

A Case Study:

**Magellan  
Health Services**

**Using the SF-BH  
Assessment to  
Measure Success  
and Prove Value**



## The Background

Almost four years ago, Magellan Health Services, a specialty healthcare manager focused on some of today's most complex and costly healthcare services, sought to better understand the impact behavioral health services have on consumers. Magellan's goals were to ensure the highest level of care, improve communication between individuals and healthcare providers, and measure success.

Magellan anticipated a major shift in healthcare policy toward demonstrating treatment effectiveness. This prescient decision to develop a comprehensive health outcomes measurement program now places Magellan at the forefront of a movement to instill more accountability and transparency into the healthcare system. This trend is evident in the recently passed American Recovery and Reinvestment Act, in which \$1.1 billion was set aside for comparative effectiveness research.

**“The SF-12 is a shorter version of the SF-36® Health Survey – the most widely used patient-reported outcome measure in the world – and uses just 12 questions to measure functional health and well-being”**

Mark Kosinski, MA  
Vice President and Senior Scientist  
QualityMetric

## Getting Started

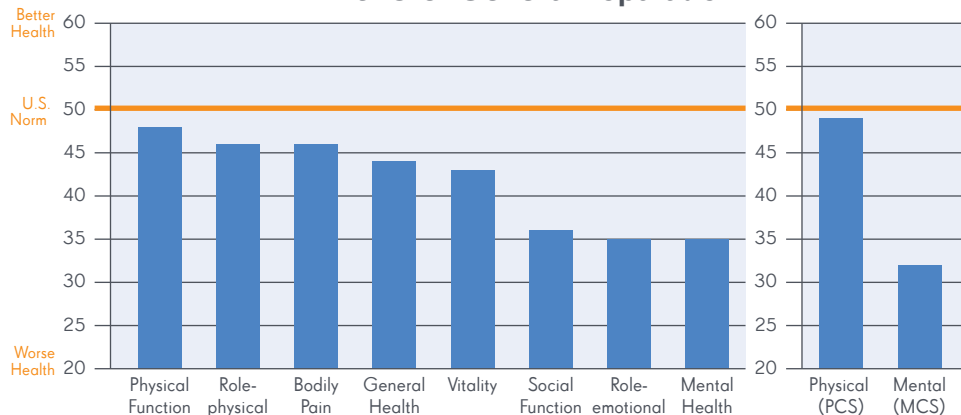
Magellan joined forces with QualityMetric Incorporated, a health outcomes measurement company, to create the SF-BH™ Assessment. The SF-BH is a self-reported assessment that measures “not only overall physical and mental health, but specific behavioral health concepts as well,” according to Joann Albright, PhD, Senior Vice President of Quality, Outcomes, and Research at Magellan. Dr. Albright led the effort to develop the SF-BH at Magellan.

The SF-BH stands for Short Form Behavioral Health Assessment. At its core is QualityMetric's SF-12® Health Survey, plus additional behavioral health questions. “The SF-12 is a shorter version of the SF-36® Health Survey – the most widely used patient-reported outcome measure in the world – and uses just 12 questions to measure functional health and well-being,” explains Mark Kosinski, MA, Vice President and Senior Scientist at QualityMetric. “It is a practical, reliable, and valid measure that is particularly useful in large populations.”

Magellan consumers complete the assessment online at baseline (upon referral) and at follow-up appointments. The technology used to administer, collect, score, and report on the SF-BH is QualityMetric's Smart Measurement® System. The data is collected and scored automatically through this technology, and a variety of customized reports are then generated for Magellan. These reports are also sent via email directly to the individual completing the assessment and the healthcare provider. The SF-BH was developed in U.S. English and U.S. Spanish to address the needs of Magellan's population.

