-0.04	0.09
	2	3	1.1	-2.6	3.0	**22.2**	0.03	-0.08

Table 3. Results of Beneficiary-Level Transition Analysis								
VR-12 Item	**Baseline Response Level**	**Follow-Up Response Level**	**Score Change Associated with Transition**		**Prevalence of Transition (%)**	**Probability of Follow-up Level Conditional on Baseline Level (%)**	**PCS Impact Index**	**MCS Impact Index**
			PCS[1]	**MCS[1]**				
	2	1	-1.6	3.8	1.8	13.5	-0.03	0.07
	4	1	-2.6	6.1	1.8	3.0	-0.05	0.11
	1	1	0.0	0.0	1.5	27.2	0.00	0.00
	1	2	1.6	-3.8	1.4	**25.9**	0.02	-0.06
	3	1	-2.7	6.4	1.3	7.3	-0.03	0.08
	1	4	2.6	-6.1	1.1	**20.3**	0.03	-0.07
	1	3	2.7	-6.4	0.8	14.6	0.02	-0.05
	4	0	-4.5	10.0	0.8	1.3	-0.04	0.08
	1	0	-1.9	3.9	0.7	12.1	-0.01	0.03
	0	0	0.0	0.0	0.6	30.4	0.00	0.00
	2	0	-3.5	7.7	0.6	4.5	-0.02	0.05
	0	4	4.5	-10.0	0.5	**22.2**	0.02	-0.05
	0	1	1.9	-3.9	0.4	**21.2**	0.01	-0.02
	3	0	-4.6	10.3	0.4	2.4	-0.02	0.04
	0	2	3.5	-7.7	0.3	16.0	0.01	-0.03
	0	3	4.6	-10.3	0.2	10.2	0.01	-0.02
Didn't do work or other activities as carefully (Role Limitations - Emotional Problems)	4	4	0.0	0.0	46.3	74.2	0.00	0.00
	4	3	0.5	-0.9	8.8	14.1	0.04	-0.08
	3	4	-0.5	0.9	7.0	**40.4**	-0.04	0.06
	3	3	0.0	0.0	5.1	29.6	0.00	0.00
	4	2	-0.1	1.0	4.7	7.5	0.00	0.05
	2	2	0.0	0.0	4.0	31.6	0.00	0.00
	3	2	-0.6	1.9	3.6	**20.8**	-0.02	0.07
	2	4	0.1	-1.0	3.5	**27.6**	0.00	-0.04

Table 3. Results of Beneficiary-Level Transition Analysis								
VR-12 Item	**Baseline Response Level**	**Follow-Up Response Level**	**Score Change Associated with Transition**		**Prevalence of Transition (%)**	**Probability of Follow-up Level Conditional on Baseline Level (%)**	**PCS Impact Index**	**MCS Impact Index**
			PCS[1]	**MCS[1]**				
	2	3	0.6	-1.9	2.9	**22.6**	0.02	-0.05
	2	1	-1.0	2.6	1.7	13.5	-0.02	0.05
	4	1	-1.1	3.6	1.7	2.7	-0.02	0.06
	1	1	0.0	0.0	1.3	26.0	0.00	0.00
	1	2	1.0	-2.6	1.2	**23.6**	0.01	-0.03
	1	4	1.1	-3.6	1.2	**23.3**	0.01	-0.04
	3	1	-1.6	4.5	1.2	6.8	-0.02	0.05
	4	0	-2.3	6.7	0.9	1.5	-0.02	0.06
	1	3	1.6	-4.5	0.7	14.4	0.01	-0.03
	0	0	0.0	0.0	0.7	29.3	0.00	0.00
	1	0	-1.2	3.1	0.7	12.7	-0.01	0.02
	0	4	2.3	-6.7	0.7	**29.1**	0.01	-0.04
	2	0	-2.2	5.7	0.6	4.7	-0.01	0.03
	3	0	-2.8	7.6	0.4	2.5	-0.01	0.03
	0	1	1.2	-3.1	0.4	18.3	0.00	-0.01
	0	2	2.2	-5.7	0.3	13.9	0.01	-0.02
	0	3	2.8	-7.6	0.2	9.4	0.01	-0.02
Felt calm and peaceful (Mental Health)	4	4	0.0	0.0	27.1	58.4	0.00	0.00
	5	5	0.0	0.0	8.1	46.6	0.00	0.00
	5	4	0.5	-1.9	6.9	**39.6**	0.03	-0.13
	4	5	-0.5	1.9	6.6	14.3	-0.03	0.13
	4	3	0.8	-2.0	6.3	13.6	0.05	-0.13
	3	4	-0.8	2.0	5.7	**39.6**	-0.05	0.11

