ACUMEN: DECISION-AWARE MULTIMODAL DISTILLATION FOR ADAPTIVE COGNITIVE ASSESSMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Diagnosing cognitive impairment requires choosing which cognitive tests to administer. A comprehensive neuropsychological evaluation can take several hours, yet omitting an informative test can leave impairment undetected or its cause unresolved. Each choice is also made from an incomplete record: only some of the current visit's tests have results, and earlier visits may be absent or may have used a different version of the test battery. Positron emission tomography (PET) and cerebrospinal fluid (CSF) biomarkers are informative but costly or invasive, and they are usually ordered only after cognitive testing. Research cohorts record these measures and magnetic resonance imaging (MRI) for a subset of participants, often at different times from the clinical visit. These modalities can therefore supervise training, but cannot be assumed when a test is chosen. ACUMEN uses them to train a learner that diagnoses and selects tests from clinical information. It represents revealed scores, prior measurements, timing, and absent modalities, and distills multimodal current-state and temporal representations into separate clinical pathways for diagnosis and acquisition. On six-category final-diagnosis classification with National Alzheimer's Coordinating Center (NACC) data, the model reaches 93.2% accuracy with full multimodal input and 90.1% with clinical inputs alone, compared with 86.2% for a clinical multilayer perceptron. Training with modality dropout reduces the full-input to clinical-only accuracy gap from 8.6 to 3.1 percentage points. To extend this transfer to test selection, we specify a privileged teacher that predicts how much each test would reduce the loss of a fixed diagnostic predictor. A clinical-only acquisition student learns from the observed reductions and from teacher predictions fitted on other participants, keeping the observed targets when richer data are absent. Our analysis separates errors that change the ranking of tests from a common offset that affects only the decision to stop. The classification results establish the predictive foundation. The effect of this supervision on test selection has not yet been measured, and we specify the study that would measure it.

1 INTRODUCTION

Selecting an informative subset of tests and examinations is central to efficient clinical diagnosis. Redundant assessments increase patient and clinician burden, whereas omitting informative measurements can prolong diagnostic uncertainty. This trade-off is especially relevant to cognitive assessment for Alzheimer's disease and related dementias, where early clinical manifestations can be subtle and heterogeneous. Early detection matters because emerging disease-modifying treatments for Alzheimer's disease are intended for the stages of mild cognitive impairment (MCI) and mild dementia, yet an analysis of United States Medicare records estimated that only about 8% of expected MCI cases are diagnosed in primary care (Liu et al., 2024).

Choosing tests well requires knowing what each test can reveal. A diagnosis of MCI requires objective evidence of lower performance in one or more cognitive domains (Albert et al., 2011), and the domains affected early differ across causes of dementia. Memory impairment is the most common presentation of Alzheimer's disease, deficits in attention, executive function, and visuoperceptual ability can be prominent early in dementia with Lewy bodies, and the criteria for behavioral variant frontotemporal dementia include executive deficits with relatively spared memory and visuospatial function (McKhann et al., 2011; McKeith et al., 2017; Rascovsky et al., 2011). Cognitive tests

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

probe these complementary domains, but a comprehensive neuropsychological evaluation typically takes several hours (Schroeder et al., 2019). Which test is informative next therefore depends on the results already observed and on the patient's earlier assessments. This motivates a sequential learning problem: estimating which remaining test is most informative at each stage and whether further testing is warranted.

Learning such a policy from retrospective longitudinal records presents two challenges. First, the records are heterogeneous, incomplete, and asynchronous. No single cognitive assessment instrument fits all circumstances (Atri et al., 2025), so patients receive different tests, and the most useful next test differs from one patient to the next. Even a research cohort with a standardized battery records uneven histories. In the National Alzheimer's Coordinating Center (NACC) Uniform Data Set, participants enter at different stages of impairment, and several tests were replaced in 2015, for example the Mini-Mental State Examination by the Montreal Cognitive Assessment (MoCA) (Weintraub et al., 2018; Lee et al., 2025). A patient's history can therefore be short, mix battery versions, or lack a given test entirely. Magnetic resonance imaging (MRI), positron emission tomography (PET), and cerebrospinal fluid (CSF) measures are recorded for only some participants, and NACC notes that scans and clinical visits often occur at different times, sometimes years apart (NACC, 2026c;d). A current score, meaning the result of a test given at the present visit such as a MoCA total, must therefore be interpreted together with the patient's earlier scores on that test, the time since they were measured, and an observation mask that records which tests have been given.

Second, diagnostic prediction and test acquisition are distinct objectives. Diagnostic models are typically trained on visits whose tests have already been administered and predict a diagnosis from those recorded results, even when they tolerate missing inputs (Xue et al., 2024). Such a model does not estimate which unobserved test would most improve its prediction, so higher diagnostic accuracy does not necessarily imply better test selection. Active feature acquisition methods learn this choice from the results observed so far (Shim et al., 2018; Kachuee et al., 2019; Yoon et al., 2019; Li & Oliva, 2021; Covert et al., 2023; Gadgil et al., 2024). They learn it from the same measurements that the policy can later request, which leaves open whether measurements available only during training, such as research imaging, can improve it.

To study this question, we propose ACUMEN, a framework for adaptive cognitive-test acquisition from longitudinal multimodal supervision. ACUMEN represents demographics, prior assessments, the current scores revealed so far, observation masks, and measurement timing. Imaging and fluid biomarkers are often unavailable when a cognitive test is chosen. Amyloid PET and CSF testing are recommended only when the diagnosis remains uncertain after a cognitive evaluation (Johnson et al., 2013; Atri et al., 2025), and their limited accessibility, invasiveness, and cost preclude routine use in most primary and secondary care settings (Hansson et al., 2022). ACUMEN therefore uses these modalities to supervise training, so that its primary deployed learner needs only clinical information. A multimodal teacher encodes MRI, PET, and CSF observations into current-state and temporal representations, which are transferred to restricted clinical pathways through feature distillation (Lopez-Paz et al., 2016; Wang et al., 2020).

To supervise acquisition directly, we introduce a privileged test-value teacher for a fixed diagnostic predictor. For each test that can still be requested, the training procedure computes the predictor's loss reduction after revealing the recorded result. A teacher with access to longitudinal multimodal descriptors estimates this gain, and a clinical-only student learns from a mixture of the observed gain and a teacher prediction fitted without that participant. Without privileged information, the observed gain is kept. At deployment, the student uses only the permitted clinical history to select the test with the largest positive estimated net value or to stop. Figure 1 places these training and deployment interfaces within the assessment workflow.

Our analysis shows that an ideal privileged teacher preserves the conditional mean of this target while changing its variance. An ideal history-only teacher does the same, so any advantage of privileged descriptors concerns finite learning. We express the bias of a fitted teacher, and we decompose gain-prediction error into errors in contrasts between tests, which change their ranking, and a common offset, which mainly changes the comparison with STOP. A uniform value-error bound also bounds one-step acquisition regret.

Our experiments on six-category NACC diagnosis establish the representation-transfer foundation (Section 4). With MRI, PET, and CSF, the model reaches 93.2% accuracy, with its largest recall

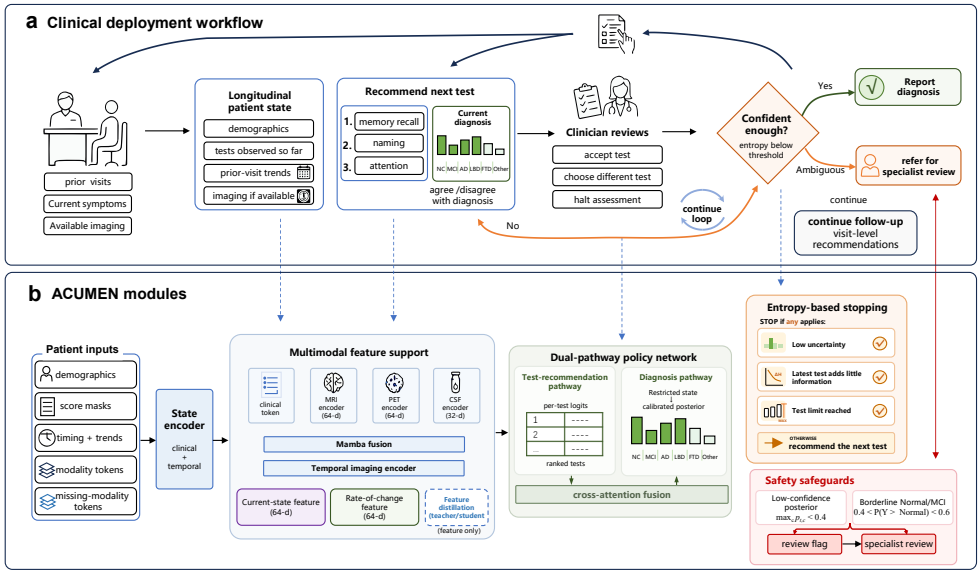


Figure 1: ACUMEN clinical workflow (a) and architecture (b), connecting multimodal state and temporal representations, feature distillation, and the diagnosis and recommendation pathways. Section 4 evaluates the diagnostic model, and Section 3.3 adds gain supervision. The entropy, calibration, and safety/review options illustrate the broader workflow, whereas the rule in Section 3.4 stops when no test has positive estimated net value. Clinician interaction is outside the computational evaluation.

gains in the categories where a clinical multilayer perceptron performs worst. With clinical inputs alone, the same model reaches 90.1%, against 86.2% for the perceptron. These results show transfer to restricted prediction, not an acquisition benefit. Measuring that benefit requires the study of test ranking, selected-test value, stopping, and diagnostic loss against the number of acquired tests described in Section 5, which we have not yet run.

2 Problem and Related Work

Adaptive cognitive assessment. At visit v, the learner chooses which cognitive tests to request before predicting a recorded diagnostic category $Y_v \in \{1, \dots, C\}$. Its information after k acquisitions is

$$h_{vk} = \left(B_v, V_v, (a_j, X_{va_j})_{j=1}^{k}\right). \tag{1}$$

Here B_v is permitted pre-visit clinical history, V_v contains modalities available when testing begins, and action a reveals a specified group of current scores X_{va}. Table 1 separates these inputs by how they change during an assessment.

Table 1: Patient information available to the learner. The primary clinical-only deployment uses $V_v = \varnothing$. Richer training-only observations do not become deployment inputs.

State block	Permitted information	Within-visit update
History B_v	Demographics, prior test values and means, slopes, observation counts, time since measurement, and visit gaps	Fixed during acquisition
Current scores	Revealed normalized scores, acquisition masks, and summaries of revealed results	Updated after each test
Modalities V_v	Already available MRI, PET, or CSF measurements, with availability and measurement ages	Fixed at episode start

Visit index v tracks longitudinal progression, whereas acquisition index k tracks information revealed within one visit. A reveal updates current scores and their dependent summaries, not the prior-visit

history. Unrevealed scores and missing historical values use masks and padding, so an observed zero remains distinct from an unavailable value. Measurement ages distinguish an old score from a current result. Diagnosis codes, label-derived inputs, future visits, and unrevealed current scores never enter the diagnosis pathway, and gain supervision also excludes them from the acquisition pathway.

The learner requests an action from the legal set $\mathcal{L}(h)$ or chooses STOP, then predicts through $q_\theta(h) \in \Delta^{C-1}$, the C-class probability simplex. Legality respects the battery, prerequisites, previous acquisitions, and remaining budget. Each action returns one or more scores, so the number of action groups need not equal the score-vector width. The objective trades terminal prediction loss against cost $c_a \geq 0$. Unit costs count groups. MRI, PET, and biomarker acquisition are outside this cognitive-test library.

Training information and retrospective scope. Training may use richer observations acquired by the episode cutoff and hidden from the deployed learner. For a fixed action library $\mathcal{A}$, gain training and replay use visits with usable results for every action, drawn under a participant-balanced law $P_{\mathcal{A}}$ (Appendix C). The historical storage mask is evaluator information, not a policy input, and this selected population does not by itself establish performance on incomplete records (von Kleist et al., 2023; 2025).

Related foundations. The ACUMEN diagnostic model builds on distillation for incomplete modalities, longitudinal displacement representations, and history distillation (Wang et al., 2020; Ouyang et al., 2022; Karimian et al., 2026). Multimodal dementia classifiers already tolerate missing inputs and predict etiology from whatever has been recorded (Xue et al., 2024), but do not choose the next test. NOCTA learns longitudinal representations and future-plan values, and REACT selects context and measurements over time (Dinh et al., 2026; Qu et al., 2026), whereas here longitudinal context informs test selection within the present assessment.

DIME learns acquisition values by regressing the prediction model's before/after loss changes while jointly training prediction and value networks (Gadgil et al., 2024). Privileged-feature distillation (PFD) trains a teacher on regular and training-only inputs, then fits a restricted student to data and teacher losses (Yang et al., 2022). Gain supervision combines the two: a privileged teacher predicts the one-test loss reduction of a fixed diagnostic predictor and supervises the restricted acquisition learner. Informed asymmetric critics also use extra training information to value a restricted policy (Ebi et al., 2026; Venkatraman et al., 2026), but they estimate policy returns rather than the one-test gains used here for myopic selection.

3 ACUMEN: FROM PATIENT REPRESENTATION TO TEST SELECTION

ACUMEN connects three operations: representing the patient's available history, transferring information from richer modalities, and deciding which additional test to request. Figure 1 shows these interfaces. Section 3.1 describes the representation of the diagnostic model evaluated in Section 4. Section 3.2 describes feature transfer, first as that model uses it and then as gain supervision extends it. Sections 3.3 and 3.4 complete gain supervision for test selection. Appendix A details the diagnostic model, and Appendix C specifies gain supervision.

3.1 REPRESENT THE PATIENT ACROSS MODALITIES AND TIME

The clinical state in Section 2 describes both what is known and how recently it was measured. Its revealed-score masks distinguish an unrequested test from an observed low score, while prior measurements distinguish a new change from a persistently low level. This state is the restricted learner's input. MRI, PET, and cerebrospinal fluid (CSF) supply additional views of the patient when they are available.

The multimodal model encodes each source separately before fusion: 89 regional MRI volumes, 19 regional PET uptake ratios, and 7 CSF biomarker values with 4 assay indicators (Appendix Table 4). The MRI encoder is a Mamba-based regional model, and PET and CSF use feedforward encoders. After projection to a common token width, a two-block Mamba stack fuses the available tokens into a 128-dimensional representation (Gu & Dao, 2023). An absent source is represented by an availability indicator and a learned missing-modality token, never by measured zeros.

The temporal module separates a current-state vector r^{cur} from a learned temporal vector r^{temp}, each with 64 coordinates. The first summarizes condition at the observation time, while the second encodes temporal structure learned from longitudinal data. These vectors provide different supervision channels for the clinical student. They are learned features, so a temporal output inferred from one scan is not an observed individual trajectory. Explicit availability and timing information are therefore part of the representation, rather than details left to imputation.

For example, consider a patient with a prior memory score but no current memory result. The history supplies the old score, its date, and any valid trend, while the current score remains masked. Requesting immediate recall reveals only that action's fields. The result can change the diagnosis and the value of a later delayed-recall action, whose prerequisite is now met, but it does not reveal any unrequested naming or attention score. A training-only MRI can likewise supervise the representation without entering this patient's clinical-only inference state.

This design separates two uses of richer data within one trained model. When MRI, PET, or CSF is available at inference, the model consumes it directly. When a modality is absent, a learned missing-modality token takes its place, and the model must rely on what training with that modality has transferred into its clinical pathway. Section 4 compares these regimes to measure how much predictive performance crosses that input boundary.

3.2 Transfer features without losing clinical context

Feature distillation trains the clinical representation to predict the teacher's current and temporal components. The diagnostic model combines feature matching with coherence and variance regularization, and uses modality dropout to expose training to absent inputs. Matching intermediate features, rather than the teacher's diagnostic probabilities, accommodates teachers and students with different label spaces.

The clinical pathways serve different purposes. The diagnosis pathway maps available evidence to diagnostic probabilities. The recommendation pathway of the diagnostic model is trained to predict which tests are administered at the next visit. On held-out NACC transitions, it places CDR Global and CDR Sum of Boxes in every top-three list and almost never recommends a memory test (Appendix B.5). Agreement with administration therefore does not measure how much a test improves diagnosis. Gain supervision replaces these recommendation scores with signed estimates of diagnostic loss reduction. Its acquisition pathway receives the candidate action as well as the clinical state. Separate trainable parameters let its objective change the selection model without changing an already fitted diagnosis model. For a pathway $r \in \{y, u\}$, let its encoder produce current and temporal segments $z_r^{\mathrm{cur}}(h)$ and $z_r^{\delta}(h)$ together with a clinical context channel $c_r(h)$. The diagnosis and gain heads have the form

$$q_\theta(h) = \mathrm{softmax}\, f_y(z_y^{\mathrm{cur}}(h), z_y^{\delta}(h), c_y(h)), \tag{2}$$

$$g_\phi(h, a) = f_u(z_u^{\mathrm{cur}}(h), z_u^{\delta}(h), c_u(h), a). \tag{3}$$

Here f_y returns C class logits, f_u returns one signed gain, and θ, ϕ include their respective encoders and heads. Feature projections from the two segments receive teacher supervision. The context channel remains outside those projections, since matching a teacher feature can discard clinical details needed for a particular test decision. No acquisition feature feeds back into q_θ.

For gain supervision, let F be a frozen multimodal encoder and W_1 the latest eligible richer observation before the episode cutoff. Define $z_T^{\mathrm{cur}} = F(W_1)$. To give the temporal target an observed-pair interpretation, encode an earlier observation W_0 and W_1 using the same nonempty modality subset M, and set

$$z_T^{\delta} = \frac{F(W_1^M) - F(W_0^M)}{\Delta\tau}, \qquad \Delta\tau = \tau_1 - \tau_0 > 0. \tag{4}$$

The pair contains distinct measurements in chronological order. Past-only alignment windows, a minimum interval, and individual measurement ages specify how asynchronously collected modalities enter each observation. This target describes feature change per unit time, rather than an identified biological progression rate. It differs from the learned temporal output r^{temp} of the diagnostic model. Appendix C gives a concrete encoder-fitting objective and uses a shared 64-dimensional output space for F.

A single observation supplies current-state supervision but no change target, and validity masks distinguish a missing pair from a valid zero change. The feature loss $\mathcal{L}_{\text{feat}}$ is masked squared error on the valid current and change targets, with separate normalizers for each. With a pair-matched current endpoint and known elapsed time, current/change coordinates carry the same information as the paired endpoints. Their purpose is to structure the representation, and any advantage must be tested against matched endpoint inputs (Appendix C).

3.3 Learn the value of a test for the deployed predictor

Feature transfer supplies a representation. Acquisition additionally requires a target expressing what a new measurement would change. Fit and freeze the diagnostic predictor q_θ. For a labeled training history h_i and legal test a, calculate

$$\ell(p, Y) = \tfrac{1}{2}\|p - e_Y\|_2^2, \tag{5}$$

$$d_{i,a} = \ell(q_\theta(h_i), Y_i) - \ell(q_\theta(h_i \oplus (a, X_{ia})), Y_i). \tag{6}$$

Here e_Y is the one-hot label, $\ell \in [0, 1]$ is normalized Brier loss, and $\oplus$ appends the revealed scores and recomputes dependent features. The gain $d_{i,a} \in [-1, 1]$ measures the realized benefit of revealing a to the same predictor. The label and candidate result enter this offline calculation, but are unavailable to the pre-acquisition decision.

Regression on these gains follows DIME's loss-reduction principle (Gadgil et al., 2024), but freezing q_θ separates acquisition supervision from changes to the diagnostic predictor, which DIME trains jointly. Richer observations can then supervise the same target. Let Z_i contain the available current and change descriptors with their validity and modality indicators, elapsed time, and measurement ages, and let $I_i = 1$ indicate that such a descriptor exists. A value teacher $m(h_i, Z_i, a)$ predicts $d_{i,a}$, the gain of the restricted predictor, which differs from the gain of a predictor that already observes the richer modalities.