Figure 1

### Magellan Patients Total Sample (n=33,362) vs. U.S. General Population



## The Results

For this analysis, QualityMetric used data from 33,362 assessments completed over the last three years. Of this population, 9,661 individuals had one follow-up administration of the assessment and 1,260 individuals had two follow-up administrations. The average number of days between the first and second assessment was 63, while the average number of days between the second and third assessments was 128. There were additional consumers who had more than three assessments, however, the sample size was too small to conduct meaningful analyses of outcomes.

Two separate approaches were taken in this analysis. First, we evaluated the health burden of the Magellan population. That is, the health status of this population was compared to the U.S. general population average to determine if Magellan's population was more or less healthy than the norm. Second, QualityMetric analyzed the health outcomes of the Magellan population, or how health improved or declined as a result of treatment intervention. Finally, we offered analysis to help place these results in context and help understand the outcomes from a return on investment perspective.

### Magellan Population's Health Burden

This analysis compared SF-12 scores between Magellan consumers and the U.S. general population norm. The SF-12 norms come from a nationally representative sample of the 1998 U.S. general population.<sup>1</sup> As can be seen in Figure 1, the greatest burden was observed in those dimensions of health most related to mental health status. In the total population (n=33,362), Magellan consumers scored significantly lower than the general population norm on the vitality, social functioning, role-emotional, mental health, and mental summary scales. On each of these scales the differences in scores were in the large effect size range (>0.8 SD's).<sup>2</sup>

The same analysis was also conducted on Samples 2 (n=9,661) and 3 (n=1,260). Sample 2 scores on the physical functioning, role-physical, bodily pain, and physical summary scales were below the general population norms, however, the differences reached statistical significance on just the role-physical scale. Sample 3 scores were significantly lower than the general population norms on all eight scales, as well as both the physical and mental summary scales.

To further illustrate this health burden, we examined the content of the SF-12 questionnaire items showing the greatest difference from the U.S. general population norm.

Figure 2 shows the percentage of Magellan consumers (n=33,362)

reporting having little or no energy during the past four weeks was more than double the percentage observed in the general population (50.1% vs. 23.1%). The percentage of individuals who reported being limited in social activities all or most of the time in the past four weeks was nearly five times higher than the general population (34.1% vs. 7.1%). The percentage of individuals who reported accomplishing less at work all or most of the time in the past four weeks was nearly four times higher than the general population (35.9% vs. 8.7%). Lastly, the percentage of Magellan consumers who reported being downhearted and depressed all or most of the time in the past four weeks was seven times higher than the general population (44.6% vs. 6.2%).

### Magellan Population's Health Outcomes

Changes in health status were investigated in several ways. First, QualityMetric computed mean changes in scores from the initial (visit 1) to the last administration (visit 2 or visit 3) of the SF-BH. Significant improvement was seen in seven of the eight SF-12 scales and the mental summary scale (see Figure 3). The greatest improvement was observed in scales most related to mental health (vitality, social functioning, role-emotional, mental health, and mental summary scales). The improvement observed on these scales ranged from five to 11 points. Score changes of this magnitude are in the moderate to large effect size range. Significant improvement was observed in the role-physical and bodily pain scales, however, the effect size was small.

Second, because mean changes in scores can mask the underlying variability of outcomes in the population, we evaluated categories of change scores on the physical summary scale (PCS) and mental summary scale (MCS). Specifically, using the minimal important difference threshold score for PCS and MCS (3.5 points), we categorized individuals as getting better if their PCS/MCS scores at the final visit were 3.5 points higher than their initial scores, worse if their final PCS/MCS scores were 3.5 points lower than their initial score, and the same if their final PCS/MCS scores were within +/-3.5 points of their initial score. Lastly, among Magellan consumers with three assessments, we evaluated the trends in PCS/MCS scores over time.

Next, we categorized the physical and mental health outcomes into the proportion of individuals who got better, got worse, or stayed the same. The majority of individuals (69%) improved in mental health status by a clinically meaningful amount, compared to 13% who got worse, and 18% who stayed the same.

