# ADAPTIVE SENSING VERSUS EVERY FIXED SCHEDULE: HOW MUCH IS ADAPTIVITY WORTH?

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Hospital staff cannot watch a patient continuously, so they measure on a schedule fixed in advance — before meals, or every four to six hours. Any fixed schedule misses events between checks, and the obvious fix is to decide each measurement from what is known at the time. But the two options are not equally tractable. A fixed schedule is a handful of clock times: a six-hour night at fifteen-minute resolution has twenty-four slots, so at most eight measurements gives $\sum_{1\le k\le 8}\binom{24}{k} = 1{,}271{,}625$ schedules and the best is found by checking every one. A policy that decides as it goes must act at every state it can reach — $(24 \text{ slots}) \times (61 \text{ belief means}) \times (25 \text{ staleness levels}) = 36{,}600$ on our grid — giving $2^{36{,}600}$ policies. **One side can be settled exactly and the other cannot** — which is why a reported gain over a fixed schedule can always be blamed on the chosen baseline. We remove that objection: a constrained partially observed Markov decision process makes the adaptive side tractable, and enumeration makes the fixed side an exact empirical optimum within its class.

**What we measure.** On 112 hospitalized patients whose blinded sensor recorded every slot, we compare a solved adaptive policy against the best of all $1{,}271{,}625$ fixed schedules at six pre-specified measurement rates. The adaptive policy detects more nocturnal hypoglycemia at every rate, by $14.2$ percentage points on average. **What we do not claim** is its size: a bootstrap that reruns selection gives $[-2.0, +39.8]$, positive in 96 percent of replicates but wide, because 93 held-out excursions carry the information of some 18 independent events. The advantage also mixes *when* to measure within a night with spending more on some nights than others.

**What generalizes beyond this dataset.** *An evaluation protocol for selected policies*: selection is rerun inside every bootstrap replicate and splits are taken on unique identifiers, the usual alternative silently mixing about one unit in six however large the sample. *A comparator that scales*: detection coverage is monotone submodular, so a greedy schedule attains $1 - 1/e$ of the exhaustive optimum wherever enumeration is infeasible, and 94 to 100 percent of it here. *Conditions under which adaptivity is worth nothing*: provably zero when a measurement carries no reward-relevant information, at most $2k\varepsilon$ when that holds to within $\varepsilon$. Zero is the right thing to characterize, because correlated sensing admits no universal constant bound on what adaptivity is worth.

## 1 INTRODUCTION

A monitoring policy decides when to spend a costly measurement on a state that is otherwise unobserved. The question is old and the machinery is well developed: structural results for partially observed Markov decision processes establish when the optimal policy is a threshold on the belief (Lovejoy, 1987; Krishnamurthy, 2025), and applied work in operations has produced closed-form monitoring schedules for progressive disease (Helm et al., 2015; Kazemian et al., 2019). Yet these results say little about *when timing is worth optimizing at all*, and in practice that is the first question a designer faces.

The difficulty is that the two natural policy classes are separated by an enormous gap in tractability. A fixed schedule of at most eight measurements over 24 slots is one of $\sum_{1\le k\le 8}\binom{24}{k} = 1{,}271{,}625$



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.

candidates, the budget cap keeping that well below the $2^{24}$ of all subsets; an adaptive policy assigns a binary action to each of the solver's $36{,}600$ grid cells. Enumeration is available for one class and unimaginable for the other.

This gap makes the value of adaptivity hard to establish rather than hard to assert: a reported gain over a fixed schedule can always be blamed on a poorly chosen baseline, and four successive comparisons in our own development failed for that reason. The nocturnal window is the largest instance we have found where the fixed-schedule class can be optimized *exhaustively*, so the objection does not arise: the comparison is against the best schedule the training sample admits. That optimum is empirical, not population-level — Appendix F.10 shows the selected schedule changes across folds — and we write *exact empirical optimum* throughout to keep the distinction.

We therefore proceed in two registers. Where the fixed class can be optimized exhaustively we benchmark against an exact empirical comparator; where it cannot — at 5 min resolution the same count exceeds $10^{10}$, and over a week-long stay $10^{22}$ — we rely on the solver alone. We know of no diagnostic that certifies in advance whether adaptivity will pay, and Proposition 1 shows why a natural candidate cannot.

**Setting.** We study inpatient glycemic monitoring, where nurses measure capillary glucose on a fixed schedule — typically before meals, or every four to six hours for patients who are not eating (American Diabetes Association Professional Practice Committee, 2026). Nocturnal events are frequently asymptomatic and carry outsized clinical and cost consequences (Edelman & Blose, 2014), and establishing what a schedule misses requires knowing glucose at the moments nobody measured. We use a public cohort of 112 hospitalized patients in which a *blinded* continuous sensor recorded every slot, including every slot the protocol skipped (Zhao et al., 2023). The sensor therefore retrospectively reveals glucose at slots a candidate schedule would have sampled but the ward did not, so detection can be scored without modeling the unmeasured trajectory. It does not reveal what would have happened under a different *treatment* history, and no clinical outcome is claimed. That is a narrower guarantee than a true counterfactual, and it is what distinguishes this benchmark from screening and surveillance studies that must calibrate a natural-history model (Ayer et al., 2012; Sandıkçı et al., 2013). Against it we can ask, for the first time on real inpatient data, what any schedule would have caught. The comparator is not a strawman: a professional society's schedule implemented verbatim does no better than this ward's own practice (Appendix E.2).

**Positioning.** The closest operations work requires a *progressing transformation*: the state must drift monotonically toward an absorbing bad state (Helm et al., 2015). Our kernel violates that condition from the hypoglycemic state, where the probability of remaining hypoglycemic decays from 0.84 to 0.40 within six hours. Glycemic excursions recur; glaucoma does not. The recurrent case has not been given structural results, and it is the case in which we find that rearranging a fixed schedule carries almost no value.

Appendix H places this against inspection scheduling, event-triggered sensing, quickest detection and the clinical literature.

**Scope.** This is a benchmark and an evaluation methodology, not a new planner. The underlying question is one machine learning asks constantly and rarely settles: when a learned policy beats a designed one, how much of the gap is the learning and how much is the chosen baseline?

**Contributions.**

1. **A benchmark whose baseline is exhaustively optimized.** Adaptivity claims are usually made against a baseline someone chose, so a gain is confounded with that choice. Removing it needs a fixed class small enough to enumerate yet worth optimizing, and that window is narrow: $1.3$ million members at $24$ slots, over $10^{10}$ at $5$-minute resolution, $10^{22}$ across a week. The nocturnal window sits inside that window, and to our knowledge it is the largest clinically meaningful setting that does. Where it is not, the comparator still is: Proposition 2 shows coverage is monotone submodular for arbitrary onset-duration mass, so greedy attains $1 - 1/e$ of the optimum, and $94$ to $100$ percent of it here. Within that benchmark we also measure what our own approximation costs — the reduced belief is compared against an exact joint filter rather than assumed harmless (Appendix G.2). Section 4.4 reports what enumeration costs and where it stops being feasible,

which is the information a reader needs to know whether their own problem admits the same treatment.

2. **An evaluation protocol for selected policies.** Comparing a solved policy against a selected baseline invites two errors that are easy to make and hard to see. We resample patients and rerun the *entire* pipeline inside each replicate, so selection and calibration variability enter the interval; and we split on *unique* identifiers, since a bootstrap draw's repeats would otherwise straddle the split (Proposition 8). The estimand is an integrated frontier gap defined before any comparison was run (Equation (5)). The protocol is independent of this cohort and of glucose.
3. **Both directions of a separation between diagnostics and schedule value.** Proposition 1 exhibits equal-budget schedules with identical one-step advantage at every reachable belief and coverage differing by one. Bounding how much the value of measuring varies across beliefs therefore says nothing about how much of the timeline a schedule covers, which rules out a natural family of shortcuts. Proposition 3 gives the converse: when measurement does not change the conditional hazard, adaptivity is provably worth nothing.

## 2 SETTING AND NOTATION

Time is indexed by $t$ on a fixed grid. A latent state $s_t \in \mathcal{S}$ evolves with transition kernel $P$; an action $a_t \in \mathcal{A} = \{0, 1\}$ decides whether to spend a measurement; an observation $o_t \in \mathcal{O}$ is returned only when $a_t = 1$. The belief is $b_t \in \Delta(\mathcal{S})$ and a policy $\phi$ maps beliefs to actions. We write $Z(a)$ for the observation kernel under action $a$, which is action-dependent by construction: this is a controlled-sensing problem, not a control problem.

**Assumption 1** (Acquisition exogeneity)**.** $P(0) = P(1) = P$*: measuring does not move the state.*

For a fingerstick this is a fact rather than a modeling choice. What a measurement can change is the treatment that follows, and that pathway is outside our estimand; we return to the consequences in Section 5. Assumption 1 also discharges the copositive-dominance condition of Krishnamurthy (2025)'s Theorem 15.8 with equality, which we use in Section 4.

**Belief update.** Since observations arrive only under $a = 1$, the filter is action-dependent:

$$b_{t+1} \;=\; F(b_t, o_t, a_t) \;=\; \begin{cases} \dfrac{\operatorname{diag}\big(Z(1)_{\cdot o_t}\big)\, P^\top b_t}{\mathbf{1}^\top \operatorname{diag}\big(Z(1)_{\cdot o_t}\big)\, P^\top b_t} & a_t = 1, \\ P^\top b_t & a_t = 0, \end{cases} \tag{1}$$

so a measurement contracts the admissible posterior set according to the observation likelihood, while waiting propagates the belief through $P^\top$ toward the stationary distribution $\pi$. That asymmetry is the whole content of Section 3.

We state the problem as reward maximization throughout, so the reward machine's $+1$ on detection and the measurement price carry consistent signs. The exact object is a belief on the joint space. With $\beta \in \Delta(\mathcal{S} \times \mathcal{U})$, horizon $H$ and price $\kappa \geq 0$ per measurement,

$$V_h^\kappa(\beta) \;=\; \max_{a \in \{0,1\}} \Big\{ r(\beta, a) - \kappa a \;+\; \gamma \sum_{o \in \mathcal{O}} p(o \mid \beta, a)\, V_{h+1}^\kappa\big(\Phi(\beta, o, a)\big) \Big\}, \tag{2}$$

with $\Phi$ the exact joint filter, $V_H^\kappa \equiv 0$ and $\gamma = 1$ throughout: the horizon is one night and nothing is discounted. For each fixed $H$ and $\kappa$ the map $\beta \mapsto V_h^\kappa(\beta)$ is piecewise linear and convex on $\Delta(\mathcal{S} \times \mathcal{U})$, being a maximum of finitely many affine functions.

**This is a Lagrangian relaxation, not a hard budget.** Pricing measurements targets an *expected* rate under the calibration distribution; it does not enforce a count pathwise, and sweeping $\kappa$ is not in general equivalent to sweeping an integer budget, since a non-convex attainable set admits a duality gap and deterministic policies can skip resource levels. What we solve and evaluate is Equation (2): Section 4.4 sweeps $\kappa$, records the realized rate, and compares classes at common rates. No experiment in this paper enforces a hard budget.

We solve Equation (2) on reduced coordinates $\widetilde{\beta} = (\hat{\mu}, g)$; Appendix G.2 measures what the projection costs and Appendix F.8 reports grid convergence, the convexity above being a property of $V^\kappa$ on

the exact simplex that we do not transfer. Writing $\widetilde{Q}_h(\widetilde{\beta}, a)$ for the bracketed term evaluated on the reduced state,

$$A_h(\widetilde{\beta}) \;=\; \widetilde{Q}_h(\widetilde{\beta}, 1) - \widetilde{Q}_h(\widetilde{\beta}, 0) \tag{3}$$

for the advantage of measuring. Proposition 5 bounds the variation of the *myopic* one-step advantage across the beliefs the process reaches, not of Equation (3) itself (Appendix C.4).

**Reward and estimand.** The per-step reward is $r(s, 1) = \mathbb{1}\{s \in \mathcal{E}\}$ on the transition that first detects an excursion and zero otherwise, and each measurement is priced at $\kappa(t)$, higher at night, so the per-step objective is $r(s, a) - \kappa(t)\, a$ with $\mathcal{E}$ the excursion states. An excursion is a maximal run of at least two consecutive slots below $70\,\mathrm{mg\,dL^{-1}}$, so at 15 min resolution the shortest event we count lasts half an hour; the threshold follows Battelino et al. (2019) and the two-slot minimum excludes single-sample dropouts, which are common in continuous sensing. For reporting we use the induced detection count: writing $Q_\phi$ for the slots at which $\phi$ measures and $E_j$ for the interval of the $j$-th excursion,

$$V(\phi) \;=\; \mathbb{E}_\tau\Big[\big|\{j : Q_\phi \cap E_j \neq \emptyset\}\big| - \sum\nolimits_{t \in Q_\phi} \kappa(t)\Big]. \tag{4}$$

This is detection on the observed trajectory. Under Assumption 1, and provided the revealed trajectory is the one the evaluated policy would have faced and no policy induces treatment, replay identifies this functional exactly: the estimand is a deterministic function of the recorded trajectory and the schedule, so no distributional argument is needed. That identification is what licenses replay; the agreement we observe between replayed and on-policy values across four schedules and sixteen trajectories each ($|z| \leq 1.46$) is an implementation check on the code, not evidence for the identification.

The classical inspection literature assumes this away: Barlow et al. (1963) require that on detection "the problem ends; specifically, no replacement or repair takes place". Clinically that assumption fails, so we price it.

**Non-Markov reward.** Equation (4) pays once per excursion, so whether the current one has been paid for depends on the action history. A three-state reward machine (Toro Icarte et al., 2022) restores the Markov property on the product of belief and automaton state (Appendix F.12). The automaton is driven by the latent state, so the exact belief lives on $\Delta(\mathcal{S} \times \mathcal{U})$ while our solver factorises it; that factorisation is exact only under a conditional independence which fails, and Appendix G.2 measures its cost at $0.018$ to $0.032$.

**A parametric belief.** We carry the belief as a mean and a variance rather than gridding $\Delta(\mathcal{S})$, standard in linear-quadratic sensor scheduling (Shi et al., 2012). Fitted on our cohort the process is close to first-order autoregressive, $x_{t+1} = a x_t + b + \epsilon_t$ with $a = r = 0.9839$. The residuals are not Gaussian: excess kurtosis $4.03$, and a standard deviation rising from $6.8$ to $12.4$ across glucose quintiles. Appendix F.9 tests whether this matters by generating traces from Student-$t$, heteroscedastic and nonparametric-bootstrap innovations and solving under the unchanged Gaussian belief: the sign of the advantage is stable across all four processes, while its magnitude ranges from $+3.8$ to $+23.0$ points, with the Gaussian case near the middle. Appendix G.4 measures the resulting miscalibration directly: the stated lower-tail probabilities overstate risk about fourfold in the middle range and understate it above $0.35$, which shifts a threshold rule without reordering states. The variance then follows a deterministic recursion given the actions, so it is a function of time since the last measurement rather than a free coordinate, and the belief reduces to a mean and an integer staleness. Its dimension is then two whatever the resolution of the underlying state.

American Diabetes Association Professional Practice Committee (2026) specify inpatient monitoring "before meals" or "every 4 h to 6 h" when the patient is not eating, which constrains the interval rather than the count. The observed protocol behaves accordingly: measurements per patient-day range from 1 to 12 with mean $4.87$ and a median gap of 3.0 h, showing none of the concentration a quota would produce. We therefore price each measurement at $\kappa$ and sweep it, which traces the same frontier a quota would while leaving the horizon free of a terminal condition. This is the economic design of Duncan (1956), whose decision variable is the sampling interval and whose three costs — sampling, false alarm, and time spent out of control undetected — correspond to measurement burden, alarm fatigue, and an undetected excursion.

# 3 WHY LOCAL BELIEF DIAGNOSTICS DO NOT CERTIFY SCHEDULE VALUE

## 3.1 A DIAGNOSTIC THAT DOES NOT DECIDE

A natural shortcut asks whether the belief forgets faster than a decision-maker can act. We report it as a diagnostic, not a criterion: the same cohort gives $2.25\,\mathrm{h}$ or $10.68\,\mathrm{h}$ depending on the discretization (Appendix F.11). The structural reason is stronger:

**Proposition 1** (A constant one-step advantage does not bound coverage)**.** *There exist a process with a time-homogeneous latent belief and a periodic onset-risk profile, and two schedules of equal budget, on which the one-step advantage of measuring is identical at every reachable belief while the fraction of excursions covered differs by one.*

*Proof.* Take $n = 12$ slots on a circle, budget $k = 2$, and deterministic excursion duration $D = 3$, with all onset mass split evenly between slots $0$ and $1$. The latent belief evolves time-homogeneously, so the one-step advantage depends on staleness and not on the slot index. Both schedules below are evenly spaced with cyclic gap $6$, so they induce the same staleness multiset and hence identical advantage multisets; equal budget alone would not suffice. The schedule $\{2, 8\}$ covers slots $\{0, 1, 2\}$ and $\{6, 7, 8\}$, hence both onsets, giving coverage $1$. The schedule $\{5, 11\}$ covers $\{3, 4, 5\}$ and $\{9, 10, 11\}$, hence neither, giving coverage $0$. □

Let $\mathcal{D}$ be the permutation-invariant functions of the advantage multiset — every shortcut we know of. The two schedules have identical multisets, so every $d \in \mathcal{D}$ values them equally while their coverage differs by one; Appendix H.9 discusses the stronger classes this leaves open. A synthetic sweep confirms that contraction does not predict adaptive gain (Appendix D.1); Appendix G.5 gives a second instance with a $49$-point gap on empirical onsets.

## 3.2 WHAT DOES GOVERN SCHEDULE VALUE

Coverage depends on where the slots fall relative to where excursions begin, and Proposition 7 turns that into a bound computable from an onset histogram alone (Appendices F.1 and G.6).

These results are elementary by design: they license the benchmark rather than stand alone. We did try the formal route — the approximate information state bound of Subramanian et al. (2022) — and report in Appendix G.3 why it holds vacuously here. Proposition 1 rules out a natural shortcut, and Proposition 7 gives a screening quantity computable from an onset histogram. Neither is deep and we do not present them as such.

*Proposition 8, in Appendix H.3*, quantifies the leakage a naive resampling split induces: about one unit in six at an even split, a fraction that does not diminish with $n$.

**Proposition 2** (Coverage is monotone submodular)**.** *Let $q(t, d)$ be the mass of excursions beginning at slot $t$ and lasting $d + 1$ slots, and let a schedule $S$ cover such an excursion if it meets the window $W(t, d) = \{t, \ldots, t + d\}$ on the circle. Then*

$$C(S) \;=\; \sum_{t,d} q(t, d)\, \mathbf{1}\{S \cap W(t, d) \neq \varnothing\}$$

*is monotone and submodular for* arbitrary *$q$, so the greedy schedule attains at least $(1 - 1/e)$ of $\max_{|S| \le k} C(S)$.*

*Proof in Appendix H.4.* $C$ is a weighted coverage function — each $(t, d)$ is an element of weight $q(t, d)$ covered by $W(t, d)$ — so monotonicity and submodularity are standard and the greedy guarantee follows (Nemhauser et al., 1978).

**Enumeration does not scale but the comparator does.** None of the assumptions of Proposition 7 are needed: no fixed duration, no divisibility, no restriction on $D$. Where the fixed class is too large to enumerate, greedy supplies a baseline with a worst-case guarantee, and in twelve instances small enough to check it matched the exact optimum every time.

**Proposition 3** (Zero value of adaptivity under clock-time hazard)**.** *Suppose that for every reward-relevant future event, the conditional law given any measurement history equals its unconditional law, so observations carry no information the reward can use; suppose acquisition is exogenous*

*(Assumption 1) and the budget of $k$ measurements binds on every path. Then*

$$\Delta_k \;=\; \sup_{\phi\in\Pi_{\text{adaptive}},\,|Q_\phi|\le k} V(\phi) \;-\; \sup_{|S|\le k} V(S) \;=\; 0.$$

*Proof in Appendix H.5.* No realisation makes one continuation preferable, so conditioning on its randomization writes any such policy as a mixture of deterministic schedules of at most $k$ slots, and a mixture cannot beat its best component. The pathwise budget is needed: expected coverage is not linear in the marginal measurement probabilities, so no polytope-vertex argument is available under an expectation constraint. □

**$\Delta_k$ is an adaptivity gap, and correlation is what makes ours hard.** For *independent* activation, Asadpour & Nazerzadeh (2016) show the adaptive-to-non-adaptive ratio is exactly $e/(e-1)$, and Gupta et al. (2017) extend constant gaps to prefix-closed constraints. Our objective is submodular and $|S|\le k$ is a uniform matroid, but one excursion spans several slots, so activation is correlated and those results do not apply. Gupta et al. (2026) treat exactly this case and show the gap is then $\Theta(d_{\max}S)$ rather than constant, for $d_{\max}$ the correlation a single latent event induces and $S$ the probing budget. **The value of adaptivity is therefore not bounded by a universal constant here**, which is why we characterize when it is zero rather than assume it is small (Appendix H.11).

The conclusion degrades gracefully: Proposition 9 bounds $\Delta_k$ by $2k\varepsilon$ when the conditional laws hold only to within $\varepsilon$, proved by a simulation lemma that needs no separability. Across 300 exactly solved instances the worst case used a fifth of it.