Randomly divide value-fitting participants into folds independently of their measurements and labels. Write $f(i)$ for example i's participant fold, and $\widehat{m}^{-f(i)}$ for a teacher fitted and selected without that fold. Its excluded-participant prediction supplies the mixed target

$$t_{i,a} = d_{i,a} + \rho I_i\big(\widehat{m}^{-f(i)}(h_i, Z_i, a) - d_{i,a}\big), \qquad 0 \le \rho \le 1. \tag{7}$$

The weight ρ controls reliance on the teacher. For $I_i = 0$, the target remains $d_{i,a}$, and all examples retain their original population weights. Thus missing privilege does not remove an episode from acquisition training.

The deployed gain model $g_\phi(h, a)$ receives only the permitted history and candidate action. It minimizes

$$\mathcal{L}_{\text{acq}} = \mathbb{E}_{i,a}\big[(g_\phi(h_i, a) - t_{i,a})^2\big] + \beta\, \mathcal{L}_{\text{feat}}, \qquad \beta \ge 0. \tag{8}$$

The expectation uses participant-balanced episodes from $P_{\mathcal{A}}$, a fixed legal-history sampler, and equal weights for legal actions within each sampled history. Figure 2 shows the resulting training and deployment paths.

This construction applies privileged-feature distillation to acquisition gains (Yang et al., 2022). For fixed teacher targets and mixture weights, squared error against $t_{i,a}$ has the same student gradients as the corresponding weighted sum of data and teacher squared losses (Appendix D.2). The loss family therefore comes from PFD, and the new element is its target: the teacher predicts the loss reduction of the deployed clinical predictor from longitudinal multimodal descriptors.

3.4 Select, reveal, and update

The desired one-test net value is

$$u_q^{\mathcal{A}}(h, a) = \mathbb{E}_{P_{\mathcal{A}}}[\ell(q_\theta(h), Y) - \ell(q_\theta(h \oplus (a, X_a)), Y) \mid h, a] - \lambda c_a, \tag{9}$$

where $\lambda \ge 0$ converts acquisition cost to predictive-loss units. The conditional law uses the same nonanticipating history/action sampler as training, whose choices depend only on information already revealed. Estimate the value by $g_\phi(h, a) - \lambda c_a$ and assign STOP value zero. Request the largest positive legal value using a fixed tie rule. After each reveal, update the patient state, diagnostic

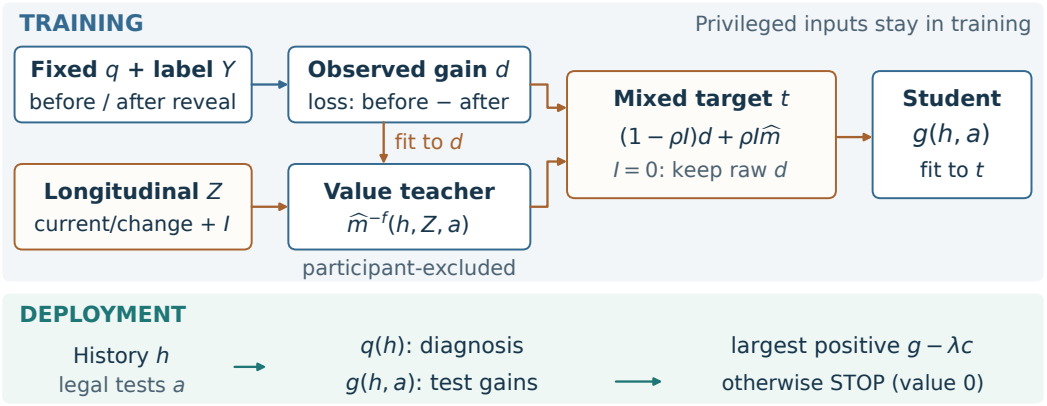


Figure 2: Gain supervision. The observed gain d and the teacher's target use the same fixed q, and each teacher predicts only for participants excluded from its fitting and selection. The student learns from the mixture set by ρ, keeping raw gains where $I = 0$. At deployment, q and g use only the permitted history, legal actions, and budget.

probabilities, and remaining test values. Stop when no legal value is positive, no action remains, or the cap is reached. This myopic rule can miss useful combinations of tests, as shown in Appendix D.

Participants are disjoint across predictor fitting D_{fit}, value fitting D_{value}, validation D_{val}, and evaluation D_{test}. The feature encoder and q_θ are fitted on D_{fit} and frozen before gains are computed, the value models are fitted within D_{value}, and settings are selected on D_{val}. Inference uses no teacher calls or cached privileged features.

3.5 What the supervision can change

The analysis links the supervision target to two decisions, which test to request and whether to stop (derivations in Appendix D).

Preserving the learning target. Fix q_θ, $P_{\mathcal{A}}$, and the nonanticipating history/action sampler. Write $G = (H, A)$ for a sampled history and legal test, let D be the realized gain in Equation (6), and put $\mu(G) = \mathbb{E}[D \mid G]$. For an ideal teacher $m^\star(G, Z) = \mathbb{E}[D \mid G, Z]$, with I included in Z, the mixed target $T_\rho = D + \rho I(m^\star - D)$ satisfies

$$\mathbb{E}[T_\rho \mid G] = \mu(G), \tag{10}$$

so the target keeps the conditional mean gain of the same fixed predictor, including where $I = 0$. A fitted teacher, selected independently of the supervised participant, has error $e = \widehat{m} - m^\star$ on applicable cases, and its practical target satisfies

$$\mathbb{E}[\widehat{T}_\rho \mid G] - \mu(G) = \rho\,\mathbb{E}[Ie \mid G]. \tag{11}$$

Cross-fitting prevents self-fitting but does not remove this bias. An ideal history-only teacher also returns $\mu(G)$, so the case for richer descriptors concerns finite learning and must be tested against matched history-only distillation.

Separating test choice from stopping. At a held-out history h with $J \geq 2$ legal tests, let $\mathbf{D}_h$ and $\mathbf{g}_h$ contain the realized and predicted gains, let $\mathbf{e}_h = \mathbf{g}_h - \mathbf{D}_h$ with mean $\bar{e}_h$, and let $\mathbf{C}_h = \mathbf{I}_J - \mathbf{1}\mathbf{1}^\top/J$ subtract a vector's mean. Orthogonality gives

$$\|\mathbf{e}_h\|_2^2 = J\bar{e}_h^2 + \|\mathbf{C}_h\mathbf{e}_h\|_2^2. \tag{12}$$

The centered term measures errors in between-test contrasts, which change the ranking. A shared offset changes only the comparison with STOP, so correcting it can reduce gain mean squared error without changing which test ranks first. Lower contrast error alone does not guarantee better decisions, so Appendix F compares test discrimination, selected-test value, and stopping separately.

4 EXPERIMENTS

We evaluate the ACUMEN diagnostic model on the National Alzheimer's Coordinating Center (NACC) Uniform Data Set. The data contain 205,908 visits from 55,004 participants, 37,633 of whom have two or more visits. MRI, PET, and CSF measurements are available at about 33%, 15%, and 22% of visits, and all three at about 8%, so most visits lack at least one modality. The task is six-category classification of the recorded final diagnosis: Normal, mild cognitive impairment (MCI), Alzheimer's disease (AD), Lewy body dementia (LBD), frontotemporal dementia (FTD), and Other Dementia. Participants are split 70/10/20 into training, validation, and test partitions, stratified by final diagnosis. On the test partition, we report accuracy, macro-averaged F_1, micro-averaged one-vs-rest area under the receiver operating characteristic curve (AUROC), and recall within each category. The evaluated model combines Mamba fusion, the learned temporal encoder, and feature transfer with matching, coherence, and variance penalties (Appendix A). The gain teacher of Section 3.3 is not part of these experiments, and Section 5 describes the acquisition study it requires.

4.1 DIAGNOSTIC PERFORMANCE AND CLINICAL-ONLY TRANSFER

Table 2 compares ACUMEN with seven diagnostic baselines. With MRI, PET, and CSF available, ACUMEN reaches 93.2% accuracy, a macro F_1 of 0.84, and an AUROC of 0.97, the highest values in the comparison. Its accuracy exceeds the clinical MLP by 7.0 percentage points and the strongest fusion baseline, a Perceiver, by 3.8 points, with macro-F_1 gains of 0.13 and 0.06, so the advantage holds when every category receives equal weight.

Table 2: Six-category diagnosis on the NACC test partition (point estimates). Subscripts give the change from the clinical-only MLP. The two ACUMEN rows are one trained model, with clinical inputs only and with MRI, PET, and CSF added. The MRI-only and PET-only baselines are separate models. AUROC is micro-averaged one-vs-rest. We **bold** the best value and underline the second best.

Diagnostic model	Accuracy (%)	Macro F_1	AUROC
Clinical-only MLP	86.2	0.71	0.91
MRI-only baseline	$78.4_{\downarrow 7.8}$	$0.58_{\downarrow 0.13}$	$0.84_{\downarrow 0.07}$
PET-only MLP	$75.1_{\downarrow 11.1}$	$0.52_{\downarrow 0.19}$	$0.82_{\downarrow 0.09}$
Early fusion, concatenation	$87.8_{\uparrow 1.6}$	$0.74_{\uparrow 0.03}$	$0.92_{\uparrow 0.01}$
Late fusion, averaging	$88.1_{\uparrow 1.9}$	$0.75_{\uparrow 0.04}$	$0.93_{\uparrow 0.02}$
Attention fusion, Perceiver	$89.4_{\uparrow 3.2}$	$0.78_{\uparrow 0.07}$	$0.94_{\uparrow 0.03}$
CLIP-Medical linear probe	$89.1_{\uparrow 2.9}$	$0.76_{\uparrow 0.05}$	$0.93_{\uparrow 0.02}$
ACUMEN, clinical inputs only	$\underline{90.1}_{\uparrow 3.9}$	$\underline{0.79}_{\uparrow 0.08}$	$\underline{0.95}_{\uparrow 0.04}$
ACUMEN, full multimodal input	$\mathbf{93.2}_{\uparrow 7.0}$	$\mathbf{0.84}_{\uparrow 0.13}$	$\mathbf{0.97}_{\uparrow 0.06}$

With no MRI, PET, or CSF at inference, the same model reaches 90.1% accuracy and 0.79 macro F_1, 3.9 points and 0.08 above the clinical MLP, and its AUROC rises from 0.91 to 0.95. Part of the benefit of multimodal training is therefore retained with clinical information alone, although full input remains 3.1 accuracy points higher.

4.2 MODALITY AVAILABILITY AND TRAINING-TIME DROPOUT

The deployment benefit depends on how the model responds to missing modalities. Training with modality dropout at $p = 0.3$ (Appendix Table 5) costs 0.6 points under full input, where accuracy falls from 93.8% to 93.2%, but gains 4.9 points with clinical inputs alone, where it rises from 85.2% to 90.1%. The gap between the two deployment conditions therefore shrinks from 8.6 to 3.1 points.

Figure 3(a) withholds every subset of modalities from the same trained model. Accuracy falls from 93.2% with all three modalities to 90.8–92.1% with two and 88.4–90.2% with one, a 4.8-point span. Adding PET or CSF alone lowers accuracy below the clinical-only condition, whereas adding MRI never lowers it (Appendix B). These results describe controlled withholding from one fitted model, which may differ from performance in patients whose modalities are naturally absent.

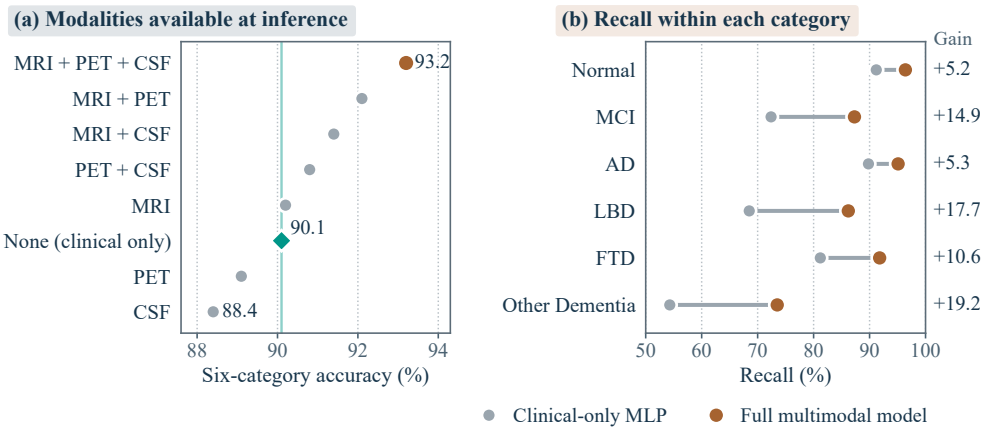


Figure 3: Diagnostic results on the NACC test partition. (a) Accuracy of one trained model with each subset of MRI, PET, and CSF available and clinical inputs always present: orange marks full input, the teal diamond and line the clinical-only condition, and gray the other subsets. PET or CSF alone lowers accuracy below the clinical-only level. (b) Recall within each category for the clinical-only MLP (gray) and the full multimodal model (orange), with gains in percentage points, which are largest where baseline recall is lowest. Exact values are in Appendix Tables 6 and 9.

4.3 Component and category results

Removing one component at a time from the full model (Appendix Table 7) lowers accuracy from 93.2% to 91.8% without the separation between the recommendation and diagnosis pathways, and to 92.5% without per-test staleness and trend modulation of the recommendation scores, while macro F_1 falls from 0.84 to 0.78 and 0.82. These whole-configuration comparisons do not isolate how recommendation-score modulation affects diagnosis. Appendix B also reports a knowledge-distillation replacement for feature transfer and single-modality teacher features.

Figure 3(b) shows that, relative to the clinical MLP, the full model gains most where the baseline is weakest: 19.2 points of recall for Other Dementia, 17.7 for LBD, and 14.9 for MCI, compared with about 5 points for Normal and AD, whose baseline recall already exceeds 89%. Other Dementia keeps the lowest recall, at 73.5%, a limitation that overall accuracy hides. These measurements do not identify which modality explains an individual error or which test should be requested next.

5 Discussion

ACUMEN connects patient history, multimodal representation, clinical feature transfer, and test selection through an explicit information boundary. With clinical inputs alone, the model improves accuracy by 3.9 percentage points over the clinical MLP while remaining 3.1 points below full input, and with full input its recall gains concentrate in the categories that clinical information separates least well. Modality dropout favors the clinical-only condition at a small cost under full input, so the representation should match the inputs available at deployment.

The gain teacher addresses the question these results leave open: whether richer observations can also improve which test is requested. Unlike the imitation target of the recommendation pathway (Appendix B.5), it supervises the loss reduction of a fixed diagnostic predictor, and Section 3.5 separates ordering errors from a shared offset that affects only stopping. Its effect on acquisition has not yet been measured. The required study (Appendix F) fixes q_θ, inputs, and tuning, and compares direct gain regression, feature transfer, matched history-only value distillation, and privileged value distillation on contrast error and offset (Equation (12)), on selected-test value, and on Brier loss against the number of acquired tests, alongside joint-training DIME (Gadgil et al., 2024).

The claims concern prediction of recorded labels and computational acquisition, not patient benefit or clinician use. Retrospective acquisition also depends on the action library, the complete-record population, and whether stored results remain valid under a simulated reveal order (Appendix E).

AI USE STATEMENT

In this work, we used generative AI tools to edit and revise the writing of the paper. This assistance included improving readability, restructuring sections and drafting revised passages, among them passages that report and interpret the results, locating and verifying references for background statements, formatting references, writing plotting code for figures of the reported results, and checking the manuscript's numbers and cross-references against the authors' records. The research questions, study design, data processing, model development, implementation, training, and experiments were carried out by the authors. We have reviewed all AI-assisted text, references, code, and figures: each added reference was checked against its source, and each plotted value was checked against the corresponding table. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work concerns adaptive measurement for cognitive assessment. It reports computational model performance against recorded clinical categories and specifies a retrospective acquisition study. It does not claim prospective patient benefit or individual-patient safety.

The diagnostic experiments use deidentified NACC data obtained through the data request process under its Data Use Agreement (NACC, 2026a;b). The contributing Alzheimer's Disease Research Centers obtained written informed consent from participants and co-participants (NACC, 2026e). This work involves no new participant recruitment or contact.

The reported results are aggregate classification and ranking measurements, and the analytical examples use explicitly constructed distributions with no participant measurements. Neither establishes equitable performance across patient groups or a clinical safety guarantee. Most NACC participants are White and highly educated, so performance in underrepresented groups requires separate evaluation before any clinical use. Recorded diagnoses also reflect clinical assessment that may itself use cognitive tests, so agreement with them is not an independent biological standard. Clinician interaction and prospective medical outcomes are outside this submission's evaluation scope.

REPRODUCIBILITY STATEMENT

Section 4 states the data, task, partitions, and metrics of the diagnostic experiments, and Appendix A specifies the evaluated model's patient state, modality encoders, temporal representation, feature-transfer objective, and prediction pathways. Table 2 and Appendix B report every diagnostic result numerically, including the recommendation pathway's evaluation split. Gain supervision is specified in Section 3 and Appendix C, its analytical results are proved under the stated assumptions in Appendix D, and Appendix F defines the acquisition protocol, including the version-specific action libraries and the returned fields of each action. The analytical examples in Appendix E use explicitly stated probability laws and need no participant data. The NACC data are available to qualified researchers through NACC's data request process (NACC, 2026a).

REFERENCES

Marilyn S. Albert, Steven T. DeKosky, Dennis Dickson, Bruno Dubois, Howard H. Feldman, Nick C. Fox, Anthony Gamst, David M. Holtzman, William J. Jagust, Ronald C. Petersen, Peter J. Snyder, Maria C. Carrillo, Bill Thies, and Creighton H. Phelps. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia*, 7(3):270–279, 2011. doi: 10.1016/j.jalz.2011.03.008.

Alireza Atri, Bradford C. Dickerson, Carolyn Clevenger, Jason Karlawish, David Knopman, Pei-Jung Lin, Mary Norman, Chiadi Onyike, Mary Sano, Susan Scanland, et al. Alzheimer's Association clinical practice guideline for the diagnostic evaluation, testing, counseling, and disclosure of suspected Alzheimer's disease and related disorders (DETeCD-ADRD): executive summary of recommendations for primary care. *Alzheimer's and Dementia*, 21(6):e14333, 2025. doi: 10.1002/alz.14333.

Ian Connick Covert, Wei Qiu, Mingyu Lu, Na Yoon Kim, Nathan J. White, and Su-In Lee. Learning to maximize mutual information for dynamic feature selection. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 6424–6447. PMLR, 2023. URL https://proceedings.mlr.press/v202/covert23a.html.

Dzung Dinh, Boqi Chen, Yunni Qu, Marc Niethammer, and Junier Oliva. Nocta: Non-greedy objective cost-tradeoff acquisition for longitudinal data, 2026. URL https://arxiv.org/abs/2507.12412.

Daniel Ebi, Damien Ernst, Klemens Böhm, and Gaspard Lambrechts. Informed asymmetric actor-critic: Leveraging privileged signals beyond full-state access. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=7G43VVT4SI.

Soham Gadgil, Ian Covert, and Su-In Lee. Estimating conditional mutual information for dynamic feature selection. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=Oju2Qu9jvn.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023. doi: 10.48550/arXiv.2312.00752. URL https://arxiv.org/abs/2312.00752.

Osman Berke Guney, Ketan Suhaas Saichandran, Karim Elzokm, Ziming Zhang, and Vijaya B. Kolachalama. Active feature acquisition via explainability-driven ranking. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 20748–20765. PMLR, 2025. URL https://proceedings.mlr.press/v267/guney25a.html.

