NATIONAL QUALITY FORUM

Person- and Family-Centered Care Measure Developers Additional Information

AACVPR
American Association of Cardiovascular and Pulmonary Rehabilitation
Promoting Health & Preventing Disease

March 27, 2015

Re: Measure 0701: Functional Capacity in COPD patients before and after Pulmonary Rehabilitation

Dear Members of the Person- and Family-Centered Care Measures Review Committee,

We would like to thank the committee for this opportunity to submit supplemental data that we hope will help to resolve some of the concerns expressed by the committee. In the comments contained within the draft report on person- and family-centered care measures, the committee expressed concern about no data being presented at the program level regarding changes in 6 minute walk test distance. This resulted in some comments asking for demonstration of a performance gap in programs that perform pulmonary rehabilitation as well as concerns regarding the reproducibility, and thus, the reliability of this measure. After considering these comments and obtaining clarification from staff at National Quality Forum, we have conducted additional analyses using data from the American Association of Cardiovascular and Pulmonary Rehabilitation registry. The results of these analyses are displayed in the Table below. The data shown in the table are based on the proportion of patients, by program, who increase their 6 minute walk test by the minimally important difference of 25 meters. For all 125 programs, the median value was 80% with an interquartile range of 70-92%, meaning that one-half of all programs had at least 80% of their patients achieve an increase in their 6 minute walk test distance of at least 25 meters, one-fourth of programs had at least 92% of their patients achieve an increase in their 6 minute walk test distance of at least 25 meters, and one-fourth of programs had fewer than 70% of their patents achieve an increase in their 6 minute walk test distance of at least 25 meters from enrollment to completion of pulmonary rehabilitation.

Additionally, because we were concerned about variation in the precision of our measurements based on the size of the program, we conducted these analyses both in the entire registry overall and then stratified by program size. As shown in the table, we did not observe significant variability in the median value for the proportion of patients in a program who increased their 6 minute walk test distance by at least 25 meters based on size of program, with the median value ranging from 75-86% by size of program. Similarly, although the minimum value varied by outlier programs, the interquartile range was relatively stable by program size, further demonstrating a lack of variability across program size.

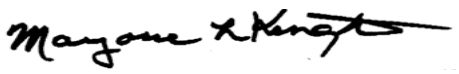
330 N. Wabash Ave., Suite 2000 • Chicago, Illinois 60611 • (312) 321-5146 • Fax (312) 673-6924 • E-mail: aacvpr@aacvpr.org • Web: aacvpr.org

We hope that these data will help to address some of the concerns expressed by the committee in their comments. First, there does appear to be a performance gap that this measure can help to address. One-fourth of all programs have fewer than 70% of their patients increasing their 6 minute walk test distance by 25 meters from enrollment to completion of pulmonary rehabilitation. Although this would seem to indicate that there is a performance gap that could be addressed in part by this performance measure, some care must be taken in interpreting these data. We are unable to risk adjust this measure as there is a lack of a suitable model for risk adjustment at the program level. Thus, patients entering pulmonary rehabilitation in one center may not be similar to those in another center.

Secondly, and most importantly, we hope that these data will help to demonstrate the reliability and reproducibility of our measure. The median percentage of patients by program who increase their 6 minute walk test distance by 25 meters was 80% overall and did not vary much by program size. In addition, the interquartile range also did not vary by program size. This provides reassurance about the precision of our measure, which is further supported by the literature where change in 6 minute walk test distance is considered to be a highly reliable and reproducible measure. One representative study by Sciurba, et al [1] conducted repeated 6 minute walk tests on successive days on the same 470 participants at 17 different clinical sites in the National Emphysema Treatment Trial, which included pulmonary rehabilitation as part of its treatment strategy. They observed an intra-class correlation coefficient of 0.88 [1], indicating a very high degree of reliability between the repeated tests. In addition, in part because of its reliability, the American Thoracic Society recommends this test as a measure of functional capacity in patients with COPD [2]. Given the strength of these data in the literature on the reliability and reproducibility of 6 minute walk test distance, we have chosen to not repeat these measures in our own patients as we feel that it would be overly burdensome to our patients to have to repeat the 6 minute walk test on multiple days simply to support this application. In addition, to test reproducibility across programs, we would have to ask patients to travel to different programs to have the test repeated, which would certainly be overly burdensome. As mentioned previously, we are reassured by these data presented below that the median and interquartile ranges for the proportion of patients in a program who increase their 6 minute walk test distance by 25 meters varies very little by size of program. We hope that the committee will view these data as evidence of the reliability and reproducibility of this test, and accept this, along with the abundance of data in the literature that also supports this.

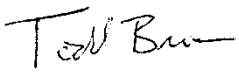
We would again like to thank the committee for this opportunity to submit supplemental data to support our application and address some of the comments presented to us in the draft report from the committee on person- and family-centered care measures. We hope that these data have addressed some of the comments contained in the report, and we would be happy to discuss this further with the committee should that be helpful.

Sincerely,



Marjorie King, MD, MAACVPR
Chair – AACVPR Quality of Care Committee

Gerene S. Bauldoff, PhD, RN, FCCP, FAACVPR, FAAN
Member – AACVPR Quality of Care Committee



Todd M. Brown, MD, MSPH
Member – AACVPR Board of Directors and
Liaison to the Quality of Care Committee

References

1. Sciurba F, Criner GJ, Lee SM, Mohsenifar Z, Shade D, Slivka W, Wise RA; National Emphysema Treatment Trial Research Group. Six-minute walk distance in chronic obstructive pulmonary disease: reproducibility and effect of walking course layout and length. *Am J Respir Crit Care Med* 2003;167:1522-1527.
2. American Thoracic Society: Statement Guidelines for the Six-Minute Walk Test. *Am J Resp Crit Care Med*, 2002;166:111-117.

Table. Descriptive Statistics for the Proportion of a Program's Pulmonary Rehabilitation Participants who Increase Their Six Minute Walk Test Distance by at Least 25 Meters Stratified by Size of Program

	N*	Minimum	25th Percentile	Median	75th Percentile	Maximum
All Programs	125	29%	70%	80%	92%	100%
Small Programs†	17	29%	67%	86%	100%	100%
Medium Programs†	50	50%	71%	79%	89%	100%
Large Programs†	50	33%	75%	82%	93%	100%
Unknown Size	8	52%	66%	75%	100%	100%

Data displayed are based on the program level data on what proportion of that program's participants increased their 6 minute walk test distance by at least 25 meters.

*N=number of programs

†Small Programs enroll <25 phase 2 patients annually; medium programs enroll 25-75 phase 2 patients annually; large programs enroll >75 phase 2 patients annually.

NQF#2624 Functional Outcome Assessment
Supplemental Information for Person and Family-Centered Care Committee

CMS and its measure development team thank you for the opportunity to submit follow-up information for NQF#2624 Functional Outcome Assessment measure as requested by the Committee during the evaluation process. The Committee specifically requested clarification of the following items:

1) The Committee would like to see information or a consideration of change to the specifications establishing a link between the assessment and the care plan- need data that clearly links the care plan with the collection of the outcomes data
2) The Committee also would like to see inter-rater reliability
3) And greater clarity on how each element of the process definition is actually measured in the field (what are coders looking for?)

We will address each item in order. It is important to note that NQF 2624 is NOT an outcomes measure. It is a process measure reported predominantly using claims data, although it can be reported via registry as well.

Item #1: Establishing a link between the assessment and care plan
The level of detail used in the G-codes to describe the clinical actions indicates there is a relationship, thus a linkage, between the assessment of the patient on the date of the encounter using a standardized tool and a care plan based on that assessment (identified deficiencies) on the date of the encounter. Please see the below G-code.
Numerator Quality-Data Coding Options for Reporting Satisfactorily:
Functional Outcome Assessment Documented as Positive AND Care Plan Documented
Performance Met: **G8539:** Functional outcome assessment documented as positive using a standardized tool **AND** a care plan based, on identified deficiencies on the date of the functional outcome assessment, is documented.

Item #2: Inter-rater reliability
Inter-rater reliability is performed by two abstractors who review each of the data elements that comprise the G-code. For example, for G8539 (defined above), the abstractor reviews the medical record to determine:

1. Date of assessment
2. Assessment tool is a standardized/validated tool
3. Deficiencies are documented
4. Care plan reflects the findings of the assessment on the date of the assessment and is documented on the date of the assessment

CMS and its developer provided the inter-rater reliability information in the NQF testing form that was submitted in our NQF submission form for #2624 (please refer to section 2a2 Reliability Testing in submitted NQF testing form for complete details). For your convenience, the inter-reliability testing results are outlined below:
<u>Inter-Rater Reliability:</u>

Numerator crude agreement: 81.3%
Prevalence Adjusted Kappa=.64 (95% CI .54-.74)
Kappa=.59(95% CI .48-.70)
Reliability testing with two independent reviewers demonstrates **substantial** agreement.

NQF#2624 Functional Outcome Assessment
Supplemental Information for Person and Family-Centered Care Committee

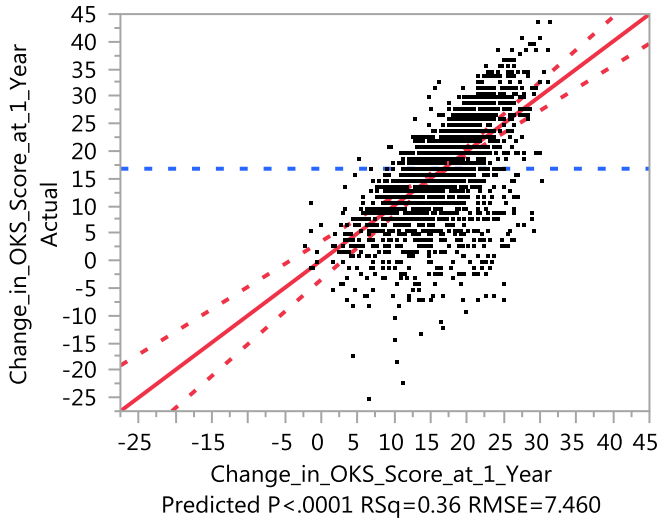
Item #3: Measurement in the field
When selecting the G-code for this measure, the coders are looking for documentation of the appropriate G-code for the clinical action, for example, G8539: Functional outcome assessment documented as positive using a standardized tool **AND** a care plan based, on identified deficiencies on the date of the functional outcome assessment, is documented. Additional G-codes are defined in the measure specifications.

Via the Physician Quality Reporting System (PQRS) claims reporting option, eligible professionals select the appropriate G code for the quality measure that relates to the claims billed for services rendered. G-codes are defined as Quality Date Codes (QDCs), which are a subset of HCPCs II codes. *QDCs are **non-billable codes** that providers use to delineate their clinical quality actions.* They are submitted with Medicare Part B Claims. The applicable G-codes were included with our initial measure submission to NQF.

As a measure developer, our focus is on defining the G-code as specifically as possible to describe the clinical quality action; we are not involved in billing codes other than using CPT codes for denominator inclusion.

Please let us know if you have any additional questions or if you require additional information that was not included in the initial submission. If so, please advise us of the specific materials we can provide to address any outstanding issues.

Response Change_in_OKS_Score_at_1_Year
Whole Model
Actual by Predicted Plot



Summary of Fit

RSquare	0.360264
RSquare Adj	0.351696
Root Mean Square Error	7.460606
Mean of Response	16.76517
Observations (or Sum Wgts)	2044

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	27	63191.42	2340.42	42.0481
Error	2016	112211.86	55.66	**Prob > F**
C. Total	2043	175403.28		<.0001*

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	1955	109613.70	56.0684	1.3164
Pure Error	61	2598.17	42.5929	**Prob > F**
Total Error	2016	112211.86		0.0847
				Max RSq

Parameter Estimates

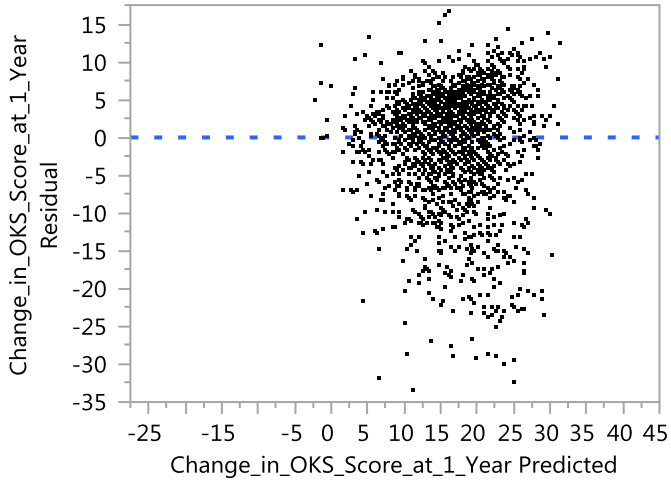
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	28.219982	2.054943	13.73	<.0001*
Preop_OKS_Score	-0.68468	0.021459	-31.91	<.0001*
Months to Procedure	0.0396532	0.146958	0.27	0.7873
Procedure_Type[1]	3.3848358	0.423612	7.99	<.0001*
Medical_Group_Name A	-2.435658	1.146636	-2.12	0.0338*
Medical_Group_Name B	-1.721775	0.717873	-2.40	0.0166*
Medical_Group_Name C	-1.602351	0.863207	-1.86	0.0636
Medical_Group_Name D	2.3361849	2.127144	1.10	0.2722
Medical_Group_Name E	1.00732	0.803359	1.25	0.2100
Medical_Group_Name F	0.1476835	2.437651	0.06	0.9517
Medical_Group_Name G	3.3995954	1.912967	1.78	0.0757
Medical_Group_Name H	-2.108872	2.750914	-0.77	0.4434
Medical_Group_Name I	-10.08843	5.086	-1.98	0.0474*
Medical_Group_Name J	0.3364536	0.77405	0.43	0.6639

Term	Estimate	Std Error	t Ratio	Prob>|t|
Medical_Group_Name K	2.6881594	0.929903	2.89	0.0039*
Medical_Group_Name L	-1.617694	2.3177	-0.70	0.4853
Medical_Group_Name M	-1.826898	1.441525	-1.27	0.2052
Medical_Group_Name N	1.4495541	1.260764	1.15	0.2504
Medical_Group_Name O	1.2852423	0.64182	2.00	0.0454*
Medical_Group_Name P	0.3776601	0.863813	0.44	0.6620
Medical_Group_Name Q	0.6789508	0.835126	0.81	0.4163
Medical_Group_Name R	-5.152488	2.965465	-1.74	0.0825
Medical_Group_Name S	-0.098323	1.306203	-0.08	0.9400
Medical_Group_Name T	2.4420679	1.216275	2.01	0.0448*
Medical_Group_Name U	-0.479812	0.879704	-0.55	0.5855
Medical_Group_Name V	5.5260396	7.178079	0.77	0.4415
Medical_Group_Name W	0.5848397	0.963037	0.61	0.5437
Medical_Group_Name X	4.2794942	4.160832	1.03	0.3038

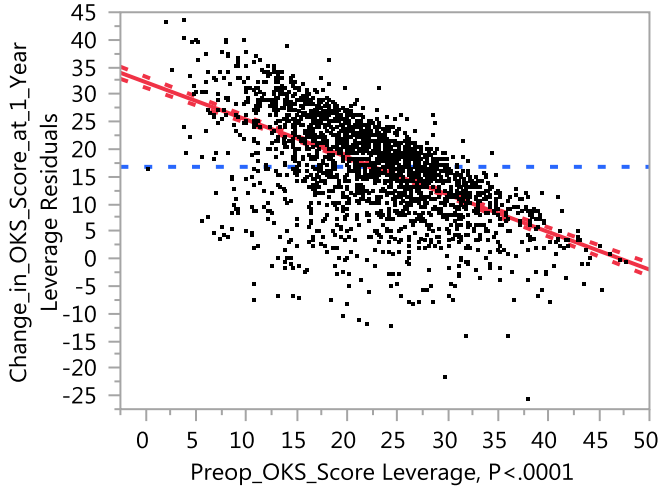
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Preop_OKS_Score	1	1	56665.665	1018.056	<.0001*
Months to Procedure	1	1	4.052	0.0728	0.7873
Procedure_Type	1	1	3553.747	63.8467	<.0001*
Medical_Group_Name	24	24	3578.214	2.6786	<.0001*

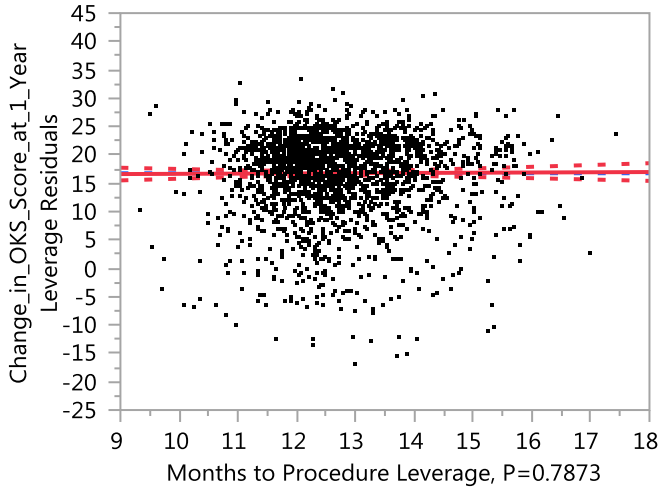
Residual by Predicted Plot



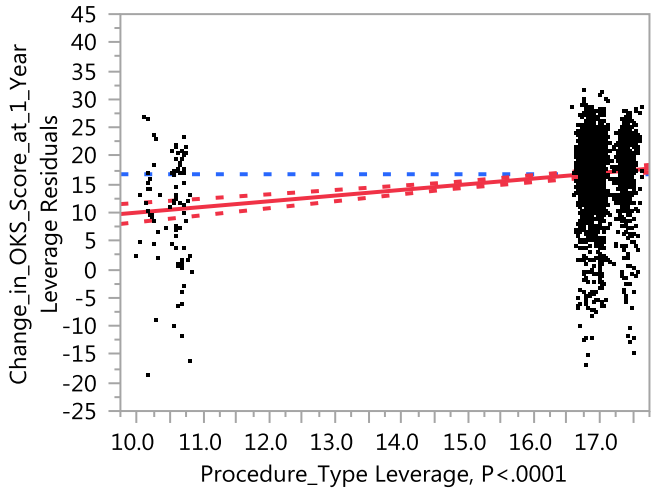
Preop_OKS_Score
Leverage Plot



Months to Procedure
Leverage Plot



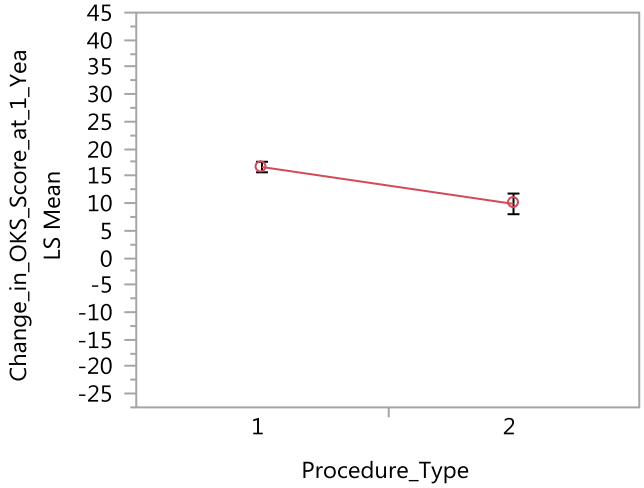
Procedure_Type
Leverage Plot



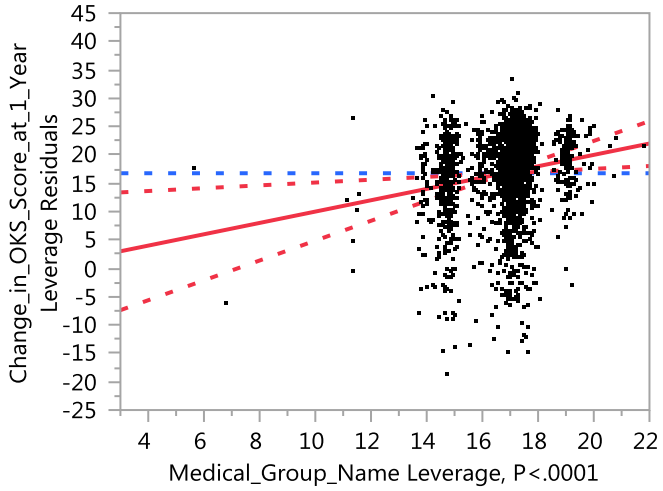
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	16.641674	0.49382938	17.0041
2	9.872003	0.96717066	11.1905

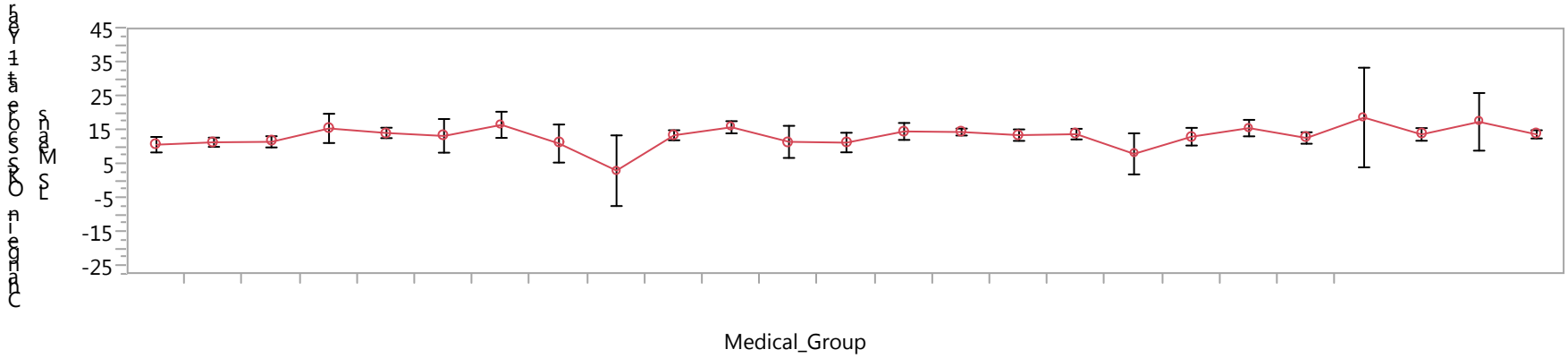
LS Means Plot



Medical_Group_Name
Leverage Plot

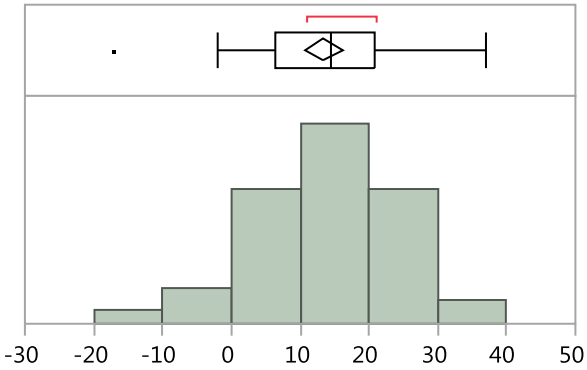


LS Means Plot by Medical Group



Distributions Medical_Group_Name=A

Change_in_OKS_Score_at_1_Year



Quantiles

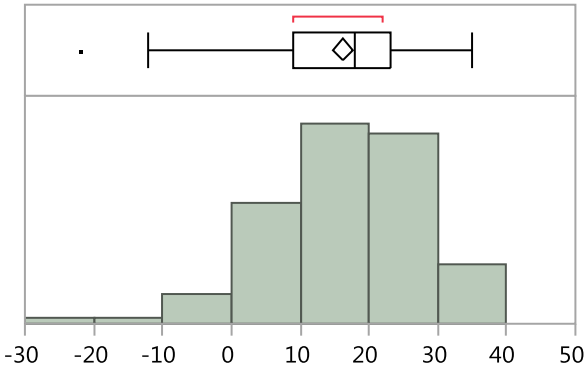
100.0%	maximum	37
99.5%		37
97.5%		36.325
90.0%		23
75.0%	quartile	20.75
50.0%	median	14.5
25.0%	quartile	6.25
10.0%		1.7
2.5%		-13.625
0.5%		-17
0.0%	minimum	-17

Summary Statistics

Mean	13.416667
Std Dev	9.6993822
Std Err Mean	1.3999852
Upper 95% Mean	16.233074
Lower 95% Mean	10.60026
N	48

Distributions Medical_Group_Name=B

Change_in_OKS_Score_at_1_Year



Quantiles

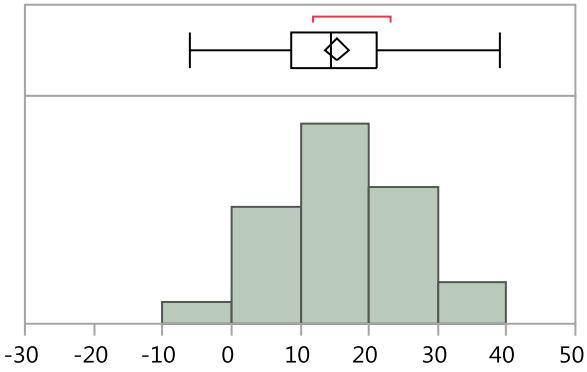
100.0%	maximum	35
99.5%		35
97.5%		34
90.0%		29
75.0%	quartile	23
50.0%	median	18
25.0%	quartile	9
10.0%		2
2.5%		-8
0.5%		-22
0.0%	minimum	-22

Summary Statistics

Mean	16.142105
Std Dev	10.405829
Std Err Mean	0.7549182
Upper 95% Mean	17.631253
Lower 95% Mean	14.652957
N	190

Distributions Medical_Group_Name=C

Change_in_OKS_Score_at_1_Year



Quantiles

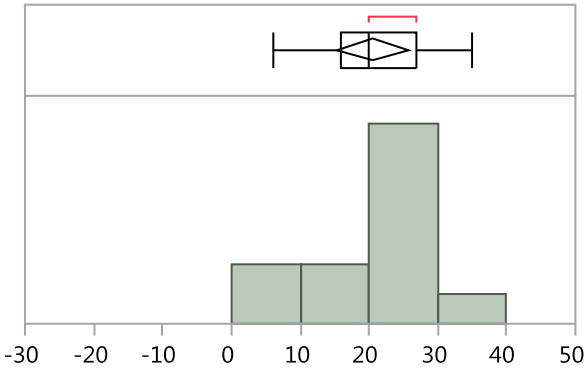
100.0%	maximum	39
99.5%		39
97.5%		35.275
90.0%		27
75.0%	quartile	21
50.0%	median	14.5
25.0%	quartile	8.75
10.0%		3.3
2.5%		-3.85
0.5%		-6
0.0%	minimum	-6

Summary Statistics

Mean	15.313725
Std Dev	9.1910206
Std Err Mean	0.9100466
Upper 95% Mean	17.119013
Lower 95% Mean	13.508438
N	102

Distributions Medical_Group_Name=D

Change_in_OKS_Score_at_1_Year



Quantiles

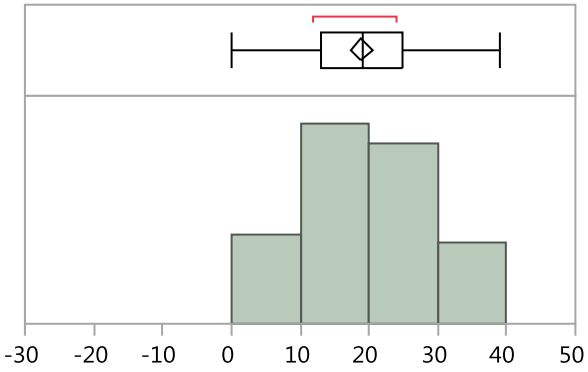
100.0%	maximum	35
99.5%		35
97.5%		35
90.0%		32.6
75.0%	quartile	27
50.0%	median	20
25.0%	quartile	16
10.0%		6.6
2.5%		6
0.5%		6
0.0%	minimum	6

Summary Statistics

Mean	20.5
Std Dev	8.2956725
Std Err Mean	2.3947544
Upper 95% Mean	25.770819
Lower 95% Mean	15.229181
N	12

Distributions Medical_Group_Name=Essentia Health - E

Change_in_OKS_Score_at_1_Year



Quantiles

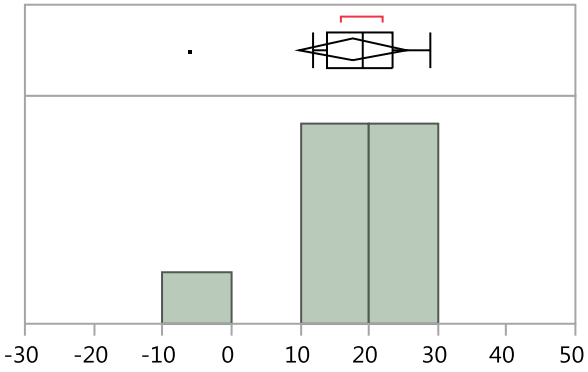
100.0%	maximum	39
99.5%		39
97.5%		33.7
90.0%		31
75.0%	quartile	25
50.0%	median	19
25.0%	quartile	13
10.0%		6.2
2.5%		1
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	18.877863
Std Dev	8.7924993
Std Err Mean	0.7682042
Upper 95% Mean	20.397663
Lower 95% Mean	17.358062
N	131

Distributions Medical_Group_Name=F

Change_in_OKS_Score_at_1_Year



Quantiles

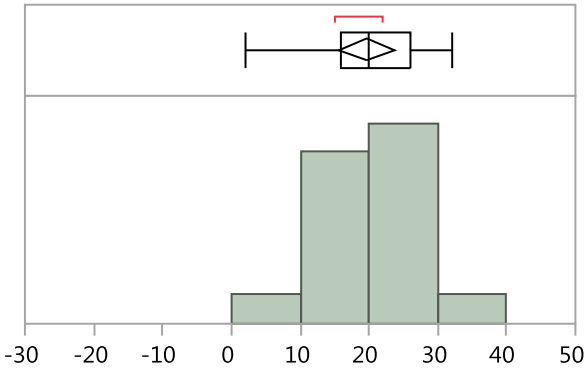
100.0%	maximum	29
99.5%		29
97.5%		29
90.0%		29
75.0%	quartile	23.5
50.0%	median	19
25.0%	quartile	14
10.0%		-6
2.5%		-6
0.5%		-6
0.0%	minimum	-6

