OPEN ACCESS

Competing Risk Trajectory Modeling of Anastomotic Leak Emergence Across Gastrointestinal Procedure Families and Postoperative Care Environments

Dimas Saputra[1], Fajar Hidayat[2], Andika Ramadhan[3]

[1] Universitas Negeri Semarang, Fakultas Teknik, Jurusan Ilmu Komputer, Sekaran Campus, Gunungpati, Semarang 50229, Jawa Tengah, Indonesia; [2] Universitas Sultan Ageng Tirtayasa, Fakultas Teknik, Program Studi Informatika, Jalan Jenderal Sudirman Km 3, Cilegon 42435, Banten, Indonesia; [3] Universitas Malikussaleh, Fakultas Teknik, Jurusan Teknik Informatika, Jalan Cot Tengku Nie, Reuleut, Aceh Utara 24355, Aceh, Indonesia;

Abstract

Postoperative anastomotic leak is usually summarized as a binary outcome, but its clinical significance depends not only on whether it occurs, but also on when it becomes detectable and under what surveillance conditions it is recognized. A leak identified on postoperative day 3 in a highly monitored environment is not operationally equivalent to a leak first detected after several days on a lower-observation ward. This study examined temporal heterogeneity in leak emergence across gastrointestinal procedure families and postoperative care environments using a retrospective multicenter electronic health record cohort. Adult patients undergoing gastrointestinal operations with restoration of continuity were followed through daily postoperative records for 21 days and outcome ascertainment for 30 days. Procedures were grouped into proximal colonic, distal colonic, low pelvic colorectal, small-bowel, and upper gastrointestinal reconstruction families. Care environment was defined by the joint combination of first postoperative bed pathway and hospital rescue capability. The primary endpoint was clinically consequential anastomotic leak, with discharge and death before leak treated as competing events. The analytic cohort included 41,286 procedures, of which 2,521 developed leak within 30 days, corresponding to an overall incidence of 6.1%. Low pelvic colorectal and upper gastrointestinal reconstructions had the highest crude event rates and the latest median day of detection. After covariate standardization, proximal colonic and small-bowel procedures showed limited temporal variation across care environments, whereas low pelvic colorectal and upper gastrointestinal procedures displayed substantial shifts in both cumulative incidence and detection timing. Ward-first pathways in standard rescue environments were associated with later recognition and higher post-detection rescue intensity for these procedures. A dynamic competing-risk model with low-rank procedure-by-environment interactions outperformed a static postoperative baseline model in discrimination, calibration, and cross-environment transport. The findings suggest that leak surveillance should be calibrated to procedural trajectory and monitoring context, rather than treated as a single undifferentiated binary risk problem.

1. Introduction

Surgical teams rarely experience anastomotic leak as a timeless event. The complication unfolds within hours and days, and its consequences depend on when physiologic drift becomes visible, when suspicion crosses the threshold for imaging or intervention, and how quickly a rescue pathway can be activated. In practice, this means that two patients with the same nominal thirty-day leak probability may occupy very different clinical situations. One patient may declare instability early in a high-observation postoperative pathway, prompting rapid imaging and source control. Another may remain on a lower-intensity ward, accumulate subtle abnormalities over several days, and reach formal diagnosis only after the septic process has become harder to contain. Yet much of the quantitative literature treats leak as a cross-sectional endpoint, asking whether it occurs, while leaving the temporal structure of its emergence largely implicit. That simplification is understandable because binary outcomes are easier to model, easier to validate, and easier to summarize. It is also limiting, especially when the question of interest is hetero-

CTASC Copyright by the Authors.
Licensed as an open access article using a CC BY 4.0 license.

geneity across surgery types and care settings rather than average risk alone.

The present study proceeds from a different premise. A leak is not only a probability attached to a patient; it is a trajectory embedded in a postoperative surveillance environment. Timing matters because early and late recognition carry different implications for imaging use, ward escalation, rescue feasibility, and mortality. Procedure type matters because the biological and mechanical pathways to failure are not exchangeable across the gastrointestinal tract. A proximal ileocolic anastomosis, a stapled colorectal anastomosis above the peritoneal reflection, a low pelvic colorectal anastomosis, a jejunojejunal or enteroenteric reconstruction, and an upper gastrointestinal reconstruction differ in blood supply geometry, local contamination burden, technical exposure, tolerance of edema, and the clinical conspicuity of early failure. Care setting matters because the same evolving physiology will be sampled, documented, and interpreted with different intensity depending on the first postoperative bed, the availability of overnight imaging, the presence of advanced endoscopic or interventional radiologic rescue, and the institutional habit of escalating from equivocal signs rather than waiting for frank decompensation.

Those considerations create a specific empirical problem. If one only estimates pooled thirty-day leak risk, it becomes difficult to distinguish a procedure that truly has higher baseline hazard from one whose apparent excess arises because detection tends to occur later and under lower-observation conditions. It also becomes difficult to separate a care environment that changes the biological likelihood of failure from one that changes the timing and threshold of recognition [1]. In reality, these mechanisms may coexist. A resource-constrained postoperative environment may indirectly elevate biological risk through delayed fluid management, less intensive hemodynamic monitoring, or slower response to ileus and distension. At the same time, it may also defer diagnosis by reducing the frequency with which warning signals are measured or acted upon. A binary end point collected at thirty days compresses those mechanisms into a single number. A time-resolved framework allows them to be partially disentangled.

Recent work using structured electronic health record data showed that postoperative white blood cell count [2], lactate, ICU admission, surgery type, and preoperative albumin carry predictive signal for leak, while also emphasizing that the available dataset represented a tertiary critical-care environment and did not include subgroup analyses by procedure type or care setting; that combination of feasibility and unresolved heterogeneity makes a time-dependent multicenter design a logical next step [2]. The methodological implication is straightforward. Once one accepts that surgery class and care setting may shape not only overall incidence but the daily hazard of recognition, a static one-time probability model becomes an incomplete description of the problem. A clinically useful analysis should represent postoperative time explicitly, accommodate competing events such as discharge and death before leak, and allow interaction between procedure family and care environment to vary across days rather than remain constant.

A second difficulty arises from observation intensity. In many hospitals, patients in higher-acuity postoperative pathways generate denser data: more frequent vital signs, more repeated laboratory values, more documentation of fluid balance, more transfer notes, and earlier imaging when suspicion arises. This richer signal can improve apparent predictive performance, but it can also mislead interpretation. Dense monitoring does not only reveal that a patient is sick; it changes what can be known about the patient at a given moment. Conversely, sparse observation on a standard ward does not only indicate lower severity; it can delay recognition of a process already underway. In statistical terms, the postoperative data stream is partly endogenous to the care setting. Models that ignore this may conflate biological hazard with the observation process itself [3]. For a study focused on heterogeneity across settings, that distinction cannot be treated as a minor technicality.

The conceptual approach adopted here is to regard leak emergence as a competing-risk process observed in discrete postoperative time. On each day after surgery, a patient remains at risk for leak detection, discharge without prior leak, or death before leak. These alternatives are not interchangeable censoring events. A patient discharged early without suspicion has left the observation window in a way that carries information about clinical stability, while a patient who dies before a documented leak has experienced a distinct terminal pathway. Modeling these states jointly provides a more coherent picture of temporal heterogeneity than simply fitting a single binary classifier at a fixed horizon. Moreover, a competing-risk framework yields quantities directly relevant to clinicians, such as day-specific hazard, cumulative incidence by postoperative day, and restricted mean day-to-detection among those who leak.

The notion of care setting in this analysis is also deliberately operational rather than administrative. Instead of classifying hospitals with broad labels such as teach-

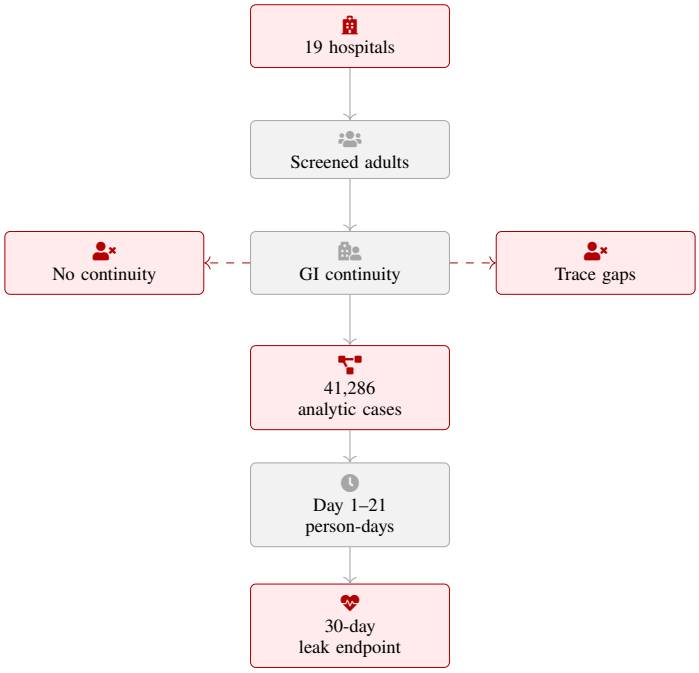


Figure 1: Study-frame overview for the multicenter gastrointestinal anastomosis cohort. The diagram emphasizes the operational sequence used in the manuscript: adult screening across nineteen hospitals, exclusion of non-restorative or poorly traceable admissions, formation of the 41,286-case analytic cohort, construction of daily postoperative records through day 21, and 30-day leak ascertainment.

ing or community, the study defines postoperative care environments from two components visible in clinical operations. The first is the initial postoperative bed pathway, distinguishing ICU-first, step-down-first, and ward-first recovery. The second is the rescue environment of the hospital, defined by whether around-the-clock advanced imaging and intervention are continuously available without interfacility transfer or daytime delay. This yields care environments that are more directly related to what matters after an anastomosis has been created: how intensely the patient is observed in the early postoperative window and how quickly a suspected failure can be confirmed and treated. That distinction is especially relevant because the same hospital may recover some patients in high-observation settings and others on standard wards, and because the effect of ward recovery plausibly depends on whether advanced rescue is readily available.

The study addresses three related questions. The first is descriptive: how do leak incidence curves differ across major procedure families and postoperative care environments when time is represented explicitly? The second is inferential: after adjustment for baseline condition, operative features, and evolving postoperative physiology, which procedure families remain sensitive to care environment, and is that sensitivity expressed more strongly as higher cumulative incidence, later detection, or both? The third is predictive: does a dynamic competing-risk model with structured procedure-by-environment interaction provide better calibration and transport than a static postoperative baseline model? These are not independent questions. A model that predicts well but offers no interpretable heterogeneity structure is limited for service design, while an interpretable model that ignores the temporal observation process may misstate the magnitude of environmental differences [4].

