February 2018

Executive Summary of the Inpatient Rehabilitation Facilities Experience of Care Survey

Prepared for

Joseph Clift, EdD, MPH, MS, PMP
Centers for Medicare & Medicaid Services
7500 Security Boulevard
Baltimore, MD 21244

Prepared by

John D. Loft, PhD
Sara Zuckerbraun, MA
Alicia Frasier, MPH
Corina Owens, PhD
Celia Eicheldinger, MS
Gina Kilpatrick, MEd
RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709

CMS Contract No. HHSM-500-2013-13015I, Task Order HHSM-500-T0003
RTI Project No. 0214077.003.002

RTI INTERNATIONAL

EXECUTIVE SUMMARY

In September 2015, CMS contracted with RTI International to develop an experience of care (EOC) survey for patients who received care in inpatient rehabilitation facilities (IRFs). CMS emphasized the importance of including the perspective of the patient as well as the family or caregiver, since these people provide important support to the patient's rehabilitation both in the IRF and after discharge. To design the survey, RTI followed all research and development protocols in the *Blueprint for the CMS Measures Management System*, and in *Getting the CAHPS Trademark: A Guide for Survey Developers*. Analysis of data collected for the mode experiment is an important stage in testing, analyzing, developing risk adjustment models, and recommending protocols for this survey.

This executive summary is divided into three domains (see ***Table ES-1***) and within each we describe the mode experiment tasks and findings.

Table ES-1
Mode experiment domains

Domain	Task in mode experiment	Objective of mode experiment
Data collection	• Conduct data collection for the IRF EOC using three modes: mail-only, phone-only and mixed-mode. • Observe key data collection outcomes.	• Determine the most appropriate protocols for a CMS-administered IRF Experience of Care Survey. • Document findings and recommended protocols in National Quality Forum (NQF) Measure Submission, Testing Attachment, and Evidence Attachment forms.
Survey instrument	• Test the reliability, validity, and consistency of the survey instrument.	• Assess the scientific validity of the IRF EOC Survey and the feasibility for measuring and comparing IRF performance in experience of care. • Document findings in NQF Measure Submission, Testing Attachment, and Evidence Attachment Forms.
Risk adjustment and nonresponse	• Analyze the impact of patient characteristics and mode of survey response on survey outcomes. • Assess and correct, if necessary, nonresponse bias.	• Create a final mode and risk adjustment model for IRF performance scores. • Document model in NQF Measure Attachment and Testing Attachment Forms.

ES.1 Data Collection

IRF participation. RTI recruited 65 IRFs to participate in the mode experiment. They represented a diverse group in terms of number of beds, geographic location, urban versus rural, and whether the IRF was a freestanding facility or a unit in an acute care hospital. Most participating IRFs were already surveying all their patients using proprietary surveys. They agreed to suspend these surveys during April and May 2017 to participate in the mode experiment.

The sample for June data collection was based on April discharges and the sample for July was based on May discharges. Using a secure web portal, each month IRFs provided a file of all eligible discharged patients. Patient eligibility criteria included: alive, over the age of 18, with a home address in the continental United States, covered by any payer (or no payer), had a stay greater than or equal to 72 hours, and was discharged to any location. Patients with unplanned discharges were eligible but were treated somewhat differently in some of the analyses. Over the 2 months of the mode experiment, IRFs uploaded data for 7,726 eligible patients. The counts of both participating facilities and eligible patients exceeded targets established in the sample design.

Three modes of data collection. Data collection relied on three modes of survey administration: mail-only, telephone-only, and mixed-mode (mail with telephone follow-up). The survey was designed to obtain a minimum of 623 completed surveys per mode, or 1,869 in total. The target sample size and allocation were determined to have sufficient power to detect a 7-percentage point difference, assuming estimates around 0.7, 95% confidence, and 80% power.

Each month, RTI randomly assigned all eligible patients to one of the three modes. The timeline for the data collection of the three modes is shown in ***Table ES-2***.

