October 2025

ADULT CARE HOME ACCREDITATION PILOT PROGRAM
Session Law 2021-180
Senate Bill 105

Final Report: Evaluation of Pilot Program

Prepared for

North Carolina Joint Legislative Oversight Committee on Health and Human Services
North Carolina Department of Health and Human Services

Prepared by

Sheryl Zimmerman, PhD
Lea Efird-Green, MSW, MPA
Philip D. Sloane, MD, MPH
John S. Preisser, PhD
Scott Davis, PhD
David Reed, PhD
Walter Moczygemba, MSW
University of North Carolina at Chapel Hill

and

Katrice M. Perry, CHES
Aja R. Johnson, BA
Toni C. Parker, BA
London Grantham, MPH
Johanna Silbersack Hickey, MSW
University of North Carolina at Chapel Hill

EXECUTIVE SUMMARY

This *Evaluation of Pilot Program* report is the second of two reports providing the results of a three-year evaluation of a pilot program studying the effectiveness of an accreditation process for adult care homes/assisted living (AL) communities across the state of North Carolina. The first report is the *Evaluation of Quality Outcome Measures*, submitted in July 2025; together, the reports provide the results of a comprehensive evaluation. The Cecil G. Sheps Center for Health Services Research at the University of North Carolina at Chapel Hill conducted data collection and analyses. Accreditation efforts were overseen by the Accreditation Commission for Health Care (ACHC), and a Stakeholder Advisory Group provided recommendations throughout the project.

Initially, 146 AL communities diverse in location and quality based on the state-assigned Star rating were recruited and randomly assigned to either an "accreditation arm" to pursue ACHC accreditation (a designation indicating a standard of care is met) or a "control arm." Quality outcome indicators were collected quarterly or biannually over 2½ years in a longitudinal parallel arms community randomized trial design. A total of 113 communities participated; of the 58 communities randomized to the accreditation arm, 25 became accredited.

New data in this *Evaluation of Pilot Program* report detail the accreditation process and compare licensure compliance between communities in the accreditation arm and control arm; and among communities in the accreditation arm, between before and after they became accredited (i.e., not accredited condition versus accredited condition). Using secondary state and county data over slightly less than four years, analyses examined standard deficiencies (including medication management as a category within standard deficiencies); Type B violations that are detrimental to resident health, safety, or welfare; Type A violations that result in risk of or actual death or serious physical harm, abuse, neglect, or exploitation; and overall deficiencies/violations. The report also includes methodologic analyses of study power and whether accreditation at least maintains resident outcomes based on quality measures.

Findings indicate that the ACHC accreditation process was efficient, with communities receiving documentation from the accrediting body within eight days. In relation to licensure compliance among AL communities in the accreditation arm, those that became accredited had significantly fewer Type B state violations, as well as notably (but not statistically significantly) fewer state total deficiencies and violations, including Type B and Type A violations.

Quality outcomes reported in the earlier *Evaluation of Quality Outcome Measures* report indicated that of the 31 outcome indicators, four showed statistically significant differences across arms or conditions, three (75%) of which favored communities in the accreditation arm or communities that became accredited. Also, 22 of the 31 outcome indicators were not statistically significantly different but were notably different due to the magnitude of their odds ratios or percent change -- 16 (73%) of which favored accreditation.

Methodologic analyses indicate that overall, the communities participating in the project – and most importantly those randomized to the accreditation arm or control arm, and those that did and did not become accredited – did not differ significantly from other communities, thus supporting the validity of the study findings based on available data. In addition, the study had good power for most outcomes and was largely successful in detecting small and moderate changes related to accreditation. In this sense, the study provided clarity on the quality of the outcome measures as well as on those outcomes that were associated with accreditation.

Because communities continued to receive regular state/county surveys throughout the project, findings reflect the potential "value added" of accreditation within the existing regulatory framework, rather than the value of accreditation itself. Other contextual factors including the selection of outcome indicators, the low incidence rates of certain indicators, baseline differences between study arms despite randomization, modest accreditation uptake, and workforce challenges may have further attenuated the potential impact of accreditation.

Considering the licensure compliance findings alongside the quality outcomes findings, the fact that the preponderance of significant and notable outcomes favored accreditation is striking. That so few reached the level of statistical significance may in part relate to the small sample size in the accredited condition, choice of outcomes to study, low event rates, and the ongoing state/county survey process that limited the amount of potential change. Thus, there may be more benefit from accreditation than was able to be determined through formal hypothesis testing that relies on sufficiently large sample size and context.

Overall, the licensure findings and quality outcome indicators are promising, but evidence is limited that accreditation improves (or worsens) outcomes due in part to the study context. Further, the study was not designed to determine whether accreditation independently maintains outcomes. Therefore, rigorous evaluation should be an integral component if accreditation is incorporated into future legislation.

CONTENTS

1. GOAL

The goal of this three-year pilot program was to evaluate the effectiveness of an accreditation process for adult care homes (also referred to as assisted living [AL]) across the state of North Carolina. The project studied the effectiveness of accreditation through an evaluation of quality outcome measures and compliance with licensure requirements to determine whether accreditation improves or maintains quality of care compared to a control group. The Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill conducted data collection and analyses. Accreditation was overseen by the Accreditation Commission for Health Care (ACHC), and a Community Stakeholder Advisory Group provided guidance.

Methodologically, the goal of the project was to engage a maximum of 150 diverse AL communities, randomize one-half to an accreditation arm and one-half to a control arm, and obtain data on change in care and resident outcomes quarterly for 2½ years (7/1/22 through 12/31/24), and change in compliance with licensure requirements for slightly less than four years using secondary data available from 11/1/21 through 8/31/25. It was intended that the number of communities in both arms be equivalent, and that diversity be reflected in payor source, Star (quality) rating, and related characteristics. Care and outcomes were evaluated in five categories: resident outcomes and care coordination/transitions (health care related), resident outcomes and care coordination/transitions (psychosocial related), person-centered care, medication management, and workforce. Compliance with licensure requirements was evaluated by adherence to mandatory staffing levels (a measure of workforce) and the identification of deficiencies by the North Carolina Department of Health and Human Services (NC DHHS) Division of Health Service Regulation and county Departments of Social Services.

This *Evaluation of Pilot Program* report provides details related to the accreditation process; data related to compliance with licensure requirements; and additional methodologic analyses of the accreditation project. The previous *Evaluation of Quality Outcome Measures* report provides additional details regarding the methods (including recruitment, enrollment, participation, and statistical analyses) and provides the quality outcome data and results, the main findings of which are summarized below.

2. SUMMARY OF PREVIOUS REPORT: EVALUATION OF QUALITY OUTCOME MEASURES

As reported in the Executive Summary of the *Evaluation of Quality Outcome Measures* report, 146 communities diverse in location and quality based on the state-assigned Star rating were recruited and randomly assigned to either an “accreditation arm” to pursue ACHC accreditation (a designation indicating that a standard of care is met) or a “control arm.” A total of 113 communities participated and provided data for 31 quality indicators in five categories: resident outcomes and care coordination/transitions (health care related); resident outcomes and care coordination/transitions (psychosocial related); person-centered care; medication management; and workforce.

Of the 58 communities in the accreditation arm, 25 became accredited. Comparisons were made between study arms and, among communities in the accreditation arm, between before and after they became accredited (referred to respectively as the not accredited and accredited condition, noting that communities that never became accredited provide only "not accredited" data).

- Of the 31 outcome indicators, four showed statistically significant differences across arms or conditions in two categories (medication management and workforce), three (75%) of which favored communities in the accreditation arm or communities that became accredited: medication administration errors, turnover of care aides, and turnover of administrators. The outcome that did not evidence favorable change was staffing sufficiency of administrators in special care units.
- Of the 31 outcome indicators, 22 were not statistically significantly different but were notable due to the magnitude of their odds ratios or percent change -- 16 (73%) of which favored accreditation in four categories: resident outcomes and care coordination/ transitions (both health care related and psychosocial related), medication management, and workforce.
- In total, 23 different outcomes were statistically significantly different or notably different.

Because communities received regular state/county surveys throughout the project, findings reflect the potential "value added" of accreditation within the existing regulatory framework, rather than the value of accreditation itself. Other contextual factors, including selection of outcome indicators, low incidence rates of certain indicators, baseline differences between study arms despite randomization, modest accreditation uptake, and workforce challenges may have further attenuated the potential impact of accreditation.

In sum, the fact that the preponderance of significant and notable outcomes favored accreditation is striking, and that so few reached the level of statistical significance may in part relate to the choice of outcomes to study, low event rates for some quality indicators, the ongoing state/county survey process that limited the amount of potential change, and the small sample size of the accredited condition. The latter resulted from (i) moderately low rate of accreditation (i.e., 25 of 58 accreditation arm communities became accredited); (ii) accreditation tending to occur late in the follow-up period; and (iii) dropout of communities prior to the end of follow-up that exacerbated the impact of (i) and (ii) on sample size. Thus, there may be more benefit from accreditation than was able to be determined through formal hypothesis testing that relies on sufficiently large sample size and context. Nonetheless, despite promising results, there is limited evidence that accreditation improves (or worsens) outcomes, due in part to the study context. Further, the study was not designed to determine whether accreditation independently maintains outcomes.

3. METHODS

Sample. As described in detail in the July 2025 *Evaluation of Quality Outcomes Measures* report, the sampling frame for the project included 564 AL communities identified from the directory of the NC DHHS that were operating and assigned a Star rating as of May 2022. To include a broad and representative sample, sampling was stratified based on both geographic

region and state-assigned quality Star rating.[1] First, the proportion of all communities as per their representation in the three adult care licensure branches -- Western Branch, Central Branch, and Eastern Branch -- was determined. Second, within each region, AL communities were further stratified based on two groups of Star ratings: 0-2 (lower quality) and 3-4 (higher quality), resulting in six strata. Then, up to 150 communities were randomly selected for participation proportionate to (a) the number of communities represented by the branch, and (b) the number of communities represented by Star rating stratum within the branch. (Note: the maximum/target number to be recruited was 150; as described later in the document, 146 were recruited during the time allotted).

Eligibility. To be eligible for participation, an AL community had to be licensed by the state of North Carolina and have a Star rating as of May 2022. Communities were not eligible for participation if they planned to close (or had already closed) at the time of recruitment, were currently accredited or were already participating in an accreditation program, were considering pursing accreditation through an alternative accrediting body during the three-year period of study, or did not agree to be randomized into the control or accreditation arm.

Recruitment. The North Carolina Senior Living Association (NCSLA) and the North Carolina Assisted Living Association (NCALA) distributed materials to their membership about the project, encouraging participation if communities were invited to participate; other stakeholder organizations and individuals did the same. Informational material included a one-page overview of the project, a more detailed two-page description, and a website that included a video presentation. It was also made clear that communities would be provided $500/quarter in recognition of the time and effort to compile information for the evaluation.

Project analysts randomly numbered AL communities within each stratum to determine the order in which to solicit participation; once selected, a mailing was sent via postal mail, followed by a telephone call from a research team member approximately five business days later. Within the ensuing four weeks, team members placed up to eight follow-up contacts by telephone and email (no more than three contacts per week) to discuss the project and solicit participation.

Communities within each stratum were recruited on a rolling basis, allowing for an initial 50% refusal rate. Based on the observed rates, additional invitations were sent to 25% of the remaining communities per stratum (excepting communities in the eastern branch, which were already reaching target numbers of participation). Once the target number of communities within a stratum was reached, active outreach was discontinued. In total, 368 of the 564 communities across the state were invited to participate in the project (65%). After 146 communities were recruited, they were randomly assigned to either the control or accreditation arm.

[1] NC Division of Health Service Regulation Adult Care Licensure Section. Star Rating Program. https://info.ncdhhs.gov/dhsr/acls/star/index.html

Accreditation Process. The Accreditation Commission for Health Care (ACHC) offered communities that were randomized to the accreditation arm the opportunity to obtain a nationally recognized validation of excellence in delivering care and services in a residential community setting. Through a comprehensive and objective survey process, ACHC evaluated adherence to established care practices, safety protocols, and compliance with laws and regulations. The process included a thorough review of communities and compliance with Life Safety Codes.

Before beginning the survey process, ACHC provided each community the accreditation standards, which are inclusive of the rules and statutes set forth by the NC DHHS. In addition, ACHC recorded a webinar for Survey Preparedness made available to each community to introduce administrators and staff to accreditation and ACHC's survey process, and developed a series of ten provider education webinars detailed in the *Evaluation of Quality Outcome Measures* report to provide tips for implementing a quality Performance Improvement Program. The goal for communities that were successful in achieving accreditation was that they use ACHC resources to improve practices and promote better resident outcomes.

ACHC's standards extend beyond the NC DHHS rules and statutes to include a section on Quality Outcomes and Performance Improvement. They apply to the planning and implementation of a performance improvement program, which must include who is responsible for the program; the activities being monitored; the process of data compilation; and corrective measures developed based on the data and outcomes. Also, ACHC requires the creation and implementation of a written Policy and Procedure document for many daily activities/tasks that may be considered common practice; this process minimizes the risk of misunderstanding a process, task, or order of events, as having a written policy or procedure leaves little room for misinterpretation.

Following the conclusion of the accreditation survey, ACHC's Clinical Review team generated a Summary of Findings that described all ACHC Accreditation Standards that were marked as a deficiency during the accreditation survey; the Summary was sent to the provider via email along with the Plan of Correction template, all of which were to occur within 10 business days from the last day of the survey. Accreditation cannot be granted until a Plan of Correction is submitted and approved. Due dates were as follows:

- The Plan of Correction was due to ACHC within 10 calendar days from the date of the organization's Accreditation Pending letter.
- If adjustments to the Plan of Correction were necessary, the organization must submit modifications to achieve an approved Plan of Correction within 10 days. Failure to submit an approved Plan of Correction within the time frame may result in a change of accreditation status from Accreditation Pending to Denial of Accreditation.
- If requested, evidence to support the implementation of the organization's Plan of Correction was due to ACHC with 45 calendar days following the date of the organization's Accreditation Pending letter. Failure to submit requested evidence results in termination of accreditation.

Licensure Data. Data in this report compare licensure compliance between study arms, and among communities in the accreditation arm, between timepoints (quarters) they were not accredited to timepoints following their accreditation (referred to respectively as the not accredited and accredited conditions). The data were derived from state and county sources.

- State data were collected from the state website that publishes the results of regularly scheduled state surveys and surveys conducted in response to a complaint or as a follow-up to a previous survey. These data were collected from Star rating worksheets that report standard deficiencies in fundamental rule areas: physical plant requirements; admission and discharge requirements; resident assessment and care plan; resident care and services; medication management; use of physical restraints and alternatives; residents' rights; and for communities with a special care unit (SCU), SCUs for Alzheimer's disease and related disorders and SCUs for mental health disorders. The Star rating worksheets also report Type B and Type A violations.
- County data were derived from Corrective Action Reports, obtained directly from the state because they are not uniformly publicly available. Most Corrective Action Reports that report deficiencies of any type are the result of a complaint to that county's Department of Social Services, but some are the result of routine monitoring. These reports contain standard deficiencies in all rule areas and Type B and A violations.

Licensure data examined four deficiencies/violations: standard deficiencies, medication-management standard deficiencies, Type B violations, and Type A violations.

- Standard Deficiency Citations: These deficiencies are reported for actions or situations that do not comply with laws and regulations related to licensure requirements; they do not result in immediate risk or detriment to the well-being of residents. (https://info.ncdhhs.gov/dhsr/acls/star/search.asp)
- Medication-Management Standard Deficiency Citations: These deficiencies are a subset of the standard deficiency citations; they were included in the examination of standard deficiencies and also examined separately. They are only labeled as such in state data, and so are not included in county-level analyses.
- Type B Violations: These violations are "detrimental to the health, safety, or welfare of any resident," but do not rise to the level of risk or harm that would result in a Type A violation. (https://info.ncdhhs.gov/dhsr/acls/star/search.asp)
- Type A Violations: These violations are the most serious, resulting in "death or serious physical harm, abuse, neglect, or exploitation of a resident" or a substantial risk of those outcomes. (https://info.ncdhhs.gov/dhsr/acls/star/search.asp)

Analyses included data from surveys conducted over slightly less than four years, from November 1, 2021 (8 months prior to the initiation of the pilot program) through August 31, 2025 (8 months after completion of the collection of quality outcome data), to allow at least one post-accreditation survey to have been completed in as many AL communities as possible.