Oskar Hansson, Rebecca M. Edelmayer, Adam L. Boxer, Maria C. Carrillo, Michelle M. Mielke, Gil D. Rabinovici, Stephen Salloway, Reisa Sperling, Henrik Zetterberg, and Charlotte E. Teunissen. The Alzheimer's Association appropriate use recommendations for blood biomarkers in Alzheimer's disease. *Alzheimer's and Dementia*, 18(12):2669–2686, 2022. doi: 10.1002/alz.12756.

Julian P. T. Higgins, Ian R. White, and Angela M. Wood. Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clinical Trials*, 5(3):225–239, 2008. doi: 10.1177/1740774508091600. URL https://journals.sagepub.com/doi/10.1177/1740774508091600.

Keith A. Johnson, Satoshi Minoshima, Nicolaas I. Bohnen, Kevin J. Donohoe, Norman L. Foster, Peter Herscovitch, Jason H. Karlawish, Christopher C. Rowe, Maria C. Carrillo, Dean M. Hartley, Saima Hedrick, Virginia Pappas, and William H. Thies. Appropriate use criteria for amyloid PET: A report of the amyloid imaging task force, the society of nuclear medicine and molecular imaging, and the Alzheimer's Association. *Journal of Nuclear Medicine*, 54(3):476–490, 2013. doi: 10.2967/jnumed.113.120618.

Mohammad Kachuee, Orpaz Goldstein, Kimmo Kärkkäinen, Sajad Darabi, and Majid Sarrafzadeh. Opportunistic learning: budgeted cost-sensitive learning from data streams. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=S1eOHo09KX.

Banafsheh Karimian, Soufiane Belharbi, Alexis Guichemerre, Luke McCaffrey, Mohammadhadi Shateri, and Eric Granger. Longitudinal risk prediction in mammography with privileged history distillation, 2026. URL https://arxiv.org/abs/2603.15814.

Rickard K. A. Karlsson, Martin Willbo, Zeshan M. Hussain, Rahul G. Krishnan, David Sontag, and Fredrik Johansson. Using time-series privileged information for provably efficient learning of prediction models. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 5459–5484. PMLR, 2022. URL https://proceedings.mlr.press/v151/k-a-karlsson22a.html.

Michael L. Lee, Seo-Eun Choi, Shubhabrata Mukherjee, Phoebe Scollard, Connie Nakano, Laura E. Gibbons, et al. Bridging the gap: addressing NACC's evolving cognition batteries across UDS versions. *Alzheimer's and Dementia*, 21(11):e70832, 2025. doi: 10.1002/alz.70832.

Yang Li and Junier Oliva. Active feature acquisition with generative surrogate models. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 6450–6459. PMLR, 2021. URL https://proceedings.mlr.press/v139/li21p.html.

Roderick J. A. Little and Donald B. Rubin. *Statistical Analysis with Missing Data*. Wiley, 3rd edition, 2019.

Ying Liu, H. Jun, A. Becker, C. Wallick, and S. Mattke. Detection rates of mild cognitive impairment in primary care for the United States Medicare population. *The Journal of Prevention of Alzheimer's Disease*, 11(1):7–12, 2024. doi: 10.14283/jpad.2023.131.

David Lopez-Paz, Léon Bottou, Bernhard Schölkopf, and Vladimir Vapnik. Unifying distillation and privileged information. In *International Conference on Learning Representations*, 2016. URL https://arxiv.org/abs/1511.03643.

Charles F. Manski. Nonparametric bounds on treatment effects. *American Economic Review*, 80(2): 319–323, 1990.

Charles F. Manski. *Partial Identification of Probability Distributions*. Springer, 2003.

Ian G. McKeith, Bradley F. Boeve, Dennis W. Dickson, Glenda Halliday, John-Paul Taylor, Daniel Weintraub, Dag Aarsland, James Galvin, Johannes Attems, Clive G. Ballard, Ashley Bayston, Thomas G. Beach, Frederic Blanc, Nicolaas Bohnen, Laura Bonanni, et al. Diagnosis and management of dementia with Lewy bodies: fourth consensus report of the DLB Consortium. *Neurology*, 89(1):88–100, 2017. doi: 10.1212/WNL.0000000000004058.

Guy M. McKhann, David S. Knopman, Howard Chertkow, Bradley T. Hyman, Clifford R. Jack, Claudia H. Kawas, William E. Klunk, Walter J. Koroshetz, Jennifer J. Manly, Richard Mayeux, Richard C. Mohs, John C. Morris, Martin N. Rossor, Philip Scheltens, Maria C. Carrillo, Bill Thies, Sandra Weintraub, and Creighton H. Phelps. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups. *Alzheimer's and Dementia*, 7(3):263–269, 2011.

NACC. *NACC Uniform Data Set: Coding Guidebook for Initial Visit Packet*. National Alzheimer's Coordinating Center, 2008. URL https://files.alz.washington.edu/documentation/uds2-ivp-guidebook.pdf. Version 2.0, February 2008. Accessed September 7, 2026.

NACC. *NACC Uniform Data Set, Follow-up Visit Packet, Form C1: Neuropsychological Battery Summary Scores*. National Alzheimer's Coordinating Center, 2015a. URL https://files.alz.washington.edu/documentation/uds3-fvp-c1.pdf. Version 3.0, March 2015. Accessed September 17, 2026.

NACC. *NACC Uniform Data Set, Initial Visit Packet, Form C2: Neuropsychological Battery Scores*. National Alzheimer's Coordinating Center, 2015b. URL https://files.alz.washington.edu/documentation/uds3-ivp-c2.pdf. Version 3.0, March 2015. Accessed September 17, 2026.

NACC. *NACC Uniform Data Set, Initial Visit Packet, Form D1: Clinician Diagnosis*. National Alzheimer's Coordinating Center, 2015c. URL https://files.alz.washington.edu/documentation/uds3-ivp-d1.pdf. Version 3.0, March 2015. Accessed September 7, 2026.

NACC. *NACC Uniform Data Set: Researchers Data Dictionary*. National Alzheimer's Coordinating Center, 2025. URL https://files.alz.washington.edu/documentation/uds3-rdd.pdf. Version 3.0, March 2015. Revised April 8, 2025. Accessed September 7, 2026.

NACC. Data request process, 2026a. URL https://www.naccdata.org/data-request-process/. Accessed September 15, 2026.

NACC. Data use agreement, 2026b. URL https://files.alz.washington.edu/documentation/nacc_data_use_agreement.pdf. Accessed September 15, 2026.

NACC. The NACC researcher's guide, 2026c. URL https://www.naccdata.org/the-nacc-researchers-guide. Accessed September 7, 2026.

NACC. Imaging data, 2026d. URL https://www.naccdata.org/about-nacc-data/imaging-data. Accessed September 22, 2026.

NACC. Longitudinal neurocognitive and clinical phenotype data, 2026e. URL https://www.naccdata.org/about-nacc-data/longitudinal-neurocognitive-and-clinical-phenotype-data. Accessed September 15, 2026.

Jiahong Ouyang, Qingyu Zhao, Ehsan Adeli, Greg Zaharchuk, and Kilian M. Pohl. Self-supervised learning of neighborhood embedding for longitudinal MRI. *Medical Image Analysis*, 82:102571, 2022. doi: 10.1016/j.media.2022.102571. URL https://doi.org/10.1016/j.media.2022.102571.

Yunni Qu, Dzung Dinh, Grant King, Whitney Ringwald, Bing Cai Kok, Kathleen Gates, Aidan Wright, and Junier Oliva. Relaxed efficient acquisition of context and temporal features, 2026. URL https://arxiv.org/abs/2603.11370.

Katya Rascovsky, John R. Hodges, David Knopman, Mario F. Mendez, Joel H. Kramer, John Neuhaus, John C. van Swieten, Harro Seelaar, Elise G. P. Dopper, Chiadi U. Onyike, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain*, 134(9): 2456–2477, 2011. doi: 10.1093/brain/awr179.

Mher Safaryan, Alexandra Peste, and Dan Alistarh. Knowledge distillation performs partial variance reduction. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-3288. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/ee1f0da706829d7f198eac0edaacc338-Abstract-Conference.html.

Ryan W. Schroeder, Phillip K. Martin, and Anne Walling. Neuropsychological evaluations in adults. *American Family Physician*, 99(2):101–108, 2019.

Hajin Shim, Sung Ju Hwang, and Eunho Yang. Joint active feature acquisition and classification with variable-size set encoding. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/e5841df2166dd424a57127423d276bbe-Paper.pdf.

Michael Valancius, Maxwell Lennon, and Junier Oliva. Acquisition conditioned oracle for nongreedy active feature acquisition. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 48957–48975. PMLR, 2024. URL https://proceedings.mlr.press/v235/valancius24a.html.

Siddarth Venkatraman, Matthieu Dinot, and Laurence Aitchison. Le critique: Privileged value functions for llm reinforcement learning, 2026. URL https://arxiv.org/abs/2608.16739.

Henrik von Kleist, Alireza Zamanian, Ilya Shpitser, and Narges Ahmidi. Evaluation of active feature acquisition methods for static feature settings. *arXiv preprint arXiv:2312.03619*, 2023. doi: 10.48550/arXiv.2312.03619. URL https://arxiv.org/abs/2312.03619v2. Version 2.

Henrik von Kleist, Alireza Zamanian, Ilya Shpitser, and Narges Ahmidi. Evaluation of active feature acquisition methods for time-varying feature settings. *Journal of Machine Learning Research*, 26 (60):1–84, 2025. URL https://www.jmlr.org/papers/v26/23-1635.html.

Qi Wang, Liang Zhan, Paul Thompson, and Jiayu Zhou. Multimodal learning with incomplete modalities by knowledge distillation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1828–1838. ACM, 2020. doi: 10.1145/3394486.3403234. URL https://doi.org/10.1145/3394486.3403234.

Sandra Weintraub, Lilah Besser, Hiroko H. Dodge, Merilee Teylan, Steven Ferris, Felicia C. Goldstein, Bruno Giordani, Joel Kramer, David Loewenstein, Daniel Marson, Dan Mungas, David Salmon, Kathleen Welsh-Bohmer, Xiao-Hua Zhou, Steven D. Shirk, Alireza Atri, Walter A. Kukull, Creighton Phelps, and John C. Morris. Version 3 of the Alzheimer Disease Centers' neuropsychological test battery in the Uniform Data Set (UDS). *Alzheimer Disease and Associated Disorders*, 32(1):10–17, 2018.

Chonghua Xue, Sahana S. Kowshik, Diala Lteif, Shreyas Puducheri, Varuna H. Jasodanand, Olivia T. Zhou, et al. AI-based differential diagnosis of dementia etiologies on multimodal data. *Nature Medicine*, 30(10):2977–2989, 2024. doi: 10.1038/s41591-024-03118-z.

Shuo Yang, Sujay Sanghavi, Holakou Rahmanian, Jan Bakus, and S. V. N. Vishwanathan. Toward understanding privileged features distillation in learning-to-rank. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26658–26670. Curran Associates, Inc., 2022. doi: 10.52202/068431-1933. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/aa31dc84098add7dd2ffdd20646f2043-Abstract-Conference.html.

Jinsung Yoon, James Jordon, and Mihaela van der Schaar. Asac: Active sensing using actor-critic models. In *Proceedings of the 4th Machine Learning for Healthcare Conference*, volume 106 of *Proceedings of Machine Learning Research*, pp. 451–473. PMLR, 2019. URL https://proceedings.mlr.press/v106/yoon19a.html.

A PATIENT REPRESENTATION AND THE DIAGNOSTIC MODEL

The ACUMEN diagnostic model connects three kinds of information: the tests currently observed, the patient's earlier assessments, and imaging or biomarker measurements that describe complementary aspects of the patient. It combines these sources through patient-state encoding, modality-specific representations, temporal feature transfer, and diagnosis and recommendation outputs. This appendix explains those interfaces, which are the ones evaluated in Section 4. We write r^{cur} and r^{temp} for its learned temporal features, reserving z_T^{cur} and z_T^{δ} for the observed-pair descriptors of the gain supervision in Appendix C. Those descriptors are an optional input to the gain teacher and do not redefine the diagnostic model.

A.1 WHAT THE CLINICAL STATE RECORDS

A patient state must distinguish a low test score from a test that has not been observed. It must also distinguish a measurement obtained at the current visit from an older result that provides context. Table 3 gives the diagnostic model's 194-dimensional state for 20 scalar cognitive-test channels. The eight demographic channels are age, sex, education, four race indicators, and a Hispanic indicator. For each test, the history block stores its last observed value, historical mean, fitted trend, and observation count. The temporal block adds time since its last administration and a trend that emphasizes recent observations. Together these fields allow two patients with the same latest score to have different histories and different amounts of supporting evidence.

Table 3: Clinical state of the diagnostic model. Its diagnosis pathway selects 88 channels: demographics, current scores, current masks, and per-test timing. Gain supervision excludes diagnosis-derived blocks from both pathways and rebuilds the state for its specified action library.

State block	Width	Information represented
Demographics	8	Age, sex, education, and demographic indicators.
Current scores	20	One scalar channel per instrument in the 20-test battery.
Current observation masks	20	Which score channels are observed at the visit.
Test history	80	Last value, mean, trend, and count for each test.
Per-test timing	40	Measurement age and time-weighted trend for each test.
Visit timing	6	Visit position, follow-up duration, and spacing of visits.
Diagnosis-code block	7	Diagnosis coding used by the recommendation pathway.
Diagnosis-history block	7	Stability and transition summaries for recommendation.
Battery summaries	6	Coverage and counts across tests and visits.
Total	194	Fixed 20-channel configuration.

The current acquisition state is a function of h_{vk}. At the start of an episode it contains permitted baseline information, strictly prior measurement summaries, and any modalities already available. For each current score field f, a reveal indicator records whether an acquired action has returned that field. Unrevealed values use fixed padding together with this indicator. Standardization uses fitting-participant statistics, so a zero standardized value can remain a legitimate observed measurement. An observation mask resolves that ambiguity.

Acquisition reveals groups of fields, not necessarily one scalar per action. Consequently, the version-specific libraries in Appendix F do not reuse the number 194 as their input dimension. Even retaining the 20 scalar channels while removing its 14 diagnosis-derived channels changes that count to 180. The current field specification determines the actual dimension and prevents a historical diagnosis code from re-entering through a copied feature slice.

A.2 ENCODING COMPLEMENTARY MODALITIES

The richer representation begins with measured regional features rather than a single undifferentiated imaging vector. Structural MRI contributes regional volumes, PET contributes regional uptake summaries, and CSF contributes analyte values with assay information. Their encoders compress different measurement spaces to compatible tokens, as summarized in Table 4. Assay indicators distinguish measurement conventions rather than acting as additional biological analytes.

Table 4: Interfaces of the diagnostic model's multimodal representation. The temporal output is a learned 128-dimensional representation. It differs from concatenating a 128-dimensional observation encoder with its 128-dimensional finite difference.

Modality	Measurement input	Representation interface
MRI	89 intracranial-volume-normalized regional measurements	Mamba regional encoder with 64 output features.
PET	19 regional standardized uptake value ratios	Regional encoder, 64-dimensional bottleneck.
CSF	7 biomarker values and 4 assay indicators	Assay-aware MLP, 32-dimensional bottleneck.
Fusion	Projected modality tokens with identity and missingness information	128-dimensional fused representation.
Temporal	Time-indexed feature observations and their intervals	64 current-state and 64 learned temporal coordinates.

Each modality bottleneck is projected to a common token width. A modality-type embedding identifies its source, while a learned missing token occupies an absent modality's slot. The four-token multimodal design processes clinical, MRI, PET, and CSF tokens with two Mamba blocks of width 128 and state dimension 16 (Gu & Dao, 2023). A separate fusion configuration first combines the three imaging and biomarker tokens, then merges the resulting representation with clinical features through a learned gate. The four-token fusion and the later gated fusion are alternative arrangements of these interfaces. They should not be interpreted as consecutive mandatory stages.

Missing tokens let a fixed interface accept different input subsets. They do not supply the patient's missing measurement. Modality dropout trains the model with selected observed modalities hidden: both the value path and its availability indicator must change. A partial modality also needs coordinate-level masks, since a missing regional value differs from the absence of the entire modality. The modality-withholding results in Section 4 measure the diagnostic model's behavior under such input restrictions. The asymptotic sequence-length advantage of Mamba does not by itself imply a material speed advantage for three or four tokens.

A.3 Current state, learned temporal structure, and observation time

The temporal interface represents a patient's current condition and temporal context in separate 64-dimensional outputs. It uses time-aware encoding and a mixture of four displacement experts. Writing u for a feature vector and Δt for a supplied interval, the displacement module has the form

$$d_\psi(u, \Delta t) = \sum_{j=1}^{4} \pi_j(u)\, d_{\psi_j}(u, \Delta t), \qquad \pi_j(u) \geq 0, \quad \sum_{j=1}^{4} \pi_j(u) = 1. \tag{13}$$

The weights adapt the combination of learned displacement functions to the input representation. Their numerical indices do not identify clinical disease stages. With repeated observations, temporal processing can combine observed endpoint differences, elapsed-time-normalized differences, and the learned displacement into r^{temp}. The current-state projection supplies r^{cur}, giving the interface $[r^{\mathrm{cur}}, r^{\mathrm{temp}}] \in \mathbb{R}^{128}$.

The teacher's training combines within-modality contrastive learning, cross-modal alignment, longitudinal consistency, and temporal displacement regularization. Cross-modal alignment uses measurements belonging to the same patient and eligible observation window, while displacement training uses ordered observations and their elapsed times. The two roles are complementary: similarity across modality views encourages shared content, whereas a displacement target models variation across time. Optional coarse-label fine-tuning does not require the teacher to share the student's complete diagnostic label space.

A single observation supports a current-state encoding. When no longitudinal pair is present, a learned missing-temporal token can represent that absence. The displacement model can alternatively predict a future feature at a chosen horizon, but that quantity is a conditional forecast. It is not observed individual change, and mixture weights alone do not establish calibrated predictive uncertainty. Both

cases remain distinct from the observed-pair z_T^{δ} in Appendix C, whose validity requires two eligible observations.

The distinction also matters when aligning modalities. Every observation retains its acquisition time, modality identity, and measurement age relative to the decision cutoff. A recent MRI and an older CSF panel must not become apparently simultaneous because their values occupy the same fused input. Past-only alignment windows determine which measurements enter an observation. Pair construction resolves valid timestamps before padding, uses comparable measurements at both endpoints, and records the time gap. A missing pair and a valid zero difference consequently have different indicators.

A.4 Feature transfer and the prediction pathways

The clinical architecture uses separate recommendation and diagnosis encoders. The recommendation encoder has hidden widths $512, 512, 256, 128$, followed by a 128-unit head and 20 recommendation outputs. The restricted diagnosis encoder has widths $256, 128, 64$ and a $64 \to 128 \to 6$ classification head. Residual connections, normalization, GELU activations, and dropout support these networks. Its auxiliary impairment and severity heads supervise coarse diagnostic structure. A separate confidence output and the entropy of the class distribution are different quantities.

Feature transfer connects the richer representation to the clinical learner without requiring identical teacher and student class logits (Wang et al., 2020). The temporal transfer objective combines current and temporal feature matching with coherence and variance regularization:

$$\mathcal{L}_{\text{transfer}}^{\text{prior}} = 0.20\,\mathcal{L}_{\text{cur}} + 0.30\,\mathcal{L}_{\text{temp}} + 0.10\,\mathcal{L}_{\text{coh}} + 0.05\,\mathcal{L}_{\text{var}}. \tag{14}$$

The first two terms match projected student features to the corresponding teacher segments using masked squared errors. The coherence term uses cosine similarity, while the variance term regularizes the learned temporal representation. These regularizers concern representation fitting. They do not make a forecast an observation or require every patient to exhibit nonzero measured change. Appendix C defines the simpler current/observed-change feature objective used in the gain-supervision specification separately.