Table ES-2.
Timeline for data collection

	June 2017		July 2017		August 2017	
April discharges mail-only mode	start 6/6/2017			finish 7/31/2017		
May discharges mail-only mode			start 7/6/2017			finish 8/31/2017
April discharges mixed-mode	start 6/7/2017				finish 8/1/2017	
May discharges mixed-mode			start 7/6/2017			finish 8/31/2017
April discharges phone-only mode	start 6/1/2017			finish 7/27/2017		
May discharges phone-only mode			start 7/3/2017			finish 8/28/2017

Mail-only mode. RTI sent a hardcopy survey, cover letter, and postage-paid return envelope to all mail-only patients. Although the survey was targeted to the patient or a proxy if appropriate, to protect patient confidentiality, most IRFs were only permitted to provide RTI with the name and contact information for the patient. The envelope was addressed to the patient's address, but the letter within was addressed to Dear [Patient Name] or Family Member. Both the letter and the survey explained that proxies were requested to complete the survey on behalf of patients unable to do so. After 4 weeks, RTI sent a second mail correspondence to nonrespondents, and data collection ended 8 weeks after the date of initial mailout.

Telephone-only mode. After completing an 8-hour training and certification, interviewers contacted either the patient or a proxy as appropriate. The initial telephone number was the one in the IRF's records for the patient. Up to 10 contact attempts were made but if an individual refused the survey, no subsequent contacts were made. Data collection ended after eight weeks. Interviews were conducted in English and Spanish.

Mixed-mode. This mode began in the same way as the mail-only mode. After 4 weeks, instead of a second mail correspondence to nonrespondents, we followed up with a telephone survey. Interviewers interviewed either the patient or a proxy as appropriate. Up to 10 contact attempts were made and data collection ended 8 weeks after the date of the initial mailout. Interviews were conducted in English and Spanish.

Data collection results. The final results are shown in ***Table ES-3***.

Table ES-3
Data collection results by mode

	Mail-only mode	Telephone-only mode	Mixed-mode	Total
Sample size	2,782	2,781	2,163	7,726
Completed surveys	897	570	830	2,297
Response rate	32.5%	20.5%	38.6%	29.9%

Data collection was completed without noteworthy problems. We encountered no distressed respondents or respondents who were upset with their IRF. In the final week of the survey, Hurricane Harvey struck and on August 24 the RTI telephone center stopped all outbound calls to affected areas in Texas and Louisiana. Mail service from the affected areas stopped and completed surveys could not reach RTI. We calculated that 0.7% of the IRF sample had addresses in Houston TX, Corpus Christi TX, Galveston TX, Pasadena TX, New Orleans LA, Beaumont LA, Lafayette LA, and Lake Charles LA. Because these were small percentages and data collection had slowed down during this final week of the survey, we estimate little to no impact from Hurricane Harvey.

At the end of data collection RTI provided each participating IRF a file showing their own survey results and a benchmark column of overall results of all participating IRFs.

ES.2 Survey Instrument Reliability, Validity, and Consistency

Measures of central tendency. The number of positive responses for some items was high, indicating a ceiling effect (lack of variance). Top-box responses on some measures skewed very positive while on other measures followed a normal distribution. This outcome is not uncommon in CAHPS studies, which often find that people rate their health care providers highly. The outcome may also be related to the many IRFs that were already surveying patients and, presumably, using this feedback to improve facility performance. Greater variability may exist in a wider range of sites and over a longer survey period.

We found missing data in all items, but among most of the core items of Q1 through Q41, the summative rate of missing responses was below 2.5%. Compared to the core questions, patient demographic questions had higher rates of missingness, notably patient's Hispanic origin, race, and language spoken at home. RTI recommends hot-deck imputation to impute missing values for patient demographic variables that are used as patient mix adjusters.

Factor structure. The factor structure was established during the field test survey conducted in 2016 (see ***Table ES-4***).