Summary Statistics

Mean	17.555556
Std Dev	10.113248
Std Err Mean	3.3710825
Upper 95% Mean	25.329286
Lower 95% Mean	9.7818253
N	9

Distributions Medical_Group_Name=G

Change_in_OKS_Score_at_1_Year



Quantiles

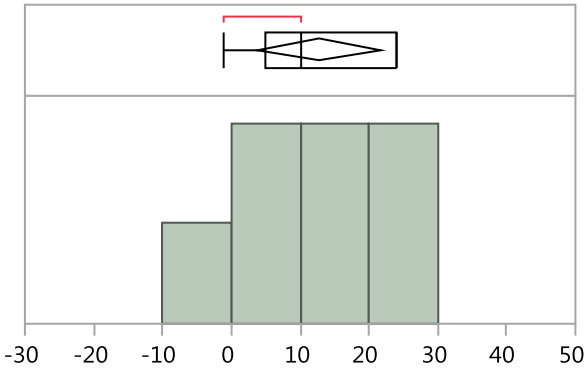
100.0%	maximum	32
99.5%		32
97.5%		32
90.0%		29
75.0%	quartile	26
50.0%	median	20
25.0%	quartile	16
10.0%		8.6
2.5%		2
0.5%		2
0.0%	minimum	2

Summary Statistics

Mean	19.6
Std Dev	7.2387844
Std Err Mean	1.8690461
Upper 95% Mean	23.608705
Lower 95% Mean	15.591295
N	15

Distributions Medical_Group_Name=H

Change_in_OKS_Score_at_1_Year



Quantiles

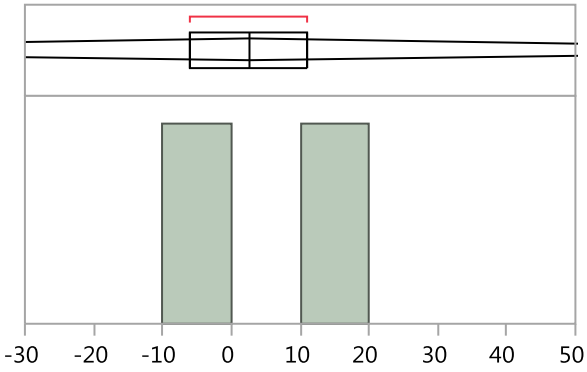
100.0%	maximum	24
99.5%		24
97.5%		24
90.0%		24
75.0%	quartile	24
50.0%	median	10
25.0%	quartile	5
10.0%		-1
2.5%		-1
0.5%		-1
0.0%	minimum	-1

Summary Statistics

Mean	12.714286
Std Dev	9.5866971
Std Err Mean	3.6234309
Upper 95% Mean	21.580502
Lower 95% Mean	3.8480696
N	7

Distributions Medical_Group_Name=I

Change_in_OKS_Score_at_1_Year



Quantiles

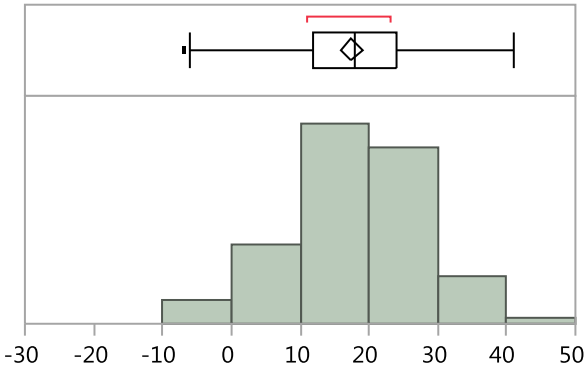
100.0%	maximum	11
99.5%		11
97.5%		11
90.0%		11
75.0%	quartile	11
50.0%	median	2.5
25.0%	quartile	-6
10.0%		-6
2.5%		-6
0.5%		-6
0.0%	minimum	-6

Summary Statistics

Mean	2.5
Std Dev	12.020815
Std Err Mean	8.5
Upper 95% Mean	110.50274
Lower 95% Mean	-105.5027
N	2

Distributions Medical_Group_Name=J

Change_in_OKS_Score_at_1_Year



Quantiles

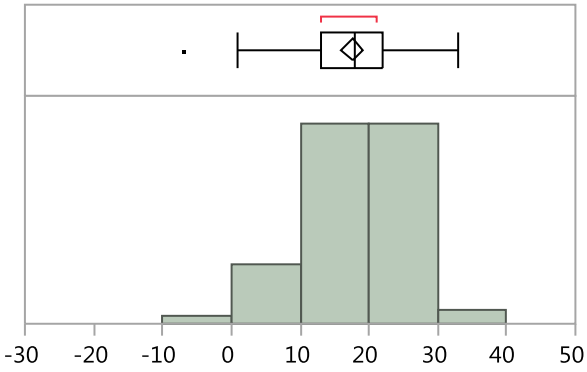
100.0%	maximum	41
99.5%		41
97.5%		34.3
90.0%		29.2
75.0%	quartile	24
50.0%	median	18
25.0%	quartile	12
10.0%		5.8
2.5%		-5.3
0.5%		-7
0.0%	minimum	-7

Summary Statistics

Mean	17.408163
Std Dev	9.5226875
Std Err Mean	0.785418
Upper 95% Mean	18.960421
Lower 95% Mean	15.855906
N	147

Distributions Medical_Group_Name=K

Change_in_OKS_Score_at_1_Year



Quantiles

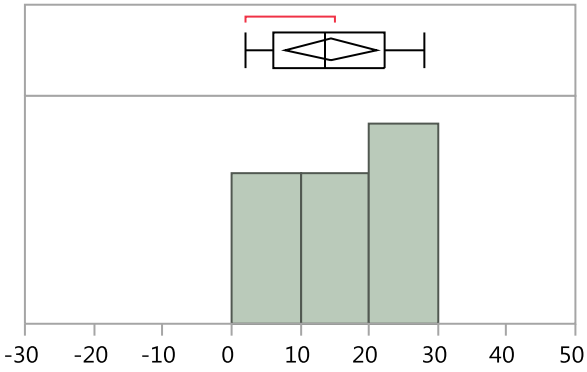
100.0%	maximum	33
99.5%		33
97.5%		30.7
90.0%		26.6
75.0%	quartile	22
50.0%	median	18
25.0%	quartile	13
10.0%		7.4
2.5%		1.1
0.5%		-7
0.0%	minimum	-7

Summary Statistics

Mean	17.60241
Std Dev	7.1208039
Std Err Mean	0.7816098
Upper 95% Mean	19.15728
Lower 95% Mean	16.047539
N	83

Distributions Medical_Group_Name=L

Change_in_OKS_Score_at_1_Year



Quantiles

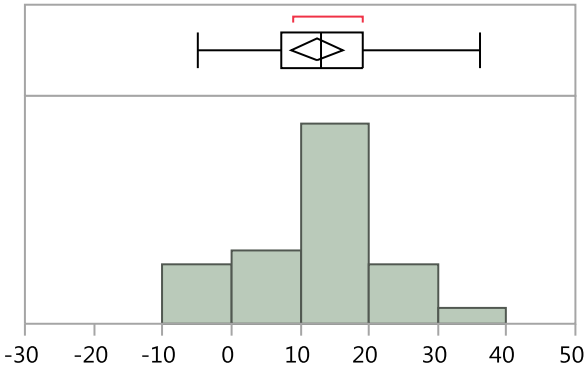
100.0%	maximum	28
99.5%		28
97.5%		28
90.0%		27.8
75.0%	quartile	22.25
50.0%	median	13.5
25.0%	quartile	6
10.0%		2.1
2.5%		2
0.5%		2
0.0%	minimum	2

Summary Statistics

Mean	14.5
Std Dev	9.1317517
Std Err Mean	2.8877134
Upper 95% Mean	21.032462
Lower 95% Mean	7.9675384
N	10

Distributions Medical_Group_Name=M

Change_in_OKS_Score_at_1_Year



Quantiles

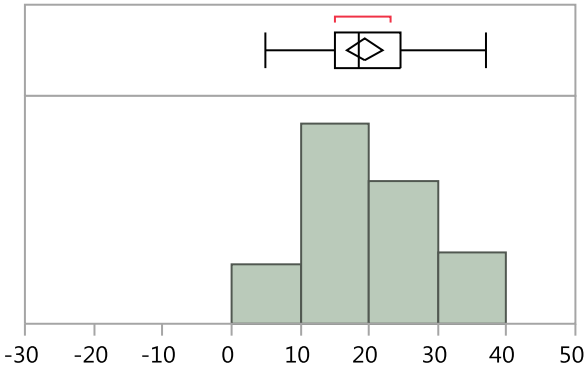
100.0%	maximum	36
99.5%		36
97.5%		36
90.0%		25.2
75.0%	quartile	19
50.0%	median	13
25.0%	quartile	7.25
10.0%		-4.1
2.5%		-5
0.5%		-5
0.0%	minimum	-5

Summary Statistics

Mean	12.5
Std Dev	9.9349738
Std Err Mean	1.8775336
Upper 95% Mean	16.352381
Lower 95% Mean	8.6476193
N	28

Distributions Medical_Group_Name=N

Change_in_OKS_Score_at_1_Year



Quantiles

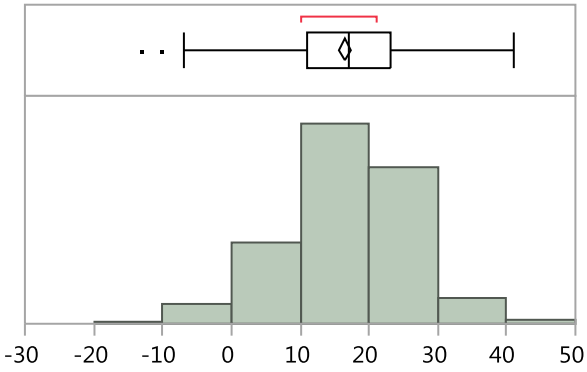
100.0%	maximum	37
99.5%		37
97.5%		36.9
90.0%		31.9
75.0%	quartile	24.5
50.0%	median	18.5
25.0%	quartile	15
10.0%		9
2.5%		5.025
0.5%		5
0.0%	minimum	5

Summary Statistics

Mean	19.375
Std Dev	7.8827546
Std Err Mean	1.2463729
Upper 95% Mean	21.896027
Lower 95% Mean	16.853973
N	40

Distributions Medical_Group_Name=O

Change_in_OKS_Score_at_1_Year



Quantiles

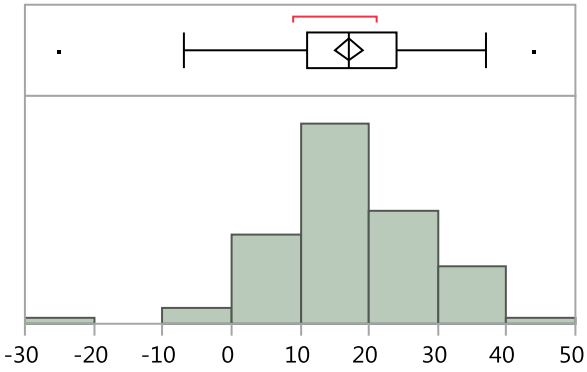
100.0%	maximum	41
99.5%		40
97.5%		32
90.0%		27
75.0%	quartile	23
50.0%	median	17
25.0%	quartile	11
10.0%		4
2.5%		-2
0.5%		-8.485
0.0%	minimum	-13

Summary Statistics

Mean	16.542
Std Dev	8.7826252
Std Err Mean	0.3927709
Upper 95% Mean	17.313689
Lower 95% Mean	15.770311
N	500

Distributions Medical_Group_Name=P

Change_in_OKS_Score_at_1_Year



Quantiles

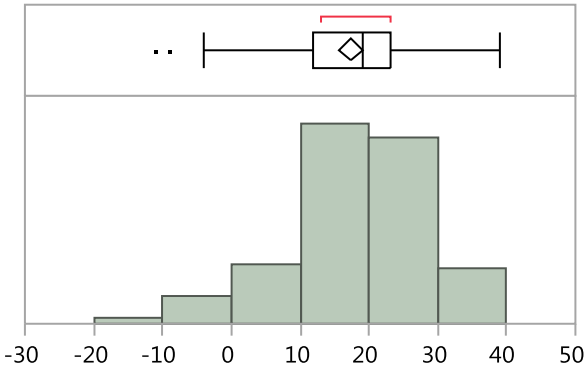
100.0%	maximum	44
99.5%		44
97.5%		36.4
90.0%		30
75.0%	quartile	24
50.0%	median	17
25.0%	quartile	11
10.0%		4.4
2.5%		-4.6
0.5%		-25
0.0%	minimum	-25

Summary Statistics

Mean	17.106796
Std Dev	10.440224
Std Err Mean	1.0287059
Upper 95% Mean	19.147229
Lower 95% Mean	15.066363
N	103

Distributions Medical_Group_Name=Q

Change_in_OKS_Score_at_1_Year



Quantiles

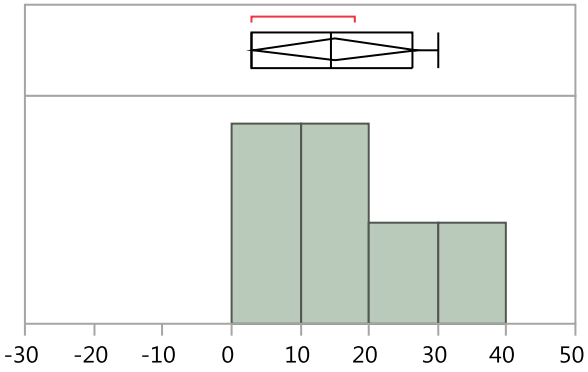
100.0%	maximum	39
99.5%		39
97.5%		34
90.0%		30
75.0%	quartile	23
50.0%	median	19
25.0%	quartile	12
10.0%		5
2.5%		-4
0.5%		-11
0.0%	minimum	-11

Summary Statistics

Mean	17.478992
Std Dev	9.5605552
Std Err Mean	0.8764147
Upper 95% Mean	19.214531
Lower 95% Mean	15.743452
N	119

Distributions Medical_Group_Name=R

Change_in_OKS_Score_at_1_Year



Quantiles

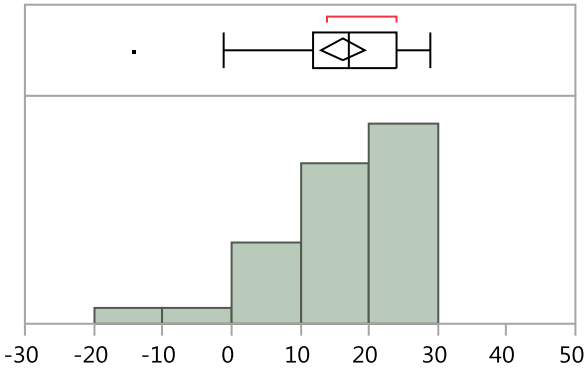
100.0%	maximum	30
99.5%		30
97.5%		30
90.0%		30
75.0%	quartile	26.25
50.0%	median	14.5
25.0%	quartile	3
10.0%		3
2.5%		3
0.5%		3
0.0%	minimum	3

Summary Statistics

Mean	15
Std Dev	11.296017
Std Err Mean	4.6115796
Upper 95% Mean	26.854443
Lower 95% Mean	3.1455572
N	6

Distributions Medical_Group_Name=S

Change_in_OKS_Score_at_1_Year



Quantiles

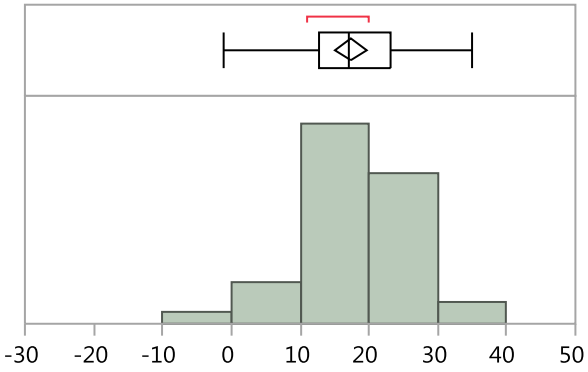
100.0%	maximum	29
99.5%		29
97.5%		29
90.0%		26.4
75.0%	quartile	24
50.0%	median	17
25.0%	quartile	12
10.0%		4.2
2.5%		-14
0.5%		-14
0.0%	minimum	-14

Summary Statistics

Mean	16.314286
Std Dev	9.1964553
Std Err Mean	1.5544847
Upper 95% Mean	19.473379
Lower 95% Mean	13.155193
N	35

Distributions Medical_Group_Name=T

Change_in_OKS_Score_at_1_Year



Quantiles

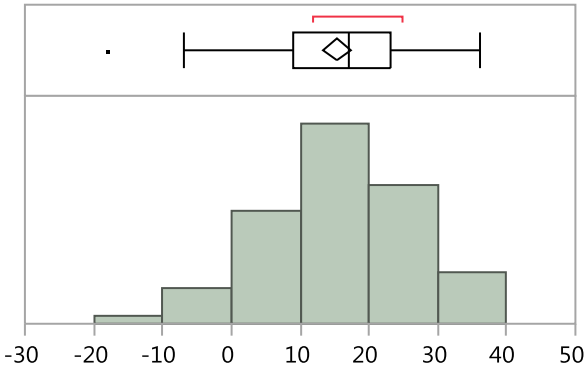
100.0%	maximum	35
99.5%		35
97.5%		34.7
90.0%		26
75.0%	quartile	23
50.0%	median	17
25.0%	quartile	12.75
10.0%		7.6
2.5%		-0.7
0.5%		-1
0.0%	minimum	-1

Summary Statistics

Mean	17.309524
Std Dev	7.5239069
Std Err Mean	1.160964
Upper 95% Mean	19.654138
Lower 95% Mean	14.964909
N	42

Distributions Medical_Group_Name=U

Change_in_OKS_Score_at_1_Year



Quantiles

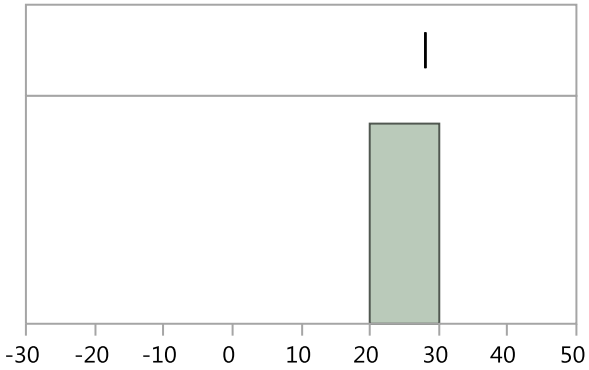
100.0%	maximum	36
99.5%		36
97.5%		33.65
90.0%		28.4
75.0%	quartile	23
50.0%	median	17
25.0%	quartile	9
10.0%		1
2.5%		-7
0.5%		-18
0.0%	minimum	-18

Summary Statistics

Mean	15.381443
Std Dev	10.277852
Std Err Mean	1.0435577
Upper 95% Mean	17.452889
Lower 95% Mean	13.309998
N	97

Distributions Medical_Group_Name=V

Change_in_OKS_Score_at_1_Year



Quantiles

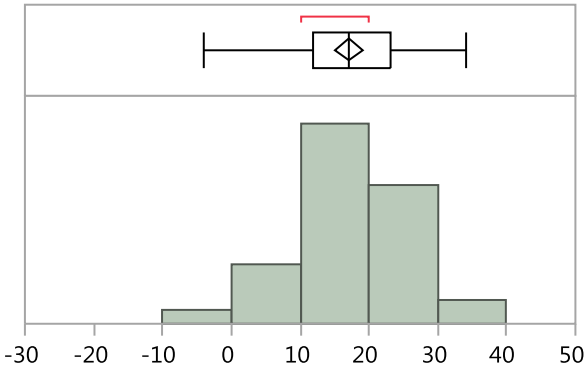
100.0%	maximum	28
99.5%		28
97.5%		28
90.0%		28
75.0%	quartile	28
50.0%	median	28
25.0%	quartile	28
10.0%		28
2.5%		28
0.5%		28
0.0%	minimum	28

Summary Statistics

Mean	28
Std Dev	.
Std Err Mean	.
Upper 95% Mean	.
Lower 95% Mean	.
N	1

Distributions Medical_Group_Name=W

Change_in_OKS_Score_at_1_Year



Quantiles

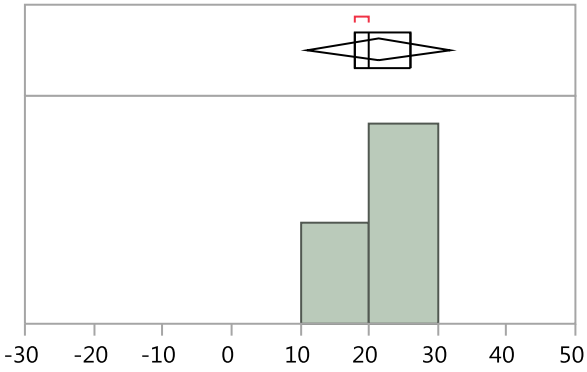
100.0%	maximum	34
99.5%		34
97.5%		34
90.0%		28
75.0%	quartile	23
50.0%	median	17
25.0%	quartile	12
10.0%		4
2.5%		-2.2
0.5%		-4
0.0%	minimum	-4

Summary Statistics

Mean	17.04
Std Dev	8.5272853
Std Err Mean	0.9846461
Upper 95% Mean	19.00195
Lower 95% Mean	15.07805
N	75

Distributions Medical_Group_Name=X

Change_in_OKS_Score_at_1_Year



Quantiles

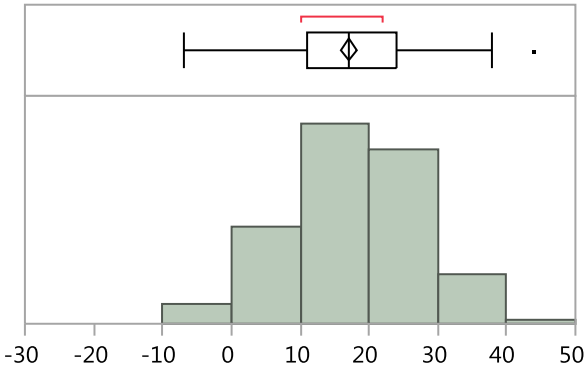
100.0%	maximum	26
99.5%		26
97.5%		26
90.0%		26
75.0%	quartile	26
50.0%	median	20
25.0%	quartile	18
10.0%		18
2.5%		18
0.5%		18
0.0%	minimum	18

Summary Statistics

Mean	21.333333
Std Dev	4.163332
Std Err Mean	2.4037009
Upper 95% Mean	31.675623
Lower 95% Mean	10.991043
N	3

Distributions Medical_Group_Name=Y

Change_in_OKS_Score_at_1_Year



Quantiles

100.0%	maximum	44
99.5%		42.8
97.5%		34
90.0%		29
75.0%	quartile	24
50.0%	median	17
25.0%	quartile	11
10.0%		4
2.5%		-2
0.5%		-6.8
0.0%	minimum	-7

Summary Statistics

Mean	17.058577
Std Dev	9.2111505
Std Err Mean	0.5958198
Upper 95% Mean	18.232331
Lower 95% Mean	15.884823
N	239

RTI INTERNATIONAL | Memorandum

Date: March 20, 2015

To: National Quality Forum Patient-and Family-Centered Care Project Panel Members and Staff

From: Tara McMullen, Centers for Medicare and Medicaid Services (CMS) and Anne Deutsch, RTI International on behalf of the Function Team

Subject: Responses to questions received during the in-person Person- and Family-Centered Care Phase 2 Standing Committee Meeting and the Follow-up Phone Meeting

Introduction

We are providing responses to questions and comments received about the LTCH and IRF functional status quality measures (Measure Steward: Centers for Medicare and Medicaid Services) that were presented and discussed during the Person- and Family-Centered Care Phase 2 Committee Meeting held on January 21st and January 22nd, 2015 and the follow-up phone call held on January 28, 2015.

First, we provide the requested information about the quality measures for which consensus for endorsement recommendation was not reached. We then provide additional information requested for the quality measures that the panel recommended for endorsement.

Quality Measures: Responses to questions and comments from the in-person and follow-up meetings.

Section 1.1 Percent of Long-Term Care Hospital (LTCH) Patients with an Admission and Discharge Functional Assessment and a Care Plan That Addresses Function (NQF #2631, under review)

Section 1.2: IRF Quality Measures

1. Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Self-Care Score for Medical Rehabilitation Patients (NQF #2633, under review)
2. Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Mobility Score for Medical Rehabilitation Patients (NQF #2634, under review)
3. Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Discharge Self-Care Score for Medical Rehabilitation Patients (NQF #2635, under review)
4. Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Discharge Mobility Score for Medical Rehabilitation Patients (NQF #2636, under review)

Quality Measures

1.1 Quality Measure: Percent of Long-Term Care Hospital (LTCH) Patients with an Admission and Discharge Functional Assessment and a Care Plan That Addresses Function (NQF #2631, under review)

This quality measure is the percentage of all Long-Term Care Hospital (LTCH) patients with an admission and discharge functional assessment and a goal that addresses function. The function goal provides evidence of a care plan that includes patient functioning.

During the in-person and follow-up meetings, panel members asked several questions about the relationship between the care plan and the functional assessment. We noted that clinicians typically conduct an admission functional assessment, and then write a care plan based on the functional assessment findings. Care plans should include a statement about the expected outcome, which can be reported as a discharge goal. The discharge goal would be established in collaboration with the patient and family.

The function goal is directly linked to the functional assessment findings. For example, if a patient required supervision (score 4) with eating on admission, the goal may be to increase the patient's functioning to a level of independence (score 6) by discharge. The quality measure requires that a goal be established for at least one self-care or mobility item, and the goal is reported using the same rating scale (level 6: Independent through level 1: Dependent) as the admission and discharge functional assessment items. One panel member highlighted that the requirement for the functional assessment and care plan/discharge goal to be linked greatly elevated the importance and relevance of the measure.

Table 1 shows an example of how the self-care items for this measure would be collected at the time of admission. For each self-care item, an admission performance score is reported. A goal for one or more self-care items can be entered for each item right next to the admission assessment score. Mobility admission and discharge goal items will be collected similarly.

Table 1. CARE Self-Care Function Items

Admission Score	Discharge Goal	CARE Function Item
4	6	A. **Eating:** The ability to use suitable utensils to bring food to the mouth and swallow food once the meal is presented on a table/tray. Includes modified food consistency.
4	5	B. **Oral hygiene:** The ability to use suitable items to clean teeth.
3	4	C. **Toileting hygiene:** The ability to maintain perineal hygiene, adjust clothes before and after using the toilet, commode, bedpan or urinal. If managing an ostomy, include wiping the opening but not managing equipment.
3	5	D. **Wash upper body:** The ability to wash, rinse, and dry the face, hands, chest, and arms while sitting in a chair or bed.

Importance of the Quality Measure
During the in-person meeting, the panel members asked many questions about the importance of this quality measure. After the follow-up phone call, the panel's vote for the measure was in the "gray zone." NQF considers the importance of a measure based on evidence, performance gap and priority.

Evidence:
Evidence supporting this quality measure includes three clinical practice guidelines (listed below) and a comprehensive literature review. We provided this information in both the Measure Information and Evidence Forms for this measure. The three clinical practice guidelines are:

I. Centre for Clinical Practice at NICE (UK). Rehabilitation after critical illness

- Centre for Clinical Practice at NICE (UK). (2009). *Rehabilitation after critical illness* (NICE Clinical Guidelines No. 83):91. Retrieved from http://www.nice.org.uk/guidance/CG83
- References and citations that support the NICE guidelines can be found at the following location: http://www.nice.org.uk/guidance/cg83/resources/cg83-critical-illness-rehabilitation-guideline2

II. National Guideline Clearinghouse. Assessing cognitive functioning. In: Evidence-based geriatric nursing protocols for best practice.

- Milisen, K., Braes, T., & Foreman, M. D. (2012). Assessing cognitive function. In M. Boltz, E. Capezuti, T. Fulmer, & D. Zwicker (Eds.), Evidence-based geriatric nursing protocols for best practice (4th ed., pp. 122–134). New York, NY: Springer.
- Adapted from: Melnyck, B. M. & Fineout-Overholt, E. (2005). Evidence-based practice in nursing & health care: A guide to best practice. Philadelphia, PA: Lippincott Williams & Wilkins and Stetler, C.B., Morsi, D., Rucki, S., Broughton, S., Corrigan, B., Fitzgerald, J., et al. (1998). Utilization-focused integrative reviews in a nursing service. Applied Nursing Research, 11(4).

III. Comprehensive assessment and management of the critically ill. In: Evidence-based geriatric nursing protocols for best practice.

- Balas MC, Casey CM, Happ MB. Comprehensive assessment and management of the critically ill. In Boltz M, Capezuti E, Fulmer T, Zwicker D, editor(s). Evidence-based geriatric nursing protocols for best practice. (4th ed., pp. 600-27). New York, NY: Springer.
- Adapted from: Melnyck, B. M. & Fineout-Overholt, E. (2005). Evidence-based practice in nursing & health care: A guide to best practice. Philadelphia, PA: Lippincott Williams & Wilkins and Stetler, C.B., Morsi, D., Rucki, S., Broughton, S., Corrigan, B., Fitzgerald, J., et al. (1998). Utilization-focused integrative reviews in a nursing service. Applied Nursing Research, 11(4).

Performance Gap:

One NQF panel member requested data on the proportion of patients with an incomplete stay, for whom discharge functional assessment data would not be required. Five hundred seventy-three patients (20.01%) in our sample had an incomplete stay, and discharge functional assessment data were not required for these patients. The NQF panel members also requested performance gap data for the measure. Accordingly, we have run additional facility-level analyses related to performance gap at the measure level.

