4.2.3a Reliability Testing Attachment (CBE ID 0531)

Table 4 (4.2.3): Distribution of component-level reliability estimates across hospitals

PSI component	N	Mean	Min	Q1	Med	Q3	Max	% Hospitals with reliability ≥ 0.6	% Hospitals with reliability ≥ 0.4
PSI03	3,145	0.52	0.00	0.26	0.57	0.77	0.98	47%	65%
PSI06	3,150	0.18	0.00	0.04	0.13	0.28	0.79	1.3%	12%
PSI08	3,150	0.12	0.00	0.03	0.08	0.18	0.70	0.10%	2.5%
PSI09	2,968	0.21	0.00	0.04	0.14	0.33	0.86	5.7%	18%
PSI10	2,821	0.12	0.00	0.01	0.05	0.17	0.89	2.5%	8.4%
PSI11	2,829	0.36	0.00	0.10	0.29	0.60	0.98	25%	40%
PSI12	2,971	0.28	0.00	0.09	0.23	0.43	0.91	11%	28%
PSI13	2,793	0.21	0.00	0.04	0.13	0.31	0.93	6.8%	18%
PSI14	2,878	0.07	0.00	0.02	0.04	0.09	0.63	0.03%	0.63%
PSI15	2,999	0.15	0.00	0.04	0.10	0.22	0.83	1.5%	9.1%

Note: N is the number of hospitals with 3+ denominator-eligible discharges for a given PSI component.

Table 5 (4.2.3): Distribution of composite-level reliability estimates across hospitals

ICC type	N	Mean	Min	Q1	Med	Q3	Max	% Hospitals with ICC ≥ 0.6	% Hospitals with ICC ≥ 0.4
ICC	2,922	0.63	0.02	0.47	0.69	0.82	0.98	62%	81%
ICC_corrected	2,922	0.75	0.04	0.64	0.82	0.90	0.99	78%	92%

Note: N is the number of hospitals with non-missing PSI 90 composite scores. ICC_corrected represents the ICCs after the Spearman-Brown correction has been applied.

Section 4.2.2: Full descriptions of reliability testing methods, including formulas

Component measure reliability:

We used a signal-to-noise approach to assess the precision of PSI component rates in distinguishing between hospitals according to their care quality. Signal-to-noise reliability assesses the extent to which variation in measure scores are attributed to true differences in quality across hospitals (signal) as opposed to error, or differences within hospitals (noise). Risk-adjusted PSI component rates were used in this analysis, as the signal-to-noise approach is not appropriate for smoothed rates, which are already corrected for each hospital's signal-to-noise reliability. All references to PSI component rates below assume risk-adjusted rates.

For each PSI component c and hospital h, we estimated a signal-to-noise reliability as: reliability = signal variance / (signal variance + noise variance). The reliability estimate can range from 0 to 1, with 0 indicating no reliability where all variation in component rates is due to noise, and 1 indicating perfect reliability where all variation in component rates is due to between-hospital differences. Noise variance ($\hat{\sigma}_{ch}^2$; one estimate per component per hospital) was estimated using the sampling variance of each hospital's PSI component rate. Signal variance ($\hat{\tau}_c^2$; one estimate per component) was computed iteratively using the empirical Bayes method (Morris, 1983) to estimate the between-hospital variance in each PSI component rate. These variances were estimated using the following formulas:

$$\hat{\sigma}_{ch}^2 = \left(\frac{\alpha_c}{n_{ch}E_{ch}}\right)^2 \sum\nolimits_{i \in A_{ch}} \hat{Y}_i\left(1-\hat{Y}_i\right),$$

where:

