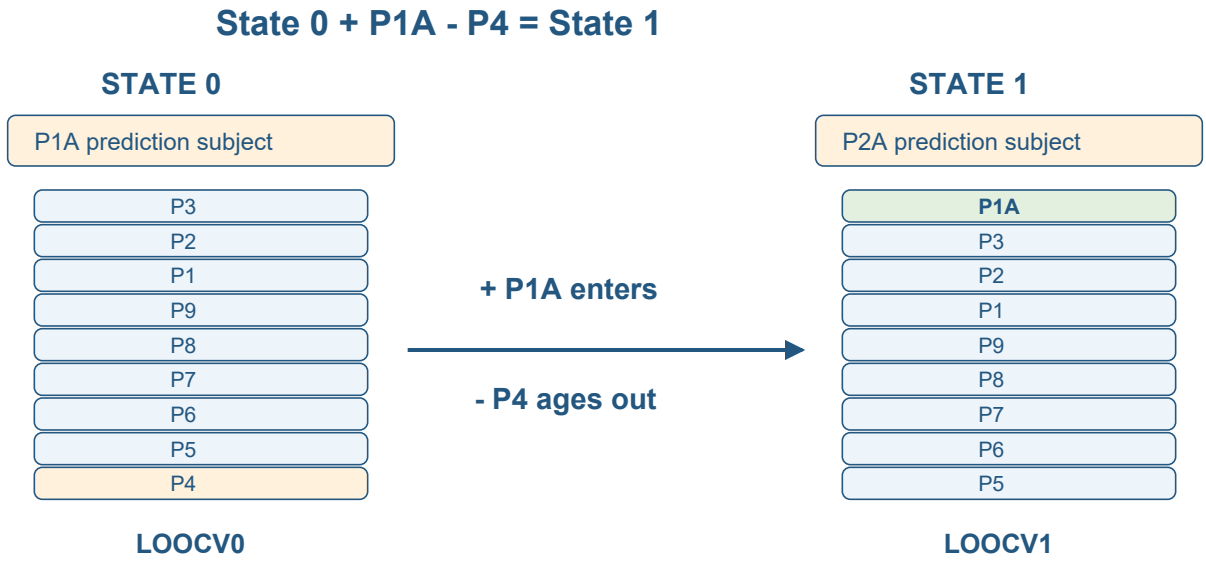
LOOCV and Evidence State Characterization

DJE | Working reader edition | October 28, 2026

A Decision Function (DF) operates on the evidence presented to it. When that evidence changes, the Decision Function's behavior and resulting outputs may change as well. The Real Property Valuation Decision Node (RPVDN) makes this issue concrete: completed sales enter a rolling eligible population, older sales leave it, and the comparable evidence supporting the next valuation evolves.



The diagram maps the transition directly examined in the laboratory. State 0 contained nine eligible completed sales. Before the P2A prediction was to be evaluated, P1A had closed and entered the eligible population, while P4 aged beyond the rolling twelve-month window. State 1 therefore also contained nine sales, but its membership had changed. This was a change in evidence eligibility, not a transition chosen for its performance consequences.

The laboratory applied the same adopted LOOCV procedure to these two identified states. It did not characterize every state encountered in the prediction experiments. The proposed operating discipline extends the principle: before a new prospective candidate prediction proceeds, establish whether characterization of the current state remains applicable and satisfies the governing empirical requirements; perform renewed characterization when required. A new candidate alone need not trigger rerunning an unchanged state's characterization.

Construct the State and Its Held-Out Tasks

An evidence state is a bounded representation of the market, constructed under specified eligibility and timing rules. RPVDN uses a rolling twelve-month window to identify the eligible completed-sale population. As evidence evolves, newly eligible sales enter and older sales age out, as shown in the State 0 to State 1 transition.

Within the identified state's eligible completed-sale population, each held-out property is estimated using the remaining properties, without excluding comparables that closed after the held-out property. The state is bounded by the rolling window; LOOCV adds no separate held-out-property closing-date restriction.

These held-out estimates differ from individual predictions, which use subject-specific evidence cutoffs. OPVAL is retrospective evaluation with an already realized Y concealed from the DF; PROSP produces and freezes a prediction before Y is known. The historical prediction-window anchors, including the single P1P case, are documented in Scope and Limitations.

The Decision Surface Snapshot (DSS) preserves the authorized evidence. A structured quantitative table, where supplied, organizes selected information within it. Identifying the population and faithfully representing its properties precede production of the held-out estimates. The state, configured DF, evidence representation, and procedure must remain identifiable so that a reviewer can determine what was characterized and whether the findings remain applicable.

Produce Fresh Held-Out Estimates

Leave-One-Out Cross-Validation (LOOCV) is the procedure for characterizing the state using the configured DF. For each fold, one completed sale becomes the held-out subject. Its realized sale price is concealed, and the remaining eligible sales provide comparative evidence.

The DF produces an estimate in each execution. Completed Experimental Records (ERs) undergo QC; ten qualifying outputs form the arithmetic-mean estimate used in this laboratory. Excluded ERs and replacements remain in the record. The estimate is compared with the held-out property's realized Y. Repeating the procedure across the population supplies the estimates, errors, and execution-variability measures from which state characterization is calculated.

When a previously predicted property enters a later characterization population, the adopted procedure uses a fresh qualifying ensemble rather than carrying forward its prediction. This constructs the population consistently under the same procedure, including when authorized inputs match earlier prediction inputs. Fresh outputs can differ because the DF is stochastic; the adopted measurement procedure is what requires fresh executions.