The data we present below describe the percent of patients with admission and discharge functional assessment data submitted from the 28 LTCHs that volunteered to participate in the Post-Acute Payment Reform Demonstration. Among the 28 LTCHs, the percent of patients with complete admission and discharge assessment data submitted was high, because:

- These LTCH facilities volunteered to participate in the demonstration and signed an agreement to provide complete data for the period of the study.
- The standardized functional assessments used in this demonstration were submitted by users via an electronic application that notified the user when an item was left blank. This design function within the application did not allow the user to advance to the next item/page without entering appropriate data; thus, increasing assessment item completion rates.

We would like to note that during site visits for the Post-Acute Care Payment Reform Demonstration, the research team noted substantial variation in the collection of functional assessment data across LTCHs. We observed that different LTCHs used different functional assessment instruments, which varied in terms of function item, item definitions and response coding.

When calculating the quality measure, we used the admission and discharge functional assessment data; however, we did not have data on the goals (care plan) related to function. Thus, the measure results presented below are based on the percent of patients who had admission and discharge functional assessment data.

Consistent with the measure specifications, complete admission functional assessment data was required for all patients (n = 2,864) and complete discharge functional assessment data was required for patients with a complete stay (n = 2,291). **Table 2** and **Figure 1** show the distribution of the provider-level data for the percent of LTCH patients with complete admission and discharge functional assessment data. The mean percent was 92.42 ± 14.73, while the median percent was 98.25. Some variability in measure performance was seen, with the minimum percent being 31.63, and 10th percentile being 78.90. As noted above, the data are from LTCHs that volunteered to participate in the demonstration and agreed to provide complete data. In addition, the data entry system used during the demonstration made it difficult to leave items blank (missing).

Table 2. Percent of LTCH Patients with an Admission and Discharge Functional Assessment.

Quality Measure	Mean	SD	Min	1st pct	10th pct	25th pct	Median	75th pct	90th pct	Max
Percent of LTCH Patients with an Admission and Discharge Functional Assessment	92.42	14.73	31.63	40.95	78.90	91.75	98.25	100	100	100

The measure requires complete admission functional assessment data for all patients; complete discharge functional assessment data are only required for 2,291 patients with a complete stay.
Pct = percentile.

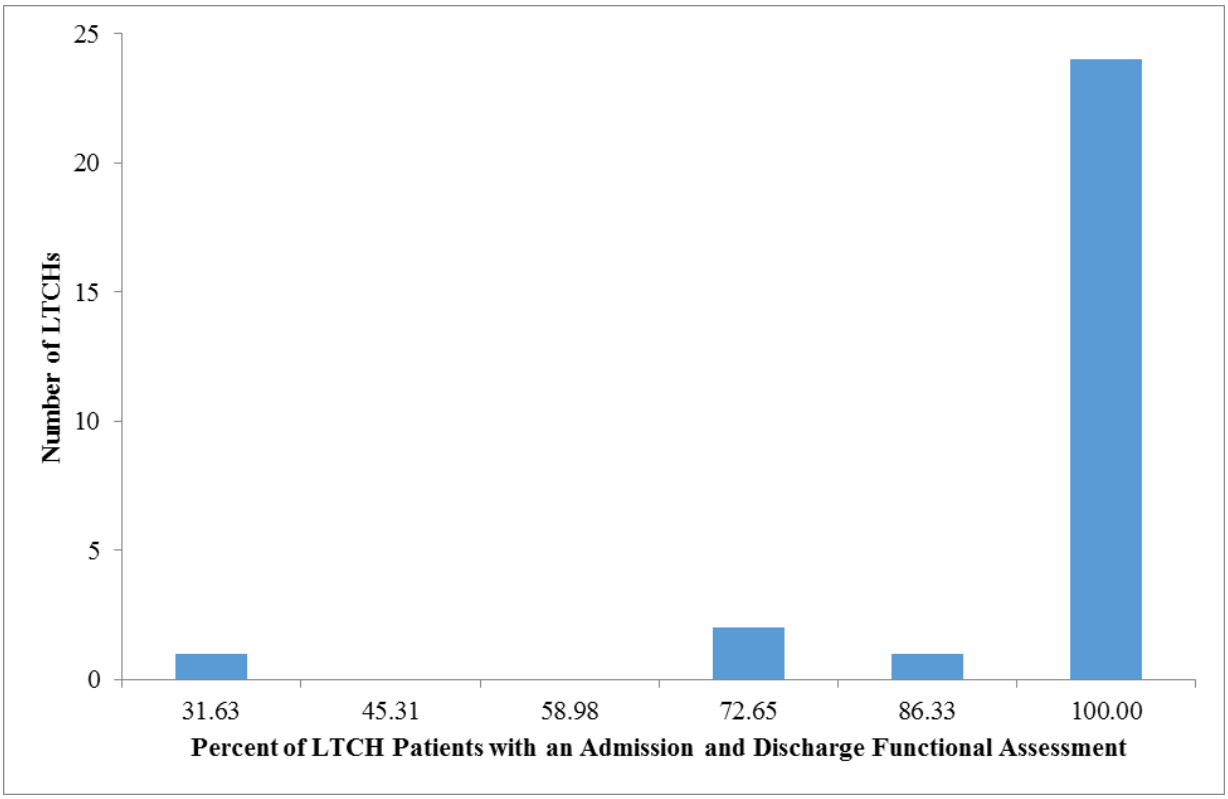


Figure 1. Percent of LTCH Patients with an Admission and Discharge Functional Assessment. The measure specifications require complete admission functional assessment data for all patients; complete discharge functional assessment data are only required for 2,291 patients with a complete stay.

Priority

In addition to the evidence provided above, we provide comments made by panel members during the in-person and follow-up meetings about the importance of this measure. These quotes are copied from the meeting transcript, and we provide a page for each comment.

NQF Panel Member Dr. Cella (page 328 of day 1) "Anyway, we're still at the level of importance and I, you know, I'll stop talking because I think it's good that there's something in this area because it's such a tough area. So I'm high on importance." [In a later part of the conversation, Dr. Cella said in response to finding out the care plan was linked] "Okay. Well, in my mind that actually elevates the importance as opposed to be...I'm glad to hear that there is a link because that strengthens it."

NQF Panel Member Dr. Bierner (page 328-329 of day 1 transcript): I just want to point out that more and more patients are getting discharged to this type of facility with pressures on the acute care side to move people into other alternatives. And so we're seeing -- this is becoming a bigger and bigger discharge disposition for a lot of acute care patients with wounds, with ventilators, and other medically complex problems.

NQF Co-Chair Dr. Stille (page 333 of day 1 transcript): I just had a comment about -- we had a fairly extensive discussion about care plan this morning and how reliable data were in detecting a care plan. One thing that's actually sort of nice about this is that there's a measure that links having something in the care plan that's related to the assessment, which wasn't in this morning. But I think, you know, data about how possible is it to measure both of those is going to be really important to look at value.

Missing Data

NQF panel members requested information on missing data for functional assessment items in the quality measure, Percent of Long-Term Care Hospital (LTCH) Patients with an Admission and Discharge Functional Assessment and a Care Plan That Addresses Function (NQF #2631, under review). Therefore, we updated the missing data information (see **Table 3**). Missing admission functional assessment data were computed for all patients, whereas missing discharge functional assessment data were computed for the 2,291 patients who had a complete stay. Items with the largest proportions of missing data were 'wash upper body', 'roll left and right', 'sit to lying', and 'walk 50 feet with two turns;' for each of these items, the percent of missing data was greater at discharge than at admission.

Table 3. Frequency and Percent of Missing Data for LTCH Patients Include in the LTCH Process Quality Measure (NQF #2631)

Items	LTCH: Admission and Discharge		LTCH Admission (n = 2,864)		LTCH Discharge for Complete Stays (n = 2,291)	
	Number missing	Percent missing	Number missing	Percent missing	Number missing	Percent missing
Eating	12	0.42%	+	+	+	+
Oral hygiene	12	0.42%	+	+	+	+
Toilet hygiene	12	0.42%	+	+	+	+
Wash upper body	303	10.58%	153	5.34%	150	6.55%
Roll left and right	303	10.58%	153	5.34%	150	6.55%
Sit to lying	303	10.58%	153	5.34%	150	6.55%
Lying to sitting on side of bed	+	+	+	+	+	+
Sit to stand	+	+	+	+	+	+
Chair/bed-to-chair transfer	+	+	+	+	+	+
Toilet transfer	+	+	+	+	+	+
Walk/Wheelchair items	+	+	+	+	+	+
Walk 50 feet with two turns	301	10.51%	153	5.34%	148	6.46%
CAM inattention	+	+	+	+	N/A	N/A
CAM disorganized thinking	+	+	+	+	N/A	N/A
CAM altered level of consciousness	+	+	+	+	N/A	N/A
Understanding verbal content	11	0.38%	+	+	+	+
Expression of ideas/wants	11	0.38%	+	+	+	+
Bladder incontinence frequency	25	0.87%	+	+	17	0.74%

*Missing admission functional assessment data were computed for all patients, whereas missing discharge functional assessment data were only computed on 2,291 patients who had a complete stay.

+ Cells based on a sample size of n < 11 are not shown. Our data use agreement does not allow us to report data when there are fewer than 11 cases.

N/A: In the Post-Acute Care Payment Reform Demonstration, data on CAM items was only collected on admission.

Activity Not Attempted – Percent of time occurring for the CARE Items in LTCHs

During the in-person meeting, one panel member requested information about the frequency of the use of the "activity not attempted" code for items in the LTCH process quality measure due to patient refusal or for other reasons.

The report, *Analysis of Crosscutting Medicare Functional Status Quality Metrics Using the Continuity and Assessment Record and Evaluation (CARE Item Set): Final Report*, includes these data in Appendices A and B. Appendix A shows graphical displays of the distribution of scores for each CARE self-care and mobility item on admission and discharge by type of provider (LTCH, Inpatient Rehabilitation Facility, Skilled Nursing Facility and Home Health Agency). These 100 percent bar charts show the percent of patients by score and the percent coded "the activity did not occur." The light purple bar labeled 'Did not occur' represents all of the Activity Did Not Occur/Not Attempted codes. Appendix B contains the quantitative data used for the graphical displays.

The report, Analysis of Crosscutting Medicare Functional Status Quality Metrics Using the Continuity and Assessment Record and Evaluation (CARE Item Set). Final Report. 2012 can be found at: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/ASPE-Report-Analysis-of-Crosscutting-Medicare-Functional-Status-Quality-Metrics-Using-the-Continuity-and-Assessment-Record-and-Evaluation-CARE-Item-Set-Final-Report.pdf

Measure Specifications: Numerator and Denominator

Based on the number of questions we received about the specifications of this quality measure, we thought it was important to describe the denominator and numerator. During the follow-up phone call for the LTCH Process quality measure discussion, we presented a simplified description of the quality measure, which is repeated here:

This quality measure has 2 components:

1) the collection of standardized functional assessment data in the areas of self-care, mobility, cognition and bladder management, and
2) the reporting, on admission, of a discharge goal (i.e., score) for one or more self-care or mobility items.

The denominator is all LTCH patients in a facility. There are no exclusion criteria for this quality measure, because there are codes that can be used to indicate the reason an activity was not attempted.

To be included in the numerator:

- Admission data are required for all function items and all patients
- At least one self-care or mobility item must have a discharge goal documented at admission for all patients
- Discharge data are required for all patients with complete stays. Discharge data are not required for patients who have incomplete stays, as it is challenging to collect accurate data during an unexpected discharge.

Patients who have incomplete stays are defined as those patients (1) with incomplete stays due to a medical emergency, (2) who leave the LTCH against medical advice, or (3) who die while in the LTCH. Discharge functional status data are not required for these patients because these data may be difficult to collect at the time of the medical emergency, if the patient dies, or if the patient leaves against medical advice.

1.2 IRF Quality Functional Change Measures: Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Self-Care Score for Medical Rehabilitation Patients (NQF #2633, under review) and Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Mobility Score for Medical Rehabilitation Patients (NQF #2634, under review)

Priority

In response to the quality measures being submitted and presented separately as Change and Discharge quality measures:

NQF Panel Member Dr. Beirner (day 2 transcript, page 176): I don't have a concern with it. I think it will be very useful for us, as an institution, to report to our consumers, the clients that we seek to attract, to present it in the percentage format or the benchmark kind of format for public purposes and then the other format for internal or for reporting purposes.

Performance Gap and Ability to Discriminate Among Facilities

For the IRF Change in Self-Care Score measure (NQF #2633, under review) and IRF Change in Mobility Score measure (NQF #2634, under review), we examined whether individual facility performance was worse than, better than, or the same as national average performance. For each facility, we calculated the 95% confidence interval for the risk-adjusted change score, and compared this with the national mean observed change score. Facilities whose confidence interval was lower than the national mean observed change score were considered to have worse performance than the national average. Facilities whose confidence interval was higher than the national mean observed change score were considered to have better performance than the national average. Facilities whose confidence interval overlapped with the national mean observed change score were considered to be similar to national average performance.

Table 4 shows that for the IRF Change in Self-Care Score measure (NQF #2633, under review), 26.3% of IRFs had 95% confidence intervals lower than the national mean change score, indicating worse than national average performance. For the IRF Change in Mobility Score measure (NQF #2634, under review), 34.2% of IRFs had worse than national average performance. For both measures, only a small proportion of IRFs had 95% confidence intervals higher than the national mean change score, indicating better than national average performance. These findings demonstrate the ability of the measures to discriminate among facilities based on facility-level measure performance. The findings also support a performance gap across facilities, justifying the importance of these measures. Figures 2 and 3 show the distribution of facility-level risk-adjusted change scores as compared with the national mean observed change score.

Table 4. Comparison of Facility-Level Measure Scores with National Average Performance for IRF Change Score Measures (NQF #2633 and NQF#2634, under review)

Measure Name	Facility Performance Worse than National Average [N (%)]	Facility Performance Better than National Average [N (%)]	Facility Performance Same As National Average [N (%)]
Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Self-Care Score for Medical Rehabilitation Patients (NQF #2633, under review)	10 (26.3%)	7 (18.4%)	21 (55.3%)
Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Mobility Score for Medical Rehabilitation Patients (NQF #2634, under review)	13 (34.2%)	6 (15.8%)	19 (50%)

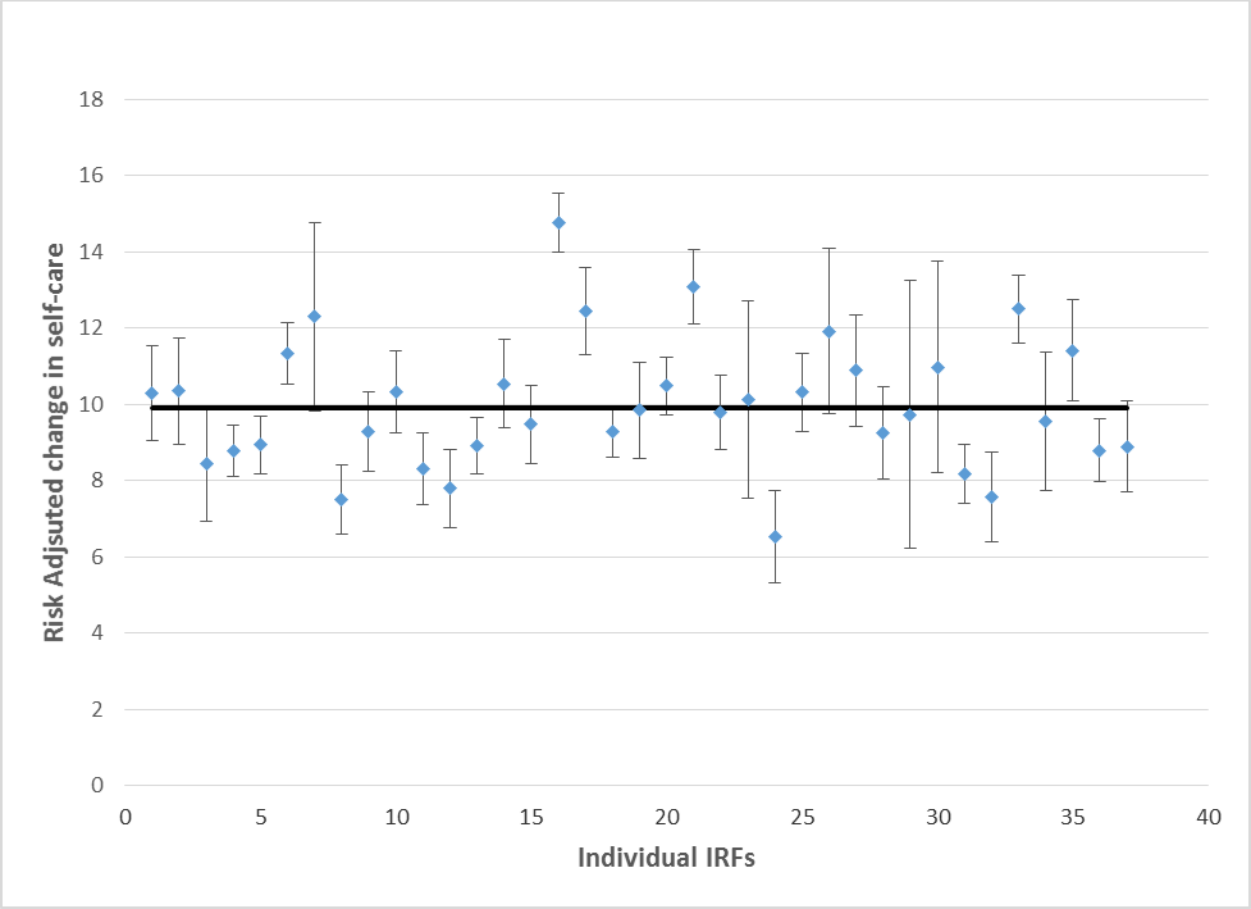


Figure 2. Facility-Level Risk-Adjusted Change in Self-Care Scores as compared with the National Mean Observed Change in Self-Care Score. The Y-axis shows the mean risk-adjusted self-care change score and 95% confidence interval for each IRF. The dark horizontal line represents the national mean observed self-care change score.

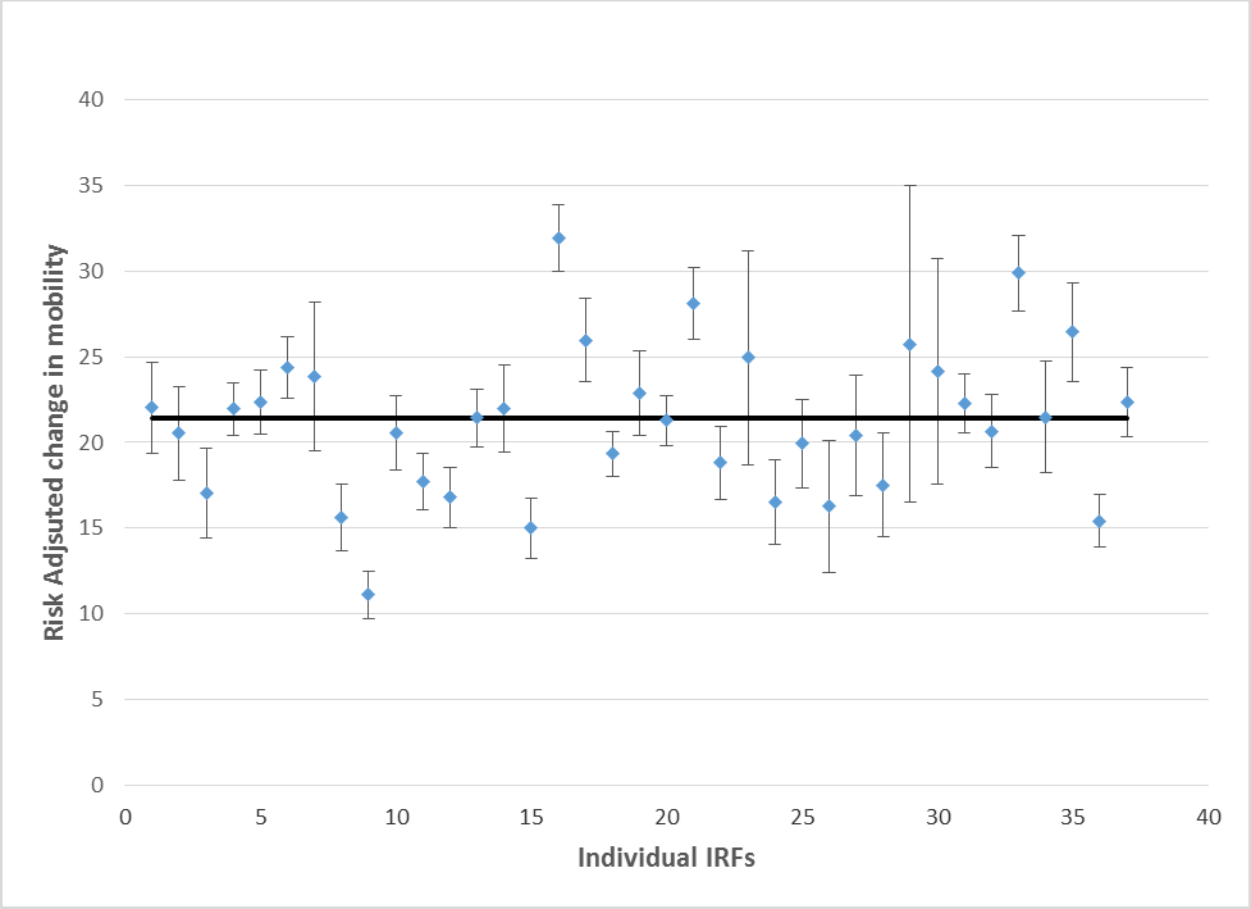


Figure 3. Facility-Level Risk-Adjusted Change in Mobility Scores as compared with the National Mean Observed Change in Mobility Score. The Y-axis shows the mean risk-adjusted mobility change score and 95% confidence interval for each IRF. The dark horizontal line represents the national mean observed mobility change score.

For the IRF Discharge Self-Care Score measure (NQF #2635, under review) and IRF Discharge Mobility Score measure (NQF #2636, under review), we examined whether individual facility performance was worse than, better than, or same as national performance. For each facility, we calculated the 95% confidence interval of the proportion of patients who meet or exceed the expected threshold, and compared the confidence interval with the national proportion of patients who meet or exceed the expected threshold. Facilities whose confidence interval was lower than the national proportion were considered to have worse than national performance. Facilities whose confidence interval was higher than the national proportion were considered to have better than national performance. Facilities whose confidence interval overlapped with the national proportion were considered to be similar to national performance.

Table 5 shows that for the IRF Discharge Self-Care Score measure (NQF #2635, under review), an equal proportion of facilities (28.9%) had 95% confidence intervals below and above the national proportion, indicating an equal proportion had performance worse than and better than

national performance. For the IRF Discharge Mobility Score measure (NQF #2636, under review), only 18.4% of IRFs had 95% confidence intervals above the national proportion, indicating better than national performance. These findings demonstrate the ability of the measures to discriminate among facilities based on facility-level measure performance. The findings also support a performance gap across facilities, justifying the importance of these measures.

Table 5. Comparison of Facility-Level Measure Scores with National Performance for IRF Discharge Score Measures (NQF #2635 and NQF#2636, under review)

Measure Name	Facility Performance Worse than National Performance [N (%)]	Facility Performance Better than National Performance [N (%)]	Facility Performance Same as National Performance [N (%)]
Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Discharge Self-Care Score for Medical Rehabilitation Patients (NQF #2635, under review)	11 (28.9%)	11 (28.9%)	16 (42.1%)
Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Discharge Mobility Score for Medical Rehabilitation Patients (NQF #2636, under review)	12 (31.6%)	7 (18.4%)	19 (50.0%)

Figures 4 and 5 show the distribution of facility-level measure scores as compared with the national proportion of patients who meet or exceed the threshold (expected).

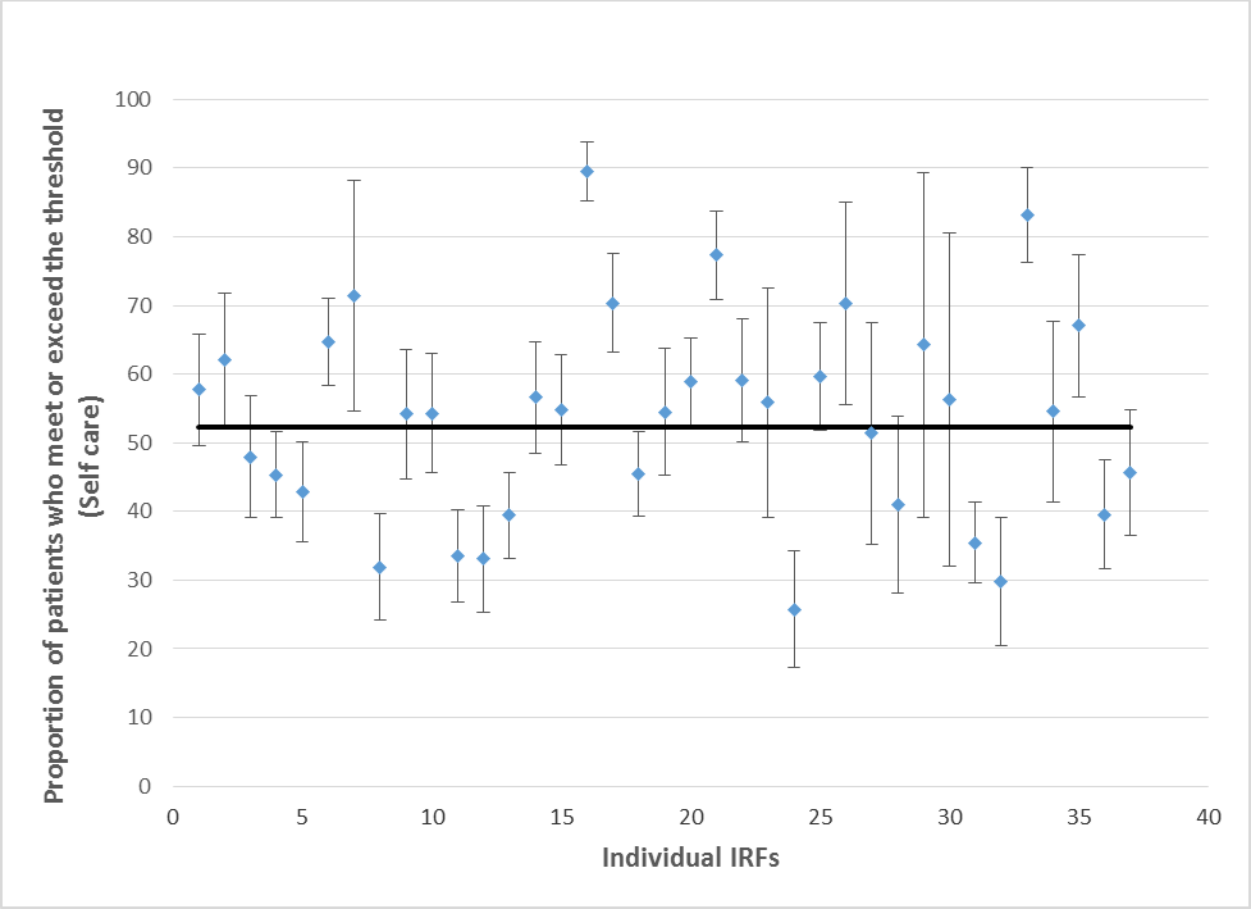


Figure 4. Facility-Level Proportion of Patients Who Meet/Exceed the Discharge Self-Care Threshold (Expected) as compared with the National Proportion of Patients Who Meet/Exceed the Threshold (Expected). The Y-axis shows the proportion of patients who meet/exceed the discharge threshold (expected) and 95% confidence interval for each IRF. The dark horizontal line represents the national proportion of patients who meet/exceed the discharge threshold (expected).

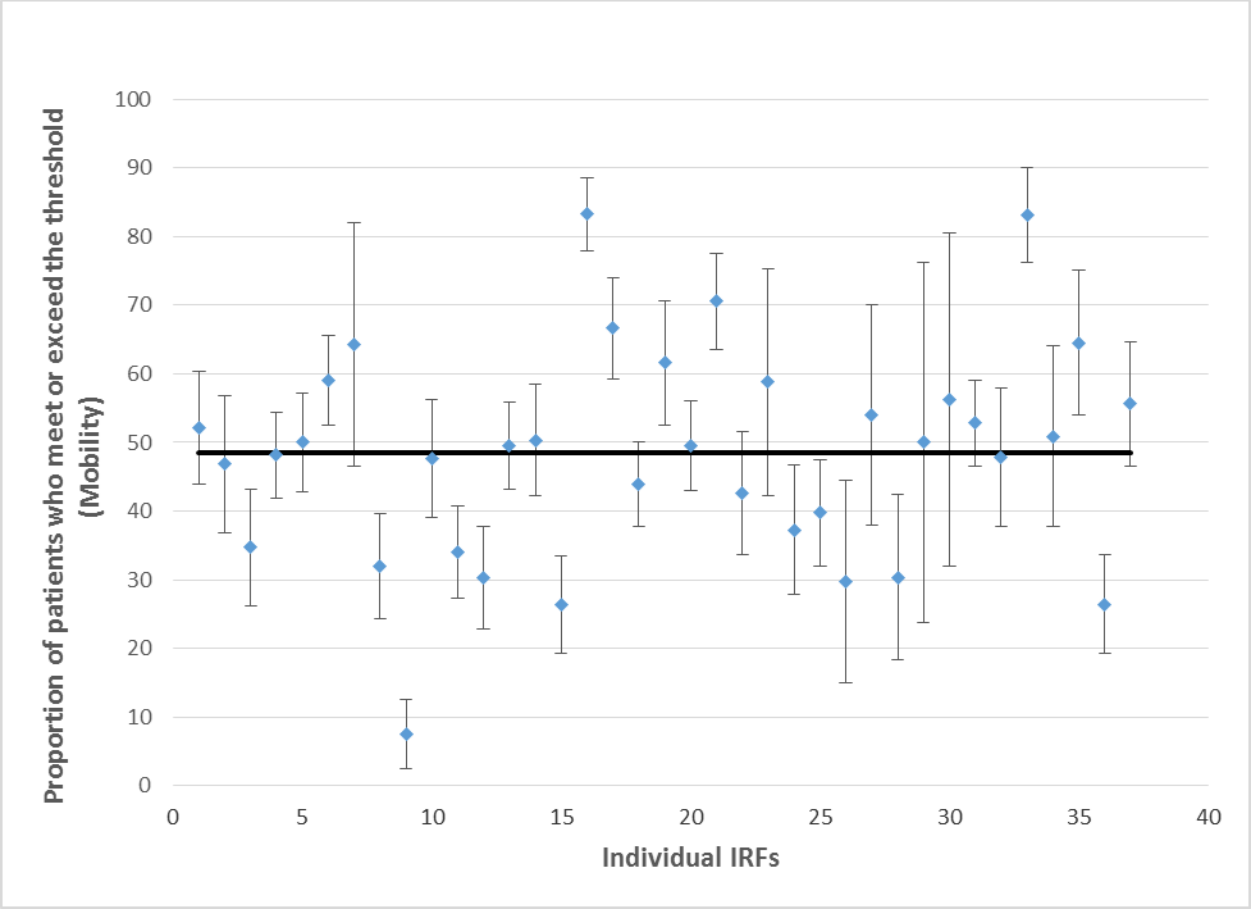


Figure 5. Facility-Level Proportion of Patients Who Meet/Exceed the Discharge Mobility Threshold (Expected) for Individual IRFs as compared with the National Proportion of Patients Who Meet/Exceed the Threshold (Expected). The Y-axis shows the proportion of patients who meet/exceed the discharge threshold (expected) and 95% confidence interval for each IRF. The dark horizontal line represents the national proportion of patients who meet/exceed the discharge threshold (expected).