Licensure Analyses. Of the target number of 150 communities, 146 communities were randomized to either the control or accreditation arm; of these, 113 participated in the project by providing data (i.e., 55 of 73 communities in the control arm, 58 of 73 in the accreditation arm). All tables and figures in the narrative portion of the licensure analyses results reflect the 113 communities that ever participated; **Appendix I** includes data from these 113 communities (Tables A1-A9), and from all 146 randomized communities (Tables A10-A18). Because there are differences in how state and county surveys are conducted, data from each type of survey (state or county) were analyzed separately.

Comparisons. Licensure periods were defined as Period 1 (November 2021 through September 2022, the time preceding and during baseline quality outcome data collection [i.e., including the baseline Quarter 1 and the 8 months preceding the baseline]); Period 2 (October 2022 to March 2023, the time before any communities were accredited [i.e., includes Quarters 2 and 3 during which quality outcome data were collected]); and Period 3 (April 2023 to August 2025 [i.e., includes Quarters 4-10 during which quality outcome data were collected and the 8 following months]). Data from multiple surveys in the same community during the same period were averaged to produce a single summarized value for each community per period. Analyses examined change from baseline in each community from Period 1 to Period 3, except for Type A violations that also included Period 2 to increase the sample size for this rare outcome.

- For state data, the overall sample size in participating communities was 361 inspections from all 113 participating communities. After combining inspections from the same community in the same period, 233 observations (community-periods) were entered into the analytic model: 112 from the control arm, 65 from intervention arm communities that never became accredited, and 56 from accredited communities (36 before those communities were accredited and 20 after).
- For county data, the overall sample size in participating communities was 238 inspections from 56 of the 113 participating communities. After combining inspections from the same community in the same period, 91 observations (community-periods) were entered into the analytic model: 47 from the control arm, 20 from intervention arm communities that never became accredited, and 24 from accredited communities (13 before those communities were accredited and 11 after).

Attribution of "accredited condition." Within the accreditation arm, accredited condition (accredited versus not accredited) was evaluated on a quarterly basis. If a community was accredited for at least 45 days during the quarter, it was considered to be accredited as of that quarter and all subsequent quarters given that no accredited communities lost their accreditation. Time-adjusted comparisons were made by arm (i.e., accreditation versus control arm, in an intent-to-treat analysis), and within the accreditation arm, of communities that had become accredited to communities that were not or not yet accredited (i.e., accredited condition versus not accredited condition; referred to as a per protocol analysis). Because no communities were accredited before Period 3, descriptive tables indicate Not Applicable (NA) in relation to "accredited condition" for Periods 1 and 2. Five communities began but did not complete the accreditation process. In all analyses and in all quarters these communities are considered to be in the accreditation arm, not accredited condition.

Determination of adjustment variables. Adjusted models accounted for the number of beds (a measure of community size) which is commonly used to reflect unmeasured sources of variation (e.g., type and variety of staff and services) and because communities that became accredited were larger than those that did not become accredited. The design variables -- region and Star rating -- were not included in the analytic models largely because Star rating was expected to be closely associated with the licensure outcome; also, because stratification reflected the combination of Star rating and region, there was no rationale to include region if not including Star rating. The percent of residents with dementia, which was included in the adjusted models for the quality outcome analyses, could not be included in the analytic models because some communities did not submit the baseline data form that reported that information.

Analytic models. To summarize the number of deficiencies or violations per community-period, means and 95% confidence intervals (CIs) were derived (i.e., the CI indicates that if the sampling process were repeated many times, 95% of the intervals created would capture the true population value). In addition, the total number of deficiencies or violations across all inspections within each community-period were calculated, and rate ratios were calculated. A rate ratio less than 1 indicates better quality (fewer deficiencies or violations) in the accreditation arm or accredited condition compared to the control arm or not accredited condition. Details related to analytic considerations and methods are available on request; herein, two are important to note.

- Because Type A violations were a rare outcome, Period 2 was included in follow-up with Period 3.
- For county data, selection bias was operative in that a community would have had at least one deficiency to be included in the data; for this reason, all types of deficiencies were combined to determine whether similar results were found in aggregate.

For most outcomes, four contrasts are reported: (i) between the accreditation arm versus the control arm in period 3 (the intent-to-treat comparison); (ii) between the accredited condition versus the control arm in period 3 (omitting from the contrast quarters from accreditation arm communities when they were in the not accredited condition); (iii) between the not accredited condition versus the control arm in period 3; and (iv) between the accredited condition versus the not accredited condition in period 3 (the per protocol comparison). Because Type A violations were rare, contrasts (ii) and (iii) could not be evaluated in the county data due to small numbers. In addition to rate ratios, 95% CIs and p-values are reported.

Two methods were used to identify potentially important change in licensure data in the longitudinal model analysis.

- *Statistically significant results*. In all models (and for all four contrasts), a statistical difference in study arms or conditions is identified for a contrast with a p-value < 0.05.
- *Other results*. In an exploratory manner, intent-to-treat and per protocol results are flagged as “notable” based on their estimated effect size irrespective of their statistical significance. Rate ratios greater than 1.25 or less than 0.8 are declared “notable,” and are equivalent in magnitude in that one is the inverse of the other.

In addition, false discovery rate (FDR) analyses were conducted for all comparisons; given the many contrasts being examined, adjustment was necessary for multiple comparisons because some p-values less than 0.05 would be expected by chance alone. The (FDR) methodology[2] with overall alpha=0.20 was used to determine which contrasts would be considered significant.

4. RESULTS

The results section includes information about the accreditation process and licensure compliance. Material later in this report includes methodologic analyses and discussion related to generalizability of the AL communities participating in the project, study power of the quality outcomes presented in the July 2025 report, and relatedly whether accreditation improves or at least maintains resident outcomes based on quality measures.

Accreditation Process. In total, 30 of the 73 communities randomized to the accreditation arm (41%) began the accreditation process; 25 completed it and became accredited (34%). Analysis examined the time involved in the accreditation process. These results are reported in **Exhibit 1** on the following page; they display the time required by ACHC to provide a Summary of Findings after a survey; the time required for a community to return their Plan of Correction; and the time required for ACHC to approve the Plan of Correction. In addition, the average number of deficiencies identified in the ACHC surveys is reported for the 25 communities that completed the accreditation process as well as the 5 that began but did not complete the process. The number of days communities required to return their Plan of Correction was highly correlated with the number of deficiencies on their survey (.70), indicating a strong association between more deficiencies and more time needed by a community to submit a Plan of Correction.

As noted in **Exhibit 1**, five of the 30 surveyed communities did not complete and submit their Plan of Correction and therefore were not accredited. Four of the five communities withdrew completely from the project; one continued submitting data for the overall project after discontinuing efforts related to accreditation. Reasons for withdrawing from the accreditation process were of three types.

- Two jointly operated communities reported that completing the corrections required for accreditation required too much staff time and effort, despite the fact that their number of deficiencies were relatively low. They withdrew completely from the project.
- One community reported it was withdrawing from the accreditation process because the Plan of Correction was not consistent with what was discussed at the ACHC survey exit conference. Although the community withdrew from the accreditation process, it did not drop out of the overall project.
- Two communities did not provide a reason for not completing the accreditation process, and withdrew completely from the project. ACHC reported that one of the communities

[2] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995;57:289-300. *Note*: The FDR methodology was applied only to analyses of participating communities; it was not applied to analyses of randomized communities because those results were expected to be strongly correlated with the results of participating communities, and FDR methodology assumes independent analyses.

appeared uninvolved when their surveyors were on site (e.g., the community staff reported they had not yet looked at the accreditation handbook, and asked surveyors how long they would be there and when they were leaving). The other community was found "deficient" in their first survey, meaning that it would have had to undergo a second survey with deficiencies corrected before proceeding with the Plan of Correction and approval process.

Exhibit 1. Process of Accreditation: Summary of Findings, Number of Deficiencies, Plan of Correction

Metric	N	Mean (SD)	Range	25th Percentile	50th Percentile	75th Percentile
Calendar days between accreditation survey and ACHC provision of Summary of Findings	30	8 (3)	2-15	6	8	10
Number of deficiencies	30	8.9 (7.6)	1-36[b]	4	7	13
Number of deficiencies in communities that did not complete process/become accredited	5	15.2 (12.1)	5-36[b]	9	13	13
Number of deficiencies in communities that did complete process/become accredited	25	7.7 (5.9)	1-19	3	6	11
Calendar days between community receiving Summary of Findings and submitting Plan of Correction[a]	25	17 (12)	0-49	10	15	21
Calendar days between receipt of Plan of Correction and ACHC approval	25	0.88 (1.8)	0-7[c]	0	0	1

[a] A strong correlation (.70) exists between the number of deficiencies and the calendar days between receiving a Summary of Findings and submitting a Plan of Correction, such that that the more deficiencies, the more days.
[b] The AL community with 36 deficiencies, an outlier in this analysis, was a "deficient" community, the only one given that designation in the study; it did not complete the accreditation process.
[c] The AL community that required seven days for ACHC to approve their Plan of Correction, an outlier in this analysis, was required to revise and resubmit it to obtain approval.

Licensure Outcomes. The mean numbers of deficiencies/violations at baseline by arm among participating communities are shown in **Exhibit 2** on the next page. The mean number of total state deficiencies is 2.26 in the control arm and 1.39 in the accreditation arm, largely composed of standard deficiencies (1.68 in the control arm and 1.08 in the accreditation arm). Rates of Type A violations are lowest, being 0.22 in the control arm and 0.15 in the accreditation arm. The reason that not all 113 communities are represented in **Exhibit 2** is that some communities did not have any inspections of the specified type (state or county) in Period 1.

Exhibit 2. Baseline Rates of Deficiencies/Violations in Participating Communities, by Arm

Deficiency/Violation Type	Control Arm		Accreditation Arm	
	N of Communities	Mean (95% CI)	N of Communities	Mean (95% CI)
Total deficiencies (State)	37	2.26 (1.28, 3.25)	42	1.39 (0.65, 2.13)
Total deficiencies (County)	11	1.61 (1.25, 1.96)	11	1.34 (0.97, 1.71)
Standard deficiencies (State)	37	1.68 (0.81, 2.54)	42	1.08 (0.49, 1.66)
Standard deficiencies (County)	11	1.36 (0.87, 1.86)	11	1.11 (0.74, 1.49)
Medication-related standard deficiencies[a]	37	0.84 (0.47, 1.21)	42	0.35 (0.14, 0.56)
Type B violations (State)	37	0.36 (0.12, 0.60)	42	0.16 (0.00, 0.34)
Type B violations (County)	11	0.24 (0.00, 0.66)	11	0.23 (0.00, 0.50)
Type A violations (State)	37	0.22 (0.05, 0.39)	42	0.15 (0.00, 0.29)
Type A violations (County)	11	0.00	11	0.00

[a] Available only for state data.

Appendix I provides data tables for state and county data in total and separately for standard deficiencies, medication-related standard deficiencies (a subset of standard deficiencies), Type B violations, and Type A violations, for all participating communities (N=113; tables A1-A9) and all randomized communities (N=146; tables A10-A18). They display means and 95% CIs for Period 1 (baseline), Period 2 (the time before any communities were accredited), and Period 3 (the time after which any communities were accredited). They also display the multivariable regression comparisons of rate ratios between communities in the accreditation arm to the control arm (intent-to-treat comparison), and (separately) per protocol comparisons between communities in the accreditation arm after they became accredited (accredited condition) and did not yet become accredited (not accredited condition) to the control arm, and finally, communities in the accreditation arm that became accredited to those that did not yet become accredited, the comparison determined most informative for the overall evaluation.

Exhibit 3 (next page) summarizes the intent-to-treat findings and per protocol findings comparing those in the accreditation arm that became accredited to those that did not yet become accredited, indicating whether they favored accreditation. The per protocol results are shown in bold because they are considered to be most informative to the overall evaluation; further, the state data are considered to be more informative than the county data because the latter are typically generated when complaints are made, thus potentially biasing estimates made from the available data. All results in **Exhibit 3** are based on the primary model that does not adjust for community size.

- Outcomes, Statistically Significant, Per Protocol. Comparing AL communities in the accreditation arm that became accredited to those that did not yet become accredited indicated statistically significant results favoring accreditation in relation to State Type B violations but not total deficiencies and violations.
- Outcomes, Otherwise Notable, Per Protocol. Comparing AL communities in the accreditation arm that became accredited to those that did not yet become accredited indicated notable results favoring accreditation in relation to Type B violations, Type A violations, and total deficiencies and violations.

Exhibit 3. Differences in Rate Ratios in Total Deficiencies and Violations and by Type, and Whether It Favored Accreditation (November 2021-August 2025)[a]

	Type of Deficiency				
	Total Deficiencies and Violations	**Standard Deficiencies**	**Medication-Related Standard Deficiencies**	**Type B Violations**	**Type A Violations**
Statistically Significant Results[b]					
Intent to treat[c]					
State (N=113)	NS	NS	Not favorable	Not favorable	NS
County (N=56)	NS	NS		NS	NS
Per protocol[d]					
State (N=113)	**NS**	**NS**	**NS**	**Favorable**	**NS**
County (N=56)	NS	Not favorable		Favorable	NS
Notable Results[e]					
Intent to treat[c]					
State (N=113)	Not favorable	NS	Not favorable	Not favorable	Favorable
County (N=56)	Favorable	NS		Favorable	NS
Per protocol[d]					
State (N=113)	**Favorable**	**NS**	**NS**	**Favorable**	**Favorable**
County (N=56)	NS	Not favorable		Favorable	Favorable

Note: Data are not available for shaded rows (medication-related standard deficiencies) because county Corrective Action Reports do not label standard deficiencies by fundamental rule area.

[a] N = total sample size (maximum number of communities with data); NS (not significant) indicates no significant or notable difference between the AL communities being compared.

[b] Due to multiple comparisons, statistically significant results are based on a false discovery rate analysis, which translated to $p < 0.05$.

[c] Intent-to-treat analysis compares AL communities randomized to the accreditation arm versus control arm.

[d] Per protocol analysis of AL communities assigned to the accreditation arm, comparing communities that became accredited versus not becoming accredited. These comparisons are considered most informative to the evaluation.

[e] Notable results are based on estimated effect size irrespective of statistical significance; they reflect a rate ratio greater than 1.25 or less than 0.8 (see text for explanation).

The same outcomes that were statistically significant in the original analysis remained statistically significant after performing the FDR analysis, and no others became statistically significant. Therefore, multiple comparisons did not inflate the number of statistically significant results.

5. ADDITIONAL METHODOLOGIC ANALYSES

Detailed methodologic analyses examined generalizability of the communities participating in the project, power of the quality outcomes presented in the July 2025 report, and whether accreditation improves or at least maintains resident outcomes based on quality measures.

Participation of AL Communities. A total of 355 eligible communities were invited to participate; of these, 146 initially agreed to participate and were randomized to the control arm or accreditation arm, and 209 did not participate, either due to refusal or because recruitment closed for the stratum before they agreed to participate. Of the 146 communities that agreed to participate, 113 participated and provided data.

Comparison of eligible communities that did and did not participate. Analyses were conducted to determine whether the 113 communities that participated differed from the 209 eligible communities that did not agree to participate, based on characteristics known of all communities. Analyses used Fisher's exact test to compare participation based on region, Star rating, and chain affiliation; a Wilcoxon rank-sum test was used to compare the number of licensed beds. Analyses of chain affiliation excluded one chain that was disproportionately represented in the sample and actively promoted participation.

Results are presented in **Exhibit 4**. They indicate that although participation rates by region ranged from 29% in the Western region to 42% in the Eastern region, this difference was not statistically significant (p=0.13). Star rating, community size, and chain affiliation also were not significantly associated with participation (excepting the one chain that was disproportionately represented in the sample and actively promoted participation; data not shown).

