Latent Anastomotic Leak Susceptibility Phenotypes Across Gastrointestinal Operations and Postoperative Care Environments

Fahad Siddiqui[1] **and Kamran Ashraf**[2]

[1]Karakoram International University, University Road, Gilgit 15100, Pakistan

[2]University of Sahiwal, Farid Town Road, Sahiwal 57000, Pakistan

ABSTRACT

Anastomotic leak after gastrointestinal reconstruction is commonly reported as an average complication rate attached to a procedure or a hospital. That practice is clinically convenient but biologically coarse, because patients who experience leak do not appear to travel through a single pathophysiologic route. Some deteriorate through early inflammatory and perfusion failure, some through technically stressed local anatomy, and others through chronic nutritional or frailty-related vulnerability that becomes visible only after the operation has already begun to tax physiologic reserve. If those hidden susceptibility states are unevenly distributed across surgery types and care environments, average rates will conceal an important layer of risk structure. This investigation estimated latent leak-susceptibility phenotypes in a large multicenter electronic health record cohort and then examined how those phenotypes were sorted across gastrointestinal procedures and postoperative care environments. The cohort contained 53,418 adult restorative gastrointestinal operations from 21 hospitals. Anastomotic leak within 30 days occurred after 3,446 procedures, producing an overall incidence of 6.5%. A joint latent class outcome model selected four phenotypes. Their estimated prevalences were 38.9%, 22.7%, 17.1%, and 21.3%, with corresponding leak incidences of 2.7%, 10.9%, 11.6%, and 7.8%. Procedure families and care environments differed substantially in phenotype composition. Low pelvic colorectal surgery was strongly enriched for a mechanically stressed phenotype, whereas esophagogastric reconstruction accumulated a depletion-dominant phenotype. Emergency overnight care contained the highest proportion of the inflammatory-perfusion phenotype. Across the procedure-by-setting matrix, 51.8% of leak-rate dispersion was attributable to phenotype sorting, 28.4% to differential phenotype expression within the same phenotype, and 19.8% to residual interaction. The findings indicate that cross-setting heterogeneity in leak is partly a problem of hidden clinical states being redistributed across pathways rather than only a problem of visible case mix.

KEYWORDS

1. Introduction

Anastomotic leak is often treated analytically as though it were a single adverse event with one dominant explanation. Surgical papers report rates by operation, by hospital, or by surgeon volume, and quantitative work often proceeds by fitting one global model that maps patient variables to a leak probability. That approach has practical advantages, yet it leaves unanswered a question that is both clinical and empirical. When leak rates differ across gastrointestinal operations and postoperative care environments, is the difference simply the visible result of procedure complexity and measured severity, or are hospitals and pathways selectively accumulating hidden susceptibility states

Article info:

Fahad Siddiqui and Kamran Ashraf. *Latent Anastomotic Leak Susceptibility Phenotypes Across Gastrointestinal Operations and Postoperative Care Environments.* Nextqueries.

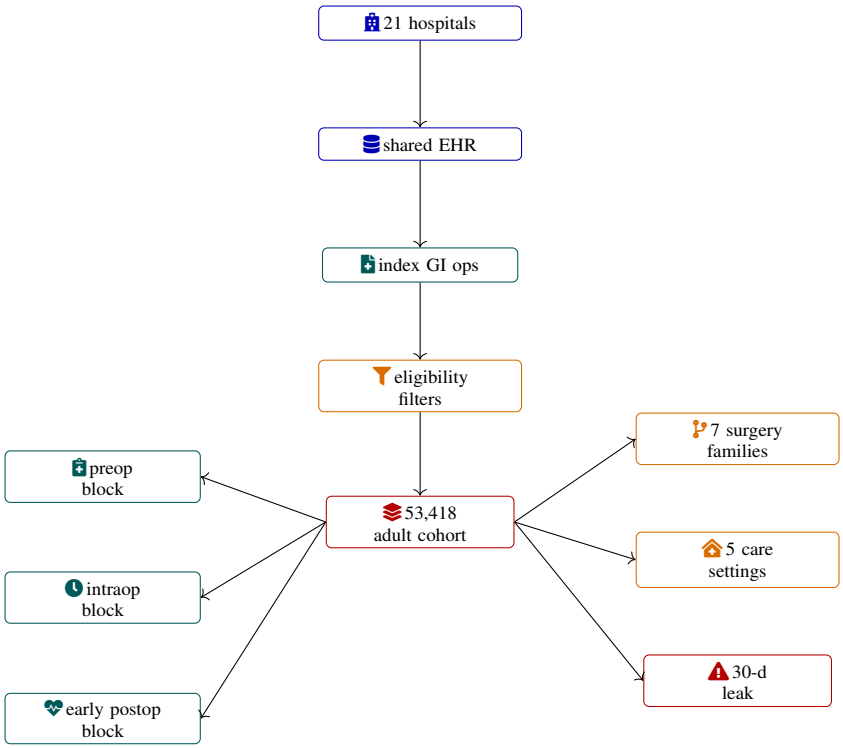

Figure 1 Cohort assembly and measurement architecture. A common EHR backbone supports episode-level extraction of the index restorative gastrointestinal procedure, the eligibility screen, the final analytic cohort, and the three feature blocks that feed later latent-state estimation. Procedure family, care environment, and 30-day leak remain separate analytic targets rather than being absorbed into a single flat case-mix summary.

that are not well described by any single average risk function? A patient with low preoperative albumin, prolonged operation, early lactate elevation, and vasopressor continuation may belong to a pathophysiologic configuration that is quite different from a patient with an ultra-low pelvic anastomosis, modest early systemic disturbance, and a technically precarious local repair. Both may leak. The observable event is shared, but the route to the event is not.

This possibility matters because a large fraction of the conversation about leak heterogeneity has been organized around the wrong unit of abstraction. Operations are treated as the natural strata, and care settings are added as contextual modifiers. In reality, procedure and setting may be acting partly as sorting mechanisms. Certain operations select into specific states of tissue perfusion, tension geometry, contamination burden, or nutritional depletion. Certain pathways then amplify or dampen the expression of those states through surveillance intensity, rescue speed, and perioperative organization. If that is true, then procedure-level averages are not merely incomplete summaries. They may also collapse clinically dissimilar subpopulations into a single number that is difficult to act on. A service caring for many low pelvic reconstructions may not simply have a higher average leak rate than a service dominated by ileocolic anastomoses. It may be caring for a distinct hidden population whose dominant vulnerability is local mechanical stress rather than broad inflammatory failure. Likewise, a rescue-intensive tertiary pathway may appear high risk in crude terms because it concentrates nutritionally depleted foregut cases, even while reducing the observed leak probability within some latent states.

Recent electronic health record studies have strengthened the idea that early structured postoperative data contains useful information for leak surveillance. Sadanandan (2024) showed that postoperative inflammatory markers, ICU admission status, surgery type, and preoperative albumin together capture meaningful signal in a proof-of-concept EHR pipeline, while also noting that aggregate model performance may conceal subgroup differences by procedure type and care environment [1]. That point is important for a reason that goes beyond predictive accuracy. If the strongest predictors are partly proxies for different hidden susceptibility states, then subgroup heterogeneity may be a reflection not only of calibration drift or case-mix imbalance, but also of unobserved population structure. A procedure family with the same average lactate and white blood cell count as another family might still carry a different latent composition because the same measured variables can arise from different physiologic configurations.

The current paper takes that idea seriously. Instead of asking only whether leak is more common in one operation or setting than another, it asks whether the observed heterogeneity can be represented through a small number of latent susceptibility phenotypes inferred from routinely collected perioperative data. The aim is not to claim that such phenotypes correspond to immutable biological entities. They are empirical constructs, estimated from covariation among nutritional markers, operative features, immediate postoperative physiology, and the leak outcome itself [2]. Their usefulness lies in whether they orga-

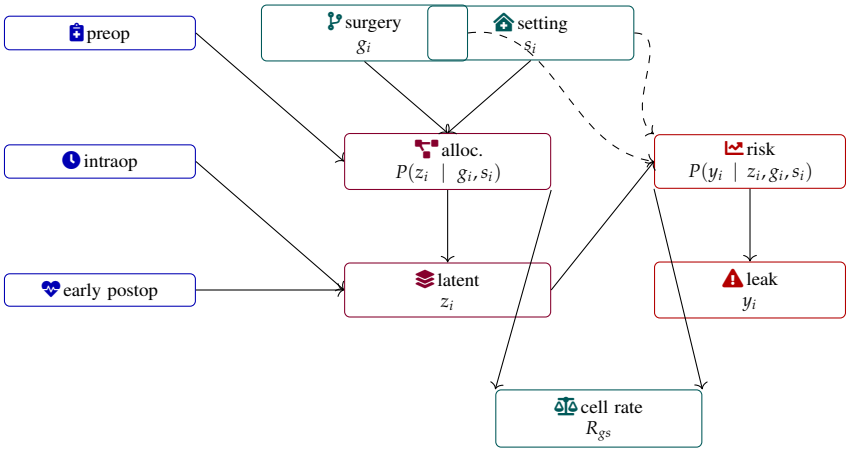

Figure 2 Joint latent phenotype formulation. Perioperative features anchor the hidden phenotype state, while surgery family and care environment modulate phenotype allocation and phenotype-conditioned leak expression. The diagram separates allocation, latent-state recovery, outcome risk, and cell-level aggregation so the model can distinguish phenotype sorting from phenotype expression.

nize the heterogeneity better than a flat collection of marginal effects. If a modest number of phenotypes explains a large share of between-cell variation across surgery families and care environments, then the phenomenon of leak heterogeneity may be more structured than the usual procedure-based summaries imply.

The study is built around two related empirical questions. The first concerns phenotype sorting. Do gastrointestinal procedures and care environments differ in how frequently they contain each latent susceptibility phenotype? A positive answer would imply that some observed variation in leak rates is attributable to hidden population composition. The second concerns phenotype expression. After conditioning on latent phenotype, does the same phenotype carry different leak probability across surgery families and care environments? A positive answer there would imply that procedures and settings do more than sort patients. They also alter the way a latent state becomes clinically consequential. The first question is about who is present in a pathway. The second is about what the pathway does to that presence.

These two dimensions of heterogeneity should not be conflated. A high-risk pathway can emerge because it concentrates phenotypes that are leak-prone everywhere, or because even shared phenotypes are expressed more severely there, or because both occur together. The distinction is clinically meaningful. If heterogeneity is mostly sorting, then the pathway's burden is driven primarily by the kinds of patients and operations it contains. If heterogeneity is mostly expression, then environment and workflow are more central. A simple overall leak rate cannot separate these explanations. Neither can a conventional fixed-effect regression that imposes additive structure on variables whose joint configuration may be more important than their individual coefficients.