Reliability

NQF panel members requested data on the reliability of the IRF quality measures for medical rehabilitation patients at the quality measure level. One panel member specifically requested data on intra-class coefficients at the quality measure level. Accordingly, we conducted split-half reliability testing for the four IRF quality measures, and present both significance testing results and intra-class correlation coefficients.

Split-half reliability: Significance Testing and Intra-Class Correlation Coefficients

We conducted split-half reliability testing to assess reliability of the four IRF quality measures for medical rehabilitation patients using facility-level and measure-level performance scores. For each facility that had a sample size of 100 or greater, we randomly split the facility sample into two halves, and recalculated each quality measure based on the split halves; thus each facility had two scores for each quality measure. The purpose of the split-half testing was to examine

whether facility performance would be similar and measure scores would be reliable when calculated using data from two different sets of patients. Twenty-six facilities were included in split-half reliability testing. For each facility, we ran an independent *t*-test to determine whether there was a difference in the two scores for each measure based on the randomly split halves. Using a 0.05 significance level, no significant difference in split-half measure scores was noted for any of the 26 facilities for any of the four IRF quality measures. These findings support the reliability of four IRF measures at the facility-level and measure-level.

Using the split-half scores, we also computed intra-class correlation coefficients for each quality measure. **Table 6** shows that both ICC (2,1) and ICC (3,1) for each measure approximated or exceeded 0.9, indicating very high reliability at the facility-level and measure-level.

Table 6. Intraclass Correlation Coefficients for the Four IRF Function Quality Measures for Medical Rehabilitation Patients

Measure Name	ICC (2,1)	ICC (3,1)
Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Self-Care Score for Medical Rehabilitation Patients (NQF #2633, under review)	0.89	0.90
Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Change in Mobility Score for Medical Rehabilitation Patients (NQF #2634, under review)	0.93	0.93
Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Discharge Self-Care Score for Medical Rehabilitation Patients (NQF #2635, under review)	0.93	0.93
Inpatient Rehabilitation Facility (IRF) Functional Outcome Measure: Discharge Mobility Score for Medical Rehabilitation Patients (NQF #2636, under review)	0.95	0.95

FOCUS ON THERAPEUTIC OUTCOMES INC

PO Box 11444
Knoxville,TN 37939
p: 1.800.482.3686
f: 865.450.9484

March 19, 2015

Dear Members of the Person and Family-Center Care Standing Committee,

Thank-you very much for the opportunity to provide additional information on FOTO's seven functional status PROM-PM measures, numbers 0422 to 0428. Members of FOTO's research advisory board have prepared these comments after carefully reviewing the transcripts of the Person and Family-Centered Care Phase 2 Standing Committee meeting and considering the committee members' feedback and requests for additional information, clarification and analyses. In the enclosed document we provide the following:

1. Additional gap analyses to illustrate the relationship between treatment interventions and outcomes and participation in the FOTO measurement program and quality improvement.
2. We present revised measure specification descriptions, which we believe are clearer.
3. We have made the decision to withdraw our request to report the lower limits of age inclusion at 14 years old. The lower limit of age is now 18 years old for all FOTO PROM-PMs.
4. We provide descriptive data on all patients who have completed the General Orthopaedic Measure (#0428). Our original application included only patients with cervical impairments. We have revised all key analyses presented in the application materials for this measure using the full group of patients including cervical, thoracic, rib, and Craniofacial impairments and show the revised tables in this document.
5. We present the results of new analyses conducted at the committee's request to examine provider reliability analyses for all FOTO PROM-PMs.
6. We present the results of new analyses conducted at the committee's request to examine validity of provider classification.
7. We present the results of new analyses conducted at the committee's request to examine components of variance analysis of discharge functional status change.
8. We also respond to the committee requests for data on the relationship between intensity and frequency of therapy visits and functional status change for all FOTO PROM-PM.
9. We provide more information on our risk adjustment models including beta coefficients for variables of the model as well as estimates of marginal means of discharge functional status by gender, age and payer group for each of the FOTO measures.

We thank the committee for their time, thoughtfulness and thoroughness in reviewing our applications and considering these additional comments. If accepted by the committee, FOTO expects to revise all application materials, including the measure specific web pages, to be consistent with the changes noted in these comments.

We very much hope that these additional materials satisfy the NQF requirements for endorsement.

Sincerely,

Ben E. Johnston, Jr. CEO

Focus On Therapeutic Outcomes
Comments on Measures 0422-0428
March 20, 2015

Table of Contents

1. Gap Analyses

The Committee stated that they were not clear about the connection between functional status outcomes and treatment interventions or how collecting the information on patient functional status outcomes and provider performance drives improvement. In this section, we present new information on the relationship between interventions and functional status outcomes and resubmit some supplemental materials that augment the rationale for the measures and show the impact of data collection over time. These supplemental materials were initially submitted prior to the in-person meeting in January, 2015. However, FOTO recognizes that some committee members may not have had sufficient time to review these materials prior to the meeting.

1a.Relationship between treatment interventions and outcomes Supplement to Measure Evidence Form 2a.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

The relationship between patient outcomes and treatment processes and interventions was clearly shown by Deutscher et al in a large prospective, observational cohort study, using data from 22, 019 patients (age mn 51.2y, sd 15.7, 58% women) seeking treatment due to lumbar spine, knee, cervical spine, or shoulder impairments from any of Maccabi Healthcare Services 54 community based outpatient physical therapy (PT) clinics in 2005-2008.[1] Functional status (FS) data were collected at intake and discharge (DC) from therapy, using FOTO's body part–specific CATs (PROMs). Associations between demographic and health characteristics at intake and treatment process variables with DC FS were evaluated using multivariable linear regression. Results are shown in Table 1a. After controlling for patient characteristics, the following treatment processes were found to be statistically significantly associated with DC FS: good compliance with attendance and home exercise program (for all impairments); waiting time between referral and initiation of PT (lumbar impairments). The following treatment interventions were positively associated with DC FS: joint mobilization (cervical and knee), stabilization exercises (lumbar), proprioceptive exercises (knee), passive movements (shoulder), group exercise (cervical), and stretching exercises (shoulder). The following treatment interventions were negatively associated with DC FS: shortwave therapy (knee, shoulder), therapeutic ultrasound (shoulder); cold packs (knee), group exercise (knee and shoulder), neural mobilization (shoulder). Although this study did not examine hip, ankle/foot, elbow wrist and hand patients, we believe that the study's results provide initial evidence of the relationship between treatment processes, interventions and patient outcomes.

	Lumbar		**Knee**		**Cervical**		**Shoulder**	
	Beta	P	Beta	P	Beta	P	Beta	P
Processes								
Attendance compliance good	2.1	<.001	2.5	<.001	2.9	<.001	2.5	<.001
Home exercise compliance good	3.0	<.001	6.6	<.001	3.1	<.001	4.2	<.001
Waiting time from referral to evaluation								
< 7 days (reference)								
7.1-14 days	-1.3	.002						
14.1-30 days	-0.7	.041						
Interventions								
Cold pack			-0.5	.017				
Consultation	-0.8	.019						
Electrotherapy for pain	-0.2	.001	-0.2	.001			-0.2	.001
Group exercise			-0.2	.006	0.2	.008	-0.2	.001
Joint mobilization			0.1	.045	0.2	.001		
Neural mobilization							-0.5	.014
Passive movements in supine or prone							0.2	0.24
Proprioceptive exercises			0.4	.025				
Reassessments	-0.3	.007						
Shortwave diathermy			-.3	<.001			-.3	<.001
Stabilization exercises	0.6	.004						
Stretching exercises							0.4	.014
Ultrasound							-.2	<.001

Table 1a. Relationships between treatment processes and interventions and discharge functional status for four patient impairment groups

*Beta indicates the coefficient that represents the amount of expected change in discharge FS given a 1-unit change in the value of the variable, given that all other variables in the model are held constant.

1. *Deutscher D, Horn SD, Dickstein R, Hart DL, Smout RJ et al. Associations Between Treatment Processes, Patient Characteristics, and Outcomes in Outpatient Physical Therapy Practice. APMR 2009;90:1349-1363.*

1b. Impact of the PRO-PM Over Time

Purpose:
To examine the impact of participation in the FOTO PROM-PM program on outcomes of clinicians over time.

Methods:
We compared clinician performance over 3 years from 2011-2013 using 2 samples: Sample 1: included all clinicians with a minimum of 10 patients/year. Sample 2 included only those clinicians who had a minimum of 10 patients/year who had participated in FOTO for each of three years.

Results
Results of clinician performance for each of the FOTO measures are shown in tables 1b1-1b7 below. In summary, for the 3 years examined, these data show 1) a dramatic growth (range 89-222%) in the number of clinicians who were eligible for FOTO quality measurement program (using threshold of 10 patients/clinician) for all measures and 2) overall improvements in clinicians' performance over time for clinicians participating in FOTO quality measurement program over time for each of three years. These improvements were particularly apparent for measures where there were greater number of clinicians participating. For the lumbar measure (675 clinicians) there was an 18% decrease in low performance level. For the shoulder (564 clinicians), knee (615 clinicians), general orthopaedic (263 clinicians), hip (75 clinicians), foot and ankle (207 clinicians), and elbow wrist and hand (130 clinicians) measures there was a 71%, 21%, 18%, 75%, 16%, and 60% increase in the number of high performance clinicians respectively.

Discussion/Conclusion:
These analyses show that participation in the FOTO quality measurement program was associated with improved clinician performance over time.

These analyses were conducted using the 2013 threshold of 10 patients per clinician per year for each measure. FOTO's recommended thresholds for minimum patients per clinician have recently changed for all but the lumbar, the elbow, wrist and hand, and the general orthopaedic measures as a result of the reliability analyses at the provider level which are reported in Table 5a on page 30. We did not update these analyses using the newer thresholds, in part, because for some measures there were too few providers (clinicians) that met the more rigorous thresholds that had participated for all three years. Participation in the FOTO database has grown steadily in the past three years, meaning that many new clinics and clinicians have begun participating in recent years. Given, that smaller numbers of patients per clinician are associated with greater "noise", we believe that the higher patients per clinician/year thresholds that FOTO plans to use moving forward will result in lower measurement error and thus will show greater differences by year. Additionally, since the reliability at the provider level is also sample dependent, once larger samples are available; the thresholds of number of patients per clinician per year needed to reach acceptable reliability at the provider level might decrease.

Lumbar - all clinicians		
Year	Performance level	N Clinicians (%)
2011	Low performance	153 (10.4)
	Average performance	1028 (69.6)
	High performance	297 (20.1)
	Total	1478 (100)
2012	Low performance	207 (11.5)
	Average performance	1251 (69.7)

Lumbar - same clinicians over time		
Year	Performance level	N Clinicians (%)
2011	Low performance	73 (10.8)
	Average performance	451 (66.8)
	High performance	151 (22.4)
	Total	675 (100.0)
2012	Low performance	67 (9.9)
	Average performance	454 (67.3)

	High performance	338 (18.8)
	Total	1796 (100)
2013	Low performance	372 (11.9)
	Average performance	2317 (73.8)
	High performance	450 (14.3)
	Total	3139 (100)

	High performance	154 (22.8)
	Total	675 (100.0)
2013	Low performance	60 (8.9)
	Average performance	465 (68.9)
	High performance	150 (22.2)
	Total	675 (100.0)

Table 1b1. Clinician Performance: Patients with Lumbar Impairments

Shoulder - all clinicians	
Performance level	**N Clinicians (%)**
Low performance	228 (18.4)
Average performance	813 (65.8)
High performance	195 (15.8)
Total	1236 (100.0)
Low performance	217 (14.3)
Average performance	1057 (69.6)
High performance	244 (16.1)
Total	1518 (100.0)
Low performance	192 (7.2)
Average performance	1987 (74.4)
High performance	491 (18.4)
Total	2670 (100.0)

Shoulder - same clinicians over time		
Year	**Performance level**	**N Clinicians (%)**
2011	Low performance	110 (19.5)
	Average performance	370 (65.6)
	High performance	84 (14.9)
	Total	564 (100.0)
2012	Low performance	80 (14.2)
	Average performance	398 (70.6)
	High performance	86 (15.2)
	Total	564 (100.0)
2013	Low performance	28 (5.0)
	Average performance	392 (69.5)
	High performance	144 (25.5)
	Total	564 (100.0)

Table 1b2 Clinician Performance: Patients with Shoulder Impairments

Hip - all clinicians		
Year	**Performance level**	**N Clinicians (%)**
2011	Low performance	26 (10.3)
	Average performance	201 (79.8)
	High performance	25 (9.9)
	Total	252 (100.0)
2012	Low performance	39 (10.4)
	Average performance	291 (77.6)
	High performance	45 (12.0)
	Total	375 (100.0)
2013	Low performance	61 (7.5)
	Average performance	693 (85.3)
	High performance	58 (7.1)
	Total	812 (100.0)

Hip - same clinicians over time		
Year	**Performance level**	**N Clinicians (%)**
2011	Low performance	9 (12.0)
	Average performance	62 (82.7)
	High performance	4 (5.3)
	Total	75 (100.0)
2012	Low performance	10 (13.3)
	Average performance	59 (78.7)
	High performance	6 (8.0)
	Total	75 (100.0)
2013	Low performance	5 (6.7)
	Average performance	63 (84.0)
	High performance	7 (9.3)
	Total	75 (100.0)

Table 1b3. Clinician Performance: Patients with Hip Impairments

Knee - all clinicians		
Year	Performance level	N Clinicians (%)
2011	Low performance	82 (6.1)
	Average performance	965 (72.2)
	High performance	289 (21.6)
	Total	1336 (100.0)
2012	Low performance	87 (5.5)
	Average performance	1168 (73.4)
	High performance	336 (21.1)
	Total	1591 (100.0)
2013	Low performance	188 (6.7)
	Average performance	2081 (73.7)
	High performance	554 (19.6)
	Total	2823 (100.0)

Knee - same clinicians over time		
Year	Performance level	N Clinicians (%)
2011	Low performance	33 (5.4)
	Average performance	444 (72.2)
	High performance	138 (22.4)
	Total	615 (100.0)
2012	Low performance	29 (4.7)
	Average performance	421 (68.5)
	High performance	165 (26.8)
	Total	615 (100.0)
2013	Low performance	28 (4.6)
	Average performance	420 (68.3)
	High performance	167 (27.2)
	Total	615 (100.0)

Table 1b4. Clinician Performance: Patients with Knee Impairments

Foot & Ankle - all clinicians		
Year	Performance level	N Clinicians (%)
2011	Low performance	32 (5.2)
	Average performance	472 (76.5)
	High performance	113 (18.3)
	Total	617 (100.0)
2012	Low performance	48 (6.3)
	Average performance	583 (76.6)
	High performance	130 (17.1)
	Total	761 (100.0)
2013	Low performance	95 (7.0)
	Average performance	1088 (79.7)
	High performance	182 (13.3)
	Total	1365 (100.0)

Foot & Ankle - same clinicians over time		
Year	Performance level	N Clinicians (%)
2011	Low performance	5 (2.4)
	Average performance	164 (79.2)
	High performance	38 (18.4)
	Total	207 (100.0)
2012	Low performance	11 (5.3)
	Average performance	150 (72.5)
	High performance	46 (22.2)
	Total	207 (100.0)
2013	Low performance	11 (5.3)
	Average performance	152 (73.4)
	High performance	44 (21.3)
	Total	207 (100.0)

Table 1b5. Clinician Performance: Patients with Foot & Ankle Impairments

General - all clinicians		
Year	Performance level	N Clinicians (%)
2011	Low performance	74 (10.5)
	Average performance	536 (75.7)
	High performance	98 (13.8)
	Total	708 (100.0)
2012	Low performance	102 (11.7)
	Average performance	645 (74.0)
	High performance	125 (14.3)
	Total	872 (100.0)
2013	Low performance	252 (15.1)
	Average performance	1231 (73.8)
	High performance	185 (11.1)
	Total	1668 (100.0)

General - same clinicians over time		
Year	Performance level	N Clinicians (%)
2011	Low performance	32 (12.2)
	Average performance	192 (73.0)
	High performance	39 (14.8)
	Total	263 (100.0)
2012	Low performance	24 (9.1)
	Average performance	190 (72.2)
	High performance	49 (18.6)
	Total	263 (100.0)
2013	Low performance	31 (11.8)
	Average performance	186 (70.7)
	High performance	46 (17.5)
	Total	263 (100.0)

Table 1b6. Clinician Performance: Patients with General Orthopaedic Impairments

Elbow Wrist & Hand - all clinicians		
Year	Performance level	N Clinicians (%)
2011	Low performance	102(31.5)
	Average performance	195(60.2)
	High performance	27(8.3)
	Total	324(100)
2012	Low performance	119(32.3)
	Average performance	223(60.6)
	High performance	26(7.1)
	Total	368(100)
2013	Low performance	117(19.1)
	Average performance	434(70.8)
	High performance	62(10.1)
	Total	613(100)

Elbow Wrist & Hand - same clinicians over time		
Year	Performance level	N Clinicians (%)
2011	Low performance	42 (32.3)
	Average performance	78 (60.0)
	High performance	10 (7.7)
	Total	130 (100)
2012	Low performance	44 (33.8)
	Average performance	72 (55.4)
	High performance	14 (10.8)
	Total	130 (100)
2013	Low performance	24 (18.5)
	Average performance	90 (69.2)
	High performance	16 (12.3)
	Total	130 (100)

Table 1b7. Clinician Performance: Patients with Elbow, Wrist & Hand Impairments

2. Revisions to Measure Specifications Document Sections

FOTO has revised the descriptions of its measures in response to the committee's request for additional information and greater clarity. For brevity of this response, we provide a unified revised definition that contains a number of details specific to each particular measure.

De.3. Brief description of measure (*including type of score, measure focus, target population, timeframe, e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year*)
A Patient Reported Outcome Measure - Performance Measure (PROM-PM) that uses as its basis a body-part specific measure of change in functional status from intake to discharge from treatment (measured with a Patient Reported Outcome Measure (PROM)) that is risk-adjusted for patient characteristics. The PROM-PM can be used as a performance measure at the patient level, the individual clinician, and at the clinic level to assess quality. The 7 FOTO PROM-PMs and their corresponding PROMs are: lumbar, shoulder, hip, knee, foot/ankle, elbow, wrist and hand, and general orthopaedic.

S.4. Numerator Statement (Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.18).

Patient Level: Cases include all patients 18 years and older within a given 12 month period who have completed the specified FOTO PROM measure (lumbar, shoulder, knee, hip, foot/ankle, elbow, wrist and hand, or general orthopaedic) at both intake and at discharge from therapy. The patient measure is the risk-adjusted change score (actual change – predicted change) for the individual patient.

Individual Clinician Level: Cases for the clinician level analysis include only those clinicians who have met or exceeded the FOTO threshold for the minimum number of patients that have completed (intake and discharge) the specified FOTO PROM measure (lumbar, shoulder, knee, hip, foot/ankle, elbow, wrist and hand, or general orthopaedic) during the previous 12 months. Based on the reliability at the clinician level (Section 5: Table 5a), the recommended threshold for patients per clinician per year varies by FOTO measure and is: 10 patients per year for lumbar, elbow wrist and hand, and general orthopaedic, 20 for knee, foot and ankle and shoulder, and 30 for hip.

Clinic Level: Cases for the clinic level analysis include those clinics that have met or exceeded the FOTO threshold for the minimum number of patients that have completed (intake and discharge) the specified FOTO PROM measure (lumbar, shoulder, knee, hip, foot/ankle, elbow, wrist and hand, or general orthopaedic) during the previous 12 months. The threshold is set at a minimum of 10 patients/therapist (small clinics) or 40 patients per year for larger clinics (5 or more clinicians)

S.6. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.18).

Patient Level: Cases include all patients 18 years and older within a given 12 month period who have completed (at intake and discharge) the specified FOTO PROM measure (lumbar, shoulder, knee, hip, foot/ankle, elbow, wrist and hand or general orthopaedic) in either paper and pencil or computer adaptive test (CAT) form. Paper and pencil (short form) measures have been developed by selecting the items from amongst

the CAT item pool that best predict the CAT score. These paper and pencil short forms provide an alternative to computer adaptive test administration, in the event that a computer system is not available or the clinic prefers this method of administration for their patients. FOTO's internal analyses demonstrate that the correlation between the short forms and the CATs is excellent (range 0.96-0.98).

The patient measure is the risk-adjusted **change score** for the individual patient derived by using a multivariate linear regression model that includes the following independent variables: intake functional status, age, symptom acuity, lumbar surgical history, payer source, gender, fear-avoidance beliefs of physical activities and number of functional comorbidities. Risk adjustment variables, other than intake functional status, are derived from items asked on FOTO's patient inquiry tool (also available in the public domain online) The risk-adjusted change score is derived by applying the statistical risk adjustment model described in S.14 and S.15 of the application, and applying steps 1-5 as described in S.18. The risk-adjusted scores can be applied to evaluate performance at the patient level using the methods described in section 2b5.1j of the measure application.

Individual Clinician Level: To maximize reliability of the benchmarking estimates performance on the PROM-PM is evaluated only for those clinicians who meet or exceed the specified threshold for minimum number of patients in the previous 12 months. Thus, clinician cases include those clinicians who have met the FOTO threshold for the minimum number of patients that have completed (both intake and discharge measures) the specified FOTO PROM measures (lumbar, shoulder, knee, hip, foot/ankle, elbow, wrist and hand, or general orthopaedic) during the previous 12 months. The recommended threshold for patients per clinician per year varies by measure and is: 10 patients per year for lumbar, elbow wrist and hand, and general orthopaedic, 20 for knee, for and ankle, and shoulder, and 30 for hip.

These patient-minimum thresholds, which are newly defined, were derived from a recent FOTO's reliability analysis (Table 5a) which showed that an average clinician reliability of 0.7 or higher was obtained using these patient-minimums. Historically, FOTO has used a standard threshold of 40 patients/clinician/year, but in 2014 changed its threshold criteria to 10 patients/clinician/year to enable participation by clinicians that did not have a sufficient volume of patients. The patient-minimum thresholds will be re-evaluated periodically (at least annually), and modified as necessary, to insure included clinicians will an average reliability of 0.7 or greater. Given that the calculation of reliability coefficients is influenced by sample size, we expect that the minimum threshold requirements may go down as more providers with larger patient populations participate in the dataset.

Clinic Level: To maximize reliability of the benchmarking estimates, performance on the PROM-PM is evaluated only for those clinics that met or exceeded the specified threshold for number of patients in the previous 12 months. [Table 5A] Thus, clinic cases include those clinics that met or exceeded the FOTO threshold for the minimum number of patients that have completed (intake and discharge) the specified FOTO PROM measures (lumbar, shoulder, knee, hip, foot/ankle, elbow, wrist and hand, or general orthopaedic) during the previous 12 months. Historically, FOTO has used a standard threshold of 40 patients/clinic/year regardless of clinic size. The new threshold is set at a minimum of 10 patients/therapist/year (small clinics) or 40 patients per year for larger clinics (5 or more clinicians), to enable participation of small clinics with few clinicians.

These patient-minimum thresholds were adopted by FOTO in 2013-2014. Reliability analyses showed that an average clinician reliability of 0.7 or higher was obtained using these patient-minimums. The patient-minimum thresholds will be re-evaluated periodically (at least annually), and modified as necessary, to insure included providers will achieve an average reliability of 0.7 or greater.

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
IF an OUTCOME MEASURE, state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.18).

All patients 18 years and older who fall within the defined impairment/diagnostic categories for each FOTO measure who have initiated rehabilitation treatment and completed the specified FOTO PROM at both intake and discharge from therapy

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

No changes proposed

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.18).

No changes proposed

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

•Patients who are not being treated for the specified impairment category.
•<18 years of age

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

- Patients who are not being treated for an eligible condition as defined in section S.9.
- Age under 18 years old.

3. Change to age exclusions

The committee requested evidence that the instrument, which was originally developed for ages 18 and over, has been tested for understandability and appropriateness for youth down to age 14, as included in the measure.

Response to NQF
FOTO justified their initial request to change the inclusion criteria for its measures from 18 to 14 years old using the results of sensitivity analyses examining the impact of changing the age exclusion criteria on the risk adjustment models. However, in light of the discussions in committee, we recognize that additional testing is necessary. Therefore, we have decided to withdraw this change and return to the 18 years and older inclusion criteria. FOTO plans to perform our own studies on the understandability and appropriateness for youth 14-18 in the future.

In the remaining comments within this document, we show the results of all new analyses for the general orthopaedic measure using the 18+ criteria. We have also recalculated the marginal means estimates by age-groups 18 and older. FOTO will update the remaining application materials for the other measures after endorsement.

4 General Orthopedic Measure

Committee members commented on the fact that only patients with cervical impairments had been included in the data tables and analyses that FOTO presented with their application, even though the measure is intended for use with patients with other types of impairments of the cervical, thoracic, rib, and cranium, mandible areas or affecting those areas.

Response:

We agree that more information is needed on the General Orthopaedic measure and the other patient types who complete this measure (craniofacial, ribs/trunk, thoracic spine). Therefore, we have revised our descriptive tables and redone all analyses in these comments using the complete sample of patients who completed the general orthopaedic measure. We have also revised the remaining relevant application materials and include below the key related methods, results tables and interpretations below from the General Orthopaedic Measure application sections.

In this section, we utilize the table numbering format that was presented with the original application sections.

a. Characteristics of patients

Table 1b2a below shows the descriptive data for all patient types that completed the general orthopaedic measure by year. Patients with cervical impairments comprise 83% of these patients in all years, followed by those with impairments related to the thoracic spine 13-14%, ribs and trunk (1%) and craniofacial (0-2%).