Table ES-4
IRF EOC survey factor structure

Factor	Component questions	Top-box categories
Measure 1, Goal Setting and Monitoring	Q1, Q2, Q25, Q26	Yes, definitely or Always
Measure 2, Communication with Staff at the IRF	Q4, Q5, Q7, Q8, Q10, Q11, Q13, Q14, Q16, Q17, Q19, Q20, Q22, Q23, Q24, Q27	Always or Yes, definitely
Measure 3, Experience at this IRF	Q28, Q29, Q30, Q31, Q32, Q33, Q35, Q36	Always or strongly agree
Measure 4, Preparing for Leaving the IRF	Q37, Q38, Q39	Yes, definitely
Global Rating 1, 0-10 rating	Q40	9 or 10
Global Rating 2, likelihood of recommending	Q41	Yes, definitely

We reassessed the survey item sets in the Table ES-4 factor structure for dimensionality using confirmatory factor analysis (CFA). Fit statistics were above acceptable values in all areas, as shown in ***Table ES-5***.

Table ES-5
Confirmatory factor analysis fit statistics from survey-based structure

Assessment criteria	Value	
	Analytic	Acceptable
Root Mean Square Error of Approximation (RMSEA)	0.058 (CI = 0.056-0.060)	< 0.08 (Cangur & Ercan, 2015)
Comparative Fit Index (CFI)	0.954	0.90 at a minimum (Hu & Bentler, 1999)
Tucker-Lewis index (TLI)	0.950	0.90 at a minimum historically, 0.95 indicates good fit
Nonsignificant factor loadings	None	—

Rasch analysis. Fit statistics were evaluated to identify any items producing unexpected response patterns (misfit values above 1.3). Only two items—Q4 and Q28—were identified. In addition, we investigated measurement redundancy, that is, the possibility that a single item did not contribute value to the full set of items in the factor. Statistical analysis suggested that only one item, Q4 (Nursing aides/assistants treat patients with courtesy and respect), might be redundant. Subject matter experts reviewed this item and advised that it could provide IRFs with valuable substantive feedback for quality improvement and therefore contributes substantively to the full set of items. For this reason, RTI recommends that this item remain in the survey. All four factor structures (measures) met assumptions of unidimensionality and local independence.

Internal consistency. We assessed reliability for the survey, both at the composite and total score levels, using the Cronbach's alpha coefficient. As shown in ***Table ES-6***, all values exceeded the cutoff criterion of 0.80, indicating adequate reliability or consistency of scores.

Table ES-6
Internal consistency estimates for the overall survey and each composite

Composite	Cronbach's alpha
Overall Survey	0.95
Goal Setting and Monitoring	0.80
Communication with Staff at the Rehabilitation Hospital/Unit	0.91
Experience at this Rehabilitation Hospital/Unit	0.87
Preparing for Leaving the Rehabilitation Hospital/Unit*	0.81

*Note. This composite was not computed for patients with unplanned discharges

Interclass reliability. This analysis determines how much of the variation in composite scores across the facilities is due to true variation versus chance or measurement error. This measure of reliability assesses the variation in responses within facilities relative to variation between facilities. We considered the top-box scores, that is, the item responses that were most

favorable to the facilities. We considered all facility scores after applying the final risk adjustment model.

We used the Spearman-Brown prophecy formula to estimate intraclass correlation coefficients (ICC) as the within-facility sample size increased. These estimates can help determine what sample size is needed to appropriately differentiate between facilities. All composites will provide acceptable ICC values equal to or above 0.70 when each facility uses the recommended risk adjustment model and has a minimum of 240 responding surveys (see ***Table ES-7***).