Two further conditions are necessary. A *reversal* — two *schedules*, not two slots, ranked differently by two distinguishable types — is needed, since without one the same schedule is optimal whatever is observed and $\Delta_k = 0$ (Appendix H.5). And $k \ge 2$ is needed: identification consumes a measurement, so with a single one there is nothing left to spend on what was learned. In 200 instances at $k = 1$ the gap was zero in every one. **Adaptivity pays only when the posterior is heterogeneous, observations resolve it, and the resolution changes which slots are worth using.** Exact solutions give a positive gap in 33 of 40 instances when all three hold and zero in all 40 when any is removed, and in 200 instances with a common ordering none had a positive gap (Appendix H.5). We establish necessity; sufficiency and the size of the gap are open. Appendix H.8 turns this into a two-axis picture: onset concentration governs what placement buys, type heterogeneity governs what adaptivity buys, and a factorial sweep shows neither predicts the other.

## 4 Experiments

We report three things about the fixed class before turning to the comparison that matters: that *placement* within a fixed schedule buys little, with detection close to the coverage null of Equation (6); that a belief on a non-mixing quantity behaves differently; and that the structural conditions we invoke hold on the data. Section 4.4 then asks the question this paper is about — whether *leaving* the fixed-schedule class pays — against a comparator computed by enumeration.

### 4.1 Data and protocol

The cohort is a public release of 112 hospitalized patients, 125 patient-periods and 1,332 patient-days at 15-minute resolution (Zhao et al., 2023), obtained from the depositors' Figshare record and not redistributed by us. Capillary measurements follow a real seven-point clinical protocol, so the behavior policy is not one we designed, and a *blinded* continuous sensor recorded every slot. There are 519 hypoglycemic excursions below 70 mg dL$^{-1}$ and 686 severe hyperglycemic excursions above 250.

Every fitted object is fitted inside the fold, under the dual split of Sergazinov et al. (2024). Every proportion carries a 95% BCa interval resampled over *patients* and is reported as a compatibility interval; comparisons are described by whether intervals overlap, with no verdict attached in either direction (Amrhein et al., 2019; Greenland et al., 2016). Excursions cluster heavily: the intra-class correlation is 0.281 and the design effect 5.27, so 519 excursions behave like 98 independent ones and the minimum detectable difference at 80% power is about 13 percentage points, declared before any policy comparison was run.

A policy is admitted only if its actions, and its alerts where it has them, are unchanged when the sensor value at every unmeasured slot is replaced by noise (Appendix B.1).

### 4.2 Rearranging a fixed schedule buys little

At a fixed budget, moving measurements within the day is worth $+2.31$ points $[-1.76, +7.31]$ by paired comparison, and detection tracks the lattice-coverage null within $0.95$–$1.10$ over a sixfold budget range. Two thresholds trade one tail against the other rather than creating detection. Appendix E.2 gives the analysis, including the exact frontier over all $\binom{96}{3}$ schedules.

### 4.3 Structural conditions

Every premise of the structural results we invoke holds on the data: the transition kernel is TP2 at every lag, the observation kernel is TP2 and full rank, and the reward is submodular in the relevant ordering. Appendix D.2 reports the tests. Verifying premises licenses the machinery; it is not itself a result.

### 4.4 Does adaptivity pay? A test where the baseline is exact

Every comparison above pits a policy against a lattice or a hand-designed rule, so a reported gain can always be attributed to a weak baseline. The nocturnal window removes that objection. Six hours from 23 h at 15 min resolution leaves 24 slots. The fixed class has $\sum_{1\le k\le 8}\binom{24}{k} = 1{,}271{,}625$ members; the alarm rates we compare at, $1.5$ to $4.0$ per night, are reached by $k \le 4$, so the frontier sweep enumerates the $12{,}950$ schedules with $k \le 4$ and every one of them exactly. The optimal fixed schedule follows by *enumeration*, so the baseline is the exact *empirical* optimum of its class on the training nights.

**Protocol.** Patients are split in half; the autoregressive kernel is fitted on the training half alone, giving $x_{t+1} = 0.9839\, x_t + 2.36$ with residual standard deviation $9.86$. The fixed schedule is chosen by enumeration on training nights and the per-measurement cost of the adaptive policy is bisected on training nights to match the alarm count; both are then scored on 646 held-out nights containing 93 hypoglycemic excursions, by the same coverage function.

A per-alarm cost bisected on training nights does not transfer pathwise: realized held-out rates are $1.99$, $2.19$, $1.71$, $3.13$ and $2.44$ against a target of two. A multiplier fitted on one sample enforces its constraint in expectation on that sample, not on another. At the one split landing on target the solved policy detects $64.5\%$ against the exhaustive optimum's $55.9\%$, a gain of $8.6$ points with an interval of $[-13.3, +32.0]$. Across five splits the mean is $+16.1$ (sd $6.9$), but four used more alarms than the target and the matched split was identified after seeing the calibration. We therefore report the frontier comparison below instead.

**What the large cohort contributes.** The 180,798 excursions from 1,720 subjects are not a second evaluation set — they carry no measurement protocol — but they supply the inputs Proposition 7 needs, and those bound the quantity the nocturnal experiment does *not* measure. Restricted to 0 h to 6 h, DiaData's onset excess mass is $0.0132$ with a mean duration of $1.42$ h, and across the eight cohorts with at least 2,000 excursions it ranges from $0.026$ to $0.054$. By Proposition 7 that caps what any rearrangement of a fixed nocturnal schedule can gain over the best lattice at roughly one point. The adaptive frontier gap we measure at the same *expected* rates — pricing controls the realized rate in expectation, not pathwise — is $+9.7$ to $+15.4$ points. Relative to the nocturnal bound that is a factor of 7 to 12; relative to the largest single-cohort bound, $5.4$ points, it is $1.8$ to $2.9$. The two quantities are different estimands — one caps rearrangement within the fixed class, the other measures leaving it — so the ratio is a magnitude comparison rather than a test, and we report its range rather than rounding it to an order of magnitude. The large cohort therefore explains why the fixed class is nearly exhausted by a lattice, which is why a procedure that conditions on observations has room to gain where rearranging within the class does not.

**The frontier comparison.** Comparing at a target alarm count requires the calibration to transfer, and it does not. We therefore sweep both classes across alarm rates and compare the resulting frontiers

by interpolation at pre-specified rates, so no policy is selected for having matched and no split is selected for being clean. The interpolation is *descriptive*: it reads each empirical frontier at a common resource level rather than constructing a policy operating there, which would require randomizing between adjacent solved policies. It compares the two estimated frontiers over the reported range; it does not establish that either lies above the other in the population, nor that any intermediate policy was run.

**The estimand.** Write $F_A(r;\mathcal{T})$ and $F_F(r;\mathcal{T})$ for the detection attained at $r$ measurements per night by the two *procedures* — fit, select, calibrate — run on a training sample $\mathcal{T}$ and scored on held-out patients. These are properties of the procedures, not of fixed policy classes: a different $\mathcal{T}$ selects a different schedule and a different price. The primary estimand averages over that randomness: $\Psi(P) = \mathbb{E}_{\mathcal{T}\sim P}[\,\Psi(\mathcal{T})\,]$, which is what a bootstrap that reruns selection inside every replicate estimates. Writing the inner quantity as

$$\Psi \;=\; \int_{r_L}^{r_U} \big\{F_A(r) - F_F(r)\big\}\, w(r)\,\mathrm{d}r, \qquad \int_{r_L}^{r_U} w(r)\,\mathrm{d}r = 1, \tag{5}$$

with $[r_L, r_U] = [1.5, 4.0]$ and $w$ uniform, both fixed beforehand. Frontiers are read at six rates by linear interpolation between attained rates and integrated by the trapezoid rule; nothing is extrapolated, so a rate a class does not attain contributes nothing.

Appendix F.13 tabulates the vertical gap in percentage points at each pre-specified rate for all five patient splits. The gap is positive at all thirty rate-by-split combinations, mean $+13.5$ points, but those cells come from five overlapping splits and six correlated rates and are descriptive rather than thirty replications.

**Optimization standards.** The fixed schedule is the exact maximizer over its class; the adaptive policy comes from an approximate solver under a model we know to be misspecified (Appendices F.9 and G.2). The gap also combines within-night timing with across-night reallocation, which we do not separate (Appendix H.15). The asymmetry does not have an unambiguous direction, and we can now price one side of it. A better solver could widen the gap, which favors the fixed arm as reported; but exact maximization over the class carries selection optimizm, and measuring it across folds gives $+10.4$ points of train-test gap for the exhaustive schedule against $-0.0$ for a lattice, which selects nothing from the data (Appendix H.12). We report the comparison as it stands and note that neither arm is the population optimum of its class. A null result here would not show that adaptivity fails.

**The evaluation protocol.** Both arms are selected on training data, so an interval that conditions on the selection understates uncertainty. We resample patients and rerun the entire pipeline inside each replicate: split, kernel fit, exhaustive selection, price calibration, rollout and interpolation. Two details are easy to get wrong. The split is over *unique* identifiers, each assigned wholly to one side with its multiplicity following it; splitting by position instead would put one patient's nights on both sides. That leakage is not small and does not shrink with sample size: Proposition 8 shows the expected fraction of distinct units straddling a split into fractions $f$ and $1-f$ converges to $L(f) = 1 - e^{f-1} - e^{-f} + e^{-1}$, which is $15.5\%$ at $f = 1/2$ and is approached quickly from below — $15.1\%$ already at $n = 10$. An assertion in the code enforces the correct split. And $\widehat{\Psi}$ is recomputed inside each replicate from its own two frontiers rather than aggregated afterwards (Appendix G). Across the $1{,}837$ of $2{,}000$ replicates in which both classes attain all six rates, the integrated gap has a $95\%$ percentile interval of $[-2.0, +39.8]$ and is positive in $96.4\%$ of them, with a Monte Carlo error of $0.25$ points on the endpoints. The classes are nested, so a negative population gap is impossible and the negative replicates are evaluation variance (Appendix H.17). The interval covers zero, so we do not claim population superiority, but zero sits at its edge with $96.4\%$ of the mass above: that is imprecision, not absence (Greenland et al., 2016). Nor is a hard comparator what puts it there — exhaustive search beats the best lattice by only $0$ to $2.7$ points here, so a weaker fixed baseline would barely move the interval. The estimates favor the adaptive procedure at every rate; $112$ patients leave the magnitude unresolved. A rerun storing each arm separately supplies the monotone-envelope and concave-hull variants: all three rules agree to within $0.8$ points and none excludes zero (Appendix G).

**A comparator that scales.** Proposition 2 licenses a greedy fixed schedule wherever enumeration is infeasible; on this cohort it recovers $94$ to $100$ percent of the exhaustive optimum while evaluating $nk$ candidates instead of $\binom{n}{k}$ (Appendix H.7).

**The magnitude is not robust.** Solving under the unchanged Gaussian belief while generating from four innovation processes leaves the sign stable in all eight settings but moves the size from $+3.8$ to $+23.0$ points (Appendix F.9). That range brackets our $14.2$: sampling uncertainty spans $41.8$ points and model uncertainty $19.2$, and only the first shrinks with more patients.

## 5 Discussion and limitations

Two of our findings look opposed and are not. Rearranging measurements within a fixed schedule is worth little; leaving that class for an adaptive policy is estimated at an integrated $14.2$ points, with a bootstrap interval of $[-2.0, +39.8]$ whose lower edge touches zero and $96.4\%$ of whose mass lies above it. The difference is the comparison, not the method: rearrangement is capped by the onset distribution, whose excess mass over the best lattice is $0.038$ across 1,720 subjects, while conditioning on what has been observed reaches a strictly larger policy class. Placement within the fixed class is cheap; the gain we measure comes from a procedure that conditions on observations, which is a larger class but also a different fitting procedure — we do not separate those two sources.

**What transfers.** The protocol and the separation result do not depend on glucose. Any comparison between a solved policy and a baseline chosen by search faces the same two errors — conditioning on the selection, and resampling units that reappear on both sides of a split — so the design applies to predictive maintenance, sensor scheduling under a communication budget, and screening intervals. Enumerability is rarer but checkable from the problem alone, and Proposition 7 now needs no divisibility assumption. A second cohort is the most important validation we have not done. The public CGM collections index outpatient studies and most inpatient cohorts sit behind data-access agreements. The standard arm of the TIGHT trial (Hirsch et al., 2025) put $50$ hospitalized patients on a blinded sensor; whether its record is downloadable with the measurement times we need is not established. Events there are scarcer still, so it would test transport rather than add power.

The deployment question this frames is not automation. A solver does not replace the clinician who decides what to do about a low reading; it proposes *when to look*, which is the part a guideline states as a fixed interval because a fixed interval is what a person can follow. Our estimates favor conditioning that decision on the current belief, and the honest summary is that they favor it by an amount this cohort cannot pin down. What the benchmark supplies either way is a way to ask the question without the answer depending on which baseline someone picked.

**Implications for guidelines.** Inpatient guidance recommends continuous monitoring with confirmatory point-of-care measurement for hypoglycemia-prone patients (Korytkowski et al., 2022; American Diabetes Association Professional Practice Committee, 2026). We agree, and would sharpen the reason. Read reactively, a sensor warns no sooner than three-hourly measurement. Its value lies in learning whom to watch, where one day of sensing recovers most of the benefit at a tenth of the sensor-days, and in conditioning each measurement on the current belief.

Two further formal routes were tried and are reported as failures: a finite-class uniform-convergence bound and the approximate-information-state bound of Subramanian et al. (2022) both hold and neither informs at our scale (Appendix G.3). Our formulation rests on Assumption 1: measuring does not alter the patient. It is false whenever a reading triggers treatment. Proposition 4 prices the violation rather than assuming it away, and the blinded design leaves our estimates unaffected, but a deployed policy would face it.

One cohort, 112 patients, two hospitals, with excursions clustered enough that 519 behave like 98. The nocturnal comparison rests on 93 held-out excursions; within a single split the patient-clustered interval crosses zero, and the claim rests on the frontier gap rather than any one point. The stated probabilities are not calibrated — in the middle range they overstate risk about fourfold, which moves a threshold without reordering states but means they are not clinical risks (Appendix G.4). We do not oversample the rare excursions. Imbalance corrections give poorly calibrated models without improving discrimination, and duplicating excursions would reuse nights (Appendix G.1). A complexity-matched comparator would sharpen it; we have not run one, and the kernel is held fixed, so this is planning (Ross et al., 2011) (Appendix C.5).

## ETHICS STATEMENT

This work uses a public, de-identified dataset of hospitalized patients released under CC BY 4.0 by its original investigators, who obtained ethical approval and consent; we obtained no new data and had no contact with patients. Two aspects warrant comment. First, our estimand is detection on an observed trajectory, not a counterfactual clinical outcome, so no result here should be read as evidence that any monitoring policy improves patient outcomes. Second, the allocation policy that performs best concentrates a fixed measurement budget on a minority of patients, which by construction reduces measurement for the rest; we report it as an analysis of what the information in the data supports, not as a recommendation, and the deployable variant we test does not reproduce the effect. The cohort is drawn from two hospitals in one city, and the published high-risk criteria we compare against do not transport to it, which is itself a caution against transporting our own findings.

## REPRODUCIBILITY STATEMENT

All quantities are computed from a public dataset whose depositors' DOI is given in Section 4; we cite the record rather than redistributing the files, so its license remains the one its depositors set. Appendix B.1 specifies the splitting, interval construction, clustering correction and leakage certification in full; Appendix A fixes the notation; Appendix G.11 reports the constants behind Proposition 5 and the measured tightness. The proof of Proposition 5 is stated in Section 3 with its assumptions and proved in Appendix F.1, as are Propositions 1, 6 and 7. Source code for every experiment, including the policy certification gate and the two deliberately leaky regression policies, is submitted as supplementary material, with `nocturnal.py` and `na_intervals.py` reproducing the headline comparison, `solver_rm.py` a hard-budget variant we do not use for the reported results, and `make_numbers.py` regenerating the macros the paper's figures are written from, so no reported number can disagree with the file it came from.

## REFERENCES

Izuwa Ahanor, Hugh Medal, and Andrew C. Trapp. DiversiTree: A new method to efficiently compute diverse sets of near-optimal solutions to mixed-integer optimization problems. *INFORMS Journal on Computing*, 36(1):61–77, 2024. doi: 10.1287/ijoc.2022.0164.

American Diabetes Association Professional Practice Committee. 16. Diabetes Care in the Hospital: Standards of Care in Diabetes—2026. *Diabetes Care*, 49, 2026. doi: 10.2337/dc26-s016.

Valentin Amrhein, Sander Greenland, and Blake McShane. Scientists rise up against statistical significance. *Nature*, 567:305–307, 2019. doi: 10.1038/d41586-019-00857-9.

Arash Asadpour and Hamid Nazerzadeh. Maximizing stochastic monotone submodular functions. *Management Science*, 62(8):2374–2391, 2016. doi: 10.1287/mnsc.2015.2254.

Turgay Ayer, Oguzhan Alagoz, and Natasha K. Stout. OR Forum—A POMDP Approach to Personalize Mammography Screening Decisions. *Operations Research*, 60(5):1019–1034, 2012. doi: 10.1287/opre.1110.1019.

Taposh Banerjee and Venugopal V. Veeravalli. Data-efficient Quickest Change Detection with Unknown Post-change Distribution. In *2014 IEEE International Symposium on Information Theory*, 2014. doi: 10.1109/isit.2014.6874931.

Richard E. Barlow, Larry C. Hunter, and Frank Proschan. Optimum Checking Procedures. *Journal of the Society for Industrial and Applied Mathematics*, 11(4):1078–1095, 1963. doi: 10.1137/0111080.

Tadej Battelino, Thomas Danne, Richard M. Bergenstal, Stephanie A. Amiel, et al. Clinical Targets for Continuous Glucose Monitoring Data Interpretation: Recommendations From the International Consensus on Time in Range. *Diabetes Care*, 42(8):1593–1603, 2019. doi: 10.2337/dci19-0028.

C. Bellinger, R. Coles, M. Crowley, and I. Tamblyn. Active Measure Reinforcement Learning for Observation Cost Minimization. In *Canadian Conference on AI*, 2021. doi: 10.21428/594757db.72846d04. URL https://arxiv.org/abs/2005.12697.

Elizabeth Bismut and Daniel Straub. Optimal adaptive inspection and maintenance planning for deteriorating structural systems. *Reliability Engineering & System Safety*, 215:107891, 2021. doi: 10.1016/j.ress.2021.107891.

Alex Carriero, Kim Luijken, Anne de Hond, Karel G. M. Moons, Ben van Calster, and Maarten van Smeden. The Harms of Class Imbalance Corrections for Machine Learning Based Prediction Models: A Simulation Study. *Statistics in Medicine*, 44, 2025. doi: 10.1002/sim.10320.

Beyza Cinar and Maria Maleshkova. DiaData: A Multi-Modal, Integrated Time-Series Dataset for Type 1 Diabetes Research. *BIO Web of Conferences*, 195:03001, 2025. doi: 10.1051/bioconf/2025 19503001.

Yubing Cui and Viswanath Nagarajan. Minimum cost adaptive submodular cover. *Mathematics of Operations Research*, 2025. doi: 10.1287/moor.2024.0548.

Yunus Emre Demirci, Ali Devran Kara, and Serdar Yüksel. Sensitivity of Filter Kernels and Robustness Bounds to Transition and Measurement Kernel Perturbations in Partially Observable Stochastic Control. *arXiv preprint 2508.10658*, 2025. doi: 10.48550/arXiv.2508.10658.

Chong-Zhi Di, Ciprian M. Crainiceanu, Brian S. Caffo, and Naresh M. Punjabi. Multilevel functional principal component analysis. *The Annals of Applied Statistics*, 3(1):458–488, 2009. doi: 10.1214/08-AOAS206.

Randal Douc, Eric Moulines, and Yaacov Ritov. Forgetting of the initial condition for the filter in general state-space hidden Markov chain: a coupling approach. *Electronic Journal of Probability*, 14, 2009. doi: 10.1214/ejp.v14-593.

Acheson J. Duncan. The Economic Design of $\bar{X}$ Charts Used to Maintain Current Control of a Process. *Journal of the American Statistical Association*, 51(274):228–242, 1956. doi: 10.2307/2281343.

Steven V. Edelman and Jean S. Blose. The impact of nocturnal hypoglycemia on clinical and cost-related issues in patients with type 1 and type 2 diabetes. *The Diabetes Educator*, 40(3):269–279, 2014. doi: 10.1177/0145721714529608.

Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, 1993. doi: 10.1007/978-1-4899-4541-9.

Hossein Esfandiari, Amin Karbasi, and Vahab Mirrokni. Adaptivity in adaptive submodularity. In *Proceedings of the 34th Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pp. 1823–1846, 2021.

Majid Forghani-Elahabad and Luiz H. Bonani. Finding all the lower boundary points in a multistate two-terminal network. *IEEE Transactions on Reliability*, 66(3):677–688, 2017. doi: 10.1109/TR.2017.2712661.

Irina Gaynanova, Naresh Punjabi, and Ciprian Crainiceanu. Modeling continuous glucose monitoring (CGM) data during sleep. *Biostatistics*, 23(1):223–239, 2022. doi: 10.1093/biostatistics/kxaa023.

Rohan Ghuge, Anupam Gupta, and Viswanath Nagarajan. The power of adaptivity for stochastic submodular cover. *Operations Research*, 72(3):1156–1176, 2024. doi: 10.1287/opre.2022.2388.