Table 1b2a. "Performance Scores for all Patients with general orthopaedic impairments who were Discharged from Treatment in 2011-2013 by type of impairment: all patients

Year	2011		2012		2013	
# of intake patients	50182		62015		105979	
# completed episodes	20608		26640		43087	
# of clinics	1177		1360		1973	
# of providers	2558		2746		3753	
# of states	41		45		49	
Gender	50181		62012		105975	
Male	16931		20977		36121	
Female	33250		41035		69854	
Average Age	51.2646		51.4231		53.05639	
Standard Deviation	16.53347		16.80125		17.1148	
Min	14		14		14	
Max	108		108		101	
14-19	942		1314		1764	
20-29	2469		3123		4162	
30-39	4196		5068		6775	
40-49	6493		7752		10311	
50-59	7067		8744		12562	
60-69	5376		6998		10977	
70-79	3215		4226		7420	
80-89	1204		1489		2746	
90-99	102		139		256	
100-109	2		1		3	
Total	31066		38854		56976	
Payer Source						
# Answered	50055		61946		105979	
Indemnity Insurance	990		1225		1584	
Litigation	430		516		677	
Medicaid	2455		2950		4297	
Medicare A	1384		1682		1344	
Medicare B	7941		10173		25013	
Patient	653		819		1510	
HMO	5939		6907		11870	
Preferred Provider	19988		25005		40319	
Workers Comp	3215		3816		5910	
No Fault	293		420		1032	
Other	5118		6089		8246	
Early Intervention	2		4		4	
School	7		7		12	
No Charge	59		107		152	
Auto Insurance	1543		2132		3848	
Medicare C	38		94		161	
FS Intake Measure						
Mean	52.314		52.18895		51.463	
Standard Deviation	14.90146		14.65108		14.0989	

Min	0.03		0.03		0.52	
Max	100		100		100	
Quartiles	43.02	Q1	43.20	Q1	42.78	Q1
	50.8	Q2	50.55	Q2	49.78	Q2
	60.58	Q3	59.96	Q3	58.75	Q3
Deciles	35.49	D1	36.09	D1	35.73	D1
	41.37	D2	41.38	D2	40.95	D2
	44.23	D3	44.32	D3	43.92	D3
	47.67	D4	47.62	D4	46.82	D4
	50.8	D5	50.55	D5	49.78	D5
	53.86	D6	53.41	D6	52.51	D6
	57.98	D7	57.7	D7	56.37	D7
	62.81	D8	62.1	D8	62.05	D8
	71.49	D9	71.34	D9	68.83	D9
FS Discharge Measure						
Mean	66.506		66.317		65.103	
Standard Deviation	17.599		17.518		17.480	
Min	0.52		3.85		0.96	
Max	100		100		100	
Quartiles	52.61	Q1	52.41	Q1	51.87	Q1
	65.67	Q2	65.67	Q2	63.55	Q2
	79.3	Q3	79.3	Q3	77.89	Q3
Deciles	44.32	D1	44.32	D1	43.86	D1
	50.48	D2	50.52	D2	49.12	D2
	55.03	D3	55.03	D3	53.85	D3
	61	D4	60.9	D4	58.58	D4
	65.67	D5	65.67	D5	63.54	D5
	70.95	D6	70.53	D6	68.22	D6
	76.6	D7	76.09	D7	73.25	D7
	84.85	D8	84.18	D8	79.63	D8
	92.28	D9	92.28	D9	92.28	D9
Patient FS Change						
Mean	14.06		14.05		13.39	
Standard Deviation	16.49		16.26		15.79	
Min	-76.90		-84.59		-72.24	
Max	97.45		92.27		92.29	
Quartiles	2.55	Q1	2.60	Q1	2.43	Q1
	11.55	Q2	11.62	Q2	10.73	Q2
	23.73	Q3	23.62	Q3	22.29	Q3
Deciles	-3.21	D1	-3.06	D1	-3.09	D1
	0.35	D2	0.42	D2	0.3	D2
	4.24	D3	4.39	D3	4.05	D3
	7.74	D4	7.96	D4	7.32	D4
	11.55	D5	11.62	D5	10.73	D5
	15.69	D6	15.61	D6	14.5	D6

	20.67	D7	20.66	D7	19.64	D7
	27.05	D8	26.91	D8	25.71	D8
	37.2	D9	36.78	D9	35.37	D9
Risk-Adjusted Patient FS Change - Residual						
Mean	0.51		0.99		0.50	
Standard Deviation	14.73		14.57		14.41	
Min	-73.54		-77.67		-71.76	
Max	69.27		60.93		62.05	
Quartiles	-9.95	Q1	-9.36	Q1	-9.79	Q1
	-1.03971	Q2	-0.58714	Q2	-1.40631	Q2
	9.9131	Q3	10.16571	Q3	9.34057	Q3
Deciles	-16.82	D1	-16.14	D1	-16.13	D1
	-11.8763	D2	-11.2569	D2	-11.5832	D2
	-8.17821	D3	-7.53988	D3	-8.06426	D3
	-4.75736	D4	-4.10431	D4	-4.78233	D4
	-1.05247	D5	-0.58714	D5	-1.40659	D5
	2.74515	D6	3.14073	D6	2.27396	D6
	7.30816	D7	7.50457	D7	6.66607	D7
	12.95935	D8	13.44376	D8	12.46367	D8
	21.38139	D9	21.66431	D9	21.05753	D9
Clinician FS Change Score						
Mean	13.98		14.12		13.13	
Standard Deviation	11.96		11.88		11.02	
Min	-49.18		-55.78		-65.91	
Max	68.93		92.27		71.61	
Quartiles	6.64	Q1	7.02	Q1	6.46	Q1
	12.84	Q2	13.04	Q2	12.1	Q2
	19.68	Q3	19.66	Q3	18.58	Q3
Deciles	0.95	D1	1.69	D1	1.28	D1
	5.09	D2	5.55	D2	4.99	D2
	7.96	D3	8.2	D3	7.7	D3
	10.38	D4	10.7	D4	9.95	D4
	12.82	D5	13.03	D5	12.09	D5
	15.35	D6	15.6	D6	14.38	D6
	17.91	D7	18.23	D7	17.08	D7
	21.92	D8	21.62	D8	20.43	D8
	28.37	D9	28.2	D9	26.06	D9
Risk Adjusted Aggregated Clinician Residual Score						
Mean	0.27		0.85		0.22	
Standard Deviation	10.64		10.42		9.91	
Min	-57.53		-50.80		-62.87	
Max	49.66		56.88		45.02	
Quartiles	-6.14	Q1	-5.43	Q1	-5.88	Q1
	-0.33614	Q2	0.17907	Q2	-0.40653	Q2
	5.87402	Q3	6.34005	Q3	5.53361	Q3

Deciles	-11.98	D1	-10.83	D1	-10.83	D1
	-7.55399	D2	-6.81249	D2	-7.15377	D2
	-4.86851	D3	-4.0928	D3	-4.57785	D3
	-2.53096	D4	-2.05551	D4	-2.55778	D4
	-0.34917	D5	0.1709	D5	-0.40981	D5
	1.86294	D6	2.39054	D6	1.71722	D6
	4.28797	D7	4.75696	D7	4.10817	D7
	7.9246	D8	8.17057	D8	7.07467	D8
	13.45868	D9	13.74679	D9	12.26269	D9
Clinic FS Change Score						
Mean	14.35		14.51		13.31	
Standard Deviation	10.31		10.34		9.50	
Min	-35.35		-26.73		-65.91	
Max	65.58		92.27		71.61	
Quartiles	8.41	Q1	8.29	Q1	8.41	Q1
	13.51	Q2	13.12	Q2	12.53	Q2
	18.91	Q3	19.16	Q3	17.49	Q3
Deciles	3.46	D1	3.85	D1	4.11	D1
	7.3	D2	7.12	D2	7.45	D2
	9.61	D3	9.57	D3	9.25	D3
	11.37	D4	11.34	D4	10.94	D4
	13.46	D5	13.1	D5	12.53	D5
	15.03	D6	15.21	D6	14.32	D6
	17.43	D7	17.81	D7	16.36	D7
	20.32	D8	20.57	D8	18.56	D8
	26.38	D9	26.22	D9	23.49	D9
Risk Adjusted Aggregated Clinic Residual Score						
Mean	0.46		1.08		0.36	
Standard Deviation	9.03		8.91		8.36	
Min	-35.32		-42.32		-62.87	
Max	35.79		56.88		41.75	
Quartiles	-4.70	Q1	-4.15	Q1	-4.03	Q1
	-0.10464	Q2	0.39654	Q2	-0.14658	Q2
	5.11063	Q3	5.65095	Q3	4.31351	Q3
Deciles	-9.44	D1	-8.61	D1	-8.71	D1
	-6.05323	D2	-5.55894	D2	-5.13623	D2
	-3.78955	D3	-3.12097	D3	-3.18216	D3
	-1.88932	D4	-1.3927	D4	-1.59412	D4
	-0.20631	D5	0.39496	D5	-0.15471	D5
	1.44757	D6	2.41754	D6	1.33858	D6
	3.88034	D7	4.37429	D7	3.38658	D7
	6.56617	D8	6.96158	D8	5.57039	D8
	11.08208	D9	11.43642	D9	9.70458	D9

Table 1b2b below shows the descriptive data for all patient types within ranked clinics that completed the general orthopaedic measure by year. Patients with cervical impairments comprise 99% of these patients in all years, followed by those with impairments related to the thoracic spine and craniofacial region (<.01%) each. There were no patients with ribs and trunk impairments included.

Table 1b2b. Performance Scores for Patients with general orthopaedic impairments who were Discharged from Treatment in 2011-2013 within Ranked Clinics 2011-2013 Continued

Year	2011		2012		2013	
# of intake patients	19809		29753		50960	
# completed episodes	10079		15520		24608	
# of clinics	141		220		332	
# of providers	19		24		31	
# of states	27		34		36	
Average Age	51.05422659		51.20832		52.16525	
Standard Deviation	16.47484415		16.81337		16.90631	
Min	14		14		14	
Max	108		108		101	
14-19	427		756		1042	
20-29	1160		1730		2404	
30-39	1927		2933		3905	
40-49	3051		4307		5854	
50-59	3255		4832		7108	
60-69	2396		3837		5919	
70-79	1454		2304		3709	
80-89	539		825		1299	
90-99	45		67		105	
100-109	1		1		2	
	14255		21592		31347	
Payer Source						
# Answered	19728		29695		50960	
Indemnity Insurance	321		625		627	
Litigation	185		244		362	
Medicaid	1150		1584		2297	
Medicare A	452		742		443	
Medicare B	3165		4698		10448	
Patient	209		312		509	
HMO	2583		3088		5918	
Preferred Provider	7374		12175		20693	
Workers Comp	1145		1680		2700	
No Fault	116		210		595	
Other	2424		3182		4550	
Early Intervention	1		1		2	
School	3		2		2	
No Charge	27		25		21	

Auto Insurance	548		1095		1723	
Medicare C	25		32		70	
FS Intake Measure						
Mean	52.20091439		52.27657		51.67632	
Standard Deviation	14.45517643		14.20253		13.74585	
Min	1.13		0.03		0.62	
Max	100		99.81		100	
Quartiles	43.40	Q1	43.63	Q1	43.16	Q1
	50.95	Q2	51	Q2	50.25	Q2
	60.27	Q3	59.96	Q3	59.16	Q3
Deciles	36.13	D1	36.58	D1	36.47	D1
	41.38	D2	41.38	D2	41.38	D2
	44.63	D3	44.9	D3	44.42	D3
	47.91	D4	48.04	D4	47.32	D4
	50.95	D5	51	D5	50.25	D5
	53.69	D6	53.53	D6	52.88	D6
	57.79	D7	57.58	D7	56.81	D7
	62.51	D8	62.19	D8	62.05	D8
	70.95	D9	70.68	D9	68.82	D9
FS Discharge Measure						
Mean	66.46514238		66.30147		65.52676	
Standard Deviation	17.36925219		17.16118		17.35659	
Min	0.52		7.32		0.96	
Max	100		100		100	
Quartiles	52.99	Q1	52.84	Q1	51.97	Q1
	65.67	Q2	65.66	Q2	64.18	Q2
	79.2	Q3	79.01	Q3	77.89	Q3
Deciles	44.87	D1	45.10	D1	44.39	D1
	51.01	D2	51.05	D2	49.82	D2
	55.25	D3	55.22	D3	54.3	D3
	61.14	D4	60.96	D4	59.33	D4
	65.67	D5	65.66	D5	64.18	D5
	70.82	D6	69.69	D6	68.64	D6
	76.48	D7	74.92	D7	73.76	D7
	84.16	D8	81.32	D8	81.21	D8
	92.13	D9	92.28	D9	92.28	D9
Patient FS Change						
Mean	13.95		13.66		13.52	
Standard Deviation	16.27		15.70		15.55	
Min	-55.42		-84.59		-54.95	
Max	92.27		92.27		92.27	
Quartiles	2.57	Q1	2.61	Q1	2.57	Q1
	11.34	Q2	11.29	Q2	10.95	Q2
	23.4	Q3	22.83	Q3	22.28	Q3
Deciles	-3.17	D1	-3.01	D1	-2.95	D1

	0.46	D2	0.59	D2	0.62	D2
	4.26	D3	4.39	D3	4.24	D3
	7.59	D4	7.76	D4	7.46	D4
	11.33	D5	11.29	D5	10.93	D5
	15.39	D6	15.07	D6	14.59	D6
	20.66	D7	19.99	D7	19.56	D7
	26.78	D8	26.07	D8	25.66	D8
	36.48	D9	35.41	D9	35.39	D9
Risk-Adjusted Patient FS Change - Residual						
Mean	0.50		0.82		0.62	
Standard Deviation	14.56		14.24		14.29	
Min	-73.25		-77.67		-55.29	
Max	52.57		56.00		55.29	
Quartiles	-9.75	Q1	-9.26	Q1	-9.70	Q1
	-1.17579	Q2	-0.77919	Q2	-1.26389	Q2
	9.74682	Q3	9.72198	Q3	9.33152	Q3
Deciles	-16.74	D1	-16.02	D1	-15.94	D1
	-11.67116	D2	-11.1498	D2	-11.5041	D2
	-8.0241	D3	-7.43336	D3	-7.93756	D3
	-4.73274	D4	-4.17378	D4	-4.58869	D4
	-1.17978	D5	-0.77919	D5	-1.26764	D5
	2.56485	D6	2.8719	D6	2.33279	D6
	7.14404	D7	7.17641	D7	6.65998	D7
	12.65131	D8	12.84117	D8	12.51039	D8
	21.05909	D9	21.08985	D9	21.32467	D9
Clinician FS Change Score						
Mean	13.10		12.61		14.06	
Standard Deviation	7.68		4.81		5.18	
Min	-1.16		2.87		2.69	
Max	41.10		23.47		27.46	
Quartiles	9.13	Q1	9.34	Q1	10.51	Q1
	10.79	Q2	12.56	Q2	13.28	Q2
	14.51	Q3	14.86	Q3	17.08	Q3
Deciles	8.05	D1	6.38	D1	7.54	D1
	8.52	D2	7.66	D2	9.67	D2
	9.61	D3	9.34	D3	10.51	D3
	10.07	D4	10.39	D4	11.77	D4
	10.79	D5	11.93	D5	12.62	D5
	11.71	D6	12.56	D6	13.28	D6
	13.67	D7	13.7	D7	14.93	D7
	14.67	D8	14.71	D8	15.48	D8
	15.49	D9	14.86	D9	17.08	D9
Risk Adjusted Aggregated Clinician Residual Score						
Mean	-1.03		-0.10		1.10	
Standard Deviation	4.78		4.37		5.12	

Min	-7.29		-7.83		-7.49	
Max	11.37		11.49		13.38	
Quartiles	-4.70	Q1	-2.68	Q1	-2.46	Q1
	-2.33319	Q2	-0.47406	Q2	-0.14175	Q2
	0.04458	Q3	1.94844	Q3	3.92107	Q3
Deciles	-6.89	D1	-5.07	D1	-4.67	D1
	-4.75336	D2	-3.71949	D2	-3.53552	D2
	-4.52369	D3	-2.68392	D3	-2.45908	D3
	-3.28659	D4	-2.00727	D4	-1.87819	D4
	-2.33319	D5	-1.36709	D5	-1.23032	D5
	-1.01975	D6	-0.47406	D6	-0.14175	D6
	-0.65465	D7	0.00792	D7	0.52896	D7
	0.52532	D8	0.72914	D8	2.64162	D8
	1.63404	D9	1.94844	D9	3.92107	D9
Clinic FS Change Score						
Mean	14.14		13.66		13.29	
Standard Deviation	5.81		4.73		4.43	
Min	3.15		3.45		4.32	
Max	42.64		37.64		33.71	
Quartiles	10.75	Q1	10.55	Q1	10.34	Q1
	13.2	Q2	13.15	Q2	12.74	Q2
	16.57	Q3	15.99	Q3	15.38	Q3
Deciles	8.70	D1	8.28	D1	7.98	D1
	10.12	D2	10.04	D2	9.74	D2
	11.12	D3	10.89	D3	10.98	D3
	12.2	D4	12.18	D4	11.86	D4
	13.2	D5	13.15	D5	12.68	D5
	14.49	D6	14.21	D6	13.9	D6
	15.38	D7	15.34	D7	14.83	D7
	17.52	D8	16.74	D8	15.99	D8
	19.6	D9	19.79	D9	18.4	D9
Risk Adjusted Aggregated Clinic Residual Score						
Mean	0.60		0.79		0.35	
Standard Deviation	4.87		4.29		4.16	
Min	-8.00		-9.65		-9.71	
Max	27.12		18.65		16.19	
Quartiles	-2.54	Q1	-2.24	Q1	-2.22	Q1
	-0.14223	Q2	0.38274	Q2	-0.11311	Q2
	2.69303	Q3	3.2497	Q3	2.44327	Q3
Deciles	-4.57	D1	-3.92	D1	-4.42	D1
	-3.23087	D2	-2.44176	D2	-2.98781	D2
	-1.99352	D3	-1.6531	D3	-1.82832	D3
	-0.76263	D4	-0.76011	D4	-1.03639	D4
	-0.14223	D5	0.38274	D5	-0.14643	D5
	0.81323	D6	1.48721	D6	0.86793	D6

	2.37494	D7	2.72476	D7	1.81387	D7
	4.03044	D8	3.98661	D8	3.5939	D8
	5.80874	D9	6.66254	D9	5.51378	D9

b. Revised Performance clinic and clinician level

Analyses of Provider Performance

We calculated patient level residual scores (residual = actual change – predicted change) after risk adjustment modeling and aggregated scores by individual clinician or clinic. At the clinic level, performance was evaluated only for large clinics (5 or more clinicians) that had a minimum of 40 patients, and small clinics (1-4 clinicians) that had a minimum of 10 patients per clinician, in the previous 12 months. At the individual clinician level, performance was evaluated only for clinicians that had a minimum of 10 patients in the previous 12 months. To examine statistical differences between entities (individual clinics or clinicians) performance scores, we plotted each entity's average aggregated patient residual scores (with their 95% confidence intervals) to examine whether or not there were statistically significant differences between clinics/clinicians, or between each clinic/clinician and the national average. Since the mean residual score is hypothetically centered at zero, each entity can be compared to that standard which is the predicted clinic aggregated outcome. When the 95% CI for a clinic/clinician crosses zero, the performance for that year is determined to be no different (statistically) than the predicted national average. If 95% CIs are below or above zero, the performance for that year is determined to be worse or better than the predicted national average, respectively.

Performance clinic/group practice level

Sample: This analysis-was conducted using data drawn from the FOTO database of all patients 18 years and older with general orthopaedic impairments in 2011-2013 in a FOTO internal study. Performance analyses utilized data from only those clinics that had a minimum of 40 patients in the prior 12 months for large clinics (5 clinicians or more) or a minimum of 10 patients per clinician for clinics with 4 or less clinicians. In 2011 there were 13494 patients, 1085 clinicians from 263 clinics and 32 states. In 2012 there were 17109 patients, 1318 clinicians from 331 clinics in 37 states. In 2013 there were 33207 patients, 2204 clinicians from 633 clinics in 42 states.

Results

Clinic performance scores with 95% CIs were classified into three groups: low performance (clinics with 95% CI of residual scores below 0), average performance (clinics with 95% CI of residual scores crossing 0), and high performance (clinics with 95% CI of residual scores above 0). The distribution of clinic performance category by year is shown in Table 2b52.b. The plot of clinics for 2013 is shown in Figure 2b5.2a.

Table 2b5.2c Distribution of Clinic Performance Categories by Year

Year	Performance level	N Clinics (%)
2011	Low performance	44 (16.7)
	Average performance	156 (59.3)
	High performance	63 (24.0)
	Total	263 (100.0)
2012	Low performance	49 (14.8)
	Average performance	210 (63.4)
	High performance	72 (21.8)
	Total	331 (100.0)
2013	Low performance	135 (21.3)
	Average performance	400 (63.2)

	High performance	98 (15.5)
	Total	633 (100.0)

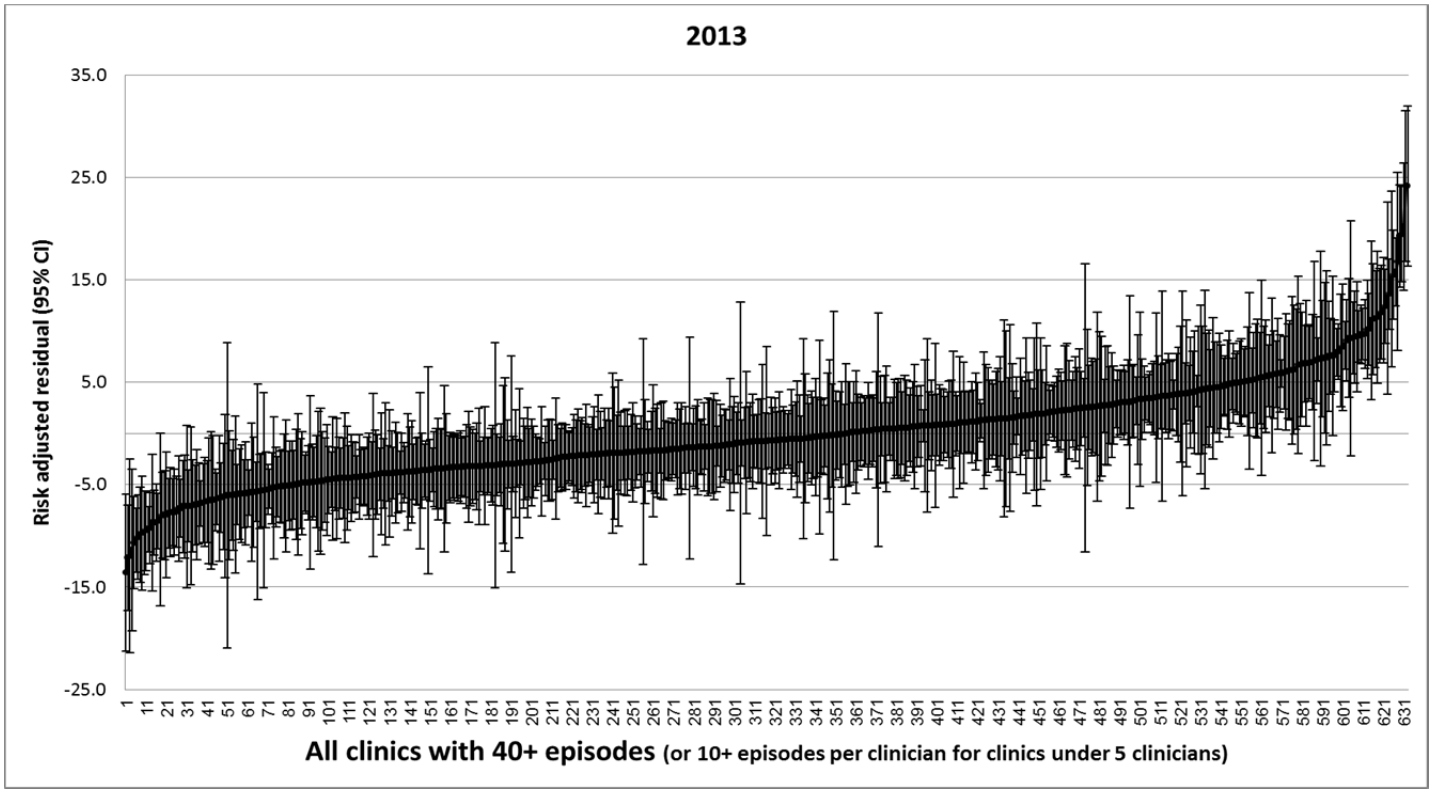


Figure 2b52a. Plot of aggregated residual scores with 95% CI bars in 2013

Performance and the individual clinician level

Sample: This analysis was conducted using data drawn from the FOTO database of all patients 18 years and older with general orthopaedic impairments in 2011-2013 in a FOTO internal study. Performance analyses utilized data from only those clinicians that had a minimum of 10 patients in the prior 12 months. In 2011 there were 12341 patients, 708 clinicians from 394 clinics and 37 states. In 2012 there were 15342 patients, 872 clinicians from 480 clinics in 39 states. In 2013 there were 30124 patients, 1668 clinicians from 857 clinics in 46 states.

Results: Clinician aggregated scores with 95% CIs were classified into three groups: low performance (clinics with 95% CI of residual scores below 0), average performance (clinicians with 95% CI of residual scores crossing 0), and high performance (clinicians with 95% CI of residual scores above 0). The distribution of clinician performance category by year is shown in Table 2b52.b.

Table 2b5.2b Distribution of Clinician Performance Categories by Year

Year	Performance level	N Clinicians (%)
2011	Low performance	74 (10.5)
	Average performance	536 (75.7)
	High performance	98 (13.8)
	Total	708 (100.0)
2012	Low performance	102 (11.7)
	Average performance	645 (74.0)
	High performance	125 (14.3)

	Total	872 (100.0)
2013	Low performance	252 (15.1)
	Average performance	1231 (73.8)
	High performance	185 (11.1)
	Total	1668 (100.0)

c. Revised Missing data analysis

Analyses of Missing Data

To improve interpretation of the impact of missing data, FOTO conducted an internal study comparing patients treated for general orthopaedic impairments in 2011-2013 with and without complete FS data at discharge. Comparisons were made using t-tests or chi-square as appropriate (See table below).

We found that patients with complete data had higher values or prevalence for characteristics that were predictive of both lower (age) and higher (Medicare B payer over the age of 60 and exercise history) FS change therefore not supporting a systematic patient selection bias. Patients with missing data had higher values or prevalence for characteristics associated with lower FS change (Chronic conditions & Medicaid payer) potentially supporting some patient selection bias. Patients with complete and missing discharge data were similar in terms of intake FS, number or comorbidities, gender distribution, surgical history, Medicare B payer under the age of 60 and high levels of fear avoidance at intake. Overall, these analyses were inconclusive and did not support a systematic patient selection bias.

		Complete data			Missing data			
	Year		Total N	% missing		Total N	% missing	P-value
Factors higher or more prevalent in patients with complete data								
Age (years): Mean(SD)	2011	**52.8 (16.1)**	21857	0.0%	**49.8 (15.5)**	12255	0.0%	<.001
	2012	**53.2 (16.2)**	26967	0.0%	**50.1 (15.6)**	16786	0.0%	<.001
	2013	**54.8 (16.4)**	47267	0.0%	**51.0 (15.9)**	31677	0.0%	<.001
Payer: Medicare B age 60 or more	2011	**17.0%**	21759	0.4%	**11.4%**	12232	0.2%	<.001
	2012	**17.4%**	26913	0.2%	**11.9%**	16777	0.1%	<.001
	2013	**24.9%**	47267	0.0%	**15.6%**	31677	0.0%	<.001
Exercise history: 1 or more / week	2011	**67.4%**	21306	2.5%	**65.7%**	11991	2.2%	0.002
	2012	**68.1%**	26548	1.6%	**66.8%**	16589	1.2%	0.005
	2013	**66.6%**	47159	0.2%	**64.9%**	31607	0.2%	<.001
Factors higher or more prevalent in patients with missing data								
Acuity: Chronic - over 3 months	2011	**50.7%**	21846	0.1%	**54.9%**	12246	0.1%	<.001
	2012	**51.0%**	26956	0.0%	**54.1%**	16782	0.0%	<.001
	2013	**51.5%**	47204	0.1%	**54.6%**	31607	0.2%	<.001
Payer: Medicaid	2011	**3.2%**	21759	0.4%	**6.3%**	12232	0.2%	<.001
	2012	**3.2%**	26913	0.2%	**6.1%**	16777	0.1%	<.001
	2013	**2.7%**	47267	0.0%	**5.2%**	31677	0.0%	<.001
Factors with similar values or prevalence between patients with complete or missing data								
Intake FS: Mean(SD)	2011	**52.3 (14.1)**	21857	0.0%	**51.8 (14.6)**	12255	0.0%	0.001
	2012	**52.1 (14.0)**	26967	0.0%	**51.5 (14.6)**	16786	0.0%	<.001

	2013	**51.4 (13.5)**	47267	0.0%	**51.2 (14.2)**	31677	0.0%	0.012
Number of comorbidities: Mean(SD)	2011	**4.0 (2.9)**	21818	0.2%	**4.1 (3.0)**	12222	0.3%	0.003
	2012	**3.9 (3.0)**	26951	0.1%	**4.0 (3.1)**	16779	0.0%	0.002
	2013	**4.4 (3.2)**	47261	0.0%	**4.3 (3.3)**	31671	0.0%	0.041
Gender (Female)	2011	**66.1%**	21857	0.0%	**66.2%**	12254	0.0%	0.830
	2012	**66.0%**	26967	0.0%	**66.5%**	16786	0.0%	0.304
	2013	**65.9%**	47267	0.0%	**65.5%**	31677	0.0%	0.235
Surgical history: 1 or more	2011	**12.5%**	21524	1.5%	**12.6%**	12093	1.3%	0.667
	2012	**13.0%**	26744	0.8%	**13.0%**	16666	0.7%	0.978
	2013	**12.9%**	47186	0.2%	**12.6%**	31606	0.2%	0.136
Payer: Medicare B age under 60	2011	**1.9%**	21759	0.4%	**2.5%**	12232	0.2%	<.001
	2012	**2.0%**	26913	0.2%	**2.8%**	16777	0.1%	<.001
	2013	**3.0%**	47267	0.0%	**3.9%**	31677	0.0%	<.001
High Fear Avoidance at intake (Physical)	2011	**68.7%**	21495	1.7%	**68.6%**	12030	1.8%	0.861
	2012	**62.8%**	26664	1.1%	**62.3%**	16646	0.8%	0.371
	2013	**58.4%**	46955	0.7%	**58.5%**	31476	0.6%	0.794
Difference not supporting potential for selection bias								
Differences supporting potential for selection bias								
Differences interpreted as not clinically important								

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

In an internal FOTO study, we examined the completeness of outcomes data of patients with general orthopaedic impairments admitted to therapy during 2011-2013. Data were extracted during July 2014 therefore all episodes of care had started at least 6 months prior to data extraction. For all years combined there were a total of 217666 patients during this time period. 98.1% (213426) had completed a FOTO (general orthopaedic) PROM at admission, while 1.9% had a non-participation audit (NPA) indicating the reason that FS was not collected. We examined whether or not these patients had FS data collected; and if no data was collected we categorized the reason for missing discharge data. Reasons for missing discharge FS data included having an "open episode", defined as patient not being discharged from therapy; having a NPA at DC, or unknown reason. For the 3 years combined the completion rate was 45%, while 55% had missing data. Open episodes accounted for 26.5% of missing DC FS. The results, stratified by year are shown in the table below.

Table 2b7.2 a Summary of FS Data Collection at Admission and Discharge

Year at episode start	Patients at admission N= 217666 Women: 66.2 % Age (Mn; SD; Range): 52.3; 16.1; 18-113		Patients after 6 months from admission N= 213426			
			Has DC FS + N (%) % represents CR	Reason for Missing FS at DC N (%)		
	FS N (%)	FS missing* (NPA) N (%)		Open episode	NPA at DC	Unknown

2011	49806 (99.5)	265 (0.5)	21857 (43.9)	15694 (31.5)	1553 (3.1)	10702 (21.5)
2012	60092 (97.7)	1426 (2.3)	26967 (44.9)	16339 (27.2)	5812 (9.7)	10974 (18.3)
2013	103528 (97.6)	2549 (2.4)	47267 (45.7)	24584 (23.7)	17034 (16.5)	14643 (14.1)
Total	213426 (98.1)	4240 (1.9)	96091 (45.0)	56617 (26.5)	24399 (11.4)	36319 (17.0)

+Only data from closed episodes with status/discharge FS are included in the aggregate dataset for calculation of risk adjustment coefficients.
*FS missing at intake with NPA completed indicating reason for missing data
CR = Completion rate = Number of discharged patients who had FS at DC/ Number of patients who had FS at intake, includes episodes that are still "open".

We also examined percent of complete outcomes data by clinic to determine the variability in completeness across clinics. There was substantial variation in the completion rate by clinic. Mean CR were 43.9%, 47.3% and 44.5% for years 2011, 2012 and 2013 respectively.