Exhibit 4. Participation of Invited Eligible Assisted Living Communities by Stratification Variables, Chain Affiliation, and Size

Characteristic	Total (N=322)	Did Not Agree to Participate (N=209) N (%) or Mean (SD)	Participated (N=113) N (%) or Mean (SD)	P-value
Region				
Western	125	89 (71%)	36 (29%)	**0.13**
Central	106	67 (63%)	39 (37%)	
Eastern	91	53 (58%)	38 (42%)	
Star Rating				
Low (0–2)	68	42 (62%)	26 (38%)	**0.32**
High (3–4)	254	167 (66%)	87 (34%)	
Chain Affiliation[a]				
No	156	118 (76%)	38 (24%)	**0.30**
Yes	112	91 (81%)	21 (19%)	
Licensed Beds	64.5 (31.8)	63.2 (32.2)	66.9 (31.2)	**0.40**

[a] Excludes the one chain that was disproportionately represented and actively promoted participation.

Comparison of randomized communities that did and did not participate. Analyses were conducted of the 146 randomized communities, to determine whether the 113 that participated differed from the 33 that did not participate, based on characteristics known of all communities. Analyses used Fisher's exact test to compare participation based on region, Star rating, chain affiliation, and treatment arm; a Wilcoxon rank-sum test was used to compare the number of licensed beds.

Results are presented in **Exhibit 5**. Four of the five comparisons indicated no statistically significant difference in participation. The one exception was chain affiliation, likely reflecting that corporate leadership of one chain strongly encouraged that its communities participate. (Of note, the statistical test ignores chain affiliation, and assumes the decision to participate is independent across all communities.)

Exhibit 5. Participation of Randomized Assisted Living Communities by Stratification Variables, Chain Affiliation, Size, and Treatment Arm

Characteristic and Treatment Arm	All Randomized Communities (N=146) N (%) or Mean (SD)	Not Participating (N=33) N (%) or Mean (SD)	Participating (N=113) N (%) or Mean (SD)	P-value
Region				
Western	50	14 (28%)	36 (72%)	0.56
Central	49	10 (20%)	39 (80%)	
Eastern	47	9 (19%)	38 (81%)	
Star Rating				
Low (0-2)	33	7 (21%)	26 (79%)	1.00
High (3-4)	113	26 (23%)	87 (77%)	
Chain Affiliation				
No	58	20 (34%)	38 (66%)	0.008
Yes	88	13 (15%)	75 (85%)	
Licensed Beds	65.9 (32.7)	62.7 (37.6)	66.9 (31.2)	0.64
Treatment Arm				
Control arm	73 (100%)	18 (25%)	55 (75%)	0.69
Accreditation arm	73 (100%)	15 (21%)	58 (79%)	

Comparison of communities that dropped out of the study over time. Throughout the study, communities stopped participating due to concerns at the community level (e.g., a COVID-19 outbreak or a change of ownership or administrator). **Exhibit 6** (next page) displays the number of communities remaining in the project and that stopped submitting data in each quarter by treatment arm. **Appendix II** includes this table and similar tables comparing dropout by region, Star rating, and chain affiliation. The columns referring to "remaining in the project" and that "stopped submitting data" are nested; for example, in **Exhibit 6**, of the communities in the control arm, 55 remained in the project as of Quarter 1, but for one community, that was their last quarter of submitting data, leaving 54 communities remaining in Quarter 2. As shown in all tables, participation rates were lower in the last two quarters of the project, perhaps because those two quarters were an extension to the project based on updated legislation.

Cox proportional hazards models were run to determine whether time to dropout differed based on treatment arm, region, Star rating, chain affiliation, and size (number licensed beds per 10 bed increase). Each variable was examined separately, and all were examined in a multivariable model. **Exhibit 7** (next page) displays the results. It indicates that when examined as separate variables, communities that were chain affiliated were more than twice as likely to drop out of the study (hazard ratio 2.04, p=0.01), but no other variables were significantly related to dropout. In multivariable models when all characteristics were examined simultaneously, region (branch), Star rating, chain affiliation and size (per 10-bed increase) related significantly to dropout, but as in the single variable model, treatment arm did not relate to drop out. That is, dropout risk was significantly higher in the Central region, low Star rating, chain affiliated, and smaller communities.

Exhibit 6. Number of Communities Remaining in the Project, by Quarter and Treatment Arm

Project Quarter	Control Arm (N=55)		Accreditation Arm (N=58)	
	N (%) Remaining in the Project[a]	N (%) that Stopped Submitting Data[a]	N (%) Remaining in the Project[b]	N (%) that Stopped Submitting Data[b]
1	55 (100)	1 (1.8)	58 (100)	6 (10.3)
2	54 (98.2)	5 (9.1)	52 (89.7)	6 (10.3)
3	49 (89.1)	0 (0)	46 (79.3)	1 (1.7)
4	49 (89.1)	2 (3.6)	45 (77.6)	3 (5.2)
5	47 (85.5)	2 (3.6)	42 (72.4)	3 (5.2)
6	45 (81.8)	1 (1.8)	39 (67.2)	1 (1.7)
7	44 (80)	10 (18.2)	38 (65.5)	7 (12.1)
8	34 (61.2)	9 (16.4)	31 (53.4)	9 (15.5)
9	25 (45.5)	4 (7.3)	22 (37.9)	2 (3.4)
10	21 (38.2)	21 (38.2)	20 (34.5)	20 (34.5)

[a]Percent is calculated as per the number initially participating in the control arm, in this case N/55.
[b]Percent is calculated as per the number initially participation in the accreditation arm, in this case N/58.

The finding with respect to chains may indicate that while chain leadership urged their communities to enroll in the study, they could not successfully induce communities to follow through with continuing participation. The multivariable model results likely arise from the fact that communities in the Central region tended to have higher Star ratings (r=0.22, p=0.02) which related to less dropout and influenced the singe variable Central branch hazard ratio, which was closer to the null. After adjusting for region, Star rating, licensed beds, and chain affiliation, the estimated hazard for the Central branch was higher (HR = 2.30, 95% CI 1.27–4.13), the association for higher Star ratings appeared stronger (HR = 0.50, 0.28–0.90), and larger bed size showed a clearer protective pattern (HR per 10 beds = 0.86, 0.78–0.95). Importantly, allocation to treatment arm was not related to dropout in either model, indicating a lack of evidence that dropout affected comparisons between arms.

Exhibit 7. Cox Proportional Hazards Model of Assisted Living Community Dropout

	Hazard Ratio	95% CI	P-value
Single Variable Model			
Central vs western branch	1.47	0.86 – 2.54	0.16
Eastern vs western branch	0.68	0.37 – 1.25	0.21
High vs low Star rating	0.82	0.48 – 1.40	0.48
Accreditation arm vs control arm	1.22	0.77 – 1.94	0.39
Chain affiliation	2.04	1.18 – 3.52	0.01
Licensed beds (per 10-bed increase)	0.94	0.87 – 1.01	0.10
Multivariable Model			
Central vs western branch	2.30	1.27 – 4.13	0.006
Eastern vs western branch	0.76	0.41 – 1.41	0.39
High vs low Star rating	0.50	0.28 – 0.90	0.02
Accreditation arm vs control arm	1.19	0.74 – 1.91	0.47
Chain affiliation	2.25	1.27 – 3.95	0.005
Licensed beds (per 10-bed increase)	0.86	0.78 – 0.95	0.002

Power Analyses of Quality Outcome Measures. The *Evaluation of Quality Outcome Measures* report submitted in July 2025 examined 31 quality indicators in five categories: resident outcomes and care coordination/transitions (health care related); resident outcomes and care coordination/transitions (psychosocial related); person-centered care; medication management; and workforce. The data were provided by the AL community, residents/families, and AL staff.

The power of a study refers to whether the data are of sufficient quality and quantity to detect specified differences in treatment arms in the assessment of outcomes. A total of 113 AL communities participated in the accreditation study, a two-arm, multi-period, cross-sectional cluster randomized trial with 58 communities in the accreditation arm and 55 in the control arm. Numerous quality outcomes were assessed over 2½ years (10 quarters), and some communities dropped out before the end of the study and did not provide data for all quarters. Power calculations take into account the rate of observed dropout, including patterns of missing data for the three types of data forms – those completed by the community, residents/ families, and staff. Also, power is determined separately when comparing outcomes between the accreditation arm and the control arm (intent-to-treat) and between the communities in the accreditation arm that did and did not yet become accredited (per protocol).

In sum, and considering dropout, power calculations indicate the extent to which the study has, for example, an 80% chance of determining a statistically significant difference between the accreditation and control arms, or between the accredited condition and the not accredited condition, regardless of whether accreditation was favorable or not favorable.

- Further, for continuous outcomes, it indicates this chance in relation to the difference between means able to be detected, measured by Cohen's d. Based on the data, Cohen's d values of 0.2, 0.5 and 0.8 were considered as small, medium and large effects, respectively.
- With respect to outcomes summarized as proportions, it is common for odds ratios of 1.25, 1.5 and 2.0 to be considered as small, medium and large effects, respectively. When the intervention is expected to decrease the outcome proportion, inverted values of these odds ratios are considered: 0.8, 0.67 and 0.5 as small, medium and large effects, respectively.

Appendix III provides details of the power analyses and the sample size for each type of measure per quarter. The material below provides a synopsis of key findings.

Statistical power for intent-to-treat analyses comparing the accreditation arm to the control arm. With respect to mean differences between the arms, **Exhibit 8** reports power for small to moderately small effects with a range of Cohen's d from 0.2 to 0.3, displaying power for one of the community-provided outcomes (staff turnover) and for the resident/family and staff questionnaires for continuous outcomes with analyses based on linear models. As a benchmark, power is also shown assuming no dropout of communities -- in which case as expected, power is always higher relative to the observed experience of community dropout. The bolded observed dropout values indicate that power is excellent to detect small to moderately small differences between the accreditation and control arms for resident outcomes and staff satisfaction and burnout from the staff questionnaire. Power was lower but also good for staff turnover, reaching 80% for a Cohen's d of 0.30 (i.e., 30%).

Exhibit 8. Percent Power[a] to Detect a Small to Moderately Small Effect Size (Cohen's d) for a Continuous Measure with Two-sided Type I Error of 0.05 for Comparing Accreditation and Control Arms

Form	Dropout Assumption	Cohen's d (small effects)		
		0.2	0.25	0.3
Community (staff turnover)	No dropout	53.0	72.2	86.4
	Observed dropout	46.3	64.7	**80.2**
Resident/family questionnaire	No dropout	80.1	93.8	98.7
	Observed dropout	69.7	**87.3**	**96.0**
Staff questionnaire	No dropout	78.6	93.0	98.4
	Observed dropout	67.2	**85.3**	**95.0**

[a]Power based on generalized estimating equations with nested exchangeable correlation structure with 0.10 within-quarter and 0.05 between-quarter intraclass correlation coefficient.

Power for dichotomous outcomes considers that some outcomes are uncommon, and so incorporates a broader range of effect sizes: small, medium, and large odds ratios as defined above. Again, as a benchmark for the impact of dropout on power, power is also shown for the case of no dropout. Based on logistic regression analyses, **Exhibit 9** indicates that power exceeded 80% for small effect sizes (OR=0.8) for the resident outcomes of falls with major injury, emergency department visit, and hospital admission; and for medium effect sizes (OR=0.67) for outcomes with fewer events (hospital readmission, advance care planning, discharge due to behavior). Power was inadequate for detecting small or medium effects for the medication management outcomes. For one such outcome – medications administered with any errors – power was 83% for detecting large effects. Similarly, related to workforce outcomes regarding staffing standards being met, power was inadequate for supervisors for all examined effect sizes, but excellent (98%) to detect large odds ratios for all other staff.

Exhibit 9. Power[a] to Detect Small, Medium or Large Odds Ratios (1.25, 1.5 or 2.0, or Inverted, 0.8, 0.67, 0.5) with Two-Sided Type I Error of 0.05 for Comparing Accreditation and Control Arms

Form	Dropout Assumption	Percent Power / Size of Effect		
		Small OR=0.8	Medium OR=0.67	Large OR=0.5
Residents (few)[b]	No dropout	81.7	>99.9	>99.9
	Observed dropout	69.3	**99.0**	**>99.9**
Residents (many)[c]	No dropout	99.0	>99.9	>99.9
	Observed dropout	**97.0**	**>99.9**	**>99.9**
Medication management	No dropout	10.2	22.0	46.5
	Observed dropout	8.6	17.3	35.9
Medication errors	No dropout	22.7	55.0	90.5
	Observed dropout	19.2	46.4	**83.4**
		OR=1.25	**OR=1.5**	**OR=2.0**
Supervisors	No dropout	8.0	15.7	32.4
	Observed dropout	7.2	13.6	27.2
Other staff	No dropout	33.8	77.8	99.3
	Observed dropout	28.6	69.4	**97.9**

[a]Power based on generalized estimating equations with nested exchangeable correlation structure with 0.01 within-quarter and 0.005 between-quarter intraclass correlation coefficient.

[b]Hospital readmission, advance care planning, discharge due to behavior.

[c]Falls with major injury, emergency department visit, hospital admission.

Statistical power for per protocol analyses comparing communities in the accreditation arm that became accredited to those that did not yet become accredited. Power is generally expected to be lower for the per protocol analysis because it included less data (i.e., only data from the accreditation arm). Due to the lower sample sizes, power considerations focus on Cohen's d values representative of moderate effect sizes. In particular, Cohen's d of 0.4, 0.5 and 0.6 were considered as "moderately small," "medium," and "moderately large" effects, respectively. For dichotomous outcomes, the same values of odds ratios for effect sizes are applied as per the intent-to-treat analyses. These analyses do not display power assuming no dropout.

The power to detect a true difference between treatment conditions for outcomes reported by the community, residents/families, and staff, respectively is shown in **Exhibit 10** (based on linear models) and **Exhibit 11** (based on logistic models, see next page). For continuous outcomes, power ranged between 64% to 73% for detecting moderate differences between accredited and not accredited conditions as given by Cohen's d of 0.5 (**Exhibit 10**). Power was approximately 80% or greater to detect a difference between conditions that was moderately large (Cohen's d of 0.6).

Exhibit 10. Percent Power[a] to Detect Small, Medium and Large Effect Sizes (Cohen's d) for a Continuous Measure with Two-Sided Type I Error of 0.05 Comparing Accredited and Not Accredited Conditions (with Dropout)

Form	Cohen's d (moderate effects)		
	0.4	**0.5**	**0.6**
Community (staff turnover)	45.2	63.7	**79.3**
Resident/family questionnaire	52.7	72.2	**86.6**
Staff questionnaire	53.4	73.0	**87.1**

[a]Power based on generalized estimating equations with nested exchangeable correlation structure with 0.10 within-quarter and 0.05 between-quarter intraclass correlation coefficient.

For outcomes that were proportions, power was always inadequate (well below 80%) to detect small effects (**Exhibit 11)**. Only for resident outcomes did power exceed 80%. In particular, for resident outcomes with fewer events per community-quarter (hospital readmission, advance care planning, discharge due to behavior), power exceeded 80% only for large effects, whereas for outcomes with more events per community-quarter (falls with major injury, emergency department visit, hospital admission), power also exceeded 80% to detect medium effects. Due to the small sample sizes or rare outcomes, power was inadequate for hypotheses related to medication management and workforce (whether standards were met) outcomes.

Exhibit 11. Power[a] to Detect Small, Medium or Large Odds Ratios (1.25, 1.5 or 2.0, or Inverted, 0.8, 0.67, 0.5) with Two-Sided Type I Error of 0.05 for Comparing Accredited and Not Accredited Conditions (with Dropout)

Form	Percent Power / Size of Effect		
	Small OR=0.8	**Medium OR=0.67**	**Large OR=0.5**
Residents (few)[b]	20.0	50.7	**89.4**
Residents (many)[c]	60.0	**97.3**	**>99.9**
Medication management	4.5	6.7	11.0
Medication errors	8.7	17.9	37.9
	OR=1.25	**OR=1.5**	**OR=2.0**
Supervisors	4.2	6.0	9.3
Other staff	9.5	21.0	47.8

[a]Power based on generalized estimating equations with nested exchangeable correlation structure with 0.01 within-quarter and 0.005 between-quarter intraclass correlation coefficient.
[b]Hospital readmission, advance care planning, discharge due to behavior.
[c]Falls with major injury, emergency department visit, hospital admission.

