CBE ID 5110, Standardized Readmission Ratio (SRR) for dialysis facilities

Table 2a. Accountable Entity Level Reliability Testing Results by Denominator, Target Population Size

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Reliability	0.336	0.081	0.120	0.175	0.216	0.256	0.291	0.326	0.360	0.397	0.445	0.546	0.898
Mean Performance Score	1.00	0	0.89	0.93	0.94	0.96	0.98	1.02	1.03	1.03	1.08	1.01	1.07
*N of Entities	7,254	1	763	688	738	806	661	753	681	736	715	713	1
**N of Persons / Encounters / Episodes	458,111	11	12,894	18,273	25,467	34,677	33,846	45,510	47,930	60,542	71,711	107,261	1,097

**N of Persons/Encounters/Episodes is the number of index discharges.

Please note: The IUR deciles were calculated based on the sample size within each facility and some facilities had the same values, so were grouped into the same decile. Due to this reason, deciles may not have a consistent distribution of facility counts.

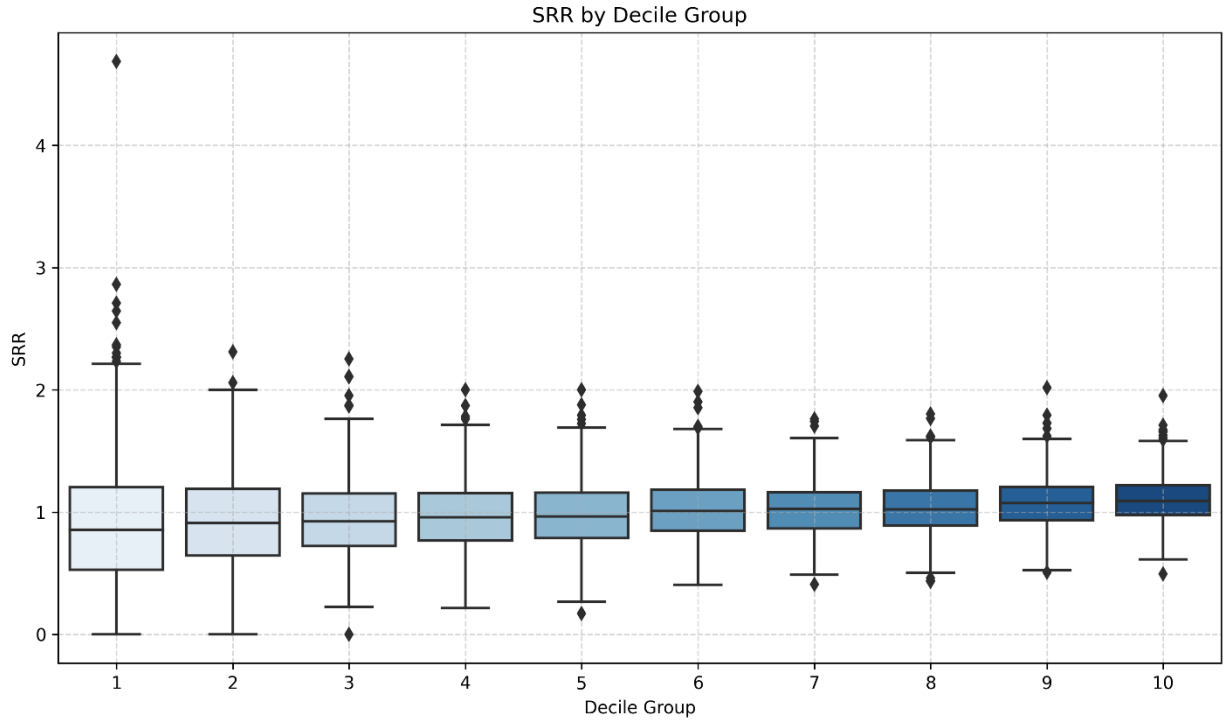
Table 2b. Accountable Entity Level Reliability Testing Results by Reliability Score

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Reliability	**0.336**	**0.081**	**0.118**	**0.175**	**0.216**	**0.255**	**0.290**	**0.325**	**0.360**	**0.396**	**0.444**	**0.536**	**0.898**

Reliability of SRR

Distribution of SRR in decile groups

To better understand the characteristics of entities with low and high reliabilities, the boxplot below displays the distribution of SRRs across decile groups. In general, the SRR distributions are similar across decile groups, with the median SRR remaining close to 1 in all cases.



Inclusion of small facilities

Reliability metrics naturally increase with the size of the entity being measured, given a fixed level of between- and within-entity variation. United States dialysis facilities are extremely small relative to measured entities in other care venues (e.g., hospitals), and the typical variation in health outcomes for dialysis patients would require an unrealistic number of very large dialysis facilities to achieve thresholds such as 0.6 for most facilities. Therefore, universal reliability standards for the entirety of care settings, when applied to this care setting, will result in the endorsement of only a small number of quality measures, and mostly those that are process or intermediate outcomes (especially unadjusted for additional risk) as reliability metrics are also known to decrease with improved risk adjustment.

Nevertheless, the readmission measure for kidney dialysis patients has proven highly useful in quality incentive programs. That said, there is an inherent trade-off between maximizing reliability and ensuring broader inclusion of facilities. In this case, including more facilities—even at the cost of lower reliability—will be preferable, as it improves participation and representation across entities and can have a greater impact on overall quality of care. If we are to include impactful quality outcomes in the portfolio of dialysis facilities, then we must accept a different standard for reliability. Failure to do so will result in a general absence of adequately risk-adjusted quality measures focused on health outcomes for the dialysis community, one of the most vulnerable patient groups within the healthcare system.