A latent phenotype approach is especially attractive in this domain because the available covariates are already known to be jointly structured. Albumin, body mass index, anemia, frailty, malignancy burden, operative duration, vasopressor use, transfusion, contamination, lactate, and early inflammatory response rarely occur in isolation. They cluster into recognizable patterns that reflect tissue quality, physiologic reserve, local technical complexity, and systemic stress. Those patterns likely differ across right-sided colectomy, low pelvic reconstruction, small-bowel anastomosis after urgent obstruction, gastrectomy for malignancy, and esophagectomy after referral to specialized care. Yet the literature usually enters them as separate features in one predictive model rather than as signals of underlying states. That modeling choice is sensible when discrimination is the only aim. It is less satisfactory when the goal is to understand how heterogeneity is assembled across operation types and care environments [3].

The empirical setting here allows that question to be addressed with more structure than is usually attempted. The cohort spans multiple hospitals and includes a wide range of restorative gastrointestinal operations. Care environment is defined at the episode level in a way that captures what clinicians would recognize as substantively different pathways: protocolized elective ward recovery, routine elective inpatient management, urgent daytime nontransfer care, emergency overnight care, and rescue-intensive complex referral care. Those settings do not merely differ in location. They differ in workflow regularity, timing of escalation, baseline patient preparation, and the degree to which complications are encountered in a planned rather than disruptive manner. Crossing those settings with operation families yields a matrix in which hidden phenotypes can be mapped, rather than simply averaged out.

The analysis uses a joint latent class model that combines perioperative features with the leak outcome. The latent classes are inferred from the data rather than imposed by clinical labels, but the resulting classes are interpreted through their feature profiles and their distribution across the procedure-setting matrix. Once estimated, they permit several quantities of interest that standard models do not produce naturally. One can estimate the posterior prevalence of each phenotype in each procedure family and care environment. One can estimate phenotype-specific leak risk by operation and setting. One can then decompose overall dispersion in leak rates into a component caused by phenotype

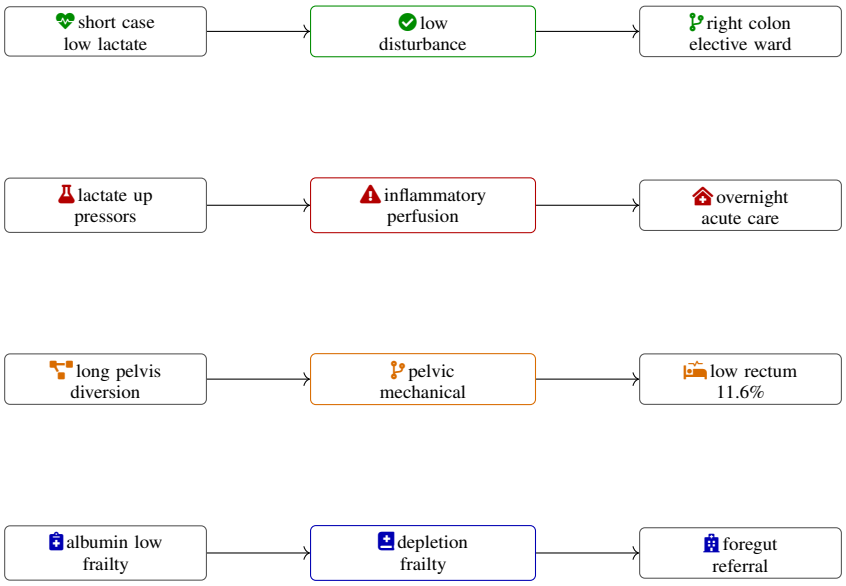

Figure 3 Recovered susceptibility phenotypes and their dominant signatures. Each phenotype is shown as a compact state node linked to a small feature summary on the left and a typical clinical concentration on the right. The panel emphasizes that the classes are empirical perioperative configurations rather than fixed procedure labels, and that their leak gradients remain materially different even before the procedure-by-setting matrix is examined in full.

sorting and a component caused by phenotype expression. That decomposition is not causal in a strict sense, but it provides a sharper vocabulary for describing why one pathway differs from another.

The objective, then, is narrower than a general theory of leak and broader than a conventional predictive exercise. The paper does not claim that latent phenotypes exhaust the biology of anastomotic failure. It asks whether they are empirically useful for organizing the heterogeneity of leak across surgery types and postoperative care environments. If they are, then the field gains a different way of looking at the problem: not only as a matter of procedures carrying different average rates, and not only as a matter of hospitals performing better or worse, but as a matter of hidden susceptibility states being unevenly sorted and differentially expressed across clinical pathways. That framing may be especially useful for pathway design, because it identifies whether a pathway differs chiefly because of who enters it or because of what happens to similar patients once they are there.

2. Cohort Assembly and Measurement Architecture

The source data came from 21 hospitals belonging to a shared surgical data network between January 2015 and December 2025. The network included tertiary referral centers, general teaching hospitals, and mixed-acuity regional hospitals using a common electronic health record backbone for operative documentation, laboratory transmission, medication administration, and patient location tracking. This did not make the cohort homogeneous. It made the cohort measurable. Common data architecture allowed features to be extracted on the same scale while preserving substantial clinical variation in procedure portfolio, perioperative organization, and postoperative recovery pathways. The present analysis used the first restorative gastrointestinal procedure within each hospitalization as the index event. This decision avoided ambiguous attribution of immediate postoperative physiology to multiple anastomoses occurring in the same admission window [4].

Eligibility required age of at least 18 years and creation of a primary gastrointestinal anastomosis involving the esophagus, stomach, small bowel, colon, or rectum. Cases involving only diversion without restored continuity were excluded because their postoperative hazard structure is fundamentally different. Additional exclusions were applied to protect interpretability of the immediate postoperative data block. Trauma laparotomy with planned staged closure, transplant surgery, palliative bypass without intended definitive restorative continuity, and admissions transferred into the network after the operation itself were removed. Procedures lacking an operative end timestamp or lacking any structured data in the first postoperative 18-hour window were also excluded. These exclusions were not intended to create an idealized cohort. They were intended to ensure that the latent phenotype model had a common temporal anchor.

After exclusions, 53,418 procedures remained. The operations were grouped into seven surgery families chosen to reflect both technical distinction and sufficient empirical support. Right-sided colectomy with ileocolic or ileotransverse reconstruction accounted for 14,226 procedures. Left or sigmoid colectomy with colocolic or colorectal continuity accounted for 11,907. Low colorectal or coloanal reconstruction after low anterior or ultra-low pelvic resection accounted for 8,735. Restorative subtotal or total colectomy accounted for 3,284. Small-bowel resection with primary continuity outside the right-sided colectomy category accounted for 8,844. Gastrectomy with gastrointestinal reconstruction accounted for 3,976. Esoph-

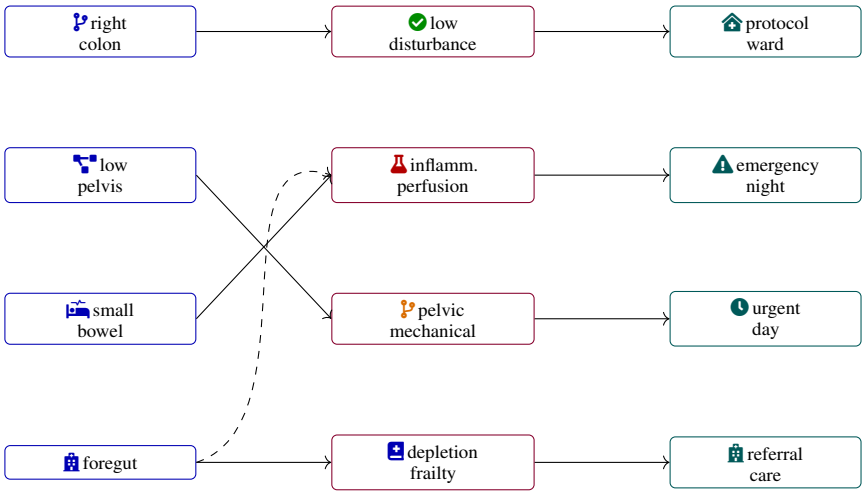

Figure 4 Phenotype sorting across operation families and care environments. The strongest redistributions are shown explicitly: right-sided colonic pathways preferentially accumulate the low-disturbance state, low pelvic reconstruction enriches the pelvic-mechanical state, emergency overnight episodes enrich the inflammatory-perfusion state, and foregut referral pathways concentrate depletion-frailty. Dashed linkage marks a secondary but still meaningful foregut route into inflammatory-perfusion.

agogastric reconstruction after esophagectomy accounted for 2,446. These frequencies created a cohort dominated numerically by colorectal and small-bowel operations, while retaining a meaningful number of foregut reconstructions whose leak biology is markedly different.

Postoperative care environment was defined at the episode level rather than as a simple hospital category. Five environments were used. The first was protocolized elective ward recovery, referring to scheduled operations managed through pathwayized perioperative programs with standardized postoperative order sets, expected ward recovery, and early mobilization. The second was routine elective inpatient care, also scheduled but lacking the same degree of explicit standardization and often involving more variable postoperative sequencing. The third was urgent daytime acute care, defined as surgery during an unplanned admission under daytime conditions, typically for obstruction, worsening symptoms, or failed nonoperative management. The fourth was emergency overnight care, defined by immediate or unscheduled operations beginning between 18:00 and 06:00 or explicitly marked as emergent. The fifth was rescue-intensive complex referral care, referring to transferred or internally escalated episodes in which the operation occurred within a pathway built around specialized services, complex disease, redo surgery, or anticipated need for higher postoperative rescue capacity [5]. These settings yielded 18,681, 13,120, 10,487, 6,744, and 4,386 procedures, respectively.

This setting definition was deliberately episode-oriented rather than purely administrative. Two patients in the same hospital could belong to different environments, and the same operation family could appear in more than one environment. That flexibility was essential because the study's aim was to examine how latent susceptibility states are distributed across real clinical pathways. A hospital-level label would have been too coarse. The pathway in which a patient is recovered, monitored, and escalated during the immediate postoperative period is more directly related to leak expression than the institution's name alone.