Non-inferiority Analyses of Quality Outcome Measures. Whereas the power analyses of the previous section relates to examining whether one arm or condition is superior to another, a non-inferiority analysis[3] asks whether accreditation is no worse than not being accredited, and so uses a directional, one-sided hypothesis at a specified non-inferiority margin defined below.

For analyses of proportions, procedures were applied to the 95% CIs of the intent-to-treat and per protocol results from the primary model using the odds ratio (OR) cut points established earlier and that can be considered as representing minimal clinically important differences: > 0.8 or < 1.25 to determine a "non-inferiority" outcome. Thus, where an OR < 1 means a better outcome, the procedure declares the accreditation arm or accredited condition as being non-inferior to the control arm or not accredited condition (in other words, maintaining or improving outcomes) if the upper confidence bound is less than 1.25. Similarly, for outcomes where an OR > 1 means a better outcome in the accredited condition, the accredited condition is considered non-inferior if the lower confidence bound is greater than 0.8.

For analyses of means, the non-inferiority margin δ was defined as a difference in means (e.g., accredited minus not accredited). When smaller values referred to a better outcome, the procedure declares the accreditation arm or accredited condition as being non-inferior to the control arm or not accredited condition if the upper confidence bound is less than δ. Similarly, for outcomes where larger values were better, the accredited condition is considered non-inferior if the lower confidence bound was greater than -δ. The specified margin δ varied across outcomes and was defined as 10% of the possible range of the continuous scale. For example, for the *Care Transition Preparation Score* that could range from 1 to 5, the non-inferiority margin was specified as δ=(5-1)/10 = 0.4; since higher scores were better, non-inferiority of the accredited condition was declared when the lower CI bound was greater than -0.4.

[3] Mascha EJ, Sessler DI. Equivalence and noninferiority testing in regression models and repeated-measures designs. Anesth Analg. 2011;112(3):678-687.

Appendix IV details the results of the non-inferiority analyses. **Exhibit 12** displays the 15 quality outcomes for which it can be ruled out that accreditation is significantly worse (according to a specified margin representing a minimally important difference) in either intent-to-treat or per protocol analyses (48% of the 31 quality outcomes). Because the study was not principally designed to address the question of maintaining outcomes, these analyses are exploratory and post-hoc. Furthermore, failure to conclude non-inferiority (i.e., not being able to rule out harm) may result from the study being underpowered for non-inferiority hypotheses such that lack of a finding of non-inferiority does not necessarily exclude that possibility.

Exhibit 12. Non-inferiority Analyses Determining Whether Accreditation Arm or Accredited Condition is No Worse than Control Arm or Not Accredited Condition According to a Non-inferiority Margin

Quality Outcome	Intent-to-treat Analysis	Per Protocol Analysis
	Can Harm be Ruled Out?	Can Harm be Ruled Out?
Hospital admission	Indeterminable	**Yes**
Hospital readmission	Indeterminable	**Yes**
Advance care planning discussion	Indeterminable	**Yes**
Discharge due to behavior	Indeterminable	**Yes**
Care transition preparation	**Yes**	**Yes**
Social activity participation	**Yes**	**Yes**
Days resident had a visitor	**Yes**	**Yes**
Resident/family satisfaction	**Yes**	Indeterminable
Person-centered care score	**Yes**	**Yes**
Medication errors	**Yes**	**Yes**
Number of different caregivers	**Yes**	Indeterminable
Care aide turnover	**Yes**	**Yes**
Medication aide turnover	Indeterminable	**Yes**
Administrator turnover	**Yes**	**Yes**
Staff satisfaction	**Yes**	**Yes**

6. DISCUSSION

This project gathered and analyzed data from 113 diverse AL communities across North Carolina to evaluate the effectiveness of accreditation in adult care homes/AL communities. Communities were stratified by region and Star rating and randomly assigned to an accreditation or control arm; those in the accreditation arm were offered the opportunity to become accredited by the Accreditation Commission for Health Care (ACHC). A previous *Evaluation of Quality Outcome Measures* report details findings related to the impact of accreditation on quality indicators including adherence to mandatory staffing levels, a workforce outcome with licensure implications.

This *Evaluation of Pilot Program* report provides details about the accreditation process overseen by ACHC; compliance with licensure requirements; and additional methodologic analyses of the quality outcome data and results that were addressed in the *Evaluation of Quality Outcome Measures* report related to participation rates; it also includes a post-hoc power analysis and non-inferiority analysis. Data related to the accreditation process were provided by ACHC. Data on compliance with licensure requirements were collected from

November 2021 through August 2025 from state and county surveys and describe total deficiencies/violations; standard deficiency citations; medication-management standard deficiency citations (a sub-type of standard deficiencies); Type B violations; and Type A violations.

Accreditation Process and Accreditation. Of the 58 communities randomized and participating in the accreditation arm at baseline, fewer than half (n=25; 43%) became accredited. Given that an initial eligibility criterion was that administrators were asked to agree to be randomized to either arm, and that the cost charged by the accrediting organization was borne by the study, some project partners and members of the Stakeholder Advisory Group anticipated more uptake. During the study, five communities withdrew their participation due to accreditation-related issues: two were concerned that completing the corrections required for accreditation required too much staff time and effort; one was concerned that the Plan of Correction was not consistent with what was discussed at the ACHC survey exit conference.

Importantly, and as noted in the *Evaluation of Quality Outcome Measures* report, communities that became accredited did not differ statistically from other communities in characteristics typically used to describe these communities, other than that they were larger and were less likely to have LPNs on site and on call and RNs on call. Community size was controlled in adjusted analyses, thereby attenuating its potential effect on study outcomes.

Regarding ACHC process efficiency, data indicate timely notification of deficiencies via Summary of Findings and prompt approval of submitted Plans of Correction. Internal ACHC benchmarking requires Summary of Findings to be provided within 10 business days after the last day of survey; in this project, the mean time was eight calendar days (SD = 3). Although ACHC does not have internal benchmarks for Plan of Correction approval timelines -- largely because clinical review teams may be required to review evidence of corrective actions on the part of AL communities -- ACHC averaged less than one calendar day when approving Plans of Correction and accrediting the communities. This short timeline suggests that community's Plans of Corrections were complete and adequately addressed deficiencies, and that ACHC was efficient in reviewing submitted documents and accrediting communities.

Licensure Outcomes. Licensure outcomes, measured by the presence of deficiencies and violations in state and county surveys, were examined for the 146 communities that initially agreed to participate and were randomized, and for the 113 communities that participated and submitted data. Data for both are provided in **Appendix I**. A comparison between these groups revealed no meaningful differences; therefore, the discussion pertains to the 113 communities that were randomized and submitted data. Analyses compared rates of change between communities in the control arm and those in the accreditation arm (intent-to-treat), and between communities in the accreditation arm that became accredited (accredited condition) and those that did not (not accredited condition; per protocol analysis). Because the comparisons considered to be most informative to the overall evaluation compare the communities in the accreditation arm that did and did not become accredited, the discussion focuses on those comparisons.

Across all deficiency and violation types, **Exhibit 13** presents the two that were statistically significant. Specifically, statistically significant results were found for standard deficiencies and Type B violations. Change favorable to accreditation was observed in both state and county data related to Type B violations. Unfavorable change was observed in county data related to standard deficiencies. However, for both state and county data, the differences in the total deficiencies and violations were not significant comparing the two arms in intent-to-treat analyses (**Exhibit 3** and **Appendix I**).

Exhibit 13 additionally presents four categories that showed notable but not statistically significant results: total deficiencies and violations; standard deficiencies; Type B violations; and Type A violations. All favored accreditation except county standard deficiencies.

Exhibit 13. Per Protocol Results Summary Table Differences in Rate Ratios in Total Deficiencies and Violations and by Type, and Whether it Favored Accreditation (November 2021-August 2025)[a,b]

	Type of Deficiency				
	Total Deficiencies and Violations	**Standard Deficiencies**	**Medication-Related Standard Deficiencies**	**Type B Violations**	**Type A Violations**
Statistically Significant Results[c]					
State (N=113)	NS	NS	NS	Favorable	NS
County (N=56)	NS	Not favorable		Favorable	NS
Notable Results[d]					
State (N=113)	Favorable	NS	NS	Favorable	Favorable
County (N=56)	NS	Not favorable		Favorable	Favorable

Note: Data are not available for shaded rows (medication-related standard deficiencies) because county Corrective Action Reports do not label standard deficiencies by fundamental rule area.

[a] Per protocol analysis of AL communities assigned to the accreditation arm, comparing communities that became accredited versus not becoming accredited. These comparisons are considered to be most informative to the overall evaluation.

[b] N = total sample size (maximum number of communities with data); NS (not significant) indicates no significant or notable difference between the AL communities being compared.

[c] Due to multiple comparisons, statistically significant results are based on a false discovery rate analysis, which translated to $p < 0.05$.

[d] Notable results are based on estimated effect size irrespective of statistical significance; they reflect a rate ratio greater than 1.25 or less than 0.8 (see text for explanation).

Considering the findings together, the fact that accredited communities showed favorable change for both Type B and Type A violations -- those having health and safety implications -- provides preliminary evidence that AL accreditation bolsters regulatory goals in important areas. This finding may in part reflect that ACHC accreditation standards were developed in alignment with state regulations, ensuring that accreditation efforts complement those related to regulatory compliance.

Other sectors and professions have increasingly adopted accreditation; examples include laboratory testing facilities,[4] opioid treatment programs,[5] higher education programs,[6] and hospitals,[7] among others. While some evidence exists that accreditation can improve compliance with ethical standards for graduates of accredited educational programs,[8] robust and reliable data on the impact of accreditation on regulatory compliance in healthcare remain limited. The lack of robust data relates in part to study designs, heterogeneity of accreditation and certification interventions, and varied financial and organizational constraints.[9] Until this study, there have been no comparative data for AL accreditation, which is unsurprising given that fewer than 1% of AL communities in the United States are accredited.[10]

The previous *Evaluation of Quality Outcome Measures* report provided data on the impact of accreditation on meeting staffing requirements across staff types, a measure of workforce with implications for licensure compliance. Overall findings were mixed, with no clear indicator of accreditation's impact. This finding is somewhat intuitive, because even if accreditation increases knowledge about and commitment to staffing requirements, ongoing workforce shortages may constrain the ability to rectify insufficient staffing regardless of accreditation status.

Methodologic Analyses: Generalizability, Power, and Non-Inferiority. Methodologic analyses of the accreditation project were conducted to determine the generalizability of results, the study's power to detect change, and whether there was sufficient evidence to at least conclude that accreditation did not cause harm.

Generalizability. Analyses were run to determine if region, Star rating, chain affiliation, and size related to project participation. The analyses compared the 113 participating communities to both the 209 eligible communities that did not agree to participate, and to the 33 communities that agreed to participate but did not provide any data. Excepting the one chain that was disproportionately represented in the sample and encouraged its communities to participate, results indicate that the 113 participating communities are representative of eligible but non-participating communities based on available data.

---

[4] Ehrmeyer SS. Satisfying regulatory and accreditation requirements for quality control. Clin Lab Med. 2013;33(1):27-40.
[5] Wickersham ME, Basey S. Is accreditation sufficient? A case study and argument for transparency when government regulatory authority is delegated. JHHSA Fall 2016. 2016; 245-282.
[6] Even TA, Robinson CR. The impact of CACREP accreditation: A multiway frequency analysis of ethics violations and sanctions. Jour of Counseling & Develop. 2013;91(1):26-34.
[7] Al-Alawy K, Moonesar IA, Mubarak Obaid HA, Al-Abed Bawadi EI, Gaafar R. Hospital accreditation: A review of evidence, regulatory compliance, and healthcare outcome measures. Dubai Med J. 2021;4(3):248-255.
[8] Even TA, Robinson CR. The impact of CACREP accreditation: A multiway frequency analysis of ethics violations and sanctions. Jour of Counseling & Develop. 2013;91(1):26-34.
[9] Brubakk K, Vist GE, Bukholm G, Barach P, Tjomsland O. A systematic review of hospital accreditation: the challenges of measuring complex intervention effects. BMC Health Serv Res. 2015;15:280.
[10] Center for Excellence in Assisted Living at the University of North Carolina at Chapel Hill (CEAL@UNC). ACHC, CARF International, and Joint Commission Present about Assisted Living Accreditation to CEAL@UNC Strategic Advisors. Dec 12, 2024. Available from: https://theceal.org/achc-carf-international-and-joint-commission-present-about-assisted-living-accreditation-to-cealunc-strategic-advisors/

Additional analyses examined whether region, Star rating, chain affiliation, and size related to dropout. Chain affiliation was associated with dropout, explained in part by leadership changes at the corporate level of the above-referenced chain during the final two quarters of the study that resulted in fewer communities submitting data. Additional findings relate to differences based on the combination of region, Star rating, chain affiliation, and size, due likely to correlations between these variables. Most importantly, whether communities were randomized to the accreditation or control arm did not relate to dropout in either model, meaning that based on known characteristics, dropout did not vary meaningfully by study arm.

Overall, the study demonstrated good generalizability and correspondingly, little evidence of selection bias in both participation and dropout between participating and eligible and non-participating communities.

Power. A post-hoc power analysis examined the study's ability to detect change in quality outcome indicators described in the *Evaluation of Quality Outcome Measures* report. In intent-to-treat comparisons between randomized study arms, statistical power was excellent to detect small to moderately small effect sizes (and by implication, larger effects) for resident outcomes and care coordination: psychosocial and healthcare related outcomes; and staff satisfaction and burnout. Power was additionally strong for differences related to workforce turnover.

Per protocol power related to comparisons between communities in the accreditation arm that became accredited and did not yet become accredited was likewise good for resident outcomes and care coordination: psychosocial outcomes; staff satisfaction; staff burnout; and turnover. Overall, analyses suggest the study had good power and was largely successful in detecting small and moderate changes related to accreditation.

Two outcome domains displayed low power: medication management and meeting staffing requirements. These comparisons were influenced by at least three considerations: only 43% of communities participating in the accreditation arm became accredited, which reduced power for per protocol evaluations; substantial community-level dropout; and low event rates for some outcomes (e.g., medication errors and supervisors not meeting staffing requirements). In this sense, the study provided clarity on the quality of outcomes as well as on those outcomes that were associated with accreditation.

Non-inferiority analysis. To further examine the impact of accreditation on quality outcomes, analyses were performed to rule out accreditation being worse (to a tolerable margin) than non-accreditation for each of the 31 quality outcomes in the *Evaluation of Quality Outcome Measures* report. Importantly, this type of analysis cannot conclude that accreditation was significantly worse (the two-sided superiority hypothesis tests of the July 2025 were designed for that), but rather is designed to rule out such a possibility. In either the intent-to-treat or per protocol comparisons, accreditation was ruled out as being worse than not being accredited for 48% of outcomes; for the remaining 52% of outcomes, the possibility of a negative impact was indeterminable. Although post-hoc and exploratory, these results indicate that, for the outcomes where inferiority was ruled out, the act of becoming accredited is not likely to have a negative impact.

Impact of Study Context on the Evaluation of Adult Care Homes Accreditation Pilot Program. Several contextual factors did, or had the potential to, influence the study findings and interpretation, both in relation to quality outcome measures and meeting licensure requirements.

State and county inspections. Throughout data collection, all communities received standard state and county inspections. NC Division of Health Services Regulation inspections are designed to "ensure the provider is operating in compliance with applicable laws and regulations, in a manner that protects the health and safety of their residents."[11] While legally and ethically necessary, these ongoing inspections -- and the potential for deficiencies to be identified and then corrected -- introduced an extraneous variable preventing the evaluation of the effect of accreditation on quality outcomes in isolation. As a result, study findings reflect the potential "value added" of accreditation within an existing regulatory framework. Similarly, the potential benefit from accreditation was likely lessened due to ongoing oversight from the state and county.

