**4.2.3 Reliability Testing Results**

**Provide the statistical results from reliability testing for each level and type of reliability testing conducted. Where applicable, include results from accountable entity-level reliability testing (e.g., signal-to-noise testing) in the table below.**

***Data Element Level.*** Table 2a2.3.a shows the four CoreQ: AL Resident Satisfaction Questionnaire items, and the response per item for both the pilot survey of 100 residents and the re-administered survey of 98 residents. The responses in the pilot survey are not statistically significant from the re-administered survey. This shows that the data elements were highly repeatable and produced the same results a high proportion of the time when assessing the same population in the same time period.

Table 2a2.3.a: CoreQ: AL Resident Satisfaction Questionnaire Responses from Pilot and Re-Administered Surveys

|  |  |  |  |
| --- | --- | --- | --- |
| Question | Response | Percent [Pilot Survey (N=100)] | Percent [Re-Administered Survey (N=98)] |
| In recommending this facility to your friends and family, how would you rate it overall? | Poor | 6% | 7% |
| Average | 9% | 9% |
| Good | 15% | 14% |
| Very Good | 32% | 33% |
| Excellent | 38% | 39% |
| Overall, how would you rate the staff? | Poor | 2% | 2% |
| Average | 10% | 9% |
| Good | 16% | 16% |
| Very Good | 41% | 42% |
| Excellent | 32% | 33% |
| How would you rate the care you receive? | Poor | 2% | 2% |
| Average | 8% | 7% |
| Good | 17% | 17% |
| Very Good | 39% | 38% |
| Excellent | 35% | 34% |
| Overall, how would you rate the food? | Poor | 11% | 10% |
| Average | 13% | 13% |
| Good | 25% | 25% |
| Very Good | 32% | 33% |
| Excellent | 22% | 21% |

Table 2a2.3.b shows the average of the percent agreement from the first survey score to the second survey score for each item in the CoreQ: AL Resident Satisfaction questionnaire. This shows very high levels of agreement.

Table 2a2.3.b: Average Percent Agreement between 1st and 2nd Administered Surveys

|  |  |
| --- | --- |
| Questionnaire Item | Percent Agreement |
| 1. In recommending this facility to your friends and family, how would you rate it overall?
 | 98% |
| 1. Overall, how would you rate the staff?
 | 97% |
| 1. How would you rate the care you receive?
 | 98% |
| 1. Overall, how would you rate the food?
 | 97% |

1. ***Person/Questionnaire Level.*** Having tested whether the data elements matched between

the pilot responses and the re-administered responses, we then examined whether the person-level results matched between the Pilot CoreQ: AL Resident Satisfaction Questionnaire responses and their corresponding re-administered responses. In particular, we calculated the percent of time that there was agreement between whether or not the pilot response was poor, average, good, very good or excellent, and whether or not the re-administered response was poor, average, good, very good or excellent.The table (2a2.3.c) shows the CoreQ: AL Resident Satisfaction Questionnaire items, and the agreement in response per item for both the pilot survey of 100 residents compared with the re-administered survey of 98 residents. The person-level responses in the pilot survey are not statistically significant from the re-administered survey. This shows that a high percent of time there was agreement between whether or not the pilot response was poor, average, good, very good or excellent, and whether or not the re-administered response was poor, average, good, very good or excellent.

Table 2a2.3.c: Average Percent Agreement between Responses per Item for Pilot and Re-Administered Surveys

|  |  |  |
| --- | --- | --- |
| Questionnaire Item | Response | Percent Person-Level Agreement in Response for the Pilot Survey (N=100) vs. Re-Administered Survey (N=98) |
| 1. In recommending this facility to your friends and family, how would you rate it overall? | Poor | 97% |
| Average | 98% |
| Good | 95% |
| Very Good | 98% |
| Excellent | 99% |
| 2. Overall, how would you rate the staff? | Poor | 99% |
| Average | 98% |
| Good | 97% |
| Very Good | 98% |
| Excellent | 98% |
| 3. How would you rate the care you receive? | Poor | 100% |
| Average | 99% |
| Good | 98% |
| Very Good | 98% |
| Excellent | 98% |
| 4. Overall, how would you rate the food? | Poor | 95% |
| Average | 96% |
| Good | 97% |
| Very Good | 98% |
| Excellent | 97% |

Table 2a2.3.d: Average Percent Agreement between Response Options for the Pilot and Re- administered Surveys

|  |  |  |
| --- | --- | --- |
|   |  | Re- administered Response |
|  |  | Poor (1) or Average (2) | Good (3), Very Good (4), or Excellent (5) |
|  | Poor (1) or Average (2) | 97% | 98% |
| Pilot Response | Good (3), Very Good (4), or Excellent (5) | 97% | 98% |

1. ***MEASURE (FACILITY) LEVEL.*** After having performed the 10,000-repetition bootstrap, 21% of

bootstrap repetition scores were within 1 percentage point of the score under the original pilot sample, 33% were within 3 percentage points, 65% were within 5 percentage points, and 95% were within 10 percentage points. For the two-level signal-to-noise analysis for AL resident, R=0.84, indicating that 84% of facilities true score can be attributed to ratings from the respondents (AL residents) and remaining 16% is due to noise and differences among respondents. This result exceeds what is generally considered a good reliability coefficient of 0.8 (Campbell et al., 2010).

In summary, the measure displays a high degree of element-level, questionnaire-level, and measure (facility)-level reliability. First, the CoreQ: AL Resident Satisfaction questionnaire data elements were highly repeatable, with pilot and re-administered responses agreeing between 95% to 100% of the time, depending on the question. That is, this produced the same results a high proportion of the time when assessed in the same population in the same time period. Second, the questionnaire level scores were also highly repeatable, with pilot and re-administered responses agreeing 98% of the time. Third, a facility drawing residents from the same underlying population only varied modestly. The 10,000-repetition bootstrap results showed that the CoreQ: AL Resident Satisfaction measure scores from the same facility are very stable.

Campbell, JA, Narayanan, A., Burford, B., Greco, MJ. Validation of a multi-source feedback tool for use in general practice.  Education in Primary Care, 2010, 21, 165-179.