This paper therefore examines leak emergence as a longitudinal event rather than a cross-sectional label. The primary analysis uses daily postoperative records over twenty-one days, because most clinically consequential leaks in routine practice are first recognized within that period and because daily hazard beyond that horizon becomes sparse. Outcome ascertainment extends to thirty days to capture the full clinical endpoint, but the dynamic model concentrates on the interval in which observation density and bed pathway are most relevant. Procedure family, care environment, baseline physiology, and time-varying postoperative measurements are combined within a weighted competing-risk framework. The results show that heterogeneity is not simply a matter of some operations leaking more often than others. For certain proce-

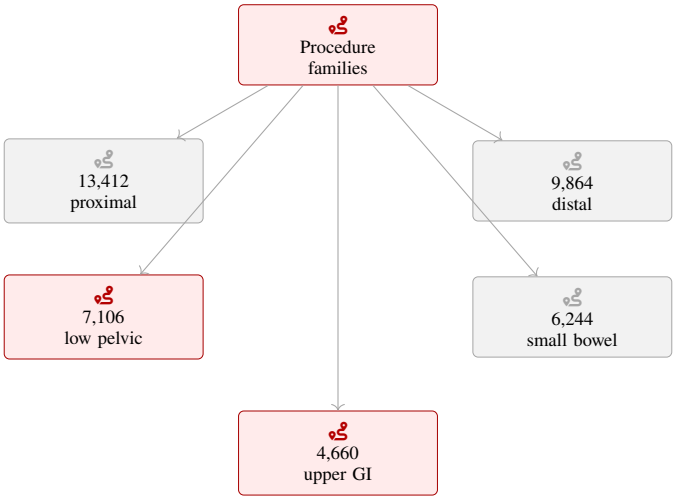


Figure 2: Procedure-family partition used for trajectory modeling. The figure preserves the paper's five clinically meaningful reconstruction groups and the reported cohort counts, highlighting the families that later show stronger temporal sensitivity to postoperative setting.

dure families, care environment changes when leaks surface, how much cumulative risk accrues before discharge, and how intensively patients require rescue once the event is recognized. For other procedures, overall heterogeneity is much smaller, suggesting that elaborate setting-specific surveillance rules may add little beyond routine observation.

2. Data Source and Analytic Cohort

The source population consisted of adult gastrointestinal surgical admissions recorded between January 2016 and December 2024 across nineteen hospitals participating in a unified enterprise electronic health record environment. The hospitals served mixed referral regions and used common data standards for medication administration, laboratory transmission, operative coding, bed-tracking events, and discharge disposition. The shared digital infrastructure made it possible to construct patient-day level postoperative records without the extensive harmonization that usually complicates multicenter EHR research. At the same time, the participating institutions differed materially in their postoperative monitoring routines and rescue capability, creating the variation needed for the present analysis. All consecutive admissions in which continuity was restored by gastrointestinal anastomosis were initially screened. To preserve a coherent postoperative time origin, only the first qualifying operation in a hospitalization was retained when more than one candidate procedure occurred.

Eligibility required age of at least eighteen years and a documented operation involving an anastomosis in the stomach, small bowel, colon, rectum, or esophagus. Operations that created only a stoma without restoration of continuity were excluded because their postoperative hazard structure is different and the concept of anastomotic leak is absent or not primary. Admissions involving transplant surgery, severe trauma with discontinuity control, or planned staged abdominal closure were excluded because their physiology and care pathways are dominated by competing postoperative syndromes. Cases lacking a complete bed-location trace for the first seventy-two postoperative hours were removed because care environment was a central exposure and could not be inferred reliably from partial data [5]. Admissions transferred into the network after the index operation were excluded because early postoperative time before transfer could not be reconstructed with sufficient fidelity. After all exclusions, the analytic cohort comprised 41,286 procedures.

To reduce overfragmentation while preserving clinically meaningful heterogeneity, operations were grouped into five procedure families. The proximal colonic family included ileocolic and other right-sided colonic reconstructions following right hemicolectomy or extended right colectomy. The distal colonic family included left colectomy and sigmoid colectomy with colocolic or upper colorectal continuity. The low pelvic colorectal family included low anterior resection and very low colorectal or

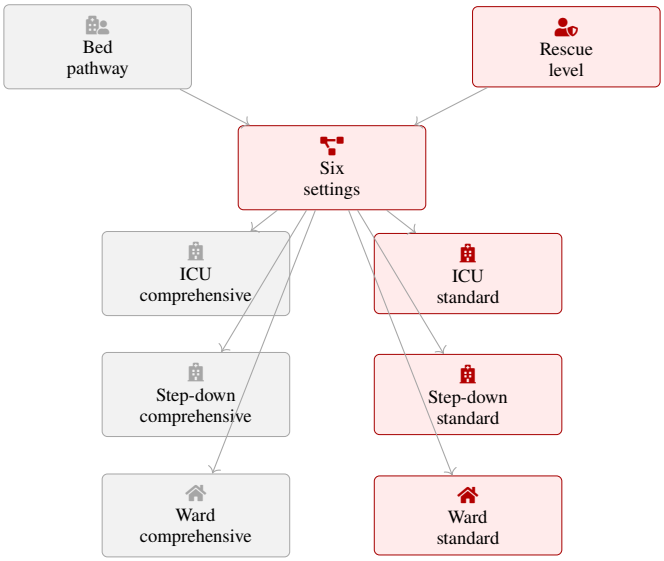


Figure 3: Operational care-environment taxonomy. The manuscript defines postoperative setting as the Cartesian product of first bed pathway and hospital rescue capability, yielding six clinically distinct monitoring-and-response environments rather than a single hospital label.

coloanal reconstruction, with or without diverting stoma. The small-bowel family included resections with enteroenteric or enterocolic anastomosis not otherwise categorized as proximal colonic. The upper gastrointestinal family included gastrectomy with gastrointestinal reconstruction and esophagectomy with esophagogastric or similar continuity restoration. This grouping was chosen because the central analytic aim was to distinguish families with plausibly different hazard shape and observation sensitivity, rather than to maximize procedure code granularity for its own sake. In total, the cohort contained 13,412 proximal colonic procedures, 9,864 distal colonic procedures, 7,106 low pelvic colorectal procedures, 6,244 small-bowel procedures, and 4,660 upper gastrointestinal reconstructions.

Care environment was defined as the Cartesian product of first postoperative bed pathway and hospital rescue environment. First postoperative bed pathway captured where the patient was initially managed after leaving the operating room and had three levels: ICU-first, step-down-first, and ward-first. Rescue environment captured the hospital's capacity to confirm and treat suspected leak without delay and was defined at the institution level as either comprehensive or standard. Comprehensive rescue environments had continuous overnight access to cross-sectional imaging, interventional radiology, advanced endoscopy, and emergency return to theatre. Standard rescue environments had routine daytime access to at least one of these components but lacked continuous same-night availability for the full rescue pathway. Combining these dimensions yielded six care environments: ICU-first comprehensive, ICU-first standard, step-down-first comprehensive, step-down-first standard, ward-first comprehensive, and ward-first standard [6]. Their respective counts in the cohort were 6,128, 4,014, 8,992, 7,486, 6,904, and 7,762.

Patient characteristics were extracted from preoperative structured data and from the operation record. Baseline variables included age, sex, body mass index, smoking history, diabetes, chronic kidney disease, malignancy status, corticosteroid exposure, immunosuppressive therapy, heart failure, chronic lung disease, prior abdominal surgery, admission source, and a deficit-based frailty score derived from functional dependency, cognitive impairment, mobility limitation, and polypharmacy. Preoperative laboratories included hemoglobin, albumin, creatinine, bilirubin, white blood cell count, and C-reactive protein where available. Operative variables included urgency, minimally invasive versus open approach, conversion status, operative duration, estimated blood loss, intraoperative transfusion, vasopressor burden, contamination class, diversion status, and anatomic family. The data were organized so that baseline covariates reflected information known by the end of the index operation.

The postoperative longitudinal record was constructed in daily intervals from postoperative day 1 through day 21. For each patient-day, the record contained the last available value, daily maximum, or burden metric for selected variables depending on their clinical meaning. White blood cell count, hemoglobin, lactate, creatinine, biliru-

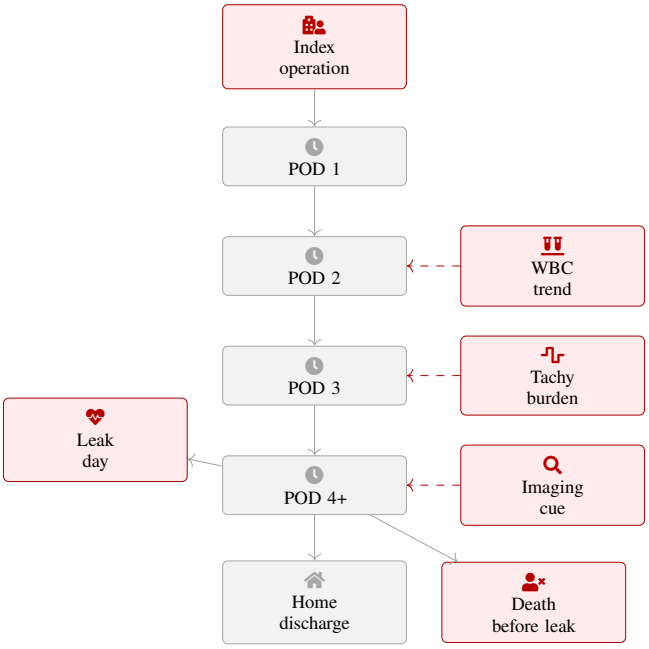


Figure 4: Discrete postoperative event-time encoding. Each patient contributes a day-by-day risk path after the index operation, with evolving physiologic evidence entering the history and three absorbing outcomes represented explicitly: clinically recognized leak, discharge without prior leak, and death before documented leak.

bin, heart rate burden above 100 beats per minute, temperature burden above 38.0 degrees Celsius, net fluid balance, vasopressor exposure, opioid administration intensity, enteral intake status, urine output, drain output when documented, and transfer events between care locations were captured. Because not every variable is sampled at the same frequency in every care environment, the longitudinal design did not require every field to be present each day. Instead, the statistical model represented both the measured values and the probability that a given day was observed with sufficient physiologic detail, thereby addressing informative differences in monitoring intensity.

