



PC4

Risk Recalibration

Leader Workbook

Understanding model updates, risk-adjusted reporting, coefficient changes, and future scope

PC4 • Post-STS Update 2025/2026

How to Use This Workbook

This workbook is intended to give PC4 leaders a practical, non-technical reference for understanding what changed during the 2025/2026 risk-model recalibration, how the reporting measures should be interpreted, how model coefficients were reviewed, and what work should—and should not—be assumed in future recalibration cycles.

Contents

- 1. Observed, Expected, O/E, and Adjusted Rates
- 2. Understanding Model Coefficients and Coefficient Magnitude
- 3. What Was Done in the 2025/2026 PC4 Recalibration
- 4. What Is Standard for Future Recalibrations
- 5. What Is Conditional or Out of Scope
- 6. Physician Leader Review Questions

1 | Understanding the Rates Shown in PC4

Observed Rate — What actually happened?

The observed rate is the hospital's actual outcome rate during the selected period. It describes the hospital's real experience and is not adjusted for differences in patient risk.

Expected Rate — What would we expect for patients like ours?

The expected rate is the outcome rate predicted if an average-performing PC4 hospital treated patients with the same measured risk profile as the hospital's patients. The risk model uses patient-level characteristics associated with the outcome to estimate each encounter's expected risk, and those predictions are aggregated to the hospital level.

Important distinction

The expected rate is not the hospital's performance score. It describes the outcome level predicted from its patient risk profile, using typical PC4 performance as the reference.

O/E Ratio — How did actual outcomes compare with expected outcomes?

$O/E = \text{Observed Rate} \div \text{Expected Rate}$

- $O/E = 1.0$: outcomes were approximately equal to what was expected based on patient risk and typical PC4 performance.
- $O/E > 1.0$: more events occurred than expected.
- $O/E < 1.0$: fewer events occurred than expected.

For example, an O/E ratio of 1.25 means the hospital experienced approximately 25% more events than would be expected if an average-performing PC4 hospital treated patients with the same measured risk profile. The distance from 1.0 reflects hospital performance that is not explained by the patient risk factors included in the model.

Adjusted Rate — A common scoreboard for hospital comparison

$\text{Adjusted Rate} = O/E \text{ Ratio} \times \text{Overall PC4 Average Expected Rate}$

The adjusted rate takes the relative performance represented by the hospital's O/E ratio and applies it to a common PC4 patient-risk reference. Conceptually, it answers: Given the performance reflected in our O/E ratio, what would our outcome rate look like if our patient population had the overall PC4 risk profile?

What the adjusted rate is—and is not

The adjusted rate is a standardized comparison rate, not the hospital's actual event rate. The observed rate tells you what actually happened. The adjusted rate is intended to support fairer comparison among hospitals with different patient populations.

Worked example

Measure	Value	Interpretation
Observed mortality	1.97%	The actual mortality rate.
Expected mortality	1.58%	Predicted for these patients if treated at an average-performing PC4 hospital.
O/E ratio	≈ 1.25	About 25% more deaths than expected for the measured patient risk profile.
Overall PC4 average expected	1.82%	The common risk reference used for standardization.
Adjusted mortality	2.28%	$1.25 \times 1.82\%$; standardized rate for peer comparison.

Four questions to remember

Observed: What happened? • Expected: What would be predicted for patients with our risk profile at a typical PC4 hospital? • O/E: How did we perform relative to that expectation? • Adjusted: What would that relative performance look like on the common PC4 patient-risk scale?

2 | Understanding Coefficients and Coefficient Magnitude

What is a model coefficient?

A coefficient is a weight assigned to a predictor in the statistical model. Its sign indicates the direction of association with the modeled outcome, while its size indicates how strongly that model term moves the model's underlying score, holding the other model terms constant.

What is coefficient magnitude?

Coefficient magnitude is the absolute value of the coefficient, written $|\beta|$. Taking the absolute value intentionally removes direction so that the size of model terms can be screened within the same model.

- Larger $|\beta|$: a one-unit change or class contrast produces a larger movement in the model's linear predictor, all else equal.
- Smaller $|\beta|$: the model term produces a smaller raw movement in that model score.
- The signed coefficient must still be reviewed to determine whether the modeled outcome increases or decreases.

How magnitude was applied in the coefficient workbook

The PC4 coefficient comparison workbook compared the previous production coefficients with the 2026 recalibrated coefficients at both the individual term level and the model-variable level.

Workbook field	How to interpret it
Old / Current coefficient	The signed model weight before and after recalibration.
Old / Current coefficient magnitude	The absolute coefficient, $ \beta $, used to compare raw effect size within that model.
Magnitude change	Current $ \beta $ – old $ \beta $. Positive = stronger raw effect; negative = weaker raw effect.
Magnitude rank	Variables/terms ranked within the same model; rank 1 has the largest raw magnitude.
Rank improvement	Old rank – Current rank. Positive means movement closer to rank 1.
Variable magnitude	For categorical predictors, the maximum $ \beta $ across the variable's class terms.
$\exp(\text{coefficient})$	For logistic models, interpretable as an odds ratio; for log-linked models, a multiplicative change in expected outcome.

What magnitude does NOT mean

Coefficient magnitude is not a probability-point change, causal effect, statistical significance, or prevalence. It also does not tell us how often a risk factor occurs. Raw magnitudes should not be compared across different models with different outcomes, coding, units, or link functions.

