Phenotype Atlas Construction and Border-Zone Outcome Definition for Anastomotic Leak in Longitudinal Surgical Electronic Health Records

Rizky Pratama[1] **and Dimas Saputra**[2]

[1]Universitas Trunojoyo Madura, Faculty of Engineering, Department of Informatics Engineering, Jalan Raya Telang, Kamal, Bangkalan 69162, East Java, Indonesia

[2]Universitas Siliwangi, Faculty of Engineering, Department of Computer Science, Jalan Siliwangi No. 24, Kahuripan, Tawang, Tasikmalaya 46115, West Java, Indonesia

ABSTRACT

Background work in postoperative informatics has expanded access to richly structured surgical records, yet the difficulty of defining clinically important complications has not diminished as datasets have grown. Anastomotic leak remains one of the clearest examples of this problem because it is recognized through evolving combinations of local findings, physiologic deterioration, procedural response, imaging practice, and delayed documentation rather than through a single stable observation. As a result, a computational leak label often compresses a much richer clinical process into a binary endpoint that hides ambiguity, severity variation, and institutional differences in recognition pathways. This paper develops a technical framework for outcome definition and anastomotic leak phenotyping based on the idea of a phenotype atlas. Instead of beginning with a fixed positive class and later subdividing it, the framework starts from a map of clinically meaningful postoperative states, including contained leak, progressive leak, catastrophic leak, unresolved leak-like courses, and non-leak mimics that occupy adjacent diagnostic territory. Outcome definition is then treated as the problem of deciding where the event boundary should be placed within that atlas for a particular purpose. The discussion emphasizes temporal narrative reconstruction, border-zone cases, annotation discipline, uncertainty retention, and cross-setting interpretation. It argues that the most consequential methodological question is not whether a dataset contains enough variables to train a predictive model, but whether the states represented by those variables have been organized in a way that preserves clinically meaningful distinctions. A phenotype-atlas approach offers a route to more transparent labels, more interpretable model behavior, and more realistic handling of difficult postoperative trajectories.

1. Introduction

Anastomotic leak is one of the most conceptually unstable outcomes in surgical data science. Everyone involved in perioperative care understands that the event is important, feared, and clinically heterogeneous, yet computational work often acts as though it were a simple target that exists in the record waiting to be retrieved. In reality, the record contains a changing mixture of observations, interpretations, tests, interventions, and abstractions that only imperfectly reveal the clinical state of the anastomosis. A leak may first appear as unexplained tachycardia, then as persistent inflammatory elevation, then as a decision to obtain imaging, then as a drain placement, then as a retrospective diagnosis code. Another leak may become obvious intraoperatively during urgent re-exploration without passing through that same chain. A third may remain clinically suspected but never fully

Article info:
Rizky Pratama and Dimas Saputra. *Phenotype Atlas Construction and Border-Zone Outcome Definition for Anastomotic Leak in Longitudinal Surgical Electronic Health Records.* Center for Theoretical and Applied STEM Research (CTASR) . Licensed under Attribution - NonCommercial - No Derivatives 4.0 International (CC BY-NC-ND 4.0)

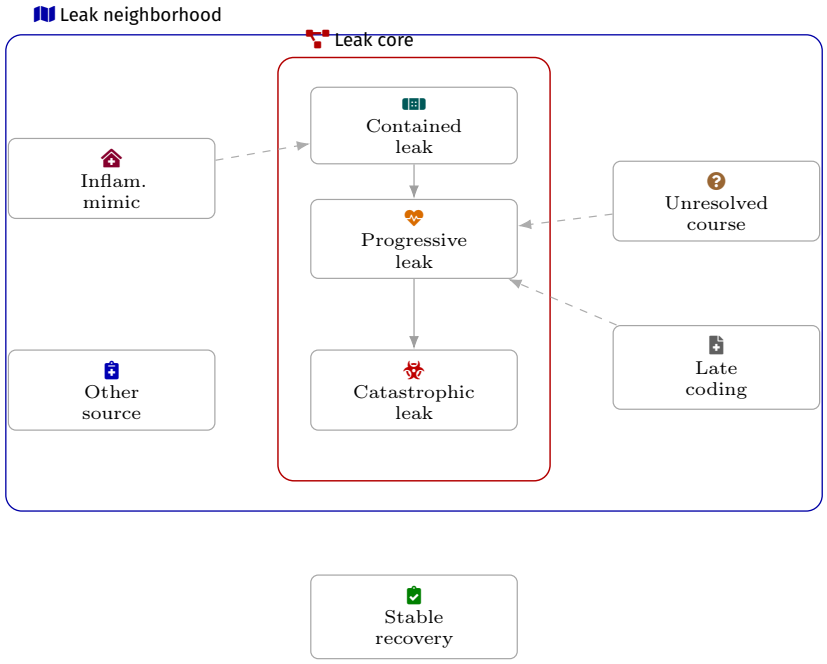


Figure 1 Phenotype-atlas overview for postoperative leak territory. The central leak region is separated from adjacent mimic, unresolved, and documentary states so that endpoint design becomes an explicit boundary-placement step rather than a hidden threshold rule.

Table 1 Core Phenotype Regions in the Atlas

Region	Clinical Profile	Evidence Pattern	Typical Management
Contained Leak	Localized failure	Imaging + mild physiology	Drainage/antibiotics
Progressive Leak	Worsening inflammation	Sequential deterioration	Escalation of care
Catastrophic Leak	Diffuse contamination	Acute instability	Urgent surgery
Unresolved State	Suspicion without proof	Sparse/mixed signals	Observation
Mimic State	Non-leak pathology	Overlapping features	Treat alternative cause

resolved in the chart. When these courses are flattened into a single positive class, the resulting outcome looks crisp even though the underlying clinical reality is graded, delayed, and path dependent.

This instability creates a familiar but underexamined problem. Models trained to predict AL are evaluated against targets that may represent very different things. One target may correspond to clinically severe leak requiring intervention. Another may correspond to any postoperative course that eventually accrues enough documentary support to be labeled as leak. Another may aim at earlier surveillance and therefore privilege subtle precursor states over later certainty. These targets are related but not equivalent [1]. A model that performs well against one of them is not automatically a good model of the others. Yet studies often discuss their results as if they addressed one common phenomenon. In doing so, they obscure the fact that outcome definition is not a neutral preprocessing step. It is the step in which the field decides what kind of leak it is willing to count.

The routine response to this problem has been to write a composite definition. Composite definitions are useful, and often necessary, because no single code or event reliably captures AL in large-scale records. However, composites usually remain threshold rules rather than conceptual frameworks. They specify what counts as positive, but they rarely describe the surrounding topography of near-positive, ambiguous, or clinically adjacent states. That omission matters because most of the difficult methodological questions live near the boundary, not in the obvious cases. There is usually little dispute about the patient with frank anastomotic disruption at reoperation and diffuse contamination. There is also little dispute about the uncomplicated recovery. The challenge lies in the middle zone: contained collections, unresolved inflammatory patterns, late documentation without strong antecedent evidence, and mimic syndromes that trigger similar workups. A binary definition has to de-

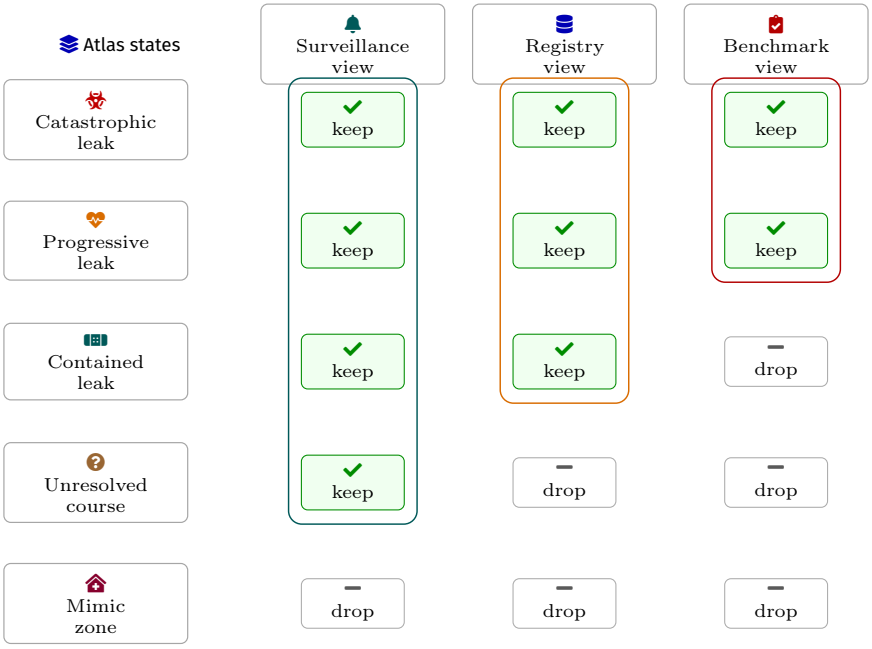


Figure 2 Task-specific binary projections from the same atlas. Surveillance reaches furthest into unresolved territory, registry keeps the managed leak spectrum, and benchmarking uses the narrowest coverage around clearly supported leak states.

