Specialized Interpretation Analysis Report

Prepared for ACME Company

September 16, 2010

QualityMetric

The information provided in this report includes fictitious data for demonstration purposes only.

SAMPLE

Contents

Background 3
Baseline Burden of Disease Analysis 4
Treatment Efficacy Analysis 6
Norm-based Interpretation of Treatment Effects 7
Categorical Change Analysis: Treatment Responders 8
Criterion-based Interpretation of Change Scores 9
Content-based Interpretation 10
Depression Screening 12
Appendix 13

Table 1. Baseline burden of disease
Table 2. Baseline burden of disease relative to age and gender
Table 3. Change in SF-36v2 subscales and summary scores
Table 4. Content-based interpretation
Figure 1. Norm-based interpretation of treatment

SAMPLE

Background

This report presents results of specialized interpretation of data from the SF-36v2® Health Survey. The SF-36v2 is a self-report measure of health-related quality of life (HRQOL).* Responses to items are computed into an eight-domain profile of scores: Physical Functioning (PF), Role Limitations-Physical (RP), Bodily Pain (BP), General Health (GH), Vitality (VT), Social Functioning (SF), Role Limitations-Emotional (RE), and Mental Health (MH). Additionally, physical and mental component summary scores (PCS and MCS, respectively) are generated.

The standardization of mean scores and standard deviations for all SF-36v2 scales was accomplished using norm-based scoring (NBS). NBS applies a linear T-score transformation (Mean = 50, SD=10), which makes it possible to meaningfully compare scores for the eight-scale profile and the physical and mental summary measures across samples. General U.S. adult population statistics were used in this standardization of the SF-36v2 scores.

The results presented in this report are from a study sponsored by ACME examining the effect of Drug X treatment for <<CONDITION>>.

This study includes the following treatment arms:

1. Drug X (Target treatment)
2. Placebo

This report presents results from seven specialized methods for data analysis and interpretation. Each analysis is described individually on the following pages, with graphical or tabular displays and text explaining each methodology and results.

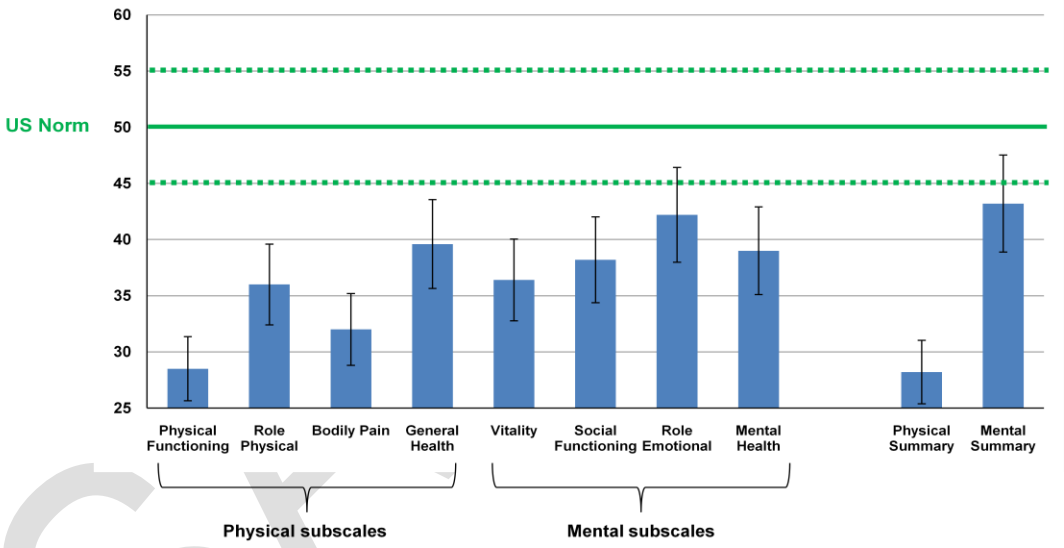
SAMPLE

* Ware J.E., Jr., Kosinski M., Bjorner J.B., Turner-Bowker, D.M., Gandek, B. and Maruish, M.E. (2007) User's Manual for the SF-36v2™ Health Survey (2nd Ed). Lincoln, RI, QualityMetric Incorporated.

Baseline Burden of Disease Analysis

1. Rationale and Method
   - This analysis assessed the functional burden associated with the patient population by comparing baseline SF-36v2 scores of study participants with general U.S. population norms
   - Comparison of the trial samples' baseline SF-36v2 scores with a benchmark sample of patients with kidney disease were also conducted (**up to two** disease-specific benchmark samples will be analyzed; the samples used will be agreed upon in conversations between ACME and QualityMetric staff).
   - Regression methods were used to adjust U.S. general population norms and disease-specific benchmark samples to the age and gender distribution of the study sample.
   - Statistical comparisons were used to test for significant differences between the study sample and each comparison sample for all scales and summary scores, while effect sizes for standardized differences between samples were calculated for all comparisons to indicate the magnitude of impact of the disease on health-related quality of life (HRQOL) scores.

2. Results: Baseline burden of disease as compared to an age- and gender-matched U.S. general population normative sample.