M. A. Girshick and Herman Rubin. A Bayes Approach to a Quality Control Model. *The Annals of Mathematical Statistics*, 23(1):114–125, 1952. doi: 10.1214/aoms/1177729489.

Styliani Giza, Eleni P. Kotanidou, Vasiliki Rengina Tsinopoulou, et al. Can Glucose Alarm Fatigue Threaten the Absolute Clinical Benefit of Continuous Glucose Monitoring? *Children*, 12:1668, 2025. doi: 10.3390/children12121668.

Daniel Golovin and Andreas Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, 42:427–486, 2011.

Sander Greenland, Stephen J. Senn, Kenneth J. Rothman, John B. Carlin, Charles Poole, Steven N. Goodman, and Douglas G. Altman. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, 31:337–350, 2016. doi: 10.1007/s10654-016-0149-3.

Anupam Gupta, Viswanath Nagarajan, and Sahil Singla. Adaptivity gaps for stochastic probing: Submodular and XOS functions. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1688–1702, 2017. doi: 10.1137/1.9781611974782.111.

Anupam Gupta, Benjamin Moseley, and Rudy Zhou. Bayesian probing on graphs. In *Integer Programming and Combinatorial Optimization (IPCO 2026)*, volume 16588 of *Lecture Notes in Computer Science*, pp. 410–425. Springer, 2026. doi: 10.1007/978-3-032-28691-8_27.

Jonathan E. Helm, Mariel S. Lavieri, Mark P. Van Oyen, Joshua D. Stein, and David C. Musch. Dynamic Forecasting and Control Algorithms of Glaucoma Progression for Clinician Decision Support. *Operations Research*, 63(5):979–999, 2015. doi: 10.1287/opre.2015.1405.

Irl B. Hirsch, Boris Draznin, John B. Buse, Dan Raghinaru, Charles Spanbauer, and Guillermo E. Umpierrez. Results From a Randomized Trial of Intensive Glucose Management Using CGM Versus Usual Care in Hospitalized Adults With Type 2 Diabetes: The TIGHT Study. *Diabetes Care*, 48(1):118–124, 2025. doi: 10.2337/dc24-1779.

International Hypoglycaemia Study Group. Glucose Concentrations of Less Than 3.0 mmol/L (54 mg/dL) Should Be Reported in Clinical Trials: A Joint Position Statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care*, 40 (1):155–157, 2017. doi: 10.2337/dc16-2215.

Arec Jamgochian, Anthony Corso, and Mykel J. Kochenderfer. Online Planning for Constrained POMDPs with Continuous Spaces through Dual Ascent. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 33, 2023. doi: 10.1609/icaps.v33i1.27195.

Naoto Kaio and Shunji Osaki. Optimal inspection policies: A review and comparison. *Journal of Mathematical Analysis and Applications*, 119(1-2):3–20, 1986. doi: 10.1016/0022-247x(86)90141-1.

Pooyan Kazemian, Jonathan E. Helm, Mariel S. Lavieri, Joshua D. Stein, and Mark P. Van Oyen. Dynamic Monitoring and Control of Irreversible Chronic Diseases with Application to Glaucoma. *Production and Operations Management*, 28(5):1082–1107, 2019. doi: 10.1111/poms.12975.

Joseph B. Keller. Optimum Checking Schedules for Systems Subject to Random Failure. *Management Science*, 21(3):256–260, 1974. doi: 10.1287/mnsc.21.3.256.

Alice P. S. Kong, Soo Lim, Seung-Hyun Yoo, Linong Ji, Liming Chen, et al. Asia-Pacific consensus recommendations for application of continuous glucose monitoring in diabetes management. *Diabetes Research and Clinical Practice*, 201:110718, 2023. doi: 10.1016/j.diabres.2023.110718.

Mary T. Korytkowski, Ranganath Muniyappa, Kellie Antinori-Lent, Amy C. Donihi, Andjela T. Drincic, Irl B. Hirsch, Anton Luger, Marie E. McDonnell, M. Hassan Murad, Craig Nielsen, Claire Pegg, Robert J. Rushakoff, Nancy Santesso, and Guillermo E. Umpierrez. Management of Hyperglycemia in Hospitalized Adult Patients in Non-Critical Care Settings: An Endocrine Society Clinical Practice Guideline. *The Journal of Clinical Endocrinology & Metabolism*, 107(8): 2101–2128, 2022. doi: 10.1210/clinem/dgac278.

M. Krale, T. D. Simao, and N. Jansen. Act-Then-Measure: Reinforcement Learning for Partially Observable Environments with Active Measuring. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 33, pp. 212–220, 2023. doi: 10.1609/icaps.v33i1.27197. URL https://arxiv.org/abs/2303.08271.

Vikram Krishnamurthy. *Partially Observed Markov Decision Processes: Filtering, Learning and Controlled Sensing*. Cambridge University Press, second edition, 2025. ISBN 978-1-009-44943-4. doi: 10.1017/9781009449441.

Vikram Krishnamurthy and Udit Pareek. Myopic Bounds for Optimal Policy of POMDPs: An Extension of Lovejoy's Structural Results. *Operations Research*, 63(2):428–434, 2015. doi: 10.1287/opre.2014.1332.

Jonathan N. Lee, Alekh Agarwal, Christoph Dann, and Tong Zhang. Learning in POMDPs is sample-efficient with hindsight observability. In *Proceedings of the 40th International Conference on Machine Learning*, 2023. arXiv:2301.13857.

Andrew C. Li, Zizhao Chen, Toryn Q. Klassen, Rodrigo Toro Icarte, and Sheila A. McIlraith. Reward Machines for Deep RL in Noisy and Uncertain Environments, 2024. arXiv:2406.00120.

William S. Lovejoy. Some Monotonicity Results for Partially Observed Markov Decision Processes. *Operations Research*, 35:736–743, 1987. doi: 10.1287/opre.35.5.736.

Curtis McDonald and Serdar Yüksel. Observability and Filter Stability for Partially Observed Markov Processes. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019. doi: 10.1109/cdc40024.2019.9029775.

Jeremy McMahan and Xiaojin Zhu. Anytime-constrained reinforcement learning. In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 4321–4329, 2024. arXiv:2311.05511.

José Luis Montiel Olea and Mikkel Plagborg-Møller. Simultaneous confidence bands: Theory, implementation, and an application to SVARs. *Journal of Applied Econometrics*, 34(1):1–17, 2019. doi: 10.1002/jae.2656.

A. G. Munford and A. K. Shahani. A Nearly Optimal Inspection Policy. *Journal of the Operational Research Society*, 23(3):373–379, 1972. doi: 10.1057/jors.1972.56.

H. A. Nam, S. L. Fleming, and E. Brunskill. Reinforcement Learning with State Observation Costs in Action-Contingent Noiselessly Observable MDPs. In *NeurIPS 34*, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/d34ab169b70c9dcd35e62896010cd9ff-Abstract.html.

Samrat Nath and Jingxian Wu. Quickest change point detection with multiple postchange models. *Sequential Analysis*, 39(4):543–562, 2020. doi: 10.1080/07474946.2020.1826795.

G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14(1):265–294, 1978. doi: 10.1007/bf01588971.

Yuqing Ni, Xiaochen Liu, and Chao Yang. Sensor Scheduling for Remote State Estimation with Limited Communication Resources: A Time- and Event-Triggered Hybrid Approach. *Sensors*, 23(21):8667, 2023. doi: 10.3390/s23218667.

Yousef Oleyaeimotlagh, Taposh Banerjee, Ahmad Taha, and Eugene John. Quickest Change Detection in Statistically Periodic Processes with Unknown Post-Change Distribution, 2023. arXiv:2303.02826.

Thierry Petit and Andrew C. Trapp. Enriching solutions to combinatorial problems via solution engineering. *INFORMS Journal on Computing*, 31(3):429–444, 2019. doi: 10.1287/ijoc.2018.0855.

Pascal Poupart, Aarti Malhotra, Pei Pei, Kee-Eung Kim, Bongseok Goh, and Michael Bowling. Approximate Linear Programming for Constrained Partially Observable Markov Decision Processes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015. doi: 10.1609/aaai.v29i1.9655.

Richard D. Riley, Kym I. E. Snell, Joie Ensor, Danielle L. Burke, Frank E. Harrell, Karel G. M. Moons, and Gary S. Collins. Minimum sample size for developing a multivariable prediction model: Part I — Continuous outcomes. *Statistics in Medicine*, 2018. doi: 10.1002/sim.7993.

Stéphane Ross, Joelle Pineau, Brahim Chaib-draa, and Pierre Kreitmann. A Bayesian Approach for Learning and Planning in Partially Observable Markov Decision Processes. *Journal of Machine Learning Research*, 12(48):1729–1770, 2011.

Burhaneddin Sandıkçı, Lisa M. Maillart, Andrew J. Schaefer, and Mark S. Roberts. Alleviating the Patient's Price of Privacy Through a Partially Observable Waiting List. *Management Science*, 59 (8):1836–1854, 2013. doi: 10.1287/mnsc.1120.1671.

Herbert Scarf. The Optimality of (S,s) Policies in the Dynamic Inventory Problem. In *Mathematical Methods in the Social Sciences*. Stanford University Press, 1960. URL https://doi.org/10.1057/9781137024381_1.

R. Sergazinov, E. Chun, V. Rogovchenko, N. Fernandes, N. Kasman, and I. Gaynanova. GlucoBench: Curated List of Continuous Glucose Monitoring Datasets with Prediction Benchmarks. In *International Conference on Learning Representations*, 2024. doi: 10.48550/arXiv.2410.05780. URL https://arxiv.org/abs/2410.05780.

Renat Sergazinov, Andrew Leroux, Erjia Cui, Ciprian Crainiceanu, R. Nisha Aurora, Naresh Punjabi, and Irina Gaynanova. A case study of glucose levels during sleep using fast function on scalar regression inference. *Biometrics*, 79(4):3873–3882, 2023. doi: 10.1111/biom.13878.

Ling Shi, Ye Yuan, and Huanshui Zhang. Sensor data scheduling for linear quadratic Gaussian control with full state feedback. In *2012 American Control Conference (ACC)*, 2012. doi: 10.1109/acc.2012.6314650.

M. Reza Skandari and Steven M. Shechter. Patient-Type Bayes-Adaptive Treatment Plans. *Operations Research*, 69(2):574–598, 2021. doi: 10.1287/opre.2020.2011.

Ewout W. Steyerberg and Frank E. Harrell. Prediction models need appropriate internal, internal-external, and external validation. *Journal of Clinical Epidemiology*, 2016. doi: 10.1016/j.jclinepi.2015.04.005.

J. Subramanian, A. Sinha, R. Seraj, and A. Mahajan. Approximate information state for approximate planning and reinforcement learning in partially observed systems. *JMLR 23(12):1-83*, 2022. doi: 10.48550/arXiv.2010.08843. URL https://arxiv.org/abs/2010.08843.

Vladislav B. Tadić and Arnaud Doucet. Exponential forgetting and geometric ergodicity for optimal filtering in general state-space models. *Stochastic Processes and their Applications*, 115(8): 1408–1436, 2005. doi: 10.1016/j.spa.2005.03.005.

Jonathan Taylor and Robert J. Tibshirani. Statistical learning and selective inference. *Proceedings of the National Academy of Sciences*, 112(25):7629–7634, 2015. doi: 10.1073/pnas.1507583112.

Ryan J. Tibshirani, Alessandro Rinaldo, Robert Tibshirani, and Larry Wasserman. Uniform asymptotic inference and the bootstrap after model selection. *The Annals of Statistics*, 46(3):1255–1287, 2018. doi: 10.1214/17-AOS1584.

Rodrigo Toro Icarte, Toryn Q. Klassen, Richard Valenzano, and Sheila A. McIlraith. Reward Machines: Exploiting Reward Function Structure in Reinforcement Learning. *Journal of Artificial Intelligence Research*, 73:173–208, 2022. doi: 10.1613/jair.1.12440.

Ruben van den Goorbergh, Maarten van Smeden, Dirk Timmerman, and Ben Van Calster. The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression. *Journal of the American Medical Informatics Association*, 29(9):1525–1534, 2022. doi: 10.1093/jamia/ocac093.

Taiyi Wang, Jianheng Liu, Bryan Lee, Zhihao Wu, and Yu Wu. OCMDP: Observation-Constrained Markov Decision Process, 2024. arXiv:2411.07087.

Ronald L. Wasserstein and Nicole A. Lazar. The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, 70(2):129–133, 2016. doi: 10.1080/00031305.2016.1154108.

Yuheng Zhang and Nan Jiang. Statistical Tractability of Off-policy Evaluation of History-dependent Policies in POMDPs. *arXiv preprint 2503.01134*, 2025. doi: 10.48550/arXiv.2503.01134.

Q. Zhao, J. Zhu, X. Shen, C. Lin, Y. Zhang, Y. Liang, and et al. (11 authors). Chinese diabetes datasets for data-driven machine learning. In *Scientific Data 10:35*, 2023. doi: 10.1038/s41597-023-01940-7. URL https://doi.org/10.1038/s41597-023-01940-7.

Yuxing Zhong, Jiawei Tang, Nachuan Yang, Dawei Shi, and Ling Shi. Event-Triggered Sensor Scheduling for Remote State Estimation With Error-Detecting Codes. *IEEE Control Systems Letters*, 2023. doi: 10.1109/lcsys.2023.3286472.

Yuxing Zhong, Lingying Huang, Yilin Mo, Dawei Shi, and Ling Shi. Event-Triggered Multi-Sensor Scheduling for Remote State Estimation Over Packet-Dropping Networks. *IEEE Transactions on Signal Processing*, 2024. doi: 10.1109/tsp.2024.3473988.

Ming J. Zuo, Zhigang Tian, and Hong-Zhong Huang. An efficient method for reliability evaluation of multistate networks given all minimal path vectors. *IIE Transactions*, 39(8):811–817, 2007. doi: 10.1080/07408170601013653.

## USE OF LARGE LANGUAGE MODELS

Generative AI was used extensively throughout this work as a research assistant. We disclose each category the policy lists as requiring disclosure.

**Used.** Generating synthetic data sets: the four innovation processes of Appendix F.9, the discrete state space of Appendix G.2, and the resampling simulations verifying Proposition 8. Developing theoretical models and conceptual frameworks: the reward-machine formulation and the exploration of the approximate-information-state route reported in Appendix G.3. Formulating mathematical claims and providing critical ingredients for their proofs: Proposition 8 was formulated with AI assistance, and its closed form and the rotation-counting argument behind Proposition 7 were derived that way. Assisting in the writing of proofs. Proposing and refining hypotheses. Designing and giving feedback on research methodology and experimental design, including the resampling protocol of Section 4.4. Implementing methods. Cleaning and reformatting the dataset. Interpreting results.

**Not used.** Assisting with translation.

**Not applicable.** Qualitative and thematic data analysis.

Among the recommended categories we also used it for writing and editing code, creating and modifying figures, suggesting experimental parameters, drafting parts of the paper, literature search and summarization, reference formatting, structural suggestions, proposing the title, and brainstorming.

**Verification.** Every factual claim was checked against primary sources, and every number reported here was produced by executed code rather than by a language model; the code is in the supplement. Each proposition was verified numerically as well as proved: Proposition 8 against an exact finite-$n$ sum and simulation, Proposition 7 against exhaustive search over $200$ instances, Proposition 5 across $48$ synthetic configurations, and Proposition 9 across $300$ exactly solved instances. Several conclusions were reversed when verification failed — comparisons across strata with different base rates, a ratio declared refuted because its interval touched the null, a policy reading sensor values it had not paid for, a value-of-information proxy largest exactly where information was worthless, and a resampling split that leaked patients across the train/test boundary. The checklist in Appendix B.1 and the leakage certification were instituted in response to those failures. The authors take full responsibility for the contents of this paper, including all text, claims and artifacts produced with AI assistance.

# A NOTATION

| | |
|---|---|
| | *POMDP* |
| $t$ | time index on a 15-minute grid |
| $s_t \in \mathcal{S}$ | latent glycemic state, discretized |
| $a_t \in \mathcal{A} = \{0, 1\}$ | measure or not |
| $o_t \in \mathcal{O}$ | observation, returned only when $a_t = 1$ |
| $P$ | transition kernel; $P(0) = P(1) = P$ by Assumption 1 |
| $Z(a)$ | observation kernel under action $a$; action-dependent |
| $b_t \in \Delta(\mathcal{S})$ | marginal belief over the glycemic state |
| $\beta \in \Delta(\mathcal{S} \times \mathcal{U})$ | joint belief over state and automaton, Equation (2) |
| $\Phi$, $\widetilde{\Phi}$ | exact joint filter and its factorised approximation |
| $F(b, o, a)$ | belief update, Equation (1) |
| $\phi$ | policy mapping beliefs to actions |
| $\gamma$ | discount factor |
| $r(s, a)$ | per-step reward; $\kappa(t)$ the price of a measurement |
| $\lambda_{\text{miss}}$, $\lambda_{\text{FA}}$ | weights on a missed excursion and a false alarm |
| $\kappa$ | price per measurement; swept to trace the frontier |
| $V_h^\kappa(\beta)$, $\widetilde{V}^\kappa$ | value under the exact and the approximate filter, Equation (2) |
| $A_h(\widetilde{\beta})$ | advantage of measuring on the reduced state, Equation (3) |
| $\pi$ | stationary distribution of $P$ |
| $F_A(r)$, $F_F(r)$ | adaptive and fixed frontiers at rate $r$ |
| $\Psi$ | integrated frontier gap, Equation (5) |
| | *Belief contraction diagnostics* |
| $\lambda_2$ | second-largest eigenvalue modulus of $P$ |
| $\tau_{1/2}$ | belief mixing half-life, $\log(1/2)/\log\lvert\lambda_2\rvert$ |
| $\Delta t$ | feasible measurement interval, in slots |
| $R$ | interval-to-half-life ratio, $\Delta t/\tau_{1/2}$ |
| $L$, $C$ | Lipschitz constant of $A$; mixing constant of $P$ |
| | *Evaluation* |
| $E_j$ | time interval of the $j$-th excursion |
| $D$ | excursion duration in slots |
| $k$ | measurements per patient-day |
| $n = 96$ | slots per day |
| $Q_\phi$ | set of slots at which $\phi$ measures; random under an adaptive policy |

# B EVALUATION PROTOCOL AND ROBUSTNESS

## B.1 EVALUATION PROTOCOL IN FULL

**Splitting.** Whole patients are held out as an out-of-distribution test set; for training patients the last 16 hours of each record are held out chronologically as an in-distribution test set, following Sergazinov et al. (2024). Scalers, transition and observation kernels, and risk models are fitted inside the fold. The two sets answer different questions and differ materially: out-of-fold AUPRC is 0.864 in-distribution against 0.767 out-of-distribution, so roughly ten points of apparent performance come from having seen the patient before.

**Uncertainty.** Intervals reported for the auxiliary analyses in this appendix are 95% BCa, resampled over patients, with acceleration from a patient-level jackknife, chosen over percentile because the statistics are proportions and ratios on skewed, clustered data. The frontier gap $\widehat{\Psi}$ carries a percentile interval from the full-pipeline bootstrap of Appendix G rather than a BCa one, because BCa's acceleration is estimated by a jackknife over units and each replicate reruns selection: leaving a patient out changes which schedule is chosen rather than perturbing a fixed statistic. $\widehat{\Psi}$ is also a difference of areas rather than a ratio, the case percentile intervals handle worst. We report the Monte Carlo error on the endpoints so their instability can be read directly. We report these as compatibility intervals and describe comparisons by whether intervals overlap. We attach no verdict in either

direction: neither a point estimate without an interval, nor a claim of no effect from an interval that touches the null (Amrhein et al., 2019; Greenland et al., 2016; Wasserstein & Lazar, 2016).

**Clustering.** 519 hypoglycemic excursions occur in 72 of 112 patients, with the top five patients holding 29.1%. The intra-class correlation of "excursion detected" is 0.281 and the effective cluster size 16.2, giving a design effect of 5.27: the excursions behave like 98 independent ones. At 80% power the minimum detectable difference is about 13 percentage points, declared before any policy comparison was run. Hour-level results are therefore not identifiable at this sample size and we report none.

**Sensor agreement.** Against 4,012 paired capillary measurements the sensor shows MARD 13.89% with a range-dependent bias: $+10.5$ mg/dL below 54, $+7.7$ between 54 and 70, $-38.3$ above 250. At most 2.9% of the discrepancy variance can be i.i.d. white noise, so the error is persistent and smoothing removes almost nothing. It meets the ADA 20/20 criterion on 76.0% of pairs. We do *not* use the empirical sensor-to-fingerstick confusion as an observation kernel: measurements near a low sensor value were often taken after treatment, so the table is contaminated. The observation kernel is built from the measured error model instead: Gaussian measurement error with standard deviation $8\,\mathrm{mg\,dL^{-1}}$. Where a discretized kernel is needed — the structural conditions of Appendix C.3 — the three bins are cut at 70 and $180\,\mathrm{mg\,dL^{-1}}$.

**Leakage certification.** A policy is admitted only if its action sequence is unchanged when the sensor value at every slot it does *not* measure is replaced by uniform noise. Policies with an alert rule must additionally leave the alert sequence unchanged; certifying the schedule alone is insufficient, because a fixed leak-free schedule can still drive an alert computed from unmeasured values. Two deliberately leaky policies are retained as regression tests: if either certifies, the gate is broken.