Additionally, not every community received the same number of surveys, and not every accredited community had a post-accreditation survey, limiting the available licensure data. State inspection data (Star rating worksheets) included reported deficiencies/violations as well as instances when no deficiencies or violations were found; on the other hand, county Corrective Action Reports were largely complaint-driven, generated when a deficiency or violation were observed. Consequently, county data apply only to cited communities, which may bias county results toward unfavorable findings. This selection bias may help explain the unfavorable differences observed for county standard deficiencies versus the lack of unfavorable results in state data.

Event/incidence rates. Across the study, event/incidence rates for several indicators were low. Examples include those related to medication management errors, which generally remained below 2% throughout the study, and the presence of Type A violations. At the other extreme, rates for certain quality outcomes and licensure indicators, such as meeting supervisory staffing requirements, were generally at or near 100 percent. Rates at either extreme, and lack of variability among participants, reduced the ability to detect statistically significant differences. In addition, high performance (such as a lack of errors and violations) at baseline and throughout the study limited the potential impact of accreditation, as communities reporting high performance had little room for improvement.

Differences at baseline. Baseline differences between study arms -- while controlled for in the analyses to the extent possible -- were notable for some indicators. For example, the unfavorable association between accreditation and meeting staffing requirements for administrators in special care (an indicator of both quality and meeting licensure requirements) may be less explained by accreditation than by being attributable to large initial differences: at

[11] NC Division of Health Service Regulation, Adult Care Licensure Section. https://info.ncdhhs.gov/dhsr/acls/star/search.asp

baseline, 93% of accreditation arm communities met requirements versus 70% of control arm communities. Thus, accreditation arm communities had potential to improve 7% while control arm communities had 30% room for improvement. In the last study quarter, both arms achieved high compliance (100% accreditation, 92% control), with the control arm outpacing improvement due to a notably lower baseline.

Accreditation uptake. As noted earlier, only 25 communities became accredited, limiting the number of communities available for comparison. Further, accreditation was largely achieved more than halfway through the study (quarter 6 or later), meaning that data assessing potential change due to accreditation was limited, lessening the ability to detect statistically significant effects. In addition, the true impact of accreditation on quality outcomes may have potentially been affected by a lack of organizational commitment. Such commitment, typically championed by a strong leader, is necessary to drive quality improvement.[12] In this study, rates of administrator turnover across both the control arm (ranging from 25-47%) and accreditation arm (ranging from 24-51%), alongside capacity issues related to staffing, may have limited communities' ability to fully commit to sustained quality improvement. Anecdotally, a champion in corporate leadership in one chain substantially increased participation by strongly encouraging accreditation and data submission.

COVID-19 and workforce instability. Data collection began in 2022. A 2023 report during the outset of this project, noted that many AL communities continued to face pandemic related recruitment challenges, staff burnout, and turnover.[13] Unsurprisingly, staffing shortages were evident in both study arms, particularly within special care units (with neither arm reporting more than 40% of shifts meeting requirements for care aides except in one instance). An intuitive and empirical relationship exists between staffing levels, job satisfaction and burnout,[14] and quality of care outcomes.[15] It is therefore likely that staffing shortages may have constrained the potential of accreditation to translate to improvements in resident and staff quality outcomes, and related compliance with regulation.

7. CONCLUSIONS AND RECOMMENDATIONS

Of the 31 quality outcomes previously reported, 3 of 4 statistically significant outcomes (75%) favored accreditation, and 16 of 22 not statistically significant but notable outcomes (73%) favored accreditation; in addition, the overall state licensure analyses in the present report favored accredited communities in per protocol analyses, and significantly favored accredited communities in Type B violations.

---

[12] Desveaux L, Mitchell JI, Shaw J, Ivers NM. Understanding the impact of accreditation on quality in healthcare: A grounded theory approach. Int J Qual Health Care. 2017;29(7):941-947.

[13] North Health Workforce Sentinel Network, Findings Brief: LTC Spring 2023 https://tinyurl.com/43xmnn45

[14] Costello H, Walsh S, Cooper C, Livingston G. A systematic review and meta-analysis of the prevalence and associations of stress and burnout among staff in long-term care facilities for people with dementia. Int Psychogeriatr. 31(8):1203-1216.

[15] Zimmerman S, Carder P, Schwartz L, et al. The imperative to reimagine assisted living. J Am Med Dir Assoc. 2022;23(2):225-234.

Considering the licensure compliance findings alongside the quality outcomes findings, the fact that the preponderance of significant and notable outcomes favored accreditation is striking. That so few reached the level of statistical significance may in part relate to the small sample size for the condition of becoming accredited, choice of outcomes to study, low event rates, and the ongoing state/county survey process that limited the amount of potential change. Thus, there may be more benefit from accreditation than was able to be determined through formal hypothesis testing that relies on sufficiently large sample size and context.

Overall, the licensure findings and quality outcome indicators are promising, but evidence is limited that accreditation improves (or worsens) outcomes due in part to the study context. Further, the study was not designed to determine whether accreditation independently maintains outcomes. Therefore, rigorous evaluation should be an integral component if accreditation is incorporated into future legislation.

APPENDICES

Appendix I: Licensure Data Tables

APPENDIX I: LICENSURE DATA TABLES

Licensure Outcomes: Participating Communities (n=113) (Tables A1-A9)[1]

Table A1. Mean (95% CI) Total Number of State Survey Deficiencies and Violations by Study Arm and Condition, Participating Communities

Period[a]	Control Arm		Accreditation Arm		Accreditation Arm: Not Accredited Condition		Accreditation Arm: Accredited Condition	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	37	2.26 (1.28, 3.25)	42	1.39 (0.65, 2.13)	42	1.39 (0.65, 2.13)	NA	NA
2	22	2.07 (0.53, 3.61)	17	1.71 (0.24, 3.17)	17	1.71 (0.24, 3.17)	NA	NA
3	53	2.13 (1.59, 2.67)	56	1.71 (1.22, 2.20)	42	1.84 (1.18, 2.49)	20	1.54 (0.89, 2.19)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	Primary Model			Adjusted Model[b]		
	Rate Ratio	95% CI	p-value	Rate Ratio	95% CI	p-value
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in Period 3	1.39	(0.93, 2.08)	0.10	1.38	(0.93, 2.05)	0.11
Per Protocol Comparisons						
Accredited Condition versus Control Arm in Period 3	1.14	(0.71, 1.84)	0.58	1.09	(0.70, 1.69)	0.72
Not Accredited Condition versus Control Arm in Period 3	1.55	(1.02, 2.34)	0.04	1.57	(1.03, 2.39)	0.03
Accredited Condition versus Not Accredited Condition in Period 3	0.74	(0.51, 1.06)	0.10	0.69	(0.49, 0.98)	0.04

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).
[b] Adjusted model adjusts for community size (see text for explanation).

[1] For state data, the overall sample size in participating communities was 361 inspections from all 113 participating communities. After combining inspections from the same community in the same period, 233 observations (community-periods) were entered into the model: 112 from the control arm, 65 from intervention arm communities that never became accredited, and 56 from accredited communities (36 before those communities were accredited and 20 after they were accredited).
For county data, the overall sample size in participating communities was 238 inspections from 56 of the 113 participating communities. After combining inspections from the same community in the same period, 91 observations (community-periods) were entered into the model: 47 from the control arm, 20 from intervention arm communities that never became accredited, and 24 from accredited communities (13 before those communities were accredited and 11 after they were accredited).

Table A2. Mean (95% CI) Total Number of County Survey Deficiencies and Violations by Study Arm and Condition, Participating Communities

Period[a]	Control Arm		Accreditation Arm		Accreditation Arm: Not Accredited Condition		Accreditation Arm: Accredited Condition	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	11	1.61 (1.25, 1.96)	11	1.34 (0.97, 1.71)	11	1.34 (0.97, 1.71)	NA	NA
2	10	1.45 (1.09, 1.81)	9	1.33 (0.79, 1.88)	9	1.33 (0.79, 1.88)	NA	NA
3	26	1.80 (1.13, 2.47)	20	1.39 (1.18, 1.60)	13	1.60 (1.10, 2.10)	11	1.23 (1.00, 1.46)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	Primary Model			Adjusted Model[b]		
	Rate Ratio	95% CI	p-value	Rate Ratio	95% CI	p-value
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in Period 3	0.79	(0.51, 1.22)	0.29	0.78	(0.51, 1.20)	0.25
Per Protocol Comparisons						
Accredited Condition versus Control Arm in Period 3	0.75	(0.49, 1.15)	0.19	0.78	(0.53, 1.14)	0.20
Not Accredited Condition versus Control Arm in Period 3	0.81	(0.51, 1.28)	0.37	0.78	(0.49, 1.26)	0.31
Accredited Condition versus Not Accredited Condition in Period 3	0.93	(0.75, 1.16)	0.53	0.99	(0.77, 1.28)	0.97

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

Table A3. Mean (95% CI) Number of State Survey Standard Deficiencies by Study Arm and Condition, Participating Communities								
Period[a]	Control Arm		Accreditation Arm		Accreditation Arm: Not Accredited Condition		Accreditation Arm: Accredited Condition	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	37	1.68 (0.81, 2.54)	42	1.08 (0.49, 1.66)	42	1.08 (0.49, 1.66)	NA	NA
2	22	1.20 (0.39, 2.02)	17	1.32 (0.14, 2.50)	17	1.32 (0.14, 2.50)	NA	NA
3	53	1.82 (1.34, 2.30)	56	1.37 (0.96, 1.77)	42	1.39 (0.87, 1.90)	20	1.48 (0.79, 1.96)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	Primary Model			Adjusted Model[b]		
	Rate Ratio	95% CI	p-value	Rate Ratio	95% CI	p-value
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in Period 3	1.22	(0.75, 1.99)	0.42	1.21	(0.75, 1.96)	0.43
Per Protocol Comparisons						
Accredited Condition versus Control Arm in Period 3	1.17	(0.67, 2.04)	0.59	1.10	(0.64, 1.87)	0.73
Not Accredited Condition versus Control Arm in Period 3	1.26	(0.77, 2.05)	0.36	1.28	(0.78, 2.10)	0.33
Accredited Condition versus Not Accredited Condition in Period 3	0.93	(0.65, 1.33)	0.68	0.86	(0.60, 1.22)	0.40

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

Table A4. Mean (95% CI) Number of County Survey Standard Deficiencies by Study Arm and Condition, Participating Communities								
Period[a]	**Control Arm**		**Accreditation Arm**		**Accreditation Arm: Not Accredited Condition**		**Accreditation Arm: Accredited Condition**	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	11	1.36 (0.87, 1.86)	11	1.11 (0.74, 1.49)	11	1.11 (0.74, 1.49)	NA	NA
2	10	0.95 (0.40, 1.50)	9	1.14 (0.64, 1.64)	9	1.14 (0.64, 1.64)	NA	NA
3	26	1.06 (0.51, 1.60)	20	1.03 (0.82, 1.25)	13	0.96 (0.50, 1.41)	11	1.12 (0.91, 1.33)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	**Primary Model**			**Adjusted Model[b]**		
	Rate Ratio	**95% CI**	**p-value**	**Rate Ratio**	**95% CI**	**p-value**
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in Period 3	0.83	(0.53, 1.29)	0.41	0.82	(0.54, 1.27)	0.38
Per Protocol Comparisons						
Accredited Condition versus Control Arm in Period 3	1.03	(0.65, 1.62)	0.90	1.07	(0.69, 1.65)	0.78
Not Accredited Condition versus Control Arm in Period 3	0.74	(0.47, 1.17)	0.20	0.72	(0.45, 1.14)	0.16
Accredited Condition versus Not Accredited Condition in Period 3	1.39	(1.06, 1.83)	0.02	1.48	(1.10, 2.01)	0.01

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

Table A5. Mean (95% CI) Number of State Survey Medication-Related Standard Deficiencies by Study Arm and Condition, Participating Communities

Period[a]	Control Arm		Accreditation Arm		Accreditation Arm: Not Accredited Condition		Accreditation Arm: Accredited Condition	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	37	0.84 (0.47, 1.21)	42	0.35 (0.14, 0.56)	42	0.35 (0.14, 0.56)	NA	NA
2	22	0.55 (0.17, 0.92)	17	0.62 (0.05, 1.19)	17	0.62 (0.05, 1.19)	NA	NA
3	53	0.61 (0.40, 0.83)	53	0.55 (0.33, 0.77)	42	0.58 (0.28, 0.87)	20	0.50 (0.19, 0.81)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	Primary Model			Adjusted Model[b]		
	Rate Ratio	95% CI	p-value	Rate Ratio	95% CI	p-value
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in Period 3	2.02	(1.09, 3.74)	0.03	2.00	(1.08, 3.71)	0.03
Per Protocol Comparisons						
Accredited Condition versus Control Arm in Period 3	2.02	(1.00, 4.07)	0.05[c]	1.93	(0.95, 3.94)	0.07
Not Accredited Condition versus Control Arm in Period 3	2.02	(1.04, 3.90)	0.04	2.04	(1.06, 3.96)	0.03
Accredited Condition versus Not Accredited Condition in Period 3	1.00	(0.56, 1.77)	1.00	0.95	(0.52, 1.72)	0.85

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

[c] p=0.049

Table A6. Mean (95% CI) Number of State Survey Type B Violations by Study Arm and Condition, Participating Communities								
Period[a]	Control Arm		Accreditation Arm		Accreditation Arm: Not Accredited Condition		Accreditation Arm: Accredited Condition	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	37	0.36 (0.12, 0.60)	42	0.16 (0.00, 0.34)	42	0.16 (0.00, 0.34)	NA	NA
2	22	0.32 (0.02, 0.62)	17	0.24 (0.00, 0.58)	17	0.24 (0.00, 0.58)	NA	NA
3	53	0.16 (0.06, 0.27)	56	0.24 (0.11, 0.38)	42	0.31 (0.12, 0.50)	20	0.10 (0.00, 0.20)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	Primary Model			Adjusted Model[b]		
	Rate Ratio	95% CI	p-value	Rate Ratio	95% CI	p-value
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in Period 3	2.97	(1.17, 7.50)	0.02	2.94	(1.16, 7.43)	0.02
Per Protocol Comparisons						
Accredited Condition versus Control Arm in Period 3	1.41	(0.48, 4.19)	0.53	1.34	(0.46, 3.93)	0.59
Not Accredited Condition versus Control Arm in Period 3	4.43	(1.69, 11.60)	0.003	4.48	(1.70, 11.78)	0.002
Accredited Condition versus Not Accredited Condition in Period 3	0.32	(0.14, 0.73)	0.007	0.30	(0.13, 0.67)	0.004

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

Table A7. Mean (95% CI) Number of County Survey Type B Violations by Study Arm and Condition, Participating Communities								
Period[a]	**Control Arm**		**Accreditation Arm**		**Accreditation Arm: Not Accredited Condition**		**Accreditation Arm: Accredited Condition**	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	11	0.24 (0.00, 0.66)	11	0.23 (0.00, 0.50)	11	0.23 (0.00, 0.50)	NA	NA
2	10	0.50 (0.00, 1.01)	9	0.19 (0.00, 0.49)	9	0.19 (0.00, 0.49)	NA	NA
3	26	0.68 (0.35, 1.00)	20	0.28 (0.13, 0.44)	13	0.53 (0.17, 0.89)	11	0.11 (0.00, 0.24)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	**Primary Model**			**Adjusted Model[b]**		
	Rate Ratio	**95% CI**	**p-value**	**Rate Ratio**	**95% CI**	**p-value**
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in Period 3	0.57	(0.20, 1.64)	0.30	0.56	(0.20, 1.62)	0.29
Per Protocol Comparisons						
Accredited Condition versus Control Arm in Period 3	0.29	(0.08, 0.97)	0.04	0.29	(0.09, 0.99)	0.05[c]
Not Accredited Condition versus Control Arm in Period 3	0.83	(0.28, 2.49)	0.74	0.81	(0.27, 2.44)	0.70
Accredited Condition versus Not Accredited Condition in Period 3	0.34	(0.14, 0.85)	0.02	0.36	(0.14, 0.94)	0.04

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

[c] p=0.048

Table A8. Mean (95% CI) Number of State Survey Type A Violations by Study Arm and Condition, Participating Communities								
Period[a]	Control Arm		Accreditation Arm		Accreditation Arm: Not Accredited Condition		Accreditation Arm: Accredited Condition	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	37	0.22 (0.05, 0.39)	42	0.15 (0.00, 0.29)	42	0.15 (0.00, 0.29)	NA	NA
2	22	0.55 (0.00, 1.13)	17	0.15 (0.00, 0.37)	17	0.15 (0.00, 0.37)	NA	NA
3	53	0.15 (0.04, 0.25)	56	0.10 (0.02, 0.17)	42	0.13 (0.01, 0.26)	20	0.06 (0.00, 0.15)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	Primary Model			Adjusted Model[b]		
	Rate Ratio	95% CI	p-value	Rate Ratio	95% CI	p-value
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in periods 2-3[c]	0.47	(0.20, 1.11)	0.08	0.48	(0.21, 1.12)	0.09
Per Protocol Comparisons[d]						
Accredited Condition versus Control Arm in periods 2-3[c]	0.30	(0.07, 1.29)	0.11	0.30	(0.07, 1.28)	0.10
Not Accredited Condition versus Control Arm in periods 2-3[c]	0.60	(0.25, 1.44)	0.25	0.62	(0.26, 1.45)	0.27
Accredited versus Not Accredited Conditions in periods 2-3[c]	0.50	(0.11, 2.25)	0.37	0.49	(0.11, 2.12)	0.34

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

[c] Due to the infrequency of events, periods 2 and 3 were combined.