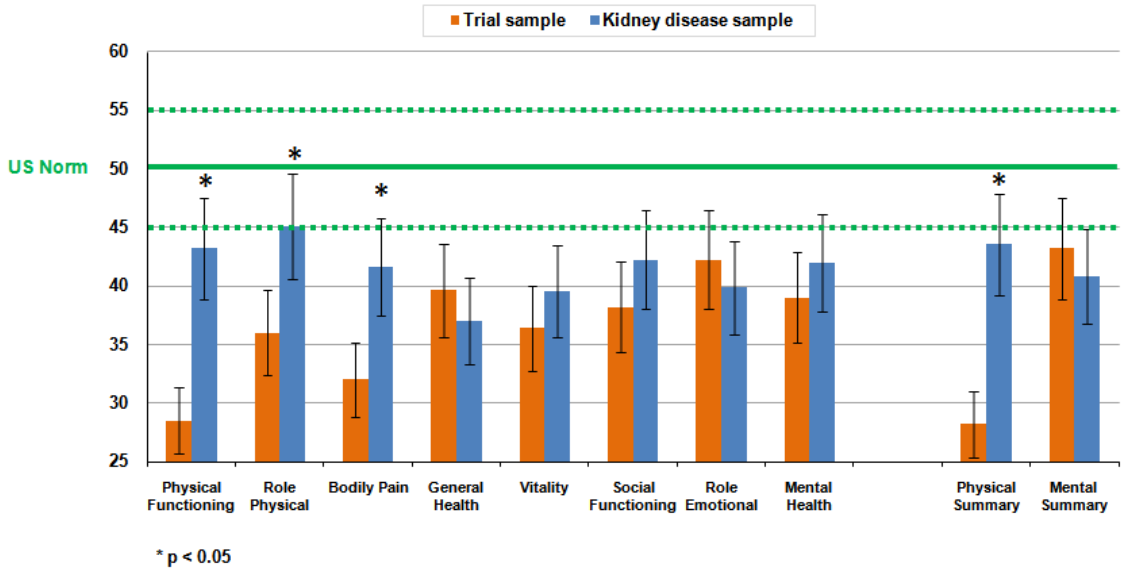


3. Interpretation
   - The burden of disease at baseline is generally largest in scales assessing physical health concepts.
   - The greatest decrement compared to U.S. general population norms is seen in the Physical Functioning score, which is more than 2 standard deviations lower than the U.S. Population average for that scale.
   - Other scales showing clinically meaningful impairment are those measuring Bodily Pain, Role Functioning problems due to physical health, Vitality, General Health, and Social Functioning. Fewer limitations in Mental Health and Role Functioning problems due to emotional health were observed.

Copyright 2010 QualityMetric Incorporated

- This pattern is clearly reflected in the summary measures, with the Physical Component Summary (PCS) showing a major decrement from the population average, and the Mental Component Summary (MCS) much closer to normal.
- Descriptive statistics, inferential statistics, and effect sizes for mean differences between the trial and normative samples are presented in **Table 1** in the appendix.
- Conclusion: the burden of disease is clearly concentrated in self-rated health concepts related to physical aspects of HRQOL.

4. Results: Baseline burden of disease relative to a sample of age- and gender-matched patients with kidney disease.



5. Interpretation

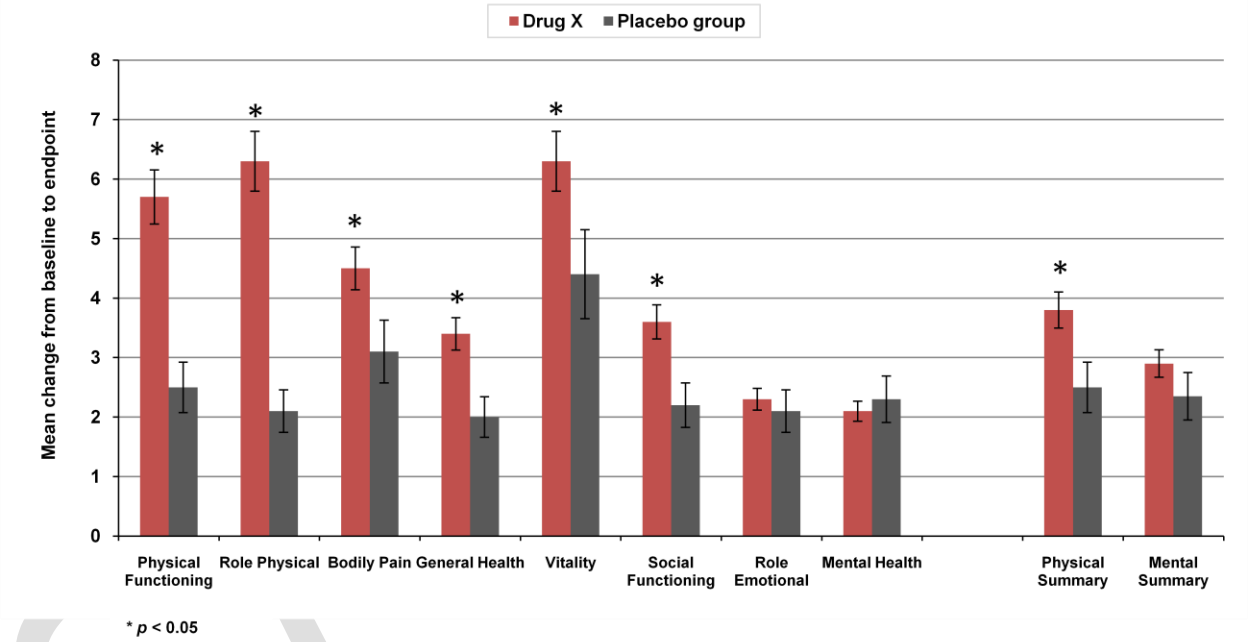
- The relative burden of disease at baseline for the trial sample as compared to the kidney disease patient sample is largest for scales assessing physical health outcomes: scores for the trial sample were significantly lower than for the kidney disease patients on Physical Functioning, Role Physical, and Bodily Pain scales.
- There were no differences between disease groups for scales measuring mental health concepts.
- The patterns of subscales is reflected in the summary measures, with the trial sample having much lower PCS scores than the kidney disease sample, but both groups have statistically comparable MCS scores.
- Descriptive statistics, inferential statistics, and effect sizes for mean differences between the trial and kidney disease samples are presented in **Table 2** in the appendix.
- Conclusion: the relative burden of disease for the trial sample as compared to patients with kidney disease is concentrated in physical health concepts.