The primary endpoint was clinically consequential anastomotic leak within thirty days of surgery. The event definition integrated diagnosis codes, radiology reports, procedure codes, escalation treatments, and charted postoperative trajectories. A leak was recorded when there was evidence compatible with failure of the anastomosis together with corroborating management or clinical response. Compatible evidence included direct operative confirmation, endoscopic visualization, imaging consistent with contrast extravasation or peri-anastomotic collection, or diagnostic language explicitly indicating leak or dehiscence. Corroboration included reoperation, image-guided drainage, endoscopic stenting or vacuum therapy, broad antimicrobial escalation linked to a compatible septic focus, or death following a trajectory strongly consistent with anastomotic failure [7]. Ambiguous records identified algorithmically underwent targeted review by trained abstractors blind to the study hypotheses. This approach was chosen to reduce dependence on any single documentation channel, especially because care environments differ in coding and imaging practices.

For the dynamic analysis, the time of leak was defined as the first postoperative day on which the event could be considered clinically recognized rather than merely biologically possible. This distinction is important. A leak may begin to form before it enters the formal record. However, from the perspective of surveillance and rescue, the actionable time point is recognition. If a patient underwent imaging on day 5 revealing leak after rising inflammatory markers on day 4, the event day was day 5, while the day 4 abnormalities entered the longitudinal history as predictors rather than event labels. This operationalization aligns the outcome with clinical decision timing. Competing events were discharge without prior leak and death without documented leak. Patients still hospitalized and leak-free after postoperative day 21 contributed to endpoint ascertainment through day 30 for descriptive totals but were administratively censored for the dynamic model because the number of patients remaining under

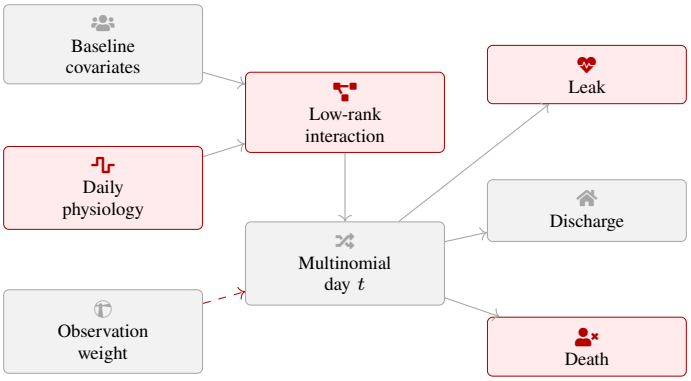


Figure 5: Dynamic competing-risk model architecture. Baseline factors, time-varying postoperative physiology, and inverse-probability observation weighting feed a structured day-specific transition model, while a low-rank procedure-by-environment interaction captures temporal heterogeneity without an unstable fully saturated design.

routine surveillance beyond day 21 was small and heavily selected.

The final cohort was clinically heterogeneous in ways relevant to both procedure family and care environment. Patients in upper gastrointestinal reconstruction were younger on average than proximal colonic cases but more likely to have malignancy, preoperative weight loss, low albumin, and prolonged operations. Low pelvic colorectal cases had high rates of stoma diversion and neoadjuvant treatment exposure. Small-bowel cases carried the largest burden of emergency presentation and contamination. Ward-first pathways were common after proximal and distal colonic surgery, whereas ICU-first pathways dominated upper gastrointestinal reconstruction. Standard rescue environments cared for a larger share of urgent small-bowel and distal colonic procedures than upper gastrointestinal operations, but still performed a meaningful number of low pelvic reconstructions [8]. These imbalances were not nuisances to be erased. They were part of the empirical structure that necessitated adjustment and time-dependent modeling.

3. Event-Time Encoding and Observation Model

The postoperative record was transformed into person-day format so that each patient contributed one row for each day from postoperative day 1 until leak, discharge, death, or administrative censoring at day 21. This representation makes the risk set explicit. A patient who remains hospitalized and leak-free through day 4 contributes four intervals of opportunity for the event to emerge, whereas a patient discharged on day 3 no longer contributes subsequent risk under inpatient observation. Such expansion is essential when the aim is to analyze not only whether leak occurs, but how its hazard accumulates under different surveillance trajectories. It also permits the incorporation of time-varying physiology without collapsing the entire postoperative course into a single summary statistic.

A central challenge is that measurement density varies by both clinical need and care environment. ICU-first pathways produce repeated laboratory and vital sign observations even in relatively stable patients. Ward-first pathways often generate sparser information, especially once early recovery appears uneventful. If one simply imputes missing values and proceeds as though each patient-day were equally observed, the model risks exaggerating the predictive value of dense monitoring and misattributing some setting differences to biology rather than surveillance. To address this, the analysis distinguished between the clinical history process and the observation process. For each patient-day, an observation indicator was defined to represent whether the record contained a minimally informative physiologic panel consisting of heart rate burden, temperature burden, at least one laboratory marker from a specified set, and medication administration continuity. This threshold was not intended to define good care; it served as a statistical marker of whether the day's physiology was visible with enough granularity for longitudinal modeling.

The observation process was then modeled separately using logistic regression with predictors available up to the preceding day. These predictors included procedure family, care environment, postoperative day, prior day physiology, transfer status, operative urgency, and whether suspicion-generating actions such as blood cultures or abdominal imaging had already occurred. The fitted probabilities from this model generated inverse prob-

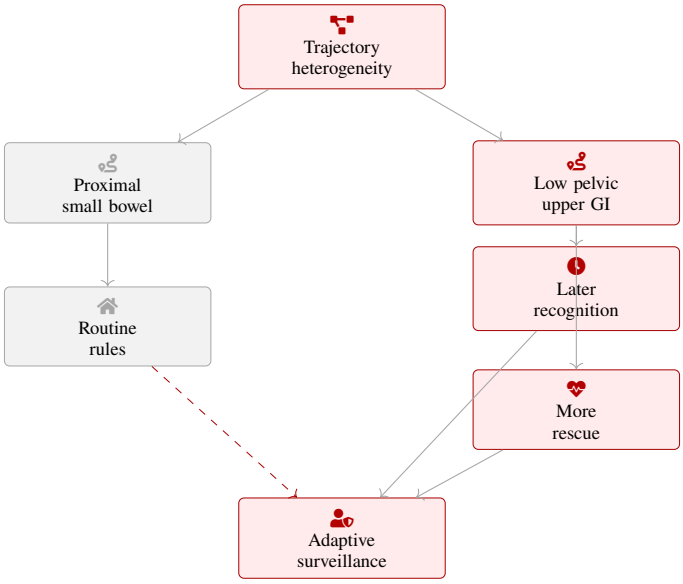


Figure 6: Clinical interpretation of the modeled trajectories. The manuscript's main practical message is selective adaptation: proximal colonic and small-bowel pathways appear comparatively stable across environments, whereas low pelvic colorectal and upper gastrointestinal reconstructions show setting-sensitive timing with later recognition and greater rescue burden, motivating targeted surveillance rather than uniform escalation.

ability weights applied in the primary competing-risk estimation [9]. Intuitively, patient-days that were unlikely to be richly observed given their history but nevertheless had informative data received lower leverage than patient-days whose detailed observation was expected. The weighting does not make sparse settings look like dense settings, nor should it. Rather, it reduces bias arising when the model learns from an observation pattern that is itself determined by care environment and evolving risk. The resulting estimand is closer to the daily hazard that would be inferred if one accounted for the fact that physiologic visibility is uneven across pathways.

Because leak risk is mediated by dynamic change rather than raw isolated values alone, the longitudinal predictors included both levels and short-term derivatives. White blood cell count entered as current value and twenty-four-hour change when two consecutive days were available. Lactate contributed as current level and persistence indicator, distinguishing transient elevation from repeated elevation across days. Heart rate and temperature entered as burden metrics, reflecting the cumulative number of hours or observations above threshold. Hemoglobin drop, creatinine rise, net positive fluid balance, failure to progress to enteral intake, and new vasopressor requirement were all represented with update terms that emphasized deviation from expected recovery rather than merely illness burden in the abstract. This approach allows the model to identify patterns such as an upper gastrointestinal patient whose lactate remains marginally elevated while heart rate burden increases, or a low pelvic patient whose inflammatory markers rise late despite otherwise modest early postoperative instability.

Another complication is that discharge competes directly with leak detection. A low-risk patient discharged on postoperative day 3 is not simply a censored observation equivalent to someone still hospitalized on day 3; the discharge itself reflects accumulated evidence of clinical stability. If early discharge is treated as noninformative censoring, the estimated late hazard among those remaining under observation becomes distorted. The present study therefore treats discharge without prior leak as a competing event rather than a censoring mechanism. This is not because discharge prevents leak biologically, but because it prevents inpatient recognition under the surveillance conditions under study. A patient may still develop a leak after discharge and return later, and such events were counted in the thirty-day endpoint if they occurred [10]. For the dynamic inpatient hazard, however, discharge marks the end of the monitored state in which bed pathway and resource environment directly shape event recognition.

Death before documented leak was also modeled as a competing event. Such deaths were uncommon, but they are substantively important because they often oc-

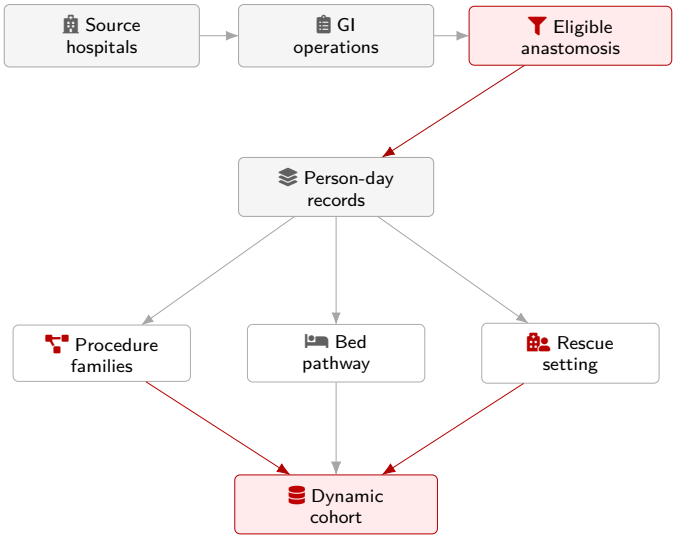


Figure 7: Cohort construction diagram. Consecutive gastrointestinal operations are screened into an eligible anastomosis cohort, converted to person-day records, and then indexed jointly by procedure family, first postoperative bed pathway, and rescue setting before dynamic competing-risk analysis.

cur in patients with severe comorbidity, contamination, or nonanastomotic postoperative complications. Treating them as simple censoring would remove a clinically distinct terminal pathway from the risk set. Moreover, the possibility of death before formal leak documentation is part of the real clinical landscape. Some patients deteriorate rapidly with mixed postoperative syndromes in which definitive attribution is difficult. Including death as a separate competing event acknowledges this complexity without requiring speculative relabeling of all such cases as leak.

