February 2018

Executive Summary of the Long-term Care Hospital 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 survey (EOC) for patients who received care in long-term care hospitals (LTCHs). CMS emphasized the importance of including the perspective the patient as well as family or caregivers, especially in situations when the patient is deceased or is unable to complete the survey due to cognitive or physical conditions. 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 LTCH EOC using three modes: mail-only, mixed-mode, and in-facility. • Observe key data collection outcomes.	• Determine the most appropriate protocols for a CMS-administered LTCH EOC 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 LTCH EOC Survey and the feasibility for measuring and comparing LTCH performance in experience of care. • Document findings in NQF 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 adjustment model for LTCH performance scores. • Document model in NQF Measure Attachment and Testing Attachment Forms.

ES.1 Data Collection

LTCH participation. RTI recruited 69 LTCHs (reporting under 62 CMS certification numbers (CCNs)) to participate in the mode experiment. They represented a diverse group in terms of number of beds, geographic location, and urban versus rural. Reflecting the LTCH population, about one-half of the participating LTCHs were part of large health systems with

multiple LTCHs. Most of the participating LTCHs were already surveying all their patients using proprietary surveys. They agreed to suspend these surveys during April, May, and June 2017 to participate in the mode experiment.

The sample for data collection was based on patients discharged in April, May, and June. Using a secure web portal, each month LTCHs provided a file of all eligible discharged patients. Patient eligibility criteria included: 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, was discharged to any location, deceased or alive at the end of the stay. Over the 3 months of the mode experiment, LTCHs uploaded data for 6,695 eligible patients. The counts of both participating hospitals 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, mixed-mode (mail with telephone follow-up), and in-facility. The survey was designed to obtain a minimum of 394 completed surveys per mode, or 1,182 in total. The target sample size and allocation were determined to have sufficient power to detect an effect size of 0.20 standard deviation (SD) in comparing survey item estimates between modes.

RTI randomly assigned all eligible April and May discharged patients to either the mail-only mode or the mixed-mode, and RTI conducted these surveys in June and July, respectively. All June discharged patients were assigned to the in-facility mode. 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 6/7/2017				finish 8/2/2017	
April discharges mixed-mode	start 6/7/2017				finish 8/3/2017	
May discharges mail-only mode			start 7/6/2017			finish 8/31/2017
May discharges mixed-mode			start 7/6/2017			finish 8/31/2017
June discharges in-facility mode	start 6/1/2017					finish 8/31/2017

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 LTCHs 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.

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. After completing an 8-hour training and certification, interviewers contacted either the patient or a proxy as appropriate. Up to 10 contact attempts were made but if an individual refused to complete the survey, no subsequent contacts were made. Data collection ended 8 weeks after the date of the initial mailout. Interviews were conducted in English and Spanish.

In-facility survey mode. From June 1st through 30th LTCH staff distributed blank surveys to all eligible patients at the time of their discharge. Including the voice of deceased patients and very ill patients was important to CMS. LTCHs used their judgement as to whether to give the survey to the patient or a family member or other responsible party instead of to the patient. Instructions and FAQs from RTI provided tips and suggested language for distributing and answering questions about the survey. Patients/family members either completed the survey on site and deposited it in an RTI-provided lockbox in the LTCH, or they brought it home to complete and mail in later.

Due to concerns with data quality and sample integrity, AHRQ'S CAHPS Consortium has not supported in-facility distribution of surveys, but is currently conducting research into this approach. CMS has not used an in-facility mode for its surveys prior to this mode experiment. Concerns with this approach include: it adds burden to staff; it is prone to systematic sampling error that can cause bias; and patients may believe that their responses are available to hospital staff, leading them to report more positive experiences than they had. However, the LTCH Technical Expert Panel (TEP) members who work in LTCHs advocated strongly for testing the in-facility distribution mode in the mode experiment due to its operational advantages. These advantages include: lower cost to LTCHs compared to printing, mailing, and telephone interviewing; faster return of results leading to more timely quality feedback; less confusion for patients with multiple facility stays (about which hospital is the subject of the survey); LTCH staff knowledge can be leveraged to physically give the survey to the most appropriate respondent (patient or the family/caregiver proxy), which may be far preferable to mailing the survey to the patient's home address (where he/she may not be living). Testing the feasibility, and potential bias of this mode, was a key objective of this mode experiment.