The opening diagram makes the distinction concrete: P1A was a prediction subject supported by State 0. After its sale completed, it entered State 1 and became a held-out subject in LOOCV1. Its prior prediction remains a prediction record; its fresh LOOCV ensemble supplies a state-characterization estimate.

Why LOOCV Was Selected Here

The RPVDN illustrates numerical predictions whose subsequent realized outcomes permit empirical comparison. That shared structure does not determine the appropriate characterization procedure. Each DN must select a method suited to its objectives, evidence population, rate of new observations, and consequences of error.

LOOCV was selected here to make systematic use of a small completed-sale population in a relatively low-volume transaction setting. Each sale supplies a held-out task with a known outcome for comparison, connecting the identified state to an inspectable set of estimates, errors, and execution-variability measures.

In higher-volume applications with frequent new outcomes, time-ordered testing—evaluating predictions on later observations—and rolling performance assessment may be more suitable. LOOCV is a plausible engineering choice for similarly bounded property-valuation applications, subject to assessment; it is neither established as optimal here nor appropriate everywhere.

Assess the State Against Selected Standards

Within the LOOCV procedure, the configured DF produces held-out estimates, QC determines which executions qualify, and calculation and aggregation produce the state-characterization metrics. Governing standards determine what those metrics must support. Engineering Sufficiency (ES) includes assessing whether the chosen method and its results provide adequate support for the intended use. A standard must be selected before applying it as a qualification rule, with its requirements and interpretation made explicit.

The presentation's Standards A and B illustrate how different requirements can lead to different qualification decisions on the same empirical results. The State 0 to State 1 comparison is especially informative: one newly eligible sale enters and an older sale leaves. Under Standard B, State 1 fails the requirement that State 0 satisfied.

That failure illustrates a prospective intervention opportunity. If Standard B governed operation, reassessment of State 1 would prevent continued automated execution on the strength of State 0's result. The intervention would occur before waiting to observe an error on the next candidate prediction.

The change also illustrates why precision and accuracy require separate attention. A more tightly clustered set of execution outputs can accompany poorer agreement with realized outcomes. An improved precision metric does not settle whether the state satisfies the selected accuracy requirements.

Use Characterization Findings in Engineering Assessment

LOOCV-derived metrics contribute to Engineering Sufficiency (ES) assessment for the configured DF, evidence state, and measurement procedure. They permit the identified state to be examined against selected empirical requirements. A favorable result does not by itself establish the adequacy of all engineering methods or grant autonomous authority.

Before a subsequent automated prediction, the proposed discipline is to establish whether the current state's characterization remains applicable and meets governing requirements, renewing LOOCV where required. If a governing empirical requirement fails, earlier favorable performance cannot substitute for reassessment. This is the prospective value of the State 0/State 1 illustration: a changed evidence population can be examined before relying on it for another prediction.

State characterization does not decide whether an individual candidate is adequately described or within the qualified operating envelope. Those are separate candidate-admission questions. The Governance companion explains Standing, continuing authority, and candidate controls; the QC companion explains validity and record handling.

Feed Findings into Continuing Decision Node Engineering

The retained estimates, execution variability, errors, exclusions, and QC findings can inform improvements beyond the current state's assessment. They may expose weaknesses in evidence selection or representation, DF instructions, the characterization procedure, ensemble design, or the governing standards. Such findings identify questions for further engineering, rather than automatically qualifying a revised approach.

Realized prediction outcomes provide a related but distinct source of feedback. A frozen prediction can later be evaluated against Y; the completed sale may then enter a subsequent state and receive a fresh held-out LOOCV estimate. Keeping those records separate preserves their purposes, evidence boundaries, and timing while allowing both to inform the broader DN engineering effort.

Scope and Limitations

The laboratory directly characterized State 0 and State 1 and examined one transition between them. It did not perform LOOCV characterization for every state encountered in operational evaluation. Reassessment of changed states before subsequent candidate predictions is the proposed RPVDN operating discipline, not a claim of complete historical implementation.

Folds share comparative evidence, and repeated executions do not create additional independent property outcomes. Outcome redaction removes the realized subject outcome from the supplied evidence; it does not establish that the underlying AI model has never encountered related information. Separate Temporary Chats were used to reduce opportunities for cross-experiment carryover. The small historical populations, imperfect QC detection, and differences between completed sales and future candidates constrain interpretation of the results.

Prediction-window anchors: OPVAL used a twelve-month window ending on the prediction subject's actual closing date. The single PROSP case, P1P, instead used its August 15, 2026 purchase contract date. That cutoff excluded 249 PMC, which closed on September 8; P1P's ERs are dated September 11. Those dates do not alone establish when the completed-sale information became available in MLS. The historical prediction is retained unchanged.

The P1P cutoff is not an established optimal prospective rule. A prospective rule must explicitly specify its anchor and admit only evidence available at the prediction freeze. An anticipated closing date is a possible anchor requiring engineering assessment. These prediction-specific choices do not add a held-out-property closing-date restriction to LOOCV.

Standards A and B illustrate selected requirements; the laboratory does not establish them as optimized or deployment-validated thresholds. Passing a standard does not guarantee accuracy for the next candidate, and a state-change test is not proven to anticipate every deterioration. The proposed controls require deployment-specific engineering and evaluation.

Qualification evidence is conditional on the state, specification, model implementation, evidence representation, ensemble design, and verification procedures used. The DF Engineering and QC companions explain these dependencies. The release supplies a concrete basis for examining governance choices, rather than a complete scientific validation or production operating protocol.