# C Results reported outside the main text

## C.1 Two-sided monitoring

Severe hyperglycemia above 250 mg/dL produces 686 excursions against 519 hypoglycemic ones, so the problem cannot be reported as one-sided. A U-shaped cost breaks monotonicity on a single glucose ordering. We use two monotone sub-problems sharing one measurement action, each satisfying the ordering conditions separately: TP2 minors $+0.000357$ on the hypoglycemic ordering and $+0.000729$ on the hyperglycemic one, total positivity being invariant under reversing rows and columns together.

Matched on realized budget the resulting two-threshold policy ties on the hypoglycemic side and loses on the hyperglycemic side. The ratio of observed detection to the coverage null is 1.20–1.24 on the hypoglycemic side and 1.24–1.30 on the hyperglycemic side across a sixfold budget range. The tails differ only through mean excursion duration, 108 minutes against 139.

A K-convexity analogue on the belief simplex would remove the need for the decomposition; Scarf (1960) solved the corresponding problem for inventory by replacing convexity with a weaker property yielding a two-threshold $(s, S)$ rule. We searched and found no analogue for beliefs. Our numerics are consistent with a two-threshold shape — as the grid refines, the non-measure region absorbs 96–97% of its mass into a single connected component while the count of components rises, the signature of a connected region with a resolution-limited boundary — but a uniform-grid solver cannot settle it. In this problem such an analogue would certify a policy class with nothing to gain.

## C.2 Learning the type, and pricing the sensing

**The type is learnable and it is not the published checklist.** A type estimated from each stay's first half predicts the second half at Spearman $+0.598$ for excursion rate and $+0.626$ for time below range. Across tertiles of the first-half rate, second-half excursion rate separates 8.5-fold: 0.090 against 0.765 per patient-day. By contrast the guideline's high-risk criteria classify 92% of this cohort as high risk, driven by a BMI cut-off of 27 derived in a different population, and separate excursion rate by only 0.461 against 0.490 per patient-day.

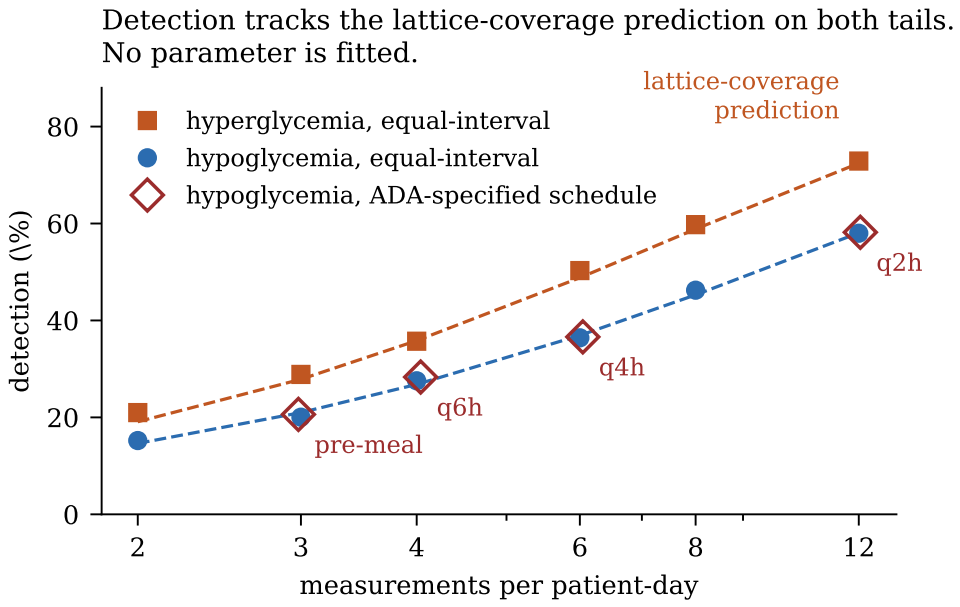



Figure 1: Detection against the coverage null. Dashed curves are the lattice-coverage reference $\min(1, Dk/n)$ averaged over the measured excursion-duration distribution, the null matching a regular schedule; no parameter is fitted to the points. Detection tracks it on both tails and on the schedules American Diabetes Association Professional Practice Committee (2026) specifies. The ratio of observed to predicted stays within 0.95–1.10 over a sixfold budget range.

**It does not survive the loss of the sensor.** Learned from fingersticks alone the type correlates $+0.460$ with later excursion rate at a 24-hour window, against $+0.581$ for the sensor-learned type, and *degrades* to $+0.359$ at 72 hours despite three times as many readings. A meal-anchored protocol resamples the same clock hours rather than adding information about the tails. We also note an estimator trap: our first statistic used the minimum reading, which drifts downward with sample size and correlates $+0.238$ with the number of readings; "fraction of readings below 100" correlates $+0.052$ with sample size and performs better.

Figure 2 places the four allocation arms on the axis that matters for cost.

**Break-even.** The gain of learned-type allocation over uniform is $+16.9$ points. At the measured budget–detection slope of 5.3–7.1 points per additional measurement per day, buying that gain with measurements would cost 2.4–3.2 per patient-day. Over a mean scored stay of 9.7 days, one day of admission sensing is worth its cost below roughly 23–31 fingerstick-equivalents. **Against the clinical comparators the figure is smaller**: the gain over observed practice is $+13.1$ points, giving 18–24 equivalents, and over the guideline's own criteria $+11.2$, giving 15–20. We supply no tariff.

### C.3 Structural conditions and the reachable belief set

With a binary state and a three-bin observation, refitted inside each of five patient-level training folds, $P$ is TP2 in 5/5 folds with minors from $+0.000747$ to $+0.000848$; on the full cohort the minor is $+0.000807$ with BCa $[+0.000566, +0.001073]$, and TP2 holds in 99.3% of patient-resampled draws. The observation kernel is TP2 ($+0.002098$) and full rank ($\sigma_{\min} = 0.589$), hence single-step revealing in the sense of Zhang & Jiang (2025), which is a binning choice we make deliberately so that their model-based result applies. Copositive dominance holds with equality by Assumption 1, and Blackwell dominance is exact by construction with residual $1.4 \times 10^{-17}$.

The region on which the myopic bound of Krishnamurthy (2025) coincides with the optimum covers 98.8% of the uniform simplex but 12.8% of the beliefs the process reaches, a gap of 86.0 points that persists under five policies spanning never-measure to always-measure. The source is the kernel: the

safe state's self-transition probability is 0.982, so beliefs concentrate within 0.018 of one vertex — inside a single cell of a 60-division grid. Optimizing a bound for uniform-simplex volume optimizes the wrong measure whenever the chain is this persistent.

### C.4 The contraction profile and the tightness of Proposition 5

**The reachable set does not contain the vertices.** Enumerating $\mathcal{B}_0 = \{F(b, o, 1)\}$ over a $1/12$-resolution grid of pre-measurement beliefs and all three observations gives 273 distinct posteriors. Their distance from stationarity ranges from $0.060$ to $\mathbf{1.878}$ in $\ell_1$, while the nearest simplex vertex sits at $1.940$ from $\pi$ by the same measure. A measurement therefore moves the belief *toward* a vertex without reaching one, which is why the bound is stated over $\mathcal{B}_0$ rather than over the vertices.

**The contraction profile, measured rather than bounded spectrally.**

| $t$ (slots) | $t$ (h) | $\varepsilon(t)$ | $\varepsilon(t)/\varepsilon(0)$ | spectral $C\lvert\lambda_2\rvert^t$ |
|---:|---:|---:|---:|---:|
| 0 | 0.00 | 1.993 | 1.000 | 2.640 |
| 4 | 1.00 | 1.558 | 0.782 | 1.685 |
| 8 | 2.00 | 1.075 | 0.540 | 1.076 |
| 12 | 3.00 | 0.709 | 0.356 | 0.687 |
| 16 | 4.00 | 0.459 | 0.231 | 0.439 |
| 24 | 6.00 | 0.189 | 0.095 | 0.179 |
| 32 | 8.00 | 0.077 | 0.039 | 0.073 |
| 48 | 12.00 | 0.013 | 0.006 | 0.012 |

The contraction half-life is $\tau_{1/2} = 9$ slots, or $2.25\,\mathrm{h}$, at the three-band resolution used throughout this section; under the continuous parametrization the same discrepancy halves in $10.68\,\mathrm{h}$ (Proposition 6). The second-largest eigenvalue modulus is $0.896$, which would give $1.57\,\mathrm{h}$; the spectral rate *understates* the half-life because the early decay is slower than geometric, while the two agree asymptotically. This is the concrete reason we estimate $\varepsilon$ rather than assuming $\|b_t - \pi\|_1 \le C|\lambda_2|^t$.

**Tightness.** Writing $L$ for the Lipschitz constant of the myopic advantage, which is $10.0$ for the reward and price used in the body:

| $k$/day | interval (h) | $R$ | $\varepsilon$ | bound $L\varepsilon$ | measured spread |
|---:|---:|---:|---:|---:|---:|
| 2 | 12.00 | 5.33 | 0.013 | 0.13 | 0.08 |
| 3 | 8.00 | 3.56 | 0.077 | 0.77 | 0.46 |
| 4 | 6.00 | 2.67 | 0.189 | 1.89 | 1.13 |
| 6 | 4.00 | 1.78 | 0.459 | 4.59 | 2.79 |
| 8 | 3.00 | 1.33 | 0.709 | 7.09 | 4.41 |
| 12 | 2.00 | 0.89 | 1.075 | 10.75 | 7.04 |
| 24 | 1.00 | 0.44 | 1.558 | 15.58 | 11.46 |

The bound holds at every budget with slack between $1.4$ and $1.7$, and it holds by construction rather than by assumption: $\varepsilon(m)$ is the measured diameter of the reachable set, so no vertex assumption and no spectral decomposition enters.

**Two statements that are proved rather than measured.** First, under Assumption 1 copositive dominance holds *with equality*, since $P(0) = P(1)$ makes both sides the same matrix. Second, continuous sensing Blackwell-dominates any intermittent schedule slot by slot: at an unmeasured slot the intermittent kernel maps every state to a single null observation, so taking $M$ to be the $|\mathcal{O}| \times 1$ all-ones column, which is stochastic, gives $(Z_{\mathrm{cont}}M)[s, \text{null}] = 1$ for every state $s$; at a measured slot $M = I$. The garbling acts on the output side, which is the direction the definition requires.

### C.5 Further limitations

Hour-level efficiency ratios have compatibility intervals wider than the quantities they estimate; at eight of twenty-two hours the coverage null is below 5%, so the ratio has a near-zero denominator and reaches values above 20 with interval widths above 15. We report none of them.

The published high-risk criteria do not transport to this cohort: 92% meet at least one, driven by a BMI cut-off of 27 derived in a different population, and they separate excursion rate by only 0.461 against 0.490 per patient-day. That is a caution against transporting our own findings as much as theirs.

The blinded sensor supplies the labels, and it is a noisy proxy rather than ground truth: excursions are defined from the CGM trace while the acquired measurement is capillary, so this is a latent-label problem. The sensor meets the ADA 20/20 agreement criterion on 76.0% of the 4,012 paired readings, with a range-dependent bias of $+10.5$ mg/dL below 54 and $-38.3$ above 250; at most 2.9% of the discrepancy variance can be white noise, so the error is persistent and smoothing removes almost nothing.

Finally, the check-rate of a patient correlates $-0.218$ with their excursion rate, surviving adjustment for length of stay. We do *not* report this as evidence that clinicians allocate away from risk: reverse causation is entirely plausible, since more measurement leads to more treatment and hence fewer recorded excursions, and the direction is not identified from observational data.

### C.6 Analyses reported here rather than in the main text

Three results support claims made in the body but were kept out of it for space. Two further analyses shaped the design of the study and are recorded because they explain choices a reader might otherwise find arbitrary.

**Costly features add nothing over the free signal.** Before settling on measurement timing we asked whether acquiring expensive contextual features — meal logs, insulin records, and additional laboratory values — improves prediction of an impending excursion over the sensor stream alone. Across twenty-four configurations spanning four targets, three horizons, and two levels of feature quality, the median change in area under the precision–recall curve was $-1.7\%$ and the best was $+0.4\%$. The mechanism is that the expensive features are *downstream* of the free one: after a hyperglycemic excursion begins, the rate of capillary measurement rises by a factor of 1.79 and the rate of correction boluses by 2.18. They are consequences of the glucose signal rather than independent evidence about it. The general criterion this suggests is that when an expensive feature is mediated by a free one, acquiring it is informationally redundant.

**Acquisition exogeneity separates the variance of two estimators.** Under Assumption 1 the joint distribution factorizes so that importance weights range only over the slots a policy actually measures, without any correction to the state distribution. Trajectory-product weights then scale as $p^{-k}(1-p)^{-(T-k)}$, exponential in the horizon, while per-decision weights scale as $p^{-k}$ and are horizon-free. Measured across forty-eight configurations, only six achieved an effective sample size above thirty under trajectory-product weighting, with weights reaching $10^{16}$ at a twenty-four hour horizon; thirty were usable under per-decision weighting. This is why the benchmark is computed by exact replay rather than by off-policy evaluation, a choice that also places us outside the hardness result of Zhang & Jiang (2025), which concerns model-free evaluation of history-dependent policies.

**The full objective surface.** Section 4 reports detection at fixed budgets. The underlying objective also carries a nighttime surcharge on measurement, a cost per missed excursion, and a cost per false alarm. Sweeping twenty-nine policies over twenty-four weightings, nine distinct policies are optimal somewhere. The false-alarm weight decides whether a model-based rule earns its keep: at a low weight an interpretable rule on the reading itself is optimal in eight of eight cells, and at a high weight a model-based threshold is optimal in five of eight. The nighttime surcharge decides whether to measure at night at all: at parity no daytime-only policy is optimal, and at a fourfold surcharge one is optimal in five of twelve cells, despite 30.6% of onsets occurring between midnight and six in the morning against 5.5% of measurements. We report the break-even device cost rather than a recommendation because the weights are placeholders for a utility model we have not built.

**Why the estimand is detection and not detection delay.** An earlier formulation treated the problem as quickest change detection, with the change point being the onset of a decline toward an excursion. That framing is natural and we abandoned it for a specific reason: with the belief placed on the current glycemic state, the posterior of "a change has occurred" is bounded above by the stationary probability of the bad state, which is 0.030 on this kernel. Every alarm threshold above that value is unreachable, and a threshold policy degenerates into a fixed staleness rule. That degeneracy is a property of where the belief was placed, not of the problem, and recognizing it is what led to Proposition 5.

**Hour-level results are not identifiable here.** We computed the ratio of observed detection to the coverage null separately for each hour of the day. At eight of twenty-two hours the null is below five per cent, so the ratio has a near-zero denominator and reaches values above twenty with interval widths above fifteen. The pooled nocturnal and daytime ratios are 2.45 and 1.06 with compatibility intervals $[1.16, 4.21]$ and $[0.75, 1.52]$. We report the two directly measured quantities instead — 30.6% of onsets and 5.5% of measurements fall between midnight and six — because they need no ratio to be interesting.

# D VALIDATION OF THE THEORETICAL CLAIMS

## D.1 SIMULATION-FIRST VALIDATION OF THE THEORETICAL CLAIMS

Every claim in Section 3 was tested on synthetic processes with known ground truth before being relied on. Four results follow, two supporting and two limiting.

**The counterexample of Proposition 1, in numbers.** On a chain of persistence 0.90 with excursions of mean duration 11.2 slots, at $k = 8$ measurements over a 96-slot day: the equal-interval schedule detects 66.4% and a schedule placing all eight in consecutive slots detects 17.3%, a gap of 49.1 points. Both occupy slots whose marginal distribution is the stationary $\pi$, so the unconditional advantage $\lambda\langle\pi, q\rangle = 2.909$ is identical. With $\varepsilon = 0$ the bound of Proposition 5 is as tight as any bound can be and the gap is unchanged, so this is not a matter of bound quality.

**Proposition 5 holds.** Across 48 configurations spanning two kernel families (one deliberately non-normal), four persistence levels, two observation-noise levels and three staleness values, the measured advantage spread never exceeds $L\varepsilon(m)$. Zero violations.

**The contraction does not predict adaptive gain.** Across 37 configurations, giving the adaptive policy a search over alert threshold and staleness cap and matching budgets to within 15%, the Spearman correlation between $\varepsilon(\Delta t)$ and the gain over open-loop is $+0.070$ ($p = 0.68$); excluding the eleven configurations where open-loop already exceeds 90% detection it is $+0.197$ ($p = 0.33$). Mean gain does rise across contraction bands ($+0.6$, $+1.2$, $+3.4$ points), so the direction is right, but the relationship is not resolved. Proposition 1 explains why no such relationship should be expected.

**Verifying the premises of Theorem 15.8 is a prerequisite, not a result.** The myopic one-step value lies below the optimal value in all eight synthetic trials *including those where total positivity fails*, because discounted future rewards are non-negative. The theorem's content is that the myopic *action* is optimal on $P_s$, which we verify holds at 100% of grid points on premise-satisfying kernels and disagrees off $P_s$ in 33–94%. The premises have content only where $P_s$ is a proper subset: at a measurement cost of 8 against a miss cost of 10, $P_s$ covers 6.3% of the simplex and the optimal policy measures at 48.9% of grid points, and agreement on $P_s$ is still 100%.

**Replay identifies the retrospective functional under acquisition exogeneity.** Under Assumption 1 the estimand is a deterministic function of the recorded trajectory and the schedule, so identification needs no distributional argument. As an implementation check we replicate both sides across sixteen trajectories: differences of $-0.17$ to $+0.77$ points at 2, 3, 6 and 12 measurements per day, $|z| \leq 1.46$. Without exogeneity the bias follows Proposition 4:

| $\varphi$ | excursions | mean duration | true detection | replay overstates by |
|---|---|---|---|---|
| 0.0 | 675 | 9.30 | 51.5% | $-1.5\%$ |
| 0.1 | 638 | 8.79 | 45.3% | $+7.9\%$ |
| 0.3 | 588 | 8.24 | 36.9% | $+31.5\%$ |
| 0.6 | 510 | 7.19 | 23.4% | $+109\%$ |
| 0.9 | 449 | 6.11 | 8.1% | $+492\%$ |

### D.2 Verifying the premises of a structural theorem is not a result

The premises of Theorem 15.8 of Krishnamurthy (2025) hold on our data, but the conclusion they yield is trivial: the premises do no work in establishing it.

**The bound is trivial.** Across eight synthetic kernels the myopic one-step value lies below the optimal value in every case, *including those where total positivity fails*, because discounted future rewards are non-negative.

**The theorem's actual conclusion holds, and is sharp.** The content is that the myopic *action* is optimal on $P_s$, the set where measuring is myopically cheaper. On premise-satisfying kernels the myopic and optimal actions agree at 100% of grid points on $P_s$ and disagree off it in 33–94%.

**But the premises do not distinguish it.** Repeating the test on kernels that *violate* TP2, in five paired trials of which three are shown:

| | TP2($P$) | TP2($Z$) | myopic = optimal on $P_s$ |
|---|---|---|---|
| trial 1, TP2-satisfying | $+0.00274$ | $+0.00377$ | 100.0% |
| trial 1, TP2-violating | $+0.00274$ | $-0.07875$ | 100.0% |
| trial 3, TP2-satisfying | $+0.00122$ | $+0.00206$ | 100.0% |
| trial 3, TP2-violating | $+0.00122$ | $-0.06096$ | 100.0% |
| trial 5, TP2-satisfying | $+0.00022$ | $+0.00335$ | 100.0% |
| trial 5, TP2-violating | $+0.00022$ | $-0.07444$ | 100.0% |

**Where the theorem does have content: the region.** $P_s$ is defined as the set where measuring is myopically cheaper, so the myopic action there is "measure" by construction and the theorem asserts the optimal action is too. When measurement is cheap, $P_s$ covers nearly the whole simplex and the optimal policy measures nearly everywhere, so agreement is close to automatic:

| $\kappa$ | $\lambda_{\text{miss}}$ | $P_s$ share | optimal measures on $P_s$ | optimal measures overall |
|---|---|---|---|---|
| 0.5 | 10 | 98.7% | 100.0% | 99.7% |
| 2.0 | 10 | 85.6% | 100.0% | 99.1% |
| **8.0** | **10** | **6.3%** | **100.0%** | **48.9%** |
| 8.0 | 30 | 75.3% | 99.6% | 97.5% |

At $\kappa = 8$ the theorem is doing real work — $P_s$ is 6.3% of the simplex while the optimal policy measures at only 48.9% of grid points — and the agreement is still exact. The theorem is true and sharp. What is not a result is reporting that its premises hold.

### D.3 Verification procedure, and the errors it was built to catch

Every comparison reported in this paper passes three checks. The checks exist because each was added after a specific failure, and we describe them in that order because the failures are what make the checks legible.

**Gate 1: implementation and estimator.** Does the code compute what the text says, and is the interval estimator right for the statistic? Percentile bootstrap intervals are unreliable for ratios and for skewed clustered data, so the auxiliary intervals reported here are bias-corrected and accelerated,

resampled over patients, with acceleration from a patient-level jackknife. *Added after* a ratio was reported without an interval and then declared refuted because a percentile interval touched the null — both errors on the same number, the second being precisely what Amrhein et al. (2019) warn against.