Table 2b7.2 b Mean and SD of average number of patients in clinics (all clinics):

Year	N Clinics	Patients with intake FS data Mean N (SD)	Has DC FS		Does not have DC FS Mean N (sd)
			Mean N (SD)	Mean CR(SD)	
2011	1341	37.1 (51.1)	16.3 (25.3)	43.9 (30.7)	20.8 (31.2)
2012	1505	39.9 (54.4)	17.9 (27.8)	47.3 (30.2)	22.0 (32.5)
2013	2258	45.8 (60.9)	20.9 (31.1)	44.5 (27.8)	24.9 (34.9)

CR = Completion rate = Number of discharged patients who had FS at DC/ Number of patients who had FS at intake, includes episodes that are still "open".

We repeated this analysis for clinics meeting the criteria selected for the performance analysis (who had at least 40 patients (for clinics with 5+ clinicians) or at least 10 patients per clinician for clinics with less than 5 clinicians) for the year tested with FS at intake & discharge treated for general orthopaedic impairment. Completion rates for these clinics were higher at 57.2 %, 56.9 % and 57.3 % for the years 2011, 2012 and 2013 respectively. See Table below:

Table 2b7.2c Mean and SD of average number of patients in clinics with 40+ complete episodes per year (or 10+ complete episodes per clinician for clinics under 5 clinicians):

Year	N Clinics	Patients with intake FS data Mean N (SD)	Has DC FS		Does not have DC FS Mean N (sd)
			Mean N (SD)	Mean CR(SD)	
2011	263	97.8 (74.8)	51.3 (37.2)	57.2 (17.9)	46.5 (46.8)
2012	331	99.0 (77.9)	51.7 (41.7)	56.9 (18.2)	47.3 (46.3)
2013	633	100.3 (84.2)	52.5 (42.9)	57.3 (17.1)	47.9 (48.8)

CR = Completion rate = Number of discharged patients who had FS at DC/ Number of patients who had FS at intake, includes episodes that are still "open".

We assessed whether missing data was a source of systematic bias by testing associations between clinic completion rates and clinic quality (as measured by clinic average residual scores after risk adjustment modeling) for clinics included in the performance analysis. Residual scores are the difference between predicted functional outcomes (given risk adjustment factors) and the actual outcomes. We expected that if

systematic bias were present we would see a statistically significant association between clinic completion rates and clinic quality. We examined Pearson Correlations between clinic CR and clinic average residual scores by year. No correlations were found between CR and residual scores except for a low (r=0.131) but statistically significant correlation (p=0.17) during 2012.

Table 2b7.2 d Relationship between Clinic Aggregated Residual Score and Clinic Completion Rate (CR)

Pearson Correlations (r)			Clinic CR
2011	Residual	r	.105
		Sig. (2-tailed)	.088
		N clinics	263
2012	Residual	r	.131*
		Sig. (2-tailed)	.017
		N clinics	331
2013	Residual	r	0.053
		Sig. (2-tailed)	.182
		N clinics	633

To examine whether there was an underlying pattern to the relationship between clinic completion rate and risk adjusted residual scores aggregated at the clinic level, we grouped clinics into 10 completion rate categories. Results shown below suggest that the relationship between CR and aggregated residual scores is not linear and has no strong pattern.

Table 2b7.2e Average residuals at the clinic level by completion rate categories				
Year	CR category	Mean residuals	N clinics	Std. Deviation
2011	Lowest to <10	3.3*	1	.
	10 to <20	5.0*	3	4.8
	20 to <30	1.7	13	6.2
	30 to <40	1.5	26	4.1
	40 to <50	-.1	55	4.6
	50 to <60	.1	42	5.5
	60 to <70	.4	59	4.5
	70 to <80	1.3	35	6.2
	80 to <90	3.1	19	4.6
	90 to Highest	4.7	10	7.5
	Total	1.0	263	5.2
2012	Lowest to <10	2.9*	2	3.5
	10 to <20	-1.5*	2	4.6
	20 to <30	-1.4	14	5.0
	30 to <40	.8	44	4.5
	40 to <50	1.5	57	5.3

	50 to <60	-.1	68	4.0
	60 to <70	1.0	68	4.9
	70 to <80	.3	44	5.2
	80 to <90	1.3	18	5.4
	90 to Highest	7.4	14	9.4
	Total	.9	331	5.3
2013	Lowest to <10	7.6*	1	.
	10 to <20	-2.7*	5	3.9
	20 to <30	1.5	27	5.2
	30 to <40	-.1	66	5.4
	40 to <50	-.5	116	4.5
	50 to <60	-.6	140	4.7
	60 to <70	-.5	128	4.1
	70 to <80	-.1	92	5.6
	80 to <90	.3	36	4.7
	90 to Highest	3.3	22	8.5
	Total	-0.2	633.0	5.0

CR = Completion rate = Number of discharged patients who had FS at DC/ Number of patients who had FS at intake, includes episodes that are still "open".
* Average residuals for CR categories that had less than 10 clinics

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Overall, the comparisons of characteristics of patients with and without complete outcomes data show no systematic pattern suggesting a selection bias in the collection of discharge FS data.

Overall completion rates (CR) by clinic were fairly stable between 2011-2013, but were lower than previously reported. We believe that this is because in our current analysis we categorized data as missing if patients had open episodes 6 months or more after admission, whereas in the earlier analyses patients with open episodes were not included in this calculation. While it is possible that some patients were still in active treatment 6 months or more after intake, we believe that this would be unusual. A more likely explanation is that these episodes were still open because in many cases the patient failed to return for a scheduled appointment and hence was not discharged and no outcome data or NPA was collected.

We also note that CRs were higher for clinics that had the minimum number of patients required to be included in the performance analysis. For all 3 years, these higher volume clinics had CRs approximately 10-13% higher than all clinics combined. This suggests that greater familiarity with the FOTO data collection system is associated with higher completion rates. In addition, although the overall percent of clinic included in the performance analysis was low (24%), it increased with time, with 20% for 2011(263 /1341), 22% for 2012 (331 /1505), and 28% for 2013 (633 /2258). This suggests an improved data collection process at the clinic level over time.

5. **Provider Reliability Analyses**

We assessed reliability at the provider level using the signal to noise approach as described below.

Method:
Reliability of each provider was calculated using the formula recommended by Adams 2009 *(Adams, John L., Ph.D. "The Reliability of Provider Profiling, A Tutorial." RAND Corporation 2009).* In summary we divided the variance (σ^2) between all providers by the sum of variance between provider groups and variance within provider divided by the number of patients (*n*) for the provider assessed (see formula below). Only providers that passed the threshold for inclusion in the FOTO benchmarking process were included in the calculation of variance between provider groups (for the clinic level, 10+ patients per clinician per clinic per year for small clinics, and 40+ patients per clinic per year for larger clinics with 5 or more clinicians. For the clinician level, at least 10 patients per clinician per year).

$$Reliabilty = \sigma^2 between\ provider\ groups \sigma 2 between\ provider\ groups + \sigma 2 wit\hbar in\ provider\ group n$$

The variance between all provider groups was assessed using a mixed linear hierarchical regression model with patients nested within the provider using the MINQUE (minimum norm quadratic unbiased estimator) variance components procedure in SPSS version 22. The dependent variable was functional status change at discharge from physical therapy, adjusting for all variables used by FOTO for risk adjustment (intake functional status, age, acuity, surgical history relevant to the impairment being treated, gender, payer, number of comorbidities and level of fear avoidance beliefs at admission to therapy).The variance within provider was calculated for each provider using the square of the standard deviation of functional status change. We then calculated the average reliability for all providers that passed the threshold described above.

Results
Because the number of providers in the FOTO database is so large, we present reliability statistics by groups of providers based on their number of patients per year in each impairment category (Table 1a). Average reliability, as well as minimum and maximum reliability coefficients and the proportion of providers that have reliability coefficients >0.7 are shown in the table below. In summary, the average reliability of clinics meeting the FOTO threshold of number of patients per clinic for quality reporting was 0.83 for lumbar, 0.67 for hip, 0.76 for knee, 0.73 for foot and ankle, 0.75 for shoulder, 0.78 for elbow, wrist and hand, and 0.77 for general orthopaedic. At the clinician level average reliability for providers with 10 or more patients per year (FOTO's threshold for 2013) was 0.70 for lumbar, 0.53 for hip, 0.62 for knee, 0.59 for foot and ankle, 0.61 for shoulder, 0.68 for elbow, wrist and hand, and 0.66 for general orthopaedic.

Interpretation
Based on these findings and using the minimum threshold of a reliability of 0.7 (after rounding to 1 decimal point) , we believe that the FOTO PROM-PM at the clinic level is reliable when used for clinics that meet the current FOTO threshold of number of patients per clinic in one year (10+ per clinician for small clinics, 40+ per clinic for large clinics (5 or more clinicians)). However, findings suggest that the 2013 threshold of 10 patients for the clinician level PROM-PM may be insufficient for all but the lumbar, the elbow, wrist and hand, and the general orthopaedic measures. Thus, FOTO will establish new thresholds for clinician PROM-PMs that will be measure-specific: 10 patients per year for lumbar, elbow, wrist and hand, and general orthopaedic, 20 for knee, foot and ankle, and shoulder, and 30 for hip. FOTO will reevaluate reliability annually, as our database grows, given that the calculation of reliability coefficients is influenced by sample size.

Table 5a. Reliability (R) at the provider level: 2011-2013

Provider Level	Threshold (pts/year)		Lumbar	Hip	Knee	Foot & Ankle	Shoulder	Elbow, Wrist Hand	General Ortho
Clinic	*FOTO	N Providers	1274	836	1158	619	1156	344	754
		Mean R	**0.83**	**0.67**	**0.76**	**0.73**	**0.75**	**0.78**	**0.78**
		Min	0.29	0.16	0.17	0.20	0.17	0.26	0.17
		Max	0.99	0.98	0.99	0.98	0.99	0.99	0.99
		% R>=0.7	83.0%	48.4%	69.1%	64.3%	66.3%	75.9%	74.0%
	20+	N Providers	1099	498	958	474	963	268	600
		Mean R	**0.85**	**0.72**	**0.79**	**0.78**	**0.77**	**0.84**	**0.81**
		Min	0.29	0.31	0.35	0.36	0.33	0.40	0.39
		Max	0.99	0.97	0.99	0.98	0.99	0.99	0.99
		% R>=0.7	88.2%	57.6%	74.5%	74.5%	69.9%	90.7%	82.5%
	30+	N Providers	953	289	763	386	769	237	490
		Mean R	**0.85**	**0.74**	**0.84**	**0.79**	**0.80**	**0.85**	**0.84**
		Min	0.53	0.40	0.47	0.47	0.42	0.47	0.52
		Max	0.99	0.97	0.99	0.97	0.99	0.99	0.99
		% R>=0.7	90.9%	61.2%	86.8%	74.9%	78.4%	94.1%	91.2%
	40+	N Providers	818	182	614	327	608	208	417
		Mean R	**0.87**	**0.78**	**0.85**	**0.80**	**0.83**	**0.86**	**0.84**
		Min	0.55	0.54	0.51	0.54	0.49	0.63	0.61
		Max	0.99	0.97	0.99	0.97	0.99	0.99	0.99
		% R>=0.7	96.5%	75.3%	91.7%	78.0%	89.3%	96.2%	93.3%
Clinician	10+	N Providers	3422	1000	3153	1718	3016	747	1939
		Mean R	**0.70**	**0.53**	**0.62**	**0.59**	**0.61**	**0.68**	**0.66**
		Min	0.15	0.15	0.13	0.17	0.15	0.24	0.17
		Max	0.99	0.95	0.98	0.97	0.99	0.98	0.98
		% R>=0.7	54.3%	16.6%	36.3%	27.5%	34.5%	50.1%	44.3%
	20+	N Providers	1854	97	1528	412	1368	358	573
		Mean R	**0.77**	**0.62**	**0.71**	**0.68**	**0.67**	**0.78**	**0.76**
		Min	0.33	0.30	0.32	0.40	0.30	0.42	0.49
		Max	0.99	0.90	0.98	0.95	0.99	0.97	0.98
		% R>=0.7	73.1%	28.9%	54.1%	41.7%	41.2%	77.7%	70.0%
	30+	N Providers	912	16	683	92	521	246	220
		Mean R	**0.81**	**0.72**	**0.73**	**0.68**	**0.72**	**0.82**	**0.80**
		Min	0.53	0.56	0.47	0.42	0.37	0.48	0.56
		Max	0.99	0.87	0.97	0.90	0.99	0.98	0.98
		% R>=0.7	88.9%	43.8%	61.5%	47.8%	56.8%	88.6%	90.9%
	40+	N Providers	451	2	279	38	200	174	76
		Mean R	**0.84**	**N/A**	**0.79**	**0.75**	**0.74**	**0.83**	**0.81**
		Min	0.60	N/A	0.56	0.53	0.50	0.66	0.66
		Max	0.99	N/A	0.95	0.91	0.99	0.98	0.98
		% R>=0.7	96.2%	N/A	86.0%	73.7%	68.0%	93.1%	94.7%

R=Reliability, *10+ per clinician for small clinics (1-3 clinicians), 40+ per clinic for large clinics (4 or more clinicians)

6. Validity of Provider Classification

The committee requested additional data on validity at the clinician and clinic level. This information was presented in part in the supplemental materials provided prior to the in-person meeting. However, that document included only data from the clinic level, but not the clinician level and did not include information on the elbow, wrist and hand and general orthopaedic measures. This document addresses these gaps, and thus replaces the supplemental document submitted prior to the January NQF in-person meeting titled, Validity of Provider Classification.

Analyses:

We performed additional analyses to examine the validity of the FOTO PROM-PM at the clinic and clinician level. Analyses utilized data from only those clinics that had a minimum of 40 patients in the prior 12 months for large clinics (5 clinicians or more) or a minimum of 10 patients per clinician for clinics with 4 or less clinicians. We utilized the new threshold for number of patients per provider to perform the clinician level analyses. We hypothesized that clinics and clinicians that were classified as higher performing using FOTO's methodology of risk-adjusted aggregated scores would also have a greater proportion of patients who had improved during treatment by at least minimal clinically important improvement (MCII). We performed ANOVAs with post-hoc analyses to test our hypothesis. Data are shown below.

RESULTS

Clinic-level PROM-PM

Overall, a higher proportion of patients in the high performing clinics experienced change greater than a minimal clinically important improvement (MCII) as compared to clinics in average and low performing clinics. This pattern was observed in every measure and in every year (2011-2013). These results support the validity of the FOTO PROM-PMs for provider classification.

Table 6 a. Clinic Validity Results: Lumbar

Year	Performance level	N Clinics (%)	% MCII or more (Mean, SD)	P value
2011	Low performance	77 (14.7)	60.1(10.9)	P<0.001
	Average performance	278 (53.1)	71.8(9.2)	
	High performance	169 (32.3)	85.4(7.1)	
	Total	524 (100)	74.5(12.3)	
2012	Low performance	94 (15.1)	57.4(10.0)	P<0.001
	Average performance	355 (57.2)	73.1(8.8)	
	High performance	172 (27.7)	85.8(7.8)	
	Total	621 (100)	74.2(12.6)	
2013	Low performance	191 (17.4)	59.5(8.8)	P<0.001
	Average performance	669 (61.0)	72.4(9.1)	
	High performance	236 (21.5)	85.1(7.2)	
	Total	1096 (100)	72.9(11.8)	

Table 6 b. Clinic Validity Results: Hip				
Year	**Performance level**	**N Clinics (%)**	**% MCII or more (Mean, SD)**	**P value**
2011	Low performance	5 (5.4)	58.2 (6.0)	P<0.001
	Average performance	65 (70.7)	66.4 (10.7)	
	High performance	22 (23.9)	84.6 (7.5)	
	Total	92 (100.0)	70.3 (12.8)	
2012	Low performance	18 (12.9)	54.7 (10.8)	P<0.001
	Average performance	99 (70.7)	69.5 (10.0)	
	High performance	23 (16.4)	83.2 (7.4)	
	Total	140 (100.0)	69.8 (12.4)	
2013	Low performance	21 (7.3)	52.6 (12.7)	P<0.001
	Average performance	216 (75.3)	68.5 (9.6)	
	High performance	50 (17.4)	79.7 (6.6)	
	Total	287 (100.0)	69.3 (11.3)	

Table 6 c. Clinic Validity Results: Knee				
Year	**Performance level**	**N Clinics (%)**	**% MCII or more (Mean, SD)**	**P value**
2011	Low performance	45(9.7)	58.5 (9.2)	P<0.001
	Average performance	262(56.7)	70.0 (9.9)	
	High performance	155(33.5)	82.8 (8.2)	
	Total	462(100)	73.1 (12.0)	
2012	Low performance	48(9.2)	56.8 (8.8)	P<0.001
	Average performance	293(56.3)	70.1 (8.6)	
	High performance	179(34.4)	81.9 (7.8)	
	Total	520(100)	72.9 (11.2)	
2013	Low performance	94(10.0)	56.7 (10.3)	P<0.001
	Average performance	548(58.5)	69.8 (9.4)	
	High performance	294(31.4)	81.3 (7.6)	
	Total	936(100)	72.1 (11.6)	

Table 6 d. Clinic Validity Results: Shoulder				
Year	**Performance level**	**N Clinics (%)**	**% MCII or more (Mean, SD)**	**P value**
2011	Low performance	127 (28.3)	56.3 (8.5)	P<0.001

	Average performance	218 (48.7)	70.8 (10.3)	
	High performance	103 (23.0)	82.6 (7.9)	
	Total	448 (100)	69.4 (13.3)	
2012	Low performance	134 (25.5)	58.9 (8.3)	P<0.001
	Average performance	265 (50.4)	70.7 (9.9)	
	High performance	127 (24.1)	83.7 (8.0)	
	Total	526 (100)	70.9 (12.6)	
2013	Low performance	116 (12.6)	61.0 (9.6)	P<0.001
	Average performance	574 (62.5)	70.7 (8.6)	
	High performance	228 (24.8)	82.2 (7.9)	
	Total	918 (100)	72.3 (10.7)	

Table 6 e. Clinic Validity Results: Foot & Ankle

Year	Performance level	N Clinics (%)	% MCII or more (Mean, SD)	P value
2011	Low performance	24 (10.2)	55.3 (12.1)	P<0.001
	Average performance	147 (62.3)	69.2 (10.3)	
	High performance	65 (27.5)	83.3 (7.6)	
	Total	236 (100.0)	71.7 (12.8)	
2012	Low performance	20 (7.2)	53.1 (10.5)	P<0.001
	Average performance	176 (63.5)	69.4 (9.0)	
	High performance	81 (29.2)	84.6 (7.9)	
	Total	277 (100.0)	72.7 (12.4)	
2013	Low performance	43 (8.8)	52.6 (11.9)	P<0.001
	Average performance	337 (68.9)	69.8 (10.1)	
	High performance	109 (22.3)	83.4 (6.7)	
	Total	489 (100.0)	71.3 (12.6)	

Table 6 f. Clinic Validity Results: Elbow Wrist and Hand				
Year	**Performance level**	**N Clinics (%)**	**% MCII or more (Mean, SD)**	**P value**
2011	Low performance	62(41.9)	47.2(11.7)	P<0.001
	Average performance	69(46.6)	66.7(9.1)	
	High performance	17(11.5)	84.1(10.6)	
	Total	148(100)	60.6(16.3)	
2012	Low performance	70(41.7)	50.8(12.7)	P<0.001
	Average performance	78(46.4)	64.9(11.0)	
	High performance	20(11.9)	84.4(9.2)	
	Total	168(100)	61.3(15.8)	
2013	Low performance	87(32.0)	54.0(10.6)	P<0.001
	Average performance	149(54.8)	68.5(9.3)	
	High performance	36(13.2)	82.1(7.9)	
	Total	272(100)	65.7(13.2)	

Table 6 g. Clinic Validity Results: General Orthopaedic				
Year	**Performance level**	**N Clinics (%)**	**% MCII or more (Mean, SD)**	**P value**
2011	Low performance	44 (16.7)	42.7 (7.0)	P<0.001
	Average performance	156 (59.3)	57.4 (10.6)	
	High performance	63 (24.0)	76.2 (9.6)	
	Total	263 (100.0)	59.5 (14.6)	
2012	Low performance	49 (14.8)	42.6 (8.2)	P<0.001
	Average performance	210 (63.4)	58.7 (10.4)	
	High performance	72 (21.8)	76.6 (11.7)	
	Total	331 (100.0)	60.2 (14.6)	
2013	Low performance	135 (21.3)	42.5 (9.7)	P<0.001
	Average performance	400 (63.2)	57.0 (11.5)	
	High performance	98 (15.5)	76.5 (11.6)	
	Total	633 (100.0)	56.9 (15.1)	

RESULTS

Clinician-level PROM-PM

Overall, a higher proportion of patients treated by high performing clinicians experienced change greater than a minimal clinically important improvement as compared to those treated by average and low performing clinicians. This pattern was observed in every measure and in every year (2011-2013). These results support the validity of the FOTO PROM-PMs for provider classification at the clinician level.

Table 6 g. Clinician Level Validity: Lumbar				
Year	**Performance level**	**N Clinicians (%)**	**% MCII or more (Mean, SD)**	**P value**
2011	Low performance	153 (10.4)	54.2 (12.3)	P<0.001
	Average performance	1028 (69.6)	72.4 (12.3)	
	High performance	297 (20.1)	89.7 (7.5)	
	Total	1478 (100)	74.0 (15.0)	
2012	Low performance	207 (11.5)	53.2 (13.1)	P<0.001
	Average performance	1251 (69.7)	72.8 (12.0)	
	High performance	338 (18.8)	89.9 (8.0)	
	Total	1796 (100)	73.8 (15.2)	
2013	Low performance	372 (11.9)	56.2 (12.3)	P<0.001
	Average performance	2317 (73.8)	73.1 (11.2)	
	High performance	450 (14.3)	89.8 (7.1)	
	Total	3139 (100)	73.5 (13.8)	

Table 6h. Clinician Level Validity: Hip				
Year	**Performance level**	**N Clinicians (%)**	**% MCII or more (Mean, SD)**	**P value**
2011	Low performance	1 (25.0)	64.7 (0.0)	P<0.001
	Average performance	3 (75.0)	59.7 (11.5)	
	High performance	0 (0.0)	0.0 (0.0)	
	Total	4 (100.0)	61.0 (9.7)	
2012	Low performance	1 (14.3)	50.0 (0.0)	P<0.001
	Average performance	4 (57.1)	72.1 (5.9)	
	High performance	2 (28.6)	89.0 (8.0)	
	Total	7 (100.0)	73.8 (14.2)	
2013	Low performance	2 (20.0)	54.4 (10.4)	P<0.001
	Average performance	6 (60.0)	75.9 (8.8)	
	High performance	2 (20.0)	75.8 (8.2)	
	Total	10 (100.0)	71.6 (12.0)	

Table 6i. Clinician Level Validity: Knee				
Year	**Performance level**	**N Clinicians (%)**	**% MCII or more (Mean, SD)**	**P value**
2011	Low performance	40 (7.9)	55.1 (10.3)	P<0.001
	Average performance	341 (67.5)	69.7 (9.6)	
	High performance	124 (24.6)	85.3 (8.2)	
	Total	505 (100.0)	72.3 (12.5)	
2012	Low performance	39 (5.7)	55.5 (11.1)	P<0.001
	Average performance	476 (69.4)	70.5 (10.2)	
	High performance	171 (24.9)	84.0 (7.6)	
	Total	686 (100.0)	73.0 (12.1)	
2013	Low performance	90 (7.1)	57.7 (9.0)	P<0.001
	Average performance	887 (69.7)	70.7 (9.8)	
	High performance	296 (23.3)	85.0 (8.0)	
	Total	1273 (100.0)	73.1 (11.9)	

Table 6i. Clinician Level Validity: Foot & Ankle				
Year	**Performance level**	**N Clinicians (%)**	**% MCII or more (Mean, SD)**	**P value**
2011	Low performance	7 (5.8)	54.5 (9.1)	P<0.001
	Average performance	81 (66.9)	66.8 (11.1)	
	High performance	33 (27.3)	82.8 (7.7)	
	Total	121 (100.0)	70.4 (12.9)	
2012	Low performance	12 (7.7)	49.4 (12.3)	P<0.001
	Average performance	106 (67.9)	68.8 (9.2)	
	High performance	38 (24.4)	84.7 (6.5)	
	Total	156 (100.0)	71.2 (12.8)	
2013	Low performance	27 (9.4)	52.5 (10.8)	P<0.001
	Average performance	204 (70.8)	70.8 (10.4)	
	High performance	57 (19.8)	85.7 (7.8)	
	Total	288 (100.0)	72.1 (13.2)	

Table 6j. Clinician Level Validity: Shoulder				
Year	**Performance level**	**N Clinicians (%)**	**% MCII or more (Mean, SD)**	**P value**
2011	Low performance	105 (22.3)	54.1 (10.4)	P<0.001
	Average performance	285 (60.5)	69.6 (10.4)	
	High performance	81 (17.2)	85.0 (8.9)	
	Total	471 (100.0)	68.8 (14.0)	
2012	Low performance	105 (18.4)	56.0 (9.1)	P<0.001
	Average performance	370 (64.8)	69.6 (10.6)	
	High performance	96 (16.8)	84.8 (9.0)	
	Total	571 (100.0)	69.6 (13.2)	
2013	Low performance	95 (8.6)	57.8 (10.4)	P<0.001
	Average performance	787 (70.9)	70.8 (9.4)	
	High performance	228 (20.5)	84.6 (7.8)	
	Total	1110 (100.0)	72.5 (11.6)	

Table 6k Clinician Level Validity: Elbow, Wrist and Hand				
Year	**Performance level**	**N Clinicians (%)**	**% MCII or more (Mean, SD)**	**P value**
2011	Low performance	102(31.5)	46.8 (12.1)	P<0.001
	Average performance	195(60.2)	65.5 (13.0)	
	High performance	27(8.3)	86.3 (8.9)	
	Total	324(100)	61.3 (16.8)	
2012	Low performance	119(32.3)	46.1 (14.2)	P<0.001
	Average performance	223(60.6)	65.4 (12.9)	
	High performance	26(7.1)	89.3 (6.6)	
	Total	368(100)	60.8 (17.6)	
2013	Low performance	117(19.1)	50.8 (12.4)	P<0.001
	Average performance	434(70.8)	67.7 (13.3)	
	High performance	62(10.1)	84.8 (8.4)	
	Total	613(100)	66.2 (15.6)	

Table 6l Clinician Level Validity: General Orthopaedic				
Year	**Performance level**	**N Clinicians (%)**	**% MCII or more (Mean, SD)**	**P value**
2011	Low performance	74 (10.5)	35.7 (9.3)	P<0.001
	Average performance	536 (75.7)	57.9 (14.8)	
	High performance	98 (13.8)	82.0 (10.6)	
	Total	708 (100.0)	59.0 (17.9)	
2012	Low performance	102 (11.7)	36.6 (10.7)	P<0.001
	Average performance	645 (74.0)	58.2 (14.0)	
	High performance	125 (14.3)	84.0 (12.0)	
	Total	872 (100.0)	59.4 (18.1)	
2013	Low performance	252 (15.1)	38.2 (12.3)	P<0.001
	Average performance	1231 (73.8)	57.2 (14.2)	
	High performance	185 (11.1)	81.8 (11.5)	
	Total	1668 (100.0)	57.0 (17.5)	

7. Components of Variance Analysis

Purpose

To determine the components of variation in outcomes attributable to the clinic, the clinician, or other sources or variance.

Method

We constructed a three level hierarchical linear mixed model with patients nested within clinicians that are nested within clinics, after controlling for patient factors and calculated the amount of variance explained by both the clinic and the clinician levels. This model included those clinics and clinicians that had the threshold number of patients in the prior 12 months for participation in the PROM-PM measures. The threshold for these analyses was: **clinics**: 10 or more patients per clinician for small clinics or a minimum of 40 patients per clinics for larger clinics with 5 or more clinicians; **clinicians**: a minimum of 10 patients in the prior 12 months.

Results:

The results of these analyses are shown in Table 4a below. In summary the clinic level accounted for between 5.9-10.7% of the variance in the model; and the clinician level accounted for between 2.1-3.3% of the variance.

	Lumbar	Hip	Knee	Foot/ Ankle	Shoulder	Elbow Wrist/ Hand	General Orthopaedic
Clinic level variance	9.5%	7.2%	6.8%	5.9%	8.4%	10.7%	6.5%
Clinician level variance	3.3%	2.9%	2.9%	3.1%	2.9%	2.1%	3.0%
Additional variance	87.2%	89.9%	90.3%	91.0%	88.8%	87.2%	90.5%

Table 7a. Percent of variance explained by each level of the hierarchical linear mixed model by FOTO measure

8. Intensity and Frequency of Visits and FOTO Measures

The committee requested additional data on validity at the clinician and clinic level and suggested that they would like to see the link between the FOTO PROM-PM and intensity and frequency of visits. They suggested that validity would be supported if patients who were seen more frequently were doing better.

Response to NQF:
If the assumption could be made that more patient visits would yield better function, then these analyses would be an appropriate method for evaluating the validity of FOTO's PROM-PMs. However, we are not comfortable with this assumption. Treatment dosage (i.e. frequency) is driven by a number of factors including reimbursement caps and allowable number of visits. FOTO is not aware of data that shows that patients receiving outpatient rehabilitation for musculoskeletal problems make more improvement if they are seen more frequently or with greater treatment intensity. In fact, some studies of orthopedic impairments have found a negative association between number of physical therapy visits and outcomes. (Deutscher, D., S. D. Horn, et al. (2009). "Associations between treatment processes, patient characteristics, and outcomes in outpatient physical therapy practice." Archives of Physical Medicine and Rehabilitation **90**(8): 1349-1363). We direct the committee to the previous section 6 titled, *Validity of Provider Classification*. This contains the supplemental material, provided prior to the in-person meeting that described our analyses of clinic validity. Some committee members indicated they had not had the opportunity to review this information prior to the meeting.

.