Table 3. Results of Beneficiary-Level Transition Analysis								
VR-12 Item	**Baseline Response Level**	**Follow-Up Response Level**	**Score Change Associated with Transition**		**Prevalence of Transition (%)**	**Probability of Follow-up Level Conditional on Baseline Level (%)**	**PCS Impact Index**	**MCS Impact Index**
			PCS[1]	**MCS[1]**				
Felt calm and peaceful (Mental Health) (continued)	2	2	0.0	0.0	5.3	36.8	0.00	0.00
	4	2	1.6	-4.2	4.5	9.7	0.07	-0.19
	3	3	0.0	0.0	3.7	25.8	0.00	0.00
	2	4	-1.6	4.2	3.5	**24.2**	-0.06	0.15
	3	2	0.8	-2.2	3.1	**21.8**	0.03	-0.07
	2	3	-0.8	2.2	2.6	17.8	-0.02	0.06
	2	1	1.0	-2.1	2.0	13.8	0.02	-0.04
	1	2	-1.0	2.1	1.8	**32.3**	-0.02	0.04
	4	1	2.6	-6.3	1.4	3.1	0.04	-0.09
	1	1	0.0	0.0	1.4	25.3	0.00	0.00
	1	4	-2.6	6.3	1.0	18.2	-0.03	0.06
	3	1	1.8	-4.3	0.9	6.3	0.02	-0.04
	5	3	1.3	-3.9	0.9	5.2	0.01	-0.04
	5	2	2.1	-6.1	0.9	5.1	0.02	-0.05
	3	5	-1.3	3.9	0.8	5.4	-0.01	0.03
	1	3	-1.8	4.3	0.7	12.2	-0.01	0.03
	2	5	-2.1	6.1	0.6	4.4	-0.01	0.04
	4	0	3.3	-7.9	0.4	0.9	0.01	-0.03
	2	0	1.7	-3.7	0.4	2.9	0.01	-0.02
	1	0	0.7	-1.6	0.4	7.0	0.00	-0.01
	0	4	-3.3	7.9	0.4	**20.9**	-0.01	0.03
	0	2	-1.7	3.7	0.3	19.8	-0.01	0.01
	5	1	3.1	-8.2	0.3	2.0	0.01	-0.03
	0	1	-0.7	1.6	0.3	18.7	0.00	0.01
	0	0	0.0	0.0	0.3	17.0	0.00	0.00

Table 3. Results of Beneficiary-Level Transition Analysis								
VR-12 Item	**Baseline Response Level**	**Follow-Up Response Level**	**Score Change Associated with Transition**		**Prevalence of Transition (%)**	**Probability of Follow-up Level Conditional on Baseline Level (%)**	**PCS Impact Index**	**MCS Impact Index**
			PCS[1]	**MCS[1]**				
	0	5	-3.8	9.8	0.3	16.3	-0.01	0.03
	5	0	3.8	-9.8	0.3	1.6	0.01	-0.03
	1	5	-3.1	8.2	0.3	5.0	-0.01	0.02
	3	0	2.5	-5.9	0.2	1.2	0.00	-0.01
	0	3	-2.5	5.9	0.1	7.3	0.00	0.01
Felt downhearted and blue (Mental Health)	5	5	0.0	0.0	30.1	67.0	0.00	0.00
	4	4	0.0	0.0	11.7	42.2	0.00	0.00
	5	4	1.0	-2.6	8.7	19.4	0.09	**-0.23**
	4	5	-1.0	2.6	8.0	**28.8**	-0.08	**0.21**
	3	3	0.0	0.0	6.8	37.9	0.00	0.00
	4	3	0.9	-2.6	5.9	**21.3**	0.05	-0.15
	3	4	-0.9	2.6	5.3	**29.9**	-0.05	0.14
	5	3	1.9	-5.2	3.7	8.2	0.07	-0.19
	3	5	-1.9	5.2	3.0	16.7	-0.06	0.15
	3	2	1.2	-3.3	1.5	8.7	0.02	-0.05
	2	3	-1.2	3.3	1.4	**35.0**	-0.02	0.05
	0	5	-4.9	14.7	1.3	**49.7**	-0.06	0.19
	5	0	4.9	-14.7	1.2	2.7	0.06	-0.18
	4	2	2.1	-5.9	1.1	3.9	0.02	-0.06
	2	4	-2.1	5.9	0.9	**23.1**	-0.02	0.06
	3	1	2.3	-6.7	0.9	4.9	0.02	-0.06
	1	3	-2.3	6.7	0.7	**26.5**	-0.02	0.05
	2	2	0.0	0.0	0.7	17.2	0.00	0.00
	5	2	3.1	-8.5	0.7	1.5	0.02	-0.06