The primary outcome was clinically consequential anastomotic leak within 30 days of surgery. An event was counted when anatomically compatible evidence of dehiscence or leak was paired with confirmatory management or direct procedural documentation. Compatible evidence included contrast extravasation, peri-anastomotic collection explicitly interpreted as leak-related, intraoperative confirmation of defect, endoscopic confirmation, or direct surgeon documentation of anastomotic failure. Confirmatory management included reoperation, image-guided drainage, endoscopic stenting or endoluminal vacuum therapy, or leak-specific broad-spectrum antimicrobial escalation linked to an anatomically plausible source. Readmissions within 30 days were incorporated so that events recognized after discharge were not missed. This composite was designed to align with clinically meaningful failure rather than coding behavior alone. In the final cohort, 3,446 leaks occurred, corresponding to an overall incidence of 6.5%.

Observed incidence varied sharply by operation family. Right-sided colectomy had 468 leaks, or 3.3%. Left or sigmoid colectomy had 566 leaks, or 4.8%. Low colorectal or coloanal reconstruction had 1,018 leaks, or 11.7%. Restorative subtotal or total colectomy had 225 leaks, or 6.9%. Small-bowel anastomosis had 416 leaks, or 4.7%. Gastrectomy had 356 leaks, or 9.0%. Esophagogastric reconstruction had 397 leaks, or 16.2%. The care environments also differed in crude incidence. Protocolized elective ward recovery had 863 leaks, or 4.6%. Routine elective inpatient care had 760 leaks, or 5.8%. Urgent daytime acute care had 802 leaks, or 7.6% [6]. Emergency overnight care had 653 leaks, or 9.7%. Rescue-intensive complex referral care had 368 leaks, or 8.4%. These figures provide context but not explanation. Their interpretation depends on whether pathways differ by visible severity alone or by hidden phenotype composition as well.

The feature space used for latent phenotype estimation was organized into preoperative, intraoperative, and early postop-

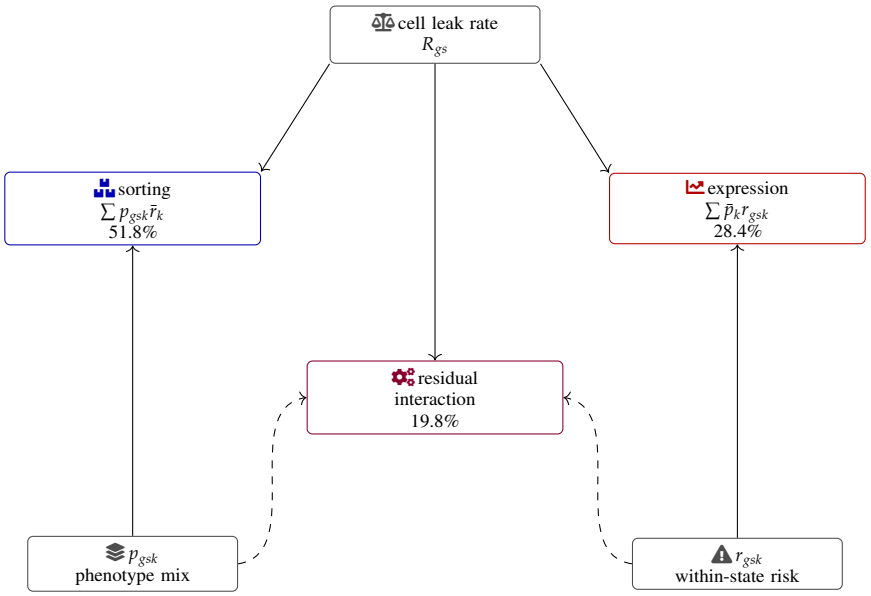

Figure 5 Variance decomposition of the procedure-by-setting matrix. The total cell-specific leak probability is partitioned into a dominant sorting contribution, a smaller but still substantial phenotype-expression contribution, and a residual interaction remainder. The lower nodes show the two primitives that generate the decomposition: posterior phenotype composition and phenotype-specific leak risk within the same cell.

erative blocks. Preoperative variables included age, sex, body mass index, smoking status, diabetes, chronic kidney disease, chronic lung disease, cardiovascular disease, immunosuppressive therapy, malignancy status, prior abdominal surgery, and a deficit-based frailty index derived from mobility limitation, dependence, cognitive impairment, and polypharmacy. Laboratory variables included albumin, hemoglobin, creatinine, bilirubin, sodium, white blood cell count, and C-reactive protein where available. Intraoperative variables included operative duration, minimally invasive versus open approach, conversion, contamination class, diversion status, estimated blood loss, transfusion, vasopressor burden, and case start time. Early postoperative variables, measured within 18 hours after closure, included peak heart rate, peak temperature, first postoperative lactate, white blood cell count, hemoglobin change, creatinine change, vasopressor continuation, ventilatory support at recovery arrival, immediate recovery location, and laboratory density.

The decision to include early postoperative variables in the latent phenotype block requires justification. A purely preoperative phenotype model would have the appeal of temporal purity, but it would miss important physiologic states that become visible only once the operation has challenged tissue perfusion, inflammatory equilibrium, and organ reserve. The present objective was not preoperative counseling alone. It was empirical organization of leak heterogeneity in the period when postoperative pathway design still matters. Including the earliest structured postoperative data therefore makes sense, provided the window remains early enough that formal leak recognition is uncommon. In this cohort, the great majority of leaks were recognized after postoperative day 3, while the phenotype window ended within the first 18 hours. The phenotype model thus used an early surveillance horizon rather than a post-diagnosis horizon.

Continuous laboratory variables were log-transformed where strongly skewed, then standardized within the derivation folds used for estimation. Operative duration and blood loss were winsorized at the 0.5th and 99.5th percentiles to reduce the influence of probable data-entry errors while preserving genuine extremes. Missingness was addressed by multiple imputation for continuous variables and explicit unknown categories for selected categorical fields, but the latent model also incorporated missingness indicators for key features whose absence was potentially informative. Albumin and body mass index were more frequently missing in urgent and emergency pathways, whereas C-reactive protein was sparse in protocolized elective pathways. Early postoperative lactate was nearly universal in rescue-intensive care and much less frequent after lower-acuity elective pathways. These absences are not random, and they can themselves signal pathway intensity or baseline concern. Modeling them separately made the latent classes more faithful to the actual clinical data environment.

Because the later analysis asks how phenotypes are distributed across operation families and care environments, it was important not to define the phenotypes using those labels directly [7]. The latent classes were therefore estimated from the perioperative covariates and the leak outcome, while procedure family and care environment entered only in the phenotype allocation and outcome components after the core measurement model was specified. This strategy prevented the classes from degenerating into mere relabelings of operation type. A latent class is only useful if it captures something more than the obvious fact that esophagectomy differs from right colectomy. The empirical test, then, is whether the classes retain meaning across families and settings and whether their distributions explain variation that is otherwise diffuse.

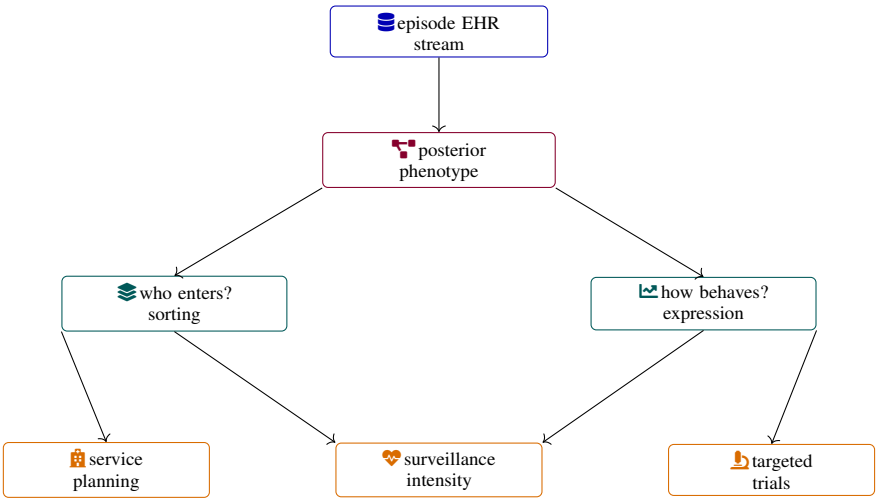

Figure 6 Phenotype-aware interpretation pathway. Once posterior latent-state membership is estimated from the structured perioperative record, two questions can be separated cleanly: which hidden states are entering a pathway and how those states are expressed after entry. That separation supports service-level planning, more interpretable monitoring intensity, and future intervention studies aimed at shifting phenotype prevalence or blunting phenotype-specific leak expression.

3. Latent Phenotype Formulation

Let $i = 1, \ldots, N$ index patients. Each patient is assumed to belong to one unobserved leak-susceptibility phenotype $z_i \in \{1, \ldots, K\}$. The latent phenotype is not an externally given biological label; it is an empirical state inferred from the joint distribution of perioperative variables and the leak outcome. The observed feature vector for patient i is partitioned into continuous variables $x_i^{(c)}$, categorical variables $x_i^{(d)}$, and missingness indicators m_i. The surgery family is denoted g_i and the care environment s_i. The leak indicator is $y_i \in \{0, 1\}$. The modeling strategy has three connected parts: a phenotype allocation model, a phenotype-specific measurement model for the observed features, and a phenotype-conditioned outcome model for leak.

Phenotype allocation was allowed to vary with surgery family and care environment because one of the central questions concerns whether pathways selectively accumulate different hidden states. The prior probability of phenotype membership was written as

$$P(z_i = k \mid g_i, s_i) = \frac{\exp\left(\alpha_k + a_{g_i k} + b_{s_i k} + c_{g_i s_i k}\right)}{\sum_{\ell=1}^{K} \exp\left(\alpha_\ell + a_{g_i \ell} + b_{s_i \ell} + c_{g_i s_i \ell}\right)} \tag{1}$$

where α_k is the baseline logit weight for phenotype k, $a_{g_i k}$ captures how surgery family g_i shifts the relative prevalence of phenotype k, $b_{s_i k}$ captures how care environment s_i shifts that prevalence, and $c_{g_i s_i k}$ captures residual interaction. The interaction term was regularized strongly so that phenotype sorting could be expressed without fitting unstable cell-specific idiosyncrasies. This allocation model does not yet make any statement about leak risk. It simply governs who is likely to occupy which hidden state.