In addition, while smaller facilities naturally have measure values that exhibit greater uncertainty and variation, statistical hypothesis testing methods account for this variability and flag only providers with truly extreme results. In addition, the QIP includes a small-facility adjustment (generally applied to facilities with 25 or fewer eligible patients), which helps mitigate the low IURs observed in the first decile that would otherwise contribute to payment reductions. Both the star ratings and the QIP further reduce random noise by combining information across multiple measures when determining payment adjustments. As a result, even a measure with a low IUR can still contribute to raising the overall reliability of the combined measure set.

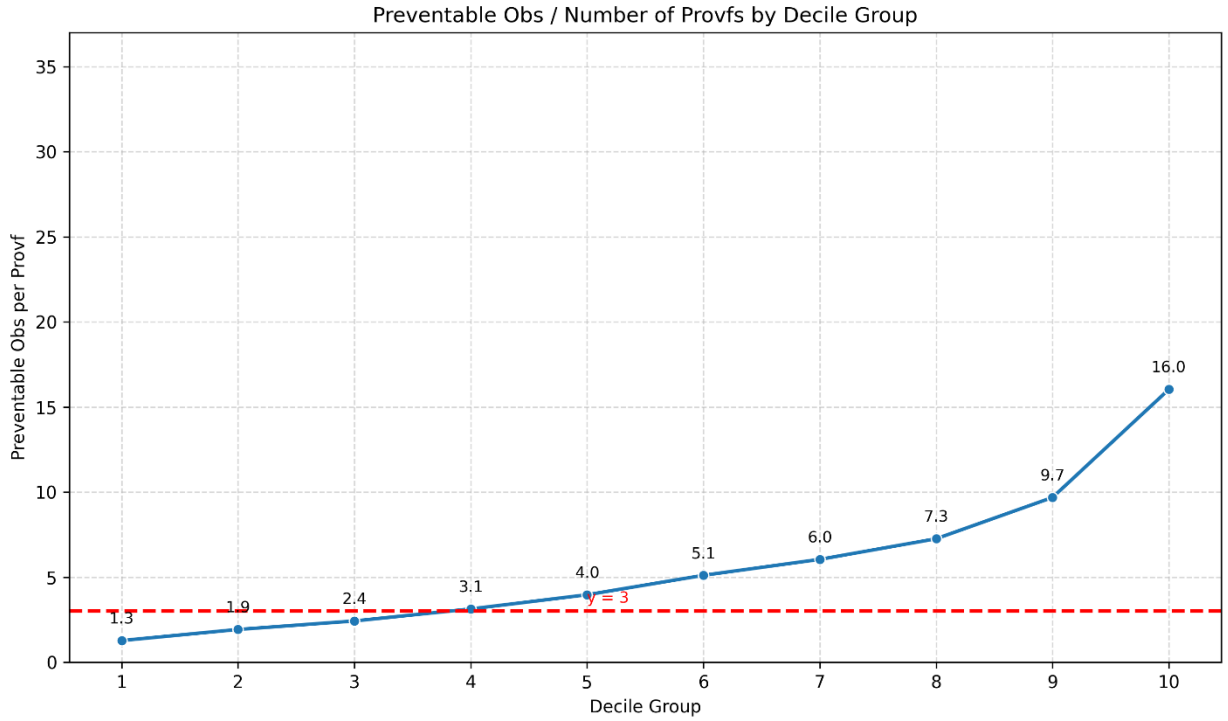
Table 1. Minimum dialysis facility size needed to achieve IUR thresholds for SRR.

Facility Size Needed		**Observed Distribution of Facility Sizes from Real Data**		
IUR=0.4	IUR=0.6	25^{th} perc.	median	75^{th} perc.
83 patients	187 patients	34	55	82

Preventable events per facility in decile groups

To illustrate the benefit of including more facilities in the SRR evaluation, we analyzed preventable events, defined as the difference between observed and benchmark event counts when the observed count exceeds the benchmark (and zero otherwise). The number of preventable events reflects the potential for improvement at a given facility. For SRR—calculated as the ratio of total observed events to total expected events over a given period, where lower values indicate better performance—the benchmark was set to the performance score of the third decile (0.70), where deciles here are defined by sorting facilities according to their performance scores.

We examined the ratio of total preventable events to the number of facilities in each decile group. A higher ratio generally indicates greater potential benefit from targeted quality improvement. The accompanying line plot shows the average preventable events per facility across decile groups, revealing a clear upward trend: facilities in higher deciles tend to have more preventable events on average. This pattern suggests that even smaller facilities in low-reliability groups still have a high average burden of adverse events, supporting the case for broader inclusion in SRR-based quality improvement programs.



Reference:

Geppert, J. (2025). Evaluating Importance (Impact) Claims about Measure Properties. MIDS Communication, Coordination, and Collaboration (C3) Forum. MIDS CORs & Contractors Quarterly Meeting.

Salerno, S., Yang, E., Dahlerus, C., Hirth, R.A., Han, P., Xu, T., Eckard, A., Agbenyikey, W., Horton, G.M., Clark, S., Messana, J.M., & Li, Y. (2025). Adding New Components to a Composite Quality Metric: How Good Is Good Enough? Medical Care. 63(4):293–299.

Nieser, K. J., & Harris, A. H. S. (2024). Comparing methods for assessing the reliability of health care quality measures. Statistics in Medicine, 43(23), 4575–4594.

Hartman, N., Shahinian, V. B., Ashby, V. B., Price, K., & He, K. (2023). Limitations of the inter-unit reliability: A set of practical examples. Health Services and Outcomes Research Methodology, 24(2), 156–169.

He, K., Kalbfleisch, J. D., Yang, Y., & Fei, Z. (2019). Inter-unit reliability for nonlinear models. Statistics in Medicine, 38(5), 844–854.