Table 3. Results of Beneficiary-Level Transition Analysis								
VR-12 Item	**Baseline Response Level**	**Follow-Up Response Level**	**Score Change Associated with Transition**		**Prevalence of Transition (%)**	**Probability of Follow-up Level Conditional on Baseline Level (%)**	**PCS Impact Index**	**MCS Impact Index**
			PCS[1]	**MCS[1]**				
Felt downhearted and blue (Mental Health) (continued)	4	1	3.2	-9.3	0.6	2.3	0.02	-0.06
	5	1	4.2	-11.9	0.6	1.2	0.02	-0.07
	1	4	-3.2	9.3	0.5	**20.0**	-0.02	0.05
	2	5	-3.1	8.5	0.5	11.6	-0.01	0.04
	1	1	0.0	0.0	0.5	17.1	0.00	0.00
	1	5	-4.2	11.9	0.4	16.1	-0.02	0.05
	2	1	1.1	-3.4	0.4	10.0	0.00	-0.01
	4	0	3.9	-12.1	0.4	1.5	0.02	-0.05
	0	4	-3.9	12.1	0.4	15.4	-0.02	0.05
	1	2	-1.1	3.4	0.4	13.0	0.00	0.01
	3	0	3.0	-9.5	0.3	1.9	0.01	-0.03
	0	0	0.0	0.0	0.3	11.9	0.00	0.00
	0	3	-3.0	9.5	0.3	10.1	-0.01	0.02
	0	1	-0.7	2.8	0.2	8.0	0.00	0.01
	1	0	0.7	-2.8	0.2	7.2	0.00	-0.01
	2	0	1.8	-6.2	0.1	3.0	0.00	-0.01
	0	2	-1.8	6.2	0.1	4.9	0.00	0.01

NOTE: Row entries are ordered within VR-12 item from most to least prevalent transition.

[1] The shaded cells indicate PCS and MCS coefficient differences that are ± 5 or more points, with red shading for items with a large negative influence and green shading for items with a large positive influence. Bolded values of ± 0.20 or greater for impact indices are designated as notable.

Table 4a. OLS Regressions Predicting Beneficiary-Level PCS Outcome from VR-12 Item Change in Linear Scores[1]

Actual PCS Outcome						
	Better (vs. Same/Worse) N=79,670			Better/Same (vs. Worse) N=79,670		
VR-12 Item	**Coefficient**	**Standard error**	***p*-value**	**Coefficient**	**Standard error**	***p*-value**
Moderate activities (Physical Functioning)	0.0904	0.0018	<0.0001***	**0.1241**	0.0019	<0.0001***
Climbing several flights of stairs (Physical Functioning)	**0.1077**	0.0018	<0.0001***	**0.1361**	0.0019	<0.0001***
Accomplished less than you would like to (Role Limitations - Physical Problems)	0.0549	0.0013	<0.0001***	0.0684	0.0014	<0.0001***
Limited in the kind of work or activities (Role Limitations - Physical Problems)	0.0457	0.0014	<0.0001***	0.0615	0.0015	<0.0001***
How much pain interferes with normal work (Pain)	0.0963	0.0011	<0.0001***	**0.1236**	0.0012	<0.0001***
In general, would you say your health is (General Health)	0.0472	0.0015	<0.0001***	0.0697	0.0016	<0.0001***
Have a lot of energy (Vitality)	0.0123	0.0010	<0.0001***	0.0189	0.0011	<0.0001***
How much time health interferes with social activities (Social Functioning)	-0.0004	0.0011	0.7105	0.0019	0.0012	0.1086
Accomplished less than you would like to (Role Limitations - Emotional Problems)	-0.0472	0.0014	<0.0001***	-0.0562	0.0015	<0.0001***
Didn't do work or other activities as carefully (Role Limitations - Emotional Problems)	-0.0261	0.0013	<0.0001***	-0.0347	0.0014	<0.0001***
Felt calm and peaceful (Mental Health)	-0.0235	0.0010	<0.0001***	-0.0274	0.0010	<0.0001***
Felt downhearted and blue (Mental Health)	-0.0308	0.0008	<0.0001***	-0.0397	0.0008	<0.0001***

[1] Linear change scores are calculated by subtracting baseline linear score for VR-12 item from follow-up linear score. For example, if you move from level 1 (baseline) to level 2 (follow-up), the change score would be +1.

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$

Table 4b. OLS Regressions Predicting Beneficiary-Level PCS Outcome from Standardized VR-12 Item Change Scores[1]