The continuous feature block was modeled using phenotype-specific Gaussian latent factor structures, rather than independent univariate normals, because perioperative physiology is strongly correlated. For continuous features of patient i under phenotype k,

$$x_i^{(c)} \mid z_i = k, h_i \sim \mathcal{N}(\mu_k + \Lambda_k h_i, \Psi_k) \tag{2}$$

where h_i is a low-dimensional within-phenotype factor score, μ_k is the phenotype-specific mean vector, Λ_k is the loading matrix, and Ψ_k is a diagonal residual covariance matrix. This structure allows, for example, albumin, body mass index, anemia, and frailty to co-move within a depletion-oriented phenotype without requiring a separate covariance parameter for every pair. It also allows early lactate, vasopressor continuation, creatinine rise, and white blood cell response to cluster within an inflammatory-perfusion phenotype. The within-phenotype factors are not interpreted clinically on their own; they serve to prevent the latent classes from being forced to explain all continuous correlation through extra classes.

The categorical feature block was modeled with phenotype-specific multinomial distributions. For a categorical variable j with category set $\mathcal{C}_j$,

$$P\left(x_{ij}^{(d)} = c \mid z_i = k\right) = \pi_{jkc}, \quad c \in \mathcal{C}_j \tag{3}$$

with Dirichlet regularization on the π_{jkc} vectors to prevent collapse in sparse categories. Missingness indicators for key variables were included in this categorical block so that informative absence could contribute to phenotype identification. For example, frequent absence of albumin in emergency surgery can co-occur with specific physiologic configurations, while dense lactate sampling may indicate a pathway whose patients are both sicker and more intensively observed.

Leak risk was then modeled conditionally on phenotype, surgery family, care environment, and a small residual covariate set not intended to define the latent state itself. The outcome model was

$$\operatorname{logit} P(y_i = 1 \mid z_i = k, g_i, s_i, r_i) = \beta_0 + \xi_k + u_{g_i} + v_{s_i} + \delta_{kg_i} + \eta_{ks_i} + \tau_{g_i s_i} + r_i^\top \beta \tag{4}$$

where ξ_k is the main effect of phenotype k, u_{g_i} and v_{s_i} are surgery-family and care-environment effects, δ_{kg_i} and η_{ks_i} are phenotype-by-family and phenotype-by-setting interaction terms, and $\tau_{g_i s_i}$ is a residual family-by-setting interaction shared across phenotypes. The vector r_i contained only a limited set of residual variables not already captured adequately in the phenotype model, mainly calendar year and hospital-level volume terms included to absorb secular change and broad center effects [8]. This outcome structure allowed the same phenotype to carry different leak probabilities in different operative and pathway contexts. That is the formal representation of phenotype expression.

The complete-data log-likelihood for one patient took the form

$$\begin{aligned}\ell_i = \sum_{k=1}^{K} \mathbf{1}(z_i = k) \Big[& \log P(z_i = k \mid g_i, s_i) \\ & + \log f_c(x_i^{(c)} \mid z_i = k) + \log f_d(x_i^{(d)}, m_i \mid z_i = k) \\ & + \log P(y_i \mid z_i = k, g_i, s_i, r_i) \Big]\end{aligned} \tag{5}$$

and the observed-data likelihood was obtained by summing over the latent classes. The model was fitted by variational expectation-maximization with coordinate ascent in the E-step and blockwise penalized optimization in the M-step. Convergence was assessed by the evidence lower bound and by stability of the posterior class prevalence vector. Random starts were used because finite mixtures can settle into local optima, especially when both the feature and outcome blocks influence class allocation.

The number of phenotypes K was not fixed a priori. Models with $K = 2$ through $K = 6$ were fitted. Selection relied on a combination of integrated completed likelihood, held-out predictive log-likelihood, posterior class separation, and clinical interpretability. The interpretability criterion was not a loose afterthought. A six-class solution with two tiny nearly redundant classes would not improve understanding even if it slightly improved fit. In the final analysis, the four-class solution represented the best balance among empirical support, stability, and substantive clarity.

To avoid overfitting in the phenotype allocation and outcome interaction terms, a hierarchical shrinkage prior was imposed on the surgery-family, care-environment, and interaction coefficients. In penalized likelihood form, the objective function was

$$\begin{aligned}\mathcal{J} = & -\sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} \exp(\ell_{ik}) \right) \\ & + \lambda_1 \sum_{g,k} a_{gk}^2 + \lambda_2 \sum_{s,k} b_{sk}^2 \\ & + \lambda_3 \sum_{g,s,k} c_{gsk}^2 + \lambda_4 \sum_{k,g} \delta_{kg}^2 \\ & + \lambda_5 \sum_{k,s} \eta_{ks}^2 + \lambda_6 \sum_{g,s} \tau_{gs}^2\end{aligned} \tag{6}$$

with tuning selected by nested cross-validation at the hospital level. Penalizing the interaction terms strongly was essential because the paper's main purpose was not to fit an unconstrained saturated table. It was to estimate a stable hidden-state structure that could be interpreted across the procedure-setting matrix.

Several derived quantities from this model are central to the empirical analysis. The posterior phenotype prevalence in a surgery-family and care-environment cell is

$$p_{gsk} = P(z = k \mid g, s) \tag{7}$$

which captures phenotype sorting. The phenotype-specific leak probability within the same cell is

$$r_{gsk} = P(y = 1 \mid z = k, g, s) \tag{8}$$

which captures phenotype expression. The overall cell-specific leak probability then follows by mixture aggregation: [9]

$$R_{gs} = \sum_{k=1}^{K} p_{gsk} r_{gsk} \tag{9}$$

The observed heterogeneity across the $G \times S$ matrix can therefore arise from changes in p_{gsk}, from changes in r_{gsk}, or from both simultaneously.

To quantify the relative importance of sorting and expression, the between-cell variance of R_{gs} was decomposed around pooled phenotype prevalence $\bar{p}_k$ and pooled phenotype-specific risk $\bar{r}_k$. Define the sorting-only surrogate

$$R_{gs}^{\text{sort}} = \sum_{k=1}^{K} p_{gsk} \bar{r}_k \tag{10}$$

and the expression-only surrogate

$$R_{gs}^{\text{expr}} = \sum_{k=1}^{K} \bar{p}_k r_{gsk} \tag{11}$$

Then, letting $V(\cdot)$ denote weighted variance across cells,

$$
\begin{aligned}
V_{\text{tot}} &= V(R_{gs}) \\
V_{\text{sort}} &= V(R_{gs}^{\text{sort}}) \\
V_{\text{expr}} &= V(R_{gs}^{\text{expr}}) \\
V_{\text{int}} &= V_{\text{tot}} - V_{\text{sort}} - V_{\text{expr}}
\end{aligned} \tag{12}
$$

The resulting shares are descriptive rather than causal. A high sorting share means that procedure-setting cells differ substantially in which phenotypes they contain. A high expression share means that even with identical phenotype composition, the same phenotype would behave differently across cells. A positive interaction remainder indicates that sorting and expression reinforce one another in specific regions of the matrix.

In addition to this variance decomposition, the study examined posterior entropy of class assignment. For patient i, the entropy was

$$
H_i = -\sum_{k=1}^{K} \gamma_{ik} \log(\gamma_{ik}) \tag{13}
$$

where $\gamma_{ik} = P(z_i = k \mid \text{data})$ is the posterior membership probability. Low entropy indicates a sharply identified phenotype assignment. High entropy indicates a patient whose feature pattern sits near the boundary between phenotypes. Reporting entropy is important because latent classes can appear clinically persuasive while remaining statistically diffuse. The empirical value of the phenotype structure depends not only on class prevalence and risk gradients, but also on whether individual allocations are meaningfully separated.

4. Estimation Strategy and Validation

The latent phenotype model was fitted on the full cohort with hospital-blocked internal validation and separate sensitivity checks for class stability. Each iteration of the estimation procedure used five random starts for the two-class model, increasing to twenty random starts for the six-class model, because class proliferation increases local-optimum risk. The evidence lower bound was monitored until relative change fell below 10^{-6}, and the best solution among converged runs was retained for each K. Hospital-blocked cross-validation was used for hyperparameter tuning so that shrinkage and covariance complexity were selected on a structure closer to cross-site deployment than to random record mixing. Although the present paper is not a transport study, hospital blocking protects against overinterpreting classes that only exist because of idiosyncratic documentation patterns at a single site.

Model choice among candidate class numbers relied on four criteria considered jointly. The first was out-of-fold predictive log-likelihood for the leak outcome, because the latent structure should improve empirical organization of the event rather than merely fit the feature distribution. The second was integrated completed likelihood, which penalizes class proliferation more strongly than ordinary information criteria when posterior assignment becomes diffuse. The third was average posterior entropy, because an additional class that lowers fit only marginally while raising ambiguity is usually not worth retaining. The fourth was phenotype distinctness under posterior predictive checks. Specifically, the derived classes were required to differ in coherent feature bundles rather than only in one extreme variable or one rare category. This last criterion was important because a clinically useless latent solution can still appear statistically competitive if it separates trivial corners of the data cloud [10].

Missing data were handled inside the model rather than through a fully separate imputation step for the primary analysis. Continuous features with missing values were integrated under the phenotype-specific Gaussian block using the observed subvector likelihood, which is a standard advantage of multivariate normal structures. Categorical missingness was represented directly through the indicator block. A secondary analysis using external multiple imputation before model fitting yielded nearly identical class prevalences and risk gradients, which supported the more integrated primary strategy. The integrated approach was preferred because it allows missingness patterns to inform phenotype structure instead of being erased by preprocessing.

Posterior uncertainty for cell-specific phenotype prevalence and phenotype-conditioned leak probability was obtained by parametric bootstrap from the fitted model. For each of 800 draws, parameters were simulated from the asymptotic covariance approximation of the penalized estimator, the latent allocation probabilities were recomputed, and all derived quantities of interest were recalculated. This produced uncertainty bands for phenotype composition in each procedure-setting cell, for phenotype-specific risk surfaces, and for the sorting-expression decomposition. Because finite mixture models can exhibit label switching across resamples, a relabeling step was applied using pairwise Kullback-Leibler distance between phenotype-specific feature profiles to align bootstrap draws with the reference solution. This prevented the uncertainty intervals from widening artificially due to arbitrary permutation of class labels.

The phenotype interpretation itself was based on standardized contrasts rather than raw means alone. For continuous variables, each phenotype’s mean was expressed as a deviation from the overall cohort mean in standard deviation units. For categorical variables, the phenotype-specific probability was compared with the pooled probability using log-odds ratios. This common scale made it easier to distinguish whether a phenotype was defined mainly by nutritional depletion, inflammatory perturbation, local technical pattern, or broad physiologic quiescence. No single variable was allowed to determine phenotype naming. Names were assigned only after reviewing the entire bundle of covariate deviations and leak behavior.

