SUPPLEMENTARY MATERIALS

Calculations for constructing each type of control chart, estimating baseline SSI rates, and computing each chart's plotted points are outlined below. For the phase-1 results in the main manuscript, these calculations were applied to all data in the identified reference baseline period. For phase-2 results shown below in Figure S1 and Table S1, calculations were applied iteratively to identify and remove from the baseline period any special cause variation, since this can artificially inflate the estimated mean and standard deviation (σ), producing wider limits and possibly reduced detection of true signals.

Baseline SSI Rate Estimation: To estimate the baseline SSI rate $\hat{\pi}_j$ for a given hospital and procedure at month $j \geq 1$, a reference data source and time interval are needed. For the former, and temporarily ignoring lags, either local hospital or network-wide data were used. For the latter, either fixed intervals (e.g., the initial year) or rolling windows of a given size w were used. The baseline SSI rate used to evaluate SSI data from month j then was estimated as

$$\hat{\pi}_j = \left\{ \begin{matrix} \frac{\sum_{i=1}^{w} x_i'}{\sum_{i=1}^{w} n_i'} & \text{if fixed} \\ \frac{\sum_{i=\max(1,j-w+1)}^{j} x_i'}{\sum_{i=\max(1,j-w+1)}^{j} n_i'} & \text{if rolling} \end{matrix} \right\}, \quad (1)$$

where x_i' and n_i' are the total numbers of SSIs and surgeries in month i in the reference data source. Calculation of rolling baselines for time periods shorter than the window size ($j < w$) used only the first j data points following standard practice.[1,2]

To avoid potential contamination of baselines estimates by potential SSI rate changes in recent months, rolling windows also were generalized to potentially be lagged by $l \geq 0$ months, where now

$$\hat{\pi}_j = \frac{\sum_{i=\max(1,j-w-l+1)}^{\max(1,j-l)} x_i'}{\sum_{i=\max(1,j-w-l+1)}^{\max(1,j-l)} n_i'}. \quad (2)$$

As above, in cases where the window size or lag required a longer time interval than exists near the start of the dataset (i.e., $j < w + l$), the window size and, if necessary, lag were decreased to a minimum of 1 and 0, respectively, in order to allow a baseline period of at least one month. Baseline alternative investigated included rolling windows of 3, 6, 9, 12, 18, or 24 months with lags of 0, 3, 6, 12, or 24 months, as well as fixed intervals of the same sizes, for a total of 36 different baseline periods.

Monthly SSI Calculation. For each month j, the observed SSI proportions were calculated as the number of infections that occurred in that month, x_j, divided by the total number of surgeries n_j performed that month,

$$p_j = \frac{x_j}{n_j}. \quad (3)$$

For ease of interpretation and adjusting for changes in the rolling baseline rates $\hat{\pi}_j$ and number of surgeries n_j, these monthly observed data were standardized as

$$z_j = \frac{p_j - \hat{\pi}_j}{\sigma_j}, \quad (4)$$

where the corresponding monthly standard deviation σ_j for month j was estimated as

$$\hat{\sigma}_j = \sqrt{\frac{\hat{\pi}_j(1-\hat{\pi}_j)}{n_j}}. \quad (5)$$

The resulting z_j scores are approximately independent and identically distributed with a mean of 0 and an estimated standard deviation 1, allowing for easy monitoring via Shewhart, MA, and EWMA p-charts. Note that to ensure z_j is well-defined when $\hat{\pi}_j = 0$ or 1, the inverse of twice the number of surgeries performed during the baseline period of j was added or subtracted, respectively, from $\hat{\pi}_j$ prior to computing σ_j following convention.[3] Similarly, for baseline periods without any surgical cases (i.e., $n_j = 0$), the prior month's baseline SSI rate, $\hat{\pi}_{j-1}$, was used instead.

Control Chart Calculation: To construct Shewhart p-charts, these standardized SSI monthly rates (z_j) were plotted against a center line equal to 0 and control limits of $\pm k$, where k is the number of standard deviations used to compute the limits in traditional charts. Note that this standardized chart is exactly equivalent to plotting the monthly non-standardized p_j terms against control limits set to

$$UCL_j, LCL_j = \hat{\pi}_j \pm k\sqrt{\frac{\hat{\pi}_j(1-\hat{\pi}_j)}{n_j}}, \tag{6}$$

although this would have a non-constant center line $\hat{\pi}_j$ for any given month j. The standardized versions thus were used for plotting convenience, visual interpretation ease, and to facilitate the below calculation of MA and EWMA charts.