Table 2 Temporal Narrative Phases in Postoperative Courses

Phase	Timing	Key Signals	Interpretation Role
Early Signals	POD 0–2	Mild labs, vitals	Low specificity
Emerging Concern	POD 2–5	Persistent abnormalities	Suspicion building
Escalation	Mid-course	Imaging, antibiotics	Turning point
Confirmation	Later	Procedure/code	Diagnostic closure
Resolution	Late	Recovery or decline	Outcome clarity

cide what to do with those cases, but if it does so silently, neither users nor readers can see which distinctions have been suppressed.

The central idea of this paper is that AL outcome definition should begin not with a binary rule but with a phenotype atlas. A phenotype atlas is a structured map of postoperative states that occupy the leak neighborhood. It includes states that are conventionally positive, states that are clearly negative but clinically adjacent, and states that are uncertain or intermediate. The atlas is not simply a taxonomy. It is a way of organizing the interpretive space within which the event boundary will later be placed. Once that space is explicit, one can ask where the boundary should lie for different purposes, which areas of the map are heavily supported by routine evidence, which are weakly observed, and which are frequently confused with one another. Under this view, binary labels become projections from the atlas rather than the primary object of analysis.

This approach changes the order of reasoning in a useful way. Instead of starting from a list of positive criteria and then trying to explain why certain errors occur, one starts by asking what kinds of postoperative courses are clinically and observationally distinct enough to merit separate representation. Only after that map is built does one ask which regions should be grouped together for surveillance, for benchmarking, or for comparative study. The result is a more transparent relationship between label design and clinical meaning. It also creates a framework in which uncertainty is not an embarrassing remainder but part of the map itself. Cases can occupy border zones without being forced immediately into either positive or negative status.

The atlas perspective also addresses a deeper problem with how EHR-based complication research often treats evidence. Not every piece of evidence carries the same significance, and not every difference in evidence corresponds to a difference in disease state. A patient may have more documented evidence because clinicians investigated more intensively, because the institution documents more thoroughly, or because the phenotype declared itself differently. Those possibilities are often

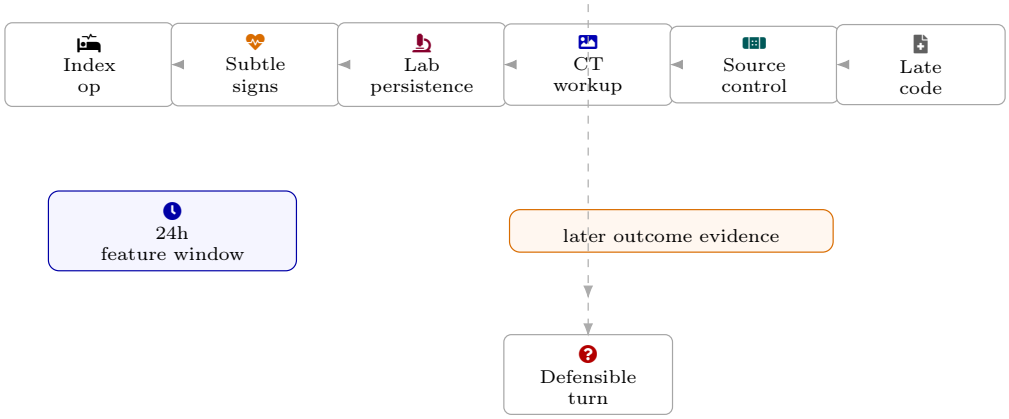


Figure 3 Temporal narrative reconstruction of a leak-like course. Early prediction features are separated from later outcome-defining evidence, and the first defensible turning point is marked explicitly instead of being collapsed into final retrospective certainty.

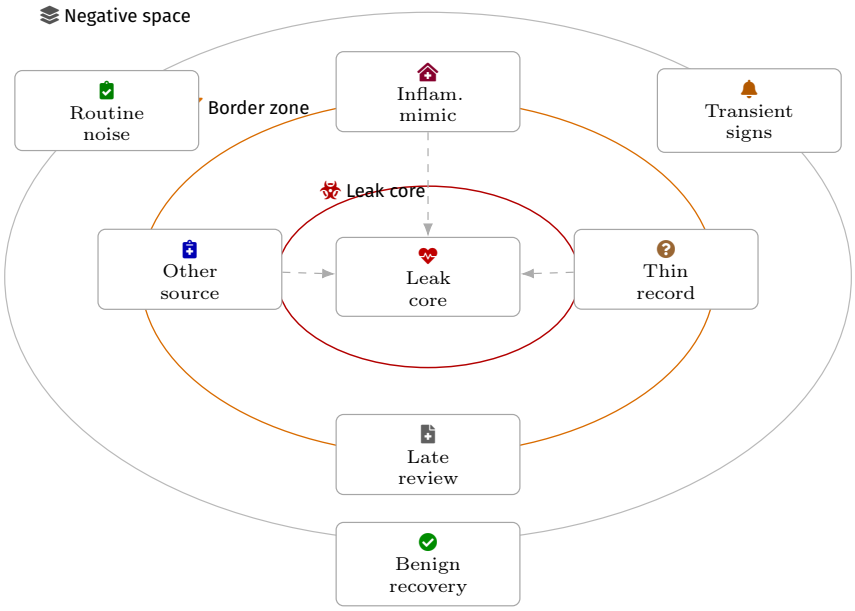


Figure 4 Structured negative space around leak. Border-zone negatives are represented as clinically near states rather than one residual class, which changes how false alarms, uncertainty, and apparent model error should be interpreted.

entangled [2]. If a model is trained on labels that ignore this entanglement, it may become a predictor of observational pathways rather than of the latent postoperative states of interest. An atlas does not solve that automatically, but it does create a place to represent different routes of appearance separately from the states those routes are meant to reveal.

Another advantage of the atlas framework is that it integrates well with human annotation and expert review. Clinicians do not naturally think only in terms of positive and negative classes. They recognize contained leaks, free leaks, delayed or occult failures, inflammatory mimics, suspicious but unresolved courses, and other nuanced states. Adjudication becomes easier and more honest when reviewers can place a case within such a map rather than being forced to answer a binary question too early. This does not mean that every study must use all atlas states in its final model. It means that the process of constructing a final endpoint benefits from a richer representational stage.