Two external validity checks were embedded within the internal analysis. First, phenotype prevalence was estimated separately by hospital, and a hospital-level random-intercept correction was applied to the allocation model to determine whether any phenotype vanished once institutional differences were absorbed. No phenotype disappeared. Second, the class solution was tested for temporal stability by fitting the model separately to the early and late halves of the study period after alignment of labels. The resulting phenotypes retained similar feature profiles and leak gradients, though the prevalence of the

Table 1 Cohort Overview and Outcome Incidence :contentReference[oaicite:0]index=0

Category	Count	Percentage	Notes
Total procedures	53,418	100%	Multicenter cohort
Leak events	3,446	6.5%	30-day incidence
Hospitals	21	–	settings
Elective pathways	31,801	59.5%	Combined elective
Urgent/emergency	17,231	32.2%	Acute care
Referral care	4,386	8.2%	Complex cases

inflammatory-perfusion phenotype declined modestly over time in elective colorectal pathways. This temporal shift is consistent with evolving perioperative care rather than with model instability, and the persistence of the class structure argues that the latent states are not merely artifacts of one historical period.

The variance decomposition between sorting and expression was evaluated both with and without weighting by cell volume. Unweighted variance answers whether the matrix is heterogeneous across cells regardless of how many operations each cell contains. Weighted variance answers how much the overall burden of heterogeneity experienced by the cohort is attributable to sorting versus expression [11]. The main text reports the weighted version because the paper is concerned with observed clinical pathways, not only abstract cell geometry. The unweighted version is noted in the results because it provides a useful comparison. In some settings, rare high-risk cells can influence the unweighted decomposition disproportionately even when they contribute less to the overall number of events.

A final methodological issue concerns the relationship between latent class modeling and ordinary regression. One might ask whether the same findings could be obtained by a richly interacted logistic model using the observed variables alone. To address that question, a comparator generalized additive logistic model was fitted with surgery-family, care-environment, and selected interactions among albumin, lactate, operative duration, contamination, and urgency. The comparator produced reasonable fit but no direct notion of hidden-state composition. More importantly, when the procedure-setting matrix was summarized through this regression alone, the resulting cell-level heterogeneity lacked the compact organization achieved by the latent phenotype model. This does not prove that latent classes are true biological entities. It suggests that they are empirically efficient summaries of high-dimensional structure that would otherwise be scattered across dozens of nonlinear interaction surfaces.

5. Results

Model comparison favored a four-phenotype solution. The two-class model collapsed distinct forms of vulnerability into one broad high-risk group and left substantial residual structure in the perioperative feature matrix. The three-class model improved outcome fit but still merged low pelvic technical stress with systemic inflammatory derangement. The five-class and six-class models split those patterns further, yet the added classes were small, poorly separated, and unstable across random starts. The four-class solution achieved the best compromise among integrated completed likelihood, held-out predictive log-likelihood, posterior entropy, and phenotype reproducibility across hospitals and calendar time. Average posterior entropy in the four-class model was 0.39 nats, corresponding to relatively sharp assignments for most patients. Seventy-two percent of patients had a posterior probability above 0.80 for their most likely phenotype, and only 11% had a maximum posterior probability below 0.60.

The first phenotype, comprising 38.9% of the cohort, was characterized by relatively preserved preoperative albumin, modest frailty burden, shorter operative duration, low transfusion use, minimal vasopressor continuation, lower early lactate, and lower postoperative laboratory density. This phenotype also had the highest probability of minimally invasive approach where applicable and the lowest burden of contamination. Its leak incidence was 2.7%. Because its feature profile suggested comparatively undisturbed perioperative recovery rather than the absence of all risk, it is described here as the low-disturbance phenotype. Patients in this phenotype were not uniformly healthy, but they showed the narrowest departure from routine perioperative physiology.

The second phenotype, accounting for 22.7% of the cohort, showed a very different profile. It was marked by elevated early postoperative lactate, higher white blood cell count, more frequent vasopressor continuation, greater creatinine rise, higher contamination burden, more open surgery, more transfusion, and longer operations [12]. Operative urgency was common but not universal. This phenotype also displayed the highest rate of immediate higher-acuity recovery and the densest early laboratory monitoring. Its leak incidence was 10.9%. The combination of systemic inflammatory response, perfusion strain, and perioperative physiologic stress makes inflammatory-perfusion phenotype the most suitable descriptive label. Importantly, this class was not defined by emergency surgery alone; a substantial fraction of major elective foregut and pelvic operations also occupied it.

The third phenotype, representing 17.1% of the cohort, was notable for concentration in low pelvic procedures, high diversion frequency, prolonged operative time, moderate blood loss,

Table 2 Leak Incidence by Surgery Family :contentReference[oaicite:1]index=1

Procedure	Cases	Leaks	Rate
Right colectomy	14,226	468	3.3%
Left colectomy	11,907	566	4.8%
Low colorectal	8,735	1,018	11.7%
Subtotal colectomy	3,284	225	6.9%
Small bowel	8,844	416	4.7%
Gastrectomy	3,976	356	9.0%
Esophagogastric	2,446	397	16.2%

Table 3 Care Environment Distribution :contentReference[oaicite:2]index=2

Environment	Cases	Leaks	Rate
Protocolized elective	18,681	863	4.6%
Routine elective	13,120	760	5.8%
Urgent daytime	10,487	802	7.6%
Emergency overnight	6,744	653	9.7%
Referral care	4,386	368	8.4%

postoperative hemoglobin decline, and a weaker systemic inflammatory signature than the second phenotype. Early lactate was often only mildly abnormal, and creatinine rise was less prominent than in the inflammatory-perfusion class. Male sex, lower anastomotic level proxies, and long pelvis-associated operations were enriched. Leak incidence in this phenotype was 11.6%, slightly higher than in the inflammatory-perfusion phenotype despite less dramatic early systemic disturbance. This profile is most coherently described as the pelvic-mechanical phenotype. It suggests a state in which local tissue geometry and technical strain dominate more than overt global physiologic collapse in the earliest postoperative window.

The fourth phenotype, accounting for 21.3% of the cohort, showed low albumin, anemia, lower body mass index or documented weight loss, higher frailty burden, malignancy concentration, and prolonged preoperative hospitalization or referral complexity. Operations were often long, but the early postoperative physiology was less acutely explosive than in the inflammatory-perfusion class. The leak incidence in this phenotype was 7.8%. Because its dominant signature was baseline reserve depletion rather than immediate inflammatory surge or localized pelvic mechanical pattern, it is referred to as the depletion-frailty phenotype. This group included many foregut and complex referral cases, as well as a smaller number of nutritionally depleted colorectal and small-bowel cases.

The distribution of these phenotypes across operation families was highly uneven. Right-sided colectomy was predominantly low-disturbance, with posterior prevalence 55.8% for phenotype one, 17.4% for phenotype two, 7.6% for phenotype three, and 19.2% for phenotype four. Left or sigmoid colectomy also remained mostly low-disturbance, though less strongly so, at 46.1%, 19.2%, 14.3%, and 20.4%, respectively. Low colorectal or coloanal reconstruction displayed the clearest enrichment for the pelvic-mechanical phenotype: 23.0% low-disturbance, 17.8% inflammatory-perfusion, 39.7% pelvic-mechanical, and 19.5% depletion-frailty. Small-bowel anastomosis split more evenly between low-disturbance and inflammatory-perfusion states, with 35.9%, 31.5%, 9.8%, and 22.8%. Gastrectomy showed a mixed pattern with substantial depletion-frailty presence at 32.4% and inflammatory-perfusion at 25.8%. Esophagogastric reconstruction was dominated by depletion-frailty and inflammatory-perfusion at 43.6% and 31.7%, respectively, with only 18.8% low-disturbance and 5.9% pelvic-mechanical. These profiles demonstrate that surgery family does not merely scale a common baseline risk. It redistributes patients into distinct hidden states [13].

Care environments also differed markedly in phenotype composition. Protocolized elective ward recovery contained 48.6% low-disturbance phenotype, 14.8% inflammatory-perfusion, 19.7% pelvic-mechanical, and 16.9% depletion-frailty. Routine elective inpatient care had 39.9%, 18.0%, 21.5%, and 20.6%. Urgent daytime acute care had 28.4%, 31.2%, 14.0%, and 26.4%. Emergency overnight care had only 16.7% low-disturbance phenotype, with 42.3% inflammatory-perfusion, 11.6% pelvic-mechanical, and 29.4% depletion-frailty. Rescue-intensive complex referral care had 20.1% low-disturbance, 24.8% inflammatory-perfusion, 12.7% pelvic-mechanical, and 42.4% depletion-frailty. In other words, emergency overnight

Table 4 Latent Phenotype Prevalence and Risk :contentReference[oaicite:3]index=3

Phenotype	Prevalence	Leak Rate	Key Trait
Low-disturbance	38.9%	2.7%	Stable physiology
Inflammatory-perfusion	22.7%	10.9%	Systemic stress
Pelvic-mechanical	17.1%	11.6%	Local strain
Depletion-frailty	21.3%	7.8%	Low reserve

Table 5 Phenotype Distribution in Selected Procedures :contentReference[oaicite:4]index=4

Procedure	Low-dist	Infl-perf	Pelvic-mech
Right colectomy	55.8%	17.4%	7.6%
Left colectomy	46.1%	19.2%	14.3%
Low colorectal	23.0%	17.8%	39.7%
Small bowel	35.9%	31.5%	9.8%
Gastrectomy	28.1%	25.8%	13.7%
Esophagogastric	18.8%	31.7%	5.9%

care was the most inflammatory-perfusion-rich environment, while rescue-intensive complex referral care was the most depletion-frailty-rich environment. This distinction is essential because crude leak rates alone would not reveal which hidden state predominates in which pathway.

The phenotype-conditioned leak probabilities showed that sorting was only part of the story. The same phenotype did not have identical risk across operation families or care environments. For the low-disturbance phenotype, leak probability remained low everywhere, but it was not flat. In right-sided colectomy under protocolized elective ward recovery it was 1.8%, while in esophagogastric reconstruction under rescue-intensive complex referral care it rose to 5.2%. For the inflammatory-perfusion phenotype, risk ranged from 7.9% in right-sided colectomy under routine elective inpatient care to 13.4% in small-bowel anastomosis under emergency overnight care and 14.1% in esophagogastric reconstruction under rescue-intensive complex referral care. For the pelvic-mechanical phenotype, the largest gradients were in low colorectal or coloanal reconstruction, where risk was 9.4% in protocolized elective ward recovery, 10.7% in routine elective care, 13.2% in urgent daytime acute care, 14.8% in emergency overnight care, and 11.1% in rescue-intensive complex referral care. For the depletion-frailty phenotype, the foregut families were especially vulnerable: gastrectomy ranged from 8.5% in routine elective care to 11.3% in rescue-intensive complex referral care, while esophagogastric reconstruction ranged from 12.1% to 14.5% across the same settings. These differences show phenotype expression directly.