Copyright 2010 QualityMetric Incorporated

Treatment Efficacy Analysis

1. Rationale and Method
   - This analysis examined the statistical significance and clinical meaningfulness of changes in SF-36v2 scores from pre- to post-treatment.
   - Within each treatment group, changes in SF-36v2 scale and summary scores were calculated by subtracting baseline (pre-treatment) scores from final visit (post-treatment, or endpoint) scores; one-sample t-tests were conducted to determine whether the changes in scores within each treatment group differ from 0.
   - A Multivariate Analysis of Variance (MANOVA) model tested for significant differences in SF-36v2 scale baseline-endpoint change scores between treated and comparison groups. A separate MANOVA model tested for group differences in changes in PCS and MCS scores. Effect sizes for standardized differences between samples were calculated for all mean comparisons to estimate the magnitude of treatment on SF-36 scale and summary scores.

2. Results: Treatment effects for SF-36v2 subscales and summary scores.



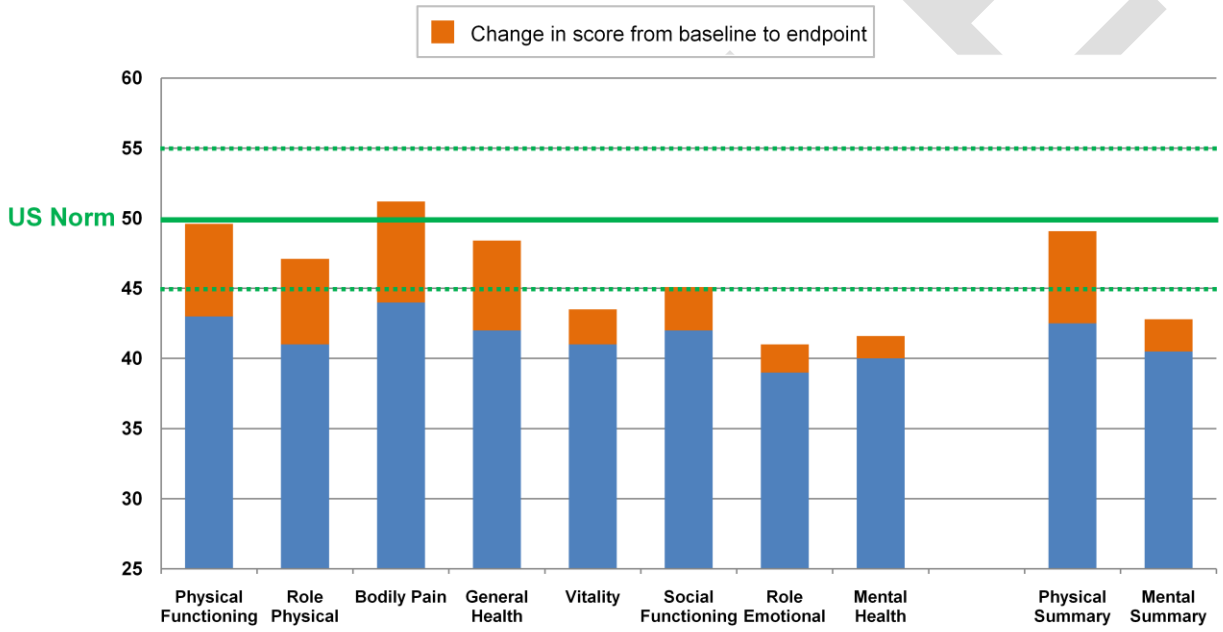
3. Interpretation
   - The benefits of treatment with the Drug X compared to placebo were most notable in the Physical summary score, where the change from baseline was significantly larger for the treatment group. No significant differences were found between groups for the mental summary score.
   - The benefits of treatment with Drug X, as compared to placebo, were most notable in the Social Functioning, Vitality, and General Health domains, where the increase from baseline for patients in the Drug X group were 1.8, 1.9, and 1.4 points greater, respectively, than for patients in the Placebo group.
   - Treatment with Drug X had no impact on RE and MH as compared to the placebo, as indicated by the lack of statistically significant differences in change scores on these scales.
   - Descriptive statistics, inferential statistics, and effect sizes for mean differences between the treatment groups are presented in **Table 3** in the appendix.