The multimodal diagnosis design can combine clinical features with current and temporal tokens through cross-attention. For clinical queries Q and projected feature keys and values K, V, this interface computes

$$\text{Attn}(Q, K, V) = \text{softmax}\Big(QK^{\mathsf{T}}/\sqrt{d_k}\Big)\,V, \tag{15}$$

where d_k is the key width. Separate token identities preserve the distinction between current and temporal features. The gated clinical-fusion variant instead blends projected clinical and multimodal features before a shared student encoder. That variant is not the isolated clinical two-encoder configuration.

The input regime determines what either fusion operation may consume. A multimodal prediction model can use measured modalities available at the decision time. A clinical-only distilled model must produce its output from clinical inputs, using transferred parameters or predicted features rather than a patient-specific hidden imaging cache. This distinction separates the benefit of direct modality access from the benefit of training with richer supervision.

A.5 From a patient state to the next acquisition

The recommendation objective learns test-administration patterns (Appendix B.5), while the gain head predicts loss reduction for a fixed diagnosis model. This changes the meaning of its score. A sigmoid recommendation score is not automatically a signed test gain, and a temporal adjustment to an imitation logit is not automatically in Brier-loss units. The gain-training objective in Section 3 therefore defines the extended acquisition output directly.

Within an episode, strictly prior history remains fixed. After an action returns its designated fields, the system changes their reveal indicators, updates current battery coverage, and recomputes all dependent representations and outputs. The mask describes what the decision maker has observed, rather than which unrequested measurements happen to exist in a retrospective record. During training,

richer teacher targets may remain available when student inputs are hidden. During deployment, teacher-target availability does not become an additional student input.

In the fixed-predictor configuration, diagnosis and acquisition parameters are separate, and acquisition fitting cannot modify the diagnosis model defining the gain targets. Feature heads retain a clinical-context route so that matching a teacher segment is not the student's only source of information. Cross-attention must respect the same input restrictions as a direct input: a forbidden result cannot enter through a key, value, temporal summary, or another pathway. These interfaces connect the diagnostic model to the gain objective while preserving the meaning of the patient's state at every reveal.

B ADDITIONAL DIAGNOSTIC RESULTS

This appendix gives the numeric results behind Section 4 and Figure 3. Unless a subsection states otherwise, every entry uses the six-category NACC final-diagnosis task, the participant-level partitions of Section 4, and the ACUMEN diagnostic model with its learned temporal encoder and feature transfer (Appendix A). The comparisons distinguish withholding inputs from one fitted model, changing a component during training, and measuring recall within a recorded category. All entries are aggregate point estimates. The aggregate summaries do not record per-condition sample sizes or how repeated assessments are weighted, so we report them as descriptive comparisons without confidence intervals or significance tests.

B.1 MODALITY DROPOUT AND INPUT AVAILABILITY

Table 5 trains the model with and without modality dropout and evaluates each trained model with full and with clinical-only input. Dropout trades 0.6 points under full input for 4.9 points with clinical inputs alone.

Table 5: Effect of modality dropout during training. Each entry is six-category accuracy. Full input includes clinical information, MRI, PET, and CSF. Each row uses one trained model in both input conditions, and subscripts give the change from training without dropout. The gap is full-input minus clinical-only accuracy, not the range across all modality subsets.

Training condition	Full input (%)	Clinical only (%)	Gap (points)
No modality dropout	93.8	85.2	8.6
Modality dropout, $p = 0.3$	$93.2_{\downarrow 0.6}$	$\mathbf{90.1}_{\uparrow 4.9}$	$\mathbf{3.1}_{\downarrow 5.5}$

Table 6 evaluates one trained model under all eight combinations of MRI, PET, and CSF availability, and Figure 3(a) plots the same values. Clinical inputs remain present throughout, so the MRI row means clinical information plus MRI. It is distinct from the separately trained MRI-only baseline in Table 2. No model is retrained for any row.

Table 6: Six-category accuracy of one trained ACUMEN model under each subset of additional modalities. Clinical inputs and model parameters are fixed. Subscripts give the change from the full-input accuracy of 93.2%, and the tinted rows match the full-input and clinical-only markers of Figure 3(a). These input conditions differ from separately trained single-modality models.

Additional modalities available	MRI	PET	CSF	Accuracy (%)
MRI + PET + CSF	✓	✓	✓	**93.2**
MRI + PET	✓	✓		$92.1_{\downarrow 1.1}$
MRI + CSF	✓		✓	$91.4_{\downarrow 1.8}$
PET + CSF		✓	✓	$90.8_{\downarrow 2.4}$
MRI	✓			$90.2_{\downarrow 3.0}$
PET		✓		$89.1_{\downarrow 4.1}$
CSF			✓	$88.4_{\downarrow 4.8}$
None, clinical only				$90.1_{\downarrow 3.1}$

The two-modality conditions reach 90.8% to 92.1%, whereas the one-modality conditions reach 88.4% to 90.2%. Clinical-only prediction reaches 90.1%, exceeding the PET and CSF conditions by 1.0 and 1.7 points, respectively. The effect of a modality therefore depends on what else is present. Adding MRI raises accuracy in every condition, by 0.1 points when it is the only addition and by 2.4 to 3.0 points when PET or CSF is also present. Adding PET or CSF alone lowers accuracy relative to clinical inputs, whereas adding either one to a condition that already contains another modality raises accuracy by 1.1 to 2.4 points. The overall range is 4.8 points. The 3.1-point quantity in Table 5 refers specifically to full input versus clinical-only prediction. Controlled withholding characterizes this model's response to absent inputs. It does not establish identical performance in patient groups whose modalities are naturally missing.

B.2 Full-model ablations

Table 7 reports modifications relative to the full model. Pathway separation distinguishes its test-recommendation and diagnosis routes. Temporal modulation here denotes per-test staleness and trend offsets on recommendation scores, not the learned multimodal temporal descriptor. The TSD condition denotes the feature-transfer recipe of Appendix A, which combines feature matching, coherence, and variance penalties. These component definitions belong to the evaluated diagnostic model, rather than to the feature loss used in the gain supervision of Appendix C.

Table 7: Ablations of the ACUMEN diagnostic model, each removing or replacing one component of the full model. Subscripts give the change from the full model. AUROC is micro-averaged one-vs-rest. The label-space alignment of the standard-KD replacement is unspecified, so that row supports only a comparison of the listed configurations.

Model configuration	**Accuracy (%)**	**Macro F_1**	**AUROC**
Full model	93.2	0.84	0.97
Without dual-pathway isolation	$91.8_{\downarrow 1.4}$	$0.78_{\downarrow 0.06}$	$0.95_{\downarrow 0.02}$
Standard KD in place of TSD	$90.6_{\downarrow 2.6}$	$0.77_{\downarrow 0.07}$	$0.94_{\downarrow 0.03}$
Without entropy stopping	$93.2_{\pm 0.0}$	$0.84_{\pm 0.00}$	$0.97_{\pm 0.00}$
Without staleness/trend modulation	$92.5_{\downarrow 0.7}$	$0.82_{\downarrow 0.02}$	$0.96_{\downarrow 0.01}$

The configuration without pathway isolation has accuracy 1.4 points lower, macro F_1 0.06 lower, and AUROC 0.02 lower than the full model. The configuration without staleness/trend modulation differs by 0.7 points, 0.02, and 0.01, respectively. These configuration-level comparisons do not establish a direct effect of recommendation-score offsets on the diagnosis pathway or isolate the value of the learned temporal representation. The standard-KD replacement is 2.6 accuracy points below the full model, but its incompletely specified label alignment precludes a general conclusion that feature transfer outperforms every response-distillation alternative.

The entropy-stopping modification leaves all three classification metrics unchanged. These aggregate scores do not establish equality of individual predictions, acquisition counts, or trajectories. In particular, that row supplies no estimate of the cost of obtaining the inputs used for classification.

B.3 Modality-specific teacher-feature configurations

A separate comparison adds teacher features from one modality to the clinical baseline. Table 8 reports these three configurations, each relative to the same baseline. They are alternative single-modality additions, not successive steps in a cumulative construction.

MRI teacher features give the largest improvement in this comparison, followed by PET and CSF. The ordering appears in accuracy and macro F_1, while PET and CSF share the displayed AUROC of 0.92. The same ordering appears when modalities are withheld from the full model, where MRI is the only modality whose addition never lowers accuracy. These results motivate retaining a strong single-modality comparator when assessing fusion. They do not isolate the contribution of each modality inside the fused model, whose inputs and training interactions differ.

Table 8: Teacher features from a single modality added to the clinical baseline. Subscripts give the change from that baseline. These training configurations differ from withholding modalities from the full model in Table 6. AUROC is micro-averaged one-vs-rest.

Teacher-feature configuration	Accuracy (%)	Macro F_1	AUROC
Clinical baseline	86.2	0.71	0.91
With MRI teacher features	**89.1**$_{\uparrow 2.9}$	**0.76**$_{\uparrow 0.05}$	**0.93**$_{\uparrow 0.02}$
With PET teacher features	88.4$_{\uparrow 2.2}$	0.74$_{\uparrow 0.03}$	0.92$_{\uparrow 0.01}$
With CSF teacher features	87.8$_{\uparrow 1.6}$	0.73$_{\uparrow 0.02}$	0.92$_{\uparrow 0.01}$

B.4 Recall within diagnostic categories

Table 9 reports recall within each of the six categories for the clinical-only MLP and the full multimodal model, and Figure 3(b) plots the same values. The comparison is between the clinical baseline and the full model, rather than between the clinical baseline and the model's clinical-only condition.

Table 9: Recall within each recorded category for the clinical-only MLP and the full multimodal model. Subscripts give the gain in recall, in percentage points. The rows summarize each category separately and do not form a confusion matrix.

Recorded category	Clinical-only MLP (%)	Full ACUMEN model (%)
Normal	91.2	96.4$_{\uparrow 5.2}$
MCI	72.4	87.3$_{\uparrow 14.9}$
Alzheimer's disease	89.8	95.1$_{\uparrow 5.3}$
Lewy body dementia	68.5	86.2$_{\uparrow 17.7}$
Frontotemporal dementia	81.2	91.8$_{\uparrow 10.6}$
Other Dementia	54.3	73.5$_{\uparrow 19.2}$

The gains are largest where the clinical baseline is weakest: Other Dementia improves by 19.2 points, Lewy body dementia by 17.7, and MCI by 14.9. Normal and Alzheimer's disease, whose baseline recall already exceeds 89%, improve by 5.2 and 5.3 points. Other Dementia keeps the lowest recall, at 73.5%, so improvement and residual difficulty coexist. These category summaries complement macro F_1 by showing how performance varies within the recorded target space. They do not identify which incorrect category received a missed case, and cannot attribute a particular correction to MRI, PET, or CSF without additional stratified comparisons.

B.5 Recommendation pathway trained on test administration

The diagnostic model's recommendation pathway (Appendix A) is trained to predict which of the 20 battery instruments are administered at a participant's next visit. It is evaluated on 15,268 held-out visit transitions from its own participant-level 70/20/10 split, stratified by each participant's most frequent diagnosis. Table 10 reports the ranking metrics and how often each instrument appears among the three recommendations.

The three recommended tests are administered 95.7% of the time, and every held-out list contains at least one administered test. These scores are high largely because CDR Global and CDR Sum of Boxes appear in every list and are administered at 15,263 of the 15,268 transitions. Across all held-out lists, a memory test is recommended 67 times, a visuospatial test never, and neither the MMSE nor the MoCA is recommended. Agreement with administration therefore measures how well the pathway reproduces routine practice, not which test would most improve the diagnosis. The gain target of Section 3 is designed to measure the latter. CDR summaries also enter the reference diagnosis, which is one reason the acquisition library in Appendix F excludes them.

Table 10: Recommendation pathway trained to predict next-visit test administration, on 15,268 held-out transitions. Left: ranking metrics against the tests actually administered. Right: number of three-item recommendation lists containing each instrument. The counts sum to $3 \times 15{,}268$. The eight instruments never recommended are the MMSE, the MoCA, both Logical Memory scores, Craft Story delayed recall, the MINT, and both Benson figure tests.

Ranking metric	Value	Instrument in top three	Lists
Precision@3	0.957	CDR Global	15,268
Recall@3	0.358	CDR Sum of Boxes	15,268
NDCG@3	0.997	Trail Making A	4,289
Hit rate@3	1.000	Animal fluency	3,520
		Digit Span, forward	2,275
		Letter F fluency	1,769
		Digit Span, backward	1,606
		Trail Making B	1,504
		Vegetable fluency	231
		Craft Story, immediate recall	67
		Letter L fluency, Boston Naming	5, 2
		Remaining eight instruments	0

C SPECIFICATION OF GAIN SUPERVISION

Gain supervision (Sections 3.3 and 3.4) uses a multimodal feature teacher to define auxiliary representations and a separate value teacher to predict gains of the deployable diagnosis model. This appendix specifies both interfaces, their fitting order, and the acquisition procedure. The reconstruction and observed-pair configuration below makes the feature interface concrete. Appendix A describes the diagnostic model's patient state, modality encoders, and learned temporal representation. Section 4 evaluates that diagnostic model with its own encoder and training losses. It does not evaluate the value-training stage specified here.

Table 11: Principal notation. Dataset partitions use D with a subscript, and reconstruction decoders use D_m. Those objects differ from the random gain D in the analysis. Additional local symbols are defined at use.

Symbol	Meaning
h, Y, C	Permitted clinical history, recorded label, and number of classes.
q_θ, g_ϕ	Restricted diagnostic predictor and deployed test-gain model.
F, z_T^{cur}, z_T^{δ}	Frozen multimodal encoder and its current/change targets.
Z, I	Training-only descriptor and its availability indicator.
d, D	Observed Brier gain and the corresponding random gain in the proofs.
$m^\star$, $\widehat{m}^{-f}$	Ideal conditional gain teacher and fitted teacher excluding fold f.
t, ρ	Mixed supervision target and teacher mixing weight.
c_a, λ	Test cost and its conversion to predictive-loss units.
$\mathcal{L}(h)$	Legal unacquired tests under prerequisites and the specified budget.
$P_{\mathcal{A}}$, $\widehat{P}_{\mathcal{A}}$	Selected population law and its empirical participant-balanced measure.
ε, η	Assumed uniform student approximation error and conditional teacher-bias bound.

C.1 EPISODE INFORMATION AND THE STUDY POPULATION

At visit v, the ordered acquisition history is

$$h_{vk} = \big(B_v, V_v, (a_j, X_{va_j})_{j \le k}\big). \tag{16}$$

The permitted prior information B_v is recorded strictly before the visit, and V_v contains modalities available when the episode begins. Action a reveals its designated group of current measurements

X_{va}. Diagnosis codes, the target Y_v, target-derived features, future visits, and unrevealed current scores are excluded from both student pathways. A prior measurement can be retained as prior information, but cannot replace that action's current result. The same restrictions apply to derived summaries and cached features.

Write $s_{vk} = \mathrm{Build}(h_{vk})$ for the clinical input vector. The map uses the returned-field specification, rather than treating each action as a scalar. It concatenates permitted baseline covariates, current field values and reveal masks, prior-history summaries with validity masks, and current coverage and timing metadata. For a field f, let $(t_{fj}, x_{fj})_{j=1}^{n_f}$ be its valid permitted observations strictly before the visit. Its count is n_f, and, when $n_f > 0$, its historical mean and latest-measurement age are

$$\bar{x}_f = \frac{1}{n_f} \sum_{j=1}^{n_f} x_{fj}, \qquad a_f = t_v - \max_j t_{fj}. \tag{17}$$

The latest observed value is stored separately from this mean. For at least two distinct observation times, define an ordinary fitted trend by

$$b_f = \frac{\sum_j (t_{fj} - \bar{t}_f)(x_{fj} - \bar{x}_f)}{\sum_j (t_{fj} - \bar{t}_f)^2}, \qquad \bar{t}_f = \frac{1}{n_f} \sum_j t_{fj}. \tag{18}$$

The denominator must be positive. A recency-weighted trend uses the same weighted least-squares construction with prespecified weights $w_{fj} = \exp(-(t_v - t_{fj})/\kappa)$, weighted means, and a time scale $\kappa > 0$ fixed using development data. Missing means, ages, or trends use fixed padding and an explicit validity indicator. In particular, a valid zero trend differs from a trend that cannot be estimated.

The new within-visit configuration recomputes current-prefix quantities after every reveal, while keeping these strictly prior summaries fixed. Any feature comparing a current score with its historical value becomes valid only after that current score is revealed. The history construction uses a common time unit and fitting-participant standardization. If instruments change between visits, incompatible scores remain separate fields rather than being interpreted as one trajectory. These rules define the acquisition state's handling of partial histories, without assuming the diagnostic model's fixed 194-channel feature layout.

The library $\mathcal{A}$ fixes jointly returned scores, protocol eligibility, prerequisites, nonnegative costs c_a, and a deterministic tie order. The legal set $\mathcal{L}(h)$ contains unacquired actions whose prerequisites are met. It also enforces the remaining hard budget, as specified below. Legality depends on the protocol and observed history, not on retrospective storage. Unit costs $c_a = 1$ count requested action groups. They do not measure test administration time. Costs incurred before the episode are excluded from its incremental acquisition cost.

Let $E_{\mathcal{A}} = 1$ mean that every action in a structurally eligible library has a usable recorded result at the visit. Define $P_{\mathcal{A}}$ as a population sampling law: draw a participant from the eligible participant population conditional on at least one complete-library visit, then sample one of that participant's eligible complete-library visits uniformly. Assume the conditioning event has positive probability. For a finite partition with $N_{\mathcal{A}}$ such participants and $n_p^{\mathcal{A}}$ complete visits for participant p, its empirical counterpart is

$$\widehat{P}_{\mathcal{A}}(p, v) = \frac{1}{N_{\mathcal{A}} n_p^{\mathcal{A}}} \quad \text{for eligible complete visits } v \text{ of } p. \tag{19}$$

The population conditional expectations in the analysis use $P_{\mathcal{A}}$, not this finite empirical measure. The empirical measure balances participants within the selected cohort. It is not generally obtained by conditioning an arbitrary visit-weighted distribution on $E_{\mathcal{A}} = 1$. All acquisition comparisons use this same law within each partition. The selection indicator and underlying storage masks remain evaluator information. Values learned on this population do not automatically transport to incomplete-library episodes or a broader clinical population.

C.2 A FROZEN MULTIMODAL FEATURE TEACHER

Let t_v be the episode's information cutoff. The richer training view contains eligible imaging and biomarker observations acquired by t_v, even when those modalities are hidden from the student. It contains no diagnostic targets or unrevealed cognitive-test scores. Modality-specific encoders

map standardized measurements into tokens with identity and observation masks. A fusion module maps these tokens to a 128-dimensional feature vector, and a shared learned linear projection maps that vector to 64 dimensions. Write F for the complete map, including the modality encoders, fusion, and projection. Current and observed-change targets therefore each have 64 coordinates. The same projection is used for both pair endpoints. Mamba is one possible fusion module, whose comparison with simpler fusion must hold inputs and feature dimension fixed (Gu & Dao, 2023). This observed-pair construction is distinct from the diagnostic model's separate learned current and temporal outputs.

A concrete label-free feature objective reconstructs masked measured coordinates. Let x_{jm} contain standardized coordinates of modality m for richer observation j, let O_{jm} mark valid measurements, and let C_{jm} select coordinates to hide. Draw coordinate or whole-modality masks only within observed data. Omit observations with fewer than two valid coordinates from reconstruction fitting. Require at least one retained coordinate and one hidden target, redrawing otherwise. Unobserved entries use fixed padding and are masked out before loss arithmetic. The corrupted view $\widetilde{W}_j$ contains retained values, explicit masks, and missing tokens for entirely hidden modalities. With modality-specific decoders D_m, fit

$$\mathcal{L}_T = \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}_C \left[\frac{\sum_m \left\| O_{jm} \odot C_{jm} \odot \left(D_m(F(\widetilde{W}_j)) - x_{jm} \right) \right\|_2^2}{\sum_m \| O_{jm} \odot C_{jm} \|_1} \right]. \tag{20}$$

The denominator is the positive number of hidden valid targets. Missing measurements are never reconstruction targets. Fit preprocessing, F, and the decoders on D_{fit}. Then freeze preprocessing and F and discard the decoders for transfer. Reconstruction establishes a feature training objective, not biological disentanglement or acquisition value.