The overall procedure-by-setting matrix of leak probabilities inherited structure from both composition and expression. The lowest observed cell-specific leak probability was 2.4% for right-sided colectomy under protocolized elective ward recovery. The highest was 15.6% for esophagogastric reconstruction under rescue-intensive complex referral care, followed closely by 14.9% for low colorectal or coloanal reconstruction under emergency overnight care. Gastrectomy under rescue-intensive complex referral care had 10.7%, and small-bowel anastomosis under emergency overnight care had 7.2%. Some cells with moderate crude risk were revealed to be heterogeneous mixtures of low-disturbance and high-risk phenotypes, especially left-sided colectomy in urgent daytime acute care and small-bowel anastomosis in routine elective inpatient care. These cells mattered because their average leak probabilities obscured a wide within-cell spread in posterior phenotype allocation.

A central aim of the study was to quantify how much of matrix dispersion was due to phenotype sorting rather than phenotype expression. In the volume-weighted variance decomposition, 51.8% of between-cell variance in leak probability was attributable to sorting, meaning that procedures and care environments differed strongly in which hidden states they contained [14]. Expression accounted for 28.4%, indicating that even if phenotype composition were held constant, the same phenotype would still behave differently across the matrix. The residual interaction share was 19.8%. In the unweighted decomposition, the sorting share fell slightly to 47.3% and the expression share rose to 31.6%, because rarer extreme cells contributed more prominently to matrix geometry when volume was ignored. The broad interpretation was stable under both approaches: most, but not all, of the heterogeneity came from differential distribution of hidden states.

The sorting component varied by operation family. Right-sided colectomy and left or sigmoid colectomy derived most of their cross-setting differences from sorting into the low-disturbance versus depletion-frailty phenotypes. Low colorec-

Table 6 Phenotype Distribution Across Care Settings :contentReference[oaicite:5]index=5

Setting	Low-dist	Infl-perf	Depletion
Protocolized	48.6%	14.8%	16.9%
Routine elective	39.9%	18.0%	20.6%
Urgent daytime	28.4%	31.2%	26.4%
Overnight	16.7%	42.3%	29.4%
Referral	20.1%	24.8%	42.4%

Table 7 Phenotype-Specific Leak Variation (Examples) :contentReference[oaicite:6]index=6

Phenotype	Low Risk	Mid Risk	High Risk
Low-disturbance	1.8%	3.5%	5.2%
Inflammatory-perf	7.9%	10.5%	14.1%
Pelvic-mechanical	9.4%	11.2%	14.8%
Depletion-frailty	6.8%	9.2%	14.5%

tal or coloanal reconstruction differed across settings mostly through a combination of sorting into the pelvic-mechanical phenotype and expression of that phenotype. In other words, some settings did not merely receive more pelvic-mechanical cases; the same pelvic-mechanical state also leaked more often in those settings. Small-bowel anastomosis differed across settings predominantly through sorting into the inflammatory-perfusion phenotype, especially in urgent and overnight pathways. Gastrectomy and esophagogastric reconstruction exhibited strong depletion-frailty sorting, but expression gradients remained meaningful, especially in rescue-intensive complex referral care where the same depletion-frailty state had higher leak probability than in routine elective pathways.

The posterior entropy results help clarify whether these phenotype distinctions were statistically meaningful at the patient level. Mean entropy was lowest in right-sided colectomy under protocolized elective pathways and in esophagogastric reconstruction under rescue-intensive complex referral care. Those were settings where the data strongly favored low-disturbance and depletion-frailty allocations, respectively. Entropy was highest in left-sided colectomy and small-bowel anastomosis under routine elective or urgent daytime care, where patients often sat between low-disturbance and inflammatory-perfusion states. This pattern is clinically plausible. Some pathways create very characteristic perioperative signatures, while others contain more blended or transitional profiles. High entropy does not invalidate the classes. It identifies where phenotype boundaries are softer and where one should be more cautious about overinterpreting a single patient's modal class.

The four phenotypes also differed in downstream non-outcome descriptors that were not used directly to define the primary endpoint. Median postoperative length of stay among non-leak cases was 5 days in the low-disturbance phenotype, 8 days in the inflammatory-perfusion phenotype, 7 days in the pelvic-mechanical phenotype, and 9 days in the depletion-frailty phenotype. Thirty-day readmission unrelated to leak was highest in depletion-frailty and inflammatory-perfusion states. Early imaging utilization within postoperative day 3 was most common in inflammatory-perfusion and rescue-intensive foregut cases. These external patterns are not proof of biological truth, but they support the idea that the latent classes are capturing clinically recognizable postoperative states rather than arbitrary statistical partitions.

A particularly informative comparison involved low colorectal or coloanal reconstruction across care environments. Its crude leak rates were 9.2% in protocolized elective ward recovery, 10.4% in routine elective inpatient care, 12.1% in urgent daytime acute care, 14.9% in emergency overnight care, and 11.6% in rescue-intensive complex referral care [15]. The phenotype distribution explained much of this gradient. Protocolized elective pathways had 27.6% low-disturbance, 13.9% inflammatory-perfusion, 41.1% pelvic-mechanical, and 17.4% depletion-frailty. Emergency overnight pathways had only 10.2% low-disturbance, 29.4% inflammatory-perfusion, 35.0% pelvic-mechanical, and 25.4% depletion-frailty. Yet even after accounting for the shift in composition, the pelvic-mechanical phenotype carried greater leak probability in the overnight pathway than in the protocolized pathway. That is a direct example of sorting and expression operating together.

Esophagogastric reconstruction displayed a different pattern. Rescue-intensive complex referral care had the highest crude incidence, but much of that was because 51.3% of those cases belonged to the depletion-frailty phenotype and 28.6% to the inflammatory-perfusion phenotype. In routine elective inpatient care, the same operation family still had many depletion-frailty cases, but the low-disturbance share was larger and the inflammatory-perfusion share smaller. Expression differences were present, though narrower than in low pelvic surgery. The

Table 8 Variance Decomposition of Leak Heterogeneity :contentReference[oaicite:7]index=7

Component	Share	Interpretation	Impact
Sorting	51.8%	Composition effect	Dominant
Expression	28.4%	Within-phenotype change	Moderate
Interaction	19.8%	Combined effects	Residual
Unweighted sorting	47.3%	Cell-based view	Slight drop
Unweighted expression	31.6%	Cell-based view	Slight rise

Table 9 Posterior Assignment Quality :contentReference[oaicite:8]index=8

Metric	Value	Meaning	Note
Mean entropy	0.39	Good separation	Low ambiguity
$>$0.80 prob	72%	Strong assignment	Majority
$<$0.60 prob	11%	Uncertain cases	Minority
Stable classes	Yes	Across hospitals	Robust
Temporal stability	Yes	Across years	Consistent

depletion-frailty phenotype under esophagogastric reconstruction leaked at 14.5% in rescue-intensive complex referral care and 12.1% in routine elective care. This suggests that foregut heterogeneity is driven more by who enters the pathway than by strong divergence in phenotype expression, although the latter is not absent.

Small-bowel anastomosis showed yet another structure. Its leak heterogeneity across settings was driven predominantly by the inflammatory-perfusion phenotype. Emergency overnight small-bowel cases had 46.1% posterior prevalence of that phenotype, compared with 24.0% in routine elective inpatient care. The phenotype-specific leak probability for inflammatory-perfusion small-bowel cases rose from 9.6% in routine elective care to 13.4% in emergency overnight care. Because the pelvic-mechanical phenotype was rare in small-bowel surgery, a large portion of the setting gradient could be understood without invoking local geometric vulnerability. This finding supports the clinical impression that contamination, global physiologic instability, and operative urgency are more central in small-bowel leak than deep local mechanical anatomy.

When the same analysis was restricted to patients whose leak required invasive source control, the broad class structure remained intact. The absolute leak probabilities fell, but the ordering of phenotype risk remained the same: low-disturbance lowest, depletion-frailty intermediate, inflammatory-perfusion high, and pelvic-mechanical highest in low colorectal surgery. The sorting share of between-cell variance rose slightly to 54.2%, while the expression share declined to 25.1%. This suggests that severe leak episodes are even more strongly tied to which hidden states accumulate in a pathway, though environment still modifies phenotype expression. The overnight low colorectal pathway and rescue-intensive foregut pathway remained the dominant high-risk regions under this stricter endpoint.

The comparator generalized additive logistic model, although reasonably well calibrated overall, did not produce as coherent an organization of the procedure-setting matrix. It assigned strong nonlinear effects to albumin, lactate, and operative duration, and it identified expected interactions with urgency and operation family [16]. However, when its fitted probabilities were aggregated across the matrix, the resulting pattern was less compact. The latent phenotype model explained the same broad gradients through four clinically interpretable states, whereas the additive model spread comparable information across many coefficient surfaces. This is not evidence that the latent model is truer in a biological sense. It is evidence that the latent representation is more efficient for the specific problem of summarizing heterogeneity across surgery families and care environments.

6. Discussion

Several conclusions follow from these findings, though they need to be stated carefully. The first is that heterogeneity of anastomotic leak across gastrointestinal operations and postoperative care environments is not well described by a single continuum running from low to high risk. The latent model instead recovered four hidden susceptibility phenotypes whose feature bundles and leak behavior were materially different. Some were dominated by early systemic derangement, some by local pelvic technical strain, some by chronic reserve depletion, and one by relative physiologic quiet. These patterns were not identical to operation families or pathway labels, though they were unevenly distributed across them. That difference is important. If the classes had simply reproduced procedure

Table 10 Length of Stay by Phenotype (Non-leak) :contentReference[oaicite:9]index=9

Phenotype	Median LOS	Readmission	Imaging Use
Low-disturbance	5 days	Low	Low
Inflammatory-perf	8 days	High	High
Pelvic-mechanical	7 days	Moderate	Moderate
Depletion-frailty	9 days	High	Moderate

categories, the latent approach would have added little. The fact that the same surgery family contained multiple phenotypes, and that the same phenotype behaved differently across settings, indicates that the heterogeneity is layered rather than flat.