**Gate 2: confounder matching.** Before comparing two policies, list every dimension on which they differ and confirm each is either matched or is the object of study. The recurring ones here are total budget, between-patient allocation, within-day placement, outcome base rate, and false-alarm rate. *Added after* a comparison that matched budget but not allocation attributed a gain to timing whose sign was reversed once the two were separated.

**Gate 3: mechanical null.** Compare against what a structureless process would give, and check the limiting case. A detection rate is compared against lattice coverage; a predictive score against the base rate; a value of information against its value at a vertex, which must be zero. *Added after* a proxy for value of information was used that was largest exactly where information was worthless.

**Leakage certification.** Separately, a policy is admitted only if its action sequence is unchanged when the sensor value at every unmeasured slot is replaced by uniform noise. Policies with an alert rule must additionally leave the alert sequence unchanged: a fixed, demonstrably leak-free schedule can still drive an alert computed from values that were never measured, and one experiment did exactly that before the check was extended. Two deliberately leaky policies are retained as regression tests; if either certifies, the gate is broken.

**What the gates cost and what they caught.** Each gate takes minutes to run. Between them they caught, in this project: a cold-start effect and an era effect that were artefacts of comparing scores across strata with different base rates; a timing effect with the wrong sign; an adaptive policy scoring 68% by reading sensor values it had not paid for; and a value-of-information proxy pointing the wrong way.

### D.4 The onset-concentration boundary

Proposition 1 says coverage is governed by the spacing of the chosen set. For a lattice under uniform onset, regular spacing attains the optimum: exhaustive search over all $k$-point schedules on a 24-slot cyclic horizon finds it optimal in all 20 combinations of $k \in \{2, 3, 4, 6\}$ and $D \in \{2, 3, 5, 8, 12\}$. Under a single duration the optimum is attained by many schedules — 228, 1088 and 2730 of them at $k = 2, 3, 4$ — but mixing over the cohort's empirical duration distribution breaks the ties: at $k = 4$ exactly six schedules attain the maximum, and all six are phase shifts of the regular lattice.

Onset in our cohort is not uniform ($\chi^2 = 102.6$ on 23 degrees of freedom, $p = 5 \times 10^{-12}$), so the question is how concentrated onset must be before departing from the lattice pays:

| onset concentration | regular | best | gain from departing | optimal schedule |
|---|---|---|---|---|
| 0.0 | 0.34167 | 0.34167 | 0.00000 | $(0, 6, 12, 18)$, regular |
| 0.2 | 0.34167 | 0.39500 | 0.05333 | $(0, 7, 9, 11)$ |
| 0.4 | 0.34167 | 0.49375 | 0.15208 | $(6, 7, 9, 11)$ |
| 0.6 | 0.34167 | 0.61250 | 0.27083 | $(6, 7, 9, 11)$ |
| 0.8 | 0.34167 | 0.73125 | 0.38958 | $(6, 7, 9, 11)$ |

Departure from regular spacing is worth something only when onset location is predictable. Proposition 7 makes that quantitative, and we verify it here.

**The bound on a constructed family.** Placing a fraction of the onset mass into a six-slot window, with $k = 4$, $D = 3$, so $m = 12$ and $m/n = 0.5$:

| concentration | lattice (best phase) | best schedule | realized gain | bound |
|---|---|---|---|---|
| 0.00 | 0.5000 | 0.5000 | 0.0000 | 0.0000 |
| 0.20 | 0.5000 | 0.6000 | 0.1000 | 0.1000 |
| 0.40 | 0.5000 | 0.7000 | 0.2000 | 0.2000 |
| 0.60 | 0.5000 | 0.8000 | 0.3000 | 0.3000 |
| 0.95 | 0.5000 | 0.9750 | 0.4750 | 0.4750 |

The bound is attained exactly on this family, which is a property of the family rather than of the bound: putting all the excess mass in one contiguous block means the $m$ likeliest onsets are exactly the ones a single well-placed cluster covers.

**The bound on random and empirical onset distributions.** On ten Dirichlet draws with concentration parameters $0.3$, $1$ and $3$, the bound holds in all ten with median slack $0.138$. On our cohort's own onset distribution by hour of day:

| $k$ | $D$ | realized gain | bound | slack |
|---|---|---|---|---|
| 3 | 3 | 0.0674 | 0.1645 | $+0.0971$ |
| 4 | 3 | 0.0771 | 0.1628 | $+0.0857$ |
| 6 | 2 | 0.1272 | 0.1628 | $+0.0356$ |

So on real onset data the bound is valid and roughly a factor of two loose. It is useful as a screening quantity — if the excess mass is small, no schedule can gain much, and no search is needed — rather than as a prediction of the gain.

**What is not established.** The optimality of the lattice under uniform onset is verified exhaustively at $n = 24$ and $k \leq 6$, not proved in general. Proposition 7 is proved, but its slack on realistic onset distributions means it bounds rather than characterizes the achievable gain.

# E Positioning and limitations

## E.1 Positioning against classical inspection scheduling

Barlow et al. (1963) state three assumptions and one consequence. We satisfy two of the assumptions and violate the consequence, and the violation is what Proposition 4 prices.

| Barlow, Hunter & Proschan (1963) | our setting |
|---|---|
| "system failure is known only through checking" | satisfied |
| "checking does not degrade the system" | satisfied, and it is Assumption 1 |
| "the system cannot fail while being checked" | satisfied |
| "upon detection of failure, the problem ends; specifically, no replacement or repair takes place" | **violated**: excursions resolve on their own, and clinically detection triggers treatment |
| objective: expected cost of the delay between failure and detection | objective: probability a measurement falls inside the excursion interval |
| assumes the failure density is a Pólya frequency function of order two | the same total positivity condition appears as (F1) and (F2) in the controlled-sensing literature |

## E.2 The forgetting side, in full

The empirical kernel has $|\lambda_2| = 0.896$, which would give a spectral half-life of 1.57 h; the measured contraction of the reachable set is slower, at $\tau_{1/2} = 2.25\,\mathrm{h}$. Both are properties of the three-band discretization rather than of the glucose process, and the continuous parametrization gives 10.68 h for the same discrepancy (Proposition 6). At the protocol's 3.01 measurements per patient-day

Table 1: **Paired** differences at a fixed budget of 3.01 measurements per patient-day. All policies run on the same episodes, so differences are computed per patient and bootstrapped over patients. No paired difference is resolved, and the intervals bound schedule effects at this budget below about eight points.

| paired comparison | difference | 95% BCa |
|---|---|---|
| observed protocol − equal-interval at each patient's own rate | $+2.31$ | $[-1.76, +7.31]$ |
| equal-interval uniform 3/day − observed protocol | $+0.77$ | $[-6.91, +7.16]$ |
| equal-interval uniform 3/day − equal-interval at own rate | $+3.08$ | $[-2.44, +8.46]$ |

the interval is 8 hours and $R = 3.6$. On the reading we once gave $R$, that would have suggested placement cannot matter; Proposition 1 shows the inference is invalid, and we report the measured result below instead.

American Diabetes Association Professional Practice Committee (2026)'s pre-meal schedule, implemented verbatim, gives 20.6% [15.5, 33.6] at 2.98 measurements per patient-day against this ward's 19.3% [12.3, 29.7] at 3.01, so what follows benchmarks published practice.

**At a fixed budget, placement moves detection little.** Table 1 gives paired differences at the observed budget. Rearranging measurements within the day is associated with $+2.31$ points $[-1.76, +7.31]$, in the direction of the clinical meal-anchored schedule rather than equal spacing, and reallocating them across patients with $+3.08$ $[-2.44, +8.46]$. None is resolved. The intervals do, however, bound schedule effects at this budget: the largest upper limit across the three comparisons is $+8.5$ points, against a minimum detectable difference of about 13. Raising the budget from 3 to 8 measurements per day is worth $+20.2$ points, about $2.4$ times the largest effect placement could contribute here.

**Detection tracks geometric coverage, with a constant ratio.** Equation (6) predicts a slope set by mean excursion duration alone: at 108 minutes, 7.5 points per additional measurement per day, and the measured slope is 5.3–7.1. The sharper test uses the null that matches the policy class. Our schedules are *regular*, so the relevant reference is not independent sampling but lattice coverage: for a lattice of period $n/k$ with uniform phase, an excursion of $D$ slots is covered with probability $\min(1, Dk/n)$. Against that null the ratio of observed to predicted detection is **0.95–1.10 across a sixfold budget range and on both tails** (Figure 1), with no parameter fitted. Against an independence null it would be 1.12–1.30, which measures the regularity advantage of a lattice over a Poisson process rather than a property of the problem.

**Two thresholds trade rather than rescue.** A U-shaped cost breaks monotonicity on a single glucose ordering, so we decompose into two monotone sub-problems sharing one measurement action, each satisfying the ordering conditions separately (Appendix C.1). Our first comparison ran the two-threshold policy at $2.4$ to $2.7$ checks per day against comparators at $3.0$ and reported it as tying; matching the realized rate instead, at $3.04$ per day it gains $3.3$ points on the hypoglycemic side and loses $4.3$ on the hyperglycemic side. The second threshold buys one tail with the other rather than creating detection where none was available.

Neither policy is optimal, however, and at this scale the exact frontier is computable: enumerating all $\binom{96}{3} = 142{,}880$ schedules and sweeping the weight on the hypoglycemic tail gives achievable pairs from $(9.8\%, 48.4\%)$ through $(19.1\%, 45.2\%)$ to $(31.8\%, 22.9\%)$. Both hand-designed policies lie well inside it, and the phase of a regular lattice alone moves hypoglycemic detection from $14.3\%$ to $28.9\%$ — more than any threshold rule we tested. The two tails differ in both quantities the bound of Proposition 7 depends on: measured over the full day across the 1,720 subjects of DiaData (Cinar & Maleshkova, 2025), hypoglycemic onsets carry excess mass $0.038$ with mean duration 1.03 h, against $0.118$ and 3.50 h for hyperglycemia. Restricting to the nocturnal window lowers the hypoglycemic figure to $0.0132$ (Section 4.4), since onsets are less concentrated within six hours than across twenty-four.

# F FORMULATION AND PROOFS

## F.1 PROOFS

**Proposition 5.**

*Proof.* Both beliefs lie in $P^{m\top}\mathcal{B}_0$ up to further propagation, and $\varepsilon$ is non-increasing, so $\|b - b'\|_1 \le \varepsilon(m)$. Lipschitz continuity gives the result. The second statement follows from $\varepsilon(\tau_{1/2}) = \varepsilon(0)/2$. □

**Proposition 6.**

*Proof.* Absent a measurement the Kalman mean recursion is affine, so $\hat{\mu}_1(t) - \hat{\mu}_2(t) = a^{\,t}(\hat{\mu}_1(0) - \hat{\mu}_2(0))$. The Gaussian tail probability has derivative in $\hat{\mu}$ bounded by the density maximum $1/\sqrt{2\pi v}$, giving the first inequality. □

**Proposition 1.**

*Proof.* The twelve-slot construction in Section 3 proves the statement with coverage differing by one. A second instance on empirical onsets gives the same conclusion with a smaller separation: two schedules of $k$ measurements per day — one spread at equal intervals, one placing all $k$ within a single hour — occupy slots with the same marginal state distribution $\pi$, so $\Delta$ is identical, while detection differs by 49 points (Appendix D.1). □

**Proposition 7.**

*Proof.* A schedule of $k$ points covers at most $m = kD$ onset slots, so $\mathrm{Cover}_p(Q^*) \le \sum_{j \le m} p_{(j)}$.

For the lattice, average over all $n$ rotations $\mathcal{L}_0 + \phi$, $\phi = 0, \dots, n-1$. Since $D \le \lfloor n/k \rfloor$ the $k$ windows within a single rotation are disjoint, so each rotation covers exactly $m$ slots, and the total number of (slot, rotation) coverings is $mn$. Rotation acts freely and transitively on the $n$ slots, so these $mn$ coverings are distributed equally: each slot is covered by exactly $m$ of the $n$ rotations. Hence

$$\frac{1}{n}\sum_{\phi=0}^{n-1} \mathrm{Cover}_p(\mathcal{L}_0 + \phi) \;=\; \sum_t p(t)\,\frac{m}{n} \;=\; \frac{m}{n}$$

for *any* onset distribution $p$, and the best phase attains at least the average. Subtracting gives the bound. Divisibility of $n$ by $k$ is never used: the argument needs only that the windows within one rotation are disjoint. □

## F.2 THE ACCUMULATING SIDE, IN FULL

A patient type does not mix. A type learned from a patient's first 24 hours separates their subsequent excursion rate 8.5-fold (0.090 against 0.765 per day across tertiles), and the correlation between first-half and second-half excursion rate is $+0.598$. Appendix C.2 places the resulting policy against the alternatives on the axis that matters for cost, sensor-days consumed.

Table 2 allocates an *identical* total budget by that type, against the guideline's own stratification. Korytkowski et al. (2022) list eleven high-risk criteria; eight are computable from this release, and they classify 92% of the cohort as high risk. They nonetheless help, by $+5.7$ points over uniform. A type learned from one day of the patient's own data adds a further $+11.2$, reaching 32.9% on 125 sensor-days against the 1,208 continuous alerting would require. **Sensor-days are consumable cost, not device count**: both arms still need a sensor on each of the 125 patient-periods, and what differs is how long each runs. Only the comparison against uniform allocation has disjoint intervals: learned-type allocation does not separate from observed practice or from the guideline's criteria, and we do not claim it does.

[†]Definitional: the alert threshold equals the event definition, so a continuous sensor fires at onset by construction. The meaningful metric for that arm is lead time.

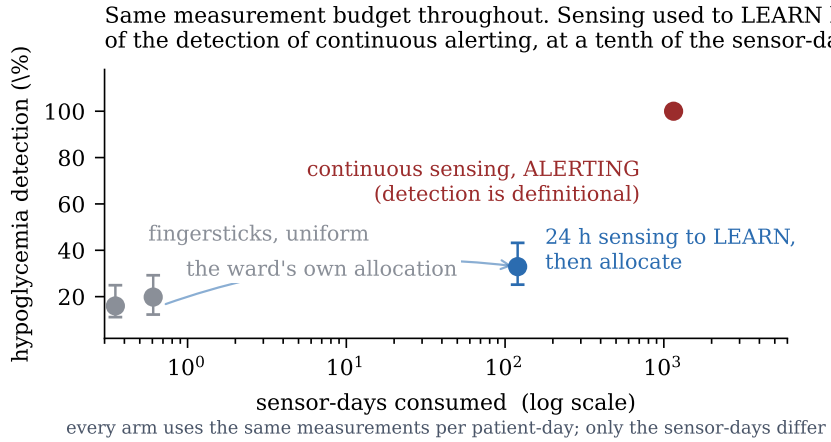



Figure 2: Detection against sensor-days consumed, at an unchanged measurement budget of 2.92 per patient-day. The guideline's own stratification is the comparator, not a uniform strawman. Bars are 95% BCa intervals resampled over patients. **The two fingerstick arms consume no sensor-days;** a logarithmic axis cannot place them at zero, so they sit at the left edge as a marker of that fact rather than at a measured abscissa.

Table 2: Allocating an unchanged measurement budget across patients. All arms at 2.92 measurements per patient-day; the type is learned from each stay's first 24 hours and scored only on the remainder.

| allocation rule | sensor-days | detection | 95% BCa |
|---|---|---|---|
| uniform | 0 | 16.0% | [11.1, 24.7] |
| observed practice | 0 | 19.8% | [12.0, 29.3] |
| guideline high-risk criteria (Korytkowski et al., 2022) | 0 | 21.7% | [16.6, 30.3] |
| **learned type, from 24 h of sensing** | **125** | **32.9%** | **[25.4, 42.7]** |
| continuous sensing, alerting | 1,208 | 100.0%[†] | — |

**Lead time separates what detection cannot.** Binary detection saturates for continuous sensing by construction, so Table 3 reports warning time instead. **Continuous sensing with a reactive threshold gives no more warning than measuring every three hours**, at a twelfth of the observation. Its value appears only when the threshold is predictive, and is bought with false alarms. A second lever appears in the last row: at a fixed budget, extending how long the alert persists more than doubles lead time with no additional measurement, working like budget by extending covered time rather than by choosing when to look.

[‡]This row's lead times are right-censored: an alert that persists for four hours covers the whole interval to the next fingerstick, so for excursions beginning inside that window the lead time is bounded below by the window rather than observed. A percentile interval on a censored quantity would understate its width, so none is given.

**The deployable version does not work, and the sensing has a price.** The same type learned from fingersticks alone is weaker ($+0.460$ against $+0.581$) and *degrades* with longer windows, because a meal-anchored protocol resamples the same clock hours rather than adding information about the tails; no fingerstick-type allocation separates from uniform. One day of admission sensing breaks even below roughly 23–31 fingerstick-equivalents at the measured budget–detection slope (Appendix C.2). We supply no tariff, only the break-even.

### F.3 Myopic versus full advantage, and tightness

**From the myopic advantage to the full one.** Proposition 5 as stated bounds the *myopic* advantage. The decision-theoretic quantity is the difference of $Q$-values, $A(\beta) = Q(\beta, 1) - Q(\beta, 0)$ as in

Table 3: Warning time before an excursion. The reactive threshold wastes continuous sensing; alert duration is a second lever at fixed budget (Appendix C.2).

| | obs/day | mean lead | 95% BCa | false alarms/day |
|---|---|---|---|---|
| continuous, alert <70 (reactive) | 96.0 | 19.2 min | [15.4, 23.3] | 2.3 |
| fingerstick q3h, alert <90 | 8.0 | 19.5 min | [15.9, 24.4] | 0.9 |
| continuous, alert <90 (predictive) | 96.0 | 65.4 min | [59.0, 72.0] | 10.5 |
| q3h, alert persisting 240 min | 8.0 | 46.4 min | ‡ | 11.2 |

Equation (3), whose first term carries the value of information and hence the POMDP structure. Because rewards are bounded and the problem discounted, $V^{\kappa}$ is convex and Lipschitz on the exact simplex, so $A$ is Lipschitz and Proposition 5 applies verbatim with a larger constant: the $2^{-R}$ decay is unchanged, which is the content, while the constant is problem-specific. On our kernel the empirical Lipschitz constant rises from $10.0$ for the myopic advantage to $32.5$ for the full one.

**Tightness, and what the constant costs.** On our kernel the myopic bound is tight to within a factor of $1.4$–$1.6$ across a sixteenfold range of $R$ (Appendix C.4). The full-advantage bound is looser — at the protocol's $R = 3.6$ it gives $4.7$ against a measured spread of $0.55$ — and becomes smaller than the measurement cost only near $R = 8$. We therefore read Proposition 5 as establishing the *rate* at which placement stops mattering, and rely on the measured spread, not the bound, for the magnitude at any particular $R$.

### F.4 Scope of Proposition 1

The proposition is stated and proved in Section 3. Its content is that equal-budget policies with identical unconditional advantage at every occupied slot can have arbitrarily different detection rates, so no bound on advantage heterogeneity constrains coverage.

### F.5 Related work, extended

**Classical inspection scheduling.** The question of when to spend a costly check on a state observable only by checking was solved for a specific objective sixty years ago. Barlow et al. (1963) minimize the expected cost of the delay between a failure and its detection; Keller (1974) and Munford & Shahani (1972) refine the schedules and Kaio & Osaki (1986) review the line. Two things separate it from ours. Their objective is detection *delay*, ours the probability that a measurement falls inside an excursion *interval*: a failure persists until detected, a glycemic excursion ends on its own. And they assume that on detection "the problem ends; specifically, no replacement or repair takes place" — the assumption that fails clinically, and which Proposition 4 prices rather than assumes away. We note one connection the literature does not appear to draw: Barlow et al. assume the failure density is a Pólya frequency function of order two, which is the total positivity condition Lovejoy (1987) and Krishnamurthy (2025) require for monotone policies. Inspection scheduling in 1963 and controlled sensing in 2025 rest on the same structural property.

**Event-triggered sensor scheduling.** The contemporary form of the same decision is scheduling transmissions under a communication budget. Zhong et al. (2024) derive the optimal MMSE estimator for multi-sensor scheduling over packet-dropping networks; Zhong et al. (2023) and Ni et al. (2023) extend the line. The objective throughout is estimation error covariance, which is a different functional of the same schedule from interval coverage.

**Adaptive inspection and maintenance.** Bismut & Straub (2021) solve inspection planning as a POMDP by direct search over heuristic policy parameters, reporting that "heuristics lead to solutions that are very close to the optimum POMDP solutions" — independent of our data, and pointing the same way as our empirical result.

**Structural results for controlled sensing.** That optimal policies for partially observed problems can be simple under ordering conditions is classical (Lovejoy, 1987). Krishnamurthy & Pareek (2015)

show that Lovejoy's cross-action conditions force either action-independent kernels or uninformative observations, and replace them with conditions checkable by linear-program feasibility; Krishnamurthy (2025) collects the modern treatment, handling action-dependent observation kernels through Blackwell dominance. We verify those premises on data rather than assuming them, and find them to hold on an abstraction where the belief carries no usable information.

**Monitoring timing in health operations.** Helm et al. (2015) solve when to monitor glaucoma through noisy tests, obtaining closed-form schedules and four structural properties; Kazemian et al. (2019) add disease control and establish a separation of estimation and control. Both require a *progressing transformation*, in which the state drifts monotonically toward an absorbing bad state — their own examples are Alzheimer's, Parkinson's, ALS and glaucoma. In our terms these are small-$R$ settings by construction: a belief on a progressing state never returns to stationarity. Our kernel violates the condition, and the recurrent case has not been given structural results.