Table A9. Mean (95% CI) Number of County Survey Type A Violations by Study Arm and Condition, Participating Communities								
Period[a]	**Control Arm**		**Accreditation Arm**		**Accreditation Arm: Not Accredited Condition**		**Accreditation Arm: Accredited Condition**	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	11	0.00	11	0.00	11	0.00	NA	NA
2	10	0.00	9	0.00	9	0.00	NA	NA
3	26	0.06 (0.00, 0.15)	20	0.07 (0.00, 0.18)	13	0.11 (0.00, 0.28)	11	0.00

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	**Primary Model**			**Adjusted Model[b]**		
	Rate Ratio	**95% CI**	**p-value**	**Rate Ratio**	**95% CI**	**p-value**
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in periods 2-3[c]	0.83	(0.21, 3.06)	0.98	0.78	(0.18, 3.20)	0.93
Per Protocol Comparisons[d]						
Accredited versus Not Accredited Conditions in periods 2-3[c]	0.43	(0.00, 2.39)	0.23	1.62	(0.00, 9.86)	0.64

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

[c] Due to the infrequency of events, periods 2 and 3 were combined.

[d] Per protocol comparisons for the accreditation condition versus the control arm, and for the not accredited condition versus the control arm could not be conducted due to the infrequency of events.

Licensure Outcomes: Randomized Communities (n=146) (Tables A10-A18)[2]

Table A10. Mean (95% CI) Total Number of State Survey Deficiencies and Violations by Study Arm and Condition, Randomized Communities								
Period[a]	**Control Arm**		**Accreditation Arm**		**Accreditation Arm: Not Accredited Condition**		**Accreditation Arm: Accredited Condition**	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	50	2.31 (1.53, 3.08)	54	1.45 (0.84, 2.06)	54	1.45 (0.84, 2.06)	NA	NA
2	28	2.30 (0.96, 3.65)	21	2.33 (0.97, 3.70)	21	2.33 (0.97, 3.70)	NA	NA
3	71	2.09 (1.65, 2.53)	71	1.72 (1.28, 2.15)	57	1.81 (1.27, 2.36)	20	1.54 (0.89, 2.19)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	**Primary Model**			**Adjusted Model**[b]		
	Rate Ratio	**95% CI**	**p-value**	**Rate Ratio**	**95% CI**	**p-value**
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in Period 3	1.32	(0.93, 1.87)	0.13	1.28	(0.91, 1.80)	0.16
Per Protocol Comparisons						
Accredited Condition versus Control Arm in Period 3	1.11	(0.73, 1.68)	0.63	1.00	(0.68, 1.48)	0.99
Not Accredited Condition versus Control Arm in Period 3	1.44	(0.98, 2.12)	0.06	1.46	(1.00, 2.14)	0.05
Accredited Condition versus Not Accredited Condition in Period 3	0.77	(0.52, 1.12)	0.17	0.69	(0.47, 1.00)	0.05

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).
[b] Adjusted model adjusts for community size (see text for explanation).

[2] For state data, the sample size in the randomized sample was 493 inspections from all 146 randomized communities. After combining inspections from the same community in the same period, 301 observations (community-periods) were entered into the model: 149 from the control arm, 96 from intervention arm communities that never became accredited, and 56 from accredited communities (36 before those communities were accredited and 20 after they were accredited).
The overall sample size for county data in the randomized sample was 304 inspections from 73 of the 146 randomized communities. After combining inspections from the same community in the same period, 119 observations (community-periods) were entered into the model: 63 from the control arm, 32 from intervention arm communities that never became accredited, and 24 from accredited communities (13 before those communities were accredited and 11 after they were accredited).

Table A11. Mean (95% CI) Total Number of County Survey Deficiencies and Violations by Study Arm and Condition, Randomized Communities								
Period[a]	**Control Arm**		**Accreditation Arm**		**Accreditation Arm: Not Accredited Condition**		**Accreditation Arm: Accredited Condition**	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	16	1.52 (1.26, 1.78)	14	1.34 (1.03, 1.65)	14	1.34 (1.03, 1.65)	NA	NA
2	14	1.56 (1.20, 1.92)	12	1.33 (0.61, 1.43)	12	1.33 (0.61, 1.43)	NA	NA
3	33	1.72 (1.19, 2.24)	26	1.39 (1.20, 1.59)	19	1.54 (1.18, 1.91)	11	1.23 (1.00, 1.46)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	**Primary Model**			**Adjusted Model**[b]		
	Rate Ratio	**95% CI**	**p-value**	**Rate Ratio**	**95% CI**	**p-value**
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in Period 3	0.82	(0.56, 1.22)	0.33	0.81	(0.54, 1.19)	0.28
Per Protocol Comparisons						
Accredited Condition versus Control Arm in Period 3	0.77	(0.52, 1.13)	0.18	0.78	(0.54, 1.12)	0.18
Not Accredited Condition versus Control Arm in Period 3	0.86	(0.57, 1.30)	0.46	0.82	(0.53, 1.27)	0.37
Accredited Condition versus Not Accredited Condition in Period 3	0.90	(0.72, 1.11)	0.32	0.95	(0.74, 1.21)	0.67

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

Table A12. Mean (95% CI) Number of State Survey Standard Deficiencies by Study Arm and Condition, Randomized Communities								
Period[a]	Control Arm		Accreditation Arm		Accreditation Arm: Not Accredited Condition		Accreditation Arm: Accredited Condition	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	50	1.69 (0.99, 2.39)	54	1.19 (0.68, 1.69)	54	1.19 (0.68, 1.69)	NA	NA
2	28	1.38 (0.58, 2.17)	21	1.88 (0.77, 2.99)	21	1.88 (0.77, 2.99)	NA	NA
3	71	1.75 (1.36, 2.13)	71	1.35 (1.01, 1.68)	57	1.36 (0.96, 1.76)	20	1.38 (0.79, 1.96)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	Primary Model			Adjusted Model[b]		
	Rate Ratio	95% CI	p-value	Rate Ratio	95% CI	p-value
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in Period 3	1.05	(0.69, 1.60)	0.83	1.02	(0.67, 1.55)	0.92
Per Protocol Comparisons						
Accredited Condition versus Control Arm in Period 3	1.05	(0.65, 1.72)	0.83	0.96	(0.61, 1.54)	0.88
Not Accredited Condition versus Control Arm in Period 3	1.05	(0.67, 1.63)	0.85	1.05	(0.68, 1.64)	0.81
Accredited Condition versus Not Accredited Condition in Period 3	1.01	(0.70, 1.45)	0.97	0.92	(0.64, 1.30)	0.62

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

Table A13. Mean (95% CI) Number of County Survey Standard Deficiencies by Study Arm and Condition, Randomized Communities								
Period[a]	Control Arm		Accreditation Arm		Accreditation Arm: Not Accredited Condition		Accreditation Arm: Accredited Condition	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	16	1.22 (0.85, 1.59)	14	1.09 (0.72, 1.45)	14	1.09 (0.72, 1.45)	NA	NA
2	14	1.20 (0.70, 1.70)	12	1.02 (0.61, 1.43)	12	1.02 (0.61, 1.43)	NA	NA
3	33	1.01 (0.57, 1.45)	26	1.08 (0.88, 1.28)	19	1.05 (0.71, 1.39)	11	1.12 (0.91, 1.33)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	Primary Model			Adjusted Model[b]		
	Rate Ratio	95% CI	p-value	Rate Ratio	95% CI	p-value
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in Period 3	0.93	(0.60, 1.43)	0.74	0.92	(0.60, 1.40)	0.69
Per Protocol Comparisons						
Accredited Condition versus Control Arm in Period 3	1.04	(0.68, 1.60)	0.85	1.06	(0.71, 1.60)	0.77
Not Accredited Condition versus Control Arm in Period 3	0.87	(0.55, 1.40)	0.57	0.85	(0.53, 1.36)	0.49
Accredited Condition versus Not Accredited Condition in Period 3	1.19	(0.89, 1.60)	0.24	1.26	(0.90, 1.76)	0.18

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

Table A14. Mean (95% CI) Number of State Survey Medication-Related Standard Deficiencies by Study Arm and Condition, Randomized Communities								
Period[a]	**Control Arm**		**Accreditation Arm**		**Accreditation Arm: Not Accredited Condition**		**Accreditation Arm: Accredited Condition**	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	50	0.87 (0.54, 1.21)	54	0.50 (0.27, 0.74)	54	0.50 (0.27, 0.74)	NA	NA
2	28	0.54 (0.23, 0.84)	21	0.79 (0.29, 1.28)	21	0.79 (0.29, 1.28)	NA	NA
3	71	0.64 (0.45, 0.82)	71	0.51 (0.33, 0.70)	57	0.52 (0.30, 0.75)	20	0.50 (0.19, 0.81)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	**Primary Model**			**Adjusted Model[b]**		
	Rate Ratio	**95% CI**	**p-value**	**Rate Ratio**	**95% CI**	**p-value**
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in Period 3	1.30	(0.78, 2.15)	0.31	1.28	(0.77, 2.12)	0.34
Per Protocol Comparisons						
Accredited Condition versus Control Arm in Period 3	1.45	(0.80, 2.62)	0.22	1.37	(0.75, 2.51)	0.30
Not Accredited Condition versus Control Arm in Period 3	1.22	(0.70, 2.13)	0.47	1.23	(0.71, 2.14)	0.46
Accredited Condition versus Not Accredited Condition in Period 3	1.18	(0.69, 2.03)	0.54	1.11	(0.64, 1.93)	0.70

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

Table A15. Mean (95% CI) Number of State Survey Type B Violations by Study Arm and Condition, Randomized Communities								
Period[a]	**Control Arm**		**Accreditation Arm**		**Accreditation Arm: Not Accredited Condition**		**Accreditation Arm: Accredited Condition**	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	50	0.37 (0.16, 0.57)	54	0.15 (0.01, 0.28)	54	0.15 (0.01, 0.28)	NA	NA
2	28	0.46 (0.13, 0.80)	21	0.33 (0.00, 0.67)	21	0.33 (0.00, 0.67)	NA	NA
3	71	0.19 (0.09, 0.28)	71	0.24 (0.12, 0.35)	57	0.29 (0.13, 0.44)	20	0.10 (0.00, 0.20)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	**Primary Model**			**Adjusted Model[b]**		
	Rate Ratio	**95% CI**	**p-value**	**Rate Ratio**	**95% CI**	**p-value**
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in Period 3	3.15	(1.42, 6.98)	0.005	3.02	(1.37, 6.68)	0.006
Per Protocol Comparisons						
Accredited Condition versus Control Arm in Period 3	1.57	(0.58, 4.27)	0.38	1.38	(0.52, 3.65)	0.52
Not Accredited Condition versus Control Arm in Period 3	4.58	(2.01, 10.43)	<0.001	4.62	(2.03, 10.52)	<0.001
Accredited Condition versus Not Accredited Condition in Period 3	0.34	(0.15, 0.79)	0.01	0.30	(0.13, 0.66)	0.003

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

Table A16. Mean (95% CI) Number of County Survey Type B Violations by Study Arm and Condition, Randomized Communities								
Period[a]	**Control Arm**		**Accreditation Arm**		**Accreditation Arm: Not Accredited Condition**		**Accreditation Arm: Accredited Condition**	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	16	0.30 (0.00, 0.59)	14	0.25 (0.00, 0.50)	14	0.25 (0.00, 0.50)	NA	NA
2	14	0.36 (0.00, 0.72)	12	0.31 (0.00, 0.71)	12	0.31 (0.00, 0.71)	NA	NA
3	33	0.62 (0.36, 0.89)	26	0.25 (0.12, 0.37)	19	0.40 (0.15, 0.66)	11	0.11 (0.00, 0.24)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	**Primary Model**			**Adjusted Model[b]**		
	Rate Ratio	**95% CI**	**p-value**	**Rate Ratio**	**95% CI**	**p-value**
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in Period 3	0.46	(0.20, 1.06)	0.07	0.44	(0.19, 1.05)	0.06
Per Protocol Comparisons						
Accredited Condition versus Control Arm in Period 3	0.25	(0.08, 0.77)	0.02	0.25	(0.08, 0.77)	0.02
Not Accredited Condition versus Control Arm in Period 3	0.63	(0.27, 1.48)	0.29	0.60	(0.25, 1.46)	0.26
Accredited Condition versus Not Accredited Condition in Period 3	0.41	(0.16, 1.03)	0.06	0.42	(0.16, 1.10)	0.08

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

Table A17. Mean (95% CI) Number of State Survey Type A Violations by Study Arm and Condition, Randomized Communities								
Period[a]	Control Arm		Accreditation Arm		Accreditation Arm: Not Accredited Condition		Accreditation Arm: Accredited Condition	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	50	0.25 (0.10, 0.40)	54	0.12 (0.00, 0.23)	54	0.12 (0.00, 0.23)	NA	NA
2	28	0.46 (0.00, 0.92)	21	0.12 (0.00, 0.29)	21	0.12 (0.00, 0.29)	NA	NA
3	71	0.19 (0.09, 0.28)	71	0.13 (0.04, 0.22)	57	0.17 (0.04, 0.29)	20	0.06 (0.00, 0.15)

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	Primary Model			Adjusted Model[b]		
	Rate Ratio	95% CI	p-value	Rate Ratio	95% CI	p-value
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in periods 2-3[c]	0.56	(0.27, 1.18)	0.13	0.58	(0.28, 1.22)	0.15
Per Protocol Comparisons[d]						
Accredited Condition versus Control Arm in periods 2-3[c]	0.30	(0.08, 1.20)	0.09	0.28	(0.07, 1.13)	0.07
Not Accredited Condition versus Control Arm in periods 2-3[c]	0.79	(0.38, 1.64)	0.52	0.87	(0.41, 1.83)	0.71
Accredited versus Not Accredited Conditions in periods 2-3[c]	0.41	(0.11, 1.60)	0.20	0.32	(0.07, 1.44)	0.14

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

[c] Due to the infrequency of events, periods 2 and 3 were combined.

Table A18. Mean (95% CI) Number of County Survey Type A Violations by Study Arm and Condition, Randomized Communities								
Period[a]	**Control Arm**		**Accreditation Arm**		**Accreditation Arm: Not Accredited Condition**		**Accreditation Arm: Accredited Condition**	
	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)	Number of Communities	Mean (95% CI)
1	16	0.00	14	0.00	14	0.00	NA	NA
2	14	0.00	12	0.00	12	0.00	NA	NA
3	33	0.08 (0.00, 0.17)	26	0.07 (0.00, 0.15)	19	0.09 (0.00, 0.20)	11	0.00

Poisson Multivariable Regression Results (Reference Categories are Period 1, Control Arm)

Comparisons from Multivariable Regression	**Primary Model**			**Adjusted Model[b]**		
	Rate Ratio	**95% CI**	**p-value**	**Rate Ratio**	**95% CI**	**p-value**
Intent-to-treat Comparisons						
Accreditation Arm versus Control Arm in periods 2-3[c]	0.89	(0.25, 2.94)	1.00	0.84	(0.23, 2.97)	0.99
Per Protocol Comparisons[d]						
Accredited versus Not Accredited Conditions in periods 2-3[c]	0.45	(0.00, 2.40)	0.24	0.61	(0.00, 4.36)	0.35

[a] Period 1 = November 2021 through September 2022, the time preceding and during baseline quality outcome data collection (i.e., includes the baseline [Quarter 1] and the 8 months preceding the baseline); Period 2 = October 2022 to March 2023, the time before any communities were accredited (i.e., includes Quarters 2 and 3 during which quality outcome data were being collected); Period 3 = April 2023 to August 2025 (i.e., includes Quarters 4-10 during which quality outcome data were being collected and the 8 following months).