To summarize heterogeneity in timing rather than only incidence, the study focused on several event-time quantities. The first was daily cause-specific hazard of leak recognition. The second was cumulative incidence of leak by postoperative day under the presence of discharge and death as competing events. The third was restricted mean day-to-detection among patients who leaked by day 21, which provides an interpretable summary of whether recognition tends to occur earlier or later in a given procedure family and care environment. These measures are complementary. Cumulative incidence may be similar in two settings even if one recognizes leak earlier and thereby alters rescue demands. Conversely, day-to-detection may be similar while cumulative incidence differs because one setting discharges lower-risk patients sooner or because baseline biological hazard is genuinely higher. A full temporal comparison therefore requires all three.

The daily time scale was chosen for clinical and statistical reasons [11]. Hourly modeling would be impractical given variable documentation frequency and would create severe sparsity for procedures with long periods of stable recovery. Weekly aggregation would erase precisely the window in which postoperative surveillance decisions are made. Daily intervals balance signal and stability. They also permit a flexible baseline hazard shape, allowing each postoperative day to carry its own intercept rather than imposing monotonic increase or decrease. This matters because leak recognition is rarely monotone. Some procedure families show early hazard driven by immediate ischemic or technical failure, followed by a plateau, whereas others show a more delayed rise as pelvic sepsis or cervical contamination becomes clinically apparent several days after surgery.

The distinction between biological emergence and clinical recognition deserves explicit attention because it affects interpretation of care-environment effects. A ward-first standard environment may appear to have later leak hazard not because tissue failure occurs later, but because suspicion is formalized later. From the standpoint of patient management, that distinction still matters. The empirical objective here is not to infer the exact unseen moment of anastomotic disruption, which routine EHR data cannot observe. It is to analyze the trajectory by which clinically consequential leak becomes visible and actionable across procedures and care environments. This is the relevant object for surveillance design, imaging thresholds, and res-

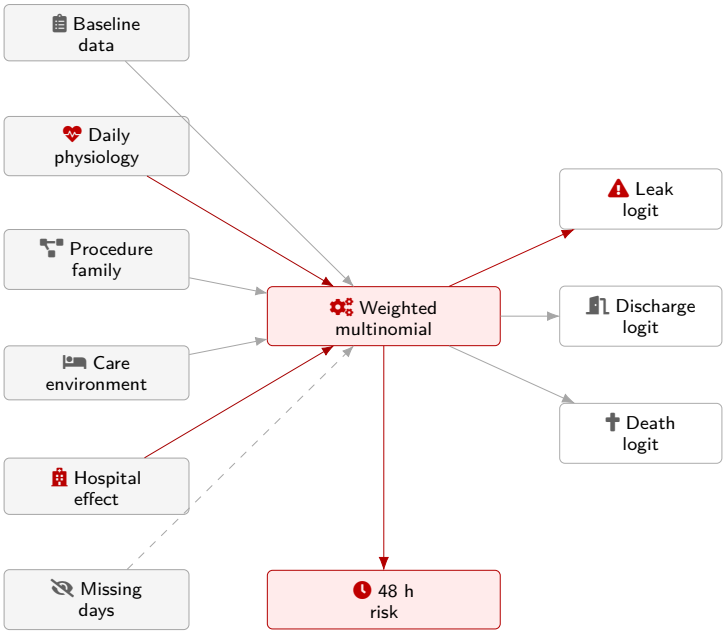


Figure 8: Dynamic model architecture. Baseline condition, evolving physiology, procedure family, care environment, hospital effects, and sparse-day indicators are fused in a weighted multinomial transition model that yields daily leak, discharge, and death probabilities together with short-horizon leak risk.

cue preparedness. If a setting consistently recognizes low pelvic leak two days later than another setting after adjustment for measured physiology, that temporal difference is operationally important whether its proximate mechanism is delayed suspicion, delayed imaging, or subtle differences in postoperative monitoring.

4. Competing Risk Estimation

Let patient i be indexed over postoperative days $t = 1, \dots, 21$. On day t, the patient can transition from the event-free postoperative state into one of three absorbing outcomes: clinically recognized leak L, discharge without prior leak D, or death without prior leak M. Denote by Y_{it} the event realized on day t if the patient is still event-free at the start of that day. The central modeling task is to estimate the daily transition probabilities conditional on baseline covariates, evolving physiology, procedure family, care environment, and hospital clustering [12]. Because these outcomes are mutually exclusive within a day and share the same risk set, a multinomial discrete-time competing-risk formulation was used.

The day-specific transition probabilities were written as

$$\begin{aligned}\pi_{it\ell} &= P\big(Y_{it} = \ell \mid Y_{i,t-1} = 0, \mathcal{H}_{it}\big) \\ &= \frac{\exp\eta_{it\ell}}{1 \;\; r\in\{L,D,M\} \exp\eta_{itr}}\end{aligned} \tag{4.1}$$

for $\ell \in \{L, D, M\}$, where $\mathcal{H}_{it}$ denotes the observed postoperative history up to day t. The no-event state probability for day t is then the denominator's reciprocal. This multinomial parameterization was preferred over separate cause-specific logits because it enforces within-day coherence across leak, discharge, and death and simplifies the calculation of cumulative incidence under competing events.

The linear predictor for the leak transition included baseline patient factors, dynamic postoperative measurements, hospital random effects, and a structured interaction that allowed procedure family and care environment to modify hazard jointly over time:

$$\begin{aligned}\eta_{itL} = \alpha_{tL} \;\; & \mathbf{x}_i^\top \beta_L \;\; \mathbf{z}_{it}^\top \gamma_L \\ & u_{gi,L} \;\; v_{ci,L} \;\; b_{ki,L} \\ & p_{gi,L}^\top \Lambda_{tL} q_{ci,L} \;\; \rho_L m_{it}\end{aligned} \tag{4.2}$$

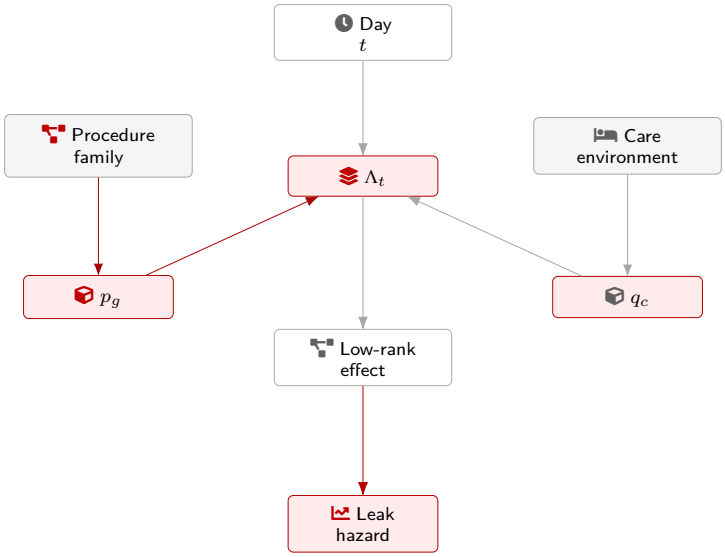


Figure 9: Low-rank interaction diagram. Procedure-family coordinates and care-environment coordinates are linked by a time-indexed interaction matrix, producing a compact day-specific effect that perturbs leak hazard without requiring a fully saturated procedure-by-environment-by-day expansion.

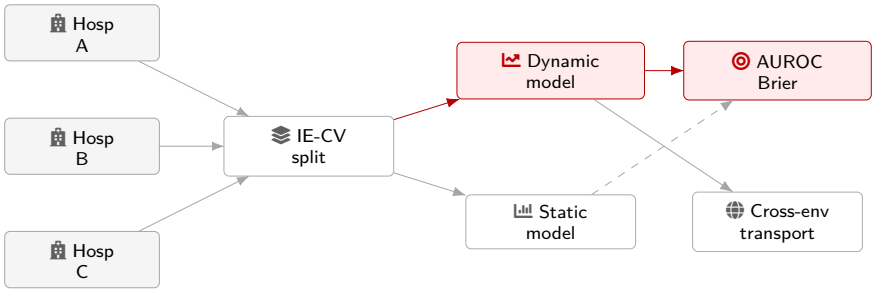


Figure 10: Validation and transport diagram. Hospital-block internal-external cross-validation compares the dynamic competing-risk model against a static benchmark and then evaluates whether gains in discrimination, calibration, and cross-environment transport persist outside the fitting hospitals.

Here α_{tL} is the unconstrained baseline hazard for postoperative day t, $\mathbf{x}_i$ contains fixed baseline covariates measured by the end of surgery, and $\mathbf{z}_{it}$ contains time-varying postoperative predictors. The term $u_{gi,L}$ is the main effect of procedure family gi, $v_{ci,L}$ is the main effect of care environment ci, and $b_{ki,L}$ is a hospital-level random intercept for hospital ki to account for residual clustering beyond the observed rescue environment classification. The quantity m_{it} denotes a vector of missingness indicators for key daily measurements, included so that sparse physiologic visibility could contribute information rather than being absorbed entirely into imputation. The most distinctive term is $p_{gi,L}^{\top}\Lambda_{tL}q_{ci,L}$, a low-rank time-varying interaction between procedure family and care environment. This tensor factorization permits the effect of care environment on leak hazard to vary across procedure families and postoperative days without estimating an unstable saturated interaction for every procedure-by-environment-by-day combination.

The same general structure was used for discharge and death transitions, though with separate parameters and less elaborate interaction for death because of event scarcity. The discharge process was expected to vary strongly by postoperative day and care environment, since ward-first pathways and lower-acuity procedures discharge earlier. The death process was dominated by baseline illness burden, contamination, emergent operation, and severe postoperative instability. Because the substantive focus lay on leak emergence, the low-rank interaction structure was fitted most flexibly for the leak transition, while the other transitions used simpler partial-pooling terms. This asymmetry is methodologically acceptable because

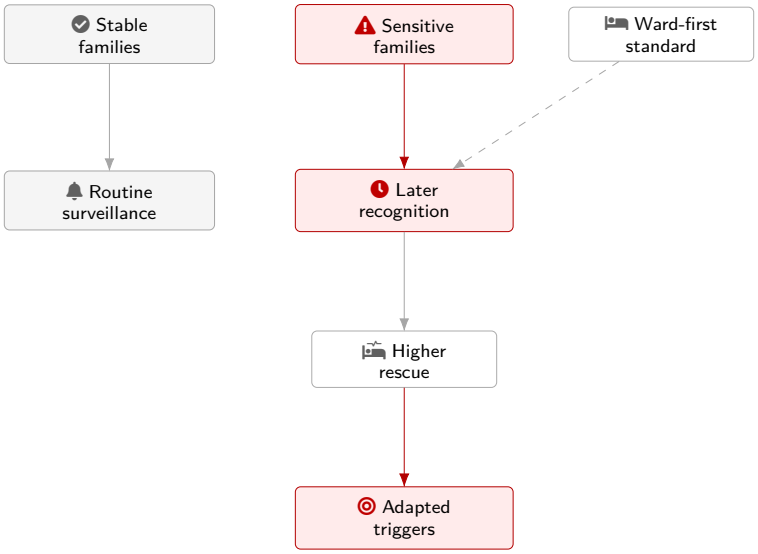


Figure 11: Clinical interpretation diagram. Procedure families with limited environment sensitivity map to routine surveillance, whereas setting-sensitive families accumulate later recognition and greater rescue burden in ward-first standard pathways, motivating context-adapted reassessment triggers rather than a single pooled leak rule.

competing events are included to preserve coherent risk-set accounting, not because the study's primary objective is to model all postoperative outcomes with equal granularity.