Actual PCS Outcome						
	Better (vs. Same/Worse) N=79,670			Better/Same (vs. Worse) N=79,670		
VR-12 Item	**Coefficient**	**Standard error**	***p*-value**	**Coefficient**	**Standard error**	***p*-value**
Moderate activities (Physical Functioning)	0.0741	0.0014	<0.0001***	**0.1013**	0.0016	<0.0001***
Climbing several flights of stairs (Physical Functioning)	0.0896	0.0014	<0.0001***	**0.1138**	0.0016	<0.0001***
Accomplished less than you would like to (Role Limitations - Physical Problems)	0.0813	0.0020	<0.0001***	**0.1015**	0.0022	<0.0001***
Limited in the kind of work or activities (Role Limitations - Physical Problems)	0.0682	0.0021	<0.0001***	0.0905	0.0023	<0.0001***
How much pain interferes with normal work (Pain)	**0.1471**	0.0018	<0.0001	**0.1850**	0.0019	<0.0001***
In general, would you say your health is (General Health)	0.0596	0.0019	<0.0001	0.0900	0.0020	<0.0001***
Have a lot of energy (Vitality)	0.0201	0.0017	<0.0001	0.0327	0.0018	<0.0001***
How much time health interferes with social activities (Social Functioning)	-0.0025	0.0012	0.0375*	-0.0021	0.0013	0.1121
Accomplished less than you would like to (Role Limitations - Emotional Problems)	-0.0885	0.0026	<0.0001	**-0.1000**	0.0028	<0.0001***
Didn't do work or other activities as carefully (Role Limitations - Emotional Problems)	-0.0436	0.0025	<0.0001	-0.0576	0.0027	<0.0001***
Felt calm and peaceful (Mental Health)	-0.0483	0.0019	<0.0001	-0.0539	0.0020	<0.0001***
Felt downhearted and blue (Mental Health)	-0.0422	0.0011	<0.0001	-0.0557	0.0011	<0.0001***

[1] Changes in standardized scores are calculated by subtracting baseline Z-score for VR-12 item from follow-up Z-score. Bolding indicates notable coefficients ± 0.10 or greater.

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$

Table 5a. OLS Regressions Predicting Beneficiary-Level MCS Outcome from VR-12 Item Change in Linear Scores[1]

Actual MCS Outcome						
	Better (vs. Same/Worse) N=79,670			Better/Same (vs. Worse) N=79,670		
VR-12 Item	**Coefficient**	**Standard error**	***p*-value**	**Coefficient**	**Standard error**	***p*-value**
Moderate activities (Physical Functioning)	-0.0511	0.0018	<0.0001***	-0.0497	0.0018	<0.0001***
Climbing several flights of stairs (Physical Functioning)	-0.0402	0.0017	<0.0001***	-0.0435	0.0018	<0.0001***
Accomplished less than you would like to (Role Limitations - Physical Problems)	-0.0140	0.0013	<0.0001***	-0.0176	0.0013	<0.0001***
Limited in the kind of work or activities (Role Limitations - Physical Problems)	-0.0181	0.0013	<0.0001***	-0.0204	0.0014	<0.0001***
How much pain interferes with normal work (Pain)	-0.0165	0.0011	<0.0001***	-0.0178	0.0011	<0.0001***
In general, would you say your health is (General Health)	-0.0064	0.0014	<0.0001***	0.0014	0.0015	0.3347
Have a lot of energy (Vitality)	0.0377	0.0010	<0.0001***	0.0367	0.0010	<0.0001***
How much time health interferes with social activities (Social Functioning)	0.0618	0.0011	<0.0001***	0.0841	0.0011	<0.0001***
Accomplished less than you would like to (Role Limitations - Emotional Problems)	0.0757	0.0013	<0.0001***	**0.1006**	0.0014	<0.0001***
Didn't do work or other activities as carefully (Role Limitations - Emotional Problems)	0.0485	0.0013	<0.0001***	0.0586	0.0013	<0.0001***
Felt calm and peaceful (Mental Health)	0.0556	0.0009	<0.0001***	0.0653	0.0010	<0.0001***
Felt downhearted and blue (Mental Health)	0.0860	0.0008	<0.0001***	0.0908	0.0008	<0.0001***

[1] Linear change scores are calculated by subtracting baseline linear score for VR-12 item from follow-up linear score. Bolding indicates notable coefficients ± 0.10 or greater.

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$

Table 5b. OLS Regressions Predicting Beneficiary-Level MCS Outcome from Standardized VR-12 Item Change Scores[1]

Actual MCS Outcome						
	Better (vs. Same/Worse) N=79,670			Better/Same (vs. Worse) N=79,670		
VR-12 Item	**Coefficient**	**Standard error**	***p*-value**	**Coefficient**	**Standard error**	***p*-value**
Moderate activities (Physical Functioning)	-0.0388	0.0014	<0.0001***	-0.0394	0.0015	<0.0001***
Climbing several flights of stairs (Physical Functioning)	-0.0318	0.0014	<0.0001***	-0.0347	0.0015	<0.0001***
Accomplished less than you would like to (Role Limitations - Physical Problems)	-0.0198	0.0020	<0.0001***	-0.0223	0.0021	<0.0001***
Limited in the kind of work or activities (Role Limitations - Physical Problems)	-0.0274	0.0021	<0.0001***	-0.0275	0.0022	<0.0001***
How much pain interferes with normal work (Pain)	-0.0260	0.0018	<0.0001***	-0.0258	0.0019	<0.0001***
In general, would you say your health is (General Health)	-0.0036	0.0019	0.0565	0.0040	0.0019	0.0379*
Have a lot of energy (Vitality)	0.0664	0.0017	<0.0001***	0.0667	0.0017	<0.0001***
How much time health interferes with social activities (Social Functioning)	0.0777	0.0012	<0.0001***	0.0989	0.0013	<0.0001***
Accomplished less than you would like to (Role Limitations - Emotional Problems)	**0.1232**	0.0026	<0.0001***	**0.1710**	0.0027	<0.0001***
Didn't do work or other activities as carefully (Role Limitations - Emotional Problems)	0.0720	0.0025	<0.0001***	0.0940	0.0026	<0.0001***
Felt calm and peaceful (Mental Health)	**0.1070**	0.0019	<0.0001***	**0.1252**	0.0020	<0.0001***
Felt downhearted and blue (Mental Health)	**0.1183**	0.0011	<0.0001***	**0.1278**	0.0011	<0.0001***