[b] Adjusted model adjusts for community size (see text for explanation).

[c] Due to the infrequency of events, periods 2 and 3 were combined.

[d] Per protocol comparisons for the accreditation condition versus the control arm, and for the not accredited condition versus the control arm could not be conducted due to the infrequency of events.

APPENDIX II: ASSISTED LIVING COMMUNITY DROPOUT TABLES

Table B1. Number of Communities Remaining in the Project, by Quarter and Treatment Arm

Project Quarter	Control Arm (N=55)		Accreditation Arm (N=58)	
	N (%) Remaining in the Project[a]	N (%) that Stopped Submitting Data[a]	N (%) Remaining in the Project[b]	N (%) that Stopped Submitting Data[b]
1	55 (100)	1 (1.8)	58 (100)	6 (10.3)
2	54 (98.2)	5 (9.1)	52 (89.7)	6 (10.3)
3	49 (89.1)	0 (0)	46 (79.3)	1 (1.7)
4	49 (89.1)	2 (3.6)	45 (77.6)	3 (5.2)
5	47 (85.5)	2 (3.6)	42 (72.4)	3 (5.2)
6	45 (81.8)	1 (1.8)	39 (67.2)	1 (1.7)
7	44 (80)	10 (18.2)	38 (65.5)	7 (12.1)
8	34 (61.2)	9 (16.4)	31 (53.4)	9 (15.5)
9	25 (45.5)	4 (7.3)	22 (37.9)	2 (3.4)
10	21 (38.2)	21 (38.2)	20 (34.5)	20 (34.5)

[a]Percent is calculated as per the number initially participating in the control arm, in this case N/55.
[b]Percent is calculated as per the number initially participation in the accreditation arm, in this case N/58.

Table B2. Number of Communities Remaining in the Project, by Quarter and Region

Project Quarter	Western Branch (N=36)		Central Branch (N=39)		Eastern Branch (N=38)	
	N (%) Remaining in the Project[a]	N (%) that Stopped Submitting Data[a]	N (%) Remaining in the Project[b]	N (%) that Stopped Submitting Data[b]	N (%) Remaining in the Project[c]	N (%) that Stopped Submitting Data[c]
1	36 (100)	0 (0)	39 (100)	6 (15.4)	38 (100)	1 (2.6)
2	36 (100)	5 (13.9)	33 (84.6)	3 (7.7)	37 (94.9)	3 (7.9)
3	31 (86.1)	0 (0)	30 (76.9)	0 (0)	34 (89.5)	1 (2.6)
4	31 (86.1)	1 (2.8)	30 (76.9)	4 (10.3)	33 (86.8)	0 (0)
5	30 (83.3)	0 (0)	26 (66.7)	3 (7.7)	33 (86.8)	2 (5.3)
6	30 (83.3)	2 (5.6)	23 (59)	0 (0)	31 (81.6)	0 (0)
7	28 (77.8)	8 (22.2)	23 (59)	6 (15.4)	31 (81.6)	3 (7.9)
8	20 (55.6)	6 (16.7)	17 (43.6)	4 (10.3)	28 (73.7)	8 (21.1)
9	14 (38.9)	1 (2.8)	13 (33.3)	4 (10.3)	20 (52.6)	1 (2.6)
10	13 (36.1)	13 (36.1)	9 (23.1)	9 (23.1)	19 (50)	19 (50)

[a]Percent is calculated as per the number initially participating in the Western branch, in this case N/36.
[b]Percent is calculated as per the number initially participating in the Central branch, in this case N/39.
[c]Percent is calculated as per the number initially participating in the Eastern branch, in this case N/38.

Table B3. Number of Communities Remaining in the Project, by Quarter and Star Rating

Project Quarter	Low Star Rating (N=26)		High Star Rating(N=87)	
	N (%) Remaining in the Project[a]	N (%) that Stopped Submitting Data[a]	N (%) Remaining in the Project[b]	N (%) that Stopped Submitting Data[b]
1	26 (100)	1 (3.8)	87 (100)	6 (6.9)
2	25 (96.2)	4 (15.4)	81 (93.1)	7 (8)
3	21 (80.8)	0 (0)	74 (85.1)	1 (1.1)
4	21 (80.8)	1 (3.8)	73 (83.9)	4 (4.6)
5	20 (76.9)	1 (3.8)	69 (79.3)	4 (4.6)
6	19 (73.1)	0 (0)	65 (74.7)	2 (2.3)
7	19 (73.1)	5 (19.2)	63 (72.4)	12 (13.8)
8	14 (53.8)	6 (23.1)	51 (58.6)	12 (13.8)
9	8 (30.8)	0 (0)	39 (44.8)	6 (6.9)
10	8 (30.8)	8 (30.8)	33 (37.9)	33 (37.9)

[a]Percent is calculated as per the number initially participating of those with low Star rating, in this case N/26.
[b]Percent is calculated as per the number initially participating of those with high Star rating, in this case N/87.

Table B4. Number of Communities Remaining in the Project, by Quarter and Chain Affiliation

Project Quarter	Chain Affiliated (N=75)		Not Chain Affiliated (N=38)	
	N (%) Remaining in the Project[a]	N (%) that Stopped Submitting Data[a]	N (%) Remaining in the Project[b]	N (%) that Stopped Submitting Data[b]
1	75 (100)	7 (9.3)	38 (100)	0 (0)
2	68 (90.1)	7 (9.3)	38 (100)	4 (10.5)
3	61 (81.3)	0 (0)	34 (89.5)	1 (2.6)
4	61 (81.3)	3 (4)	33 (86.6)	2 (5.3)
5	58 (77.3)	3 (4)	31 (81.6)	2 (5.3)
6	55 (73.3)	1 (1.3)	29 (76.3)	1 (2.6)
7	54 (72)	16 (21.3)	28 (73.7)	1 (2.6)
8	38 (50.7)	12 (16)	27 (71.1)	6 (15.8)
9	26 (34.7)	6 (8)	21 (55.3)	0 (0)
10	20 (26.7)	20 (26.7)	21 (55.3)	21 (55.3)

[a]Percent is calculated as per the number initially participating of those that were chain affiliated, in this case N/75.
[b]Percent is calculated as per the number initially participating of those that were not chain affiliated, in this case N/38.

APPENDIX III: POWER ANALYSES OF QUALITY OUTCOME MEASURES

Background and Methods

The accreditation study was a two-arm, multi-period, cross-sectional cluster randomized trial with 113 participating assisted living (AL) communities, 58 in the accreditation and 55 in the control arm. It examined numerous measures from a total of sixteen topics from five domains (resident outcomes, care coordination and transitions, medication management, workforce, person-centered care). Measures were collected at the community level (e.g., from administrative chart/record review), resident level (from resident/family questionnaires), and staff level. The purpose of this appendix is to provide statistical power for a post-study evaluation to address the question of whether the accreditation study generated data of sufficient quality and quantity to assess whether: (i) in modified intent-to-treat analyses, differences in outcomes between intervention and control arms could be adequately assessed; and (ii) in per protocol analyses, differences in the time-varying accredited versus not accredited conditions could be adequately assessed.

Aim 1. Compare outcomes of communities between the accreditation and the control arms

Aim 2. Among those communities in the accreditation arm, compare outcomes pre- and post-accreditation.

For each aim, power is provided separately for representative outcome type (instead of each outcome). Most outcomes involve averages or proportions over residents or staff. While the number of residents and staff providing data varied across quarters and communities, they are set to constants (approximately equal to observed averages) in power calculations. Tables that are referenced below refer to those in the July 2025 Final Report Appendix.

- *Resident Outcomes and Care Coordination/Transitions* collected quarterly over 10 occasions: these *Health Care Related* outcomes in Tables A1-A6 (Form B) are administrator-reported proportions (numerator/denominator) for each AL-quarter analyzed with longitudinal logistic regression.
    - *Hospital readmission, Advance Care Planning, Discharge due to behaviors* have 5 residents ("few") per AL-quarter representing select groups of residents.
    - *Falls with major injury, ED visits,* and *Hospital Admissions* have 35 residents ("many") per AL-quarter representing essentially all residents.
- *Resident Outcomes and Care Coordination/Transitions* collected twice yearly (for 5 quarters): These are continuous scaled *Psychosocial* resident/family questionnaire data (Tables A8-A11; Form E/F) that were collected for 8 residents per AL-quarter (even though the study design allowed for collection from up to 15 residents). These power results also apply to twice-yearly outcomes "Care transition preparation" (a *Health Care Related* outcome) and "Person-centered care score" (Table B1).

- *Medication Management* collected quarterly (up to 10 occasions): these are administrator-reported proportions (numerator/denominator) for residents analyzed with longitudinal logistic regression models. Two cases are considered:
    - Outcomes where "Number of Residents" is the denominator set equal to 5 residents (Tables C1, C2 and C4);
    - The outcome *Medications Administered with Any Errors* has denominator "Number of Medication Administrations" set equal to 35 administrations (Table C3).
- *Workforce* outcomes are collected bi-annually and include continuous variables (*Number of Different Caregivers* and *turnover,* Forms D1 and D3, respectively) and proportions (*Staffing requirements met*, Form D2):
    - *turnover* is based on a single administrator-reported value per quarter (Tables D9-D13);
    - Outcomes for the percent of *Staff* meeting requirements by staff-types (Tables D2-D8) are based on (approximately) two shifts reported per AL-quarter analyzed with logistic regression;
    - *Number of Different Caregivers* is based on 30+ residents per AL and quarter (Table D1). Power is not reported for the latter but expected to exceed power for continuous E/F outcomes.
- *Staff satisfaction and burnout* (Form G) are continuous outcomes collected from an average of 7 staff per community, twice a year (for a total of five quarters); a constant of 7 staff is used in power calculations.

Calculation of Statistical Power

The SAS macro CRTFASTGEEPWR was used to calculate power based on generalized estimating equations analysis of longitudinal cluster randomized trials.[3] The macro performs non-simulation-based fast power calculations and accounts for general patterns of missing outcome data. The macro was applied for two-sided hypothesis tests with Type I error at 5%, with consideration for the frequency of data collection of the measure. The calculations indicate the extent to which the study has, say, an 80% chance of determining a statistically significant difference between the accreditation and control arms (regardless whether accreditation was favorable or not favorable) assuming specified differences. Otherwise, if, in fact, there was no difference between arms, our statistical analysis procedures would incorrectly conclude there as a difference between arms only 5% or less or the time, i.e., a "Type I error".

[3] Zhang Y, Preisser JS, Li F, Turner EL, Rathouz PJ. %CRTFASTGEEPWR: A SAS macro for power of generalized estimating equations analysis of multi-period cluster randomized trials with application to stepped wedge designs. *J Statist Soft. Code Snippets*, 2024; 108(1), 1–27. https://doi.org/10.18637/jss.v108.c01

Missing Data and Community Dropout in the Accreditation Trial

The power calculations take into account the pattern of community-level dropout **observed in the completed accreditation study.** Dropout causes information loss and thus decreased power. While there are instances of communities not providing data at an earlier visit, only to provide data at one or more subsequent visits, the major reason for missing data in the accreditation study is dropout. For convenience, power calculations assume that missing data is due solely to dropout. Further, we specify sample sizes for each study arm and quarter for outcomes grouped by "type" rather than for each of the 31 outcomes reported in Appendix II of the July 2025 Final Report. In particular, sample sizes for Forms E/F (Resident/Family questionnaires) and G (Staff Questionnaire) were less than those for Forms B and C (Administrative questionnaires).

Power is provided in separate tables for continuous and dichotomous outcomes. For the statistical analysis procedures of the trial based on data aggregation at the community-quarter level, these correspond to the analysis of means and proportions, respectively. Power is presented for various effect sizes expressed in terms of Cohen's d for the standardized difference in means whereas odds ratios were used as effects based on the analysis of proportions. In particular, the minimally clinically important difference between means is indicated by Cohen's d, which is the ratio of the expected mean difference in outcomes between treatment arms (or conditions) to the standard deviation of the outcome. It is common for Cohen's d values of 0.2, 0.5 and 0.8 to be considered as small, medium and large effects, respectively. With respect to outcomes summarized as proportions, we consider odds ratios of 1.25, 1.5 and 2.0 as small, medium and large effects, respectively; when the intervention is expected to decrease the outcome proportion, inverted values of the aforementioned odd ratios are considered: 0.8, 0.67 and 0.5 as small, medium and large effects, respectively.

Statistical Power for Aim 1. With respect to mean differences between treatment arms, we report power for small to moderately small effects with a range of Cohen's d from 0.2 to 0.3. As some dichotomous outcomes are uncommon, we consider power for a broad range of effect sizes in terms of small, medium and large odds ratios as defined above. As a benchmark for the impact of dropout on power, we also present power for the case of no dropout.

Statistical Power for Aim 2. Power is generally expected to be lower for Aim 2 than Aim 1, as the Aim 2 analysis only utilize data from the intervention arm. The number and pattern of AL-quarters under the not accredited versus accredited conditions are based on the actual trial numbers; the case of no dropout is not considered. Due to the lower sample sizes for Aim 2 analyses, power considerations focus on Cohen's d values representative of moderate effect sizes. In particular, Cohen's d of 0.4, 0.5 and 0.6 were considered as "moderately small", "medium", and "moderately large" effects, respectively. For dichotomous outcomes, we consider the same values of odds ratios for effect sizes as in Aim 1.

Results: Aim 1

Table C1 shows the sample sizes in terms of the number of communities having data at each quarter used in the power analysis for **Aim 1**. Administrator-reported outcomes (Form B *Resident Outcomes and Care Coordination/Transitions: Health Care Related* outcomes, and Form C *Medication Management* outcomes) were reported quarterly across 10 quarters. Other outcomes were reported bi-annually in the five odd-numbered quarters (i.e. Q1, Q3, Q5, Q7, and Q9). The number of communities per quarter in the treatment arms decreases over time, reflecting the pattern of dropout observed in the trial. Sample sizes are smallest for Form E/F and G outcomes.

Table C1. Total Number of Participating Communities per Quarter for Aim 1 used in power calculations

Form	Arm	Frequency of data collection	Number of AL Communities									
			Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
B[1] C[2]	Control	Quarterly	55	54	49	49	47	45	44	34	25	21
	Accreditation	Quarterly	58	52	46	45	42	39	38	31	22	20
	Total	Quarterly	113	106	95	94	89	84	82	65	47	41
D[3]	Control	Twice a year	49		43		43		42		36	
	Accreditation	Twice a year	53		43		42		37		35	
	Total	Twice a year	102		86		85		79		71	
E/F[4]	Control	Twice a year	34		27		21		20		13	
	Accreditation	Twice a year	33		31		17		16		11	
	Total	Twice a year	67		58		36		38		24	
G[5]	Control	Twice a year	30		24		22		21		14	
	Accreditation	Twice a year	38		28		17		17		14	
	Total	Twice a year	68		52		39		38		28	

[1]B, [2]C: for these administrative data, the number of communities per quarter are the number of clusters providing any data in the trial; in actuality, the number providing Form B (or C) data for any particular outcome may be slightly less.
[3]Form D number of clusters are pulled as a composite from Tables D1 and D9 to D13. The number of clusters is not reported in Tables D2-D8 in Appendix II of the July Final Report.
[4]E/F, Resident/Family/Staff Questionnaire. [5]G is staff questionnaire (satisfaction and burnout)

The power to detect a true difference between treatment arms for outcome means from administrators, residents/families, and staff, respectively is shown in **Table C2** (next page) for continuous outcomes with analysis based on linear models. The bolded power values in **Table C2** indicate that power is excellent (exceeding 80%) to detect ***small to moderately small*** differences between the Accreditation and Control arms for Resident Outcomes (Forms B, E/F) and Staff satisfaction and burnout from the Staff Questionnaire. Power was lower but good for staff turnover reaching 80% for a Cohen's d of 0.30 (i.e., 30%). As a benchmark, power is also shown assuming no dropout of communities. As expected, power is always higher relative to the observed experience of community dropout.