Skandari & Shechter (2021) place the belief on a static patient type learned by monitoring, and obtain monotonicity in (health, type) with a monotone stopping policy. That is the $R = 0$ end of our axis, and it is what motivated our allocation experiment. Ayer et al. (2012) personalise mammography screening where the test never reveals the disease state exactly; Sandıkçı et al. (2013) obtain a threshold belief *vector* under the MLR partial order on a hybrid state space, showing threshold structure survives well beyond two states.

**Two-sided costs.** A U-shaped cost is canonical in operations rather than exotic. Girshick & Rubin (1952) drive a control limit from the posterior of an unobserved out-of-control state; Scarf (1960) replaces convexity with K-convexity and obtains a two-threshold $(s, S)$ policy. We searched for an analogue on the belief simplex and found none; we decompose into two monotone sub-problems instead, and note that in this problem such an analogue would certify a policy class with nothing to gain.

**Costly observation in learning.** Acquisition under a budget has been studied as active feature acquisition and as observation-cost reinforcement learning (Nam et al., 2021; Krale et al., 2023). That literature is largely instance-wise, assumes acquisition reveals a feature exactly, and reports no structural results. Both assumptions matter here: an exact reveal collapses the belief to a vertex, which is precisely the modeling choice that makes timing look worthless for the wrong reason. Zhang & Jiang (2025) show that model-free off-policy evaluation of history-dependent policies in POMDPs is statistically hard, and give a model-based route under a single-step revealing condition; our three-bin observation is chosen so that condition holds.

**Machine learning for glucose.** The CGM literature is dominated by forecasting, consolidated by Sergazinov et al. (2024), whose split protocol we adopt. The nocturnal window specifically has been modeled before: Gaynanova et al. (2022) give a multilevel functional model of overnight CGM profiles, building on the multilevel functional PCA of Di et al. (2009), and Sergazinov et al. (2023) a function-on-scalar analysis of the same period. Both describe the overnight process; neither asks when to measure it, which is the question here. That benchmark standardised how to predict given the observations one has; the question of whether to pay for an observation at all has no benchmark, and the clinical literature on scanning frequency asks how many scans per day rather than when.

### F.6 Limitations, extended

**Other limitations.** One cohort, 112 patients, two hospitals. Excursions cluster heavily enough that 519 of them behave like 98, so fine-grained results are beyond this sample and we report none. Learned-type allocation separates from uniform allocation but not from what the ward actually did, and the version learned from fingersticks alone does not work.

**What the nocturnal result rests on.** The comparison against an exhaustive baseline uses 646 held-out nights containing 93 hypoglycemic excursions. That is few: within a single split the patient-clustered interval on the gain is $[-13.3, +32.0]$ and crosses zero. We therefore do not rest the claim there: the primary estimand is the integrated frontier gap of Equation (5), whose bootstrap interval covers zero (Appendix G). The matched-count rows below are retained as a post-hoc sensitivity analysis. Budgets of five or more alarms per night are not reported here, because the matched-count

comparison needs a bisection at each budget and we ran it only to four; the enumeration itself reaches $k = 5$ (Appendix H.7). Only the two-alarm row has exactly matched alarm counts, and the larger gains at one, three and four alarms are partly bought by $0.23$ to $0.47$ extra alarms.

**The kernel is fitted, not learned.** We estimate an autoregressive kernel on training patients and then hold it fixed, which makes this a planning problem. The clinically realistic version is different: a new patient arrives with an unknown kernel, so each measurement carries a second value — information about the kernel for later decisions — that our value function does not contain. The corresponding formulation is a Bayes-adaptive POMDP, whose state carries a posterior over transition and observation parameters as well as over the physical state (Ross et al., 2011). We do not solve it. Our residual diagnostics also show the linear-Gaussian assumption is approximate rather than exact: excess kurtosis $4.03$ and a standard deviation rising from $6.8$ to $12.4$ across glucose quintiles.

**Scope of the solver.** The solver uses backward induction on a discretized belief with linear interpolation, which is adequate here because the action space is binary and the state space small enough to sweep. For continuous action spaces or larger horizons the current approach is Monte Carlo tree search with dual ascent and progressive widening (Jamgochian et al., 2023); we do not adopt it and have not tested whether it would change these results. Our two-tailed formulation at fine resolution projects to roughly ten hours by exact sweep, which is where a sampler would become the right tool. Appendix C.5 collects the rest, and Appendix C.6 reports analyses that shaped the design but are not needed for the argument.

### F.7 Replay bias under therapeutic measurement

**Proposition 4** (Replay bias under therapeutic measurement)**.** *Suppose measurement triggers treatment that ends an excursion with probability $\varphi > 0$, and let a policy be evaluated by replaying it against traces recorded under a different policy. Coverage estimated by replay is then biased, and the bias grows with $\varphi$: excursions detected under the recording policy are truncated by treatment, so the recorded population is enriched for excursions that went undetected, and any policy replayed against it inherits that enrichment.*

Measured across five feedback levels, replay overstates coverage by $7.9\%$ at $\varphi = 0.1$, $31.5\%$ at $0.3$, and $109\%$ at $0.6$. Under exogeneity replay identifies the retrospective functional; the check across four policies and sixteen trajectories ($|z| \leq 1.46$). This is the assumption Barlow et al. (1963) exclude by requiring that the problem end on detection, and it is why the blinded-sensor design reveals glucose at unsampled slots rather than reconstructing it, though only under the treatment history that was actually administered.

### F.8 Solver convergence

The belief mean is carried on a grid and the posterior after an observation is placed by linear interpolation between adjacent nodes, following the convex combination used by Poupart et al. (2015). The choice matters. With nearest-neighbour projection the value *diverges* under refinement, rising by $+0.013$, $+0.014$ and $+0.024$ as the grid doubles, because projection on a convex value function is biased and the reward $\Pr(x < \tau \mid \hat{\mu}, v)$ is steepest exactly at the threshold, which a coarse grid under-resolves. With linear interpolation, evaluating at a fixed belief rather than at whichever node is nearest:

| grid points | $V(\hat{\mu} = 70)$ | change |
|---|---|---|
| 41 | 0.853812 | |
| 81 | 0.854665 | $+0.000853$ |
| 121 | 0.854832 | $+0.000167$ |
| 161 | 0.854992 | $+0.000160$ |

The increments fall by roughly half at each refinement and the value is stable to four decimal places. One caution: `searchsorted` placed the probe at $72.500$, $72.500$ and $70.333$ for the three resolutions, comparing values at three different beliefs where $\Pr(x < \tau)$ is steep.

### F.9 MISSPECIFICATION OF THE INNOVATION MODEL

The belief is Gaussian, but the residuals are not: fitted on the training half they have excess kurtosis $4.03$ and a standard deviation rising from $6.8$ to $12.4$ across glucose quintiles, and a Student-$t$ fit returns $\nu = 2.58$ — a tail heavy enough that the variance is barely finite. Whether that matters for the comparison is a separate question from whether the model is right, so we test it directly.

We generate nocturnal traces from four innovation processes fitted to the same training patients, solve under the *unchanged* Gaussian belief in every case, and evaluate on held-out synthetic nights. The solver is therefore misspecified in three of the four settings by construction.

| innovation process | $k$ | fixed | adaptive | gap | realised |
|---|---|---|---|---|---|
| Gaussian AR(1) | 2 | 55.3% | 63.8% | +8.6 | 1.95 |
| | 3 | 67.8% | 77.0% | +9.2 | 2.87 |
| Student-$t$, $\nu = 2.58$ | 2 | 56.8% | 65.3% | +8.5 | 1.86 |
| | 3 | 70.3% | 75.4% | +5.1 | 2.69 |
| heteroscedastic, $\sigma = 0.018\,x + 6.6$ | 2 | 49.2% | 72.2% | +23.0 | 1.86 |
| | 3 | 61.1% | 83.3% | +22.2 | 2.84 |
| nonparametric residual bootstrap | 2 | 55.1% | 58.9% | +3.8 | 1.85 |
| | 3 | 62.0% | 74.7% | +12.7 | 2.89 |

The realised column is the measured rate, not the nominal budget: the price is bisected on training nights and the policy spends between $1.85$ and $2.89$ where $k$ is 2 or 3. The adaptive arm is therefore at or below the budget the fixed arm spends exactly, so these gaps are not bought with extra measurements. Two things follow, and the second is a limitation.

**The Gaussian assumption does not inflate the result.** Its gaps of $+8.6$ and $+9.2$ sit in the middle of the range, not at the top; the heteroscedastic process gives more than twice as much and the residual bootstrap less than half.

**But the magnitude is sensitive to the generating process**, ranging from $+3.8$ to $+23.0$ points across the four. The sign is stable — adaptive sensing is ahead in all eight settings — while the size is not, so the point estimates in the main text should be read as one draw from that range rather than as a transportable effect size. This is why the frontier comparison, which reports a gap at each alarm rate rather than a single number, is the form we rest on.

### F.10 SELECTION INSTABILITY OF THE EXHAUSTIVE FIXED SCHEDULE

Enumeration returns the maximizer on *training* nights, which is not the population-optimal fixed schedule. Over eight patient folds:

| | $k = 2$ | $k = 3$ |
|---|---|---|
| distinct schedules selected | 7 of 8 | 8 of 8 |
| most frequently chosen slot | 8 (6 folds) | 8 (7 folds) |
| mean training coverage | 62.7% | 74.5% |
| mean held-out coverage | 51.1% | 60.0% |
| mean oracle coverage on held-out nights | 56.6% | 71.1% |
| optimizm, training minus held-out | +11.6 | +14.6 |
| selection penalty, oracle minus held-out | +5.6 | +11.1 |

The selected schedule is unstable: at $k = 3$ no two folds choose the same one, though slot 8 (2 h) appears in seven. Training coverage overstates held-out coverage by 12 to 15 points, and the oracle schedule chosen on the held-out nights themselves would do 6 to 11 points better than the one selected on training. **The "exact optimum" is exact with respect to the training sample, not the population**, and the main text should not be read as claiming otherwise.

This does not distort the comparison. The adaptive policy's cost parameter is calibrated on the same training nights and evaluated on the same held-out nights, so both sides pay a selection penalty. **The fixed side pays the larger one**: it maximizes over $1{,}271{,}625$ candidates where the adaptive side

fits a single price, and the comparator ladder puts exhaustive selection at $+10.4$ points of train–test optimizm against $-0.0$ for a lattice that selects nothing. It does mean the absolute coverage levels reported for both arms are optimiztic relative to a population deployment.

### F.11 CONTRACTION OF THE REACHABLE BELIEF SET

This section defines the reachable post-measurement set, the contraction profile and its half-life, and states the two propositions that bound how much the advantage of measuring can vary across stale beliefs.

**The advantage of measuring is governed by the reachable belief set.** Let $\Delta(b, \tau)$ denote the advantage of measuring now over waiting, when the consequence of a measurement persists for $\tau$ slots. The optimal policy is sensitive to *where* a measurement is placed only if $\Delta$ varies across the beliefs the process actually visits. That set is much smaller than the simplex, and it is what these diagnostics are computed over.

**Definition 1** (Reachable post-measurement set)**.** *Let* $\mathcal{B}_0 = \{F(b, o, 1) : b \in \Delta(\mathcal{S}), o \in \mathcal{O}, \sigma(b, o, 1) > 0\}$ *be the set of posteriors attainable immediately after a measurement, where* $\sigma(b, o, a) = \mathbf{1}^\top \operatorname{diag}(Z(a)_{\cdot o}) P^\top b$ *is the observation likelihood.*

We do *not* assume a measurement reveals the state. On our kernel $\mathcal{B}_0$ does not reach the simplex vertices: its maximum distance from stationarity is $1.878$ in $\ell_1$ against $1.940$ for the nearest vertex (Appendix C.4).

**Definition 2** (Contraction profile and half-life)**.** *Let* $\varepsilon(t) = \sup_{b,b' \in \mathcal{B}_0} \|P^{t\top} b - P^{t\top} b'\|_1$*, the diameter of the reachable set after* $t$ *unmeasured steps. The* contraction half-life *is* $\tau_{1/2} = \inf\{t : \varepsilon(t) \le \varepsilon(0)/2\}$.

Markov kernels are non-expansive in total variation, so $\varepsilon(t+1) \le \varepsilon(t)$ and $\tau_{1/2}$ is finite whenever the filter forgets its initial condition, which holds under standard observability and ergodicity conditions (Douc et al., 2009; Tadić & Doucet, 2005; McDonald & Yüksel, 2019). We estimate $\varepsilon$ directly rather than bounding it by the second eigenvalue: on our kernel the spectral rate understates the half-life, giving $1.57\,\mathrm{h}$ against a measured $2.25\,\mathrm{h}$, because the early decay is slower than geometric.

Both figures are properties of the three-band discretization and not of the glucose process itself. Binning discards differences in the belief mean smaller than a band width, so two beliefs whose means differ by tens of $\mathrm{mg\,dL^{-1}}$ can fall in the same band and register as identical. Under the continuous parametrization of Section 2, where the belief is carried as a mean and a variance, the same discrepancy decays at the autoregressive rate $a$ and the corresponding half-life is $10.68\,\mathrm{h}$. The coarse figure is smaller by construction, and every contraction half-life reported here should be read as a statement at three-band resolution. That the same data give 2.25 h and 10.68 h depending on how the state is discretized is itself the finding: a quantity that moves by a factor of four with a modeling choice cannot serve as a decision threshold. We therefore report the contraction profile as a diagnostic of the belief representation rather than as a criterion for whether to solve, and Proposition 1 gives the separate and stronger reason why no such criterion can work.

**Definition 3** (Interval-to-half-life ratio)**.** *For a policy whose measurement intervals have distribution* $\Delta t$*, write* $R = \mathbb{E}[\Delta t]/\tau_{1/2}$*. For an irregular schedule we report the staleness distribution rather than a single ratio.*

The interpretation is direct, and it can be made precise. Write $\Delta(b) = \mathbb{E}_b[\, r(s, 1) - r(s, 0)\,]$ for the advantage of measuring at belief $b$.

**Proposition 5** (Reachable-set contraction bounds advantage heterogeneity)**.** *Suppose* $\Delta$ *is* $L$*-Lipschitz in the belief under the* $\ell_1$ *norm. Then for any two beliefs reached at staleness at least* $m$ *after a measurement,*

$$\big|\Delta(b) - \Delta(b')\big| \;\le\; L\,\varepsilon(m).$$

*If in addition* $\varepsilon(t) \le \varepsilon(0)\rho^t$ *for some* $\rho < 1$*, then with* $\tau_{1/2} = \log(1/2)/\log\rho$ *the half-life* of that decay*,* $\varepsilon(m) \le \varepsilon(0) 2^{-m/\tau_{1/2}}$ *and the bound decays geometrically in* $R$*.*

*Proof in Appendix F.1.*

Proposition 5 is stated for a belief on a finite simplex, where $\varepsilon$ is an $\ell_1$ diameter over a reachable set and must be measured. Under the parametric belief of Section 2 neither object exists, but the same statement holds with a shorter proof and an exact rate.

**Proposition 6** (Contraction under a parametric belief)**.** *Let the latent state follow $x_{t+1} = ax_t + b + \epsilon_t$ with $|a| < 1$ and let the belief be carried as a mean $\hat{\mu}_t$ and variance $v_t$. For two beliefs sharing a variance and differing only in mean, and for the reward $r(\hat{\mu}, v) = \Pr(x < \tau \mid \hat{\mu}, v)$ at any fixed threshold $\tau$,*

$$\left|r(\hat{\mu}_1, v) - r(\hat{\mu}_2, v)\right| \;\le\; \frac{|\hat{\mu}_1 - \hat{\mu}_2|}{\sqrt{2\pi v}} \;=\; \frac{a^{\,t}\,|\hat{\mu}_1(0) - \hat{\mu}_2(0)|}{\sqrt{2\pi v}}$$

*after $t$ slots without a measurement.*

*Proof in Appendix F.1.*

The contraction rate is now the autoregressive coefficient itself rather than a measured profile: no reachable set, no spectral decomposition, and no assumption beyond $|a| < 1$. On our cohort $a = 0.9839$, so a discrepancy in the belief mean halves in 10.68 h, against 2.25 h for distinguishability at three-band resolution. The two differ because binning discards mean differences smaller than a band width, and the coarser figure is the smaller by construction.

**What the proposition does not say, and why no repair is available.** It bounds how much the advantage of measuring can *differ* across stale beliefs, not the difference in value between two budget-matched schedules. This is a mismatch of objects rather than a missing lemma: the advantage $\Delta(b)$ is a marginal expectation at a single slot, whereas detection depends on the *spacing* of the chosen set relative to excursion durations. A construction with zero advantage spread and a 49-point coverage gap makes the separation exact (Appendix F.1), and a synthetic sweep confirms that contraction does not predict adaptive gain (Appendix D.1). The natural reading of Proposition 5 — that a small advantage spread renders schedules interchangeable — is therefore false.

### F.12 The reward machine, in full

**The reward is not Markov in the state.** Equation (4) pays once per excursion however many measurements land inside it, so whether the current excursion has already been paid for is a property of the action history rather than of $s_t$. We restore the Markov property by the reward machine construction of Toro Icarte et al. (2022). The automaton has states $\mathcal{U} = \{u_0, u_1, u_2\}$ — outside an excursion, inside and undetected, inside and detected — and transition

$$u_{t+1} \;=\; \delta(u_t, s_{t+1}, a_t) \;=\; \begin{cases} u_0 & s_{t+1} \notin \mathcal{E}, \\ u_2 & s_{t+1} \in \mathcal{E} \text{ and } (u_t = u_2 \text{ or } a_t = 1), \\ u_1 & \text{otherwise,} \end{cases}$$

where $\mathcal{E}$ is the excursion set, paying $+1$ on every transition *into* $u_2$ from a state other than $u_2$ — on $u_1 \to u_2$ and on $u_0 \to u_2$, the latter an excursion detected by a measurement in the slot before it begins. The reward is Markov on the product $(s, u)$, and since $\delta$ is a deterministic function of $(u_t, s_{t+1}, a_t)$, the joint filter over $(s, u)$ is a sufficient statistic of the history for this reward.

One consequence matters and we state it as a limitation rather than resolve it. Because $\delta$ depends on the latent $s_{t+1}$, the automaton state is unobserved, so the belief lives on $\Delta(\mathcal{S} \times \mathcal{U})$ rather than on $\Delta(\mathcal{S}) \times \mathcal{U}$. Our solver carries the marginal belief over $s$ together with $\Pr(u = u_2)$, which is exact only when the two are conditionally independent given the observation history. That does not hold in general, so the implementation is an approximation of the joint filter and we label it as such; Appendix F.8 reports the resulting policy, and Li et al. (2024) show that lacking direct access to the labeling function can change optimal behavior under partial observability, though not when the state is fully observed.

### F.13 Frontier gaps at pre-specified alarm rates

| alarms/night | mean gap | sd | min | max |
|---|---|---|---|---|
| 1.5 | $+14.5$ | 6.7 | $+8.6$ | $+25.1$ |
| 2.0 | $+13.4$ | 7.0 | $+7.8$ | $+24.9$ |
| 2.5 | $+15.4$ | 4.9 | $+10.7$ | $+22.9$ |
| 3.0 | $+14.9$ | 3.2 | $+10.8$ | $+18.2$ |
| 3.5 | $+12.8$ | 3.1 | $+8.0$ | $+16.3$ |
| 4.0 | $+9.7$ | 4.4 | $+4.6$ | $+15.3$ |

Positive at all thirty rate-by-split combinations; mean $+13.5$ points.

### F.14 Random durations in Proposition 7

The bound is stated per $D$ and does not survive substituting a single $m$. Averaging the per-$D$ bounds is valid, since the inequality holds conditionally; replacing $m$ by $\lceil k\,\mathbb{E}[D] \rceil$ or by $\mathbb{E}[\min(kD, n)]$ is not. In a search over 300 random onset distributions with two-point duration mixtures, both substitutions were violated in 10 cases, the largest excess being $0.009$. We therefore report $\mathbb{E}_D\big[\sum_{j \le kD} p_{(j)} - kD/n\big]$, the expectation of the per-$D$ bound, wherever durations vary.

## G Uncertainty for the frontier gap

The thirty rate-by-split cells in Section 4.4 are descriptive: five overlapping patient splits and six correlated rate points are not thirty independent replications. The inference the comparison needs is a bootstrap of the entire pipeline with the patient as the resampling unit, and we implement it.

**Design.** Each replicate redraws patient identifiers with replacement. The train/test split is then taken over *unique* identifiers, with each assigned wholly to one side and its multiplicity following it, so no patient's nights appear on both sides; an assertion in the code enforces this. Within the replicate we rerun the autoregressive fit, exhaustive selection of the fixed schedule, price calibration, rollout of both policies, and frontier interpolation. Selection and calibration variability therefore enter the interval rather than being conditioned away. Acceleration for a bias-corrected and accelerated interval comes from leave-one-patient-out reruns of the same pipeline; the simultaneous band would be the sup-$t$ band of Montiel Olea & Plagborg-Møller (2019), the narrowest in its family attaining simultaneous coverage.