The second conclusion is that a large share of cross-setting and cross-procedure variation in leak arose because hidden states were sorted unevenly across the matrix. More than half of the between-cell variance was attributable to differential phenotype composition. This is a stronger statement than the familiar observation that some pathways care for sicker patients. The phenotypes were not defined by any single severity metric. They captured multidimensional perioperative configurations. A rescue-intensive referral pathway did not simply have lower albumin or longer operations. It accumulated a specific hidden state in which nutritional depletion, malignancy burden, frailty, and prolonged perioperative stress traveled together. Emergency overnight care did not merely exhibit higher urgency. It accumulated a state in which contamination, lactate elevation, vasopressor continuation, and inflammatory activation co-occurred. Seeing the problem this way helps distinguish true pathway differences from crude average severity.

The third conclusion is that sorting is not the whole story. The same hidden state did not carry identical leak risk across all operation families and care environments [17]. This was especially clear for the pelvic-mechanical phenotype in low colorectal surgery and for the inflammatory-perfusion phenotype in emergency bowel surgery. A patient whose feature pattern places them in a pelvic-mechanical state is not equally likely to leak in every recovery environment. Likewise, an inflammatory-perfusion phenotype after small-bowel surgery is not equivalent in a protocolized daytime pathway and an emergency overnight pathway. This expression gradient matters because it points toward pathway-level factors that are not reducible to composition. Technical concentration, surveillance intensity, timing of escalation, and local rescue capacity likely contribute, even though they are only partially captured by the structured data. The current analysis cannot separate those mechanisms causally, but it does show that phenotype expression varies in a way that composition alone cannot explain.

The operation families offer a useful lens through which to interpret these results. Low pelvic colorectal reconstruction emerged as the family in which local geometric and technical vulnerability was most visible in latent form. The pelvic-mechanical phenotype was both highly prevalent and highly expressive there, especially in urgent and overnight settings. This suggests that the well-known vulnerability of low pelvic anastomosis is not simply a single fixed procedure effect. It is partly the result of a distinct hidden state that can be concentrated or diluted depending on the pathway and can be made more or less clinically hazardous by the same pathway. That is a more granular statement than saying that low pelvic surgery is high risk. It says that within low pelvic surgery, a particular latent state drives much of the heterogeneity.

Foregut reconstruction told a different story. Esophagogastric and gastric operations were strongly enriched for depletion-frailty and inflammatory-perfusion states. The expression gradient was present but smaller relative to the sorting gradient than it was in low pelvic surgery. This suggests that foregut heterogeneity is shaped more by who reaches the pathway and in what baseline reserve condition than by large divergence in hidden-state behavior once there. In practical terms, referral concentration and nutritional depletion may explain more of the foregut signal than setting alone. That is consistent with how specialized foregut services often function: they accumulate difficult cancer and redo cases whose hidden vulnerability is evident before any question of rescue pathway even arises.

Small-bowel and right-sided colonic surgery occupied a middle ground that is analytically informative. Their average leak probabilities are lower than those of low pelvic and foregut operations, but they are not uniform. Small-bowel surgery in particular showed substantial influx of the inflammatory-perfusion phenotype under urgent and overnight conditions. That pattern suggests that in some families, acute systemic stress rather than local anastomotic geometry is the main hidden driver of leak heterogeneity. This distinction matters for pathway design. A surveillance rule built around pelvic-specific concern signals will not necessarily map well onto small-bowel emergency surgery, where the key hidden state looks more like diffuse physiologic strain and contamination-related instability [18].

The phenotype perspective also clarifies why some apparently modest average-risk pathways remain clinically important. Consider left-sided colectomy under routine elective care. Its mean leak rate is not extreme, yet the posterior phenotype distribution is mixed and entropy is relatively high. This means the pathway contains patients crossing between low-disturbance and higher-risk states, and the average obscures that spread. A flat procedure-based estimate can therefore make such a pathway look unremarkable when, in reality, it contains meaningful hidden heterogeneity. That heterogeneity may not justify an entirely separate service, but it does suggest that the same operation family can harbor quite different postoperative profiles that a single global rate cannot display.

From a modeling standpoint, the results argue for caution in how risk tools are interpreted. A conventional predictive model that uses albumin, lactate, white blood cell count, operation type, and urgency can produce a leak probability with reasonable accuracy. What it does not automatically reveal is whether that probability arises because the patient appears to belong to a depletion-frailty state, a pelvic-mechanical state, or an inflammatory-perfusion state. Those distinctions may matter for action. A risk estimate driven mainly by depletion-frailty may imply different monitoring priorities than one driven by localized technical strain. The current analysis does not claim that every hospital should deploy a phenotype engine in real time. It does suggest that future surveillance tools may be more clinically intelligible if they indicate not only how much risk is present, but what hidden configuration is thought to underlie it.

The findings also have implications for comparative interpretation across care environments. It is easy to view overnight emergency pathways as uniformly adverse and specialized rescue pathways as uniformly protective. The phenotype analysis suggests a more nuanced picture. Emergency overnight care appears adverse partly because it accumulates the inflammatory-perfusion phenotype and partly because that phenotype is expressed more strongly there. Rescue-intensive complex referral care appears high risk chiefly because it accumulates depletion-frailty and foregut cases, though some expression gradients remain. These are different kinds of heterogeneity. One points toward acute pathway disruption and possibly modifiable process factors. The other points more toward concentration of difficult baseline states. Treating them as the same because both produce high average leak rates would erase clinically relevant distinctions.

There is also a conceptual lesson about what surgery type means in observational data. Operation family is often entered into a model as a fixed categorical predictor, as though its meaning were exhausted by anatomy or technical category [19]. The latent results imply that surgery family is also a sorting device for hidden physiologic configurations. Low colorectal surgery is not only a pelvic operation. It is also a route into a mechanical-strain state. Foregut surgery is not only a foregut operation. It is also a route into depletion-frailty and inflammatory-perfusion states. This way of thinking could be useful beyond leak, because many postoperative complications likely arise from hidden bundles of reserve, technical burden, and systemic response that are unevenly selected by operations.

One might ask whether the latent classes should be interpreted as mechanistic endotypes or only as statistical conveniences. The prudent answer is somewhere in between. The model does not observe perfusion directly, nor does it measure tension geometry or microcirculatory healing. It infers hidden states from structured proxies. The resulting classes are therefore empirical phenotypes, not experimentally verified biologic endotypes. Yet the fact that they align with clinically recognizable bundles and produce stable gradients across operation families and care environments gives them more meaning than purely arbitrary clusters. Their value lies not in claiming biological purity, but in organizing heterogeneity in a way that is more coherent than a long list of marginal coefficients.

This interpretive caution does not weaken the main empirical conclusion. Whether one calls them phenotypes, hidden states, or latent susceptibility profiles, the data support the existence of a small number of recurring perioperative configurations that are distributed unevenly across surgery families and care environments and that contribute substantially to observed leak heterogeneity. That statement is useful even without strong causal claims. It invites a different set of questions for future work. Instead of asking only whether one can predict leak better, one might ask whether targeted pathway interventions change the prevalence of the inflammatory-perfusion phenotype in emergency care, whether prehabilitation shifts patients out of the depletion-frailty state before foregut reconstruction, or whether pelvic technical innovations reduce the expression of the pelvic-mechanical state. Those are intervention questions, not merely prediction questions.

7. Limitations

Several limitations shape the interpretation of the latent phenotype structure. The first is that latent classes are model-dependent. A finite mixture with four classes is a useful summary of the observed covariation, but it is not the only possible summary. A continuous latent variable model or a mixed-membership formulation might represent the same data differently. Some patients likely contain blended features from more than one hidden state, and the posterior entropy results confirm this. The class solution should therefore be read as a practical approximation to structured heterogeneity rather than as a claim that the clinical world consists of four sharply bounded leak endotypes [20].

A second limitation concerns the feature set itself. The model relied on structured variables routinely available in the electronic record. It did not have direct access to anastomotic level details beyond those implied by procedure family, intraoperative perfusion imaging, operative note nuance, drain strategy, exact stapling technique, conduit quality, or continuous hemodynamic data. Some of the hidden-state structure probably reflects these omitted variables only indirectly. The pelvic-mechanical phenotype, for example, likely stands in for a richer set of local geometric and technical conditions than the model could observe explicitly. Similarly, the depletion-frailty phenotype probably compresses several distinct forms of baseline vulnerability into one state because structured data cannot fully separate cancer cachexia, chronic inflammatory wasting, and frailty-related reserve loss.

Third, the inclusion of early postoperative variables makes the phenotype concept partly dynamic rather than purely preoperative. This was intentional, but it narrows the interpretation. The model is best understood as identifying early perioperative susceptibility states, not baseline patient phenotypes in a timeless sense. A patient may move from a relatively quiescent preoperative condition into an inflammatory-perfusion state once the operation and initial recovery reveal stress that was not previously visible. That is clinically relevant, yet it means the classes are not suitable for every question. They are more useful for early surveillance and pathway analysis than for purely

preoperative counseling.

Fourth, because the study is observational, phenotype expression differences across care environments are not causal estimates. A higher leak probability for the same phenotype in emergency overnight care could reflect unmeasured confounding rather than pathway effect. Some relevant variables, such as contamination severity granularity, staffing composition, delay to source control before surgery, and surgeon-specific experience, were only partially captured. The expression gradients observed here identify where hidden states behave differently, but they cannot by themselves determine why. Any claim that a particular environment amplifies a phenotype would require stronger causal design than the current data permit.

Fifth, the endpoint emphasized clinically consequential leak and depended on management-linked recognition. This is appropriate for many practical questions, but it means that latent states influencing detection and intervention thresholds may partly shape the observed event. A pathway with denser surveillance could identify more leaks of a given biologic severity than a pathway with sparser surveillance. Some of the phenotype expression gradient may therefore reflect differences in recognition timing or management behavior rather than only differences in biological failure. That possibility is particularly relevant for rescue-intensive referral pathways. The stricter sensitivity analysis using invasive source control reduced but did not eliminate this concern.

Sixth, the cohort came from a common electronic network [21]. This is useful for reducing technical inconsistency, but it may limit external generalizability to systems with very different documentation habits or perioperative structures. The latent classes were stable across hospitals inside the network, yet a different health system might exhibit somewhat different class prevalence or class interpretation. For example, hospitals with routine perfusion imaging, different enhanced recovery adoption, or alternative critical care triage could alter the empirical boundaries between low-disturbance and inflammatory-perfusion states. External replication would therefore be necessary before treating the exact four-class structure as universally fixed.