[1] Changes in standardized scores are calculated by subtracting baseline Z-score for VR-12 item from follow-up Z-score. Bolding indicates notable coefficients ± 0.10 or greater.

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$

Table 6. OLS Regressions Predicting Contract-Level Percent of Beneficiaries with Actual PCS Outcome from VR-12 Item Change in Standardized Scores[1,2] and Mortality

Actual PCS Outcome						
	Better (vs. Same/Worse) N=381			Better/Same (vs. Worse) N=381		
VR-12 Item	**Coefficient**	**Standard error**	***p*-value**	**Coefficient**	**Standard error**	***p*-value**
Moderate activities (Physical Functioning)	0.0165	0.0214	0.4402	0.0149	0.0271	0.5841
Climbing several flights of stairs (Physical Functioning)	**0.1183**	0.0207	<0.0001***	**0.1536**	0.0263	<0.0001***
Accomplished less than you would like to (Role Limitations - Physical Problems)	0.0768	0.0244	0.0018**	0.0335	0.0309	0.2801
Limited in the kind of work or activities (Role Limitations - Physical Problems)	0.0481	0.0219	0.0285*	0.0682	0.0277	0.0145*
How much pain interferes with normal work (Pain)	**0.1348**	0.0191	<0.0001***	**0.1972**	0.0242	<0.0001***
In general, would you say your health is (General Health)	**0.1852**	0.0196	<0.0001***	-0.0699	0.0248	0.0051**
Have a lot of energy (Vitality)	0.0063	0.0181	0.7276	**0.1439**	0.0229	<0.0001***
How much time health interferes with social activities (Social Functioning)	-0.0024	0.0162	0.8843	-0.0465	0.0205	0.0239*
Accomplished less than you would like to (Role Limitations - Emotional Problems)	**-0.1709**	0.0318	<0.0001***	0.0051	0.0402	0.8990
Didn't do work or other activities as carefully (Role Limitations - Emotional Problems)	0.0743	0.0206	0.0003**	-0.0635	0.0261	0.0153*
Felt calm and peaceful (Mental Health)	-0.0985	0.0217	<0.0001***	0.0015	0.0274	0.9566
Felt downhearted and blue (Mental Health)	-0.0574	0.0137	<0.0001***	-0.0507	0.0173	0.0037**
2-year Mortality Rate	-0.0836	0.0277	0.0027**	**-0.2232**	0.0351	<0.0001***

[1] Contract-level change in standardized scores is the plan-level mean of the beneficiary-level standardized change scores; the beneficiary-level standardized change score is obtained by subtracting baseline Z-score for VR-12 item from follow-up Z-score. Bolding indicates notable coefficients ± 0.10 or greater.
[2] Contract-level VR-12 change scores are the mean change in standardized scores for a given VR-12 Item.
* $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$

Table 7. Contract-Level Intra-class Correlation Coefficients for VR-12 Change Scores and Mortality Using Four Methods[1]

VR-12 Item	Linear Score	Z-Score	PCS Coefficient	MCS Coefficient
Moderate activities (Physical Functioning)	0.07	0.08	0.07	0.07
Climbing several flights of stairs (Physical Functioning)	**0.13**	**0.13**	**0.12**	**0.11**
Accomplished less than you would like to (Role Limitations - Physical Problems)	**0.12**	0.09	**0.15**	**0.15**
Limited in the kind of work or activities (Role Limitations - Physical Problems)	**0.13**	**0.11**	**0.15**	**0.15**
How much pain interferes with normal work (Pain)	0.09	**0.10**	0.08	0.08
In general, would you say your health is (General Health)	**0.13**	**0.13**	**0.13**	**0.13**
Have a lot of energy (Vitality)	**0.15**	**0.14**	**0.15**	**0.14**
How much time health interferes with social activities (Social Functioning)	0.09	0.07	0.09	0.08
Accomplished less than you would like to (Role Limitations - Emotional Problems)	**0.14**	**0.16**	**0.17**	**0.16**
Didn't do work or other activities as carefully (Role Limitations - Emotional Problems)	**0.16**	**0.16**	**0.14**	**0.15**
Felt calm and peaceful (Mental Health)	0.08	0.07	0.07	0.08
Felt downhearted and blue (Mental Health)	-0.01	-0.02	-0.01	-0.01
2-year Mortality Rate	**0.94**	n.a.	n.a.	n.a.