**Determinism and failures.** Every replicate and every jackknife fold is seeded individually — replicate $b$ by $\text{SEED} \times 10^5 + b$, fold $i$ by $\text{SEED} \times 2 \times 10^5 + i$ — so a run resumed across sessions reproduces what an uninterrupted one would, and a fold recomputed later is identical. Schedule ties in the exhaustive search are broken by strict improvement, so the first combination attaining the maximum in enumeration order is kept; the enumeration order is fixed, so ties resolve identically across replicates. Every completed unit is written as it finishes, and units that fail to produce a frontier at all reported rates are recorded rather than silently dropped. In the runs completed so far, no fold or replicate failed and none produced a missing rate.

**Replicate count and endpoint stability.** Efron & Tibshirani (1993) give 50–200 replicates as adequate for standard errors and at least 500–1000 for percentile intervals. We ran 2,000. Because the endpoints of a percentile interval are themselves estimated, we report how they move as $B$ grows rather than asserting that 2,000 suffices:

| $B$ | usable | 2.5% | 97.5% | MC error |
|---|---|---|---|---|
| 200 | 186 | $+2.5$ | $+38.1$ | 0.72 |
| 500 | 466 | $+0.6$ | $+39.9$ | 0.48 |
| 1000 | 929 | $-0.6$ | $+39.6$ | 0.34 |
| 2000 | 1837 | $-2.0$ | $+39.8$ | 0.25 |

The upper endpoint is stable from $B = 500$ onward, moving by less than its Monte Carlo error. **The lower endpoint is not: it fell by** $1.5$ **points between** $B = 1000$ **and** $B = 2000$**, against a Monte Carlo error of** $0.25$**.** That is a left tail still being resolved, and it is the reason we describe the interval as covering zero rather than quoting $-2.0$ as a settled bound.

**Sensitivity to the grid and to the frontier rule.** $\widehat{\Psi}$ is a trapezoid area over six points, so it depends on the grid and on what counts as the frontier. Both were fixed before any comparison; we report the sensitivity so the choices can be judged.

| grid | mean $\widehat{\Psi}$ | $2.5\%$ | $97.5\%$ |
|---|---|---|---|
| all six rates (pre-specified) | 18.3 | $-2.0$ | $+39.8$ |
| every other rate | 19.1 | $-1.5$ | $+41.0$ |
| endpoints only | 17.1 | $-2.0$ | $+37.5$ |
| dropping the lowest rate | 17.8 | $-2.2$ | $+39.5$ |
| middle four rates | 18.7 | $-2.0$ | $+41.3$ |

The estimate moves between $17.1$ and $19.1$ across grids and every interval still covers zero, so the six-point choice is not doing the work.

**On the frontier rule.** A defensible variant applies a monotone upper envelope to each *frontier* before differencing, on the ground that a class attaining a detection level at one rate can attain it at a higher one. Computing it needs both frontiers per replicate, which the bootstrap stores:

| frontier rule | mean | 2.5% | 97.5% | positive |
|---|---|---|---|---|
| raw (reported in the body) | 18.3 | $-2.0$ | 39.8 | 96.4% |
| monotone envelope per arm | 18.3 | $-2.0$ | 39.7 | 96.4% |
| upper concave hull per arm | 17.5 | $-2.6$ | 38.6 | 96.1% |

The three rules agree to within $0.8$ points and none excludes zero, so the result does not turn on which reading of the frontier one adopts. The concave hull is the strictest of the three because it is what a decision maker who may randomize between adjacent solved policies would face, and it moves the estimate down rather than up. Imposing monotonicity on the *gap* instead is a different operation with no such justification — the gap may legitimately shrink as budget grows, and in our data it does, from $+20.4$ points at $1.5$ alarms to $+13.2$ at $4.0$ — so we do not report that either.

**Replicates that do not yield all six rates.** A replicate contributes at a rate only if both classes attain it, since we do not extrapolate. At $1.5$ alarms per night $163$ of $2{,}000$ replicates fall short, at $2.0$ alarms $32$, at $2.5$ alarms $4$, and none at the three higher rates; $1{,}837$ replicates yield all six and enter $\widehat{\Psi}$. The shortfall is concentrated where the adaptive class must be driven to its sparsest, which is where a resampled cohort is most likely to leave the rate unattainable. We report these counts rather than dropping them silently, and no replicate failed to run.

**What the estimates show.** The gap is positive at all six pre-specified rates, from $+13.2$ to $+20.4$ points on the bootstrap mean, and positive in $90$ to $96\%$ of replicates depending on the rate. The integrated estimate is $\widehat{\Psi} = 14.2$, computed once on the cohort as observed, with a $95\%$ percentile interval of $[-2.0, +39.8]$ and $96.4\%$ of replicates positive. **The replicates average** $18.3$**, not** $14.2$: resampling shifts the estimate upward by $4.1$ points, the bootstrap's estimate of its own bias. Tables summarizing replicates therefore carry the larger figure; the body quotes the sample estimate. The sup-$t$ simultaneous band of Montiel Olea & Plagborg-Møller (2019) has a critical value of $2.31$ against a pointwise $1.96$ and does not exclude zero at any rate.

**We therefore do not claim population superiority.** The interval covers zero, and per Greenland et al. (2016) what that reports is imprecision rather than absence of an effect: the data are compatible with anything from a small disadvantage to a gain of nearly forty points. The point estimates favor the adaptive procedure at every rate and in the large majority of resamples; $112$ patients do not settle by how much.

### G.1 WHY WE DO NOT CORRECT THE CLASS IMBALANCE

Hypoglycemic excursions are rare, and a natural instinct is to oversample them. The methodological literature is against it. Van den Goorbergh et al. (2022) compared no correction, random undersampling, random oversampling and SMOTE, and found that all three corrections gave *poorly calibrated* models with strongly overestimated minority-class probabilities and no gain in discrimination; at a one percent event fraction the median calibration intercept fell below $-4.5$ against $-0.05$ to $0.03$ without correction. Their conclusion is that outcome imbalance is not a problem in itself and that correcting it can make performance worse. What improvement the corrections did produce in the sensitivity-specificity balance was obtainable by shifting the decision threshold instead. Carriero et al. (2025) extend the finding to machine-learning models.

Our situation makes the case stronger rather than weaker. The estimand is coverage of excursions, not a predicted probability, so there is no class ratio to correct. And duplicating excursions would reuse nights from the same patients, manufacturing independence that the data do not contain — the opposite of what resampling patients is for.

We instead follow the small-sample practice these authors recommend: decide the required sample from a precision target rather than repairing a small one after the fact (Riley et al., 2018), and validate internally by a bootstrap that re-runs every model-building step inside the resample (Steyerberg & Harrell, 2016), which is what Appendix G does.

### G.2 HOW MUCH THE FACTORISED REWARD-MACHINE BELIEF COSTS

The solver carries a marginal belief over glucose together with a scalar $\Pr(u = u_2)$, rather than a joint belief on $\mathcal{S} \times \mathcal{U}$. That factorisation is exact only under a conditional independence which does not hold, so it cannot be proved; it can be measured. On a discrete state space small enough that the exact joint filter is computable — twelve glucose bins, three automaton states, sixteen steps — we run both filters on the same trajectories and compare.

| mixing $\rho$ | measure prob. | TV on $x$ | mean $\lvert\Delta \Pr(u_2)\rvert$ | max |
|---|---|---|---|---|
| 0.70 | 0.15 | 0.000 | 0.018 | 0.092 |
| 0.70 | 0.35 | 0.000 | 0.032 | 0.173 |
| 0.85 | 0.15 | 0.000 | 0.020 | 0.072 |
| 0.85 | 0.35 | 0.000 | 0.032 | 0.165 |
| 0.95 | 0.15 | 0.000 | 0.026 | 0.091 |
| 0.95 | 0.35 | 0.000 | 0.026 | 0.118 |

The glucose marginal is exact, since both filters propagate $x$ identically; the approximation lives entirely in the detected-flag probability, which is what the reward reads. Its mean absolute error is $0.018$ to $0.032$ and its worst case $0.17$, and the error grows with the measurement rate rather than with mixing speed — measuring is what couples $x$ and $u$, so measuring more often makes the independence assumption less tenable. We report this rather than assume it away: the factorisation is a genuine approximation whose cost is small at the measurement rates we operate at (one to four per night, well below the $0.35$ per slot in the worst row) but is not zero.

### G.3 WHY WE MEASURE THE APPROXIMATION INSTEAD OF BOUNDING IT

The natural formal route is the approximate information state framework of Subramanian et al. (2022). Their Theorem 13 states that if $\{\hat{\sigma}_t, \hat{P}_t, \hat{r}_t\}$ is an $(\varepsilon_t, \delta_t)$-AIS generator, then the value computed on the compressed state differs from the true value by at most $\alpha_1 = \sum_{t=1}^{H} \big(\varepsilon_t + \rho_F(\hat{V}_{t+1})\,\delta_t\big)$, where $\varepsilon_t$ bounds the reward approximation and $\delta_t$ the dynamics approximation.

We can measure both on the discrete space of Appendix G.2, where the exact joint filter is computable. The dynamics term is *exactly* zero: both filters propagate the glucose marginal through the same kernel, so $\delta_t = 0$ for all $t$. The reward term has mean $0.029$ and supremum $0.579$ over $60$ trajectories.

**The resulting bound is valid and useless.** With $H = 16$ and $\delta \equiv 0$, taking the supremum gives $\alpha_1 \le 16 \times 0.579 = 9.26$, against a reward bounded by $1$ per excursion. The mean error would

give $\alpha_1 \leq 0.46$, inside the useful range, but the theorem takes a supremum. A useful bound needs $\varepsilon < 1/H = 0.0625$ uniformly; our mean of $0.029$ is comfortably under it while our supremum is twenty times the mean.

**What this is and is not.** It is not a limitation of our simulation: we measured $\varepsilon$ and $\delta$ exactly. It is not a limitation of the cohort: the bound uses no data at all, and more patients would not change it. It is the worst-case form of the bound meeting a heavy-tailed per-step error. We therefore report the measured error distribution (Appendix G.2) rather than a bound that holds vacuously. The one genuine gap is that $\varepsilon$ is measured on twelve glucose bins, since the exact joint filter is not computable at clinical resolution, so we do not know that the same distribution holds there.

### G.4 Calibration of the lower-tail probability

Detection is driven by $\Pr(x < \tau \mid \hat{\mu}, v)$, so the Gaussian belief is being asked for exactly the quantity where its assumption is weakest. Coverage results cannot settle this: a policy can detect well while the probabilities it states are wrong. We therefore measure calibration directly on held-out patients, binning the stated probabilities and comparing each bin's mean against the fraction of slots in it where glucose really was below $70\,\mathrm{mg\,dL^{-1}}$.

| stated probability | slots | mean stated | observed |
|---|---|---|---|
| 0.00–0.02 | 8188 | 0.003 | 0.001 |
| 0.02–0.05 | 1929 | 0.033 | 0.005 |
| 0.05–0.10 | 1717 | 0.072 | 0.018 |
| 0.10–0.20 | 1696 | 0.144 | 0.037 |
| 0.20–0.35 | 1078 | 0.262 | 0.072 |
| 0.35–0.60 | 493 | 0.453 | 0.556 |
| 0.60–1.00 | 403 | 0.785 | 0.906 |

Two-hourly measurement; the four- and six-hourly schedules give the same pattern. The logistic recalibration intercept is $-0.16$ with slope $1.38$ (four-hourly $-0.16$ and $1.38$, six-hourly $-0.40$ and $1.09$), and the Brier score is $0.030$ to $0.036$.

**The belief is not well calibrated, and the direction is systematic.** In the middle range it *overstates* risk by roughly fourfold — a stated $0.26$ corresponds to an observed $0.07$ — while above $0.35$ it *understates*, stating $0.45$ where $0.56$ occurs. A slope above one says the stated probabilities are not extreme enough; the negative intercept says they are too high on average. This is what a Gaussian belief does against heavy-tailed, heteroscedastic residuals: it spreads mass into a middle range the process rarely occupies.

**What this does and does not affect.** A monotone distortion cannot reorder two states at a single decision, which is why detection can be good while calibration is poor. **That argument is one-step.** The solver is a backward induction and compares continuation values, which are expectations over stated probabilities, and a nonlinear monotone map does not commute with expectation: under the logistic fitted here, outcome sets with means $0.382$ and $0.348$ recalibrate to $0.332$ and $0.335$, reversing the order. What we claim empirically is the detection measured on held-out patients, not that the solved policy is optimal for the recalibrated process. It does mean the stated probabilities should not be read as risks by a clinician, and that a deployment wanting calibrated risk output would need a recalibration layer — a two-parameter logistic map would suffice, using the slope and intercept above. We report the miscalibration rather than leaving it for a reader to discover.

### G.5 The counterexample behind Proposition 1, in full

The proposition asserts that two equal-budget schedules can have an identical one-step advantage at every reachable belief while covering very different amounts of excursion mass. The construction is explicit.

Take a process whose latent belief is time-homogeneous, so that the belief reached at a given staleness does not depend on when the staleness began, and let both schedules use the same number of measurements. Because the latent belief is time-homogeneous and the advantage $\Delta(b) = \lambda\langle \pi, q\rangle$ depends only on the belief and not on the slot index, every schedule with the same budget induces the

same multiset of advantages: the *spread* is exactly zero. In our instance the advantage is $2.909091$ for both schedules and the spread is $0.000000$ to six decimal places.

Detection, however, depends on where the measurements sit relative to onsets. The spread-out schedule detects $66.4\%$ of excursions and the clustered one $17.3\%$, a gap of $49.1$ percentage points at zero advantage spread.

A smaller construction makes the same point with less arithmetic. Take $n = 12$ slots on a circle, $k = 2$ measurements, deterministic duration $D = 3$, and onset mass $\frac{1}{2}$ on each of slots $0$ and $1$. The schedule $\{2, 8\}$ covers both onsets and attains coverage $1$; the schedule $\{5, 11\}$ covers neither and attains $0$. Both use the same budget and, under a time-homogeneous latent belief, present the same one-step advantage at every slot. The gap is a full $100$ points.

The lesson is the same at either size. A bound on advantage *heterogeneity* constrains how much the value of measuring varies across beliefs; coverage depends on the *spacing* of the chosen set relative to excursion durations. Driving the first to zero leaves the second untouched.

### G.6 Statement of Proposition 7

**Proposition 7** (Excess-mass bound on the value of departing from a lattice)**.** *Let $\mathcal{L}_0 = \{\lfloor in/k \rfloor : i = 0, \dots, k-1\}$ be the near-uniform lattice, whose consecutive gaps are $\lfloor n/k \rfloor$ or $\lceil n/k \rceil$, and fix a deterministic duration $D \le \lfloor n/k \rfloor$ so that its coverage windows do not overlap. Set $m = kD \le n$.* No divisibility assumption is needed. *For a schedule $Q$ of $k$ slots let $C_D(Q) = \bigcup_{q \in Q} \{q - D + 1, \dots, q\} \bmod n$ be the set of onsets it covers, so $|C_D(Q)| \le m$. Let $\mathcal{L}$ be the regular lattice at its best phase and $Q^*$ a coverage-maximizing schedule of $k$ slots. Then*

$$\mathrm{Cover}_p(Q^*) - \mathrm{Cover}_p(\mathcal{L}) \;\le\; \sum_{j=1}^{m} p_{(j)} \;-\; \frac{m}{n},$$

*where $p_{(1)} \ge \cdots \ge p_{(n)}$ are the sorted onset probabilities.*

### G.7 Structural conditions, verified

With a binary state and a three-bin observation, every premise of Theorem 15.8 of Krishnamurthy (2025) holds on the data: $P$ is TP2 at every lag (minimum minor $+0.000807$, BCa $[+0.000566, +0.001073]$, holding in 99.3% of patient-resampled draws and in 5/5 training folds); the observation kernel is TP2 and full rank, hence single-step revealing in the sense of Zhang & Jiang (2025); copositive dominance holds with equality by Assumption 1; Blackwell dominance is exact by construction; and both linear programs are feasible (Appendix C.3).

Two observations follow from doing this on data rather than on a constructed example. First, the region where the myopic bound coincides with the optimum covers 98.8% of the uniform simplex but only 12.8% of the beliefs the process reaches — a gap that persists under every policy we tried, including never-measure and always-measure. **The tightness of an approximation bound should be assessed on the reachable belief measure.** Second, total positivity holds at every lag with two states and at none with eleven, while out-of-fold AUPRC rises from 0.514 [0.426, 0.598] to 0.703 [0.604, 0.774], disjoint; estimation error contributes a belief band three orders of magnitude smaller than that gap. **The abstraction on which the conditions hold is the abstraction on which the belief carries no usable information.**

### G.8 The forgetting side, expanded

On the moving glucose state $R = 3.6$, which on the reading we once gave it would have suggested placement cannot matter. Measurement agrees with that conclusion here, but the inference is invalid in general (Proposition 1): at a fixed budget, rearranging measurements within the day is worth $+2.31$ points $[-1.76, +7.31]$ by paired comparison, and detection behaves as lattice coverage, with the ratio to the correct combinatorial null within $0.95$–$1.10$ across a sixfold budget range. Two thresholds trade one tail against the other rather than creating detection. Appendix E.2 gives the full analysis, including the exact frontier over all $\binom{96}{3} = 142{,}880$ schedules and the two-tail onset statistics from DiaData.

### G.9 Enumerability and the nocturnal result, in one figure

Figure 3 shows both facts the nocturnal experiment turns on: how fast the fixed-schedule class outgrows enumeration, and what the two policy classes detect at each alarm budget on held-out patients.

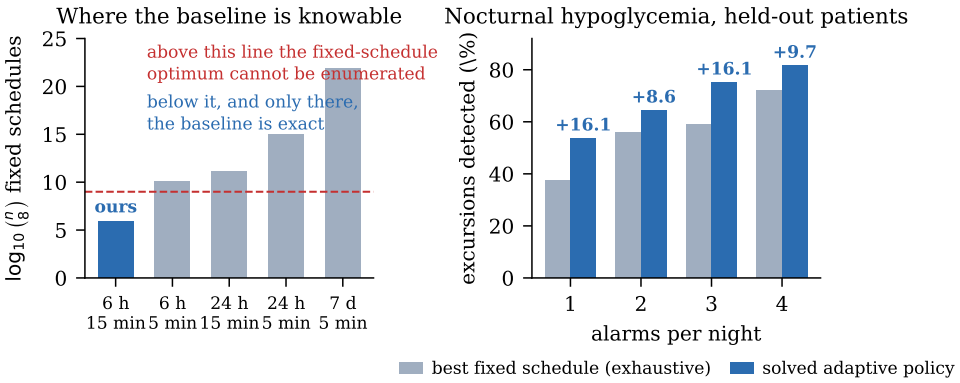



Figure 3: **Left:** the number of fixed schedules of eight measurements grows past enumeration almost immediately, so the exact optimum of that class is knowable only in a narrow window. A six-hour night at fifteen-minute resolution is the largest such setting we found. **Right:** on held-out patients, the solved adaptive policy against the exhaustive fixed-schedule optimum at each alarm budget. Only the two-alarm column has a matched alarm count. At one, three and four alarms it fires $0.23$ to $0.47$ extra alarms, so part of those gains is bought.

### G.10 The solver, in pseudocode

Algorithm 1 gives the procedure summarized in Appendix H.6.

**Algorithm 1** Solve, or decide it is not worth solving

**Require:** historical trajectories; a horizon of $H$ slots; a per-measurement price $\kappa$ (Equation (2))

fit $x_{t+1} = ax_t + b + \epsilon_t$ and the observation noise on the training fold
$E \leftarrow \mathbb{E}_D\big[\sum_{j \le kD} p_{(j)} - kD/n\big]$ where $p$ is the onset distribution ▷ Proposition 7; substituting $k\,\mathbb{E}[D]$ is not valid
**if** $E$ is small **then**
    within the fixed class, use the best lattice: no schedule beats it by more than $E$ ▷ says nothing about the adaptive class
**end if**
$v_0 \leftarrow \sigma^2_{\text{obs}}$; tabulate $v_g$ for staleness $g = 0, \dots, G$ ▷ deterministic given the actions
**for** $h = H - 1$ **down to** $0$ **do**
    **for** each $(\hat{\mu}, g)$ on the grid, with $p_{\text{det}}$ derived **do**
        $Q_1 \leftarrow \Pr(x < \tau \mid \hat{\mu}, v_g)\,(1 - p_{\text{det}}) - \kappa + \mathbb{E}_y[V_{h+1}]$ ▷ quadrature over the observation
        $Q_0 \leftarrow V_{h+1}(\hat{\mu}', g{+}1, p'_{\text{det}})$
        $V_h \leftarrow \max(Q_0, Q_1)$; $\phi_h \leftarrow \arg\max$
    **end for**
**end for**
sweep $\kappa$ and record the realized rate; the budget is met in expectation, not pathwise
evaluate on held-out patients against the exhaustive fixed-schedule optimum

### G.11 Tightness of the bound in Proposition 5

On the estimated kernel $|\lambda_2| = 0.89579$, so the spectral half-life is $6.30$ slots ($1.57\,\text{h}$); the identity $|\lambda_2|^{\tau_{1/2}} = 1/2$ checks to four decimal places. The measured contraction of the reachable set is

slower, at 9 slots (2.25 h), because the early decay is sub-geometric. The empirical mixing constant, taken as the maximum of $\|b_t - \pi\|_1/|\lambda_2|^t$ over vertices and over $t \le 60$, is $C = 2.640$. With the reward and price used in the body the advantage is linear in the belief, giving $L = 10.0$.

| $R$ | staleness (h) | bound $LC2^{-R}$ | measured variation | slack |
|---|---|---|---|---|
| 0.5 | 0.75 | 18.66 | 13.32 | 1.4× |
| 1.0 | 1.50 | 13.20 | 9.13 | 1.4× |
| 2.0 | 3.00 | 6.60 | 4.50 | 1.5× |
| 3.0 | 4.75 | 3.30 | 2.03 | 1.6× |
| 5.2 | 8.00 | 0.72 | 0.47 | 1.5× |
| 8.0 | 12.25 | 0.10 | 0.07 | 1.5× |

The slack is essentially constant at 1.4–1.6 across a sixteenfold range of $R$, so the bound is not merely valid but tight up to a small constant. The constant is attributable to $C$, which we take as a worst case over vertices.