Construct richer observations at measurement-time anchors $\tau \le t_v$. For each modality, retain its most recent eligible measurement in a prespecified past-only alignment window $[\tau - w, \tau]$, where $w \ge 0$ is fixed before evaluation. Measurements outside that window are absent from the observation $W_{v,\tau}$. Record each retained measurement's actual timestamp. The primary rule selects the latest eligible nonempty anchor and uses its frozen encoding as z_T^{cur}. For change supervision, pair this observation at τ_1 with an eligible earlier observation at $\tau_0 < \tau_1 \le t_v$, using a fixed selection rule and a nonempty common modality subset M^{pair}. Encode both observations with exactly this subset and the same preprocessing:

$$z_T^{\delta} = \frac{F(W_{v,\tau_1}^{M^{\text{pair}}}) - F(W_{v,\tau_0}^{M^{\text{pair}}})}{\Delta\tau}, \qquad \Delta\tau = \tau_1 - \tau_0 > 0. \tag{21}$$

Each paired modality must have two distinct measurements in chronological order. Reusing one measurement at both anchors is not a measured pair. Choose the most recent earlier anchor satisfying these conditions and the prespecified positive minimum interval. Use a common time unit. A fixed pair-selection rule and compatible measurement definitions prevent choosing pairs retrospectively for favorable changes. Matching modalities removes one source of artificial displacement, but does not eliminate scanner effects, assay differences, or measurement noise. The target is a learned feature change per unit time, not an identified biological progression rate.

Validity indicators m^{cur} and m^{δ} distinguish the two supervision sources. A single eligible observation supplies no measured individual change target. A valid zero displacement remains valid. Resolve observation indices before padding, and do not construct synthetic individual trajectories to replace missing pairs. Record current-observation and pair endpoint ages, including per-modality measurement ages, relative to t_v. The anchor interval in the difference is not necessarily each modality's measurement interval. These metadata expose that alignment approximation instead of presenting old measurements as contemporaneous observations.

When the current encoding uses the latest member of the same masked pair, current state and change with known $\Delta\tau$ are invertible coordinates of the endpoint encodings. They add no information to those endpoints. A matched comparison using additional current modalities gives their encoding to both arms as a shared covariate and retains the pair-matched current endpoint in the current/change arm. A richer encoding alone need not recover the pair-matched encoding. Equation (80) specifies the invertible comparison. Any advantage of those coordinates is therefore an empirical question about representation and fitting under matched information.

C.3 STUDENT REPRESENTATIONS AND LEGAL PREFIX VIEWS

The diagnosis encoder g_y and acquisition encoder g_u have separate parameters. For $r \in \{y, u\}$, encoder g_r produces segments $z_r^{\mathrm{cur}}(h)$ and $z_r^{\delta}(h)$ and a clinical-context channel $c_r(h)$. The context channel lies outside the feature-matching projections:

$$q_\theta(h) = \mathrm{softmax}\, f_y(z_y^{\mathrm{cur}}(h), z_y^{\delta}(h), c_y(h)), \tag{22}$$

$$g_\phi(h, a) = f_u(z_u^{\mathrm{cur}}(h), z_u^{\delta}(h), c_u(h), a). \tag{23}$$

The parameters ϕ include g_u and the gain head. The context channels permit retention of useful history that is absent from conditional means of the teacher features. No acquisition representation enters q_θ in this instantiation, so value fitting cannot change the predictor defining the gain targets.

Prefix generation is independent of the learned policy. Fix a maximum training depth $K_{\max} \geq 1$ and training cost budget $C_{\mathrm{train}} = \sum_{a \in \mathcal{A}} c_a$ for the fixed finite library. The sampler uses $\mathcal{L}(h)$ with this budget, which permits every protocol-legal sequence. Deployment may impose a smaller cost budget. For value fitting, draw a requested depth uniformly from $\{0, \ldots, K_{\max} - 1\}$. Starting from the empty acquisition history, repeatedly choose uniformly among currently legal actions, reveal its result, and continue until the requested depth or an empty legal set. Choices never inspect an unrevealed value or label. On complete-library episodes, every legal next action at every reachable prefix before the cap has positive sampling probability. Prerequisites mean that this procedure is not generally uniform over complete permutations.

Diagnosis fitting on D_{fit} can also use naturally incomplete records. It samples requested depths from $\{0, \ldots, K_{\max}\}$, including the extra endpoint evaluated after a value-fitting prefix's final reveal. Its offline prefix generator reveals only legally reachable actions with usable recorded outcomes. The generator may exclude an unobserved action from replay, but must not present this storage restriction to the student as clinical unavailability. It ends if no recorded legal reveal remains and does not impute a target for a missing result. In contrast, D_{value} uses the complete-library law above, where every legal next-action target is evaluable. The diagnosis-fitting prefix law need not equal the value-fitting law, but must be held fixed across matched predictor comparisons.

A sampled prefix exposes only its revealed tests. Student modality dropout can additionally hide measured modalities while retaining eligible teacher targets. Hiding an input removes every dependent summary and cached feature. Student observation masks describe what the prefix reveals, not what the original record happens to store. Entirely hidden modalities receive missing tokens. Zeroing a value while leaving it marked observed is not masking.

For a training collection with nonnegative example weights ω_i, define ordinary feature regression for segment $b \in \{\mathrm{cur}, \delta\}$:

$$S_b = \sum_i \omega_i m_i^b, \tag{24}$$

$$\mathcal{L}_{r,b} = \frac{\sum_i \omega_i m_i^b \left\| P_{r,b} z_r^b(h_i) - \mathrm{sg}(z_{T,i}^b) \right\|_2^2}{d_b S_b} \quad \text{if } S_b > 0, \tag{25}$$

$$\mathcal{L}_{r,\mathrm{feat}} = \alpha_{\mathrm{cur}} \mathcal{L}_{r,\mathrm{cur}} + \alpha_\delta \mathcal{L}_{r,\delta}. \tag{26}$$

Set a term to zero when $S_b = 0$. Here $P_{r,b}$ is a trainable projection, d_b is the target dimension, $\alpha_b \geq 0$ weights the segment, and sg stops teacher gradients. Separate valid-target normalizers make the feature weights explicit. These terms are a masked block-weighted squared error on concatenated predictions, not a new loss family. Their projected outputs predict the named targets, whereas unrestricted projections do not identify the raw latent segments as biological state or change.

Fit the diagnosis pathway on D_{fit} by weighted categorical cross-entropy plus $\beta_y \mathcal{L}_{y,\mathrm{feat}}$, with $\beta_y \geq 0$. The acquisition encoder may be initialized by the same feature transfer and retain the auxiliary loss during value fitting. This conventional transfer supplies a representation comparator for the value-target extension. Teacher features are detached supervision and never patient-specific inputs to a clinical-only student at deployment.

C.4 A PRIVILEGED TEACHER FOR THE SAME PREDICTOR'S GAIN

Select and freeze q_θ, including its probability transformation, before constructing action labels. For prefix i and legal action a, use the normalized multiclass Brier gain

$$\ell(p, Y) = \tfrac{1}{2}\|p - e_Y\|_2^2 \in [0, 1], \tag{27}$$

$$d_{i,a} = \ell(q_\theta(h_i), Y_i) - \ell(q_\theta(h_i \oplus (a, X_{ia})), Y_i) \in [-1, 1]. \tag{28}$$

The reveal operator $\oplus$ updates every dependent input before the second prediction. Both predictions use the same restricted q_θ. The label and the candidate's recorded scores enter this offline calculation, not the student's pre-acquisition input. Direct regression to $d_{i,a}$ is the loss-reduction principle used by DIME (Gadgil et al., 2024).

The proposed extension adds a training-only value teacher. Set $I_i = \mathbf{1}\{m_i^{\mathrm{cur}} + m_i^{\delta} > 0\}$ and define its auxiliary view as

$$Z_i = \big(I_i, m_i^{\mathrm{cur}}, m_i^{\delta}, m_i^{\mathrm{cur}} z_{T,i}^{\mathrm{cur}}, m_i^{\delta} z_{T,i}^{\delta}, M_i^{\mathrm{cur}}, M_i^{\mathrm{pair}}, \Delta\tau_i, \mathrm{ages}_i\big). \tag{29}$$

Unavailable feature and timing fields use fixed zero encodings with their validity indicators. The ages identify the current observation and the pair endpoints. M_i^{cur} identifies modalities in the current encoding, and M_i^{pair} identifies the shared pair modalities. Each is empty when its corresponding target is invalid. These indicators distinguish missing change from valid zero change. Every underlying observation satisfies the episode cutoff.

A separate regressor receives (h_i, Z_i, a) and predicts $d_{i,a}$. It can concatenate an encoded prefix, available auxiliary fields, and an action embedding. It cannot receive Y_i, the unrevealed candidate's raw scores, post-reveal diagnosis outputs, or a target-bearing residual as inputs. The frozen embeddings in Z_i use eligible richer modalities only. Their association with diagnosis is the intended privilege, not access to the recorded diagnostic label. This value teacher estimates the gain of q_θ, not the gain of its own richer classifier or the return of a different policy. Cost is absent from the regression target.

Randomly partition participants in D_{value} into $F_{\mathrm{cv}} \geq 2$ folds independently of their measurements, labels, and gain targets. Let $f(i)$ identify the fold containing example i's participant. All visits, prefixes, and candidate-action targets of that participant stay in the same fold. Let $\mathcal{I}_+$ contain sampled value prefixes with $\mathcal{L}(h_i) \neq \varnothing$. Let w_i be their weights under $P_{\mathcal{A}}$ and the fixed prefix sampler, normalized once over $\mathcal{I}_+$. Equal weights apply to examples sampled directly from this law. Enumeration instead retains the participant, visit, and prefix probabilities. Each prefix's weight is distributed uniformly over its legal actions.

For each fold f, fit a value regressor on the other folds by minimizing

$$\frac{1}{S_{-f}} \sum_{i:f(i)\neq f} \frac{w_i I_i}{|\mathcal{L}(h_i)|} \sum_{a\in\mathcal{L}(h_i)} \big(m(h_i, Z_i, a) - d_{i,a}\big)^2, \qquad S_{-f} = \sum_{i:f(i)\neq f} w_i I_i. \tag{30}$$

Only $I_i = 1$ examples are needed to fit this teacher. Its objective keeps the same relative prefix and action weights as direct gain regression, rather than oversampling a preferred action or temporal subgroup. Require $S_{-f} > 0$ for every fitted teacher. If a randomly assigned fold fails this requirement, report the extension as unavailable on that split and use the raw-target baseline. Do not redraw folds based on target availability or outcomes. Clip fitted predictions to $[-1, 1]$ and denote the resulting excluding-fold model by $\widehat{m}^{-f}$. Clipping respects the known gain range and cannot increase squared error to an individual bounded gain target. Teacher configuration selection uses only its fitting folds and external validation participants. No excluded participant may influence that model's fitting or configuration selection through another trained component.

For a validation-selected global mixture coefficient $\rho \in [0, 1]$, form

$$t_{i,a} = d_{i,a} + \rho I_i \big(\widehat{m}^{-f(i)}(h_i, Z_i, a) - d_{i,a}\big). \tag{31}$$

This is a convex mixture in $[-1, 1]$ where privilege is available, and the original target exactly where $I_i = 0$. For those examples, define the unused teacher prediction as zero without querying a teacher. Fit the restricted acquisition model on all value examples:

$$\mathcal{L}_u = \sum_{i\in\mathcal{I}_+} \frac{w_i}{|\mathcal{L}(h_i)|} \sum_{a\in\mathcal{L}(h_i)} \big(g_\phi(h_i, a) - t_{i,a}\big)^2 + \beta_u \mathcal{L}_{u,\mathrm{feat}}. \tag{32}$$

The student receives neither Z_i nor I_i. Crucially, teacher-valid and teacher-missing groups are not separately reweighted to equal total mass. The main-text feature term is $\mathcal{L}_{\text{feat}} = \mathcal{L}_{u,\text{feat}}$, with weight $\beta = \beta_u \geq 0$. The per-target normalization in the auxiliary feature loss does not change these value-loss weights. Setting $\rho = 0$ recovers direct gain regression with the selected feature-transfer setting.

The ideal conditional teacher for this same gain preserves the student's conditional mean under the mixture, including when privilege availability is informative. A fitted teacher can introduce conditional bias, and cross-fitting does not remove approximation error. It also does not make the out-of-fold targets mutually independent, since fold models share training participants. Section 3.5 and Appendix D separate the ideal identity from these finite-teacher limitations. Neither identity establishes improved acquisition from current/change coordinates.

C.5 Fitting order and deployment

Table 12 summarizes the sequence described below.

Table 12: Proposed fitting and deployment sequence. Privileged feature and value teachers supervise the student during training and are absent from its deployment interface.

Stage	Required information flow
Feature fit	Fit F and its reconstruction decoders on fitting participants. Freeze F, form valid current and paired-change targets, and fit q_θ from permitted prefixes.
Value fit	Freeze the selected q_θ. Generate all legal gains on complete-library prefixes. Fit richer value teachers on participant folds, form excluding-fold mixtures, and fit g_ϕ with all examples' original weights.
Selection	Select settings on validation participants. A changed predictor requires regenerating the entire dependent value-training stage.
Deployment	Use only the student history, q_θ, g_ϕ, legal prerequisites, costs, and caps. Acquire the best positive feasible action or stop, then return the terminal prediction.

Partition participants into D_{fit}, D_{value}, D_{val}, and D_{test} before fitting. All visits of one participant remain together. Fit preprocessing, the feature teacher F, and q_θ using D_{fit}. Use D_{val} to select the predictor configuration and probability transformation, then freeze them. Construct gains, fit excluding-fold value teachers, and fit the restricted gain model using D_{value} with the selected population law.

Validation selects model sizes, feature weights, value-teacher settings, ρ, cost coefficient λ, and deployment caps. The prefix sampler is fixed across the matched target comparisons. Validation examples may be scored for selection but never added to fold training or student fitting. Keep D_{test} untouched until the complete procedure is fixed. If q_θ changes, rebuild $d_{i,a}$, refit all value teachers, regenerate $t_{i,a}$, and refit g_ϕ. A changed feature teacher or preprocessing also requires rebuilding its dependent targets and models. External teacher data require participant-overlap checks against every held-out partition.

At deployment, let $C(h) = \sum_{j \leq k} c_{a_j}$ be incurred cost, let $C_{\max} \leq C_{\text{train}}$ be the nonnegative hard cost budget, and let $K_{\text{dep}} \leq K_{\max}$ be a finite test-count cap. This depth restriction keeps each queried prefix within the training sampler's designed depth range. Let $\mathcal{L}_{\text{proto}}(h)$ contain unacquired actions meeting protocol eligibility and prerequisites before applying the cost cap. The legal set is

$$\mathcal{L}(h) = \{a \in \mathcal{L}_{\text{proto}}(h) : C(h) + c_a \leq C_{\max}\}. \tag{33}$$

Thus an action that would exceed the remaining budget cannot be selected. The test-count cap also bounds episodes with zero-cost actions. For feasible actions, compute $\widehat{u}(h,a) = g_\phi(h,a) - \lambda c_a$. STOP has value exactly zero. Stop at the test-count cap, an empty feasible set, or a largest net utility at most zero. Otherwise request the maximizing action, apply the fixed tie rule, append its measured result, and recompute both student outputs. Return the terminal distribution and stopping reason. No label, feature teacher, value teacher, or cached patient-specific privileged feature is accessed by this deployment path. Record mutually exclusive stopping reasons in this order: test-count cap, no remaining protocol-legal action, no action within the remaining cost budget, then nonpositive maximum utility. This reporting order does not change the selected actions or terminal prediction.

The retained overview depicts the broader framework, including cross-attention options, an entropy-based illustration, and clinician review. The precise instantiation here has no policy-to-diagnosis path and uses one-step net gain for primary stopping. An entropy stopping baseline or a clinician override must be evaluated as its own policy. Myopic positive-gain selection does not optimize every future test combination, and neither its stopping decision nor its hard budget certifies diagnostic correctness. The comparisons specified in Appendix F would test the proposed supervision interface under these restrictions.

D Analysis of privileged acquisition supervision

This appendix supplies the loss equivalence and analytical rationale in Section 3.5. The acquisition objective instantiates privileged-feature distillation (Yang et al., 2022). Related analyses study learning with privileged information (Karlsson et al., 2022), variance in distillation (Safaryan et al., 2023), and training-only information for value estimation (Ebi et al., 2026). The elementary arguments below identify the quantity transferred, the effect of finite teacher error, and diagnostics for test choice and stopping. They do not establish a new general theory or an empirical benefit of particular multimodal descriptors.

D.1 Fixed objects and the loss being transferred

Fix the deployable predictor q_θ, the complete-library study law $P_{\mathcal{A}}$, and the prefix and legal-action sampling procedure from Section 3. The study law includes its participant and visit weights. All expectations use the resulting joint law. Let H be a sampled permitted history, A a legal candidate, and $G = (H, A)$. The candidate sampler does not inspect unrevealed values or labels. Define the random gain

$$D = \ell(q_\theta(H), Y) - \ell(q_\theta(H \oplus (A, X_A)), Y), \qquad \mu(G) = \mathbb{E}[D \mid G]. \tag{34}$$

Its observed realization is $d_{i,a}$ in Equation (6). The target net utility is $\mu(h, a) - \lambda c_a$, as defined in Equation (9). The mixing weight ρ is distinct from the cost coefficient λ.

For normalized multiclass Brier loss, $\ell(p, Y) = \frac{1}{2}\|p - e_Y\|_2^2$, the probability-simplex constraints give $0 \le \ell \le 1$, so $D \in [-1, 1]$. In the binary case, writing p for the probability of label one gives $\ell(p, Y) = (p - Y)^2$. The Bayes risk of a fair label is therefore $1/4$. If the fixed predictor is the Bayes posterior under the same study law before and after a reveal, then

$$\mathbb{E}[D \mid H = h, A = a] = \tfrac{1}{2}\mathbb{E}\left[\|q_\theta(h \oplus (a, X_a)) - q_\theta(h)\|_2^2 \mid H = h, A = a\right]. \tag{35}$$

Indeed, the conditional Bayes risk at class-probability vector p is $\frac{1}{2}(1 - \|p\|_2^2)$, and the posterior after the reveal has conditional mean $q_\theta(h)$. Expanding the squared difference proves the identity. An imperfect fixed predictor need not satisfy it and can have negative expected gain. Brier improvement is also distinct from the conditional mutual-information interpretation of log-loss acquisition (Gadgil et al., 2024).

D.2 Mixed targets and weighted distillation losses

For a fixed example and candidate test, let d be the observed gain, $\widehat{m}$ the excluded-participant teacher prediction, and $\alpha = \rho I \in [0, 1]$. Put $t = (1 - \alpha)d + \alpha\widehat{m}$. For every student prediction g,

$$(1 - \alpha)(g - d)^2 + \alpha(g - \widehat{m})^2 = (g - t)^2 + \alpha(1 - \alpha)(d - \widehat{m})^2. \tag{36}$$

To verify the identity, expand the left side as $g^2 - 2gt + (1 - \alpha)d^2 + \alpha\widehat{m}^2$ and use $(1 - \alpha)d^2 + \alpha\widehat{m}^2 - t^2 = \alpha(1 - \alpha)(d - \widehat{m})^2$.

Hold q_θ, the teacher encoder, teacher predictions, mixture coefficients, input histories, and their weights fixed when optimizing the student. Then the last term is independent of every student parameter. The mixed-target objective and the weighted data/teacher loss consequently have identical student gradients and minimizers, even for a restricted student class. Summing with the same fixed nonnegative example/action weights and adding the same feature auxiliary loss preserve this conclusion. This is the squared-loss form of privileged-feature distillation (Yang et al., 2022), with regular input (h, a), privileged input Z, and response d_q.