[1]All ICC entries in the table are statistically significant except for the four ICC entries related to the "Felt downhearted and blue (Mental Health)" item. Note that ICCs can be negative, although the negative values in this table are not statistically distinguishable from zero.
Bolding indicates notable coefficients of 0.10 or greater.

Table 8a. OLS Regressions Predicting Contract-Level Linear Score Changes in Each of 13 Variables (VR-12 Items and Mortality) from the Other 12 in the Set (N=381 Contracts)[A]

	PF02	PF04	RP2	RP3	BP2	GH1	VT2	SF2	RE2	RE3	MH3	MH4	Death
PF02	--	0.57***	-0.04	-0.07	0.51***	0.14	-0.11	0.08	-0.06	0.06	0.23*	-0.36**	0.05
PF04	0.51***	--	0.19*	0.34**	-0.21*	-0.17	0.36**	-0.19	-0.05	0.04	-0.15	0.47**	0.05
RP2	-0.02	0.08*	--	0.58***	0.01	0.21**	0.04	0.39***	0.23***	-0.04	0.06	-0.17*	0.06*
RP3	-0.02	0.11**	0.45***	--	0.16**	-0.11*	0.39***	-0.25***	-0.13**	0.15*	-0.06	-0.11	-0.04
BP2	0.12***	-0.05*	0.00	0.12**	--	-0.06	0.13*	0.26***	0.07	-0.02	0.11	0.19**	0.01
GH1	0.04	-0.06	0.17**	-0.11*	-0.08	--	0.19**	0.10	0.07	-0.10	0.06	-0.23**	-0.04
VT2	-0.02	0.06**	0.02	0.21***	0.09*	0.10**	--	0.05	-0.00	0.00	0.21***	0.00	-0.03
SF2	0.02	-0.05	0.26***	-0.22***	0.29***	0.09	0.08	--	0.05	0.07	0.28***	0.24**	-0.09**
RE2	-0.03	-0.02	0.27***	-0.21**	0.13	0.10	-0.01	0.09	--	1.07***	0.14	0.29**	0.09**
RE3	0.01	0.01	-0.03	0.11*	-0.02	-0.08	0.00	0.06	0.54***	--	-0.09	-0.30***	-0.07**
MH3	0.05*	-0.03	0.03	-0.04	0.09	0.04	0.26***	0.22***	0.06	-0.08	--	0.18**	0.04
MH4	-0.06**	0.08**	-0.07*	-0.06	0.13**	-0.12**	0.00	0.15**	0.10**	-0.20***	0.14**	--	0.01
Death	0.05	0.06	0.18*	-0.14	0.02	-0.16	-0.18	-0.39**	0.21**	-0.34**	0.22	0.05	--
R-square	*0.41*	*0.47*	*0.65*	*0.54*	*0.43*	*0.20*	*0.40*	*0.57*	*0.74*	*0.69*	*0.40*	*0.27*	*0.10*

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$

[A] Key to Table Names (columns indicate 13 models with parameter coefficients and R-square)
PF02 = Moderate activities (Physical Functioning), **PF04** = Climbing several flights of stairs (Physical Functioning),
RP2 = Accomplished less than you would like to (Role Limitations - Physical Problems),
RP3 = Limited in the kind of work or activities (Role Limitations - Physical Problems),
BP2 = How much pain interferes with normal work (Pain), **GH1** = In general, would you say your health is (General Health),
VT2 = Have a lot of energy (Vitality), **SF2** = How much time health interferes with social activities (Social Functioning),
RE2 = Accomplished less than you would like to (Role Limitations - Emotional Problems),
RE3 = Didn't do work or other activities as carefully (Role Limitations - Emotional Problems),
MH3 = Felt calm and peaceful (Mental Health), **MH4** = Felt downhearted and blue (Mental Health),
Death = 2-year Mortality rate

Table 8b. OLS Regressions Predicting Contract-Level Linear Score Changes in Each of 13 Variables (VR-12 Items and Mortality) from the Other 12 in the Set for Contracts with 30 or More Respondents (N=370 Contracts)[A]