Figure 2

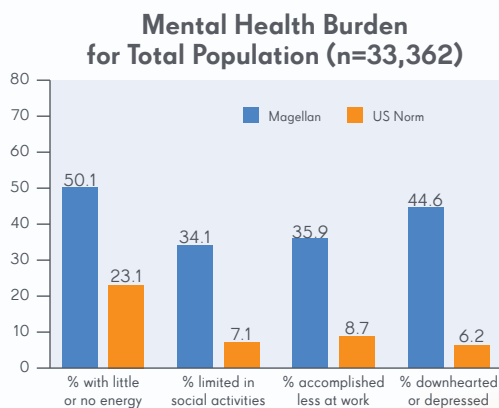
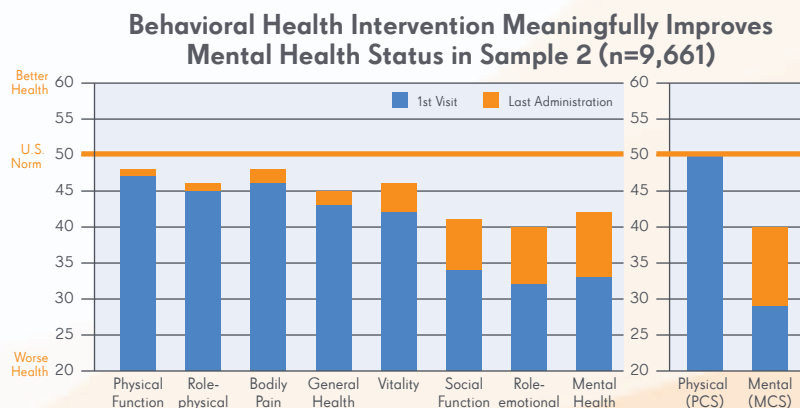


Figure 3



Content-based interpretations were applied to provide meaningful interpretations of the mental health outcomes observed in Sample 2 (see Figure 4). As shown, the percentage of Magellan consumers who reported having little or no energy all or most of the time during the past four weeks dropped from 56.5% to 33.8%. The percentage of individuals who reported limitations in social activities dropped from 41.8% to 17.1%, and the percentage of individuals who reported accomplishing less at work all or most of the time dropped from 42.2% to 17.2%. Lastly, the percentage of consumers who reported being downhearted and depressed all or most of the time during the past four weeks dropped from 52.8% to 16.3%.

Important consequences of the mental health outcomes observed in Sample 2 were significant reductions in the percentage of consumers who screened positive for depression, a significant reduction in predicted total annual medical expenses, and a significant reduction in reported missed work days.

In addition, QualityMetric identified that 83.3% of Magellan's population screened positive for depression at the initial visit, using applied published interpretation guidelines for the mental summary scale.<sup>1</sup> At the second visit, this percentage dropped to 55.3%. There was an even greater reduction in Sample 3: from 93.4% to 37.3%.

## Return on Investment

An estimate of the reduction in annual medical expenses is presented in Figure 5. Using an algorithm tested and validated by QualityMetric and used by the federal government's Agency for Healthcare Research and Quality (AHRQ), QualityMetric was able to translate PCS and MCS scores into predicted medical expenses.<sup>3</sup> The predicted medical expenses estimated with initial visit scores were compared against the predicted medical expenses estimated with final visit scores. The improvement in reported health status for the Magellan population (N=9,661) theoretically translates into an annual cost savings of around \$300 per individual for an annual savings of more than \$3 million in the measurement cohort.

There was also a significant reduction in the mean number of missed work days reported by Magellan consumers. At the initial visit the mean number of missed work days reported was 3.3 days in the past four weeks. This was reduced to a mean of 1.7 days in the past four weeks at the follow-up assessment. For Sample 3, the reduction was 4.7 to 0.9 days in the past four weeks.

