Public

Identifying Illness Episode Misclassification in ADaM Using AI-Assisted Temporal Logic and Explainable Decision Tree Models

Purvi Kalra | Ephicacy Lifescience Analytics

PharmaSUG Japan 2026 | Paper AI-004

Public

The QC Problem

A new illness number does not always mean a new illness.

- Symptoms may be reported on several dates across a longitudinal study.
- ADaM derivations can create sequential records such as COVID_1_ILLNESS and COVID_2_ILLNESS.
- The QC question is whether the transition is consistent with the protocol/SAP temporal rule.
- Review becomes harder when similar derivations must be checked across multiple studies.

QC focus: identify transitions that deserve closer review — without changing the clinical rule.

Public

Protocol / SAP Rule → A Deterministic QC Screen

Example rule used in the synthetic SARS-CoV-2 scenario

Previous illness resolution

GAP_DAYS
Current onset − previous resolution

Current illness onset

GAP_DAYS ≤ 4

Temporal-review candidate

GAP_DAYS > 4

Four-day screening criterion not triggered

Important: GAP_DAYS is a screening variable — not the final clinical classification.

Public

Why Add an AI-Assisted QC Layer?

Deterministic SAS screen

- Transparent and traceable to the protocol/SAP
- Identifies the candidate population
- Consistent across records and studies
- Defines the screening boundary

Explainable ML layer

- Uses symptom-level evidence within candidates
- Organizes transitions by QC evidence
- Provides an interpretable if–then rule
- Helps prioritize human review

Protocol/SAP defines the rule → AI assists the reviewer

Public

Hybrid Protocol-Driven + AI-Assisted QC Framework

1. Protocol / SAP 4-day rule
2. SAS temporal screen
3. Symptom-level features
4. HPSPLIT explainable ML
5. QC review prioritization

Clinical rule remains authoritative.
Machine learning prioritizes candidate transitions for QC review.

Public

Synthetic Demonstration Design

30
synthetic illness transitions

20
GAP_DAYS ≤ 4 review candidates

10
GAP_DAYS > 4 not triggered

- SUBJ001–SUBJ030; two synthetic illness records per subject.
- 153 synthetic symptom-level records.
- 20 candidate transitions: 12 HIGH_CONFIDENCE and 8 REVIEW.
- No actual participant identifiers, study dates, or clinical outcomes.

Labels are synthetic QC reference labels — not clinical ground truth.

Public

What Evidence Does the Model See?

GAP_DAYS Current onset − previous resolution

PREVIOUS_DURATION Days in previous illness

CURRENT_DURATION Days in current illness

PREV_N_SYMPTOMS
Distinct symptoms in previous illness

CURRENT_N_SYMPTOMS
Distinct symptoms in current illness

N_SAME_SYMPTOMS
Shared symptom codes

SYMPTOM_CONTINUATION
Timing supports symptom continuation

The final model uses all seven predictors. SAME_SYMPTOM was omitted because it is redundant with N_SAME_SYMPTOMS > 0.

Public

Explainable HPSPLIT Model

20
candidate transitions

→

Entropy
tree growth

→

Cost-complexity
pruning

→

5-fold
random CV

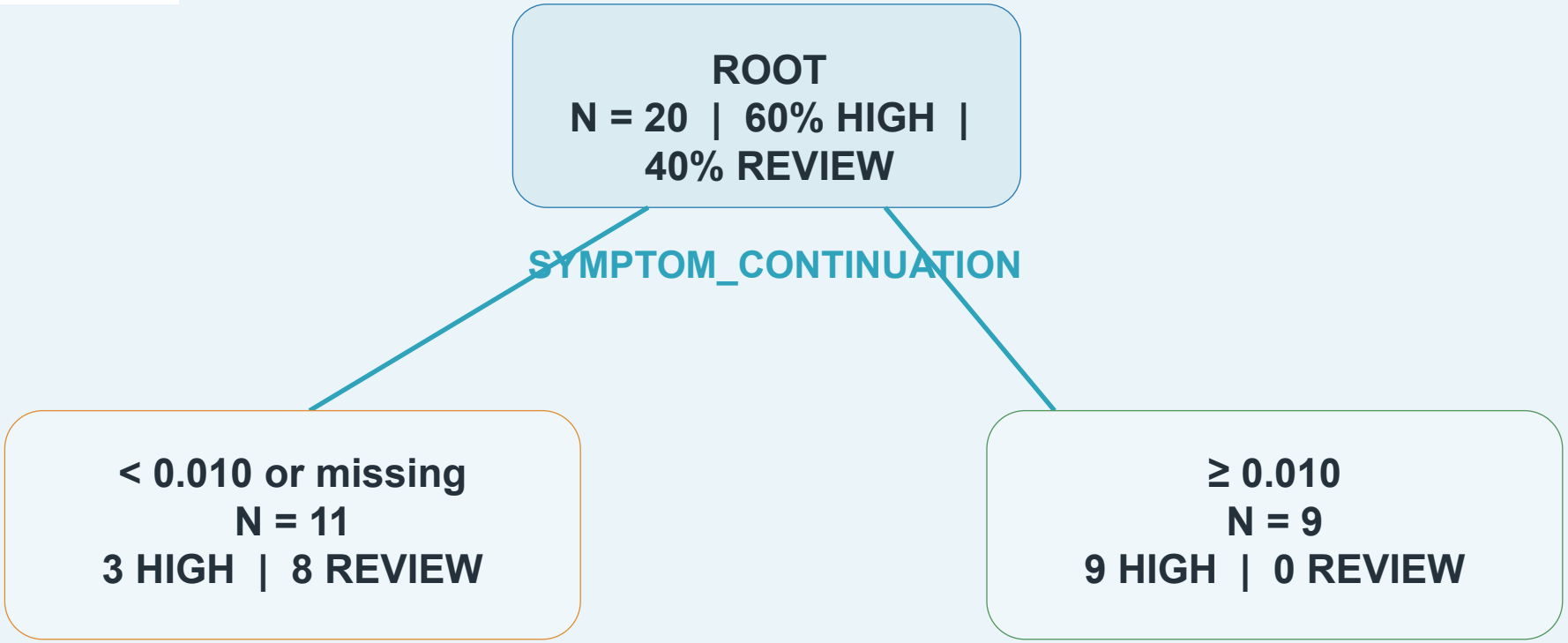
Target: QC_REFERENCE HIGH_CONFIDENCE | REVIEW