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

Table ES-3
Data collection results by mode

	Mail-only mode	Mixed-mode	In-facility mode	Total
Sample size	2,156	2,160	2,379	6,695
Completed surveys	318	517	529	1,364
Response rate	14.7%	23.9%	22.2%	20.4%

Telephone interviewers encountered a few respondents who were upset with their LTCH and, per protocol, gave respondents the CMS Support number. 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 2.6% of the LTCH 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 a small percentage and data collection slowed down during this final week of the survey, we estimate little to no impact from Hurricane Harvey.

The in-facility data collection encountered one significant problem. LTCH staff were asked to write the patient's Medical Record Number (MRN) on the blank surveys they handed to the patients. Since the MRN appeared on the returned survey (which had no other identifying information) and in the LTCH's files of eligible patients, it served as the only link between patients' survey responses and sample frame data that the LTCH reported: age, admission function score, length of stay, and primary medical condition category. Among the 529 returned in-facility surveys, we were unable to use the MRN to link 184 completed surveys to the patient's frame data. These errors occurred when an MRN was not included on the survey prior to in-facility distribution, when the MRN was illegible, or when the MRN did not match up to any MRN in the file from the LTCH. These appeared to be random errors that occurred for most LTCHs. Because of the inability to link to the sample frame data, some cases could not be included in the differential item functioning analysis, patient mix analysis, and nonresponse bias analyses. At the end of data collection RTI provided each participating LTCH a file showing their own survey results and a benchmark column of overall results of all participating LTCHs.

ES.2 Survey Instrument Reliability, Validity, and Consistency

Measures of central tendency. In general, responses to the survey items were distributed across negative and positive responses though positive responses are prevalent.

We found missing data in all items to some extent, but among most core items of Q1 through Q31, 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 overall health and patient's overall mental or emotional health. RTI recommends hot-deck imputation to impute missing values for patient demographic variables 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
LTCH EOC survey factor structure

Factor	Component questions	Top-box categories
Measure 1, Goal Setting and Monitoring	Q1, Q2, Q3, Q14	Yes, definitely
Measure 2, Communication with Staff at the LTCH	Q5, Q6, Q8, Q9, Q11, Q12, Q13, Q15, Q16, Q17	Always or yes, definitely
Measure 3, Experience at this LTCH	Q18, Q19, Q20, Q21, Q22, Q24, Q25, Q26,	Always or strongly agree
Measure 4, Preparing for Leaving the LTCH	Q28, Q29, Q30, Q31	Yes, definitely
Global Rating 1, 0-10 rating	Q32	9 or 10
Global Rating 2, likelihood of recommending	Q33	Yes, definitely

We used confirmatory factor analysis (CFA) to determine if the LTCH factor structure in Table ES-4 was the best representation of the data structure. We found in the specification sample that the fit statistics were above acceptable in all areas, as shown in ***Table ES-5***. The validation sample was then used to confirm the fit statistics of the factor structure. Again, fit statistics exceeded acceptable values in all areas, as shown in ***Table ES-6.***

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

	Value	
Assessment criteria	Analytic	Acceptable
Root Mean Square Error of Approximation (RMSEA)	0.070 (CI = 0.066-0.075)	< 0.08 (Cangur & Ercan, 2015)
Comparative Fit Index (CFI)	0.973	0.90 at a minimum (Hu & Bentler, 1999)
Tucker-Lewis index (TLI)	0.970	0.90 at a minimum historically, 0.95 indicates good fit
Nonsignificant factor loadings	None	—

Table ES-6
Confirmatory factor analysis fit statistics from factor structure with validation sample

Assessment criteria	Value	
	Analytic	Acceptable
Root Mean Square Error of Approximation (RMSEA)	0.066 (CI = 0.062-0.070)	< 0.08 (Cangur & Ercan, 2015)
Comparative Fit Index (CFI)	0.981	0.90 at a minimum (Hu & Bentler, 1999)
Tucker-Lewis Index (TLI)	0.979	0.90 at a minimum historically, 0.95 indicates good fit
Nonsignificant factor loadings	None	—

Rasch analysis. We found no evidence of measurement redundancy. Fit statistics were evaluated to determine any items producing unexpected response patterns (misfit values above 1.3), and only four items—Q5, Q14, Q18, and Q19—were identified. Subject matter experts reviewed these and advised that all four items could provide LTCHs with valuable feedback for quality improvement. Based on this advice, RTI recommends they 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-7***, nearly all values exceeded the cutoff criterion of 0.80, indicating adequate reliability or consistency of scores.