## Why It Works

Through the use of the SF-BH, Magellan is able to demonstrate the value of its services to payors, consumers, and providers. The customized reports allow Magellan to share results with both consumers and providers to improve communication, assist in self-care management, and improve program transparency.

Measuring these concepts has made it possible for Magellan to obtain a comprehensive understanding of the health of their population – vital information not measurable previously. Results of the data are also ideal for use in comparative effectiveness research or analysis. Magellan can now answer vital questions about their programs, such as how the mental health burden of their population compares to that of the U.S. general population, whether interventions are resulting in positive or negative outcomes, and even how one approach compares to another.

## Moving Forward

Magellan and QualityMetric have been successfully working together now for well over three years and plan to continue working together. Magellan still uses the SF-BH and plans to expand both the frequency and type of use in other programs. In an effort to increase compliance and reduce the burden to those consumers who are unable to complete the SF-BH online, Magellan's first item of expansion will be the June 2009 implementation of a fax mode of administration. "Increasingly, consumers and providers recognize the value of the SF-BH," says Dr. Albright, "The field of health outcomes measurement is evolving and growing, and we definitely look forward to future ventures with QualityMetric."

**"Increasingly, consumers and providers recognize the value of the SF-BH"**

Joann Albright, PhD  
Senior Vice President of  
Quality, Outcomes, and Research  
Magellan Health Services

Figure 4

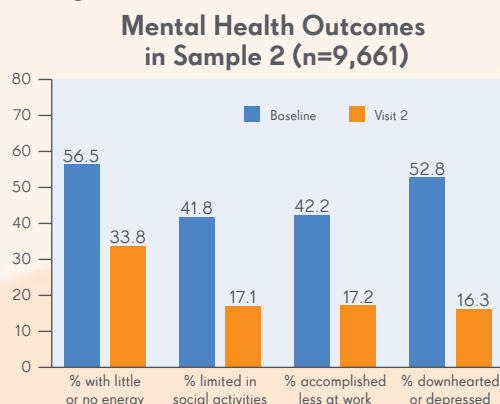
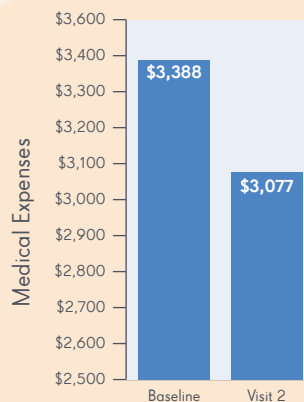


Figure 5

### Decreased Medical Expenses in Sample 2 (n=9,661)



## About QualityMetric Incorporated

QualityMetric works with the world's largest healthcare and life sciences companies to measure health outcomes. Our products help market new drugs and biologics, track health improvement or decline, screen populations for disease, and identify future health risks. QualityMetric's patient-reported outcome (PRO) surveys provide scientifically valid assessments of both physical and mental health. Our PROs include the SF-36v2®, SF-12v2®, and SF-8™ Health Surveys, SF-10™ Health Survey for Children, Asthma Control Test (ACT™), Pain Impact Questionnaire (PIQ-6™), and DYNHA® Dynamic Health Assessments. Our Outcomes Insight Consulting™ Division can develop customized PROs and provide in-depth analysis of the results. QualityMetric's founder and Chief Science Officer, John E. Ware, Jr., PhD is a thought-leader in the field of psychometrics and health outcome measurement. Visit us at [www.qualitymetric.com](http://www.qualitymetric.com).

### References

1. Ware JE, Kosinski M, Turner-Bowker DM, Gandek B. How to Score Version 2 of the SF-12 Health Survey. Second Edition. 2004. QualityMetric Incorporated. Ref Type: Serial (Book, Monograph).
2. Fleischman JA, Cohen JW, Manning W, Kosinski M. Using the SF-12 Health Status Measure to Improve Predictions of Medical Expenditures. *Med Care* 2006; 44(5):54-63.
3. Cohen J. Statistical power for the behavioral sciences. 1988.