At $I = 0$, both formulations reduce to the observed-gain loss. Participant cross-fitting and missing-privilege handling change how targets are constructed, but not this algebra. The equality alone does not give identical joint optimization if a target-producing network or α is optimized with the student, because the final term can then depend on an optimized parameter. Likewise, different data/teacher normalizers or different auxiliary penalties need not implement the same objective.

D.3 Ideal privileged targets with incomplete availability

Let $I \in \{0, 1\}$ indicate that the privileged descriptor is usable. The descriptor Z includes I, valid current-state and change features, their masks, and the relevant timing metadata. It need not contain every raw privileged measurement. Its observations obey the episode cutoff. A missing change target and a valid zero change have different masks. The deployed student observes H, not this additional descriptor.

For the analysis define

$$m^*(G, Z) = \mathbb{E}[D \mid G, Z], \qquad \xi = D - m^*(G, Z), \qquad \sigma^2(G, Z) = \mathrm{Var}(D \mid G, Z). \tag{37}$$

Then $\mathbb{E}[\xi \mid G, Z] = 0$. The conditional mean at $I = 0$ is a mathematical convenience and need not be fitted. For fixed $\rho \in [0, 1]$, the ideal version of Equation (7) is

$$T_\rho = D + \rho I(m^* - D) = m^* + (1 - \rho I)\xi. \tag{38}$$

It retains the raw gain wherever privileged information is unavailable.

Proposition 1 (Mean preservation and target variance)**.** *Under the fixed law above, for almost every G,*

$$\mathbb{E}[T_\rho \mid G] = \mu(G), \tag{39}$$

$$\mathrm{Var}(D \mid G) - \mathrm{Var}(T_\rho \mid G) = \rho(2 - \rho)R(G), \qquad R(G) = \mathbb{E}[I\sigma^2(G, Z) \mid G]. \tag{40}$$

The variance reduction is strict if $\rho > 0$ and $R(G) > 0$.

Proof. Because I is measurable with respect to Z, the tower property gives

$$\mathbb{E}[T_\rho \mid G] = \mathbb{E}[m^* \mid G] + \mathbb{E}[(1 - \rho I)\mathbb{E}[\xi \mid G, Z] \mid G] \tag{41}$$

$$= \mu(G). \tag{42}$$

The residual is conditionally orthogonal to functions of (G, Z), giving

$$V_\rho(G) := \mathrm{Var}(T_\rho \mid G) = \mathrm{Var}(m^* \mid G) + \mathbb{E}[(1 - \rho I)^2\sigma^2(G, Z) \mid G], \tag{43}$$

$$\mathrm{Var}(D \mid G) = \mathrm{Var}(m^* \mid G) + \mathbb{E}[\sigma^2(G, Z) \mid G]. \tag{44}$$

Subtract the first expression from the second and use $I^2 = I$. This proves Equations (10) and (40). □

No missing-at-random condition on I is required for this identity. Instead, it requires the same population weights in both targets and raw gains for $I = 0$. Restricting student training to $I = 1$ generally changes the target to $\mathbb{E}[D \mid G, I = 1]$. Separately normalizing available and unavailable groups to equal weight likewise changes the study law unless that weighting was part of its definition. The proposition concerns conditional target variance, not the variance of a fitted estimator, stochastic gradients, or final policy performance.

More privileged information need not give more denoising. Hold applicability I and the study law fixed. Suppose Z_2 includes Z_1 and both contain I. Let $m_j^* = \mathbb{E}[D \mid G, Z_j]$ and $R_j = \mathbb{E}[I\,\mathrm{Var}(D \mid G, Z_j) \mid G]$. Conditional total variance gives

$$R_1 - R_2 = \mathbb{E}[I\,\mathrm{Var}(m_2^* \mid G, Z_1) \mid G] \geq 0. \tag{45}$$

The richer ideal teacher therefore leaves weakly more variance in the mixed target at the same positive ρ. If Z determines D, no residual target variance can be removed. Conversely, when $I = 1$ for every example and the teacher uses only G, its ideal prediction is $\mu(G)$ and has zero conditional variance. Privileged state/change information must consequently be justified through the learnability or approximation of finite teachers, rather than through an assertion that more conditioning information always improves ideal denoising.

D.4 APPROXIMATE TEACHERS INTRODUCE BIAS AND COVARIANCE TERMS

For this subsection, condition on the frozen predictor and feature encoder, and on an auxiliary model whose complete fitting and configuration selection are independent of the evaluated participant. This requires independent participant sampling, data-independent fold allocation, and exclusion from model selection as well as fitting. The model prediction is then a function of (G, Z) and the fixed training sample. Write

$$\widehat{m} = m^* + e \quad \text{on } I = 1, \qquad \widehat{T}_\rho = T_\rho + \rho I e, \tag{46}$$

and set $e = 0$ on $I = 0$. Assume $\mathbb{E}[e^2 \mid G] < \infty$. Conditional expectations below also hold this excluding-participant fit fixed. Define

$$b(G) = \mathbb{E}[Ie \mid G], \qquad C(G) = \mathbb{E}[(m^* - \mu(G))Ie \mid G]. \tag{47}$$

Proposition 2 (Finite-teacher target error)**.** *The mixed target has bias*

$$\mathbb{E}[\widehat{T}_\rho \mid G] - \mu(G) = \rho b(G). \tag{48}$$

Its variance and squared error about the desired mean satisfy

$$\operatorname{Var}(\widehat{T}_\rho \mid G) = V_\rho(G) + 2\rho C(G) + \rho^2 \operatorname{Var}(Ie \mid G), \tag{49}$$

$$\mathbb{E}[(\widehat{T}_\rho - \mu(G))^2 \mid G] = V_\rho(G) + 2\rho C(G) + \rho^2 \mathbb{E}[Ie^2 \mid G]. \tag{50}$$

Proof. Mean preservation of T_ρ gives the bias expression, which is Equation (11). Since e is a function of the fixed model and (G, Z), $\mathbb{E}[\xi Ie \mid G] = 0$. Expanding the conditional variance of $T_\rho + \rho Ie$ therefore leaves the covariance $\operatorname{Cov}(m^*, Ie \mid G) = C(G)$. Adding the squared bias to the variance yields Equation (50). □

These terms need not favor distillation. Relative to the raw gain's conditional variance, the target squared-error difference is

$$-\rho(2 - \rho)R(G) + 2\rho C(G) + \rho^2 \mathbb{E}[Ie^2 \mid G]. \tag{51}$$

The covariance can be positive or negative. For $\rho > 0$, improvement occurs only if $(2 - \rho)R(G) > 2C(G) + \rho\mathbb{E}[Ie^2 \mid G]$. Writing $p_I(G) = \Pr(I = 1 \mid G)$, conditional Cauchy–Schwarz also gives

$$|\rho b(G)| \le \rho\sqrt{p_I(G)\mathbb{E}[Ie^2 \mid G]}. \tag{52}$$

In particular, $|e| \le \eta$ on applicable cases implies $|\rho b(G)| \le \rho p_I(G)\eta \le \rho\eta$.

The mixture therefore trades retained raw-target variation against teacher error. It cannot be selected by the oracle variance expression alone. When its denominator is positive, the hypothetical conditional squared-error-minimizing mixture is

$$\rho^*(G) = \operatorname{clip}_{[0,1]} \frac{R(G) - C(G)}{R(G) + \mathbb{E}[Ie^2 \mid G]}. \tag{53}$$

This follows by minimizing the quadratic in Equation (51). Its unknown population terms make it an explanatory expression, not a claimed implementation. The proposed procedure selects its mixture using held-out validation.

What participant cross-fitting does and does not establish. For each utility-data fold, fit the auxiliary model on the other participant folds and predict the excluded fold. All visits and prefixes of an excluded participant must remain outside that auxiliary fit. The preceding proof can then condition on that participant's excluding-fold model. It cannot condition on the entire collection of fold models and still assert independence, because other models can have used that participant. Out-of-fold targets also share training data across folds, so their aggregate uncertainty is not that of independent targets. Averaging over fitted teachers can add variation in their conditional target means. The per-fit identities are therefore not a formula for the empirical variance of a pooled set of cross-fitted targets. The identities hold for a fixed mixing weight and model configuration. Conditioning on a globally selected configuration or mixture after inspecting student performance does not automatically preserve the independence premise.

If the auxiliary model fits the same participant it supervises, $\mathbb{E}[I\xi e \mid G]$ need not vanish. Before any conditioning that would disclose that participant, a direct expansion adds $2\rho(1-\rho)\mathbb{E}[I\xi e \mid G]$ to the displayed second-moment decomposition. Cross-fitting addresses this reuse, but does not ensure small e, zero $b(G)$, or a beneficial covariance. The population conditional-mean target for student regression also requires its specified weights and function class. Positive auxiliary regularization or a restricted student class need not attain that mean.

D.5 Contrasts between tests and comparison with STOP

Fix the fitted diagnostic predictor q_θ, every gain model being compared, and their hyperparameters before drawing held-out participants. Use the same selected population $P_{\mathcal{A}}$, participant and visit weights, and held-out prefix sampler for every comparison. The sampler is nonanticipating: each reveal choice and its randomness use only information available at that prefix, without inspecting hidden results or labels. A common sampler may be fixed after training, but must not be separately adapted to each compared model's held-out outcomes. All expectations below condition on the fitted models and use this same population and prefix law. Training randomness is therefore not left implicit in a conditional expectation over a test record.

At history $H = h$, let $\mathcal{L}(h)$ contain the $J \geq 1$ legal tests, ordered deterministically. Their identities and costs are known at h. The complete-library population supplies a held-out result for every test in this set. Define

$$L_0 = \ell(q_\theta(h), Y), \qquad L_a = \ell(q_\theta(h \oplus (a, X_a)), Y), \tag{54}$$

$$\mathbf{D}_h = (L_0 - L_a)_{a \in \mathcal{L}(h)}, \qquad \boldsymbol{\mu}_h = \mathbb{E}[\mathbf{D}_h \mid H = h], \tag{55}$$

$$\mathbf{g}_h = (g_\phi(h, a))_{a \in \mathcal{L}(h)}, \qquad \mathbf{C}_h = \mathbf{I}_J - \frac{1}{J}\mathbf{1}\mathbf{1}^\mathsf{T}. \tag{56}$$

Here $\mathbf{I}_J$ is the identity matrix and $\mathbf{1}$ is the length-J vector of ones. The notation $\boldsymbol{\mu}_h$ denotes a vector, whose action-a component $\mu(h, a)$ is the conditional mean gain for that action under the specified history law. Because $\mathbf{C}_h\mathbf{1} = 0$,

$$\mathbf{C}_h\mathbf{D}_h = -\mathbf{C}_h(L_a)_{a \in \mathcal{L}(h)}. \tag{57}$$

Thus between-test contrasts cancel the shared loss before acquisition.

For any vector $\mathbf{e} \in \mathbb{R}^J$, put $\bar{e} = \mathbf{1}^\mathsf{T}\mathbf{e}/J$. The centering matrix is symmetric and idempotent, and $\mathbf{e} = \bar{e}\mathbf{1} + \mathbf{C}_h\mathbf{e}$ is an orthogonal decomposition. Its squared norm proves Equation (12). Equivalently, $\sum_{a<b}(e_a - e_b)^2 = J\|\mathbf{C}_h\mathbf{e}\|_2^2$. The centered term measures pairwise differences, while the common component shifts every test value equally. Adding a common offset to predicted gains leaves every ranking among tests unchanged, including rankings after subtracting the same known action costs. It can still change whether the best predicted net value exceeds STOP's fixed value of zero.

What a held-out contrast loss estimates. Conditional on $H = h$, $\mathbf{g}_h$ and $\mathbf{C}_h$ are fixed. Write $\mathbf{g}_h - \mathbf{D}_h = (\mathbf{g}_h - \boldsymbol{\mu}_h) - (\mathbf{D}_h - \boldsymbol{\mu}_h)$ and expand the centered squared norm:

$$\begin{aligned}\mathbb{E}[\|\mathbf{C}_h(\mathbf{g}_h - \mathbf{D}_h)\|_2^2 \mid H = h] = \|\mathbf{C}_h(\mathbf{g}_h - \boldsymbol{\mu}_h)\|_2^2 \\ + \mathbb{E}[\|\mathbf{C}_h(\mathbf{D}_h - \boldsymbol{\mu}_h)\|_2^2 \mid H = h].\end{aligned} \tag{58}$$

The cross term vanishes because $\mathbb{E}[\mathbf{D}_h - \boldsymbol{\mu}_h \mid H = h] = 0$. The second term is common to gain models scored on the same held-out law. Expected contrast-loss differences therefore equal differences in squared error for the conditional mean contrasts. This loss identifies contrasts, not the common level of all gains. If histories have different numbers of legal actions, division by J provides a per-action normalization. The chosen normalization and history weights must remain the same across compared models. At $J = 1$, the centered loss is zero and supplies no information about the remaining test-versus-STOP decision. At $J = 0$, there is no test to compare.

These are statements about the population risk of a fixed fitted model. They do not make repeated prefixes independent observations, remove the need for participant-level uncertainty, or imply that a lower average contrast loss changes the maximizing action correctly. Each participant's realized best test also need not be the optimal conditional action, since it uses hidden results and the held-out label.

Measuring the selected action's value. Let $\pi(h)$ be a fixed decision rule using only h, with STOP permitted, and set $D_{\mathrm{STOP}} = c_{\mathrm{STOP}} = \mu(h, \mathrm{STOP}) = 0$. For a deterministic rule, iterated expectation gives

$$\mathbb{E}[D_{\pi(H)} - \lambda c_{\pi(H)}] = \mathbb{E}[\mu(H, \pi(H)) - \lambda c_{\pi(H)}]. \tag{59}$$

The same identity holds for randomized rules whose action randomization is independent of hidden results and labels conditional on H. This justifies evaluating realized selected gains and net values on a common held-out prefix law. It does not identify each individual's conditional regret. Full policies induce their own history distributions, so terminal loss and acquisition cost require a separate rollout comparison on the same participant population. The contrast and STOP diagnostics explain possible sources of an effect rather than replacing that comparison.

D.6 Consequences for a single acquisition decision

Proposition 3 (Elementary utility stability)**.** *Fix a history, predictor, study law, and finite feasible set including STOP. If estimated net values satisfy* $\max_a |\widehat{u}_a - u_a| \le \kappa$*, a maximizing estimated action* $\widehat{a}$ *has regret at most* 2κ *relative to a true maximizer* $a^\star$*. A unique optimum with gap greater than* 2κ *is unchanged.*

Proof. By the two maximizing properties,

$$0 \le u_{a^\star} - u_{\widehat{a}} = (u_{a^\star} - \widehat{u}_{a^\star}) + (\widehat{u}_{a^\star} - \widehat{u}_{\widehat{a}}) + (\widehat{u}_{\widehat{a}} - u_{\widehat{a}}) \tag{60}$$

$$\le \kappa + 0 + \kappa. \tag{61}$$

Choosing a different action under the stated strict gap would contradict the inequality. □

For the estimated privileged target, suppose the student uniformly approximates its conditional mean $\mu(h, a) + \rho b(h, a)$ within ε, and $|b(h, a)| \le \eta$ over feasible tests. Its gain error relative to μ is then at most $\varepsilon + \rho\eta$. Subtracting the same known costs leaves that error unchanged. Applying Proposition 3 gives

$$u_q^{\mathcal{A}}(h, a^\star) - u_q^{\mathcal{A}}(h, \widehat{a}) \le 2(\varepsilon + \rho\eta). \tag{62}$$

Because STOP has exact value zero, a mistake involving STOP has regret at most $\varepsilon + \rho\eta$. This implication requires simultaneous error bounds. Average training loss does not establish them, and the result does not bound the risk of a full adaptive trajectory.

D.7 What the state/change parameterization changes

For one student pathway, concatenate the projected heads and targets as $\widehat{t}_i = (P_{\mathrm{cur}} z_i^{\mathrm{cur}}, P_\delta z_i^\delta)$ and $t_i = (t_i^{\mathrm{cur}}, t_i^\delta)$. With the example weights and valid-target normalizers from Appendix C, put $w_{ib} = \alpha_b \omega_i m_i^b / (d_b S_b)$ if $S_b > 0$, and zero otherwise. The feature objective is exactly

$$\sum_{i,b} w_{ib} \|\widehat{t}_i^b - t_i^b\|_2^2 = \sum_i \|\Omega_i(\widehat{t}_i - t_i)\|_2^2, \tag{63}$$

$$\Omega_i = \mathrm{diag}(\sqrt{w_{i,\mathrm{cur}}}\mathbf{I}_{d_{\mathrm{cur}}}, \sqrt{w_{i,\delta}}\mathbf{I}_{d_\delta}). \tag{64}$$

Invalid target blocks can have arbitrary algebraic placeholders because their weights are zero. No fabricated trajectory is used for supervision. Thus segmented MSE and the matching weighted concatenated MSE are the same objective. A comparison must also match the projection architecture and residual channels, since changing these can alter the function class.

For an observed temporal pair with the same modality subset, put $r_j = F(W_{\tau_j}^M)$ and $\Delta\tau = \tau_1 - \tau_0 > 0$. The map from paired features to current/change coordinates and its inverse are

$$(r_1, r_0) \mapsto (r_1, \Delta r), \qquad \Delta r = (r_1 - r_0)/\Delta\tau, \qquad r_0 = r_1 - \Delta\tau\, \Delta r. \tag{65}$$

At a known interval, it creates no additional information. With endpoint prediction errors e_1, e_0 and nonnegative current and change weights α, β, current/change squared error becomes

$$\alpha\|e_1\|_2^2 + \frac{\beta}{(\Delta\tau)^2}\|e_1 - e_0\|_2^2. \tag{66}$$

The parameterization therefore selects a metric on endpoint errors, including interval-dependent weighting. Whether that preference aids finite learning is an empirical question. A fair paired-endpoint comparator receives the same observations, modality masks, and elapsed times. If the current-state target uses additional modalities beyond the pair's common subset, its encoding is a shared additional covariate. The current/change arm still retains the pair-matched endpoint r_1, which is necessary for the displayed inverse. Equation (80) gives the two matched input tuples.

Finally, for invertible Q, replacing a student segment z by Qz and its projection P by PQ^{-1} preserves feature matching. The fitted projections therefore do not identify the raw latent coordinates as biological state or change. Frozen teacher coordinates make targets consistent across examples, but do not remove this interpretive limitation. Their role here is to supply candidate descriptors for the richer gain model, whose acquisition relevance is separately tested.

D.8 Supporting examples and limits of myopic selection

Different predictors can have different information needs.

Proposition 4 (Different information can imply different test values)**.** *There is a classification problem in which a test has strictly positive net value for the student, while its net value for a richer Bayes classifier is strictly negative at the same cost.*

Proof. Let Y be a fair binary label. The student initially observes an empty history, the privileged variable is $W = Y$, and a test reveals $X = Y$. The student's Bayes posterior changes from $1/2$ to the correct label. Its realized and expected Brier gains are both $1/4$. The richer classifier already knows Y from W, so its loss is zero before and after the test. For effective cost $c \in (0, 1/4)$, the student's net value is $1/4 - c > 0$, whereas the richer classifier's net value is $-c < 0$. □

The example concerns target alignment. A richer model predicting the same D as the student has ideal target $1/4$ here. A model supplying its own diagnostic gain instead supplies zero and answers the wrong acquisition question. Since D is already constant in this construction, the example does not demonstrate a variance benefit from distillation. It is a caution about changing the target predictor, not evidence of a failure in gain regression or privileged value methods that already define values for the intended predictor or policy (Gadgil et al., 2024; Ebi et al., 2026).