Seventh, although hospital-blocked estimation reduced overfitting to local practice, the model still did not include explicit surgeon-level or unit-level random effects. Some of the apparent phenotype sorting into operation families and care environments may therefore partly reflect concentration of cases within specific surgical teams or specialty units. In principle that concentration is part of the clinical pathway and not necessarily something to be removed. Even so, a richer multilevel design could clarify how much of the hidden-state structure belongs to patient-operation configuration and how much to provider concentration.

Finally, latent classes can invite overinterpretation because they produce intuitively attractive labels. Terms such as inflammatory-perfusion or pelvic-mechanical are descriptive shorthand, not pathologic proof. They summarize clusters of structured variables and leak behavior. They do not reveal the microscopic tissue processes of anastomotic healing. The labels are useful because they organize the data, but they should not be mistaken for directly observed mechanisms. Their real empirical value lies in clarifying that heterogeneity across surgery families and care environments is more structured than average rates imply.

8. Conclusion

The multicenter evidence presented here suggests that heterogeneity of anastomotic leak across gastrointestinal operations and postoperative care environments is organized partly by hidden susceptibility phenotypes. Four latent states provided a stable and interpretable summary of perioperative variation: a low-disturbance state, an inflammatory-perfusion state, a pelvic-mechanical state, and a depletion-frailty state. These states were not distributed uniformly across the clinical landscape. Low pelvic reconstruction concentrated the pelvic-mechanical phenotype, foregut surgery concentrated depletion-frailty and inflammatory-perfusion phenotypes, emergency overnight care concentrated inflammatory-perfusion, and rescue-intensive referral care concentrated depletion-frailty.

Most of the dispersion in leak probability across the procedure-by-setting matrix arose because these hidden states were sorted unevenly across pathways. A smaller but still substantial share came from the fact that the same phenotype carried different leak probability in different operative and care contexts. That combination implies that observed pathway differences reflect both who arrives and how the pathway shapes the expression of that arrival.

Viewed this way, average leak rates are incomplete descriptions of a layered phenomenon. Procedure family, care environment, and immediate postoperative physiology do not merely add independent increments of risk. They help reveal and redistribute distinct perioperative vulnerability states. A phenotype-based perspective does not replace conventional risk estimation, but it offers a more structured account of why leak heterogeneity looks the way it does across gastrointestinal surgery [22].

Acknowledgments

We would like to thank the reviewers of this research for their helpful comments and suggestions.

References

[1] B. Sadanandan, “Data-driven risk stratification for anastomotic leak: A proof-of-concept study using electronic health records,” *Journal of Data Science, Predictive Analytics, and Big Data Applications*, vol. 9, no. 6, pp. 1–27, 2024.

[2] H. Linde, V. Bartusevicius, Åke Norberg, F. Klevebro, and J. Grip, “Frequency of blood lactate elevation following esophagectomy and its association to postoperative complications.,” *Acta anaesthesiologica Scandinavica*, vol. 67, pp. 277–283, 12 2022.

[3] E. L. Watkins, N. Schellack, V. Abraham, and B. Bebington, “Men and those with a history of smoking are associated with the development of postoperative ileus following

elective colorectal cancer resection at a private academic hospital in johannesburg, south africa: A retrospective cohort study.," *Frontiers in surgery*, vol. 8, pp. 667124–, 6 2021.

[4] A. Holmén, M. Hayami, E. Szabo, I. Rouvelas, T. Agustsson, and F. Klevebro, "Nutritional jejunostomy in esophagectomy for cancer, a national register-based cohort study of associations with postoperative outcomes and survival," *Langenbeck's archives of surgery*, vol. 406, pp. 1415–1423, 11 2020.

[5] S. Engelsgjerd, S. Kodres-O'Brien, E. Choudhury, B. M. Garijo, J. B. Mason, and K. J. Kowalczyk, "Outcomes and safety of suprapubic vs urethral catheterization following pelvic fascia–sparing robotic prostatectomy.," *Urology practice*, vol. 11, pp. 376–384, 12 2023.

[6] J. H. Law, C. Z.-M. Ng, S.-K. Lauw, J. B. Y. So, G. W. Kim, and A. Shabbir, "384. a 10-year experience with anastomotic leaks in upper gastrointestinal surgery," *Diseases of the Esophagus*, vol. 35, 9 2022.

[7] V. Bencurik, M. Skrovina, L. Martínek, J. Bartoš, M. Machackova, M. Dosoudil, E. Štěpánová, L. Přibylová, R. Bris, and K. Vomackova, "Intraoperative fluorescence angiography and risk factors of anastomotic leakage in mini-invasive low rectal resections.," *Surgical endoscopy*, vol. 35, pp. 5015–5023, 9 2020.

[8] S. Kirmiz, S. K. Babitz, S. Linsell, J. Qi, C. M. Brede, D. C. Miller, J. E. Montie, B. R. Lane, and M. U. S. I. Collaborative, "Regular vs. selective use of closed suction drains following robot-assisted radical prostatectomy: results from a regional quality improvement collaborative.," *Prostate cancer and prostatic diseases*, vol. 23, pp. 151–159, 8 2019.

[9] B. K. Sah, Z. Yang, L. Jian, Y. Chao, L. Chen, Z. Huan, Y. Min, and Z. Z. Gang, "Early diagnosis of anastomotic leakage after gastric cancer surgery," 4 2020.

[10] Y. L. Verschoor, D. M. J. Lambregts, J. van den Berg, B. A. Grotenhuis, A. Aalbers, B. V. Triest, R. G. Beets-Tan, M. van de Belt, S. Dokter, S. Balduzzi, E. E. Voest, J. B. A. G. Haanen, M. E. van Leerdam, G. Beets, and M. Chalabi, "Radiotherapy, atezolizumab, and bevacizumab in rectal cancers with the aim of organ preservation: The tarzan study.," *Journal of Clinical Oncology*, vol. 41, pp. 158–158, 2 2023.

[11] G. L. Baiocchi, S. Giacopuzzi, D. Reim, G. Piessen, P. Costa, J. V. Reynolds, H. J. Meyer, P. Morgagni, I. Gockel, L. L. Santos, L. S. Jensen, T. B. Murphy, D. D'Ugo, R. Rosati, U. F. Romario, M. Degiuli, W. Kielan, S. Mönig, P. Kołodziejczyk, W. Polkowski, M. Pera, P. M. Schneider, B. P. L. Wijnhoven, W. O. de Steur, S. S. Gisbertz, H. H. Hartgrink, J. W. van Sandick, M. Botticini, A. H. Hölscher, W. H. Allum, and G. de Manzoni, "Incidence and grading of complications after gastrectomy for cancer using the gastrodata registry: A european retrospective observational study.," *Annals of surgery*, vol. 272, pp. 807–813, 9 2020.

[12] N. T. Nguyen, P. D. Rudersdorf, B. R. Smith, K. M. Reavis, X.-M. T. Nguyen, and M. J. Stamos, "Management of gastrointestinal leaks after minimally invasive esophagectomy: conventional treatments vs. endoscopic stenting.," *Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract*, vol. 15, pp. 1952–1960, 9 2011.

[13] A. Maher, P. D. Coppi, S. Blackburn, S. Loukogeorgakis, S. Eaton, K. Cross, S. Giuliani, J. Curry, and D. Mullassery, "Short and medium term outcomes of open and laparoscopic assisted oesophageal replacement procedures.," *Journal of pediatric surgery*, vol. 59, pp. 192–196, 10 2023.

[14] H. Cantay, U. Aydin, I. Ozaydin, T. Anuk, S. K. Tasci, U. Yildiz, D. Ermutlu, and O. Aksoy, "Different surgical methods in colon anastomosis: Experimental study.," *Arquivos brasileiros de cirurgia digestiva : ABCD = Brazilian archives of digestive surgery*, vol. 35, pp. e1662–e1662, 6 2022.

[15] L. Marano, R. Porfidia, M. Pezzella, M. Grassia, M. Petrillo, G. Esposito, B. Braccio, P. Gallo, V. Boccardi, A. Cosenza, G. Izzo, and N. D. Martino, "Clinical and immunological impact of early postoperative enteral immunonutrition after total gastrectomy in gastric cancer patients: a prospective randomized study.," *Annals of surgical oncology*, vol. 20, pp. 3912–3918, 7 2013.

[16] G. Tan, A. K. Srivastava, S. Grover, D. Peters, P. Dorsey, A. Scott, J. Jhaveri, D. Tilki, A. E. Te, and A. Tewari, "Optimizing vesicourethral anastomosis healing after robot-assisted laparoscopic radical prostatectomy: Lessons learned from three techniques in 1900 patients," *Journal of endourology*, vol. 24, pp. 1975–1983, 10 2010.

[17] G. Zhang, C. Zhang, N. Sun, L. Xue, Z. Yang, L. Fang, Z. Zhang, Y. Luo, S. Gao, Q. Xue, J. Mu, Y. Gao, F. Tan, and J. He, "Neoadjuvant chemoradiotherapy versus neoadjuvant chemotherapy for the treatment of esophageal squamous cell carcinoma: a propensity score-matched study from the national cancer center in china," *Journal of cancer research and clinical oncology*, vol. 148, pp. 1–12, 5 2021.

[18] B. Kumar and P. K. Mishra, "Early closure of loop ileostomy: Is it feasible and safe?," *INDIAN JOURNAL OF APPLIED RESEARCH*, pp. 67–69, 11 2020.

[19] J. Slieker, F. Daams, I. M. Mulder, J. Jeekel, and J. F. Lange, "Systematic review of the technique of colorectal anastomosis.," *JAMA surgery*, vol. 148, pp. 190–201, 2 2013.

[20] A. Chatterjee, N. Tuteja, R. Tanger, P. K. Gupta, and A. K. Shukla, "Hirschsprung's disease: Our institutional experience," *INDIAN JOURNAL OF APPLIED RESEARCH*, pp. 68–71, 4 2021.

[21] M. K. Hebbar, W. Riaz, P. Sains, M. K. Baig, and M. S. Sajid, "Meta-analysis of randomized controlled trials only exploring the role of single incision laparoscopic surgery versus conventional multiport laparoscopic surgery for colorectal resections.," *Translational gastroenterology and hepatology*, vol. 3, pp. 30–30, 5 2018.

[22] C. Richards, R. C. McMurray, E. T. Criman, M. E. Clark, and S. M. Gillern, "An unusual presentation of a rare disease: posterior reversible encephalopathy syndrome following abdominal sepsis," *Journal of surgical case reports*, vol. 2016, pp. rjw184–, 11 2016.