The sections that follow develop this idea in a form suitable for highly technical work with routine surgical records but without relying on formal mathematical notation. The discussion first describes the atlas as an organizing principle for leak-related postoperative states. It then turns to temporal narrative case construction, emphasizing that AL is often best understood through sequence and trajectory rather than through isolated variables. It next examines the border zone of difficult, ambiguous, and mimic cases, because that zone is where phenotype-aware outcome design is most needed. The paper then treats annotation and review as scientific ac-

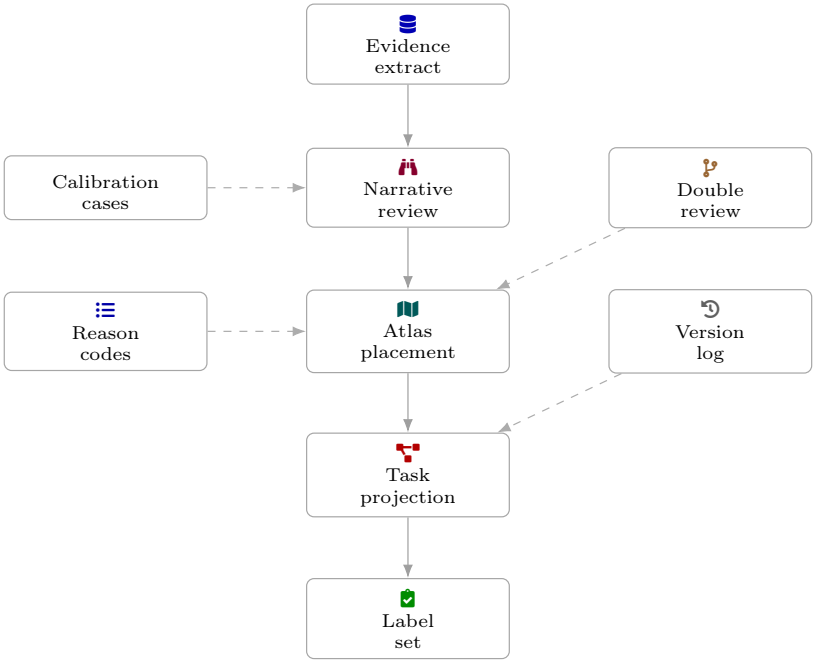


Figure 5 Annotation workflow for phenotype-atlas construction. Evidence extraction, narrative reconstruction, atlas placement, and task projection are kept distinct while calibration, reason coding, disagreement review, and protocol versioning remain visible parts of the scientific process.

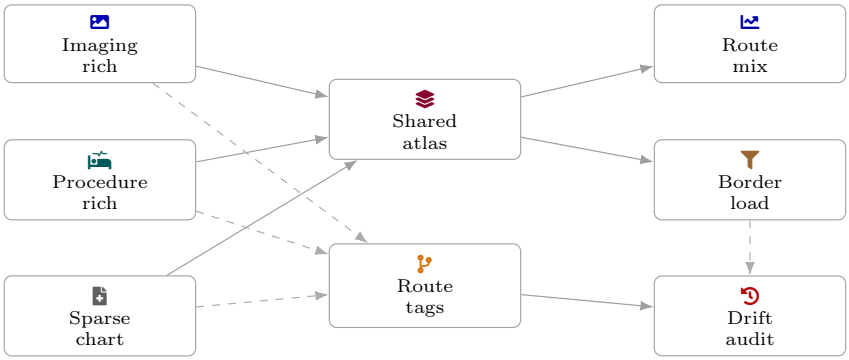


Figure 6 Cross-setting translation and drift monitoring. Different local evidence pathways feed a shared atlas, while route tagging and atlas-occupancy review make institutional variation and temporal drift auditable rather than silent.

tivities in their own right, rather than as hidden support work for modeling. After that, it considers how phenotype atlases should travel across institutions and across changing documentation regimes. The overall goal is to provide a different route to AL outcome definition, one that is less dependent on the fiction of a naturally occurring binary event and more attentive to the complex landscape of postoperative states from which such events are operationally drawn [3].

2. The Phenotype Atlas as an Organizing Principle

A phenotype atlas is best understood as a structured map rather than as a checklist. It does not begin by asking whether a patient is positive or negative for leak. It begins by asking what clinically meaningful postoperative states appear near the conceptual territory of leak, how those states differ in manifestation and trajectory, and which of them are easy or difficult to observe in routine records. The purpose of the atlas is to make the terrain visible before the event boundary is imposed. This is a reversal of the usual workflow, but it is a productive one because it forces the definition problem to be framed in terms of state space rather than in terms of arbitrary thresholding over codes and procedures.

The most immediate consequence of this shift is that leak is no longer treated as a monolithic category. Instead, it occupies a region that contains several subterritories. One part of that region may correspond to contained local failures, where the clinical picture is dominated by pelvic or peri-anastomotic collection, modest

Table 3 Border-Zone Case Categories

Category	Description	Evidence Strength	Common Issue
Near-Leak	Almost meets criteria	Moderate	Missing confirmation
Ambiguous	Conflicting signals	Mixed	Attribution unclear
Late-Code	Retrospective labeling	Weak early	Timing distortion
Mild Course	Limited severity	Low	Under-captured
Overlap Mimic	Competing diagnosis	Moderate	Misclassification

Table 4 Mimic States Near Leak Territory

Mimic Type	Clinical Features	Overlap Dimension	Distinguishing Clue
Ileus	Delayed bowel function	Inflammation	No focal source
Non-AL Infection	Intra-abdominal infection	Sepsis pattern	Different site
Bleeding	Hemodynamic change	Instability	Hemoglobin drop
Pulmonary Complication	Hypoxia, stress	Systemic response	Imaging chest
Ischemia	Tissue compromise	Pain/inflammation	Vascular evidence

or evolving physiologic disturbance, and management through drainage, antibiotics, or close observation. Another part may correspond to progressive leaks, where inflammatory and hemodynamic abnormalities intensify over time and the patient is moving toward more obvious source-control needs. Another may correspond to catastrophic or disseminated failure, where contamination is generalized, operative findings are overt, and the event is recognized through major clinical deterioration. These regions are linked, but they are not equivalent. They differ in onset, evidence density, management response, and risk of misclassification.

The atlas must also include adjacent non-leak territory. This is not a matter of politeness toward negatives. It is methodologically essential [4]. The cases that most often destabilize outcome definitions are not uneventful recoveries. They are conditions that overlap with leak clinically and operationally. Persistent ileus with inflammatory response, intra-abdominal infection not clearly attributable to the anastomosis, postoperative bleeding with secondary physiologic derangement, pulmonary complications producing systemic deterioration, and generalized sepsis from another abdominal source all belong close to the leak neighborhood. In binary datasets these cases are often treated as residual error. In a phenotype atlas they become named regions with their own observational signatures. This reframing helps explain why certain false positives recur and why the positive class sometimes looks heterogeneous. The issue is not only label noise. It is neighborhood structure.

A well-designed atlas also includes unresolved territory. Not every admission can be responsibly placed in a definitive phenotype. Some charts support suspicion but not confirmation. Some contain evidence that points in several directions at once. Some are too thinly observed to justify strong claims. These states are often forced into the negative class because analysts prefer clean denominators. That decision is understandable, but it is not neutral [5]. It effectively asserts that lack of decisive evidence is equivalent to absence of leak. In many postoperative datasets that equivalence is untenable. An atlas solves this by providing a region for unresolved or indeterminate courses. The existence of that region does not weaken the system. It makes the system more faithful to what the record can actually support.

The atlas perspective changes how outcome definitions are evaluated. Instead of asking only whether the binary event rule has reasonable face validity, one asks which regions of the map it captures and which it excludes. A narrow rule may cover only the more severe or intervention-defined areas of leak territory. A broader rule may extend into contained and unresolved states. A surveillance-oriented rule may intentionally include some border-zone cases because the cost of delayed recognition is high. A benchmarking rule may intentionally exclude them to improve specificity and comparability. Once the atlas is explicit, these choices can be discussed as decisions about territorial coverage rather than hidden inside a single performance statistic. This is a more transparent language for the real trade-offs involved.