To construct moving average (MA) p-charts with a window of size s, the same z_j standardized SSI rates were used to compute a moving average statistic, m_j, plotted for each month j and its corresponding standard deviation σ_{m_j} following established procedures,[1]

$$m_j = \frac{\sum_{i=\max(1,j-s+1)}^{j} z_i}{\min(j,s)}, \tag{7}$$

and

$$\sigma_{m_j} = \frac{1}{\sqrt{\min(j,s)}}, \tag{8}$$

where s is the MA span, set here at 3, 6, 9, or 12 months, and the numerator in equation (8) is $\sigma_{z_j} = 1$. Similarly, to construct exponentially-weighted moving average charts, the standardized z_j terms were used to compute an EWMA statistic, e_j, plotted for each month j,

$$e_j = \lambda \sum_{i=1}^{j} z_i (1-\lambda)^{j-i} = \lambda z_j + (1-\lambda) e_{j-1}, \quad (9)$$

and a corresponding standard deviation σ_{e_j} of

$$\sigma_{e_j} = \sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2j}]}, \quad (10)$$

where $0 \leq \lambda \leq 1$ is the EWMA weight parameter, varied in our analysis set equal to 0.2, 0.4, 0.6, or 0.8.

Signal Detection. To explore tradeoffs between the number of signals and their average magnitude, control limits for each type of chart were set to $k = 0.5$, 1, 2, 3, or 4 standard deviations (σ_{z_j}, σ_{m_j}, or σ_{e_j}). For Shewhart p-charts, statistical signals were defined as any point outside the control limits or violating any commonly-used supplementary within-limit rules.[4] Since our primary focus was on detecting significant SSI rate increases, signals were examined further if one or more of the following conditions were met:

- standardized SSI z_j scores greater than the upper control limit,
- two of three consecutive z_j scores greater than 2/3 of the distance from the centerline to the upper control limit,
- four of five consecutive z_j scores greater than 1/3 the distance to the upper control limit,
- nine consecutive z_j scores rates greater than 0, or
- six consecutive z_j scores exhibiting a monotonously increasing trend.

For MA and EWMA p-charts, only the first rule was used (a plotted point above the upper control limit), since the supplementary rules classically do not apply to these other types of control charts due to the non-independence of consecutive plotted points.

Performance Metrics. To evaluate the predictive performance of individual chart variations and dual-chart combinations, we computed the below standard binary classification metrics. This assessment was restricted to the subset of statistical signals reviewed by epidemiologists during first stage of the study, since there was no apparent way to estimate the clinical relevance of the remaining initial signals or additional signals generated during the second stage, and was conducted separately for training and validation subsets. In each case, we computed:

- True positives (TP) = number of signals occurring during months ranked by clinicians as moderate or high concern, pooled across all hospital-procedure combinations;
- False positives (FP) = number of signals occurring in months rated as low or no concern;
- True negatives (TN) = number of months rated as low or no concern with no signals;
- False negatives (FN) = number of moderate or high concern months with no signals.

We then computed the following performance measures:

- Sensitivity $(Sens) = TP/(TP + FN)$;
- Specificity $(Spec) = TN/(FP + TN)$;
- Positive predictive value $(PPV) = TP/(TP + FP)$;
- Negative predictive value $(NPV) = TN/(TN + FN)$;
- Positive log-likelihood ratio $(LR+) = Sens/(1 - Spec)$;
- Negative log-likelihood ratio $(LR-) = (1 - Sens)/Spec$;
- Youden's index (informedness) $= Sens + Spec - 1$;

- Markedness $= PPV + NPV - 1$;
- Diagnostic odds ratio $= (LR+)/(LR-)$;
- Accuracy (fraction correct) $= (TP + TN)/(TP + FP + TN + FN)$; and
- F_1 score $= 2\ PPV\ Sens\ /(PPV + Sens)$.

To estimate the level of over-signaling within the remaining (non-rated) data points, we computed the average number of signals per month per hospital-procedure.