9. Risk Adjustment

The committee raised some concerns that risk adjustment for gender and payer might actually mask disparities in care and requested more information and a justification for the risk adjustment variables, especially gender and payer. FOTO's estimates of discharge functional change are made from multivariable linear models which control for potential confounders at the patient level. Thus the results can be interpreted as the risk adjusted change scores, conditional on the means (or reference categories) of the covariates in the model. However, the beta coefficients from the risk adjustment model can be used to estimate the impact of each covariate on the dependent variable. The beta coefficients for each risk adjustment model are provided in the application and are available to the public on FOTO's website. The beta coefficients from our risk models, calculated using the most recent FOTO data are summarized below. The interpretation of the beta coefficient is that each 1 unit change in the independent variable is associated with a 1 point change in the FOTO PROM. As an example, a beta coefficient of 1 for male indicates that men have, on average, a one point greater change score as compared to women, holding all other control variables constant.

Beta coefficients							
Variable	**Lumbar**	**Shoulder**	**Hip**	**Knee**	**Foot/ankle**	**Elbow, wrist/hand**	**General orthopaedic**
Age groups							
18-<45 (Reference)							
45-<65	-2.1	-2.0	-2.9	-3.4	-2.6	-1.4	-2.3
65 or more	-2.5	-2.2	-5.0	-3.7	-3.6	-0.4	-1.9
Gender							
Male (Reference)							
Female	-1.1	-1.8	-1.5	-2.3	-0.7	-2.1	-0.8
Payer							
HMO/PPO (Reference)							
Litigation, WC, No fault, Auto	-4.7	-5.1	-4.1	-5.2	-4.1	-4.1	-4.5
All Medicare	-2.1	-1.6	-2.1	-2.5	-3.6	-2.3	-1.9
Medicaid	-4.6	-4.3	-3.6	-5.5	-4.4	-3.4	-3.2
Patient	-0.4	-0.5	-0.2	-0.9	-1.1	-0.9	-0.1
Indemnity	0.1	1.0	0.7	0.0	-1.6	0.8	0.0
No charge/other/school/Early intervention	-0.9	-0.5	-1.0	-1.2	-1.3	-1.4	-0.6
Intake FS (Predicting FS change)	-0.4	-0.7	-0.4	-0.5	-0.4	-0.6	-0.4
Acuity:							
0-7 days (REF)							
8-14 days	-3.3	-1.8	-1.2	-2.2	-1.5	-0.2	-2.4
15-21 days	-4.9	-2.8	-2.1	-3.3	-2.9	-1.8	-3.7
22-90 days	-7.8	-4.1	-3.7	-4.4	-4.8	-3.2	-5.9
91 days to 6 months	-10.3	-4.6	-5.3	-5.4	-7.1	-5.3	-8.2
Over 6 months	-12.6	-5.8	-6.4	-7.0	-9.7	-7.1	-10.6
Surgical history:							
No surgery (REF)							
1 or more surgeries	-2.7	2.4	-0.8	0.9	-1.6	1.1	-3.5

Fear Avoidance-Physical:							
Not Elevated (REF)							
Elevated	-0.8	-0.5	-1.0	-1.0	-0.8	-0.8	-1.0
Number of comorbidities:							
None (REF)							
One	-1.4	-1.2	-1.3	-1.3	-0.9	-0.8	-1.5
Two	-2.0	-2.0	-2.4	-2.2	-1.5	-1.3	-2.2
Three or more	-4.3	-4.0	-4.7	-4.6	-3.8	-3.0	-4.4
Constant	53.4	63.0	49.7	55.3	51.2	55.8	48.8

Table 9a. Beta coefficients from FOTO's risk-adjustment models predicting Functional Status Change: Key Variables

Marginal Means Estimation

Given the request of the committee we believe that it is informative to examine our calculations of estimated marginal means of FS at discharge by age groups, gender and payer. These documents were previously submitted as supplemental materials, however the tables have been since updated to include elbow wrist and hand data. The marginal mean estimates control for intake FS, acuity, surgical history, fear avoidance from physical activity, number of comorbidities were first estimated for all years combined and then the analysis was repeated by year for each measure and data shown in the Tables below.

	N	%	Marginal means	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age groups								
18-<45	45837	26.0	65.8	0.07	928.3	0.000	65.6	65.9
45-<65	64477	37.0	63.6	0.06	1099.2	0.000	63.5	63.8
65 or more	65974	37.4	63.3	0.07	915.3	0.000	63.1	63.4
Gender								
Male	70891	40.2	64.7	0.05	1285.4	0.000	64.6	64.8
Female	105397	59.8	63.6	0.04	1545.3	0.000	63.5	63.7
Payer								
HMO/PPO	76827	43.6	65.4	0.05	1218.4	0.000	65.3	65.5
Litigation, WC, No fault, Auto	16443	9.3	60.8	0.11	548.7	0.000	60.6	61.0
All Medicare	58582	33.2	63.4	0.07	850.9	0.000	63.2	63.5
Medicaid	6485	3.7	60.8	0.17	357.5	0.000	60.5	61.2
Patient	1406	0.8	65.0	0.36	182.9	0.000	64.3	65.7
Indemnity	2222	1.3	65.6	0.28	231.7	0.000	65.0	66.1
No charge, other, school, early intervention	14323	8.1	64.5	0.11	571.1	0.000	64.3	64.7

HMO=health maintenance organization, PPO=Preferred provider organization, WC=worker's compensation

Table 9b. Marginal Means for 2011-2013: Lumbar Impairments

	N	%	Marginal means	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age groups 2011								
18-<45								
45-<65	11323	28.3	66.7	0.14	472.0	0.000	66.4	67.0
65 or more	15501	39.0	64.7	0.12	551.6	0.000	64.4	64.9
Age groups 2012	13216	33.0	64.2	0.16	409.4	0.000	63.9	64.5
18-<45								
45-<65	13569	27.9	66.6	0.13	513.3	0.000	66.3	66.8
65 or more	18345	38.0	64.5	0.11	598.9	0.000	64.3	64.7
Age groups 2013	16712	34.4	64.0	0.14	461.1	0.000	63.7	64.2
18-<45								
45-<65	20945	23.9	64.9	0.11	617.4	0.000	64.7	65.1
65 or more	30631	35.0	62.7	0.08	742.7	0.000	62.5	62.9
Gender 2011	36046	41.1	62.4	0.09	680.3	0.000	62.3	62.6
Male								
Female	16141	40.3	65.8	0.11	619.7	0.000	65.6	66.0
Gender 2012	23899	59.7	64.6	0.09	742.5	0.000	64.4	64.8
Male								
Female	19658	40.4	65.6	0.10	683.8	0.000	65.4	65.8
Gender 2013	28968	59.6	64.5	0.08	818.5	0.000	64.3	64.6
Male								
Female	35092	40.1	63.7	0.07	896.0	0.000	63.6	63.9
Payer 2011	52530	60.0	62.7	0.06	1081.6	0.000	62.6	62.8
HMO/PPO								
Litigation, WC, No fault, Auto	18147	45.3	66.3	0.11	605.3	0.000	66.1	66.5
All Medicare	4008	10.0	62.4	0.23	276.7	0.000	61.9	62.8
Medicaid	11412	29.0	64.4	0.17	376.0	0.000	64.1	64.8
Patient	1760	4.4	61.8	0.33	187.7	0.000	61.1	62.4
Indemnity	312	0.8	65.6	0.76	86.4	0.000	64.1	67.0
No charge, other, school, early intervention	573	1.4	66.7	0.56	119.0	0.000	65.6	67.8
Payer 2012	3828	10.0	65.6	0.22	300.1	0.000	65.2	66.1
HMO/PPO								
Litigation, WC, No fault, Auto	22045	45.3	66.0	0.10	666.5	0.000	65.8	66.2
All Medicare	4827	9.9	61.4	0.20	299.9	0.000	61.0	61.8
Medicaid	14369	29.6	64.6	0.15	426.1	0.000	64.3	64.9
Patient	1902	3.9	61.8	0.31	196.4	0.000	61.2	62.4
Indemnity	378	0.8	65.4	0.69	95.2	0.000	64.1	66.8
No charge, other, school, early intervention	684	1.4	65.9	0.51	128.9	0.000	64.9	66.9
Payer 2013	4421	9.1	65.4	0.20	321.9	0.000	65.0	65.8
HMO/PPO								
Litigation, WC, No fault, Auto	36635	41.8	64.8	0.08	825.4	0.000	64.6	64.9

All Medicare	7608	8.7	59.7	0.16	368.1	0.000	59.4	60.0
Medicaid	32801	37.4	62.3	0.10	638.0	0.000	62.1	62.5
Patient	2823	3.2	59.7	0.26	233.0	0.000	59.2	60.2
Indemnity	716	0.8	64.5	0.50	130.1	0.000	63.5	65.5
No charge, other, school, early intervention	965	1.1	64.9	0.43	151.8	0.000	64.0	65.7

HMO=health maintenance organization, PPO=Preferred provider organization, WC=worker's compensation

Table 9c. Marginal Means by Year: Lumbar Impairments

Supplemental Disparities Data
Shoulder Impairments

The estimated marginal means of FS at discharge by age groups, gender and payer, controlling for intake FS, acuity, surgical history, fear avoidance from physical activity, number of comorbidities for 2011-2013 are shown below. The estimated marginal means analysis was repeated by year and data shown below.

	N	%	Marginal means	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age groups								
18-<45	29140	20.8	69.2	0.09	805.0	0.000	69.0	69.3
45-<65	64028	46.0	67.2	0.06	1146.2	0.000	67.1	67.3
65 or more	46831	33.5	67.0	0.08	814.6	0.000	66.8	67.2
Gender								
Male	63752	45.5	68.5	0.05	1282.6	0.000	68.4	68.6
Female	76247	54.5	66.7	0.05	1367.8	0.000	66.6	66.8
Payer								
HMO/PPO	67444	48.2	68.7	0.06	1211.7	0.000	68.5	68.8
Litigation, WC, No fault, Auto	14688	10.5	63.6	0.12	550.9	0.000	63.3	63.8
All Medicare	40067	28.6	67.0	0.09	743.9	0.000	66.9	67.2
Medicaid	3044	2.2	64.4	0.24	263.5	0.000	63.9	64.8
Patient	1222	0.9	68.2	0.38	178.5	0.000	67.4	68.9
Indemnity	1910	1.4	69.7	0.31	227.7	0.000	69.1	70.3
No charge, other, school, early intervention	11624	8.3	68.1	0.12	545.2	0.000	67.9	68.4

HMO=health maintenance organization, PPO=Preferred provider organization, WC=worker's compensation

Table 9d Marginal Means 2011-2013: Shoulder Impairment

	N	%	Marginal means	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age groups 2011								
18-<45	7165	22.4	68.8	0.17	400.0	0.000	68.5	69.2
45-<65	15424	48.0	66.7	0.12	562.9	0.000	66.5	67.0
65 or more	9416	29.4	66.5	0.19	358.0	0.000	66.1	66.9
Age groups 2012								
18-<45	8827	22.5	69.1	0.16	434.2	0.000	68.8	69.4
45-<65	18349	47.0	67.0	0.11	605.8	0.000	66.8	67.2
65 or more	12138	30.9	66.8	0.17	403.5	0.000	66.5	67.2
Age groups 2013								
18-<45	13148	19.1	69.4	0.13	548.4	0.000	69.2	69.7
45-<65	30255	44.0	67.5	0.08	796.3	0.000	67.3	67.7
65 or more	25277	36.8	67.3	0.11	617.1	0.000	67.0	67.5
Gender 2011								
Male	14805	46.3	68.1	0.11	610.1	0.000	67.9	68.3
Female	17200	53.7	66.3	0.10	641.1	0.000	66.1	66.5
Gender 2012								
Male	18109	46.1	68.4	0.10	666.5	0.000	68.1	68.6
Female	21205	53.9	66.6	0.09	704.8	0.000	66.5	66.8
Gender 2013								
Male	30838	44.9	68.8	0.08	912.6	0.000	68.7	69.0
Female	37842	55.1	66.9	0.07	984.6	0.000	66.8	67.0
Payer 2011								
HMO/PPO	16154	50.5	67.9	0.11	592.5	0.000	67.7	68.1
Litigation, WC, No fault, Auto	3458	10.8	63.2	0.24	265.2	0.000	62.7	63.6
All Medicare	7680	24.0	67.2	0.21	322.2	0.000	66.8	67.6
Medicaid	823	2.6	64.0	0.47	135.2	0.000	63.1	64.9
Patient	231	0.7	67.0	0.88	75.9	0.000	65.3	68.8
Indemnity	482	1.5	70.7	0.61	115.3	0.000	69.5	71.9
No charge, other, school, early intervention	3177	10.0	67.8	0.24	282.7	0.000	67.3	68.3
Payer 2012								
HMO/PPO	19766	50.3	68.4	0.11	647.4	0.000	68.2	68.6
Litigation, WC, No fault, Auto	4176	10.6	63.5	0.22	288.2	0.000	63.0	63.9
All Medicare	9993	25.4	67.2	0.18	363.8	0.000	66.8	67.6
Medicaid	915	2.3	64.2	0.45	141.2	0.000	63.3	65.1
Patient	283	0.7	67.9	0.81	83.8	0.000	66.3	69.5
Indemnity	601	1.5	68.4	0.56	123.0	0.000	67.3	69.5
No charge, other, school, early intervention	3580	9.1	67.9	0.23	296.2	0.000	67.5	68.4
Payer 2013								
HMO/PPO	31524	45.9	69.2	0.08	838.4	0.000	69.0	69.4

Litigation, WC, No fault, Auto	7054	10.3	63.8	0.16	388.1	0.000	63.5	64.1
All Medicare	22394	32.6	66.9	0.12	570.0	0.000	66.7	67.1
Medicaid	1306	1.9	64.7	0.37	176.5	0.000	64.0	65.4
Patient	708	1.0	68.6	0.49	139.0	0.000	67.6	69.5
Indemnity	827	1.2	70.0	0.46	153.2	0.000	69.1	70.9
No charge, other, school, early intervention	4867	7.1	68.6	0.19	359.9	0.000	68.2	68.9

HMO=health maintenance organization, PPO=Preferred provider organization, WC=worker's compensation

Table 9e. Marginal Means by Year: Shoulder Impairment

Supplemental Disparities Data
Knee Impairments

The estimated marginal means of FS at discharge by age groups, gender and payer, controlling for intake FS, acuity, surgical history, fear avoidance from physical activity, number of comorbidities for 2011-2013 are shown below. The estimated marginal means analysis was repeated by year and data shown below.

	N	%	Marginal means	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age groups								
18-<45	35704	24.8	64.0	0.08	762.2	0.000	63.9	64.2
45-<65	58729	41.0	60.6	0.06	957.5	0.000	60.5	60.8
65 or more	49354	34.3	60.3	0.08	720.2	0.000	60.2	60.5
Gender								
Male	55980	38.9	62.8	0.06	1057.8	0.000	62.7	62.9
Female	87807	61.1	60.5	0.05	1283.1	0.000	60.4	60.6
Payer								
HMO/PPO	72393	50.4	62.7	0.06	1088.5	0.000	62.6	62.8
Litigation, WC, No fault, Auto	9636	6.7	57.5	0.15	393.2	0.000	57.2	57.8
All Medicare	42456	29.5	60.2	0.09	658.3	0.000	60.0	60.4
Medicaid	3535	2.5	57.2	0.24	241.8	0.000	56.7	57.7
Patient	1174	0.8	61.8	0.40	152.7	0.000	61.0	62.6
Indemnity	2089	1.5	62.7	0.30	206.2	0.000	62.1	63.2
No charge, other, school, early intervention	12504	8.7	61.5	0.13	489.4	0.000	61.2	61.7

HMO=health maintenance organization, PPO=Preferred provider organization, WC=worker's compensation

Table 9f. Marginal Means 2011-2013: Knee Impairments

	N	%	Marginal means	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age groups 2011								
18-<45	8773	27.4	64.4	0.17	382.5	0.000	64.1	64.7
45-<65	13789	43.0	61.6	0.13	472.1	0.000	61.3	61.8
65 or more	9491	29.6	60.6	0.20	310.2	0.000	60.2	61.0
Age groups 2012								
18-<45	10733	26.7	64.7	0.15	419.7	0.000	64.4	65.0
45-<65	16955	42.0	61.1	0.12	518.7	0.000	60.9	61.4
65 or more	12586	31.3	61.1	0.17	362.8	0.000	60.7	61.4
Age groups 2013								
18-<45	16198	22.7	63.5	0.12	509.7	0.000	63.2	63.7
45-<65	27985	39.0	59.9	0.09	654.1	0.000	59.7	60.1
65 or more	27277	38.2	59.8	0.11	543.8	0.000	59.6	60.0
Gender 2011								
Male	12506	39.0	63.5	0.13	499.5	0.000	63.3	63.8
Female	19547	61.0	61.2	0.10	604.4	0.000	61.0	61.3
Gender 2012								
Male	15717	39.0	63.5	0.11	560.2	0.000	63.3	63.7
Female	24557	61.0	61.2	0.09	678.3	0.000	61.0	61.3
Gender 2013								
Male	27757	38.8	62.0	0.08	746.1	0.000	61.9	62.2
Female	43703	61.2	59.8	0.07	906.9	0.000	59.7	59.9
Payer 2011								
HMO/PPO	16861	52.6	63.0	0.12	531.0	0.000	62.8	63.3
Litigation, WC, No fault, Auto	2381	7.4	58.2	0.30	195.8	0.000	57.6	58.8
All Medicare	7914	25.0	61.5	0.22	284.0	0.000	61.0	61.9
Medicaid	940	2.9	57.6	0.46	123.9	0.000	56.7	58.5
Patient	240	0.8	62.1	0.91	68.5	0.000	60.3	63.9
Indemnity	502	1.6	64.1	0.63	102.3	0.000	62.9	65.4
No charge, other, school, early intervention	3215	10.0	62.3	0.25	249.6	0.000	61.8	62.8
Payer 2012								
HMO/PPO	21225	52.7	63.4	0.11	598.7	0.000	63.2	63.6
Litigation, WC, No fault, Auto	2808	7.0	57.3	0.27	210.3	0.000	56.8	57.9
All Medicare	10442	25.9	61.2	0.19	327.6	0.000	60.9	61.6
Medicaid	1056	2.6	57.4	0.44	131.2	0.000	56.6	58.3
Patient	282	0.7	60.9	0.83	73.0	0.000	59.3	62.6
Indemnity	633	1.6	63.0	0.56	112.9	0.000	61.9	64.0
No charge, other, school, early intervention	3828	9.5	61.7	0.23	269.7	0.000	61.3	62.2
Payer 2013								
HMO/PPO	34307	48.0	62.1	0.08	741.1	0.000	62.0	62.3
Litigation, WC, No fault, Auto	4447	6.2	57.3	0.21	268.7	0.000	56.9	57.7

All Medicare	24100	33.7	59.3	0.12	500.2	0.000	59.1	59.5
Medicaid	1539	2.2	56.9	0.35	160.9	0.000	56.2	57.6
Patient	652	0.9	62.0	0.54	115.5	0.000	61.0	63.1
Indemnity	954	1.3	61.7	0.44	139.1	0.000	60.9	62.6
No charge, other, school, early intervention	5461	7.6	60.9	0.19	323.6	0.000	60.5	61.2

HMO=health maintenance organization, PPO=Preferred provider organization, WC=worker's compensation

Table 9g. Marginal Means by Year: Knee Impairments

Supplemental Disparities Data
Hip Impairments

The estimated marginal means of FS at discharge by age groups, gender and payer, controlling for intake FS, acuity, surgical history, fear avoidance from physical activity, number of comorbidities for 2011-2013 are included below. The estimated marginal means analysis was repeated by year and data shown below.

	N	%	Marginal means	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age groups								
18-<45	8774	16.5	64.4	0.16	398.3	0.000	64.1	64.7
45-<65	19415	37.0	61.5	0.11	550.2	0.000	61.3	61.7
65 or more	24915	46.9	59.4	0.11	545.6	0.000	59.2	59.7
Gender								
Male	17419	32.8	62.0	0.10	605.5	0.000	61.8	62.2
Female	35685	67.2	60.5	0.07	850.4	0.000	60.4	60.7
Payer								
HMO/PPO	24351	45.9	62.1	0.10	615.3	0.000	61.9	62.3
Litigation, WC, No fault, Auto	1279	2.4	58.0	0.38	152.0	0.000	57.3	58.8
All Medicare	21342	40.2	60.0	0.12	498.8	0.000	59.8	60.2
Medicaid	980	1.9	58.5	0.43	134.5	0.000	57.7	59.4
Patient	379	0.7	61.9	0.69	89.7	0.000	60.5	63.2
Indemnity	702	1.3	62.8	0.51	123.6	0.000	61.8	63.8
No charge, other, school, early intervention	4071	7.7	61.1	0.21	285.4	0.000	60.7	61.6

HMO=health maintenance organization, PPO=Preferred provider organization, WC=worker's compensation

Table 9h. Marginal Means 2011-2013: Hip Impairments

	N	%	Marginal means	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age groups 2011								
18-<45	1948	17.6	64.7	0.34	190.0	0.000	64.0	65.4
45-<65	4347	39.0	61.7	0.24	260.8	0.000	61.3	62.2
65 or more	4802	43.3	59.9	0.25	235.4	0.000	59.4	60.4
Age groups 2012								
18-<45	2560	17.5	65.1	0.30	215.8	0.000	64.5	65.7
45-<65	5569	38.0	62.3	0.21	298.1	0.000	61.9	62.7
65 or more	6530	44.6	59.6	0.21	277.5	0.000	59.2	60.0
Age groups 2013								
18-<45	4266	15.6	63.9	0.23	275.5	0.000	63.5	64.4
45-<65	9499	35.0	61.0	0.16	383.0	0.000	60.7	61.3
65 or more	13583	49.7	59.2	0.14	408.0	0.000	58.9	59.4
Gender 2011								
Male	3610	32.5	62.5	0.23	277.0	0.000	62.0	62.9
Female	7487	67.5	61.0	0.16	391.2	0.000	60.7	61.3
Gender 2012								
Male	4824	32.9	62.7	0.20	315.3	0.000	62.3	63.1
Female	9835	67.1	61.1	0.14	441.4	0.000	60.8	61.3
Gender 2013								
Male	8985	32.9	61.4	0.14	436.8	0.000	61.2	61.7
Female	18363	67.2	60.1	0.10	613.4	0.000	59.9	60.3
Payer 2011								
HMO/PPO	5230	47.1	62.1	0.22	286.5	0.000	61.7	62.5
Litigation, WC, No fault, Auto	297	2.7	58.9	0.80	74.0	0.000	57.4	60.5
All Medicare	3990	36.0	60.9	0.29	213.3	0.000	60.3	61.4
Medicaid	231	2.1	58.8	0.90	65.3	0.000	57.0	60.6
Patient	70	0.6	62.6	1.61	38.9	0.000	59.4	65.8
Indemnity	172	1.6	63.4	1.03	61.5	0.000	61.3	65.4
No charge, other, school, early intervention	1107	10.0	61.5	0.41	149.7	0.000	60.7	62.3
Payer 2012								
HMO/PPO	7119	48.6	62.2	0.19	335.6	0.000	61.9	62.6
Litigation, WC, No fault, Auto	364	2.5	57.6	0.73	79.2	0.000	56.2	59.0
All Medicare	5327	36.3	61.0	0.24	250.8	0.000	60.5	61.5
Medicaid	303	2.1	60.1	0.80	75.5	0.000	58.5	61.7
Patient	86	0.6	61.0	1.47	41.4	0.000	58.1	63.9
Indemnity	209	1.4	61.9	0.95	65.4	0.000	60.1	63.8
No charge, other, school, early intervention	1251	8.5	61.9	0.39	158.0	0.000	61.1	62.6
Payer 2013								
HMO/PPO	12002	43.9	62.0	0.14	429.9	0.000	61.7	62.3
Litigation, WC, No	618	2.3	57.8	0.54	106.6	0.000	56.8	58.9

fault, Auto								
All Medicare	12025	44.0	59.2	0.16	376.9	0.000	58.9	59.5
Medicaid	446	1.6	57.3	0.64	90.1	0.000	56.1	58.6
Patient	223	0.8	61.9	0.89	69.7	0.000	60.2	63.7
Indemnity	321	1.2	63.2	0.74	85.0	0.000	61.7	64.7
No charge, other, school, early intervention	1713	6.3	60.4	0.33	184.6	0.000	59.8	61.1

HMO=health maintenance organization, PPO=Preferred provider organization, WC=worker's compensation

Table 9i. Marginal Means by Year: Hip Impairments

Supplemental Disparities Data
Foot/Ankle Impairments

The estimated marginal means of FS at discharge by age groups, gender and payer, controlling for intake FS, acuity, surgical history, fear avoidance from physical activity, number of comorbidities for 2011-2013 are shown below. The estimated marginal means analysis was repeated by year and data shown below.

	N	%	Marginal means	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age groups								
18-<45	24359	33.0	66.1	0.10	674.8	0.000	65.9	66.3
45-<65	30540	41.0	63.4	0.08	747.9	0.000	63.3	63.6
65 or more	18820	25.5	62.5	0.14	438.6	0.000	62.2	62.8
Gender								
Male	26326	35.7	64.5	0.09	756.1	0.000	64.4	64.7
Female	47393	64.3	63.8	0.06	1008.6	0.000	63.7	64.0
Payer								
HMO/PPO	41189	55.9	65.5	0.08	868.9	0.000	65.3	65.6
Litigation, WC, No fault, Auto	5371	7.3	61.4	0.19	316.9	0.000	61.0	61.8
All Medicare	17231	23.4	61.8	0.15	411.1	0.000	61.6	62.1
Medicaid	1800	2.4	61.1	0.33	187.0	0.000	60.5	61.8
Patient	698	1.0	64.4	0.52	124.1	0.000	63.4	65.4
Indemnity	1045	1.4	63.9	0.42	150.5	0.000	63.1	64.8
No charge, other, school, early intervention	6385	8.7	64.2	0.17	369.8	0.000	63.8	64.5

HMO=health maintenance organization, PPO=Preferred provider organization, WC=worker's compensation

Table 9j. Marginal Means 2011-2013: Foot and Ankle Impairments

	N	%	Marginal means	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age groups 2011								
18-<45	5968	35.6	66.8	0.20	341.4	0.000	66.4	67.1
45-<65	7222	43.0	64.1	0.17	369.1	0.000	63.8	64.4
65 or more	3598	21.4	63.1	0.33	190.0	0.000	62.5	63.8
Age groups 2012								
18-<45	7323	35.3	67.0	0.18	376.0	0.000	66.6	67.3
45-<65	8696	42.0	64.2	0.16	406.9	0.000	63.9	64.5
65 or more	4726	22.8	62.9	0.28	222.6	0.000	62.3	63.4
Age groups 2013								
18-<45	11068	30.6	65.3	0.15	446.6	0.000	65.0	65.5
45-<65	14622	40.0	62.7	0.12	508.5	0.000	62.4	62.9
65 or more	10496	29.0	62.0	0.19	328.7	0.000	61.6	62.4
Gender 2011								
Male	5955	35.5	65.4	0.18	361.7	0.000	65.0	65.7
Female	10833	64.5	64.6	0.13	484.7	0.000	64.3	64.8
Gender 2012								
Male	7500	36.2	65.5	0.16	405.5	0.000	65.2	65.8
Female	13245	63.9	64.5	0.12	534.3	0.000	64.3	64.8
Gender 2013								
Male	12871	35.6	63.6	0.12	525.8	0.000	63.4	63.8
Female	23315	64.4	63.1	0.09	705.2	0.000	62.9	63.3
Payer 2011								
HMO/PPO	10814	59.5	65.9	0.15	430.2	0.000	65.6	66.2
Litigation, WC, No fault, Auto	1379	7.6	62.4	0.39	160.9	0.000	61.6	63.1
All Medicare	3168	17.4	63.1	0.36	176.1	0.000	62.4	63.8
Medicaid	579	3.2	61.5	0.65	94.9	0.000	60.2	62.7
Patient	168	0.9	64.9	1.13	57.6	0.000	62.7	67.1
Indemnity	284	1.6	65.7	0.85	77.6	0.000	64.1	67.4
No charge, other, school, early intervention	1799	9.9	64.9	0.34	188.5	0.000	64.2	65.6
Payer 2012								
HMO/PPO	13374	59.3	66.1	0.14	480.0	0.000	65.8	66.4
Litigation, WC, No fault, Auto	1565	6.9	61.5	0.36	169.9	0.000	60.8	62.2
All Medicare	4194	18.6	63.2	0.30	208.9	0.000	62.6	63.8
Medicaid	686	3.0	61.9	0.60	103.5	0.000	60.8	63.1
Patient	170	0.8	64.7	1.09	59.4	0.000	62.6	66.8
Indemnity	350	1.6	63.8	0.78	82.2	0.000	62.3	65.4
No charge, other, school, early intervention	2199	9.8	64.8	0.31	207.0	0.000	64.2	65.4
Payer 2013								
HMO/PPO	21713	55.6	65.0	0.11	585.3	0.000	64.7	65.2
Litigation, WC, No fault, Auto	2469	6.3	60.8	0.28	213.7	0.000	60.3	61.4

All Medicare	9878	25.3	60.7	0.20	309.6	0.000	60.3	61.1
Medicaid	1013	2.6	60.4	0.49	123.4	0.000	59.4	61.4
Patient	425	1.1	63.9	0.69	92.3	0.000	62.6	65.3
Indemnity	505	1.3	63.0	0.63	99.5	0.000	61.8	64.3
No charge, other, school, early intervention	3063	7.8	63.4	0.26	242.1	0.000	62.9	63.9

HMO=health maintenance organization, PPO=Preferred provider organization, WC=worker's compensation

Table 9k Marginal Means by Year: Foot and Ankle Impairments

Supplemental Disparities Data
General Orthopaedic

The estimated marginal means of FS at discharge by age groups, gender and payer, controlling for intake FS, acuity, surgical history, fear avoidance from physical activity, number of comorbidities for 2011-2013 are included below. The estimated marginal means analysis was repeated by year and data shown below.