	PF02	PF04	RP2	RP3	BP2	GH1	VT2	SF2	RE2	RE3	MH3	MH4	Death
PF02	--	0.55***	0.00	0.10	0.30**	0.11	0.16	0.01	-0.10	0.03	0.03	-0.15	0.08
PF04	0.47***	--	0.19**	0.10	-0.15	-0.06	0.18	-0.02	-0.01	0.07	-0.01	0.13	0.02
RP2	0.00	0.11**	--	0.61***	-0.10	0.04	0.35***	0.05	0.32***	-0.06	-0.05	-0.09	-0.01
RP3	0.04	0.05	0.52***	--	0.33***	-0.05	-0.06	0.12*	-0.10*	0.10	0.16*	-0.13	-0.06
BP2	0.09**	-0.06	-0.06	0.25***	--	0.08	0.14*	0.08	0.05	0.01	0.02	0.18**	0.02
GH1	0.05	-0.03	0.04	-0.05	0.11	--	0.26**	0.07	0.03	0.02	0.08	-0.00	-0.05
VT2	0.03	0.05	0.15***	-0.03	0.09*	0.13**	--	0.12**	-0.05	0.01	0.23***	0.01	0.00
SF2	0.00	-0.01	0.03	0.09*	0.08	0.05	0.17**	--	0.05	0.04	0.08	0.24**	-0.10**
RE2	-0.04	-0.01	0.31***	-0.11*	0.08	0.03	-0.10	0.08	--	0.65***	0.10	0.15	0.12***
RE3	0.01	0.03	-0.04	0.09	0.02	0.01	0.02	0.04	0.50***	--	0.07	0.05	0.04
MH3	0.01	-0.00	-0.02	0.08*	0.01	0.04	0.23***	0.05	0.05	0.04	--	0.14**	0.00
MH4	-0.03	0.03	-0.04	-0.07	0.12**	-0.00	0.01	0.16**	0.07	0.03	0.14**	--	-0.03
Death	0.11	0.03	-0.01	-0.18	0.09	-0.14	0.00	-0.39**	0.34***	0.14	0.03	-0.18	--
R-square	*0.43*	*0.44*	*0.64*	*0.60*	*0.32*	*0.14*	*0.39*	*0.30*	*0.57*	*0.50*	*0.25*	*0.16*	*0.16*

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$

[A] Key to Table Names (columns indicate 13 models with parameter coefficients and R-square)

PF02 = Moderate activities (Physical Functioning), **PF04** = Climbing several flights of stairs (Physical Functioning),
RP2 = Accomplished less than you would like to (Role Limitations - Physical Problems),
RP3 = Limited in the kind of work or activities (Role Limitations - Physical Problems),
BP2 = How much pain interferes with normal work (Pain), **GH1** = In general, would you say your health is (General Health),
VT2 = Have a lot of energy (Vitality), **SF2** = How much time health interferes with social activities (Social Functioning),
RE2 = Accomplished less than you would like to (Role Limitations - Emotional Problems),
RE3 = Didn't do work or other activities as carefully (Role Limitations - Emotional Problems),
MH3 = Felt calm and peaceful (Mental Health), **MH4** = Felt downhearted and blue (Mental Health),
Death = 2-year Mortality rate

Table 8c. OLS Regressions Predicting Contract-Level Linear Score Changes in Each of 13 Variables (VR-12 Items and Mortality) from the Other 12 in the Set for Contracts with 100 or More Respondents (N=328 Contracts)[A]

	PF02	PF04	RP2	RP3	BP2	GH1	VT2	SF2	RE2	RE3	MH3	MH4	Death
PF02	--	0.51***	0.03	0.04	0.20*	0.13	0.19	0.08	-0.11	0.05	-0.17	-0.10	0.00
PF04	0.47***	--	0.19**	0.02	0.09	-0.13	-0.03	-0.03	0.07	0.07	0.02	0.03	0.03
RP2	0.02	0.12**	--	0.58***	0.08	0.04	0.17*	0.09	0.30***	-0.02	-0.11	0.04	-0.01
RP3	0.02	0.01	0.47***	--	0.22***	0.02	0.08	0.06	-0.13*	0.17**	0.09	-0.07	0.00
BP2	0.09*	0.04	0.06	0.21***	--	0.10*	0.15*	0.09	-0.02	0.02	0.15*	0.02	0.03
GH1	0.08	-0.08	0.04	0.02	0.14*	--	0.34***	0.08	0.06	-0.04	-0.17*	0.09	0.03
VT2	0.06	-0.01	0.09*	0.05	0.10*	0.17***	--	0.09	0.07	-0.03	0.32***	-0.15*	0.01
SF2	0.03	-0.01	0.05	0.05	0.07	0.05	0.11	--	-0.04	0.08	0.11	0.20**	-0.02
RE2	-0.05	0.04	0.25***	-0.13*	-0.02	0.05	0.11	-0.05	--	0.59***	0.12	0.24**	0.01
RE3	0.02	0.03	-0.01	0.15**	0.02	-0.03	-0.04	0.10	0.53***	--	0.11	-0.03	-0.01
MH3	-0.05	0.01	-0.05	0.05	0.10*	-0.08*	0.29***	0.09	0.07	0.07	--	0.18**	0.00
MH4	-0.02	0.01	0.01	-0.03	0.01	0.03	-0.11*	0.13**	0.11**	-0.01	0.15**	--	0.00
Death	0.01	0.11	-0.07	0.04	0.22	0.18	0.07	-0.25	0.10	-0.07	0.07	0.05	--
R-square	*0.41*	*0.41*	*0.62*	*0.58*	*0.41*	*0.20*	*0.43*	*0.23*	*0.55*	*0.52*	*0.31*	*0.14*	*0.03*