Table ES-7
Internal consistency estimates for the overall survey and each composite

Composite	Cronbach's alpha
Overall Survey (Q28)*	0.96
Overall Survey (Q29 and Q30)*	0.97
Goal Setting and Monitoring	0.89
Communication with Staff at the LTCH	0.93
Experience at this LTCH	0.92
Preparing for Leaving the LTCH (Q28)*	0.83
Preparing for Leaving the LTCH (Q29 and Q30)*	0.78

*Note. The reliabilities for the overall survey were computed separately for patients discharged to different places. Patients discharged to the community were asked one set of discharge questions while patients discharged to another facility were asked a separate set.

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. In our assessment of within versus between variation among 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 us 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 110 responding surveys (see ***Table ES-8***).

Table ES-8
ICCs reliability estimates for each composite (risk-adjusted scores)

Composite measures	ICC (N=least 20 per facility)	ICC (N = 110 per facility)
Goal Setting and Monitoring	0.63	0.80
Communication with Staff at the LTCH	0.61	0.79
Experience at this LTCH	0.50	0.71
Preparing for Leaving the LTCH	0.55	0.74

Differential item functioning. RTI further assessed the survey instrument to determine if results were comparable across different patient groups. We looked at the patient subgroups formed by age (64 and younger, 65 to 74, 75 and older), medical condition type (acute onset versus chronic), patient functioning (above versus below the population median), gender, and type of respondent (patient versus proxy). Q8 and Q12 showed difference by age and Q18 showed difference by type of respondent.

In addition, we also conducted a separate examination assessing the impact of survey administration mode on survey scores. We conducted six multivariate analyses, which examined each composite item and the two global ratings as dependent variables; survey administration method was used as a predictor in the model. Covariates included gender, patient age, highest education level, respondent type, patient functioning, and patient's primary medical condition. Based on a review of all six models, the survey administration mode was significant, while holding all variables present in the model constant. Those who were given the survey from the LTCH staff at discharge had more positive overall rating compared to those who received their surveys in the mail or by phone.

The findings in these analyses informed us that these patient factors and survey mode were likely to be impactful in the patient mix (risk) analysis.

ES.3 Patient Mix and Nonresponse Bias Analysis

For the LTCH performance scores to provide objective estimates of experience of care and meaningful comparisons between LTCHs, adjustments are needed to account for significant sources of bias in the survey results that are outside the LTCHs' control. As was shown in the differential item functioning analysis, potential sources of bias include data collection mode chosen by the LTCH (in-facility, mail-only, mixed-mode); variability in various patient risk factors such as demographic characteristics, disease status, and health status; and variation in response propensity across patients within facilities.

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. As part of this goal, we explored whether a model without sample frame data would perform at acceptable levels compared to a model that included sample frame data. We did this analysis because of the difficulty in obtaining accurate and complete sample frame data for surveys completed using the in-facility mode.

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 28 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 model and mode and patient risk factors from the LTCH's own characteristics of providing care. We sequentially dropped independent variables that were not statistically significant in any of the 28 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 8 patient risk variables: patient age, patient sex, overall health, overall mental health, marital status, education, race, and type of respondent. Note that all variables come from the survey data eliminating the need to match a patient back to his/her frame data for patient mix adjustment. 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 left these questions unanswered.

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 number of beds and urban/rural. This logistic regression revealed that

younger patients, female patients, patients with length of stay longer than 33 days, patients with admission functioning scores equal or below median, patients from LTCHs in urban locations, and larger sized LTCHs 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 28 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 nonresponse adjusted weights and the residuals from regression models including mode and the final set of patient risk factors for all 28 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.