The low-rank interaction deserves explanation because it underpins the heterogeneity analysis. A full procedure-by-environment-by-day interaction would require estimating $5 \times 6 \times 21 = 630$ leak-specific deviations before considering shrinkage, which is feasible numerically but difficult to interpret and prone to instability in lower-frequency cells. Instead, each procedure family and care environment was assigned latent coordinates, and the time-varying interaction was expressed through a small matrix Λ_{tL} acting between them. In practice, a rank-2 representation was sufficient. This means that heterogeneity could be summarized by two latent axes varying over postoperative time. One axis contrasted procedure families whose leak trajectory was particularly sensitive to rescue environment, while the other contrasted families whose hazard shifted primarily with monitoring intensity and bed pathway. The low-rank approach thus achieved two goals simultaneously: regularization and interpretation [13].

To reduce bias from unequal physiologic visibility, inverse probability of observation weights were estimated for each patient-day. If $O_{it} = 1$ denotes that day t had a minimally informative physiologic panel, the stabilized weight was

$$w_{it} = \prod_{s=1}^{t} \frac{1}{\widehat{P}O_{is} = 1 \mid O_{i,s-1} = 1, \mathcal{G}_{is-1}} \tag{4.3}$$

where $\mathcal{G}_{is-1}$ contained history available before day s, including prior bed location, previous day physiology, procedure family, care environment, and actions suggestive of increasing clinical concern. These weights were multiplied across days to form the contribution of each patient-day to the weighted multinomial log-likelihood. Truncation at the 1st and 99th percentiles prevented a small number of sparsely observed trajectories from dominating estimation. Sensitivity analyses using unweighted estimation and simpler last-observation-carried-forward imputation were also performed to evaluate robustness.

The weighted objective function combined multinomial likelihood with penalization for spline smoothness and random-effect shrinkage:

$$\begin{aligned}\mathcal{L} = &- \sum_{i=1}^{N} \sum_{t=1}^{21} w_{it} \log\left(PY_{it} \mid \mathcal{H}_{it}\right) \\ &\lambda_{\mathrm{s}} \sum_{r} \|D\psi_r\|_2^2 \;\; \lambda_{\mathrm{u}} \sum_{g} u_{g,L}^2 \\ &\lambda_{\mathrm{v}} \sum_{c} v_{c,L}^2 \;\; \lambda_{\mathrm{b}} \sum_{k} b_{k,L}^2 \\ &\lambda_{\Lambda} \sum_{t=1}^{21} \|\Lambda_{tL}\|_F^2\end{aligned} \tag{4.4}$$

where the spline penalty governed nonlinear continuous covariate effects and the Frobenius norm penalty

on Λ_{tL} regularized the time-varying interaction matrices. Penalty tuning was selected using blockwise internal-external validation across hospitals, so that the chosen model balanced fit and transport rather than maximizing within-site performance alone.

From the fitted transition probabilities, cumulative incidence of leak for procedure family g in care environment c was derived as

$$S_{gc}t = \prod_{s=1}^{t} \left(1 - \pi_{gc,s,L} - \pi_{gc,s,D} - \pi_{gc,s,M}\right)$$
$$F_{gc}^{L}t = \sum_{s=1}^{t} S_{gc}s - 1\pi_{gc,s,L} \tag{4.5}$$

where $S_{gc}t$ is the probability of remaining under inpatient observation without leak, discharge, or death through day t. These quantities were computed after standardizing over the empirical distribution of baseline covariates and dynamic histories through Monte Carlo plug-in prediction. The resulting $F_{gc}^{L}t$ curves answer a clinically relevant question: if the same distribution of patients and postoperative physiologic trajectories were hypothetically placed into a given procedure family and care environment, how would their leak recognition accumulate over postoperative days under competing discharge and death?

Restricted mean day-to-detection among those leaking by day 21 was also computed:

$$\mu_{gc}^{L} = \frac{\sum_{s=1}^{21} s\,\Delta F_{gc}^{L}s}{F_{gc}^{L}21} \tag{4.6}$$

where $\Delta F_{gc}^{L}s$ is the increment in cumulative incidence on day s. This quantity is especially useful when comparing settings with similar cumulative incidence but different temporal concentration. A lower restricted mean day-to-detection indicates that leaks, conditional on occurring, are recognized earlier.

Model assessment prioritized dynamic metrics rather than single-horizon discrimination alone. The primary performance statistic was time-dependent area under the receiver operating characteristic curve for leak within the next forty-eight hours among patients still under inpatient observation. Integrated Brier score over postoperative days 1 through 14 quantified combined discrimination and calibration in the interval with highest event density. Calibration was examined by comparing predicted and observed cumulative incidence across deciles of risk within each care environment and by inspecting observed-to-expected ratios for day-7 and day-14 leak incidence. A static benchmark model was fitted for comparison using only baseline and postoperative day-1 information in a conventional thirty-day binary logistic framework [14]. This benchmark reflects the common clinical use case in which a single early postoperative risk score is computed once and not updated.

Two additional analyses were performed to probe the meaning of the procedure-by-environment interaction. First, a severity-restricted endpoint limited leaks to those followed by reoperation, image-guided drainage, or endoscopic intervention within forty-eight hours of recognition. This tested whether heterogeneity was confined to lower-grade events or persisted for failures requiring active source control. Second, post-detection rescue intensity was modeled among patients with leak, using ordinal regression for the maximum intervention level within seventy-two hours after recognition. This step was not part of the primary endpoint model, but it clarified whether delayed recognition in certain care environments coincided with greater downstream intervention burden. These analyses allowed the paper to move beyond incidence curves and ask whether procedure-specific differences in timing corresponded to clinically distinct rescue trajectories.

5. Results

Among 41,286 procedures, 2,521 clinically consequential leaks occurred within thirty days, for an overall incidence of 6.1%. The crude event distribution differed sharply across procedure families. Proximal colonic procedures had 416 leaks among 13,412 operations, corresponding to 3.1%. Distal colonic procedures had 454 leaks among 9,864 operations, corresponding to 4.6%. Low pelvic colorectal procedures had 812 leaks among 7,106 operations, corresponding to 11.4%. Small-bowel procedures had 281 leaks among 6,244 operations, corresponding to 4.5%. Upper gastrointestinal reconstructions had 558 leaks among 4,660 operations, corresponding to 12.0%. Death before documented leak occurred in 214 admissions, or 0.5% of the cohort, and was concentrated in emergency small-bowel and contaminated upper gastrointestinal cases. Median postoperative length of stay among patients without leak was 5 days for proximal colonic procedures, 6 days for distal colonic procedures, 8 days for low pelvic colorectal procedures, 7 days for small-bowel procedures, and 10 days for upper gastrointestinal reconstruction [15].

The temporal concentration of leak recognition was clinically informative. Across the full cohort, 76.0% of leaks were first recognized by postoperative day 7, 91.7%

Table 1: Cohort composition by procedure family

Procedure family	Procedures	Leaks (n)	Incidence (%)
Proximal colonic	13,412	416	3.1
Distal colonic	9,864	454	4.6
Low pelvic colorectal	7,106	812	11.4
Small-bowel	6,244	281	4.5
Upper gastrointestinal	4,660	558	12.0

Table 2: Care environment distribution

Environment	Count	Share (%)	Pathway type
ICU-first comprehensive	6,128	14.8	High observation
ICU-first standard	4,014	9.7	High observation
Step-down comprehensive	8,992	21.8	Intermediate
Step-down standard	7,486	18.1	Intermediate
Ward-first comprehensive	6,904	16.7	Low observation
Ward-first standard	7,762	18.8	Low observation

by day 14, and 97.6% by day 21. The overall median day of recognition was postoperative day 5 with an interquartile range of 4 to 8 days. However, this central summary concealed major family-level differences. Proximal colonic leaks were recognized earliest, with a median day of 4. Distal colonic and small-bowel leaks had median recognition on day 5. Low pelvic colorectal leaks had a median recognition day of 6, and upper gastrointestinal leaks a median of 5 but with a broader right tail extending into the second postoperative week. The baseline hazard curve for proximal colonic reconstruction peaked on days 3 and 4 before declining steadily, whereas low pelvic colorectal hazard rose more gradually and remained elevated through days 6 to 8. Upper gastrointestinal reconstruction showed a bimodal pattern, with one peak on days 3 to 4 and a second smaller shoulder on days 6 to 7. These curves were visible in crude estimates and persisted after multivariable adjustment.

Procedure family and care environment were strongly associated at baseline. ICU-first comprehensive pathways were dominated by upper gastrointestinal reconstruction and a selected subset of low pelvic colorectal cases with prolonged operations or major comorbidity. Ward-first standard pathways contained a large proportion of proximal and distal colonic procedures, but they also included a nontrivial number of low pelvic reconstructions, particularly in hospitals where routine postoperative ICU use was uncommon. Standard rescue environments cared for more emergency small-bowel procedures and had higher crude contamination rates. Comprehensive rescue environments cared for more elective cancer surgery and longer operations. These imbalances made crude comparisons of care environment difficult to interpret and reinforced the need for dynamic standardization [16].

The observation model confirmed large differences in physiologic visibility across pathways. On postoperative day 1, a minimally informative physiologic panel was present in 94.2% of ICU-first comprehensive records, 90.5% of ICU-first standard records, 81.7% of step-down-first comprehensive records, 76.3% of step-down-first standard records, 62.1% of ward-first comprehensive records, and 55.4% of ward-first standard records. By postoperative day 4, these differences widened further as stable ward patients were sampled less frequently. The strongest predictors of detailed observation on a given day were prior-day lactate elevation, persistent tachycardia burden, transfer to a higher-acuity bed, and upper gastrointestinal procedure family. Importantly, care environment remained a strong predictor of observation even after conditioning on these variables. This justified the weighting strategy and indicated that physiologic visibility was not reducible to evolving illness severity alone.