Another important benefit is that the atlas separates state identity from route of recognition. A contained leak may become visible through early imaging in one institu-

Table 5 Annotation Workflow Stages

Stage	Task	Output	Purpose
Extraction	Identify signals	Event list	Data grounding
Narrative Build	Sequence events	Timeline	Context creation
Atlas Mapping	Assign region	Phenotype label	State placement
Projection	Binary decision	Outcome label	Task-specific use
Review	Resolve conflicts	Consensus/meta-data	Quality control

Table 6 Reason Codes for Phenotype Assignment

Reason Type	Example Signal	Role	Reliability
Imaging	CT collection	Structural evidence	High
Procedure	Drain/reoperation	Intervention marker	High
Labs	Persistent elevation	Supportive	Moderate
Antibiotics	Escalation	Response signal	Moderate
Coding	Diagnosis entry	Summary evidence	Variable

tion, through delayed drainage in another, and through coded documentation after prolonged antibiotics in a third [6]. These are different observational routes into the same clinical territory. If the outcome system equates route with state, then institutional differences in workup become phenotype differences by accident. An atlas discourages that collapse. It allows route of appearance to be recorded as a descriptor of how a state became legible, while preserving the possibility that the underlying state is similar across sites. This distinction is particularly valuable in multicenter work, where documentation style and clinical thresholds vary markedly.

Constructing the atlas is not the same as inventing a taxonomy from scratch. It is an iterative process that draws on clinical reasoning, chart review, and the recurrent patterns visible in the data. One begins with provisional regions that are broad enough to be recognizable and narrow enough to be distinct. As more cases are reviewed, some regions may prove stable while others blur together or split into meaningful subregions. For instance, contained leak may turn out to include one cluster of image-confirmed but minimally symptomatic courses and another of persistent inflammatory courses that never receive clean structural confirmation. Whether those should remain together depends on their observational stability and on the intended downstream use. The atlas is therefore not static. It is an evolving scientific object.

Because the atlas is richer than a binary label, it also supports more realistic descriptions of uncertainty [7]. Cases near the border between contained leak and inflammatory mimic can be recognized as such without pretending that one side of the border is always correct. Over time, the density of cases in these border regions can inform both phenotype refinement and model calibration. If a large number of cases repeatedly accumulate near one boundary, that is an indication that the boundary is either clinically subtle or operationally poorly supported. A binary system usually hides this by converting all such cases into either false positives or false negatives. An atlas makes the pattern visible.

The organizing principle of the atlas has implications for feature interpretation as well. A variable may not simply predict leak in the abstract. It may pull cases toward one region of the atlas more than another. Elevated inflammatory markers may distinguish both progressive leak and some mimic states from uncomplicated recovery, while imaging-supported procedural signals may be more specific to contained or managed leak territory. Thinking in atlas terms encourages interpretation that is local rather than global. This is often closer to clinical reality. Surgeons do not ask only whether a variable increases the probability of leak in a pooled population. They ask what sort of postoperative course that variable suggests and how it changes concern within a given differential neighborhood.

Most importantly, the phenotype atlas makes visible the fact that outcome definition is a boundary-placement exercise. The clinical world supplies a terrain of states [8]. The analyst or governance group decides where on that terrain to draw the event border. There is no way to avoid that decision, but there is every reason to make it explicit. Once the terrain is mapped, boundary choices become understandable, contestable, and revisable. That

Table 7 Route-to-Recognition Variability

Route	Entry Signal	Site Variation	Risk
Imaging-first	Early CT	High imaging centers	Detection bias
Procedure-first	Drain/surgery	Aggressive care	Severity skew
Code-first	Admin coding	Documentation-heavy	Delay distortion
Lab-driven	Persistent labs	Limited imaging	Ambiguity
Hybrid	Mixed pathway	Variable	Inconsistency

Table 8 Negative Class Structure

Negative Type	Description	Distance to Leak	Modeling Impact
True Recovery	Uneventful course	Far	Easy separation
Mimic	Overlapping syndrome	Near	False positives
Unresolved	Incomplete evidence	Near	Label noise
Underobserved	Sparse data	Variable	Uncertainty
Transient Issue	Self-limited abnormality	Moderate	Boundary blur

is a major improvement over approaches in which the boundary is built first and the terrain inferred afterward from model errors. In AL research, where difficulty tends to concentrate in the middle and not at the extremes, beginning with the atlas is often the more honest and more informative path.

3. Temporal Narrative Case Construction

Anastomotic leak is rarely best understood as a single charted moment. It is more often a narrative that unfolds through sequence, accumulation, and reinterpretation. For this reason, case construction for AL should be temporal and narrative before it is categorical. By narrative, I do not mean anecdotal storytelling. I mean that the clinically relevant object is often the ordered relation among observations rather than the isolated presence of any one of them. A patient's trajectory through the postoperative record may convey more about leak status than the final set of coded facts considered without chronology. This is why many static composite rules feel both useful and inadequate: they capture some of the important pieces but not the path by which those pieces came to matter.

Temporal narrative construction begins with the recognition that postoperative evidence changes role over time. A white blood cell rise in the first postoperative day may be a weak and common signal. The same or greater elevation several days later, especially if persistent and accompanied by localized symptoms or management changes, carries different meaning [9]. Broad-spectrum antibiotics given routinely or prophylactically have a different status from antibiotics escalated after a pattern of clinical concern has already emerged. A diagnosis code entered at discharge often summarizes a history rather than serving as a new observation. Imaging or intervention events may function as turning points in the narrative, converting suspicion into concrete management response. When time is ignored, these differences collapse into a flat bag of evidence. When time is preserved, the same evidence can be understood as part of a developing story.

One of the main strengths of temporal narrative construction is that it gives proper weight to sequence. The order in which events occur is often clinically informative. A patient who first develops inflammatory persistence, then undergoes CT imaging, then receives drainage, and finally accrues a diagnostic code has followed one narrative. A patient who deteriorates abruptly, returns to the operating room, and is coded afterward has followed another. A patient who receives prolonged antibiotics without strong structural evidence and is later coded has followed yet another. These paths are not mere administrative variants. They correspond to different regions of the phenotype atlas and to different degrees of certainty and severity. If a model sees only the final label, it cannot know which narrative it is learning. If the underlying outcome construction preserves the narrative, both phenotyping and error analysis become more meaningful.

This perspective also clarifies the role of hindsight [10]. Retrospective chart review often makes leak appear cleaner than it felt prospectively because later evidence retroactively organizes earlier uncertainty. Once a drain is placed or a surgeon documents concern explicitly, earlier subtle signs become legible as part of an

Table 9 Atlas Drift Factors Across Settings

Factor	Change Mechanism	Effect on Atlas	Example
Imaging Access	Increased scans	More contained leaks	Earlier CT use
Documentation	Template shifts	Coding timing change	Earlier diagnosis
Practice Style	Intervention threshold	Severity redistribution	More reoperations
Data Quality	Missingness	More unresolved cases	Sparse labs
Protocols	ERAS adoption	Faster recovery patterns	Shorter stays

evolving leak story. This is a clinically reasonable way to reconstruct truth, but it can create problems for early surveillance modeling if the retrospective endpoint is treated as though it were recognizable at the earlier time. Temporal narrative construction addresses this by distinguishing between final retrospective interpretation and earliest defensible narrative turning point. The latter is the moment at which the chart, without future information, begins to support meaningful concern for a leak-like trajectory. Preserving that distinction is essential if the outcome system is meant to support timely prediction rather than merely retrospective classification.

The relevance of this approach can be seen in Sadanandan (2024) [11], where the leak endpoint was intentionally defined as a composite thirty-day outcome requiring both diagnostic coding and corroborating clinical evidence, while feature extraction was restricted to the first twenty-four postoperative hours specifically to preserve a temporal gap between the prediction window and the later clinical declaration of the event. That design choice illustrates an important principle for narrative case construction: the event may be defined across a broad postoperative course, but the predictive use of the label depends on separating early narrative signals from later confirming chapters of the record. A phenotype atlas strengthens this principle by making it possible to describe not only that a later leak-related narrative occurred, but which type of narrative unfolded and when it began to take shape.

Narrative construction is also valuable for handling ambiguous cases. Consider a patient with persistent inflammatory elevation, delayed bowel function, and one course of escalated antibiotics, but no clear imaging confirmation and no procedure. A static rule may force such a case into negative or indeterminate status. A narrative view can preserve more nuance. It can characterize the admission as following a leak-adjacent story that never achieved structural or management confirmation. That classification may not satisfy a strict event rule, but it is still highly relevant to surveillance, mimic analysis, and phenotype development [12]. In other words, narrative thinking allows the record to retain clinically meaningful incompleteness without pretending that every path must end in decisive proof.