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$

[A] Key to Table Names (columns indicate 13 models with parameter coefficients and R-square)

PF02 = Moderate activities (Physical Functioning), **PF04** = Climbing several flights of stairs (Physical Functioning),
RP2 = Accomplished less than you would like to (Role Limitations - Physical Problems),
RP3 = Limited in the kind of work or activities (Role Limitations - Physical Problems),
BP2 = How much pain interferes with normal work (Pain), **GH1** = In general, would you say your health is (General Health),
VT2 = Have a lot of energy (Vitality), **SF2** = How much time health interferes with social activities (Social Functioning),
RE2 = Accomplished less than you would like to (Role Limitations - Emotional Problems),
RE3 = Didn't do work or other activities as carefully (Role Limitations - Emotional Problems),
MH3 = Felt calm and peaceful (Mental Health), **MH4** = Felt downhearted and blue (Mental Health),
Death = 2-year Mortality rate

Table 9. Low, Non-, and High Outliers from OLS Regressions Predicting Contract-Level Linear Score Changes in VR-12 Items, Contracts with No Exclusions, Contracts with 30 or More Respondents, and Contracts with 100 or More Respondents

VR-12 Item	% Low outliers (n)			% Non-outliers (n)			% High outliers (n)		
	All	30+	100+	All	30+	100+	All	30+	100+
Moderate activities (Physical Functioning)	2.9% (11)	2.4% (9)	1.5% (5)	97.4% (371)	97.6% (361)	98.5% (323)	0% (0)	0.0% (0)	0% (0)
Climbing several flights of stairs (Physical Functioning)	1.3% (5)	0.8% (3)	1.8% (6)	98.7% (377)	98.9% (366)	98.2% (322)	0% (0)	0.3% (1)	0% (0)
Accomplished less than you would like to (Role Limitations - Physical Problems)	4.5% (17)	3.2% (12)	1.5% (5)	94.5% (361)	96.8% (358)	98.2% (322)	1.0% (4)	0.0% (0)	0.3% (1)
Limited in the kind of work or activities (Role Limitations - Physical Problems)	1.8% (7)	3.0% (11)	2.4% (8)	97.4% (372)	96.5% (357)	97.5% (320)	0.7% (3)	0.5% (2)	0% (0)
How much pain interferes with normal work (Pain)	2.4% (9)	3.5% (13)	0.6% (2)	95.8% (366)	95.7% (354)	99.1% (325)	1.8% (7)	0.8% (3)	0.3% (1)
In general, would you say your health is (General Health)	4.7% (18)	3.2% (12)	1.2% (4)	94.0% (359)	96.5% (357)	99.8% (324)	1.3% (5)	0.3% (1)	0% (0)
Have a lot of energy (Vitality)	3.4% (13)	3.0% (11)	1.5% (5)	95.3% (364)	96.5% (357)	98.5% (323)	1.3% (5)	0.5% (2)	0% (0)
How much time health interferes with social activities (Social Functioning)	4.5% (17)	3.0% (11)	2.1% (7)	95.3% (364)	96.8% (358)	97.5% (320)	0.3% (1)	0.3% (1)	0.3% (1)
Accomplished less than you would like to (Role Limitations - Emotional Problems)	3.1% (12)	2.4% (9)	2.4% (8)	96.6% (369)	97.3% (360)	97.3% (319)	0.3% (1)	0.3% (1)	0.3% (1)
Didn't do work or other activities as carefully (Role Limitations - Emotional Problems)	5.0% (19)	2.7% (10)	2.4% (8)	94.0% (359)	97.0% (359)	97.5% (320)	1.0% (4)	0.3% (1)	0% (0)
Felt calm and peaceful (Mental Health)	5.5% (21)	5.4% (20)	2.4% (8)	94.0% (359)	94.1% (348)	97.5% (320)	0.5% (2)	0.5% (2)	0% (0)
Felt downhearted and blue (Mental Health)	2.4% (9)	2.2% (8)	2.1% (7)	97.4% (372)	97.8% (362)	97.9% (321)	0.3% (1)	0.0% (0)	0% (0)

Figure 1. Residual Analysis Results from OLS Regressions Predicting Contract-Level Linear Score Changes in Each of 13 Variables (VR-12 Items and Mortality) from the Other 12 in the Set limited to Contracts At Recommended MSS of 30 or more Members (N=370 Contracts)