Copyright 2010 QualityMetric Incorporated

Norm-based Interpretation of Treatment Effects

1. Rationale and Method
   - This analysis interpreted the HRQOL benefits of treatment with Drug X from a normative perspective by showing the degree to which burden is reduced, or possibly even eliminated, in each subscale and summary score following treatment.
   - Changes in the SF-36v2 scale and summary measure scores are displayed in relation to the average scores from the U.S. general population norms.
   - Thus, the final health status of the study participants was evaluated in terms of their change from baseline as well as the residual burden of disease following treatment.

2. Results: Change in baseline to endpoint SF-36v2 scale and summary scores for patients receiving Drug X relative to U.S. general population norms.



3. Interpretation
   - The reduction of burden in physical health status due to treatment with Drug X is indicated by the fact that patients' Physical Summary component score following treatment is equivalent to that of the US general population normative sample. The burden in mental health status was not eliminated, however, as increase in patients' Mental Component scores following Drug X treatment were still below those of the general population.
   - In the SF-36v2 profile, the greatest benefits of treatment were seen in scales assessing physical health concepts: scores on all 4 physical outcome scales indicated a return to normal functioning following treatment. Smaller improvements were observed in mental health concepts: following treatment with Drug X, patients' scores remained below those of the general population on all 4 mental outcomes scales.
   - Burden was not eliminated in patients receiving the placebo, as indicated by the fact that all scale and summary scores for patients in this group remained below those of the general population (see **Figure 1** in the Appendix).
   - Conclusion: Following treatment with Drug X, burden of disease was eliminated for physical health concepts, but not for mental health concepts.

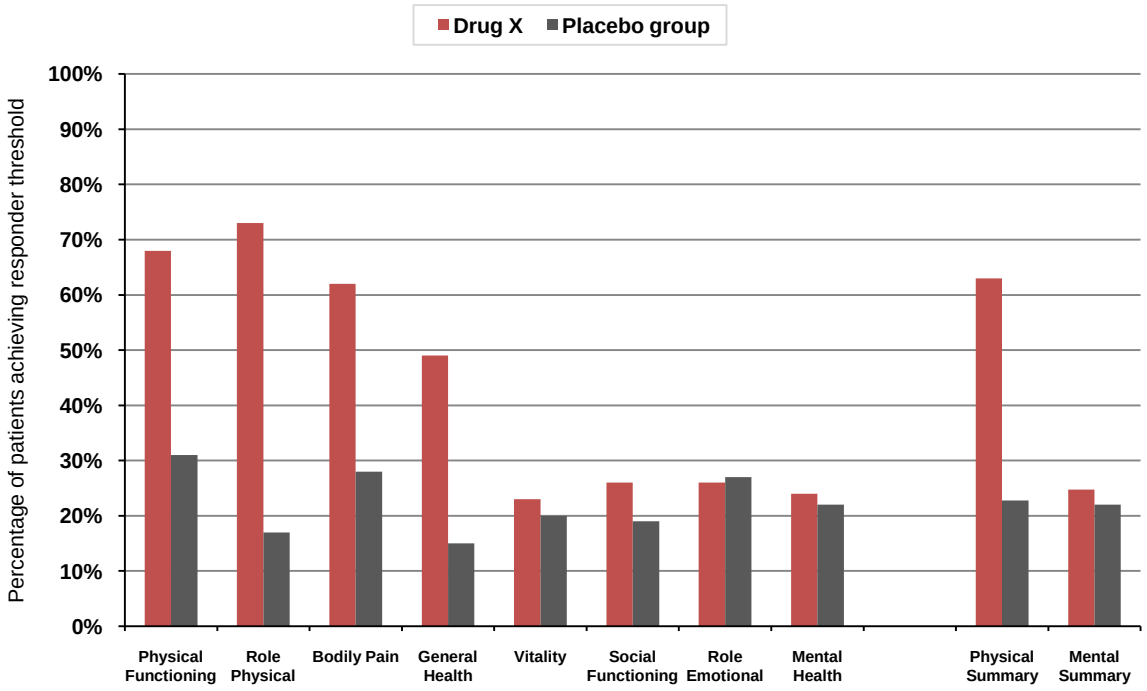
Copyright 2010 QualityMetric Incorporated

Categorical Change Analysis: Treatment Responders

1. Rationale and Method

- This responder analysis shows the percentage of patients in each treatment group who achieved a pre-defined magnitude of change from baseline to endpoint for each SF-36v2 scale and summary score.
- Responder classification was based on empirically established "minimal important difference" (MID) thresholds for the SF-36v2 scales and summary scores.*

2. Results: Percentage of patients achieving responder threshold for each SF-36v2 subscale and summary score.



3. Interpretation

- There was little treatment benefit on mental health outcomes: for the mental summary score and all mental scales, only 20-30% of patients in the Drug X group showed greater than MID improvement, similar to what was seen for patients receiving placebo.
- A clear treatment benefit emerged for physical health outcomes, with over 60% of patients in the Drug X group showed clinically significant improvement on the physical summary score (as compared to less than 25% of patients receiving placebo), and with between 50% and 75% of Drug X patients showing significant improvement on each of the 4 physical outcome scales, as compared to only approximately 15-30% for the placebo group.