In internal-external validation across hospital blocks, the dynamic competing-risk model achieved a mean time-dependent area under the receiver operating characteristic curve of 0.812 for leak within the next forty-eight hours and an integrated Brier score of 0.046 over postoperative days 1 through 14. Calibration by care environment was stable, with day-7 observed-to-expected ratios ranging from 0.93 to 1.08 across the six environments. The static postoperative baseline model, which used only day-1 data and a single thirty-day binary outcome, achieved a mean area under the receiver operating characteristic curve of 0.741 and a Brier score of 0.058. More impor-

Table 3: Temporal distribution of leak recognition

Time point	Proportion (%)	Median day	IQR
By day 7	76.0		
By day 14	91.7	5	4–8
By day 21	97.6		
Overall	100	5	4–8

Table 4: Median day of leak detection by procedure

Procedure family	Median day	Early peak	Late tail
Proximal colonic	4	Yes	Minimal
Distal colonic	5	Moderate	Limited
Low pelvic colorectal	6	Gradual	Pronounced
Small-bowel	5	Moderate	Limited
Upper gastrointestinal	5	Bimodal	Present

tantly, the static model systematically underpredicted leak in low pelvic colorectal patients recovering in ward-first standard environments and overpredicted risk in proximal colonic patients managed in ICU-first comprehensive settings. In other words, the performance loss from ignoring temporal structure was not uniform. It was concentrated precisely where heterogeneity across procedure family and care environment was clinically largest.

The low-rank interaction structure captured a substantial share of this heterogeneity. A rank-2 time-varying interaction explained 92% of the out-of-sample deviance reduction obtained by a far less parsimonious full interaction. The first latent axis separated procedure families whose leak trajectories were strongly sensitive to rescue environment from those that were not. Low pelvic colorectal and upper gastrointestinal families loaded positively on this axis, while proximal colonic and small-bowel families loaded near zero. Care environments in standard rescue hospitals also loaded positively, especially when combined with ward-first recovery [17]. The second latent axis separated pathways with dense early monitoring from those with delayed recognition. ICU-first environments loaded negatively and ward-first environments positively, while step-down pathways lay between them. Over postoperative days 1 to 3, the first axis contributed more to hazard variation in upper gastrointestinal surgery, whereas over days 5 to 8 the second axis dominated in low pelvic colorectal surgery. This indicates that upper gastrointestinal heterogeneity was driven more by rescue environment early after surgery, while low pelvic colorectal heterogeneity manifested as delayed recognition under lower-observation pathways later in the week after surgery.

After covariate standardization over the empirical patient distribution, proximal colonic procedures exhibited limited sensitivity to care environment. Day-7 cumulative incidence ranged from 2.5% in ward-first comprehensive settings to 2.9% in ICU-first standard settings, with all six environments falling within a span of 0.4 percentage points. Day-21 cumulative incidence ranged from 2.8% to 3.3%. Restricted mean day-to-detection among proximal colonic leaks varied narrowly from 4.0 to 4.5 days. These findings suggest that, conditional on the kinds of baseline and dynamic physiologic patterns observed in the cohort, proximal colonic leak recognition is relatively stable across monitoring contexts. Distal colonic procedures showed slightly more variation, but still modest in absolute terms. Day-7 cumulative incidence ranged from 4.2% in step-down-first comprehensive settings to 4.9% in ward-first standard settings, and restricted mean day-to-detection ranged from 4.6 to 5.2 days.

The low pelvic colorectal family showed a different pattern. Standardized day-7 cumulative incidence was 9.1% in ICU-first comprehensive settings, 10.2% in ICU-first standard settings, 9.6% in step-down-first comprehensive settings, 10.8% in step-down-first standard settings, 11.3% in ward-first comprehensive settings, and 12.7% in ward-first standard settings. By day 21, these values increased to 10.0%, 11.4%, 10.8%, 12.3%, 12.9%, and 14.6%, respectively. The absolute difference between the two extreme environments at day 21 was therefore 4.6 percentage points, large enough to alter both surveillance burden and clinical expectations. Timing differed as much as cumulative incidence [18]. Restricted mean day-to-detection among low pelvic colorectal leaks was 4.4 days in ICU-first comprehensive settings and 6.8 days in ward-first standard settings. The largest increments in

Table 5: Model performance comparison

Model	AUROC	Brier score	Calibration range
Dynamic model	0.812	0.046	0.93–1.08
Static baseline	0.741	0.058	Wider variation
Improvement	+0.071	-0.012	Improved
Validation	Internal-external	Stable	Consistent

Table 6: Standardized day-21 incidence by environment (low pelvic)

Environment	Incidence (%)	Mean day	Pattern
ICU-comp	10.0	4.4	Early
ICU-std	11.4	5.1	Moderate
Step-comp	10.8	5.3	Moderate
Step-std	12.3	6.0	Delayed
Ward-comp	12.9	6.4	Delayed
Ward-std	14.6	6.8	Latest

leak recognition for ward-first standard low pelvic cases occurred on days 6 through 8, whereas ICU-first comprehensive cases showed a sharper concentration on days 3 through 5. This pattern is compatible with both delayed recognition and prolonged inpatient exposure in lower-observation settings, though the model cannot assign those mechanisms causal weights.

Upper gastrointestinal reconstruction also showed marked care-environment dependence, but with a temporal signature distinct from low pelvic colorectal surgery. Standardized day-7 cumulative incidence was 8.3% in ICU-first comprehensive settings, 9.5% in ICU-first standard settings, 8.7% in step-down-first comprehensive settings, 10.1% in step-down-first standard settings, 9.4% in ward-first comprehensive settings, and 10.9% in ward-first standard settings. By day 21 the range widened to 9.1% through 12.1%. Restricted mean day-to-detection ranged from 4.1 days in ICU-first comprehensive settings to 6.1 days in ward-first standard settings. Unlike low pelvic colorectal cases, upper gastrointestinal reconstructions retained a visible early hazard even in high-observation pathways, suggesting that some failures are driven by early conduit or perfusion vulnerability not easily altered by surveillance intensity. Nevertheless, comprehensive rescue environments appeared to blunt the later component of the hazard curve and reduced the accumulation of late-recognized cases.

Small-bowel procedures occupied an intermediate position. Standardized day-7 cumulative incidence ranged from 4.1% in ward-first comprehensive settings to 4.7% in ICU-first standard settings, and day-21 cumulative incidence ranged from 4.3% to 5.0%. Timing differences were modest, with restricted mean day-to-detection between 4.5 and 5.1 days. This family had substantial emergency burden and contamination, yet after adjustment its leak trajectory was less sensitive to bed pathway and rescue environment than low pelvic colorectal or upper gastrointestinal reconstruction. One interpretation is that the biology and clinical expression of failure in these cases are sufficiently overt that monitoring context has less room to alter recognition timing. Another possibility is that the emergency physiology driving ICU admission already dominates the signal, leaving less residual heterogeneity attributable to postoperative setting.

The dynamic coefficients for postoperative physiology aligned with clinical expectations but varied materially by procedure family [19]. Rising white blood cell count and persistent lactate elevation were associated with higher next-forty-eight-hour leak hazard in all families, but their magnitude was larger for upper gastrointestinal and low pelvic colorectal procedures. A one standard deviation increase in the daily white blood cell slope was associated with a 22% increase in leak odds over the next forty-eight hours in proximal colonic procedures, compared with 37% in low pelvic colorectal procedures and 34% in upper gastrointestinal reconstruction. Persistent lactate elevation across consecutive days increased leak odds by 18% in proximal colonic procedures but by 41% in upper gastrointestinal reconstruction. New tachycardia burden without concomitant lactate rise had limited incremental value in proximal colonic and distal colonic procedures yet remained informative in upper gastrointestinal reconstruction, likely reflecting different postoperative physiologic baselines. Failure to progress to any enteral intake by postoperative day 3 was independently associated with later leak recognition in upper gastroin-

Table 7: Standardized day-21 incidence by environment (upper GI)

Environment	Incidence (%)	Mean day	Pattern
ICU-comp	9.1	4.1	Early
ICU-std	10.4	4.8	Moderate
Step-comp	9.6	5.0	Moderate
Step-std	11.2	5.6	Delayed
Ward-comp	10.5	5.8	Delayed
Ward-std	12.1	6.1	Latest

Table 8: Observation density by care pathway (day 1)

Environment	Observed (%)	Monitoring level	Trend
ICU-comp	94.2	Very high	Stable
ICU-std	90.5	Very high	Stable
Step-comp	81.7	High	Declining
Step-std	76.3	Moderate	Declining
Ward-comp	62.1	Low	Sparse
Ward-std	55.4	Lowest	Sparse

testinal cases but not in low pelvic colorectal cases once other covariates were included.

Bed transfers provided additional signal beyond the initial environment assignment. Transfer from ward to step-down or ICU was associated with sharp short-term increases in leak hazard across all families, but the predictive meaning of transfer differed. In proximal colonic procedures, ward-to-ICU transfer often occurred in the context of nonleak cardiopulmonary instability and had lower positive predictive value for subsequent leak. In low pelvic colorectal and upper gastrointestinal procedures, upward transfer was more specifically associated with leak-related deterioration. However, even after incorporating transfer events and dynamic physiology, care-environment effects remained substantial for these families. This indicates that the observed heterogeneity was not merely a function of patients being moved to higher-acuity beds when they began to fail.

The severity-restricted endpoint yielded lower absolute incidence, as expected, but preserved the central pattern. When leak was limited to cases followed by drainage, endoscopic intervention, or reoperation within forty-eight hours, the overall event count fell to 1,756. Proximal colonic and distal colonic families showed little change in the ranking of care environments. Low pelvic colorectal day-21 cumulative incidence still ranged from 7.1% in ICU-first comprehensive settings to 10.9% in ward-first standard settings. Upper gastrointestinal reconstruction ranged from 6.9% to 9.4% [20]. The persistence of substantial differences under a stricter endpoint argues against the interpretation that care-environment heterogeneity was driven only by differential coding or heightened detection of lower-grade leaks. The gradient narrowed but remained clinically meaningful for the procedure families already identified as setting-sensitive.

Post-detection rescue intensity further clarified the consequences of timing differences. Among patients with leak, the proportion requiring invasive source control within seventy-two hours of recognition was 61.8% in comprehensive environments and 76.9% in standard environments after adjustment for procedure family and measured physiology at the time of recognition. This contrast was largest in ward-first pathways. Leak-associated mortality within thirty days of recognition was 6.4% in comprehensive environments and 10.9% in standard environments, with the widest gap in upper gastrointestinal reconstruction. These post-detection differences do not prove that delayed recognition caused worse outcomes, because hospitals and patients differed in many ways. They do, however, fit the broader temporal pattern: settings in which leak tended to be recognized later also showed greater rescue intensity after recognition.