A further advantage of the narrative view is that it can integrate absence as well as presence. The fact that no confirmatory imaging was obtained despite persistent symptoms may mean one thing in a center with low imaging thresholds and another in a center with selective imaging practice. The fact that a patient improved rapidly without source-control intervention after transient concern may pull the narrative away from leak territory. The fact that deterioration occurred after discharge or outside the visible record may leave the narrative genuinely truncated. Static endpoint logic often treats absences simplistically, as if missing evidence were always neutral. Narrative construction encourages a more contextual reading in which absences themselves become interpretable components of the case.

Narrative case construction also changes the way phenotype labels can be justified. Instead of assigning a phenotype solely because certain codes or procedures are present, one can ask which story the record most strongly supports. Is it the story of early contained local failure that became radiographically visible before generalized decline? Is it the story of progressive systemic disturbance culminating in source control? Is it the story of recurrent concern without resolution? Is it the story of another postoperative complication that repeatedly mimicked leak before declaring its own identity? These are distinct narratives, and their distinction often matters more than any individual variable. The phenotype atlas gains depth when states are linked to typical narrative arcs rather than merely to static feature patterns.

Importantly, this narrative emphasis does not require abandoning structured data. On the contrary, structured data often provide the backbone of the story: timing of labs, antibiotic changes, ICU transfer, imaging, procedures, and coded diagnoses. Notes and reports refine the story, but the temporal relationships among structured events already carry substantial narrative information. What matters is that those relationships are treated as constitutive of the outcome rather than as incidental metadata [13]. A dataset that stores only whether each event ever happened has lost part of the clinical object. A dataset that preserves ordered occurrence, turning points, and late consolidation retains more of what makes AL definable.

Narrative construction has practical consequences for annotation as well. Reviewers asked to classify cases often perform this work implicitly, reconstructing stories from chart fragments and then translating them into a single label. Making the narrative layer explicit can improve consistency. Reviewers can be trained to identify phases of suspicion, escalation, confirmation, and resolution, rather than jumping directly to a verdict. The resulting annotations are more verbose, but they also create a richer resource for later refinement of the phenotype atlas. Over time, repeated narratives can reveal which leak states recur and which boundary cases remain persistently difficult. This is much harder to discover when review outputs are binary from the outset.

A final virtue of temporal narrative construction is that it aligns more closely with how clinicians actually reason. Surgeons do not ordinarily decide whether a complication exists by inspecting a final unordered set of chart variables. They interpret the evolving sequence of events in relation to anatomy, operative details, and prior expectations. Outcome systems that ignore sequence often feel unnatural to clinical users for this reason. By contrast, narrative-aware systems can provide outputs that refer to evolving leak-like trajectories rather than to static hidden classes. This does not solve the challenge of generalization, but it does reduce the conceptual distance between chart abstraction and bedside reasoning [14].

For AL phenotype work, then, temporal narrative construction is not a stylistic preference. It is a methodological necessity when the event of interest is itself a trajectory of emergence. It enables more faithful handling of hindsight, more realistic description of ambiguity, and more interpretable links between evidence and phenotype states. Most importantly, it turns outcome definition from a flat aggregation problem into a problem of reconstructing clinically meaningful stories from routinely recorded traces. That reconstruction is where much of the real scientific work lies.

4. Border Zones, Mimic States, and the Negative Space of Leak Phenotyping

The most revealing region of an outcome system is often not its center but its edge. In AL phenotyping, the edge consists of border-zone cases: admissions that are neither cleanly inside nor cleanly outside leak territory, and whose final status depends heavily on how evidence is weighted, how uncertainty is tolerated, and which adjacent postoperative syndromes are recognized. These border zones are methodologically important because they expose the hidden assumptions of the event definition. They are also clinically important because many practical decisions occur there. A bedside team is not usually challenged by the obvious catastrophic leak or the clearly uneventful recovery. The challenge is the patient whose trajectory is concerning but inconclusive, or whose record resembles leak closely enough to trigger action without ever achieving definitive diagnostic closure.

A major weakness of standard binary labeling is that it erases this negative space. All non-leak cases are often treated as members of one vast residual class. That makes classification look simpler than it is. In reality, some negative cases are far from leak and others are very near it [15]. The near cases may differ from true leaks only by the absence of one crucial finding, by a different final interpretation of the same imaging pattern, or by a favorable response to treatment before the chart could settle the question. Such cases are not ordinary negatives. They are part of the leak neighborhood, and the phenotype atlas should show them as such. If it does not, the resulting models may appear to have unexplained false positives when they are actually identifying clinically meaningful near-leak states.

Mimic states deserve especially careful representation. A mimic is not simply any false positive. It is a postoperative syndrome that overlaps with leak in the dimensions that matter for recognition: inflammatory pattern, abdominal findings, hemodynamic change, imaging ambiguity, or management response. Some mimics are primarily inflammatory, such as postoperative ileus with associated systemic stress. Others are infectious but anatomically distinct from the anastomosis. Others are consequences of bleeding, ischemia, or pulmonary complications that disturb the postoperative course enough to resemble leak in broad terms. Because these states share observational features with leak, they often occupy the exact terrain where predictive systems and adjudicators struggle. Their inclusion as named neighboring regions is therefore essential for realistic phenotyping.

One advantage of explicitly modeling negative space is that it changes the interpretation of false alarms. In a surveillance setting, a model that frequently flags contained mimics or unresolved inflammatory courses may still be clinically useful if those courses are exactly the ones that warrant closer attention [16]. In contrast, a model whose false alarms are concentrated in entirely benign recoveries is of a different character. Aggregate precision does not distinguish these situations. A phenotype-aware evaluation does. It can report not only that a prediction was wrong relative to the final binary label, but also how close it was in atlas space and whether it erred into a clinically important neighboring state. This distinction matters for clinician trust, because not all false positives impose the same burden or reveal the same model weakness.

Border zones also force a more honest treatment of uncertainty. In many datasets, ambiguous or near-leak cases are either discarded, adjudicated aggressively into a hard class, or absorbed into the negative class because the positive rule is stricter than the evidence. Each of those options hides something important. Discarding the cases hides that the boundary is real and recurrent.

Forcing them into a class hides the degree of uncertainty. Absorbing them into negative status hides that the negative class is internally structured. The phenotype atlas provides a more honest alternative by assigning such cases to unresolved or mimic-adjacent regions when the evidence supports only proximity, not full inclusion. This does not eliminate the need for a final binary rule in some settings, but it preserves the structure that the binary rule will otherwise destroy.

There is also a strategic reason to study border zones carefully. Definitions are easiest to revise when one knows where they fail [17]. If a certain cluster of border-zone cases keeps drifting between contained leak and mimic status depending on the institution or adjudicator, that is a signal that the phenotype boundary needs clarification or that the evidence ontology is underspecified. If another cluster repeatedly receives late diagnosis codes without corresponding early narrative support, that may indicate an administrative pathway that should be treated differently from physiologically evident leak. If certain mimic states are always mistaken for leak because the record lacks adequate discriminating features, then the limits of routine data should be acknowledged rather than papered over by a sharp label. Border zones, in this sense, are where the definition learns about itself.

The negative space around leak is also crucial for fairness and transport. Subgroups that are observed differently may accumulate more unresolved or mimic-like cases near the event border. A hospital with richer imaging access may move some border-zone cases into contained leak territory that would remain unresolved elsewhere. A service line with more aggressive intervention may convert progressive concerns into management-defined positives more often. Without an atlas, these differences manifest as unexplained variation in prevalence or model calibration. With an atlas, they appear as shifts in border-zone occupation and in the pathways by which cases leave uncertainty and enter positive territory. This is a more diagnosable and more clinically intelligible description of variation.

Importantly, negative space is not merely a defensive addition to protect against overfitting. It also enriches positive phenotype design. A contained leak phenotype becomes more meaningful when one can say how it differs from nearby unresolved inflammatory states. A severe leak phenotype becomes more interpretable when separated from other catastrophic postoperative syndromes that share some physiological signals but not the same anatomical logic [18]. Border-zone modeling therefore sharpens the interior of the atlas as well as its edges. It forces phenotype classes to be defined relationally rather than in isolation.