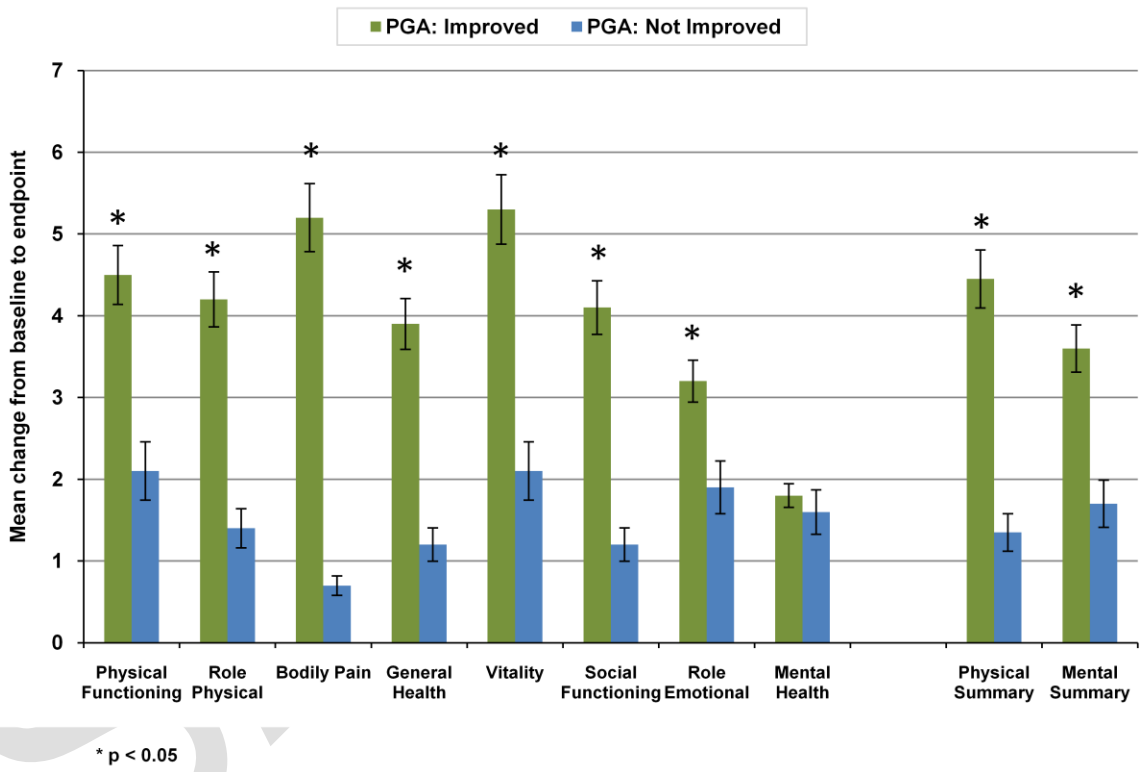
* Ware J.E., Jr., Kosinski M., Bjorner J.B., Turner-Bowker, D.M., Gandek, B. and Maruish, M.E. (2007) User's Manual for the SF-36v2™ Health Survey (2nd Ed). Lincoln, RI, QualityMetric Incorporated.

Copyright 2010 QualityMetric Incorporated

Criterion-based Interpretation of Change Scores

1. Rationale and Method
   - Differences or changes in SF-36v2 scores in relation to external measures, such as other PRO instruments and/or clinical outcomes, were analyzed.
   - The association between trial samples' SF-36v2 change scores with Physician Global Assessment ratings of improvement was determined using independent sample t-tests to comparing mean SF-36v2 scores between patients rated as "improved" and patients rated as "not improved" (**up to two** criterion measures will be analyzed; the criterion measures used will be agreed upon in conversations between ACME and QM staff).

2. Results: Baseline-endpoint change in SF-36v2 scores for patients classified as improved or not improved according to the endpoint Physician Global Assessment measure.



3. Interpretation
   - Across treatment conditions, patients who were classified by their physician as showing overall improvement had significantly higher mental and physical summary scores than did patients classified as not improved.
   - Patients classified as improved also had significantly better scores on all physical domains and on three of the 4 mental domains (with the only exception being the MH scale).
   - Conclusion: Patients' HRQOL very closely corresponds with their physicians' assessment of patient health following treatment.