REFERENCES

1. Montgomery DC. *Introduction to Statistical Quality Control.* 6th ed. New York, NY: John Wiley & Sons, Inc.; 2009.
2. DeVor RE, Chang T, Sutherland JW. *Statistical Quality Design and Control: Contemporary Concepts and Methods*. 2nd ed. Upper Saddle River, NJ: Prentice Hall, Inc.; 2006.
3. Basu AP, Gaylor DW, Chen JJ. Estimating the probability of occurrence of tumor for a rare cancer with zero occurrence in sample. *Regul Toxicol Pharmacol*. 1994; 23(2):139-144.
4. The Western Electric Company. *Statistical Quality Control Handbook.* New York, NY: Western Electric Co., Inc.; 1958.

Figure S1. Impact of chart design parameters on chart performance when using phase-2 baseline and control limit calculations. Plots of sensitivity versus specificity for individual charts (**A-E**) and dual-chart combinations (**F**), with color and symbol indicating different design parameter values (**A-E**) or types of combinations (**F**). In contrast to our primary analysis (Figure 3), baseline rates were recalculated after excluding any out-of-control points located within the baseline periods; this process was repeated until all remaining points were in control with respect to their mean (0-7 iterations, average = 1.4). **A:** Baseline data source; DICON, Duke Infection Control Outreach Network; Local, individual hospital. **B:** Control limit (CL) width; σ, estimated standard deviation. **C:** Moving average (MA) span and equivalent EWMA averaging horizon; Med., medium. **D:** Baseline window size; mo., months. **E:** Baseline lag length. **F:** Types of dual-chart combinations; Disj., disjunction; Conj., conjunction; Excl., exclusive disjunction; Diff., set difference; Indiv., individual charts (shown for reference).

Table S1. Optimized moving average charts with phase-2 baseline calculations. Chart parameters and performance for the optimal dual-chart approach, each of these two charts individually, and optimal individual charts (overall, and relying only on local data). In contrast to our primary analysis (Table 3), baseline rates were recalculated iteratively excluding out-of-control points located within the baseline periods. Parameter values are listed as median and, in brackets, range over the best 20 solutions in each category. Performance measures and ranks (out of 3,600 individual and 32.4 million dual charts) are based on the training dataset, with values in square brackets based on the validation dataset.

Approach	**Solution**	**Disjunctive (“or”) combinations**			**Optimal individual chart**
		Optimal combination	**Each chart individually**		
Parameters	**Baseline data**	N/A	Network	Local	Network
	Baseline size	N/A	6 [3-24]	3 [3]	12 [3-24]
	Baseline lag	N/A	3 [0-24]	3 [3-6]	6 [0-24]
	MA span	N/A	12 [9-12]	6 [6-12]	12 [12]
	CL width (σ)	N/A	2 [1-2]	1 [1-3]	1 [1]
Chart performance	**Sensitivity**	0.84 [0.74]	0.49 [0.47]	0.63 [0.52]	0.74 [0.80]
	Specificity	0.72 [0.79]	0.90 [0.96]	0.79 [0.82]	0.76 [0.82]
	PPV	0.58 [0.64]	0.71 [0.87]	0.58 [0.59]	0.59 [0.70]
	NPV	0.91 [0.85]	0.79 [0.78]	0.82 [0.76]	0.86 [0.89]
	Positive LLR	2.95 [3.49]	5.14 [13.1]	3.01 [2.82]	3.04 [4.44]
	Negative LLR	0.22 [0.33]	0.56 [0.55]	0.47 [0.59]	0.35 [0.25]
	Youden’s index	0.56 [0.53]	0.40 [0.44]	0.42 [0.33]	0.49 [0.62]
	Markedness	0.49 [0.50]	0.50 [0.65]	0.41 [0.36]	0.45 [0.58]
	Diagnostic OR	13.3 [10.7]	9.16 [23.9]	6.45 [4.75]	8.75 [18.1]
	Accuracy	0.76 [0.77]	0.77 [0.80]	0.74 [0.71]	0.75 [0.81]
	F_1 score	0.69 [0.69]	0.58 [0.61]	0.61 [0.55]	0.65 [0.74]
	Signals / month	0.19 [0.15]	0.11 [0.09]	0.11 [0.10]	0.21 [0.21]
	Rank	#9 [#161607]	#264 [#322]	#198 [#613]	#19 [#1]

MA, moving average; CL, control limits; σ, estimated standard deviation; PPV, positive predictive value; NPV, negative predictive value; LLR, log likelihood ratio; OR, odds ratio.