Table ES-7
ICCs reliability estimates for each composite (risk-adjusted scores)

Composite measures	ICC (N = least 40 per facility)	ICC (N = 240 per facility)
Goal Setting and Monitoring	0.54	0.76
Communication with Staff at the Rehabilitation Hospital/Unit	0.46	0.70
Experience at this Rehabilitation Hospital/Unit	0.54	0.76
Preparing for Leaving the Rehabilitation Hospital/Unit	0.53	0.75

Differential item functioning. RTI further assessed the survey instrument to determine if results were comparable across different patient groups. We examined patient subgroups formed by age (64 and younger, 65 to 74, 75 and older), primary impairment category (neural versus orthopedic versus other), patient admission self-care functioning (three groups divided by percentiles), patient admission mobility functioning (three groups divided by percentiles), gender, and type of respondent (patient versus proxy). Q1, Q17, and Q22 showed difference by age. The findings in these analyses informed us that this patient factor (age) is likely to be impactful in the patient mix (risk) analysis.

ES.3 Patient Mix and Nonresponse Bias Analysis

For the IRF performance scores to provide objective estimates of experience of care and meaningful comparisons between IRFs, adjustments are needed to account for significant sources of bias in the survey results that are outside the IRFs' control.

We conducted statistical analyses to evaluate the relative impact of each potential factor in the presence of the other patient factors. The goal was to determine an appropriate statistical adjustment protocol that adjusts facility performance scores up or down based on the characteristics of the responding patient population from each facility. We conducted a correlation analysis on the independent (i.e., all the patient risk) variables. Highly correlated independent variables can cause problems for estimating regression models when both of the correlated variables are included in the models. We calculated both Pearson correlation

coefficients and variance inflation factor (VIF) statistics to identify changes needed in the proposed set of patient mix variables.

We estimated 33 multivariate regression models—one for each survey item comprising the four composites plus the two global rating items. The individual patient-respondent was the unit of analysis. From the outset we included mode and all potential patient risk variables in all regression models. We fit a facility indicator variable as a fixed effect in all models to isolate the effects of potential mode and patient risk factors from the IRFs' own characteristics of providing care. We sequentially dropped independent variables that were not statistically significant in any of the 33 regression models or were statistically significant in only one or two of the regression models. To determine the best model, we created a set of facility-level scores from the predicted values of each regression model and compared the results to determine the impact of dropping variables on the facility-level predicted values.

The final recommended model includes mode of data collection and the following 13 patient risk variables: patient age, length of stay, overall health, overall mental health, marital status, education, ethnicity, race, language spoken at home, patient self-care functioning, patient mobility functioning, impairment category, and type of respondent. These variables were derived from both from the frame data provided by the IRFs and from the survey. When applying the statistical adjustments to the facility-level scores, we recommend using hot-deck imputation to fill missing data values for any respondents who had missing data for these patient risk variables.

Nonresponse bias analysis. We conducted a logistic regression analysis that included all patient variables known for both respondents and nonrespondents, as well as facility stratification variables of patient volume, urban/rural, and freestanding/unit. The logistic regression analysis revealed that older patients, males, patients with more than a 16-day stay, patients with lower mobility functioning, patients whose impairment group is neural, and patients from freestanding IRFs, IRFs in urban locations, and smaller IRFs had statistically significant lower response propensity. Thus, these predictors should be included in the calculation of the nonresponse-adjusted weights. We included these variables in the final logistic regression model and output each respondent's predicted response propensity. We calculated each respondent's nonresponse-adjusted weight as the reciprocal of the predicted response propensity. Finally, we calculated the Pearson correlation coefficients between the nonresponse-adjusted weights and the residuals from regression models including mode and the final set of patient risk factors for all 33 survey items.

This correlation analysis examined if patient and facility factors significantly affecting nonresponse should also be used in creating patient risk-adjusted scores with nonresponse-adjusted weights. We found no statistically significant correlations between the non-response weights and the residuals from regression models including mode and the final set of patient risk factors for all 33 survey items. We conclude that, when using our final risk adjustment model, nonresponse-adjusted weights are not needed to further adjust the patient risk-adjusted facility scores.