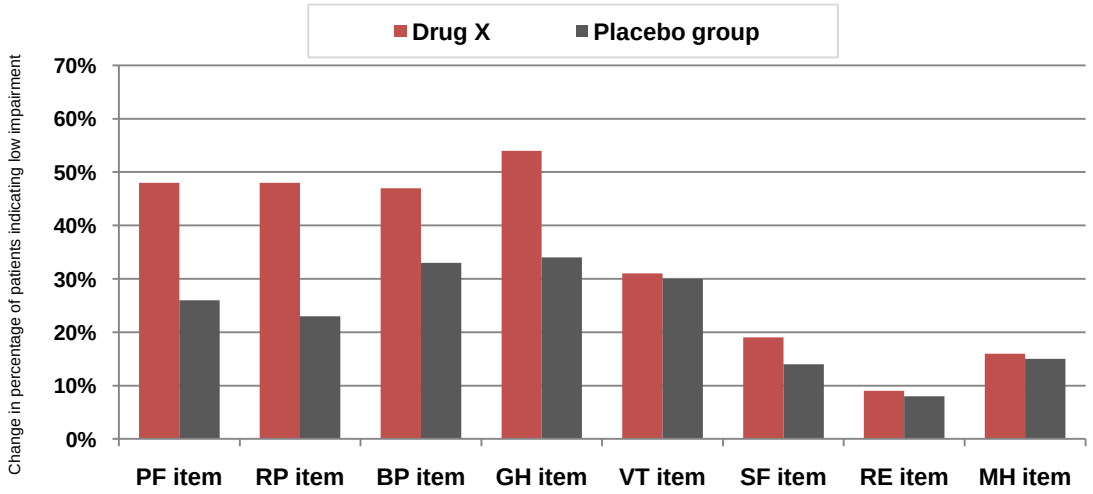
Copyright 2010 QualityMetric Incorporated

Content-based Interpretation

1. Rationale and Method
   - This approach used a descriptive analysis to examine changes in responses to individual items on the SF-36v2 to provide interpretative meaning to any potential improvements conferred by Drug X.
   - Responses to one item from each of the 8 SF-36v2 scales were dichotomized into 'high impairment' and 'low impairment' categories.
   - The change from baseline to endpoint in percentage of patients classified as showing low impairment on each item at baseline and endpoint was calculated for both Drug X and placebo groups.

Scale	Summary of Item content	"High impairment" responses	"Low impairment" responses
PF	Does your health now limit you in moderate activities (e.g., moving a table, bowling)?	Yes, limited a little; Yes, limited a lot	No, not limited at all
RP	How often did you accomplish less at work or other activities as a result of physical health?	All of the time; Most of the time; Some of the time	A little of the time; None of the time
BP	How much did pain interfere with work?	Moderately; Quite a bit; Extremely	A little bit; Not at all
GH	In general, your health is ____	Fair; Poor	Good; Very good; Excellent
VT	How often did you have a lot of energy?	Some of the time; A little of the time; None of the time	Most of the time; All of the time
SF	How often has physical health or emotional problems interfered with social activities?	Most of the time; All of the time; Some of the time	A little of the time; None of the time
RE	How often did you accomplish less at work or other activities as a result of emotional problems?	All of the time; Most of the time; Some of the time	A little of the time; None of the time
MH	How often have you felt downhearted and depressed?	Most of the time; All of the time; Some of the time	A little of the time; None of the time

2. Results: The change from baseline to endpoint in the percentage of patients indicating low impairment.



SAMPLE

3. Interpretation
   - The percentages of patients indicating low impairment for each item at baseline and endpoint are presented in **Table 4** in the Appendix.
   - Patients in the Drug X group appear to have larger improvement than patients in the placebo group in impairment as measured by physical scale items:
     - Fewer patients are limited in moderately physical activities
     - Fewer patients accomplished less as a result of physical problems
     - Fewer patients had pain interfere with work
     - Fewer patients reported fair or poor health
   - Patients in the Drug X group appear to have similar levels of improvement as patients in the control group in impairment as measured by mental scale items:
     - Similar numbers of patients had a lot of energy
     - Similar numbers of patients had health problems interfere with social activities
     - Similar numbers of patients accomplished less as a result of emotional problems
     - Similar numbers of patients reported feeling depressed