Why this was useful for the recalibration review

The coefficient workbook provides a transparent way to see which predictors were retained, added, or changed; how their model weights moved; and which terms have the largest raw effects within each model. It is best used as a model-review and screening tool rather than as a claim that a ranked variable is clinically the 'most important' risk factor.

The comparison covered 23 model specifications. The previous coefficient file contained 374 numeric coefficient rows and the current file contains 388.

3 | What Was Done for the 2025/2026 PC4 Recalibration

The recalibration was more than a simple coefficient refresh. It combined model fitting and validation with additional investigation required because the STS update changed or affected model inputs and definitions.

Core modeling and validation work

- Validated model specifications and confirmed required variables were available in the analytic dataset.
- Performed two-stage feature selection: univariate fitting followed by multivariate stepwise selection.
- Used an 80/20 train/test split in the later modeling workflow.
- Fit logistic GLMs for binary outcomes and Poisson GLMs for LOS/ventilation-duration outcomes.
- Generated out-of-sample predictions and model-specific validation statistics on held-out data.
- Refit approved model specifications on the full available population after validation.
- Calculated overall observed, expected, and O/E results after final fitting.
- Extracted and persisted recalibrated model parameters for production use.
- Generated encounter-level predictions and aggregated results to the site level.
- Compared current-production versus recalibrated observed, expected, and O/E results for each site and modeled outcome.
- Produced Excel outputs supporting site-level impact review and coefficient/model-driver review.

Why site-level O/E comparison matters

Aggregate model performance can look reasonable even when changes in expected values materially redistribute results across hospitals. Comparing old and new site-level O/E therefore serves as a regression and impact test: it shows whether recalibration creates clinically or operationally meaningful changes in hospital performance reporting.

Additional investigation required this cycle

- Investigated the effect of STAT score changes, including alternative model specifications with STAT-score predictors removed and reintroduced.
- Investigated changed or newly available variable definitions associated with the STS update.
- Identified processing implications where variable changes required corresponding surgical/analytic XML updates.
- Addressed variables that no longer had useful variation in the new data; for example, `flg_version3` was dropped from the unplanned-complication model when it had only one value.
- Performed additional root-cause review where new versus production O/E results shifted unexpectedly.

4 | What Should Be Expected in Future Recalibrations?

Standard recurring work — in scope every cycle

Recurring workstream	Purpose
Input/specification validation	Confirm required model variables exist, are usable, and have reasonable distributions.
Model recalibration	Refit the defined production logistic and Poisson GLMs using the approved current data.
Train/test validation	Evaluate model performance on held-out observations before final refitting.
Final full-population fit	Refit the approved specification using the complete modeling population.
Overall O/E validation	Confirm observed versus expected behavior for every model.
Parameter generation	Create and hand off the approved production coefficients/parameters.
Site-level old/new O/E comparison	Compare hospital-level observed, expected, and O/E before and after recalibration.
Prior-cycle comparison	Review model performance and material changes relative to the prior production cycle.

Conditional / not automatically in scope

The following work should not be assumed to be part of every routine recalibration. It becomes necessary when the underlying specification, source data, or validation results indicate a problem that requires investigation.

Conditional work	When it becomes in scope
Changed or renamed variable investigation	Only when specifications or source-data definitions change.
Alternative model specifications	Only when validation fails, a variable becomes unusable, or a specification change requires testing alternatives.
STAT-score / no-STAT-score experiments	Specific to this cycle's STS-related investigation; not an automatic annual task.

Repeated model proving/re-proving after unexpected O/E shifts	Exception work triggered when standard regression/impact tests identify unexplained movement.
Tracing variable changes into XML/analytic processing	Engineering/change-impact work required when upstream definitions or availability change.
Extensive site-level root-cause analysis	Triggered when the standard old-versus-new comparison identifies material unexplained changes.

Planning principle

Future cycles should be planned around a repeatable core package: data/specification validation → model fit → validation → O/E impact review → production handoff. Investigation caused by a major source-data or specification change should be estimated separately rather than treated as routine recalibration effort.

5 | Leader Review Guide

The following questions can help leaders focus review on the clinical meaning of the recalibration without requiring detailed statistical implementation knowledge.

Review area	Question for physician leaders
Expected rates	Do changes in expected rates make clinical sense given the risk characteristics of the populations being treated?
O/E movement	Are there hospitals or outcomes with large changes in O/E that warrant closer review?
Adjusted rates	Do changes in adjusted rates appear consistent with the changes in O/E and the updated collaborative benchmark?
Model drivers	Do the highest-magnitude predictors within a model make clinical sense, and are there surprising additions, removals, or large changes?
STS-related variables	Where an STS definition or score changed, is the resulting model behavior clinically plausible?
Site impact	Are there hospitals whose relative standing changes substantially after recalibration, and can that change be explained?
Interpretation	Are users distinguishing actual observed outcomes from expected risk and standardized adjusted performance?

Key takeaways

- Expected outcomes reflect patient risk relative to typical PC4 performance; they are not themselves a hospital performance score.
- O/E describes how observed outcomes compare with what was expected for that hospital's measured case mix.
- Adjusted rates put hospitals onto a common PC4 patient-risk scale for fairer peer comparison.
- Coefficient magnitude is a useful within-model screening measure, but it should not be interpreted as causal or as a universal measure of clinical importance.
- Routine recalibration should include model validation and site-level impact testing; extensive root-cause and specification redesign work is conditional.