Table C2. Percent Power[1] for *Aim 1* to detect a *small to moderately small effect size* (Cohen's d) for a continuous measure with two-sided Type I error of 0.05 for comparing control and intervention communities

Form	Description	Frequency of data collection	Number of observations per timepoint	Total Number of AIs at baseline		Cohen's d (small effects)		
					Total n	0.2	0.25	0.30
D3	Workforce - turnover	Bi-annual	1 administrator	102	510	53.0	72.2	86.4
		w/ dropout[3]	1 administrator	102	423	46.3	64.7	**80.2**
E/F[2]	Resident/Family Questionnaire	Bi-annual	8 residents	67	2,680	80.1	93.8	98.7
		w/ dropout[3]	8 residents[2]	67	1,784	69.7	**87.3**	**96.0**
G	Staff Questionnaire	Bi-annual	7 staff	68	2,380	78.6	93.0	98.4
		w/ dropout[3]	7 staff	68	1,575	67.2	**85.3**	**95.0**

[1]Power based on GEE with nested exchangeable correlation structure with 0.10 within-quarter and 0.05 between-quarter ICC. [2]Power for the D1 outcome (*Number of Different Caregivers*) is similar to Form E/F variables, since it is a continuous variable like them – producing means - collected bi-annually; however, its power is greater than shown as it is based on a greater number of residents. [3]Assumes the number of communities per quarter is as shown in Table 1.

Based on logistic regression analysis for AL-level longitudinal proportions (i.e., corresponding to dichotomous outcomes at the resident or staff level), power again focused on the actual dropout scenarios, shown in bold in **Table C3**, next page). Power exceeded 80% for small effect sizes (OR=0.8) only for the three Form B resident outcomes with 35 residents per AL-quarter, and for medium effect sizes (OR=0.67) for the three Form B outcomes with 5 residents per AL-quarter. Power was clearly inadequate for detecting small or medium effects for the medical management outcomes. For one such outcome – Medications Administered with Any Errors – power was 83% for detecting large effects. Similarly, upon considering workforce outcomes about standards being met, power was inadequate for Supervisors for all examined effect sizes. However, power was excellent (98%) to detect large odds ratios for all other staff types (**Table C3**).

Table C3. Power[1] provided by logistic model for *Aim 1* to detect small, medium or large odds ratios (1.25, 1.5 or 2.0, or inverted, 0.8, 0.67, 0.5) with two-sided Type I error of 0.05 for comparing control and intervention communities

Form/ Description	Frequency of data collection	Number of observations per timepoint	Total Number of ALs at baseline	Total observ-ations	Baseline preval-ence	Percent Power / Size of Effect		
						Small, OR=0.8	Medium, OR=0.67	Large, OR=0.5
B. Residents (few)[4]	Quarterly	5 residents	113	5650	0.20	81.7	>99.9	>99.9
	w/ dropout[3]	5 residents	113	4080	0.20	69.3	**99.0**	**>99.9**
Residents (many)[5]	Quarterly	35 residents	113	39,550	0.15	99.0	>99.9	>99.9
	w/ dropout[3]	35 residents	113	28,560	0.15	**97.0**	**>99.9**	**>99.9**
C. Medication - Management	Quarterly	5 residents	113	5650	0.01	10.2	22.0	46.5
	w/ dropout[3]	5 residents	113	4080	0.01	8.6	17.3	35.9
- Administered w/ Any Errors	Quarterly	35 medications	113	39,550	0.01	22.7	55.0	90.5
	w/ dropout[3]	35 medications	113	28,560	0.01	19.2	46.4	**83.4**
						OR=1.25	OR=1.5	OR=2.0
D2. Supervisors	Bi-annual	2 shifts	102	1020	0.97	8.0	15.7	32.4
	w/ dropout[3]	2 shifts	102	838	0.97	7.2	13.6	27.2
Other staff	Bi-annual	2 shifts	102	1020	0.70	33.8	77.8	99.3
	w/ dropout[3]	2 shifts	102	838	0.70	28.6	69.4	**97.9**

[1]Power based on GEE with nested exchangeable correlation structure with 0.01 within-quarter and 0.005 between-quarter ICC. [3]Assumes the number of communities per quarter is as shown in Table 1. [4]Hospital readmission, advance care planning, discharge due to behavior; [5]Fall w/major injury, ED visit, Hospital admission.

Results: Aim 2

Table C4 shows the sample sizes in terms of the number of communities having data at each quarter and condition (Not accredited, accredited) used in the power analysis for **Aim 2**. Notably, less than 25% of communities in the Accreditation arm ever became accredited. Also, dropout was substantial by Quarter 8 among communities that never became accredited.

Table C4. Total Number of Participating Communities per Quarter for Aim 2 used in power calculations

Form	Arm	Frequency of data collection	Number of AL Communities									
			Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
B[1] C[2]	Not Accredited	Quarterly	58	52	46	43	33	19	17	9	4	2
	Accredited	Quarterly	0	0	0	2	9	20	21	22	18	18
	Total	Quarterly	58	52	46	45	42	39	38	31	22	20
D[3]	Not Accredited	Twice a year	53		43		33		13		12	
	Accredited	Twice a year	0		0		9		24		23	
	Total	Twice a year	53		43		42		37		35	
E/F[4]	Not Accredited	Twice a year	33		31		13		2		1	
	Accredited	Twice a year	0		0		4		14		10	
	Total	Twice a year	33		31		17		16		11	
G[5]	Not Accredited	Twice a year	38		28		12		2		1	
	Accredited	Twice a year	0		0		5		15		13	
	Total	Twice a year	38		28		17		17		14	

The power to detect a true difference between treatment conditions for outcomes from administrators, residents/families, and staff, respectively is shown in **Table C5** (based on linear models) and **Table C6** (based on logistic models). For continuous outcomes, power ranged between 63% to 73% for detecting moderate differences between Accredited and Not Accredited conditions as given by Cohen's d of 0.5 (**Table C5**). Power was approximately 80% or greater to detect a difference between conditions that was moderately large (Cohen's d of 0.6).

Table C5. Percent Power[1] for *Aim 2* to detect *small, medium and large effect sizes* (Cohen's d) for a continuous measure with two-sided Type I error of 0.05 comparing Not Accredited and Accredited conditions with dropout

Form	Description	Frequency of data collection	Number of observations per timepoint	Total Number of ALs at baseline		Cohen's D (moderate effects)		
					Total n	**0.4**	**0.5**	**0.6**
D3	Workforce - turnover	Bi-annual	1 administrator	53	210	45.2	63.7	79.3
E/F[2]	Resident/Family Questionnaire	Bi-annual	8 residents	33	864	52.7	72.2	**86.6**
G	Staff Questionnaire	Bi-annual	7 staff	38	798	53.4	73.0	**87.1**

For outcomes that were proportions, power was always inadequate (well below 80%) to detect small effects (**Table C6**). Only for residents outcomes (Form B), did power exceed 80%. In particular, for resident outcomes with 5 residents per community-quarter power exceeded 80% only for large effects, whereas for residents with 35 residents per community-quarter, power also exceeded 80% to detect medium effects. Due to the aforementioned difficulties with either small sample sizes or rare outcomes, power was inadequate for Aim 2 hypotheses for Medication Management and Workforce (whether standards were met) outcomes.

Table C6. Power[1] provided by logistic model for *Aim 2* to detect small, medium or large odds ratios with two-sided Type I error of 0.05 for comparing Not Accredited and Accredited conditions with dropout

Description	Frequency of data collection	Number of observations per timepoint	Total Number of ALs at baseline	Total observ-ations	Baseline preval-ence	Percent Power / Size of Effect		
						Small, OR=0.8	Medium, OR=0.67	Large, OR=0.5
B. Residents (few)[4]	Quarterly	5 residents	58	1965	0.20	20.0	50.7	**89.4**
Residents (many)[5]	Quarterly	35 residents	58	13,755	0.15	60.0	**97.3**	**>99.9**
C. Medication Management	Quarterly	5 residents	58	1965	0.01	4.5	6.7	11.0
Medications Administered w/ Any Errors	Quarterly	35 medications[2]	58	13,755	0.01	8.7	17.9	37.9
						OR=1.25	OR=1.5	OR=2.0
D2. Supervisors	Bi-annual	2 shifts	53	420	0.97	4.2	6.0	9.3
Other staff	Bi-annual	2 shifts	53	420	0.70	9.5	21.0	47.8

Discussion

This power analysis conducted upon completion of the trial suggests that the study was largely successful in addressing the goals set at the study's initiation. In particular,

- statistical power for the intent-to-treat **Aim 1** based on the randomized groups was excellent for the detection of small to moderately small effect sizes for resident/family (Form E&F combined resident and family questionnaires - *resident outcomes and care coordination: psychosocial outcomes*) and Staff (Form G: satisfaction and burnout) continuous outcomes. Moreover, power was at least 80% for turnover (a workforce outcome) for detection of a difference in means that is 30% or greater than the standard deviation of the outcome as measured by Cohen's d. Power was also excellent for the comparison of intervention and control arms - based on odds ratios of small or medium effect size - for *resident outcomes and care coordination: healthcare related outcomes*.
- Power for **Aim 2** evaluation of becoming accredited was likewise good (at or exceeding 80% for moderately large effects) for *resident outcomes and care coordination: psychosocial outcomes, worker satisfaction, burnout and turnover.*

There were, however, two types of outcomes for which power was mostly inadequate:

- For four medication management outcomes, power never exceeded 80% for Aim 2. In Aim 1, power only exceeded 80% to detect large effects (odds ratio of 0.5) for the outcome *Number of Medications administered with Any Errors*.
- For outcomes on *whether staffing requirements were met*, power never exceeded 80% for Aim 2. In Aim 1, power exceeded 80% to detect large effects (odds ratio of 2.0) for other staff types, but not for supervisors.

These instances of lower power were impacted by multiple factors: (i) only about 25% of communities in the Accreditation arm ever became accredited, which reduced power in general for Aim 2; and (ii) substantial dropout at the community level. Additionally, outcomes on *medication management* and *whether staffing requirements were met* had intrinsic limitations, as some events almost never happened, namely, supervisors not meeting staffing expectations and medication errors. In this sense, the study provided clarity on the quality of outcomes as well as on those outcomes that were associated with the intervention and/or accreditation itself.

APPENDIX IV: NON-INFERIORITY ANALYSES OF QUALITY OUTCOME MEASURES

Table D1. Non-inferiority Analyses Determining Whether Accreditation Arm or Accredited Condition is No Worse than Control Arm or Not Accredited Condition According to a Non-inferiority Margin

Quality Outcome	Intent-to-Treat Analysis			Per Protocol Analysis		
	Odds Ratio/Mean (CI)	Non-inferiority Margin	Can Harm be Ruled Out?	Odds Ratio/Mean (CI)	Non-inferiority Margin	Can Harm be Ruled Out?
Fall with major injury	1.00 (0.55, 1.8)	1.25[a]	Ind	.95 (.53, 1.7)	1.25[a]	Ind
Emergency department visit	1.26 (.82, 1.9)	1.25[a]	Ind	.93 (.69, 1.3)	1.25[a]	Ind
Hospital admission	1.38 (.88, 2.1)	1.25[a]	Ind	.86 (.71, 1.1)	1.25[a]	**Yes**
Hospital readmission	.76 (.38, 1.5)	1.25[a]	Ind	.71 (.45, 1.1)	1.25[a]	**Yes**
Advance care planning discussion	1.13 (.44, 2.9)	0.8[b]	Ind	1.75 (1, 3.1)	0.8[b]	**Yes**
Discharge due to behavior	1.22 (.34, 4.4)	1.25[a]	Ind	.39 (.14, 1.1)	1.25[a]	**Yes**
Care transition preparation	.15 (-.14, .44)	-0.3[b]	**Yes**	.02 (-.24, .29)	-0.3[b]	**Yes**
Social activity participation	1.32 (-9.9, 12.6)	-10[b]	**Yes**	3.81 (-5.1, 12.7)	-10[b]	**Yes**
Days spoke on phone with family/friend	-.028 (-2.02, 1.45)	-1.4[b]	Ind	.11 (-1.52, 1.75)	-1.4[b]	Ind
Days resident had a visitor	-.08 (-1.36, 1.19)	-1.4[b]	**Yes**	-.33 (-1.36, 1.19)	-1.4[b]	**Yes**
Resident/family satisfaction	-.07 (-.39, .24)	-0.4[b]	**Yes**	-.18 (-.53, .17)	-0.4[b]	Ind
Person-centered care score	.01 (-.028, .29)	-0.5[b]	**Yes**	-.04 (-.32, .25)	-0.5[b]	**Yes**
Medication: Identity not confirmed	.46 (.08, 2.02)	1.25[a]	Ind	.93 (.02, 17.9)	1.25[a]	Ind
Ordered medications not administered	.78 (.33, 1.83)	1.25[a]	Ind	.70 (.12, 2.91)	1.25[a]	Ind
Medication errors	.27 (.17, .43)	1.25[a]	**Yes**	.39 (.10, 1.17)	1.25[a]	**Yes**
Significant medication errors	2.15 (.31, 23.8)	1.25[a]	Ind	.35 (0, 2.07)	1.25[a]	Ind
Number of different caregivers	.97 (.75, 1.24)	1.25[a]	**Yes**	1.1 (.94, 1.3)	1.25[a]	Ind
Staffing - Care aides traditional unit	1.02 (.45, 2.31)	0.8[b]	Ind	.61 (.32, 1.16)	0.8[b]	Ind
Staffing - Supervisors traditional	1.04 (.31, 3.51)	0.8[b]	Ind	.24 (.02, 1.47)	0.8[b]	Ind
Staffing - Administrators traditional	.56 (.18, 1.77)	0.8[b]	Ind	1.42 (.67, 3.02)	0.8[b]	Ind
Staffing - Care aides special care unit (SCU)	.85 (.28, .60)	0.8[b]	Ind	1.45 (.68, 3.09)	0.8[b]	Ind
Staffing - Supervisors SCU	.33 (.01, 3.41)	0.8[b]	Ind	1.52 (.12, 81.9)	0.8[b]	Ind
Staffing - Administrators SCU	.22 (.07, .72)	0.8[b]	Ind	1.08 (.43, 2.75)	0.8[b]	Ind
Staffing - Care coordinator SCU	.24 (.03, 2.22)	0.8[b]	Ind	.70 (.17, 2.8)	0.8[b]	Ind
Care aide turnover	.74 (.57, .95)	1.25[a]	**Yes**	.77 (.55, 1.08)	1.25[a]	**Yes**
Medication aide turnover	.84 (.53, 1.35)	1.25[a]	Ind	.76 (.53, 1.09)	1.25[a]	**Yes**
Nurse turnover	1.55 (.36, 6.66)	1.25[a]	Ind	.76 (.40, 1.44)	1.25[a]	Ind
Administrator turnover	.47 (.20, 1.11)	1.25[a]	**Yes**	.32 (.16, .62)	1.25[a]	**Yes**
Care coordinator turnover	.66 (.29, 1.52)	1.25[a]	Ind	.66 (.32, 1.37)	1.25[a]	Ind
Staff satisfaction	.04 (-.15, .23)	-0.3[b]	**Yes**	-.14 (-.31, .02)	-0.3[b]	**Yes**
Staff burnout	.92 (.50, 1.69)	1.25[a]	Ind	1.37 (.78, 2.43)	1.25[a]	Ind

Note: "Ind" refers to indeterminable.

[a]Upper limit of confidence interval had to be less than this value to rule out harm.

[b]Lower limit of confidence interval had to be greater than this value to rule out harm.