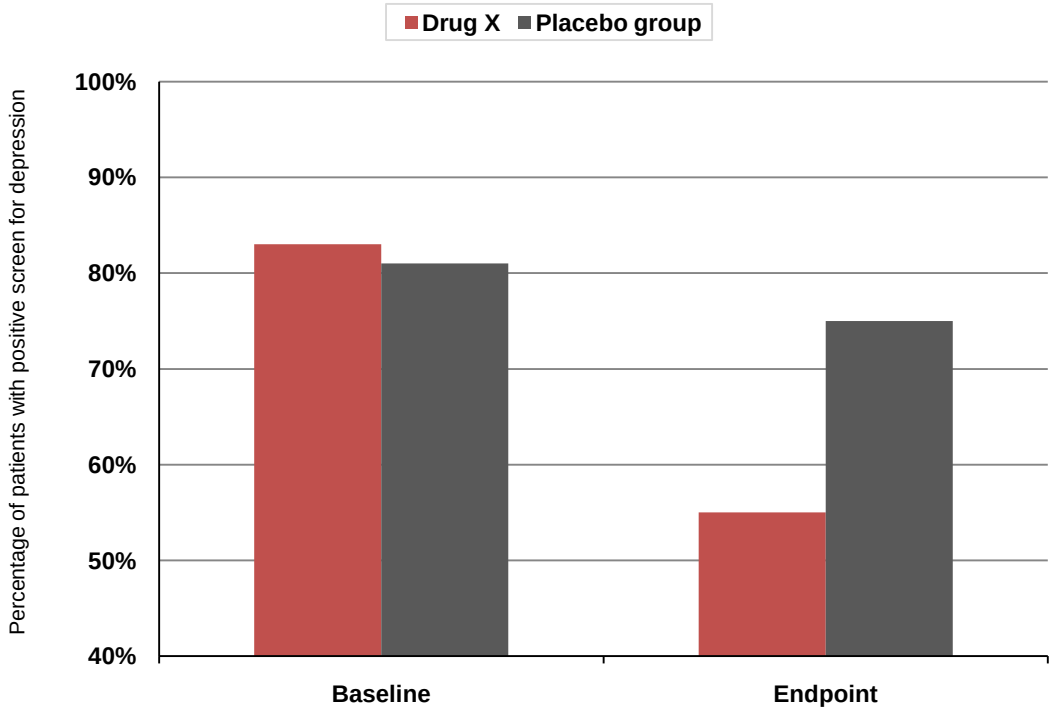
SAMPLE

Depression Screening

1. Rationale and Method

- This approach used SF-36v2 scores to estimate the prevalence of depression for patients in each treatment condition.
- The five-item Mental Health (MH) scale has been shown to be an effective screener for depression.*
- Cross-calibration formulas classified the MH scores into one of four levels of severity:
  - Not depressed, mild-moderate depression, moderate-severe depression, and severe depression.
  - The levels correspond with severity staging from the Beck Depression Inventory®.

2. Results: Percentage of patients in each treatment condition with positive screening for depression.



3. Interpretation
   - At baseline, more than eight out of ten patients were found to have a positive screen for depression.
   - After receiving treatment with Drug X, the percent of patients who screened positive for depression had declined substantially to 55%.
- The placebo group showed only a minor decline in the percentage of patients screening positive for depression.
- Conclusion: Treatment with Drug X had a significant favorable impact on the overall mental health status of the Target population.

* Bjorner, J, Kosinski M, Raju A. (2009). Evaluation of the Validity of BDI Scores Estimated from the SF-36 Mental Health Scale. Paper presented at International Society for Quality of Life Research, New Orleans.

Copyright 2010 QualityMetric Incorporated

Appendix

Table 1. Baseline burden of disease relative to age and gender-matched U.S. general population normative sample.

	Trial sample			Matched U.S. general population normative sample					
	N	***Mean***	***SD***	***N***	***Mean***	***SD***	***t***	***p-value***	***Effect size (Cohen's d)***
Physical functioning	186	28.50	15.09	7058	50.75	12.61	-23.62	0.0000	-1.75
Role physical	188	31.10	16.95	7059	50.74	22.70	-11.78	0.0000	-0.87
Bodily pain	184	32.00	19.02	7069	50.28	31.53	-7.83	0.0000	-0.58
General health	189	39.60	18.26	7056	50.38	25.22	-5.83	0.0000	-0.43
Vitality	186	36.40	18.11	7052	50.06	10.09	-17.73	0.0000	-1.32
Social functioning	185	38.20	19.79	7051	50.04	20.18	-7.88	0.0000	-0.59
Role emotional	186	42.20	18.82	7069	50.42	13.87	-7.89	0.0000	-0.59
Mental health	184	44.70	18.87	7051	49.75	18.92	-3.57	0.0004	-0.27
PCS	182	28.20	13.20	7054	50.80	12.10	-24.82	0.0000	-1.86
MCS	179	44.40	12.10	7051	49.83	9.80	-7.27	0.0000	-0.55

SAMPLE

Copyright 2010 QualityMetric Incorporated

Table 2. Baseline burden of disease relative to age and gender-matched kidney disease patient sample.

	Trial sample			Matched kidney disease patient sample					
	N	***Mean***	***SD***	***N***	***Mean***	***SD***	***t***	***p-value***	***Effect size (Cohen's d)***
Physical functioning	186	28.50	15.09	87	43.21	24.40	-6.10	0.0000	-0.80
Role physical	188	31.10	16.95	87	45.10	22.57	-5.71	0.0000	-0.74
Bodily pain	184	32.00	19.02	86	41.57	15.55	-4.07	0.0001	-0.53
General health	189	39.60	18.26	87	36.96	16.37	1.15	0.2507	0.15
Vitality	186	36.40	18.11	87	39.52	20.95	-1.26	0.2083	-0.16
Social functioning	185	38.20	19.79	85	42.21	28.01	-1.35	0.1787	-0.18
Role emotional	186	42.20	18.82	87	39.84	31.36	0.77	0.4415	0.10
Mental health	184	44.70	18.87	87	41.97	25.68	0.99	0.3250	0.13
PCS	182	28.20	13.20	84	43.54	21.47	-7.15	0.0000	-0.95
MCS	179	44.40	12.10	83	40.80	19.44	1.83	0.0682	0.24

SAMPLE

Copyright 2010 QualityMetric Incorporated

Table 3. Change in SF-36v2 subscales and summary scores from baseline to endpoint for each treatment condition.