Conditional feature means need not determine the useful action. Let an observed fair bit H be independent of a fair sign S, set $Y = \mathbf{1}\{S = +1\}$, and consider abstract teacher targets $T^{\mathrm{cur}} = S$ and $T^{\delta} = (1 + H)S$. Their population-optimal squared-error predictions given H are both zero at both contexts. This follows from

$$\mathbb{E}[(T - f(H))^2 \mid H] = \mathrm{Var}(T \mid H) + (f(H) - \mathbb{E}[T \mid H])^2. \tag{67}$$

Nevertheless, for tests $X_a = (1 - H)S$ and $X_b = HS$, only a is informative at $H = 0$ and only b at $H = 1$. With a Bayes predictor and equal effective cost $c \in (0, 1/4)$, the informative test has net value $1/4 - c$ and the other has value $-c$. An actor seeing only the identical feature means and initial posterior cannot select both distinct optima. This concerns an exclusive bottleneck, not the full ACUMEN architecture, which retains clinical context. It also does not establish an advantage of the particular temporal descriptor over other privileged inputs.

Complementarity survives perfect gain prediction. Let X_a, X_b be independent fair bits and let their exclusive OR be Y. The Bayes posterior is $1/2$ both initially and after either single test, so each individual Brier gain is zero. At effective cost $0 < c < 1/8$ per test, a perfect one-step policy chooses STOP, while acquiring both tests reduces risk from $1/4$ to zero at net value $1/4 - 2c > 0$. Privileged target alignment does not repair this horizon limitation. Nongreedy acquisition addresses a separate question (Valancius et al., 2024).

E Limits of retrospective policy evaluation

An acquisition evaluation must distinguish a measurement that a policy could request from a measurement whose result a historical record stores. This appendix makes that distinction precise using

established missing-data and partial-identification arguments (Manski, 1990; 2003; Little & Rubin, 2019). These arguments describe limits of the evaluation design. They are not a claimed innovation of ACUMEN or evidence for its representation architecture.

E.1 Information access and the population being evaluated

Let B contain information available before the acquisition episode, let $Y \in \{1, \dots, C\}$ be an observed reference diagnosis, and let $X = (X_f)_f$ contain the complete measurement fields. Each structurally eligible action $a \in \mathcal{A}$ reveals $X_a = (X_f : f \in F_a)$. Its prerequisites are fixed by the action specification. For instance, an instrument absent from a battery version is outside that version's action library.

The historical field mask D_f indicates whether a result is stored and usable. An action is fully observed if $M_a = \prod_{f \in F_a} D_f = 1$. The full-data record is (B, Y, D, X), whereas the evaluator observes

$$O = (B, Y, D, X_{\text{obs}}), \qquad X_{\text{obs}} = \{(f, X_f) : D_f = 1\}. \tag{68}$$

Thus the record retains known fields of a partly observed action even when $M_a = 0$. An absent result limits replay. It does not establish that the test would be unavailable if requested prospectively.

A fixed policy chooses legal actions from the ordered revealed history $h_t = (B, (a_i, X_{a_i})_{i=1}^t)$. It cannot use the reference label, unrevealed values, or the evaluator's storage mask to select requests. Legitimate initial information, such as modality availability known before the episode, may belong to B. Fix the predictor and policy randomness throughout this analysis. At a requested depth k, retain the terminal history if the policy has already stopped, and denote the resulting top-one diagnostic error by

$$L_{\pi,k} = \mathbf{1}\{\widehat{Y}(h_{\pi,k}) \neq Y\}. \tag{69}$$

This binary evaluation loss differs from the Brier loss used to train the one-step utilities in Section 3. The sensitivity model below does not certify those utilities or the stopped policy's diagnostic risk.

For a fixed sampling law, fix a common allowed action set $A \subseteq \mathcal{A}$ and define $E_A = \{M_a = 1 \text{ for all } a \in A\}$. For policies restricted to A, the complete-record and broad-population targets are

$$\psi_A^{\text{complete}}(\pi, k) = \mathbb{E}[L_{\pi,k} \mid E_A], \qquad \psi^{\text{broad}}(\pi, k) = \mathbb{E}[L_{\pi,k}], \tag{70}$$

where the conditional target requires $\Pr(E_A) > 0$. Complete-record replay reveals stored results without imputation, but its target concerns the selected population. It also assumes that those static results remain applicable under the simulated ordering, rather than establishing equivalence to prospective clinical testing (von Kleist et al., 2023; 2025).

The formulas in this appendix refer to that specified law. If complete records are reweighted to the participant-balanced law $P_{\mathcal{A}}$ in Section 2, expectations must use those weights. Conditioning an arbitrary visit-weighted law on E_A does not itself implement participant balancing. The distinction between selected and broad populations remains, with the same weighting convention used for both compared policies.

Let $R_{\pi,k} = 1$ when all actions requested in the policy's prefix through depth k have usable results. Then E_A implies $R_{\pi,k} = 1$, but the latter group can be larger and policy-specific. For an adaptive policy, replay either finishes the prefix or encounters a first missing request. That request already establishes $R_{\pi,k} = 0$, regardless of the unknown continuation. Hence evaluability is observable even when the remaining trajectory is not. Treating that missing request as STOP would change the policy under evaluation.

E.2 Identical records can conceal opposite rankings

The following construction keeps labels, prediction rules and feature relationships shared across policies. Take constant B, two prerequisite-free actions a, b, and $Y = 1$ with $C \geq 2$. For a binary variable Z, let $X_a = Z$ and $X_b = 1 - Z$. Policy A acquires a and policy B acquires b. Both use the same predictor, returning class 1 for acquired value zero and class 2 for acquired value one. Their losses are therefore $L_A = Z$ and $L_B = 1 - Z$.

Let R be Bernoulli with probability $1/2$ and set $D_a = D_b = R$, so both results are stored together. In both complete-data laws, set $\Pr(Z = 1 \mid R = 1) = 1/2$. In world 1, let $\Pr(Z = 1 \mid R = 0) = 1/4$,

and in world 2 let it equal $3/4$. The two laws agree on B, Y, D and on every stored feature value. When $R = 0$, changing the law of Z changes no component of the observed record in Equation (68). Yet total expectation gives

	$\psi^{\mathrm{broad}}(A,1)$	$\psi^{\mathrm{broad}}(B,1)$
World 1	$3/8$	$5/8$
World 2	$5/8$	$3/8$

(71)

because, for example, $\Pr_1(L_A = 1) = \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{4} = 3/8$. Thus no functional of the observed-data law can recover the correct ranking in both worlds. The observed loss odds equal one, while the missing loss odds are $1/3$ or 3. Both worlds consequently satisfy an odds restriction with factor three for both policies. This is an existence example, not a claim that every policy comparison is ambiguous or that this factor is clinically plausible.

E.3 An explicit sensitivity model for missing binary losses

For a fixed policy and depth, suppress their subscripts and write L, R. Let S index prespecified finite strata of observed metadata, with $w_s = \Pr(S = s)$, $r_s = \Pr(R = 1 \mid S = s)$ and, when $r_s > 0$, $p_s = \Pr(L = 1 \mid R = 1, S = s)$. The unknown missing-group risk is $v_s = \Pr(L = 1 \mid R = 0, S = s)$. A missing binary outcome sensitivity model restricts its odds relative to the observed risk (Higgins et al., 2008):

$$\Lambda^{-1} \le \frac{v_s/(1-v_s)}{p_s/(1-p_s)} \le \Lambda, \qquad \Lambda \ge 1. \tag{72}$$

For interior p_s, define $\omega_t(p) = tp/(1-p+tp)$. The allowed interval for v_s is $[\omega_{\Lambda^{-1}}(p_s), \omega_\Lambda(p_s)]$. At $p_s = 0$ or 1, we use this map's continuous extension, which fixes $v_s = p_s$ at every finite Λ. With $r_s = 0$, there is no observed risk to tilt and v_s remains in $[0,1]$. With $r_s = 1$, no missing-group risk is needed.

Total expectation yields the bounds

$$\begin{aligned} \underline{\psi}(\Lambda) &= \sum_{s:r_s>0} w_s\{r_s p_s + (1-r_s)\omega_{\Lambda^{-1}}(p_s)\}, \\ \overline{\psi}(\Lambda) &= \sum_{s:r_s>0} w_s\{r_s p_s + (1-r_s)\omega_{\Lambda}(p_s)\} + \sum_{s:r_s=0} w_s. \end{aligned} \tag{73}$$

They are sharp for the coarsened observation (S, R, RL) with no cross-stratum restrictions. To see this, retain its observed law and assign a Bernoulli loss with any allowed probability v_s on each missing branch. Choosing all lower or all upper probabilities attains the displayed endpoints. Interpolating the probabilities attains every intermediate risk.

That construction need not preserve the full record's joint feature and label distribution or the fixed predictor. Equation (73) therefore provides valid *outer bounds* for the full feature-acquisition model whenever the loss-odds restriction holds. Partial results or deterministic predictor relationships can narrow them. Full-record sharpness requires its own compatibility argument, as in the explicit constructions here.

At $\Lambda = 1$, supported strata equate observed and missing loss risks. This is an assumption about policy-induced loss, not a missing-at-random model for all measurement fields or a model of clinician ordering propensities. Unsupported strata retain their full mass interval even at this baseline. The bounds are nested in Λ and contained in the unrestricted interval

$$\left[\sum_{s:r_s>0} w_s r_s p_s, \quad \sum_{s:r_s>0} w_s r_s p_s + \sum_s w_s(1-r_s)\right]. \tag{74}$$

Its width is the missing-prefix fraction. Convergence to both unrestricted endpoints requires interior observed risk in every partially observed stratum with positive missing mass. Boundary risks remain fixed under the finite-index odds model.

E.4 Marginal intervals and joint policy comparisons

Assume the complete-data model satisfying all stated restrictions is nonempty. If $\overline{\psi}_A < \underline{\psi}_B$, policy A has lower risk under every compatible law. Overlap does not establish an attainable reversal, because the policies share features and labels. The relevant object is the joint difference set $\{\psi_A(P) - \psi_B(P) : P \text{ is compatible}\}$. Subtracting separate marginal intervals only gives an outer bound on that set.

Two exact examples make this distinction concrete. They share stored fraction $1/2$, observed risks $p_A = 2/5, p_B = 3/5$, and the fixed one-action policies and predictor above. All values follow from specified probability laws.

Complementary losses. Retain $X_a = Z$, $X_b = 1 - Z$, but set $\Pr(Z = 1 \mid R = 1) = 2/5$. The odds restriction for both policies permits exactly

$$a_L(\Lambda) = \frac{2}{3\Lambda + 2} \le v_A \le a_U(\Lambda) = \frac{2\Lambda}{3 + 2\Lambda}, \qquad v_B = 1 - v_A. \tag{75}$$

Every such value defines a compatible Bernoulli missing-group law. Consequently $\psi_A = 1/5 + v_A/2$ and $\psi_B = 1 - \psi_A$. Policy A is strictly better for every compatible law precisely when $a_U < 3/5$, equivalently $\Lambda < 9/4$. A tie becomes possible at $9/4$ and reversal above it. At $\Lambda = 3$, $v_A = 2/11$ yields risks $(16/55, 39/55)$, whereas $v_A = 2/3$ yields $(8/15, 7/15)$. Both laws preserve every observed record distribution.

Nested losses. Instead impose $X_a \le X_b$ on all complete records, including missing ones. In the stored group assign probabilities $(2/5, 1/5, 2/5)$ to $(X_a, X_b) = (0,0), (0,1), (1,1)$. This preserves the same observed marginal risks. The structural ordering is an additional assumption, not a conclusion from the observed absence of $(1,0)$. The missing risks must satisfy $v_A \le v_B$, so

$$\psi_A - \psi_B = -\frac{1}{10} + \frac{1}{2}(v_A - v_B) \le -\frac{1}{10}. \tag{76}$$

The marginal odds intervals are $v_A \in [a_L, a_U]$ and $v_B \in [1 - a_U, 1 - a_L]$. Their joint difference set is

$$\left[a_L - \frac{3}{5}, \quad \min\left\{a_U - \frac{3}{5}, -\frac{1}{10}\right\}\right]. \tag{77}$$

Indeed, any allowed $v_A \le v_B$ is attained by missing-group masses $(1 - v_B, v_B - v_A, v_A)$ on the three ordered pairs. The lower endpoint uses $(v_A, v_B) = (a_L, 1 - a_L)$. The upper endpoint uses $(a_U, 1 - a_U)$ for $\Lambda \le 3/2$ and $(1/2, 1/2)$ otherwise. Convex interpolation supplies every intermediate difference. At $\Lambda = 3$, subtracting marginal risk intervals gives $[-23/55, 1/15]$, while the sharp joint range is $[-23/55, -1/10]$. A strict advantage therefore survives despite marginal overlap.

E.5 Estimation and implications for the acquisition study

Plugging empirical stratum weights, support fractions and observed error rates into Equation (73) estimates population bounds. It does not produce confidence limits. A sample stratum with no evaluable records must contribute its full empirical mass interval. An empirical zero error rate also need not imply a population boundary risk, although a direct odds-model plug-in fixes its missing risk to zero. Small-stratum uncertainty must remain visible. For repeated visits, an uncertainty analysis should preserve participant clusters and match the target sampling design. Separate percentile intervals for two estimated endpoints do not automatically give simultaneous coverage of the population identified set.

These distinctions constrain interpretation of ACUMEN's retrospective evaluation. Common-complete replay supports a comparison on its selected population. Model-based completion adds assumptions about unrecorded results. Neither establishes broad-population policy benefit without further support. The representation and acquisition ablations must therefore hold the evaluation population and information access fixed, report evaluable support, and separate sampling uncertainty from sensitivity to missing results. This makes the evaluation more informative without turning partial identification into a claimed contribution of the acquisition architecture.

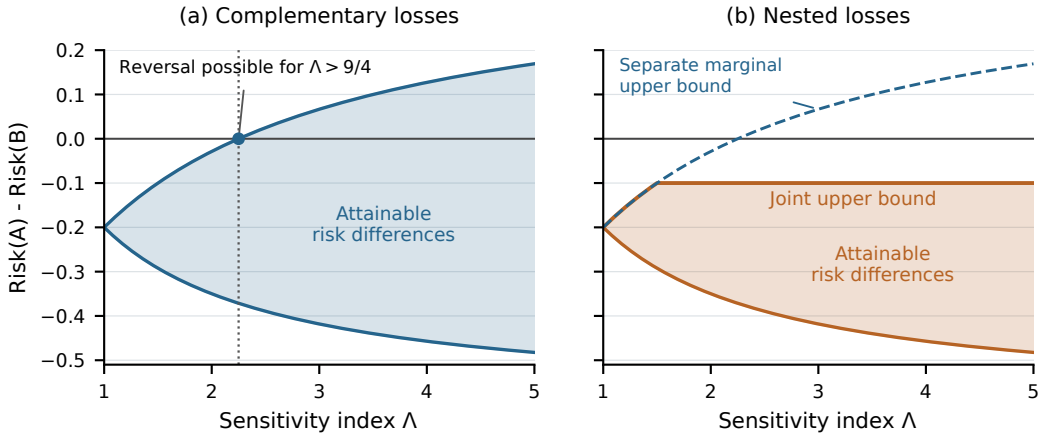


Figure 4: Exact sensitivity examples with identical observed risks and stored fractions. Negative differences favor policy A. Complementary losses permit reversal above $\Lambda = 9/4$, whereas nested losses preserve a strict advantage despite overlapping marginal intervals. Shading denotes attainable population ranges under the specified models, not confidence intervals or clinical results.

F Computational acquisition evaluation

Section 4 evaluates diagnostic classification and feature transfer for the diagnostic model. This appendix specifies the current-visit acquisition study that gain supervision still requires, whose results cannot be inferred from those classification metrics. Its central comparison holds the diagnostic predictor fixed and changes the acquisition supervision, separating better test selection from better diagnosis given the same inputs. Complete-system comparisons then assess their combined effect on diagnostic loss and acquisition count. The scope is computational acquisition performance. Clinician interaction and human-in-the-loop outcomes belong to a separate medical study and are not prerequisites for this comparison. The protocol below defines these acquisition experiments, which have not yet been run.

F.1 Reference diagnosis and eligible episodes

The primary endpoint is a categorical recorded diagnosis at visit v: Normal, mild cognitive impairment (MCI), Alzheimer's disease (AD), Lewy body dementia (LBD), frontotemporal dementia (FTD), or Other Dementia. Cognitive status is resolved first using the release-specific NACC definitions. Under the documented derived status coding, normal cognition maps to Normal and MCI to MCI. Impairment that does not meet MCI criteria is outside this endpoint. Only visits classified as dementia enter the etiologic mapping (NACC, 2025).

For those dementia visits, AD, LBD, and FTD require a documented, unambiguous primary etiology in the corresponding group. FTD is the display label for dementia attributed to a primary frontotemporal lobar degeneration etiology. A syndrome such as primary progressive aphasia does not by itself establish that etiology. Other Dementia requires an explicitly documented primary cause outside the three named groups. It is not a residual label for an unknown, pending, contradictory, or unresolved mixed diagnosis. These distinctions follow the separation of cognitive status, syndrome, and primary versus contributing etiologies in the diagnostic form (NACC, 2015c).

The proposed mapping excludes records with multiple conflicting primary causes or an unresolved mixed primary diagnosis. Additional contributing conditions remain recorded for subgroup interpretation. Their presence does not silently change a documented primary label or imply a biologically pure disease. Unknown status, missing primary cause in dementia, and other exclusions receive separate counts. The completed cohort description must report retained and excluded participants and visits by battery version and class, together with the actual data release and observation period.

This target measures agreement with a recorded clinical diagnosis, not an independent biological gold standard. The clinical assessment may itself use cognitive tests. Direct diagnosis fields, diagnostic-role codes, target-defining status summaries, and their derived features are excluded from both decision pathways. The primary action library also excludes CDR summaries. Predicting this recorded endpoint must therefore be distinguished from establishing a new biological etiology or improving patient outcomes.

F.2 Version-specific actions and burden

Table 13 fixes the returned fields for two proposed libraries. The C1 library is a nine-action legacy core, including when a later C1 form records additional tests outside that core. The C2 library contains thirteen actions. The administered form and packet determine the applicable library, rather than the calendar year alone. C1 remained an option for some follow-up assessments after C2 was introduced. Telephone or video C2T batteries require a separate action specification and are not pooled with these libraries. Retain form version, administration mode, and language as protocol metadata, and report their composition (NACC, 2008; 2025; 2015a;b).

C1 attention uses Digit Span, whereas C2 uses Number Span. Their different trial and span ranges are retained, without recoding them as the same measurement. MMSE and MoCA, Logical Memory and Craft recall, and Boston Naming and MINT are likewise distinct instruments. Their common positions in the table identify the acquisition role, not score equivalence.

The returned variables are prespecified summary scores. MoCA uses its raw `MOCATOTS` total, without substituting the education-corrected `NACCMOCA` or revealing its item scores. Craft uses verbatim recall only, so the paraphrase alternatives `CRAFTURS` and `CRAFTDRE` are not additional returns. Trails returns only its recorded time score. Commission errors and correct-line counts in `TRAILARR`, `TRAILALI`, `TRAILBRR`, and `TRAILBLI` are excluded. The primary current-visit inputs also exclude Benson recognition `UDSBENRS`, MINT cue and component scores, fluency repetition and error counts, combined F/L totals, and recall-delay metadata. These unselected fields cannot leak a hidden action through another input channel. Alternative score bundles would define additional, separately specified comparisons.

The table combines the dictionary revision dated 8 April 2025 with the administered forms' scoring rules. One discrepancy requires an explicit decision: the dictionary permits `DIGIBLEN` up to 8, whereas the version 2 C1 guide and version 3 follow-up C1 form specify a maximum of 7. The primary protocol uses $[0, 7]$. An exported value of 8 is flagged for source clarification, excluded from the primary analysis, and counted separately. It is neither clipped nor silently accepted (NACC, 2025; 2008; 2015a).