- α_c is the observed rate of component c in the reference population
- E_{ch} is the expected rate of component c for hospital h, defined as: $\frac{1}{n_h}\sum_{i \in A_{ch}} \hat{Y}_i$
- n_{ch} is the number of denominator-eligible discharges for component c in hospital h
- $\hat{Y}_i$ is the discharge-level predicted probability of experiencing the PSI component event for discharge i from the risk adjustment model
- A_{ch} is the set of denominator-eligible discharges for component c in hospital h

$$\hat{\tau}_c^2 = \frac{\sum_{h=1}^{H} \frac{1}{\left(\hat{\tau}_c^2 + \sigma_{ch}^2\right)^2} \left\{ \frac{H}{H-1}\left(\hat{p}_{ch} - \hat{\mu}_c\right)^2 - \hat{\sigma}_{ch}^{2} \right\}}{\sum_{h=1}^{H} \frac{1}{\left(\hat{\tau}_c^2 + \sigma_{ch}^2\right)^2}},$$

where:

- H is the number of hospitals
- $\hat{p}_{ch}$ is the risk-adjusted rate for component c for hospital h
- $\hat{\mu}_c$ is the mean risk-adjusted component rate for component c across hospitals
- $\hat{\sigma}^2_{ch}$ is the noise variance estimated for component c for hospital h above

In this computation, we assume that each hospital's true risk-adjusted rate for component c is normally distributed around the reference population rate with a spread of $\hat{\tau}^2_c$ (signal variance), and that each hospital's observed risk-adjusted rate is normally distributed around its true risk-adjusted rate with a spread of $\hat{\sigma}^2_{ch}$ (noise variance). An iterative estimation procedure is required for estimating the signal variance because $\hat{\tau}^2_c$ appears on both sides of the equation above.

Morris, C. N. (1983). Parametric empirical Bayes inference: theory and applications. Journal of the American statistical Association, 78(381), 47-55.

Composite measure reliability:

The signal-to-noise approach is not appropriate for assessing the composite reliability, because the PSI 90 composite is constructed from smoothed PSI component rates, which already account for each hospital's signal-to-noise reliability (see section 1.18, step #4 for an explanation of the smoothing process). Therefore, we used a split-half reliability approach and calculated an intraclass correlation coefficient (ICC) for each hospital has:

$$ICC_h = \frac{\hat{\sigma}^2_b}{\hat{\sigma}^2_b + \frac{\hat{\sigma}^2_e}{n_h}},$$

where $\hat{\sigma}^2_b$ is the between-hospital variance component, $\hat{\sigma}^2_e$ is the error variance component, and n_h is the number of unique denominator-eligible discharges across all PSI components for hospital h. Similar to signal-to-noise, the ICC quantifies the amount of variation in measure scores due to between-hospital differences rather than within.

We obtained estimates of the between-hospital and error variance components from a simple, intercept-only random effects model with no predictors. This random-effects model was fit to data made multilevel by creating random split-half samples, without replacement, of discharges (denominator-eligible for at least one PSI component) within each hospital and thereby producing two PSI 90 composite scores per hospital. The random effects model can be expressed using the following equation and distributional assumptions:

$$Y_{ht} = \mu + \alpha_h + \varepsilon_{ht}$$

$$\alpha_h \sim N\left(0, \sigma^2_b\right); \varepsilon_{ht} \sim N\left(0, \frac{\sigma^2_e}{n_{ht}}\right)$$

where Y_{ht} is the PSI 90 composite score for hospital h for subsample t, μ is a fixed effect or intercept indicating the mean composite score, α_h is a hospital-level random effect for hospital h, and ε_{ht} is the residual subsample-level random effect for hospital h's subsample t.

With maximum likelihood estimates $\hat{\sigma}_b^2$ and $\hat{\sigma}_e^2$ obtained from the random effects model, we plugged these values into the formula above to compute an ICC per hospital. ICCs range from 0 to 1, where 0 indicates no agreement or reliability, and 1 indicates perfect agreement or reliability. ICCs were additionally corrected using the Spearman-Brown formula to account for the fact that composite scores Y_{ht} were computed using only half of the discharges from each hospital.