	Drug X group			Placebo group					
	N	***Mean***	***SE***	***N***	***Mean***	***SE***	***F***	***p-value***	***Effect size (Cohen's d)***
Physical functioning	64	3.6	2.6	59	2.5	1.0	9.26	0.0029	0.55
Role physical	63	3.1	3.2	60	2.1	0.7	5.59	0.0196	0.43
Bodily pain	64	4.2	1.5	60	3.1	1.6	15.64	0.0001	0.71
General health	64	3.4	1.9	60	2.0	0.9	27.40	0.0000	0.94
Vitality	63	6.3	2.0	60	4.4	1.4	38.77	0.0000	1.13
Social functioning	62	3.6	1.2	59	2.2	1.2	40.70	0.0000	1.16
Role emotional	64	2.3	1.1	58	2.1	0.8	1.28	0.2608	0.21
Mental health	64	2.1	0.8	60	2.3	1.1	1.31	0.2540	0.21
PCS	63	3.8	2.6	59	2.5	1.1	13.09	0.0004	0.66
MCS	62	2.9	1.0	58	2.4	1.0	8.90	0.0035	0.55

SAMPLE

Copyright 2010 QualityMetric Incorporated

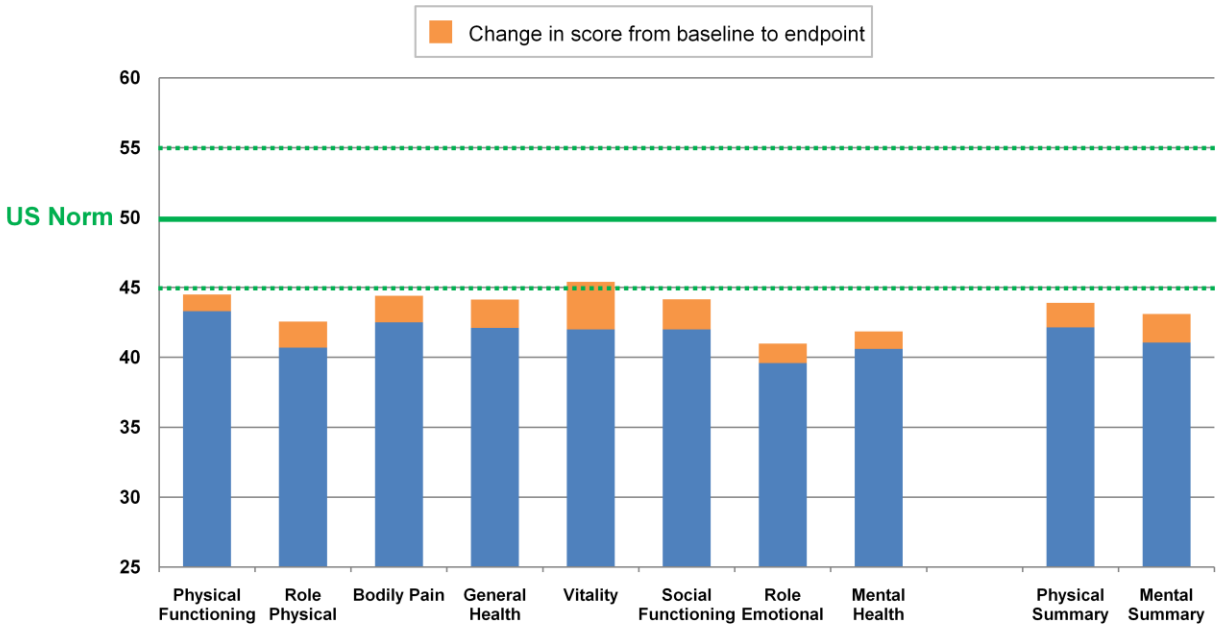
Table 4. Content-based interpretation: the percentage of patients within each condition indicating low impairment for each item.

	Drug X group			Placebo group		
	Baseline	***Endpoint***	***Difference***	***Baseline***	***Endpoint***	***Difference***
Physical functioning item	20%	68%	***48%***	30%	56%	***26%***
Role physical item	33%	81%	***48%***	34%	57%	***23%***
Bodily pain item	27%	74%	***47%***	29%	62%	***33%***
General health item	23%	77%	***54%***	34%	68%	***34%***
Vitality item	38%	69%	***31%***	39%	69%	***30%***
Social functioning item	24%	43%	***19%***	35%	49%	***14%***
Role emotional item	35%	44%	***9%***	22%	30%	***8%***
Mental health item	31%	47%	***16%***	26%	41%	***15%***

SAMPLE

Copyright 2010 QualityMetric Incorporated

Figure 1. Norm-based interpretation of treatment effects for patients receiving placebo treatment.



SAMP