Another consequence is that negative classes should not be homogeneous even in the final analytic projection. When a study ultimately requires a binary endpoint, it can still preserve structured negatives during development and validation. This allows analysts to ask whether a model is distinguishing leak from uncomplicated recovery, from mimics, from underobserved courses, or from unresolved concern. Those are different tasks. A single AUC over all negatives hides the fact that some regions of the negative class are easy and others are very near the positive boundary. Phenotype-aware negative space makes these local tasks visible and helps prevent overinterpretation of global performance metrics.

The border zone is also where human review and automated systems are likely to disagree most often. That disagreement should not be seen as a nuisance. It is evidence that the system is operating in precisely the region where definitions are hardest. If reviewers are forced to collapse their judgments into binary labels, one learns only whether agreement was high or low. If they can describe a case as lying between contained leak and mimic, or between progressive leak and unresolved concern, the disagreement becomes informative. It can reveal not only which boundary is unstable, but why. Over time, this can guide revision of both the atlas and the operational outcome rule [19].

For AL phenotyping, then, negative space is not an afterthought. It is part of the phenomenon. Leak is recognized against a background of postoperative states that are abnormal, overlapping, and variably observed. A system that does not represent that background will eventually rediscover it as error. A system that does represent it can use it to build more honest labels, more interpretable phenotypes, and more realistic performance claims. In practical terms, this means that difficult cases should be studied as residents of a structured border zone rather than treated as nuisances at the margins of an otherwise clean event definition.

5. Annotation Science, Review Workflow, and Label Discipline

The construction of a robust phenotype atlas depends on annotation, but annotation in this setting is not a clerical process. It is a scientific activity that deserves its own methods, controls, and documentation. Too often, AL label development relies on hidden annotation labor: a few experts review charts, a consensus label appears, and the analytic paper moves on. This compresses a large amount of interpretive work into an invisible background step. It also makes it difficult for later researchers to understand how the label was actually stabilized. A more disciplined approach treats annotation as a structured workflow with explicit goals, calibrated raters, captured rationales, and preserved uncertainty.

The first requirement of such a workflow is task separation. Reviewers should not be asked one vague question such as whether the patient had a leak. They should be guided through a staged interpretation [20]. One stage can focus on evidence extraction, identifying which signals were present and when they appeared. Another can focus on narrative interpretation, reconstructing the

likely course of postoperative concern. A later stage can focus on atlas placement, determining which phenotype region or regions the case most plausibly occupies. Only after that should a task-specific projection be made into a binary label, if needed. This staging matters because it preserves intermediate judgments that would otherwise be lost. It also makes disagreement easier to localize. Reviewers may agree on the evidence and on the narrative but disagree on the final projection, which suggests a definitional issue rather than a chart-reading one.

Reviewer calibration is equally important. Surgeons, intensivists, radiologists, coders, and data scientists do not enter the chart with identical assumptions. Even within surgery, differences in training and local practice shape what counts as persuasive evidence. Calibration sessions should therefore precede large-scale review. These sessions are not merely to align raters on the current rule. They are opportunities to refine the rule itself by surfacing cases that reveal hidden ambiguities. The most useful calibration cases are often not the obvious positives and negatives, but the difficult border-zone cases where reviewers are forced to articulate why they are leaning one way or another. Capturing those discussions helps build the rationale for the phenotype atlas and clarifies which dimensions matter most in practice [21].

A disciplined annotation workflow should also record reason codes. When a reviewer places a case in contained leak, progressive leak, mimic, or unresolved territory, the system should capture which elements of the narrative were decisive. Did the decision depend on imaging language, on operative findings, on the sequence of antibiotics and drainage, on persistent systemic response, on exclusion of another source, or on late coding? Such reason codes are valuable for several purposes. They create an audit trail. They reveal which evidentiary streams dominate certain phenotypes. They provide a way to train future annotators. They also offer insight into what the eventual model is actually expected to learn. A label without reasons is a verdict; a label with reasons is part of a cumulative scientific record.

One of the most important features of label discipline is the preservation of ambiguity. Annotation systems often push reviewers toward decisive answers because undecided labels are inconvenient for modeling. This is a methodological mistake. Ambiguous cases do not disappear when the reviewer is forced to choose. They only disappear from view. A better annotation protocol offers reviewers a structured way to represent uncertainty. That can take the form of confidence levels, multiple permissible phenotype assignments, or explicit unresolved categories [22]. These outputs may seem harder to use computationally, but they are more faithful to the chart and more useful for future refinement. They allow the dataset to contain not just labels, but information about where the labels are least secure.

Temporal annotation should be built directly into the workflow. Reviewers need to indicate not just what the final case status appears to be, but when the record first supports that interpretation and what later evidence solidified it. This is especially important when the atlas will later support surveillance tasks. A case that is retrospectively a clear progressive leak may have been practically indistinguishable from an inflammatory mimic until much later. If annotation captures only the final judgment, the dataset will suggest a level of early recognizability that did not exist. By asking reviewers to mark turning points in the narrative, one preserves a better correspondence between retrospective knowledge and prospective usefulness.

Review workflow should also anticipate disagreement as informative rather than pathological. Some cases should be double-reviewed deliberately because they are expected to fall near phenotype boundaries. Others may need panel review because they reveal mismatches between the ontology and routine chart language. Instead of resolving all disagreements immediately into one consensus answer, the workflow should retain records of where the disagreement occurred. Did reviewers disagree about evidence extraction, about attribution to the anastomosis, about the severity of the state, or about whether the record was sufficient to decide? These are very different kinds of discordance. When preserved, they become data about the definition system rather than mere obstacles to producing a training set.

Another important aspect of label discipline is adjudication of negatives [23]. In many studies, positives receive careful review while negatives are accepted by default. This creates asymmetry in the dataset. The negative class becomes a mixture of true recoveries, unreviewed mimics, underobserved cases, and unresolved postoperative abnormalities. A phenotype atlas cannot be built responsibly on that basis. At least a targeted sample of negatives near the border should undergo the same level of review as positives. This is not simply for fairness to the negative class. It is necessary for understanding where the event boundary actually lies and for identifying the most common leak-adjacent states that might otherwise be hidden.

Annotation science also includes version control of the annotation protocol itself. As reviewers learn, they may request clarifications, new reason codes, or refined phenotype descriptions. These changes should be documented rather than silently introduced. Otherwise, the meaning of a label can shift during the annotation campaign, producing an internally inconsistent dataset. Versioning the review manual, the phenotype atlas, and the adjudication rubric allows later analysts to know which definitions were in force for which cases. This is particularly important in large or longitudinal projects where annotation occurs over months or years.

There is also a training function here [24]. New annotators should not simply be given the latest label definitions. They should be shown representative cases from

each atlas region, including borderline and mimic cases, together with examples of disagreement and its resolution. This helps them understand not only the categories but the logic of the categories. It also reduces the risk that annotation devolves into surface matching based on keywords or procedures. Because AL often requires reasoning across time and across evidentiary channels, training by category description alone is usually insufficient. Annotators need to see how the narrative structure of cases maps onto the atlas.

Well-disciplined annotation can also improve future automation without turning the process into a simplistic labeling factory. The point is not to mimic human decisions mechanically. It is to identify which patterns of reasoning recur consistently enough to be operationalized. If reviewers repeatedly cite the same combinations of temporal sequence, imaging language, and management response when assigning a phenotype, those patterns can inform better structured extraction and better model features. Conversely, if certain boundaries remain unstable despite careful review, automation should not be expected to solve them by force. In this way, annotation science contributes not only labels but knowledge about the achievable limits of the task.

For AL research, then, label discipline is central. A phenotype atlas built without structured annotation quickly becomes a rhetorical device rather than a functioning analytic object. The atlas needs reviewers who can place cases within it consistently, document why they did so, and signal where they could not [25]. Only then can the resulting states support reliable modeling, meaningful validation, and honest discussions of uncertainty. Annotation is not what happens before the science begins. In outcome definition, it is a large part of the science itself.