A field is usable only when its administered form supports that field, the stored value is a valid score, and it refers to the current assessment. Blank entries, structural code -4, and documented missing or nonadministration codes are not scores. In particular, 88 invalidates the selected MMSE or MoCA total. Codes 95–98 indicate noncompletion reasons for the relevant bounded count scores, whereas Trails uses 995–998. Thus a Trails time of 95–98 seconds remains valid. Values of 150 and 300 seconds retain the instruments' recorded time-limit convention and are not discarded merely because the participant did not finish within that limit (NACC, 2025; 2015b).

Apply the source release's field-specific codebook before numerical standardization. If another release changes a field or its coding, reconcile that difference explicitly rather than silently applying these ranges. A span bundle is usable only when all four selected scores are valid. A partly stored bundle retains its known fields for the observation audit but cannot be replayed as a complete action. By contrast, an unselected auxiliary score does not affect completeness. Previous-administration fields such as `LOGIPREV` never replace `LOGIMEM` as a current-action outcome. Any permitted previous result belongs to dated baseline history instead.

Immediate recall precedes delayed recall, and Benson copy precedes delayed recall. These restrictions define legal reveal sequences without reconstructing clinical waiting times. The primary cost assigns $c_a = 1$ to each newly acquired action, including a complete span bundle. Report realized action count and the imposed cap separately. Neither count is converted to minutes, financial cost, or patient burden without independently justified weights. Baseline information already present before the episode incurs no incremental test charge.

Table 13: Exact primary action returns. Each row is one acquisition action in its library. Brackets denote inclusive integer score ranges, while braces specify noncontiguous valid sets. The attention action returns all four listed scores jointly. "Not selected" denotes an action outside the proposed library, not a missing patient result. Field definitions follow the NACC dictionary and forms (NACC, 2025; 2015a;b). The backward Digit Span length uses the form-specific restriction described in this section.

Action	C1 field and valid scores	C2 field and valid scores
General cognition	NACCMMSE $[0, 30]$	MOCATOTS $[0, 30]$
Immediate verbal recall	LOGIMEM $[0, 25]$	CRAFTVRS $[0, 44]$
Delayed verbal recall	MEMUNITS $[0, 25]$	CRAFTDVR $[0, 44]$
Benson copy	Not selected	UDSBENTC $[0, 17]$
Benson delayed recall	Not selected	UDSBENTD $[0, 17]$
Attention span bundle	DIGIF, DIGIB: $[0, 12]$ DIGIFLEN: $[0, 8]$ DIGIBLEN: $[0, 7]$	DIGFORCT, DIGBACCT: $[0, 14]$ DIGFORSL: $\{0, 3, \ldots, 9\}$ DIGBACLS: $\{0, 2, \ldots, 8\}$
Trails A	TRAILA $[0, 150]$	TRAILA $[0, 150]$
Trails B	TRAILB $[0, 300]$	TRAILB $[0, 300]$
Animal fluency	ANIMALS $[0, 77]$	ANIMALS $[0, 77]$
Vegetable fluency	VEG $[0, 77]$	VEG $[0, 77]$
Naming	BOSTON $[0, 30]$	MINTTOTS $[0, 32]$
Letter F fluency	Not selected	UDSVERFC $[0, 40]$
Letter L fluency	Not selected	UDSVERLC $[0, 40]$

F.3 A COMMON POPULATION AND A CONSISTENT CLOCK

For each fixed action library $\mathcal{A}$, first apply the clinical eligibility and label rules, then retain visits with usable results for every action. Define $P_{\mathcal{A}}$ by drawing from the eligible participant population conditional on having at least one complete-library visit, then sampling one of that participant's eligible complete visits uniformly. Its empirical counterpart weights retained participants equally, then their retained visits equally. This law is not automatically obtained by conditioning an arbitrary visit distribution on complete storage. If retained participant i has n_i eligible complete visits, an episode outcome M_{iv} is summarized as

$$\widehat{M} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{n_i} \sum_{v=1}^{n_i} M_{iv}. \tag{78}$$

All competing policies use the same retained episodes and weights. A smaller library defines a separate comparison population and must be fixed before test evaluation, rather than chosen for a favorable ranking. Report retention and class composition against the broader clinically eligible cohort. A visit-weighted analysis is secondary and estimates a different target. Participants can contribute to separate version-specific analyses, but results from those populations are not interchangeable.

At each episode, $h_{vk} = (B_v, V_v, (a_j, X_{va_j})_{j \le k})$ contains only the permitted baseline history, initially available modalities, and acquired current-visit results. Later visits cannot supply either a predictor input or the current outcome. Masking a current result also removes its descendants from aggregates, trend features, cached representations, and fusion inputs. Offline target construction and the evaluator can access withheld results and reference labels. The deployed policy cannot. Historical storage indicators cannot determine its legal action mask.

All primary replay losses concern $P_{\mathcal{A}}$. Extrapolating to incomplete records requires additional assumptions or a separately justified estimator. Imputation and exact stored-result replay are reported as different procedures. Appendix E details the corresponding support and identification limits.

F.4 PARTITIONS, FITTING, AND SELECTION

Partition participants into D_{fit}, D_{value}, D_{val}, and D_{test} before fitting. Keep all visits from one participant together. Fit preprocessing, the feature encoder F, and the diagnosis pathway on D_{fit}.

External pretraining data require overlap checks against the held-out participants. Rich observations remain subject to the episode cutoff even when hidden from the student.

Select the diagnostic configuration and any probability transformation using D_{val}, then freeze q. Generate its one-test gain labels on the retained complete-library episodes in D_{value}, using the same q before and after each legal reveal. Within this partition, randomly divide participants into folds independently of their measured data and gain targets. Fit each value teacher on the other folds and use it to predict the examples in the omitted fold. The held-out participant contributes no visit, prefix, or action target to the teacher producing that participant's predictions. Cross-fitting the value teacher is separate from fitting the feature encoder F on D_{fit}. Its configuration selection likewise excludes the omitted participants, using only the fitting folds and external validation participants.

Let I_{iv} indicate that the rich descriptors required by the selected teacher configuration are available. For an observed gain d_{ivka} and its out-of-fold teacher prediction $\widehat{m}_{ivka}^{(-i)}$, the student target is

$$t_{ivka} = d_{ivka} + \rho I_{iv}\big(\widehat{m}_{ivka}^{(-i)} - d_{ivka}\big). \tag{79}$$

The superscript $(-i)$ means that the teacher was fitted without participant i's fold. When $I_{iv} = 0$, the target is d_{ivka} and no teacher prediction is required. Rich-target availability therefore changes the supervision, not the selected study population. The student retains every episode with its original participant and visit weight. The teacher is fitted on its valid rich-input examples with those weights, normalized over that fitting subset.

Average candidate-action losses uniformly within each sampled prefix, and use the same prefix distribution and sampling effort across variants. This prevents episodes with more legal actions or more sampled prefixes from receiving unintended extra weight. Teacher validity indicators may enter the training-only teacher, but they cannot reveal hidden modality availability to the deployed student.

Validation selects ρ, feature-transfer weights, teacher and student capacity, and acquisition settings under matched tuning opportunities. Include $\rho = 0$, which recovers raw gain supervision. If q changes, regenerate its gains and refit the value teachers and acquisition student. Freeze the complete procedure before examining D_{test}. Cross-fitting avoids fitting a teacher to the very participant it supervises. It does not remove teacher approximation bias or establish a generalization advantage. Repeated model seeds likewise do not increase the number of independent participants.

F.5 Comparisons that isolate value supervision

The primary contrasts keep q, student inputs, action rules, episode weights, prefix sampling, and tuning budgets common. Compare the following acquisition learners:

1. Direct gain regression, using the fixed-predictor adaptation of DIME's objective on observed before/after loss differences (Gadgil et al., 2024). This mechanism control is distinct from the original joint-training DIME comparator described below.
2. Ordinary feature transfer followed by the same direct gain regression. Comparison against this condition isolates the benefit of adding a privileged value target to feature supervision already supplied to the acquisition representation.
3. Value distillation from a teacher receiving only the student's prefix and the candidate action, retaining the feature-supervised student. Match its capacity, teacher-fitting participants, cross-fitting, availability gate, and blending procedure to the rich teacher.
4. Privileged value distillation with current-state descriptors, retaining the same student as in the feature-transfer baseline.
5. Privileged value distillation with the two paired endpoint descriptors and their elapsed time.
6. Privileged value distillation with the corresponding current-state and change coordinates, with the same observations and elapsed time supplied.

The prefix-only value teacher is essential because target smoothing and extra computation could help without privilege. Both teachers fit the same valid training participants and supervise the same $I = 1$ examples. The prefix-only teacher therefore estimates $\mathbb{E}[d \mid h, a, I = 1]$, rather than the unconditional-population gain $\mathbb{E}[d \mid h, a]$. The unchanged raw targets on $I = 0$ preserve the common

population in the ideal conditional-mean case. A separate prefix-only teacher fitted and applied to all episodes changes both the training data and supervision coverage, so it is a secondary comparison. The blended squared-loss target follows privileged-feature distillation (Yang et al., 2022), with related information asymmetry in privileged critics (Ebi et al., 2026; Venkatraman et al., 2026). The question here is whether the additional observations help a finite teacher estimate this fixed predictor's acquisition gains.

For the temporal comparisons, first hold the actual dated observations and their applicability gate common. Compare current-only descriptors with descriptors from two distinct observed visits, rather than a change vector inferred from a single scan. Restrict both conditions' teacher supervision to the same valid-pair examples, retain raw gain targets elsewhere, and provide the same timing and availability metadata. Match encoder families, output dimensions, parameter budgets, and training effort. This comparison tests the additional earlier observation without confounding it with more teacher training examples.

Next compare paired endpoints with current/change coordinates using the same modality subset M and the same elapsed time. Put $u_j = F(W_j^M)$ for $j \in \{0, 1\}$, and let Q collect any shared additional covariates, including a richer current-state encoding with more modalities. The two teacher inputs are

$$Z_{\text{end}} = (u_0, u_1, \Delta\tau, Q), \qquad Z_{\text{change}} = (u_1, (u_1 - u_0)/\Delta\tau, \Delta\tau, Q). \tag{80}$$

Since $\Delta\tau > 0$, these inputs contain the same information. In particular, the current/change arm retains the pair-matched endpoint u_1 even when Q contains a richer current encoding. That richer encoding does not generally determine u_1, so it cannot replace u_1 in an equal-information comparison. A difference between the specified arms concerns finite models and optimization, not information created by subtraction. Longitudinal representation and history-distillation methods provide relevant precedents (Ouyang et al., 2022; Karimian et al., 2026).

Fusion requires a separate comparison against strong single-modality teachers. For each temporal condition, compare MRI-only, PET-only, and biomarker-only descriptors with their fusion on the same training examples, using a common applicability gate and matched capacity and tuning. Select the strongest single-modality comparator on validation data. Include simple concatenation followed by an MLP and an attention-fusion control with the same measured inputs when assessing a particular fusion architecture. All acquisition students retain the same clinical-only deployment inputs and raw targets outside the common gate. A secondary comparison may use every available example for each teacher, but then any gain combines modality information with supervision coverage. It does not isolate fusion.

A matched block-weighted concatenation of the segment errors is algebraically the same feature loss as the separate terms. Verify agreement of their losses and gradients as an implementation check. This verifies equivalence rather than establishing a new structured objective. To test patient-specific temporal supervision instead, compare true pairs with a prespecified within-stratum pairing perturbation on the same supported examples. Preserve the observation cutoff, modality masks, elapsed-time distribution, and sample weights. This control tests dependence on correct temporal correspondence, without assigning a causal biological meaning to the learned displacement.

An ideal teacher fitted and applied to all prefixes could return the student's conditional mean gain without privileged inputs. Within the matched $I = 1$ subset, the corresponding target is its subset-conditional mean. More privileged information does not monotonically reduce variation in ideal teacher targets within that subset. The empirical case for richer inputs must therefore rest on finite-sample learnability or improved estimated supervision. Vary the number of value-training participants, availability of valid pairs, teacher capacity, and blend weight to identify both benefit and negative transfer. Thin privileged supervision within the training data while keeping the evaluation population fixed. Naturally different availability groups require separate population descriptions.

F.6 Acquisition alternatives and complete systems

Learned static orders and uniform random legal selection provide simple references. A complete-battery predictor gives a maximum-input comparison. Historical-administration imitation, when compatible labels exist, predicts recorded clinical behavior rather than the diagnostic value of a test. Stronger acquisition alternatives include nongreedy and explanation-based methods with explicit adaptations to the same actions, inputs, costs, and recorded endpoint (Valancius et al., 2024; Guney et al., 2025).

NOCTA and REACT address acquisitions across future visits or feature-time plans (Dinh et al., 2026; Qu et al., 2026). Their original problem differs from choosing cognitive tests within the present visit. Any within-visit adaptation must state which planning capability is retained, use the same fixed q when isolating selection, and receive comparable tuning. Published results from their original longitudinal tasks are not numerical baselines for this assessment protocol.

After isolating acquisition, compare complete systems with their own fitted diagnosis and selection components. Gains in that comparison can arise from prediction, selection, or both, so interpret them alongside the fixed-q contrasts.

The full-system DIME comparator retains its original classification procedure: random-mask predictor pretraining, joint predictor/value training with exploratory policy rollouts, cross-entropy loss-reduction targets, and its published value parameterization and stopping rules (Gadgil et al., 2024). Adapt feature groups to the action bundles and enforce their prerequisites and the common hard cap. Supply the permitted baseline history to both networks, without privileged deployment inputs. Make the same development participant pool available across D_{fit} and D_{value}, with joint-training rollouts restricted to the complete-library episodes. If naturally incomplete episodes contribute to predictor pretraining, give both systems the same observed-prefix access and never fabricate missing gain labels. Retain common validation and test participants and comparable capacity and tuning budgets. Report actual training resources and the partition used for each fitted component. Because its predictor is updated during training, this comparator does not share the frozen q or gain labels used in the mechanism comparison. Evaluate all systems with the same held-out Brier loss, accuracy, and realized action count. Select each method's stopping settings on validation data and compare matched budgets or loss-versus-count curves. Equal numerical penalties need not represent equal operating points across different training losses.

These comparisons evaluate computational selection and prediction. Clinician override, referral, and the human review options in the overview are outside the evaluated policies.

F.7 Deployment inputs and action-target diagnostics

The primary deployment condition is clinical-only, with $V_v = \varnothing$. Its baseline B_v retains permitted dated clinical and cognitive history, but excludes patient-specific MRI, PET, and biomarker values and all their derived features or caches. Its execution also omits feature-teacher targets and value-teacher outputs. Merely zeroing an imaging input tensor does not establish this restriction. Compare this condition with explicitly defined modality-subset deployment, which supplies its allowed modalities through V_v, on a common evaluable participant set. Naturally different modality-availability groups are separate population comparisons. Artificially hiding an observed modality tests controlled input removal and does not reproduce every mechanism of naturally missing imaging.

For the fixed-q mechanism comparison, the evaluator can enumerate each legal one-step reveal on complete held-out episodes and calculate

$$d_{vka} = \ell(q(h_{vk}), Y_v) - \ell(q(h_{vk} \oplus (a, X_{va})), Y_v). \tag{81}$$

Use one prespecified, nonanticipating prefix sampler for all fitted policies and teachers. It may depend on already revealed results and protocol metadata, but cannot inspect hidden values or the reference label when generating a history. Fix the remaining budget as well as the prefix so every method faces the same legal set. Average over sampled prefixes within each episode, then use the participant and visit weights in Equation (78). This common-history comparison precedes on-policy evaluation, where different methods can visit different histories.

Contrast and offset errors. For a prefix h with $J \geq 2$ legal tests, form the realized-gain vector $\mathbf{D}_h$, predictions $\mathbf{g}_h$, and residual $\mathbf{e}_h = \mathbf{g}_h - \mathbf{D}_h$. Let $\bar{e}_h = \mathbf{1}^\top \mathbf{e}_h / J$ and $\mathbf{C}_h = \mathbf{I}_J - \mathbf{1}\mathbf{1}^\top / J$. Report the two normalized components

$$E_{\mathrm{offset}}(h) = \bar{e}_h^2, \qquad E_{\mathrm{contrast}}(h) = \frac{1}{J} \|\mathbf{C}_h \mathbf{e}_h\|_2^2. \tag{82}$$

Their sum equals per-action gain MSE by Equation (12). Mean signed error is an additional bias summary, not a substitute for squared offset error, since positive and negative offsets can cancel. A shared correction can improve total gain MSE while leaving every ranking among tests unchanged. The centered component tests whether relative test values improve. Appendix D.5 shows why its

expected difference between fixed models on a common history law removes a shared outcome-noise term. Individual realized contrasts still do not reveal a patient's conditional expected gains. Report singleton legal sets separately, since their contrast error is identically zero. Empty sets contribute only to termination outcomes.

Test selection with stopping disabled. On each common prefix with at least one legal test, record the preferred test $\widehat{a}(h) = \arg\max_a\{g(h,a) - \lambda c_a\}$, using the common tie rule even when the deployed policy would stop. Report between-policy test agreement and the realized net gain $d_{h,\widehat{a}(h)} - \lambda c_{\widehat{a}(h)}$ under that forced one-step selection. This separates choosing among tests from deciding whether any test is worth its cost. The selected action uses only the permitted prefix. Consequently its mean realized net gain estimates the corresponding mean conditional utility under the fixed history law, as in Equation (59). A hindsight maximizer of realized gains uses hidden results and the label, so agreement with that maximizer is only an information-rich diagnostic, not agreement with the optimal deployable action.

STOP and the complete decision. For the same prefixes, compare $\max_a\{g(h,a) - \lambda c_a\}$ with zero and report stopping frequency, pairwise stop/continue disagreement, and realized net gain of the complete one-step rule. Assign gain and cost zero to STOP. Break down these summaries by prefix size, budget, class, and teacher availability. Also report the preferred test's realized net gain within stop and continue groups. These group summaries can expose aggregate overvaluation or undervaluation, but a negative realized gain on one record does not make its acquisition decision conditionally wrong. Fixed bins of the predicted best-test net gain can compare predicted with average realized selected-test net gain, with bin boundaries chosen on validation data. Neither these bins nor action agreement certify individual optimality or clinical safety.

Apply the same offset and contrast diagnostics to the rich and prefix-only teachers on common held-out rich-input examples. Any teacher ensemble is fixed using development data and used only for evaluation, never deployment. Follow the common-prefix analysis with complete policy trajectories to assess terminal loss and cost. Exact synthetic models may additionally test true conditional action rankings, which are known there but not observed directly in the clinical records.

F.8 Outcomes, uncertainty, and interpretation

Report participant-weighted normalized Brier loss and top-one accuracy against realized action count across prespecified caps. Include per-class recall, macro classification summaries, and sample support so majority-class performance cannot conceal subtype failures. For utility stopping, also report terminal risk, mean actions, and the fractions ending through nonpositive utility, budget exhaustion, or an empty legal set. A budget stop is not a confidence certificate. Referral or clinician override requires a separately defined combined policy and is outside this computational evaluation.

Use the same test participants for paired contrasts. Resample whole participants with all eligible visits and recompute Equation (78) within each bootstrap sample. Keep model-seed variation distinct from participant sampling uncertainty. Designate the primary contrast and budget before test inspection. Multiple-budget or multiple-class claims require their own joint inferential treatment or an explicitly descriptive interpretation. Selecting the most favorable test-set operating point is not matched-budget evaluation.

The central hypothesis requires an advantage of privileged value supervision over direct gain regression, feature transfer alone, and a matched prefix-only value teacher. A temporal claim additionally requires the current-only and equal-information endpoint comparisons. Better full-input classification supports a different claim. Report unresolved contrasts as unresolved, and do not use prediction preservation imposed by a removal rule as independent non-inferiority evidence. A gain confined to the common offset supports a different explanation from improved test discrimination. A lower contrast error without higher selected-test value does not establish better decisions. The final evidence for acquisition performance is the supported loss-versus-cost comparison of the complete policies.