Purpose: prioritize QC evidence within the protocol-defined candidate population

Sensitivity model: GAP_DAYS excluded to test whether symptom-level evidence retains discriminative value.

Public

Selected Two-Leaf Tree: One Transparent Split



In this synthetic example, SYMPTOM_CONTINUATION is the dominant split.

Public

Model Performance: Illustrative, Not Generalizable

75.0%

CV accuracy

15 / 20

66.7%

CV sensitivity

HIGH_CONFIDENCE

87.5%

CV specificity

REVIEW

0.2300

CV misclassification

model assessment

Training AUC = 0.875

9 / 12 HIGH_CONFIDENCE correctly classified
8 / 8 REVIEW correctly classified

Because only 20 observations are modeled, these metrics demonstrate the workflow — not predictive performance in a real study.

Public

Two Transitions: Same Temporal Screen, Different Evidence

Example A — SUBJ007

GAP_DAYS = 0

SYMPTOM_CONTINUATION = 1

Temporal rule triggers review
+ symptom evidence supports continuity

Synthetic class: HIGH_CONFIDENCE

Example B — SUBJ014

GAP_DAYS = 4

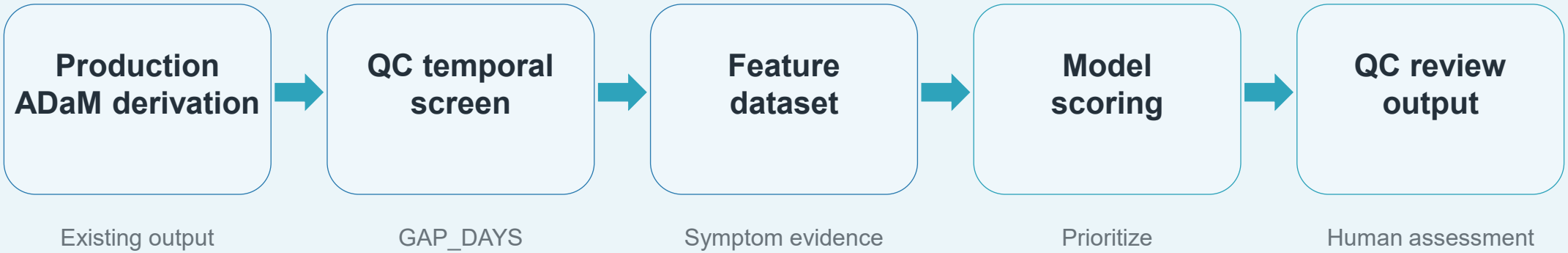
SYMPTOM_CONTINUATION = 0

Temporal rule triggers review
+ symptom evidence is insufficient

Synthetic class: REVIEW

Public

How This Fits Into a Real QC Workflow



Suggested QC dataset fields

Previous/current dates • GAP_DAYS • symptom features • model class • review status

The production derivation itself does not need to be changed for the QC demonstration.

Public

Governance: What the Model Does — and Does Not Do

DOES

- Prioritize temporal-review candidates
- Use traceable symptom-level features
- Provide an interpretable rule for review focus
- Support reproducible QC documentation

DOES NOT

- Redefine the protocol/SAP rule
- Override the clinical endpoint
- Make a diagnosis
- Establish clinical truth from synthetic labels

Human/QC assessment remains authoritative.

Public

Limitations and Next Steps

What limits this demonstration?

- Synthetic data and simulated QC reference labels
- Only 20 observations in the supervised population
- Decision-tree results may be unstable with small samples
- SYMPTOM_CONTINUATION is highly informative by construction

What would be needed next?

- Independent real study data
- Documented QC / clinical-adjudication labels
- Additional temporal and symptom-level predictors
- Assessment of impact on manual QC effort

Goal: scale the QC workflow without compromising traceability or clinical governance.

Public

Key Takeaways

1 Protocol/SAP defines the temporal boundary.

2 SAS makes the temporal screen and features reproducible.

3 Explainable ML helps prioritize candidate QC transitions.

4 In the synthetic demonstration, SYMPTOM_CONTINUATION was the sole split.

The model tells the reviewer where to look; it does not tell the reviewer what the clinical truth is.

Public

Thank You

Questions & Discussion

Name:	Purvi Kalra
Organization:	Ephicacy Lifescience Analytics
Address:	2nd Floor, 06, 2nd Cross Rd, Sarvobhogam Nagar, Arekere, Bengaluru, Karnataka 560076
Work Phone:	
E-mail:	purvi.kalra@ephicacy.com
Web:	www.ephicacy.com
LinkedIn:	Purvi Kalra