6. Cross-Setting Translation, Atlas Drift, and Clinical Use

A phenotype atlas that functions well in one setting is not automatically portable. Translation across hospitals, surgical services, and time periods raises a distinct set of problems because the atlas is built from the record, and records differ in what they make visible. A site with frequent postoperative imaging may populate contained leak and mimic regions differently from a site where imaging is delayed or used sparingly. A center with lower thresholds for reintervention may shift progressive courses more quickly into severe or management-defined territory. Documentation culture can alter how unresolved cases appear. Even if the underlying clinical states are similar, their record signatures may not be. This means that the atlas itself may drift when moved, not because the concept of leak has changed, but because the observation surface has.

Cross-setting translation therefore requires explicit mapping between atlas states and local evidence pathways. The goal should not be to enforce identical raw criteria everywhere. That is rarely realistic. Instead, the goal is to preserve the intended state distinctions while allowing different institutions to manifest those states through different combinations of evidence. A contained leak remains a contained leak even if one hospital documents it primarily through radiology language and another through drain placement plus narrative concern [26]. Translation succeeds when those different pathways still land in the same atlas neighborhood. It fails when local documentation habits become the dominant determinants of phenotype membership.

This is where atlas descriptors become important. Each state in the atlas should be defined not only by one operational signature but by a core clinical profile and a set of common observational realizations. The clinical profile anchors the state concept. The observational realizations describe how the state usually appears in a given data environment. This separation makes it possible to adapt the atlas without abandoning comparability. External collaborators can ask whether they can instantiate the same state concepts using their own realizations, rather than asking whether they can reproduce every field and every extraction rule identically. In AL work, this is often the only practical route to multicenter consistency.

Translation also benefits from preserving route-to-atlas information. If a case enters the contained leak region through imaging support at one site and through procedure-plus-note evidence at another, the atlas should be able to record that difference without splitting the state itself. Over time, route patterns can be compared across sites to detect where the same phenotype is being observed differently. This is valuable both scientifically and operationally. Scientifically, it clarifies whether reported differences in phenotype prevalence reflect clinical reality or observation policy. Operationally, it helps institutions understand which kinds of cases their current documentation practices render most visible or least visible [27].

Atlas drift over time is another major issue. Hospitals change. Enhanced recovery pathways, note templates, coding support, imaging access, and antimicrobial practices all evolve. These changes can alter the way cases move through the atlas even within a single center. A label system that is not monitored for drift may continue to produce apparently stable outputs while slowly changing meaning. One year's unresolved inflammatory courses may become the next year's contained leaks simply because imaging was incorporated earlier. Conversely, a reduction in documentation detail may cause some formerly distinct phenotypes to blur together. Drift is not always harmful; sometimes it reflects genuine improvement in recognition. The key is that it should be visible, not silent.

Monitoring atlas drift requires more than checking model AUC. It requires inspecting how cases occupy the atlas over time, how often certain border zones are used, which routes of recognition dominate, and whether the

prevalence of unresolved or mimic states changes. If a major workflow intervention suddenly reduces unresolved classifications and increases image-supported contained leak cases, that may represent more definitive clinical workup rather than more leak. If a documentation change causes diagnostic codes to appear earlier or later in the timeline, the apparent onset of certain narratives may shift. A governance-aware atlas should incorporate routine review of such patterns and should treat major shifts as prompts for definitional audit.

Clinical use introduces another layer of translation because the atlas must eventually communicate with humans under time pressure [28]. Clinicians are unlikely to embrace a phenotype system that feels alien to the way they conceptualize postoperative deterioration. The atlas therefore needs outputs that are interpretable at the bedside. This does not mean simplistic one-line explanations. It means that the state descriptions and border zones should correspond to recognizable clinical situations. If the system identifies a patient as occupying a region consistent with contained leak versus unresolved inflammatory mimic, a clinician should be able to understand what that distinction means and why it matters. This is one of the reasons atlas construction should remain close to narrative and operational realities rather than drifting into abstract clustering.

Different clinical uses may require different projections of the same atlas. A surveillance tool may want to highlight movement toward leak-adjacent territory even before definitive confirmation, which means unresolved and progressive border states must remain visible. A registry may want only a narrower projection focused on well-supported clinically significant leak states. A quality review committee may want access to the full atlas because it is interested not only in final events but also in near misses and difficult postoperative courses. The strength of an atlas approach is that these different projections can coexist without requiring entirely different conceptual systems. Binary labels are then understood as task-specific slices through a shared map rather than as competing and disconnected definitions.

There is also a cultural consequence to this. When institutions adopt a phenotype-aware view of AL, discussions of postoperative complications become more precise. Instead of arguing over whether a case counts as a leak in the abstract, teams can discuss which region of postoperative behavior it occupied, whether the current institutional projection from the atlas is appropriate, and whether the evidence route that made the case visible reflects local practice more than underlying pathology [29]. This can improve case review and quality improvement even before any predictive model is deployed. The atlas becomes a shared language rather than merely a modeling convenience.

Cross-setting translation and clinical use therefore reinforce the same lesson. Outcome definitions are not self-sustaining. They need a representational framework that can tolerate local variation without losing conceptual coherence. For AL, the phenotype atlas provides such a framework by keeping clinical states distinct from their observational routes, by allowing border zones to remain explicit, and by supporting multiple operational projections for different tasks. A binary event may still be necessary in many contexts, but it becomes much less fragile when it is understood as one output of a maintained and interpretable state map.

7. Conclusion

Outcome definition for anastomotic leak becomes more coherent when it is approached as map-making rather than as threshold setting. A routine surgical record does not present leak as a single clean datum. It presents traces of evolving postoperative states, some clearly compatible with leak, some clearly distant from it, and many occupying the difficult middle ground of suspicion, mimicry, partial confirmation, or incomplete observation. A phenotype atlas makes that terrain visible before the event boundary is drawn. In doing so, it exposes the fact that binary labels are projections from a richer clinical and observational landscape, not natural facts waiting to be extracted.

This perspective changes several things at once. It shifts attention from isolated criteria to phenotype regions and border zones. It treats temporal narrative as central rather than incidental [30]. It elevates annotation and adjudication into explicit scientific activities with their own discipline. It recognizes unresolved and mimic states as part of the task rather than as inconvenient leftovers. It also provides a more durable basis for translation across institutions and over time, because route of recognition can vary without destroying the coherence of the underlying atlas. Many applications still require them. The value is that it clarifies what those endpoints include, what they exclude, and which clinically meaningful distinctions have been compressed in the process. A contained leak, a progressive leak, a catastrophic leak, and a leak-like postoperative mimic should not become exchangeable simply because they all lie near the same event boundary. A phenotype-aware framework keeps those differences visible and makes model behavior easier to interpret. For AL research and deployment, the main methodological challenge is therefore not only to improve predictive performance. It is to ensure that the target being predicted corresponds to a defensible and inspectable region of postoperative state space. A phenotype atlas offers a way to do that. It does not promise perfect certainty, and it does not remove the need for chart review, governance, or external testing. What it provides is a more honest representational starting point for defining leak in records that are heterogeneous, delayed, and path dependent by nature. In that sense, the atlas is not an embellishment of the outcome. It is the structure that makes the outcome meaningfully definable [31].

References

[1] T. Sammour, L. Cohen, A. I. Karunatillake, M. Lewis, M. Lawrence, A. Hunter, J. Moore, and M. L. Thomas, "Validation of an online risk calculator for the prediction of anastomotic leak after colon cancer surgery and preliminary exploration of artificial intelligence-based analytics," *Techniques in coloproctology*, vol. 21, pp. 869–877, 10 2017.

[2] P. Allan, S. Reddy, B. J. Moran, G. Vrakas, H. Giele, F. Mohamad, J. Gilbert, S. Sinha, A. Vaidya, T. Cecil, and P. J. Friend, "Ptu-105 cytoreductive surgery and small bowel transplantation is a feasible option for patients with end-stage pseudomyxoma peritonei," *BSG 2016*, vol. 65, pp. A106.2–A107, 8 2016.