Observation weighting had measurable effects on calibration. In unweighted models, ward-first standard pathways appeared to have slightly lower hazard on early postoperative days than in the primary analysis, followed by a steeper late increase. After weighting, the early hazard rose modestly and the late hazard attenuated, suggesting that sparse measurement had masked some earlier physiologic signal. The weighted model therefore supported a nuanced interpretation. Lower-observation environments did not create a completely different haz-

Table 9: Post-detection outcomes by rescue environment

Outcome	Comprehensive	Standard	Difference
Invasive intervention (%)	61.8	76.9	+15.1
Mortality (%)	6.4	10.9	+4.5
Early detection	More frequent	Less frequent	–
Severity at detection	Lower	Higher	–

ard shape from nothing; rather, they shifted a portion of risk recognition later than it might have appeared under more uniform physiologic visibility. Even after that correction, however, substantial late hazard remained for low pelvic colorectal and upper gastrointestinal procedures in standard rescue settings. Thus, unequal monitoring intensity explained part, but not all, of the procedure-by-environment heterogeneity.

The static benchmark model illustrates why timing matters analytically. When a single day-1 score was used to predict thirty-day leak, patients in low pelvic colorectal and upper gastrointestinal families were often assigned similarly elevated risk regardless of whether they entered ICU-first comprehensive or ward-first standard pathways [21]. That approach compressed heterogeneity into a family main effect. The dynamic model, by contrast, separated early risk concentration from delayed accumulation. Two patients could begin with similar day-1 risk but diverge substantially by day 4 as their environments generated different observation density, transfer patterns, and evolving physiologic trajectories. This divergence is not a peripheral modeling detail. It determines when imaging, antibiotic escalation, drain reassessment, or senior review might reasonably be triggered.

Overall, the results identify three broad empirical patterns. First, procedure family strongly shapes baseline leak hazard and its temporal concentration. Second, care environment matters unevenly, with low pelvic colorectal and upper gastrointestinal reconstruction displaying considerably more sensitivity than proximal colonic, distal colonic, or small-bowel procedures. Third, the most pronounced environmental differences concern not only how much risk accumulates, but when clinically recognized leak enters the record and how intensive rescue becomes after recognition. These patterns remained visible after extensive covariate adjustment, after accounting for competing discharge and death, after weighting for differential physiologic visibility, and after narrowing the endpoint to more severe leak events.

6. Discussion

The main contribution of this analysis is not simply the observation that some procedures leak more often than others. That fact is already well understood. What the present results show is that heterogeneity across surgery types and care settings has a temporal form that is obscured when leak is reduced to a static binary label. For proximal colonic and many distal colonic procedures, cumulative incidence and day-to-detection varied little across postoperative care environments after adjustment. For low pelvic colorectal and upper gastrointestinal reconstruction, by contrast, care environment changed both the amount of cumulative risk visible by a given postoperative day and the time at which clinically consequential leak was typically recognized. This distinction matters because a surveillance system designed around pooled thirty-day incidence will miss precisely the dimensions that clinicians must manage in real time [22].

The low pelvic colorectal pattern is especially instructive. Even after accounting for baseline condition, operative complexity, daily physiologic changes, transfer events, and differential observation density, ward-first pathways in standard rescue environments showed later recognition and higher cumulative incidence than ICU-first or step-down pathways in comprehensive environments. The most likely interpretation is not that the underlying biology of pelvic anastomotic failure becomes entirely different across hospitals. Rather, low pelvic leaks appear unusually sensitive to the combination of monitoring intensity and rescue readiness. Pelvic sepsis can emerge gradually, with inflammatory markers rising before overt hemodynamic collapse. If physiologic drift is sampled less often or if the threshold for imaging is operationally higher because overnight rescue is less available, formal recognition may shift by one or two days. In a complication where source control timing matters, that shift is not trivial.

Upper gastrointestinal reconstruction showed a related but not identical pattern. Unlike low pelvic colorectal surgery, upper gastrointestinal procedures retained substantial early hazard even in high-observation environ-

ments. This suggests that some failures are driven by early conduit perfusion, anastomotic tension, or other mechanisms that surveillance intensity cannot substantially prevent. However, the late component of the hazard curve still varied by environment, and post-recognition rescue burden was greater in standard settings. One plausible interpretation is that care environment does not eliminate the intrinsically high-risk component of upper gastrointestinal reconstruction, but it can influence how quickly evolving leak is confirmed and how severe the clinical state is by the time rescue begins. The distinction between altering biological incidence and altering clinical trajectory is difficult to draw sharply with observational EHR data, but it is precisely why time-resolved analysis is more informative than pooled incidence alone.

The relative stability of proximal colonic and small-bowel trajectories is equally important because it prevents overgeneralization. It would be easy to conclude from the setting-sensitive families that all gastrointestinal anastomoses require elaborate environment-specific surveillance logic. The data do not support that conclusion [23]. For procedures with lower baseline mechanical stress and more overt clinical expression of failure, temporal heterogeneity across care environments was limited after adjustment. In those families, a common surveillance framework may be sufficient, and aggressive differentiation by bed pathway would add complexity without much clinical return. This asymmetry suggests that the value of context-aware surveillance is itself procedure-specific. Precision in postoperative monitoring should therefore be selective rather than universal.

The observation-process adjustment helps clarify what kind of heterogeneity is being estimated. Dense ICU monitoring can make a model appear more accurate simply by supplying more repeated measurements. That alone does not imply a genuine difference in underlying leak trajectory. By weighting for the probability of detailed observation, the analysis reduced the extent to which richer monitoring mechanically inflated or deflated hazard estimates. The weighted results indicate that some early signal in ward-first pathways was indeed being underseen. Yet even after accounting for that, later hazard accumulation remained disproportionately high in low pelvic colorectal and upper gastrointestinal procedures managed in standard rescue environments. This implies that unequal surveillance density is only part of the explanation. The residual heterogeneity likely reflects a combination of delayed diagnostic action, longer dwell time before definitive imaging, and perhaps structural factors not fully visible in routine EHR data, such as staffing ratios, overnight senior coverage, or procedural experience with subtle early signs of failure.

The competing-risk structure also changes interpretation of discharge. In conventional binary models, early discharge is often invisible except indirectly through length of stay. Here, discharge is treated as a state transition that removes patients from inpatient surveillance. This matters because the day-by-day opportunity to recognize leak is partly determined by how quickly low-risk patients leave the hospital [24]. A setting that discharges proximal colonic patients early may have less late inpatient leak recognition without any true difference in biological failure. Conversely, a setting that keeps patients longer may observe more late events simply because the monitored window is extended. By explicitly modeling discharge as a competing event, the analysis avoids attributing all differences in cumulative inpatient leak recognition to hazard alone. The fact that low pelvic colorectal and upper gastrointestinal differences persisted after this accounting strengthens the argument that their heterogeneity is not merely an artifact of length-of-stay policy.

From a prediction standpoint, the superiority of the dynamic model over the static benchmark has practical implications. The static model was not useless. Day-1 information did identify high-risk families and high-risk physiologic states. But it flattened trajectories that diverged meaningfully after the first postoperative day. A clinically implemented system based only on an early fixed score would therefore miss the opportunity to re-estimate risk as postoperative evidence accumulates and as the influence of bed pathway becomes clearer. In high-sensitivity families, risk appears to be less a fixed quantity set at the end of surgery and more a path-dependent process. That does not mean every patient requires continuous algorithmic updating. It means that any surveillance framework treating day-1 risk as complete is likely to underperform where heterogeneity is most pronounced.

The rescue-intensity analysis adds a further layer of interpretation. Settings with later recognition also showed a higher probability of invasive intervention after recognition. This pattern does not establish causality, but it aligns with the intuitive clinical idea that delayed recognition may narrow the window for conservative management. If a contained collection is identified early, antibiotics, drainage, or limited endoscopic therapy may suffice. If diagnosis occurs later, operative source control may become more likely [25]. Even without causal proof, the coexistence of delayed recognition and greater rescue intensity suggests that time-to-detection is not merely a descriptive nicety. It has downstream operational consequences

relevant to staffing, interventional availability, and patient counseling.

These findings also bear on benchmarking and service design. Hospitals are often compared using pooled postoperative outcomes, but the present analysis indicates that such comparisons can be misleading in at least two ways. First, aggregate leak incidence obscures whether a hospital's profile is driven by a case mix concentrated in timing-sensitive procedures. Second, even within the same procedure family, care-environment composition matters. A service that routinely places low pelvic patients in ward-first pathways without comprehensive rescue access may face a different temporal risk profile than a service that uses step-down or ICU-first recovery for similar patients. Thus, benchmarking should incorporate both procedure family and postoperative environment rather than relying on a single hospital-wide leak rate.

Several operational implications follow. For low pelvic colorectal and upper gastrointestinal reconstruction, surveillance thresholds may need to be lower in ward-first standard environments than in ICU-first or comprehensive settings, because the expected day-to-detection is later and the post-recognition rescue burden is greater. That does not necessarily imply more imaging for every patient. It may instead support a structured trigger system emphasizing repeated inflammatory trends, unexplained tachycardia burden, failed enteral progression, or early transfer decisions. For proximal colonic and small-bowel procedures, by contrast, there is less evidence that setting-specific thresholds would materially improve detection efficiency. The practical message is differentiation, not maximalism. Resources should be concentrated where the trajectory is most sensitive to context.

The study also illustrates a broader point about EHR-based surgical modeling [26]. When the clinical phenomenon of interest is shaped by observation intensity, one cannot treat missingness and measurement frequency as merely technical inconveniences. They are part of the underlying system. The bed pathway determines not only who is being watched, but what can be known at each time point. This is especially true for complications like leak, where subtle physiologic changes often precede definitive diagnosis. Future work in this area would benefit from making the observation process explicit rather than assuming that data absence is random or that dense monitoring can be generalized without adjustment. The current results suggest that doing so changes substantive conclusions, not just statistical precision.

There are also limits to what can be learned from this design. The model tracks clinical recognition, not the exact biological onset of failure. As a result, care-environment effects combine differences in underlying hazard with differences in detection timing. For surveillance planning, that combined estimand is useful. For causal claims about tissue failure, it is insufficient. Likewise, hospital rescue environment was summarized using a binary operational descriptor. This captured meaningful variation in availability of imaging and intervention, but not all relevant aspects of postoperative care. Technical skill, nursing expertise, overnight staffing, protocol adherence, and surgeon volume are only partially visible through the available covariates. Some of the residual environment sensitivity almost certainly reflects these unmeasured factors.