	N	%	Marginal means	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age groups								
18-<45	26821	28.8	66.8	0.10	690.0	0.000	66.6	67.0
45-<65	39488	42.0	64.5	0.08	822.2	0.000	64.4	64.7
65 or more	26907	28.9	64.8	0.12	543.0	0.000	64.6	65.1
Gender								
Male	31681	34.0	65.8	0.08	807.7	0.000	65.7	66.0
Female	61535	66.0	65.0	0.06	1116.0	0.000	64.9	65.1
Payer								
HMO/PPO	45111	48.4	66.4	0.07	893.5	0.000	66.3	66.6
Litigation, WC, No fault, Auto	11028	11.8	62.0	0.14	428.6	0.000	61.7	62.2
All Medicare	23954	25.7	64.5	0.13	504.0	0.000	64.2	64.8
Medicaid	2722	2.9	63.3	0.28	225.6	0.000	62.7	63.8
Patient	951	1.0	66.3	0.47	141.7	0.000	65.4	67.2
Indemnity	1434	1.5	66.5	0.38	174.3	0.000	65.7	67.2
No charge, other, school, early intervention	8016	8.6	65.9	0.16	404.5	0.000	65.5	66.2

HMO=health maintenance organization, PPO=Preferred provider organization, WC=worker's compensation

Table 9l. Marginal Means 2011-2013: General Orthopaedic Impairments

	N	%	Marginal means	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age groups 2011								
18-<45	6426	30.8	68.1	0.20	343.6	0.000	67.7	68.5
45-<65	9247	44.0	65.4	0.16	401.1	0.000	65.1	65.7
65 or more	5201	24.9	65.2	0.28	233.8	0.000	64.7	65.8
Age groups 2012								
18-<45	7889	30.2	67.4	0.18	378.7	0.000	67.1	67.8
45-<65	11345	43.0	65.2	0.15	447.0	0.000	65.0	65.5
65 or more	6906	26.4	65.5	0.24	277.6	0.000	65.0	65.9
Age groups 2013								
18-<45	12506	27.1	65.8	0.14	465.2	0.000	65.5	66.1
45-<65	18896	41.0	63.7	0.11	563.4	0.000	63.5	63.9
65 or more	14800	32.0	64.3	0.16	406.6	0.000	64.0	64.6
Gender 2011								
Male	7070	33.9	66.8	0.18	381.2	0.000	66.5	67.2
Female	13804	66.1	65.9	0.12	527.3	0.000	65.6	66.1
Gender 2012								
Male	8906	34.1	66.4	0.15	429.0	0.000	66.1	66.7
Female	17234	65.9	65.7	0.11	593.7	0.000	65.5	66.0
Gender 2013								
Male	15705	34.0	65.0	0.11	569.4	0.000	64.8	65.3
Female	30497	66.0	64.1	0.08	785.6	0.000	64.0	64.3
Payer 2011								
HMO/PPO	10539	50.5	67.1	0.15	434.1	0.000	66.8	67.4
Litigation, WC, No fault, Auto	2455	11.8	62.9	0.31	202.8	0.000	62.3	63.5
All Medicare	4502	22.0	65.7	0.30	216.7	0.000	65.1	66.3
Medicaid	679	3.3	64.2	0.57	112.7	0.000	63.1	65.3
Patient	212	1.0	65.4	1.01	64.9	0.000	63.4	67.3
Indemnity	356	1.7	67.6	0.78	86.9	0.000	66.1	69.1
No charge, other, school, early intervention	2131	10.0	66.8	0.32	209.0	0.000	66.2	67.5
Payer 2012								
HMO/PPO	13027	49.8	67.0	0.14	489.9	0.000	66.7	67.3
Litigation, WC, No fault, Auto	3140	12.0	62.4	0.27	229.8	0.000	61.9	63.0
All Medicare	5862	22.4	65.5	0.26	253.2	0.000	65.0	66.0
Medicaid	840	3.2	64.0	0.51	125.9	0.000	63.0	65.0
Patient	262	1.0	67.2	0.90	75.0	0.000	65.5	69.0
Indemnity	455	1.7	66.0	0.68	96.9	0.000	64.6	67.3
No charge, other, school, early intervention	2554	9.8	66.5	0.29	229.2	0.000	65.9	67.1
Payer 2013								
HMO/PPO	21545	46.6	65.8	0.11	610.2	0.000	65.6	66.0
Litigation, WC, No fault, Auto	5433	11.8	61.3	0.20	300.1	0.000	60.9	61.7
All Medicare	13590	29.4	63.6	0.17	381.4	0.000	63.2	63.9
Medicaid	1203	2.6	62.2	0.42	149.3	0.000	61.4	63.0

Patient	477	1.0	66.1	0.65	101.2	0.000	64.8	67.3
Indemnity	623	1.4	66.2	0.57	115.9	0.000	65.1	67.3
No charge, other, school, early intervention	3331	7.2	64.9	0.25	259.7	0.000	64.4	65.4

HMO=health maintenance organization, PPO=Preferred provider organization, WC=worker's compensation

Table 9m. Marginal Means by Year: General Orthopaedic Impairments

Supplemental Disparities Data
Elbow, Wrist and Hand

The estimated marginal means of FS at discharge by age groups, gender and payer, controlling for intake FS, acuity, surgical history, fear avoidance from physical activity, number of comorbidities, for 2011-2013 are included below. The estimated marginal means analysis was repeated by year and data shown below.

	N	%	Marginal means	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age groups								
18<45	16620	29.3	66.0	0.11	600.2	0.000	65.8	66.3
45-<65	25658	45.0	64.6	0.09	745.4	0.000	64.4	64.8
65 or more	14426	25.4	65.6	0.15	440.3	0.000	65.3	65.9
Gender								
Male	22293	39.3	66.6	0.09	757.1	0.000	66.4	66.7
Female	34411	60.7	64.5	0.07	917.8	0.000	64.3	64.6
Payer								
HMO/PPO	25634	45.2	66.8	0.09	774.5	0.000	66.6	67.0
Litigation, WC, No fault, Auto	10623	18.7	62.7	0.13	471.3	0.000	62.5	63.0
All Medicare	12961	22.9	64.5	0.16	406.4	0.000	64.2	64.8
Medicaid	1651	2.9	63.4	0.32	197.7	0.000	62.8	64.0
Patient	744	1.3	65.9	0.47	139.4	0.000	64.9	66.8
Indemnity	746	1.3	67.6	0.47	143.3	0.000	66.7	68.5
No charge, other, school, early intervention	4345	7.7	65.4	0.20	332.9	0.000	65.0	65.8

HMO=health maintenance organization, PPO=Preferred provider organization, WC=worker's compensation

Table 9n. Marginal Means 2011-2013: Elbow, Wrist and Hand Impairments

	N	%	Marginal means	Std. Err.	Z	P>|z|	[95% Conf. Interval]	
Age groups 2011								
18-<45	4176	31.6	65.8	0.21	312.0	0.000	65.3	66.2
45-<65	6154	47.0	64.6	0.17	378.6	0.000	64.3	64.9
65 or more	2892	21.9	65.0	0.33	199.2	0.000	64.3	65.6
Age groups 2012								
18-<45	4898	31.2	65.5	0.19	337.2	0.000	65.1	65.9
45-<65	7142	45.5	64.1	0.16	410.1	0.000	63.8	64.4
65 or more	3661	23.3	65.3	0.28	229.3	0.000	64.7	65.8
Age groups 2013								
18-<45	7546	27.2	66.5	0.17	393.1	0.000	66.2	66.9
45-<65	12362	44.5	64.9	0.13	500.0	0.000	64.7	65.2
65 or more	7873	28.3	66.1	0.21	320.5	0.000	65.6	66.5
Gender 2011								
Male	5204	39.4	66.5	0.18	374.5	0.000	66.1	66.8
Female	8018	60.6	64.1	0.14	451.6	0.000	63.8	64.4
Gender 2012								
Male	6207	39.5	66.0	0.16	412.7	0.000	65.7	66.3
Female	9494	60.5	64.0	0.13	499.0	0.000	63.8	64.3
Gender 2013								
Male	10882	39.2	66.9	0.13	516.5	0.000	66.6	67.2
Female	16899	60.8	64.9	0.10	628.9	0.000	64.7	65.1
Payer 2011								
HMO/PPO	6123	46.3	66.1	0.17	390.1	0.000	65.7	66.4
Litigation, WC, No fault, Auto	2696	20.4	62.5	0.26	243.8	0.000	62.0	63.0
All Medicare	2474	18.7	65.1	0.36	183.1	0.000	64.4	65.8
Medicaid	415	3.1	63.5	0.62	101.9	0.000	62.3	64.7
Patient	159	1.2	64.5	1.00	64.7	0.000	62.5	66.4
Indemnity	170	1.3	69.5	0.96	72.2	0.000	67.6	71.4
No charge, other, school, early intervention	1185	9.0	65.6	0.37	179.2	0.000	64.8	66.3
Payer 2012								
HMO/PPO	7275	46.3	66.2	0.15	430.0	0.000	65.9	66.5
Litigation, WC, No fault, Auto	3098	19.7	62.3	0.24	264.6	0.000	61.8	62.8
All Medicare	3205	20.4	64.3	0.31	209.8	0.000	63.7	64.9
Medicaid	514	3.3	62.5	0.55	113.2	0.000	61.5	63.6
Patient	185	1.2	65.4	0.91	71.9	0.000	63.6	67.2
Indemnity	199	1.3	67.6	0.88	77.2	0.000	65.9	69.3
No charge, other, school, early intervention	1225	7.8	65.0	0.35	183.3	0.000	64.3	65.7
Payer 2013								
HMO/PPO	12236	44.0	67.5	0.13	518.8	0.000	67.3	67.8

Litigation, WC, No fault, Auto	4829	17.4	63.1	0.20	309.6	0.000	62.7	63.5
All Medicare	7282	26.2	64.3	0.22	297.3	0.000	63.9	64.8
Medicaid	722	2.6	63.9	0.50	128.2	0.000	63.0	64.9
Patient	400	1.4	66.5	0.66	100.2	0.000	65.2	67.8
Indemnity	377	1.4	66.8	0.68	97.5	0.000	65.4	68.1
No charge, other, school, early intervention	1935	7.0	65.6	0.30	216.1	0.000	65.0	66.2

HMO=health maintenance organization, PPO=Preferred provider organization, WC=worker's compensation

Table 9o. Marginal Means by Year: Elbow, Wrist and Hand Impairments

Q1: Intraclass correlations at the scale and practice level

This question is difficult to interpret. Exactly what is meant by the scale level? Both the Oxford Knee Score and the Oswestry Disability Index are established instruments with strong, published evidence of their psychometric properties. Both are widely used in clinical practice so it is difficult to understand the need to establish intraclass correlation coefficients (ICCs) for these submissions.

The second part of the question "at the practice level" is equally troublesome. ICCs are statistics for measuring homogeneity for pairs of measurements. ICCs measure the proportion of a variance that is attributable to the objects (or targets) of measures. In the case of the two measures under consideration ("Average Change in Functional Status Following Total Knee Replacement Surgery" and "Average Change in Functional Status Following Lumbar Spine Fusion Surgery") the objects are patients that can be attributed to clinics. Any model to which ICCs are defined requires randomly selected objects of measurement (McGraw and Wong, 1996), a requirement that cannot be met for the two measures at either the patient or clinic level, therefore, further clarification of this request is needed.

Q2: Standard error of measurement

This, too, is unclear. We can only assume that the intent of this question is to understand the dispersion of scores around the mean, i.e., the standard deviation. A low standard deviation means that a set of scores is not very widely dispersed around the mean, while a high standard deviation indicates that the scores are more widely dispersed. in a normal distribution, about 68% of scores can be expected to fall in the range of scores between minus one standard deviation below the mean and plus one standard deviation above the mean and that about 95% of scores can be expected to fall in the range of scores between minus two standard deviations below the mean and plus two standard deviations above the mean.

A second, related, dispersion measure is the standard error of the means. Multiple random samples for which the same instrument has been administered results in multiple means, which if plotted would, because of random selection, approximate a normal distribution and the means would have a mean (the mean of means). The standard deviation of this distribution of means would be the standard error of the mean and would represent the distribution of errors (or random fluctuations) in estimating the population mean. The same percentage estimates--68% of the errors to be distributed within one standard error plus or minus of the population mean, and 95% to be distributed within two standard errors plus or minus of the population mean--would apply.

To address this question pre-test score means, post-test score means and the mean difference with standard deviations are provided for each of the two measures. SAS v9.4 (SAS Institute, Cary NC, USA) was used to calculate the standard deviation and standard error of measurement at the surgeon and clinic level.

Total Knee Replacement : *Standard deviation and standard error of means calculation*

N	pre_ mean	post_ mean	Diff_ Mean (pre-post)	pre_std	post_std	Diff_std	pre_std err	post_std err	Diff_std err
2044	22.5905	39.3557	-16.7652	7.8109	8.0484	9.2658	0.1728	0.178	0.2049

Lumbar Spinal Fusion Surgery : *Standard deviation and standard error of means calculation*

N	pre_mean	post_mean	Diff_mean (pre-post)	pre_std	post_std	Diff_std	pre_std err	post_std err	Diff_std err
341	50.1554	32.9091	17.2463	15.9299	22.4077	19.3901	0.8627	1.2134	1.05

Q3: Interpretation of performance scores

The Oxford Knee Score (OKS) is a 12-item patient reported scale originally developed and validated to assess function and pain in patients undergoing total knee replacement. It is considered a joint-specific outcome measure and has been considered the patient-reported outcome measure (PROM) of choice in England and Wales for performance assessment purposes. OKS scores range from 0-48 with higher scores considered worse.

To determine the minimally important clinical difference—the MCID-- (i.e., the minimal change in a scoring measure that is perceived by the patient to be beneficial or harmful) for most health-related quality of life instruments has been to halve the standard deviation of the change (Norman et al, 2003). Across published studies of the OKS this has been estimated to be between 3 and 5 points. A recent study to determine the minimally important clinical difference (i.e., the minimal change in a scoring measure that is perceived by the patient to be beneficial or harmful) following total knee arthroplasty reported the MCID using the anchor method for pain (6-items) to be 5.0 points and for functional outcome (6 items) 4.3 points (Clement et al, 2014).

Oswestry Disability Index (ODI) is one of several numerical rating scales that have become standards for the clinical assessment and evaluation of treatment effectiveness in patients with low-back pain.

MCID values are similar for patients undergoing revision and primary lumbar fusion procedures, even when subgroup analysis is done for the different diagnostic etiologies. This simplifies the interpretation of clinical improvement, since the thresholds are similar in patients undergoing primary compared with revision surgery (Carreon et al, 2013). Patients perceive clinical improvement with a change of 12 points in the ODI (Carreon et al, 2013).

Q4a: Proportion of standard deviation represented by a 14-17 point spread:

By this question we are assuming that the underlying interest is the relationship of the standard deviation of the mean difference to the mean change.

For Total Knee Replacement the mean difference is -16.765 (that is post-scores are higher than pre-scores) and the standard deviation of the mean difference is 9.2658. Using the previously identified percentage interpretation 68% of change scores would be estimated between 7.50 and 26.03; approximately 95% of change scores would be estimated to be between and – 1.77 and 35.30. The MCID using the generally estimated halve s.d. rule would be 4.63, within the MCID estimated range published in literature.

For Lumbar Spinal Fusion Surgery the mean difference is 17.2463 (that is post-scores are lower than pre scores) and the standard deviation of the mean difference is 19.3901. Previously identified percentage interpretations for the standard deviation would apply. The MCID using the generally estimated halve s.d. rule would be 9.70.

Q4b: Is that a meaningful effect size:
Yes for Total Knee Replacement since the mean difference would meet the threshold of an estimated minimum clinically important difference and is comparable to published literature.
Equivocal for Lumbar Spinal Fusion Surgery. The mean difference would meet the threshold of an estimated minimum clinically important difference using the "halve s.d." rule but the estimate from this population is lower than the MCID reported in literature.

Q4c: How usable is the standard deviation to discriminate against groups.
We assume the underlying question is discrimination between groups. The standard deviation cannot be used to discriminate between groups. This requires the use of 95% confidence intervals calculated for each clinic and comparable.

When confidence intervals are calculated at the medical group level for the data sets for these two measures, the 95% confidence intervals can be used to determine statistically significantly different medical groups at the extremes of the ordered groups high to low.

A follow-up analysis using an ANOVA technique clearly demonstrates that there are medical groups that can be statistically significantly differentiated. The ANOVA technique is a recommended technique for this determination.

REFERENCES:

Carreon LY, Bratcher KR, Canan CE, et al. Differentiating minimum clinically important difference for primary and revision lumbar fusion surgeries. *J Neurosurg Spine* 2013; 18: 102-106.

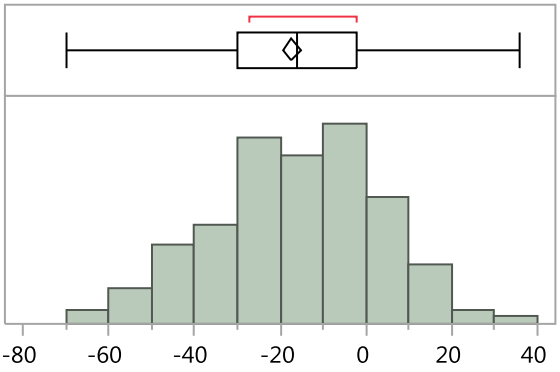
Clement ND, MacDonald D, Simpson AHRW. The minimal clinically imporant difference in the Oxford knee score and Short Form 12 score after total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc* 2014; 22: 1933-1939.

McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1996; 1 (1): 30-46.

Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003; 41: 582-92.

Distributions

Change in ODI Score



Quantiles

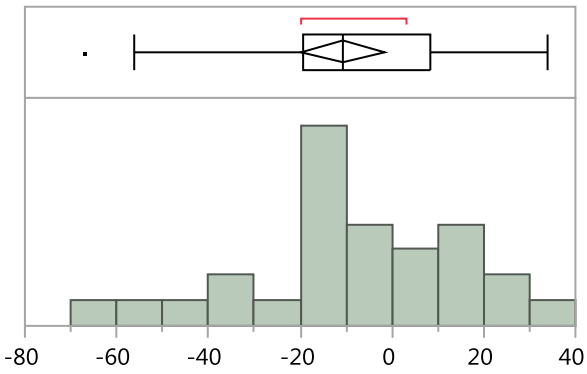
100.0%	maximum	36
99.5%		34.58
97.5%		17.45
90.0%		5.8
75.0%	quartile	-2
50.0%	median	-16
25.0%	quartile	-30
10.0%		-44
2.5%		-58
0.5%		-70
0.0%	minimum	-70

Summary Statistics

Mean	-17.24633
Std Dev	19.39006
Std Err Mean	1.0500306
Upper 95% Mean	-15.18096
Lower 95% Mean	-19.31171
N	341

Distributions Medical Group_Name=A

Change in ODI Score



Quantiles

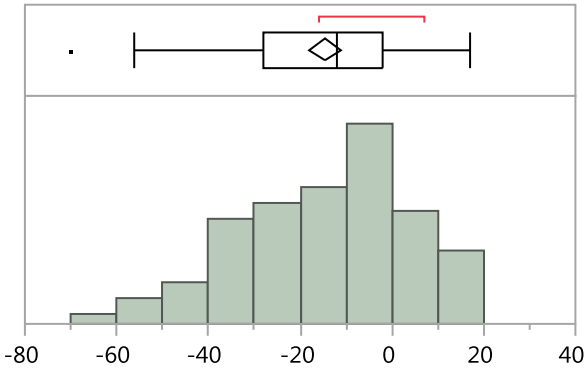
100.0%	maximum	34
99.5%		34
97.5%		34
90.0%		20
75.0%	quartile	8.25
50.0%	median	-10.5
25.0%	quartile	-19.5
10.0%		-45.2
2.5%		-67
0.5%		-67
0.0%	minimum	-67

Summary Statistics

Mean	-10.67857
Std Dev	23.467951
Std Err Mean	4.4350258
Upper 95% Mean	-1.57865
Lower 95% Mean	-19.77849
N	28

Distributions Medical Group_Name=B

Change in ODI Score



Quantiles

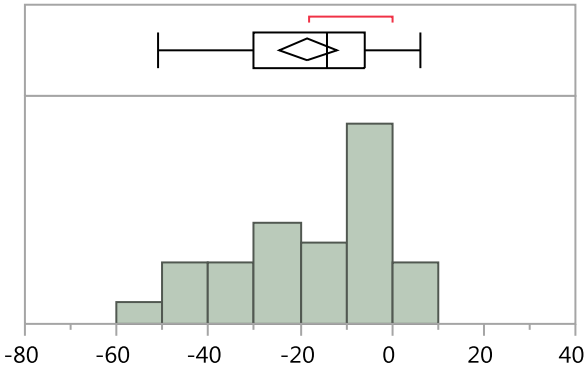
100.0%	maximum	17
99.5%		17
97.5%		15.425
90.0%		7.7
75.0%	quartile	-2
50.0%	median	-12
25.0%	quartile	-28
10.0%		-39.7
2.5%		-54.275
0.5%		-70
0.0%	minimum	-70

Summary Statistics

Mean	-14.60784
Std Dev	18.335788
Std Err Mean	1.8155136
Upper 95% Mean	-11.00635
Lower 95% Mean	-18.20933
N	102

Distributions Medical Group_Name=C

Change in ODI Score



Quantiles

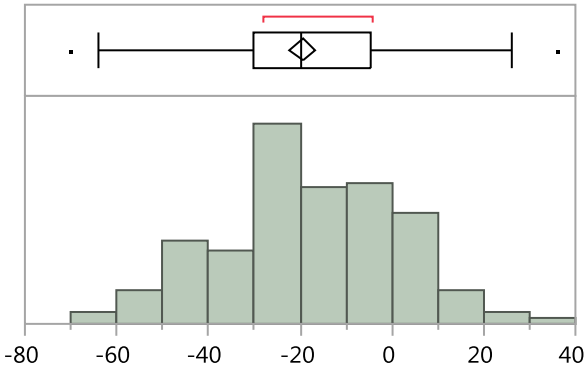
100.0%	maximum	6
99.5%		6
97.5%		6
90.0%		0
75.0%	quartile	-6
50.0%	median	-14
25.0%	quartile	-30
10.0%		-46
2.5%		-51
0.5%		-51
0.0%	minimum	-51

Summary Statistics

Mean	-18.48276
Std Dev	16.513243
Std Err Mean	3.0664322
Upper 95% Mean	-12.20146
Lower 95% Mean	-24.76406
N	29

Distributions Medical Group_Name=D

Change in ODI Score



Quantiles

100.0%	maximum	36
99.5%		36
97.5%		18.425
90.0%		4
75.0%	quartile	-4.75
50.0%	median	-20
25.0%	quartile	-30
10.0%		-48
2.5%		-60
0.5%		-70
0.0%	minimum	-70

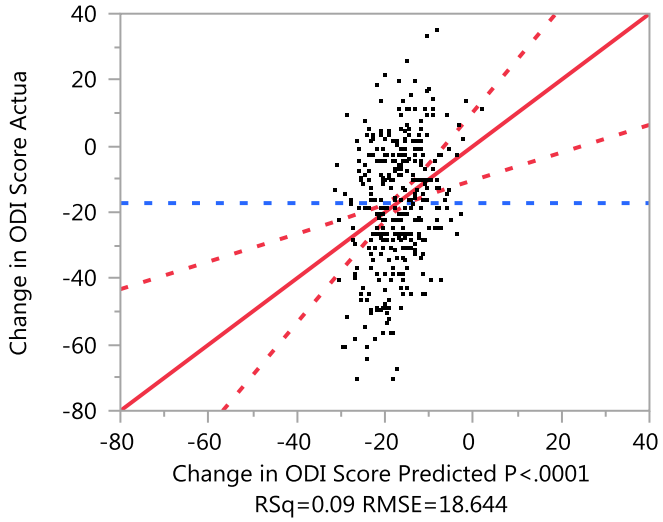
Summary Statistics

Mean	-19.53846
Std Dev	19.446943
Std Err Mean	1.4415033
Upper 95% Mean	-16.69415
Lower 95% Mean	-22.38277
N	182

Response Change in ODI Score

Whole Model

Actual by Predicted Plot



Summary of Fit

RSquare	0.089109
RSquare Adj	0.075514
Root Mean Square Error	18.64358
Mean of Response	-17.2463
Observations (or Sum Wgts)	341

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	11390.95	2278.19	6.5544
Error	335	116440.36	347.58	**Prob > F**
C. Total	340	127831.31		<.0001*

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	319	112971.69	354.143	1.6336
Pure Error	16	3468.67	216.792	**Prob > F**
Total Error	335	116440.36		0.1260
				Max RSq

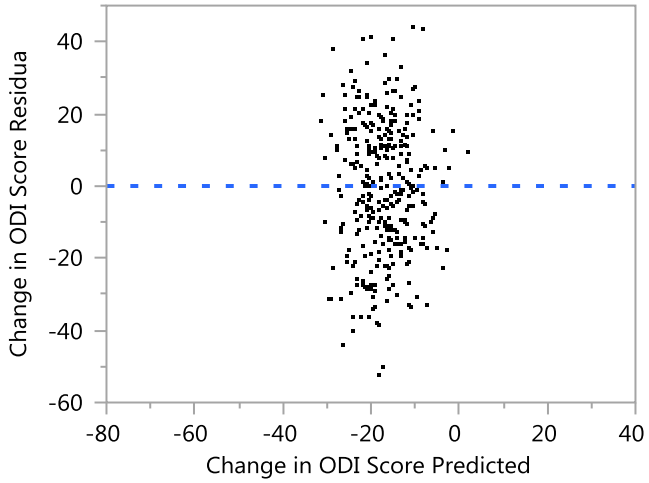
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	-6.764914	4.112211	-1.65	0.1009	-14.85392	1.324096
Preop_ODI_Summary_Score	**-0.267077**	**0.06498**	**-4.11**	**<.0001***	**-0.394898**	**-0.139256**
Months to Prodedure	**3.0936525**	**1.358673**	**2.28**	**0.0234***	**0.421047**	**5.7662579**
Medical Group_Name A	4.6661071	2.844735	1.64	0.1019	-0.929687	10.261902
Medical Group_Name B	2.8349152	1.91636	1.48	0.1400	-0.9347	6.6045303
Medical Group_Name C	-3.26608	2.815123	-1.16	0.2468	-8.803625	2.271465

Effect Tests

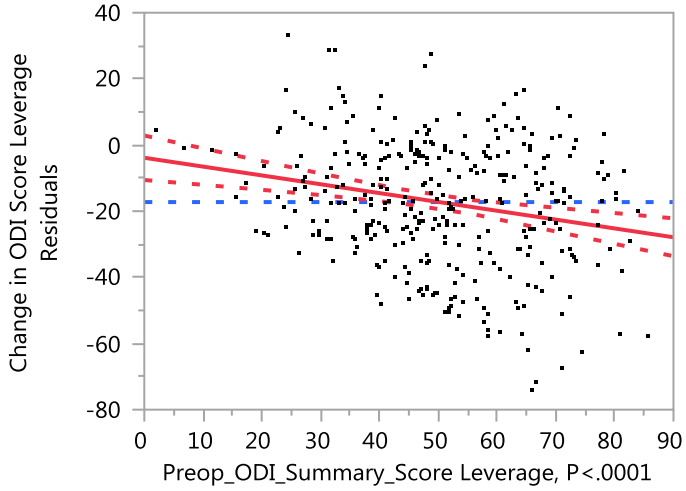
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Preop_ODI_Summary_Score	1	1	5871.7574	16.8931	<.0001*
Months to Prodedure	1	1	1802.0714	5.1846	0.0234*
Clinic_Name	3	3	4320.3830	4.1433	0.0067*

Residual by Predicted Plot



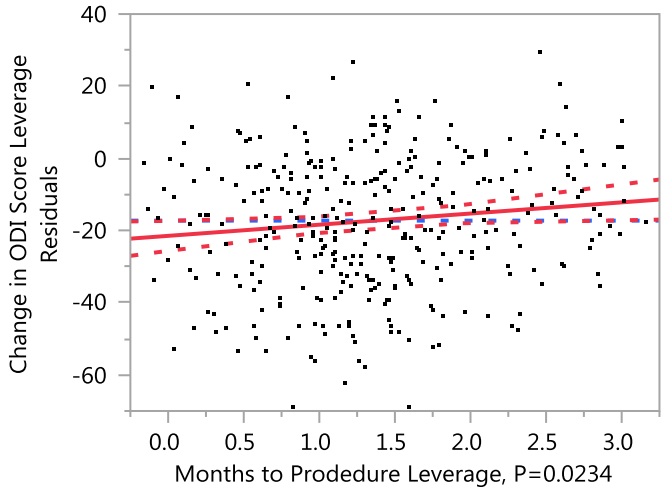
Preop_ODI_Summary_Score

Leverage Plot



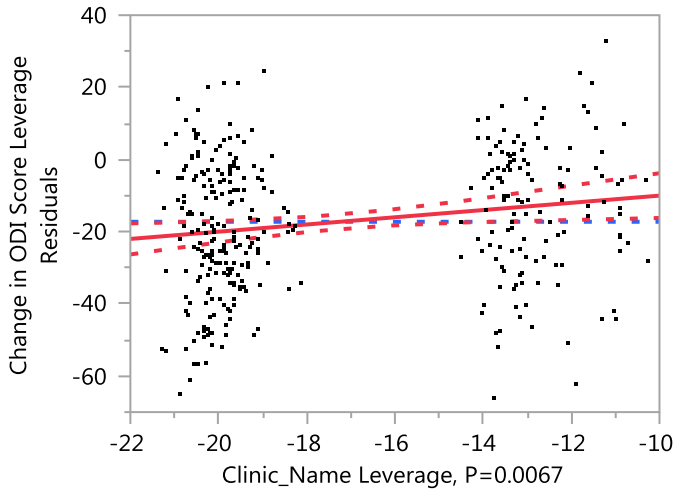
Months to Prodedure

Leverage Plot



Clinic_Name

Leverage Plot



Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
A	**-11.27330**	3.5269256	-10.679
B	**-13.10449**	1.8712146	-14.608
C	**-19.20548**	3.4761178	-18.483
D	**-20.17435**	1.3879773	-19.538

LS Means Plot by Medical Group