[3] E. Nilsson, M. F. Olsén, U. Trölle, F. Eriksson, and M. Hermansson, "295 prediction of morbidity and mortality after thoracoabdominal esophageal surgery," *Diseases of the Esophagus*, vol. 34, 9 2021.

[4] E. Smith, S. Daly, B. R. Smith, M. W. Hinojosa, and N. T. Nguyen, "The role of endoscopic stent in management of postesophagectomy leaks.," *The American surgeon*, vol. 86, pp. 1411–1417, 10 2020.

[5] J. C. Y. Ngu and A. S. Y. Wong, "Transanal natural orifice specimen extraction in colorectal surgery: bacteriological and oncological concerns," *ANZ journal of surgery*, vol. 86, pp. 299–302, 11 2015.

[6] A. S. Soares, S. Bano, N. T. Clancy, L. B. Lovat, D. Stoyanov, and M. Chand, "Fluorescence angiography classification in colorectal surgery – a preliminary report," 6 2022.

[7] M. S. Renna, M. T. Grzeda, J. Bailey, A. Hainsworth, S. Ourselin, M. Ebner, T. Vercauteren, A. Schizas, and J. Shapey, "Intraoperative bowel perfusion assessment methods and their effects on anastomotic leak rates: meta-analysis.," *The British journal of surgery*, vol. 110, pp. 1131–1142, 5 2023.

[8] F. Friedmacher and P. Puri, "Delayed primary anastomosis for management of long-gap esophageal atresia: a meta-analysis of complications and long-term outcome," *Pediatric surgery international*, vol. 28, pp. 899–906, 8 2012.

[9] C. J. R. Schlick, T. K. Yuce, A. D. Yang, M. F. McGee, D. J. Bentrem, K. Y. Bilimoria, and R. P. Merkow, "A postdischarge venous thromboembolism risk calculator for inflammatory bowel disease surgery.," *Surgery*, vol. 169, pp. 240–247, 10 2020.

[10] K. Stormark, P.-M. Krarup, A. Sjövall, K. Søreide, J. T. Kvaløy, A. Nordholm-Carstensen, B. S. Nedrebø, and H. Kørner, "Anastomotic leak after surgery for colon cancer and effect on long-term survival.," *Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland*, vol. 22, pp. 1108–1118, 2 2020.

[11] 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.

[12] L. K. Temple and R. S. McLeod, "A meta-analysis comparing functional outcome following straight coloanal anastomosis versus a colonic j pouch," *Seminars in Colon and Rectal Surgery*, vol. 13, pp. 62–66, 3 2002.

[13] E. Hannan, G. Feeney, M. F. Ullah, E. Condon, J. C. Coffey, and C. Peirce, "Robotic colorectal surgery in elderly patients: A single-centre experience.," *The international journal of medical robotics + computer assisted surgery : MRCAS*, vol. 18, pp. e2431–e2431, 6 2022.

[14] M. G. Ruiz, M. L. Escribano, C. C. Fernández, L. C. Poch, and S. S. Martínez, "Robotic surgery for colorectal cancer.," *Annals of gastroenterological surgery*, vol. 4, pp. 646–651, 12 2020.

[15] F. Rebecchi, E. Ugliono, M. E. Allaix, M. Toppino, A. Borello, and M. Morino, "Robotic roux-en-y gastric bypass as a revisional bariatric procedure: a single-center prospective cohort study," *Obesity surgery*, vol. 30, pp. 11–17, 8 2019.

[16] O. Backman, J. Freedman, R. Marsk, and H. Nilsson, "Laparoscopic roux-en-y gastric bypass without division of the mesentery reduces the risk of postoperative complications," *Surgical endoscopy*, vol. 33, pp. 2858–2863, 11 2018.

[17] S. Somashekhar, G. R. K. Reddy, A. Y. Deshpande, K. R. Ashwin, and R. Kumar, "A prospective study of real-time identification of line of transection in robotic colorectal cancer surgery by icg," *Journal of robotic surgery*, vol. 15, pp. 369–374, 6 2020.

[18] R. Rajan, A. Arachchi, M. Metlapalli, J. Lo, R. Ratinam, T. C. Nguyen, W. M. K. Teoh, J. T.-T. Lim, and H. Chouhan, "Ileocolic anastomosis after right hemicolectomy: stapled end-to-side, stapled side-to-side, or handsewn?," *International journal of colorectal disease*, vol. 37, pp. 673–681, 2 2022.

[19] C. Hao, X. Su, J. Ji, X.-F. Huang, and B. Xing, "Stomach-interposed cholecystogastrojejunostomy: a palliative approach for periampullary carcinoma.," *World journal of gastroenterology*, vol. 11, pp. 2009–2012, 4 2005.

[20] V. C. Shakya, C. S. Agrawal, and S. Adhikary, "Initial experience with soave's transabdominal pull-through: An observational study," *International journal of surgery (London, England)*, vol. 8, pp. 225–228, 1 2010.

[21] N. Hanna, Z. M. Mir, E. Williams, S. J. Merchant, B. Zevin, and W. Chung, "Laparoscopic ischaemic conditioning of the gastric conduit prior to a hybrid mckeown esophagectomy may not decrease the risk of anastomotic leak," *Videosurgery and Other Miniinvasive Techniques*, vol. 16, pp. 669–677, 4 2021.

[22] M. Brito, G. Nunes, C. Luz, G. Oliveira, P. P. Marques, and J. Fonseca, "Niti-s esophageal mega-stent: An emerging endoscopic tool with different applications in the management of surgical anastomotic leaks.," 4 2022.

[23] A. R. Baqar, S. Wilkins, W. C. Wang, K. Oliva, S. Centauri, R. Yap, and P. McMurrick, "A comparison of extracorporeal side to side or end to side anastomosis following a laparoscopic right hemicolectomy for colon cancer.," *ANZ journal of surgery*, vol. 92, pp. 1472–1479, 4 2022.

[24] 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.

[25] J. Crane, M. Hamed, J. Borucki, A. El-Hadi, I. Shaikh, and A. Stearns, "852 complete mesocolic excision (cme) versus conventional surgery for colon cancer: A systematic review and metaanalysis," *British Journal of Surgery*, vol. 108, 5 2021.

[26] A. Brownlee, E. Bromberg, and M. Roslin, "Outcomes in patients with helicobacter pylori undergoing laparoscopic sleeve gastrectomy," *Obesity surgery*, vol. 25, pp. 2276–2279, 5 2015.

[27] R. Cirocchi, G. Cochetti, J. J. Randolph, C. Listorti, E. Castellani, C. Renzi, E. Mearini, and A. Fingerhut, "Laparoscopic treatment of colovesical fistulas due to complicated colonic diverticular disease: a systematic review," *Techniques in coloproctology*, vol. 18, pp. 873–885, 5 2014.

[28] Y. Wan, S. Lim, J. Riebman, N. Jamous, and X. Gao, "Pwe-024 clinical and economic burden associated with anastomotic leak after colorectal surgeries in the united kingdom," *Gut*, vol. 63, pp. A132.1–A132, 6 2014.

[29] C. Papadimitriou, W. Martin, A. E. Dellis, C. Deliveliotis, and I. Mitsogiannis, "Twenty-four-month efficacy of ureteroureterostomy combined with unilateral nephrostomy following radical cystectomy.," *Cureus*, vol. 14, pp. e30478–e30478, 10 2022.

[30] N. Malibary, S. Manfredelli, A. Almuttawa, J.-B. Delhorme, B. Romain, C. Brigand, and S. Rohr, "Evaluating the surgeon's experience as a risk factor for post-esophagectomy chylothorax on a four-year cohort.," *Cureus*, vol. 12, pp. e8696–, 6 2020.

[31] N. Z. Teo and J. C.-Y. Ngu, "A comparison between the da vinci xi endowrist stapler and a conventional laparoscopic stapler in rectal transection: A randomized controlled trial," 9 2022.