Even with those limitations, the overall empirical pattern is difficult to dismiss [27]. Heterogeneity across surgery types and care settings is not just a matter of some procedures starting at a higher point on the same risk scale. The shape of the postoperative trajectory changes as well. In some procedure families, care environment exerts little influence beyond what is already explained by baseline risk and routine physiology. In others, especially low pelvic colorectal and upper gastrointestinal reconstruction, environment shifts the hazard curve forward or backward in postoperative time and changes the burden of rescue after recognition. That is a different kind of heterogeneity from the one usually summarized by odds ratios or aggregate event rates, and it is the kind most relevant to postoperative monitoring strategy.

7. Conclusion

A time-resolved competing-risk analysis showed that anastomotic leak heterogeneity across gastrointestinal surgery types and care settings is expressed through both cumulative incidence and day-to-detection. Proximal colonic, distal colonic, and most small-bowel reconstructions exhibited relatively stable trajectories across postoperative environments after adjustment. Low pelvic colorectal and upper gastrointestinal reconstructions did not. In those families, ward-first pathways and standard rescue environments were associated with later recognition, greater cumulative inpatient leak incidence, and higher post-recognition rescue intensity.

These findings suggest that leak surveillance should be organized around procedural trajectory rather than a single pooled thirty-day probability. A dynamic framework that updates risk over postoperative days and accounts for competing discharge and death captures clinically relevant differences that a one-time baseline score

compresses. The practical implication is selective calibration. High-sensitivity procedure families may benefit from care-environment-specific triggers for reassessment, imaging, or escalation, whereas lower-sensitivity families may not require such differentiation.

When postoperative data density varies systematically across care environments, the observation process becomes part of the phenomenon being studied. Representing time, competing outcomes, and observation intensity together yields a more credible account of how leak becomes visible in routine practice. Future work can build on this structure by adding surgeon-level effects, protocol adherence measures, and prospective implementation testing, but the present analysis already indicates that the timing of leak recognition is a central dimension of heterogeneity, not a secondary detail [28].

References

[1] X. Serra-Aracil, A. García-Nalda, B. Serra-Gómez, A. Serra-Gómez, L. Mora-López, A. Pallisera-Lloveras, V. Lucas-Guerrero, and S. Navarro-Soto, "Experimental study of the quantification of indocyanine green fluorescence in ischemic and non-ischemic anastomoses, using the sergreen software program.," *Scientific reports*, vol. 12, pp. 13120–13120, 7 2022.

[2] 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.

[3] R. Chen, H. Jiang, W. Jiang, K. Luo, H. Zhang, and F. Gao, "Effects of different inferior mesenteric artery ligation levels on the prognosis of patients with low rectal cancer.," *Advances in clinical and experimental medicine : official organ Wroclaw Medical University*, vol. 32, pp. 733–739, 6 2023.

[4] F. Al-Herz, D. Healey, T. Sammour, J. Turagava, B. Rhind, and M. Young, "Short and long term outcomes of oesophagectomy in a provincial new zealand hospital.," *The New Zealand medical journal*, vol. 125, pp. 30–39, 4 2012.

[5] K. Pather, A. D. Ghannam, S. Hacker, C. Guerrier, E. M. Mobley, R. Esma, and Z. T. Awad, "Reoperative surgery after minimally invasive ivor lewis esophagectomy.," *Surgical laparoscopy, endoscopy & percutaneous techniques*, vol. 32, pp. 60–65, 9 2021.

[6] M.-A. Martínez-Serrano, D. Parés, M. Pera, M. Pascual, R. Courtier, M. J. G. Egea, and L. Grande, "Management of lower gastrointestinal bleeding after colorectal resection and stapled anastomosis.," *Techniques in coloproctology*, vol. 13, pp. 49–53, 3 2009.

[7] A. Ramirez, M. Roy, J. E. Hidalgo, S. Szomstein, and R. Rosenthal, "Outcomes of bariatric surgery in patients >70 years old," *Surgery for obesity and related diseases : official journal of the American Society for Bariatric Surgery*, vol. 8, pp. 458–462, 4 2012.

[8] H. K. Bangash, M. Rahmatzeh, and B. P. Mwipatayi, "Successful endovascular management of endoleak-like phenomenon following open abdominal aortic aneurysm repair.," *International journal of surgery case reports*, vol. 29, pp. 168–171, 11 2016.

[9] S. K. Kamarajah, A. Madhavan, J. Chmelo, M. Navidi, S. Wahed, A. Immanuel, N. Hayes, S. M. Griffin, and A. W. Phillips, "Impact of smoking status on perioperative morbidity, mortality, and long-term survival following transthoracic esophagectomy for esophageal cancer," *Annals of surgical oncology*, vol. 28, pp. 4905–4915, 3 2021.

[10] J. Machowicz, M. Wołkowski, B. Jabłońska, and S. Mrowiec, "Ileocolonic anastomosis-comparison of different surgical techniques: A single-center study.," *Medicine*, vol. 101, pp. e31582–e31582, 12 2022.

[11] B. Allin, M. Knight, P. Johnson, and D. M. Burge, "Outcomes at one-year post anastomosis from a national cohort of infants with oesophageal atresia.," *PloS one*, vol. 9, pp. e106149–, 8 2014.

[12] A. Broadbent, S. Rahman, B. Grace, R. Walker, F. Noble, J. Kelly, J. Byrne, and T. Underwood, "The effect of surgical complications on long-term prognosis following oesophagectomy.," *European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology*, vol. 49, pp. 106930–106930, 5 2023.

[13] L. Park, K. Zaghiyan, and P. Fleshner, "Transanal ileal pouch: Is it better?," *Journal of laparoendoscopic & advanced surgical techniques. Part A*, vol. 31, pp. 898–904, 6 2021.

[14] J. C. Sebastián-Tomás, A. Martínez-Pérez, E. Martínez-López, N. de'Angelis, M. G. Ruiz, and E. García-Granero, "Robotic transanal total mesorectal excision: Is the future now?," *World journal of gastrointestinal surgery*, vol. 13, pp. 834–847, 8 2021.

[15] R. Jaibaji, M. Ibrahim, A. Tsai, and C. Peters, "Ogc p67 extended use of feeding jejunostomy and reduction of weight loss in patients with total gastrectomy for malignancy," *British Journal of Surgery*, vol. 110, 11 2023.

[16] L. Castagneto-Gissey, M. F. Russo, J. Casella-Mariolo, A. Serao, R. Marcellinaro, V. D'Andrea, M. Carlini, and G. Casella, "The role of antibiotic prophylaxis in anastomotic leak prevention during elective colorectal surgery: Systematic review and meta-analysis of randomized controlled trials.," *Antibiotics (Basel, Switzerland)*, vol. 12, pp. 397–397, 2 2023.

[17] L. Zhong, J. Zhong, Z. Tan, Y. Wei, X. D. Su, Z.-S. Wen, T. Rong, Y. Hu, and K. Luo, "An approach to accelerate healing and shorten the hospital stay of patients with anastomotic leakage after esophagectomy: An explorative study of systematic endoscopic intervention.," *Frontiers in oncology*, vol. 11, pp. 657955–657955, 5 2021.

[18] A. Montedori, R. Cirocchi, E. Farinella, F. Sciannameo, and I. Abraha, "The cochrane library - covering ileo- or colostomy in anterior resection for rectal carcinoma," *The Cochrane database of systematic reviews*, vol. 2010, pp. CD006878–, 5 2010.

[19] L. Huang, Y. Yin, Y. Liao, J. Liu, K. Zhu, X. Yuan, L. Xue, and H. Pan, "Risk factors for postoperative urinary retention in patients undergoing colorectal surgery: a systematic review and meta-analysis.," *International journal of colorectal disease*, vol. 37, pp. 2409–2420, 11 2022.

[20] J. Deng, Q. Su, Z. Ren, J. Wen, Z. Xue, L. Zhang, and X. Chu, "Comparison of short-term outcomes between minimally invasive mckeown and ivor lewis esophagectomy for esophageal or junctional cancer: a systematic review and meta-analysis.," *OncoTargets and therapy*, vol. 11, pp. 6057–6069, 9 2018.

[21] A. Wallace, S. Houlton, and J. Garner, "Gastrointestinal procedures and anastomoses can be safely performed during complex abdominal wall reconstruction.," *Hernia : the journal of hernias and abdominal wall surgery*, vol. 27, pp. 439–447, 11 2022.

[22] Y. Ishikawa, A. Hider, P. Zhang, B. Askar, A. Haveric, A. Brandser, K. Lagisetty, E. Wakeam, W. Lynch, B. Enchakalody, J. Sullivan, M. Orringer, A. Chang, S. Wang, R. Reddy, and J. Lin, "290. analytic morphomic index and quantitative gastric perfusion assessment for risk stratification of anastomotic leaks following esophagectomy," *Diseases of the Esophagus*, vol. 35, 9 2022.

[23] C. E. Graves, C. Co, R. S. Hsi, D. Kwiat, J. Imamura-Ching, M. R. Harrison, and M. L. Stoller, "Magnetic compression anastomosis (magnamosis): First-in-human trial," *Journal of the American College of Surgeons*, vol. 225, pp. 676–681.e1, 8 2017.

[24] G. Valero, J. Luján, Q. Hernández, M. de las Heras, E. Pellicer, A. Serrano, and P. Parrilla, "Neoadjuvant radiation and chemotherapy in rectal cancer does not increase postoperative complications.," *International journal of colorectal disease*, vol. 18, pp. 495–499, 7 2003.

[25] B. Zhan, J. Chen, S. Du, Y. Xiong, and J. Liu, "Using the hand-sewn purse-string stapled anastomotic technique for minimally invasive ivor lewis esophagectomy.," *The Thoracic and cardiovascular surgeon*, vol. 67, pp. 578–584, 6 2018.

[26] J. Gardner-Thorpe, R. H. Hardwick, and S. J. Dwerryhouse, "Salvage oesophagectomy after local failure of definitive chemoradiotherapy.," *The British journal of surgery*, vol. 94, pp. 1059–1066, 7 2007.

[27] D. S. Kumar, S. N. Noushad, M. P. Viswanathan, and V. D. S, "Surgical and survival outcomes of operable gastric cancer-experience from a tertiary care center in south india.," *Indian journal of surgical oncology*, vol. 14, pp. 48–52, 8 2022.

[28] Z. Liao, W. Zheng, W. Huang, and L. Hu, "The value of ionic-contrast esophagogram in diagnosis of intrathoracic anastomotic leakage after oesophagectomy," *Chinese Journal of Primary Medicine and Pharmacy*, vol. 24, pp. 1985–1988, 7 2017.