**Two further statements that are proved rather than measured.** First, under Assumption 1 copositive dominance holds *with equality*: since $P(0) = P(1)$, both sides of the condition are the same matrix. Second, continuous sensing Blackwell-dominates any intermittent schedule slot by slot. At an unmeasured slot the intermittent kernel maps every state to one null observation, so taking $M$ to be the $|\mathcal{O}| \times 1$ all-ones column, which is stochastic, gives $(Z_{\text{cont}}M)[s, \text{null}] = \sum_o Z_{\text{cont}}[s, o] = 1$ for every $s$; at a measured slot $M = I$. The garbling therefore acts on the output side, which is the direction the definition requires, and the reverse factorisation would not suffice.

### G.12 Where the belief is placed

$R$ is not a property of a disease. It is a property of the pair (state variable, measurement interval), and the state variable is a modeling choice. The same patient admits beliefs with radically different half-lives:

- A belief on the **current physiological state** mixes at the rate of the physiological process. It *forgets*.
- A belief on a **static patient type** — a parameter that does not evolve — has $|\lambda_2| = 1$ and hence $\tau_{1/2} = \infty$, so $R = 0$ for any interval. It *accumulates*: every observation sharpens it and none of the sharpening decays.

This is a modeling observation rather than a decision rule. Where timing appears not to help, it is worth asking whether the belief has been placed on a quantity that forgets — but Proposition 1 shows that the answer does not settle whether a better scheduler exists, so the two questions must both be asked. Skandari & Shechter (2021) place the belief on a patient type and obtain monotone structure; Helm et al. (2015) place it on a progressing disease state, which by construction does not return to stationarity, and obtain closed-form schedules. Both sit at small $R$. Our setting, with the belief on a recurrent physiological state, sits at large $R$, and behaves accordingly.

## H Related work

**Scheduling under an observation cost.** Inspection scheduling begins with Barlow et al. (1963), who minimize expected detection delay for a process observable only by checking, and whose assumption that the problem ends on detection is the one clinical monitoring violates. Event-triggered sensing (Zhong et al., 2024; Ni et al., 2023) and adaptive inspection (Bismut & Straub, 2021) relax the fixed schedule; Krishnamurthy (2025) gives the structural conditions under which a myopic or threshold policy is optimal in controlled sensing, which we verify rather than assume. In health operations the timing question appears as screening and surveillance interval design (Helm et al., 2015; Kazemian et al., 2019; Ayer et al., 2012). Costly-observation formulations in learning (Bellinger et al., 2021; Wang et al., 2024) price the measurement but retain a per-step reward, and glucose-specific machine learning (Wang et al., 2024) predicts the trajectory rather than choosing when to look.

**Sequential detection with a cost of observing.** The closest formal setting is quickest change detection under an observation constraint. Banerjee & Veeravalli (2014) treat exactly the combination we need — a cost on observations together with finitely many possible post-change distributions — and Nath & Wu (2020) generalize to several post-change models, of which our hypoglycemic and hyperglycemic tails are the two-alternative case. Oleyaeimotlagh et al. (2023) add statistical periodicity, which a daily glucose cycle has, and formulate both joint detection-and-classification and a multislot problem in which "the change occurs in only a subset of the $T$ time slots in each period" — precisely our onset structure.

Two differences remain. Their observation constraint is an average fraction of time sampled *before the change*, so that dropping a fixed fraction of samples costs nothing asymptotically; ours binds throughout, since a ward has a fixed measurement rate whether or not anything is happening. And their estimand is detection delay under a false-alarm constraint, whereas ours is interval coverage: an excursion ends on its own whether or not it is detected, so there is no stopping time to optimize. That second distinction separates us from Barlow et al. (1963) in the same way, and we arrived at it independently in both literatures.

**Clinical evidence does not settle alert design.** Consensus documents standardise what to measure and how to report it (Battelino et al., 2019; International Hypoglycaemia Study Group, 2017; Kong et al., 2023), but none prescribes an alert threshold or how long an alert should persist. Giza et al. (2025) state the position directly: "there are a few data concerning the frequency, consequences and methods of eliminating alarm fatigue", and note that alarms drive patients to abandon sensing, which is what makes alarms-per-night a real constraint rather than a modeling convenience; the Asia-Pacific consensus lists alarm fatigue and hospitalised patients together as future work (Kong et al., 2023). The one inpatient trial does not fill the gap. Hirsch et al. (2025) randomized hospitalised adults with type 2 diabetes to a standard or an intensive glycemic target guided by continuous sensing. They found no difference in mean glucose — 170 against $175\,\mathrm{mg\,dL^{-1}}$, adjusted difference $-7$ with $[-19, +5]$ — and only $7\%$ of the intensive arm reached its target. It compared *targets*, using the sensor as the instrument, and did not vary alert thresholds, persistence, or an alarm budget; its null result therefore constrains what target intensification achieves, not what alert design achieves.

If placement carried no information, detection would reduce to timeline coverage, $k\,\mathbb{E}[D]/n$ under independent placement and $\min(1, kD/n)$ for a lattice. We use these as reference quantities against which measured detection is compared (Equation (6), Appendix H.1); they are a null model, not a consequence of any diagnostic.

### H.1 The coverage null model in full

$$\text{detection rate} \;\approx\; \frac{k\,\mathbb{E}[D]}{n} \tag{6}$$

If placement carried no information, detection would reduce to how much of the timeline a budget covers: independent placement of $k$ measurements over $n$ slots gives $1-(1-k/n)^D$ for an excursion of duration $D$, which for small $k$ linearises to detection rate $\approx k\,\mathbb{E}[D]/n$ (Equation (6)), while a regular lattice with non-overlapping windows covers exactly $\min(1, kD/n)$. That is a null model, not a consequence of any diagnostic. We use it as a reference quantity: Section 4 reports the ratio of observed detection to it across budgets and both tails, and a ratio near one says the data are consistent with placement carrying little information at that budget. It predicts nothing about the adaptive class.

### H.2 Three resources, three symbols

Write $M$ for measurements acquired, $A$ for alerts emitted and $D$ for excursions detected. Here a single action acquires an observation and detection is credited when it falls inside an excursion, so $A = M$: there is no separate alerting step. We report rates in $M$ per night and use "alarms" only where the clinical framing demands it, meaning the same action. That identity is a property of this resource model; under continuous sensing with selective alerting $M$ is fixed and $A \ll M$, and our results do not transfer there unchanged.

### H.3 LEAKAGE UNDER A NAIVE SPLIT: NUMBERS

**Proposition 8** (Leakage under a naive resampling split)**.** *Resample $n$ units with replacement and split the resampled list into fractions $f$ and $1-f$. Say a unit* straddles *if at least one copy falls on each side. As $n \to \infty$ the expected fraction of distinct units that straddle converges to*

$$L(f) \;=\; 1 - e^{f-1} - e^{-f} + e^{-1},$$

*which at $f = \frac{1}{2}$ equals $1 - 2e^{-1/2} + e^{-1} \approx 0.155$.*

*Proof of Proposition 8.* A unit with $k$ copies avoids straddling exactly when all $k$ land on one side, with probability $f^k + (1-f)^k$. The multiplicity of a fixed unit converges to $\mathrm{Poisson}(1)$, and for $x \in [0,1]$, $\sum_{k\geq 2} e^{-1}x^k/k! = e^{x-1} - e^{-1} - xe^{-1}$. Applying this at $x = 1, f, 1-f$ and collecting terms gives $L(f)$. The restriction to $k \geq 2$ matters: a unit with fewer than two copies cannot straddle. □

The result is verified three ways in `src/leakage_bound.py`: the closed-form limit, the exact finite-$n$ sum over the binomial multiplicity distribution, and direct simulation.

| $n$ | $f$ | exact | limit | simulated |
|---|---|---|---|---|
| 50 | 0.5 | 0.1542 | 0.1548 | 0.1573 |
| 112 | 0.5 | 0.1545 | 0.1548 | 0.1558 |
| 500 | 0.5 | 0.1548 | 0.1548 | 0.1552 |
| 1000 | 0.5 | 0.1548 | 0.1548 | 0.1550 |
| 112 | 0.7 | 0.1302 | 0.1305 | 0.1322 |
| 1000 | 0.7 | 0.1304 | 0.1305 | 0.1307 |

The largest discrepancy between the exact value and simulation is $0.0035$. A naive split on this cohort gives a measured crossing of $17.4$ patients against the $17.3$ this proposition predicts. **The denominator is the** $112$ **distinct patients**, not the resampled draws: $L(f)$ is a fraction of the units that exist, and a unit drawn several times straddles or does not exactly once.

### H.4 SUBMODULARITY AND THE GREEDY COMPARATOR: NUMBERS

*Proof of Proposition 2.* $C$ is a weighted coverage function: each pair $(t, d)$ is an element of weight $q(t, d)$ with covering set $W(t, d)$. Monotonicity is immediate. For $A \subseteq B$ and $x \notin B$, the windows newly covered by adding $x$ to $A$ are those meeting $x$ but not $A$; those newly covered by adding $x$ to $B$ are those meeting $x$ but not $B$. The latter is contained in the former, so the marginal gain at $A$ is at least that at $B$. □

`src/submodular_coverage.py` checks the result two ways. Over $300$ random onset-duration masses, drawn with Dirichlet concentrations from $0.3$ to $3.0$ on circles of $6$ to $10$ slots with durations up to three, there were zero violations of monotonicity and zero of submodularity. Against exhaustive search on twelve instances small enough to enumerate, greedy matched the exact optimum in every case:

| $n$ | $k$ | greedy | exhaustive | ratio |
|---|---|---|---|---|
| 8 | 2 | 0.4863 | 0.4863 | 1.000 |
| 12 | 3 | 0.7704 | 0.7704 | 1.000 |
| 9 | 3 | 0.9058 | 0.9058 | 1.000 |
| 11 | 2 | 0.4713 | 0.4713 | 1.000 |
| 12 | 2 | 0.5645 | 0.5645 | 1.000 |

The worst ratio over all twelve was $1.000$ against a guaranteed $1 - 1/e \approx 0.632$. We report the guarantee rather than the observed ratio as the transferable claim, since twelve small instances do not establish that greedy is exact in general.

### H.5 THE ADAPTIVITY GAP: WHICH INGREDIENTS ARE NECESSARY

*Proof of Proposition 3.* Under the hypothesis, measuring conveys no information about any reward-relevant future event, so the conditional law of those events given the history equals its unconditional law at every $t$. A policy's decision at $t$ can therefore depend only on $t$ and on the slots it has already used, and no realisation of the observation process makes one continuation preferable to another. Fix a policy $\phi$ with at most $k$ measurements on every path. Conditioning on its internal randomization writes $\phi$ as a mixture over deterministic schedules, each of at most $k$ slots, and its value as the corresponding mixture of their values; a mixture cannot exceed its best component, so some deterministic schedule of at most $k$ slots does at least as well. Hence $\Delta_k = 0$. □

**The argument needs a pathwise budget.** If the constraint is only $\mathbb{E}[\text{measurements}] \leq k$, a randomized policy need not be a mixture of schedules with at most $k$ slots and the reduction does not apply. Expected coverage is also not linear in the marginal probabilities $p_t$ when two measurements can cover the same excursion — interpolating between two single-slot policies loses $0.074$ on one small instance we checked — so a polytope-vertex argument is unavailable. We searched numerically for a randomized policy under an expectation budget that beats the best deterministic schedule and found none across the instances we tried, but we do not claim that case. □

`src/adaptivity_gap_probe.py` solves the adaptive problem exactly by backward induction over slot, measurements still available and belief over latent types, and compares against exhaustive fixed schedules on the type-marginalised hazard. Forty instances per regime, $n = 6$, $k = 2$, two types:

| regime | instances with a positive gap | largest gap |
|---|---|---|
| identical types, informative observations | 0 / 40 | 0.0000 |
| differing types, uninformative observations | 0 / 40 | 0.0000 |
| differing types, informative observations | 33 / 40 | 0.1409 |

**Necessity of a reversal, and where the argument stops.** If no two slots are ranked differently by two distinguishable types, every type shares one ordering. *When detection is additive in slots* the top-$k$ set is then identical across types, the optimal policy selects it whatever it observes, and $\Delta_k = 0$. Across 200 constructed instances with a common ordering none had a positive gap.

**The step does not survive multi-slot coverage.** Once a measurement covers a window rather than a point, the best $k$-set is not the top $k$ slots: it trades individual mass against overlap. Two types can share a per-slot ordering exactly and still have different optima — with onset mass $(0.30, 0.25, 0.20, 0.13, 0.08, 0.04)$ for both, a type of duration one is best served by slots $\{0, 1\}$ while a type of duration three is best served by $\{0, 3\}$, which covers the whole circle. **A common per-slot ordering is therefore not sufficient for a zero gap when coverage is submodular**, and our 200 instances verify the additive case only. The condition should be read as ranking over the *sets* a policy might choose, not over slots.

The first row is Proposition 3. The second is the same conclusion by a different route: heterogeneity that measurement cannot resolve is heterogeneity the policy cannot act on. Only the third gives adaptivity anything to do. We report this as evidence for the necessity of both conditions rather than as a proof of it; the sufficiency direction, and a bound on the size of the gap, are open.

### H.6 THE SOLVER

The formulation of Section 2 is not solvable by gridding the belief simplex, whose dimension is $|\mathcal{S}| - 1$. Two properties make it tractable: the belief is carried as a mean and a staleness index, so its dimension is two regardless of state resolution, and the variance is an index rather than a coordinate. The procedure is backward induction on that parametric belief, with linear rather than nearest-neighbour interpolation between grid points — nearest-neighbour projection on a convex value function fails to converge under refinement while linear interpolation converges to four decimal places. Appendix G.10 gives the pseudocode and Appendix F.8 the refinement study.

### H.7 THE GREEDY COMPARATOR ON THIS COHORT

Proposition 2 licenses a greedy fixed schedule wherever enumeration is infeasible. On this cohort, where both are computable, greedy recovers $94$ to $100$ percent of the exhaustive optimum while evaluating $nk$ candidates instead of $\binom{n}{k}$ — 120 against $42{,}504$ at $k = 5$:

| $k$ | greedy | exhaustive | ratio | evaluations (greedy / exact) |
|---|---|---|---|---|
| 1 | 0.388 | 0.388 | 1.000 | 24 / 24 |
| 2 | 0.546 | 0.579 | 0.943 | 48 / 276 |
| 3 | 0.683 | 0.694 | 0.984 | 72 / 2,024 |
| 4 | 0.792 | 0.792 | 1.000 | 96 / 10,626 |
| 5 | 0.847 | 0.853 | 0.994 | 120 / 42,504 |

### H.8 TWO AXES, FOUR REGIMES

Comparing a mixing half-life with an action interval collapses two separate questions into one number, which is why Proposition 1 can refute it. The mechanisms separate along two axes: how concentrated event onset is in clock time, which governs what *placement* can buy, and how dispersed the conditional hazard is across latent types, which governs what *adaptivity* can buy. A factorial sweep with everything computed exactly ($n = 6$, $k = 2$, two types, twelve instances per cell) populates the grid.

| clock concentration | heterogeneity | placement gain | adaptivity gain |
|---|---|---|---|
| high | none | 0.1125 | 0.0000 |
| high | medium | 0.0932 | 0.0089 |
| high | high | 0.0843 | 0.0615 |
| medium | high | 0.0689 | 0.0475 |
| low | none | 0.0623 | 0.0000 |
| low | high | 0.0536 | 0.0255 |

Placement gain is the best schedule minus the best lattice; adaptivity gain is the exact adaptive value minus the best schedule. Averaged over the sweep, placement is worth $0.097$ under concentrated onsets against $0.055$ under spread ones, while adaptivity is worth $0.045$ under heterogeneous types against exactly $0.000$ under identical ones. **The axes do not predict each other**: the column with the largest placement gain has zero adaptivity gain, and adaptivity is worth most where placement is worth least relative to its own row. This is the four-regime picture: neither matters, placement matters, adaptivity matters, both matter. The sweep is synthetic and does not place our cohort on the grid; what it does is separate two mechanisms that a single mixing half-life conflates, so that a study can ask which of them its own setting has.

### H.9 WHAT THE DIAGNOSTIC CLASS DOES AND DOES NOT COVER

Proposition 1 rules out permutation-invariant functions of the advantage multiset. That class is chosen because it contains the shortcuts actually in use – a contraction half-life, an advantage spread, a ratio of measurement interval to mixing time – all of which discard clock labels. It is a fair objection that discarding clock labels is exactly what makes the result available: coverage depends on *where* the slots fall, so any diagnostic blind to position cannot see it.

Two stronger classes remain open. A *time-indexed* diagnostic could read the advantage at each slot separately; a *finite-window* one could read a neighbourhood. We do not know whether either can certify coverage, and Proposition 7 suggests what a sufficient object would look like – the convolution of onset mass with the duration kernel, which is not a function of advantages at all.

### H.10 THE APPROXIMATE ZERO-GAP BOUND

`src/approx_zero_gap.py` builds instances by perturbing a common base hazard by $\pm d$ with $|d|_\infty = \varepsilon$ and $\sum_t d_t = 0$, so the two latent types differ by at most $\varepsilon$ while the marginal is unchanged.

The adaptive value is solved exactly; the fixed value is the best schedule against the marginalised hazard.

| $\varepsilon$ | mean gap | max gap | $\max/(2k\varepsilon)$ |
|---|---|---|---|
| 0.00 | 0.00000 | 0.00000 | — |
| 0.02 | 0.00020 | 0.00788 | 0.099 |
| 0.05 | 0.00239 | 0.02349 | 0.117 |
| 0.10 | 0.01161 | 0.05935 | 0.148 |
| 0.20 | 0.04421 | 0.10604 | 0.133 |

Across 300 instances with $n$ from 4 to 7, $k$ from 1 to 3 and $\varepsilon$ from $0.01$ to $0.30$, the bound $\Delta_k \le 2k\varepsilon$ was never violated; the ratio averaged $0.030$ and peaked at $0.204$. **The counting argument does not establish this.** Each of the $k$ measurements gaining and losing at most $\varepsilon$ treats the perturbations as separable, and under submodular coverage they are not: two measurements covering one excursion do not contribute independently.

**A simulation lemma does, and does not need separability.** Under Assumption 1 the transition kernel is identical in the two models, so they differ only in the observation kernel, and only where a measurement is taken — at most $k$ steps. The simulation lemma for partially observed models (Lee et al., 2023) bounds the value difference of *any* history-dependent policy by $V_{\max}$ times the summed $\ell_1$ deviation of the kernels along the trajectory. Our value is a coverage fraction, so $V_{\max} = 1$, and the sum is at most $k\varepsilon$; every policy's value therefore moves by at most $k\varepsilon$, each arm's supremum by at most $k\varepsilon$, and their difference by at most $2k\varepsilon$. Since $\Delta_k = 0$ in the unconditional model by Proposition 3, the perturbed gap is at most $2k\varepsilon$.

**What this argument bounds is the path measure, not the reward decomposition**, which is exactly why submodular coverage does not obstruct it. **The coefficient deserves care.** Lee et al. (2023) state the lemma with an $H$ in front because they take a per-step reward in $[0, 1]$, so $V_{\max} = H$ through their bound $\|\alpha\|_\infty \le GH$ on the $\alpha$-vectors. Our total reward is a coverage fraction bounded by one, so $V_{\max} = 1$ and the coefficient is 1, not the $H = 24$ of our horizon. Their Theorem 6.3, where the lemma is applied, allows infinite latent and observation spaces, so a continuous belief is no obstacle.

**Proposition 9** (Graceful degradation)**.** *Under Assumption 1, with a policy's value the expected fraction of excursions covered: if replacing every reward-relevant conditional law by its unconditional value perturbs the observation kernel by at most $\varepsilon$ in $\ell_1$, then $\Delta_k \le 2k\varepsilon$.*

*Proof of Proposition 9.* Write $v(\phi)$ and $v_0(\phi)$ for the value of a history-dependent policy $\phi$ under the model and the unconditional model. By Assumption 1 the two models share a transition kernel, so they differ only in the observation kernel, and an observation is drawn only when $a_t = 1$, which happens at most $k$ times on any path. The simulation lemma therefore gives

$$|v(\phi) - v_0(\phi)| \;\le\; V_{\max}\, \mathbb{E}_\phi \sum_{t\,:\,a_t=1} \big\| Z(\cdot \mid s_t) - Z^0(\cdot \mid s_t) \big\|_1 \;\le\; V_{\max}\, k\varepsilon \;=\; k\varepsilon,$$

since the value is a coverage fraction and $V_{\max} = 1$. The bound holds for every $\phi$, so for any family $\Pi$, $|\sup_\Pi v - \sup_\Pi v_0| \le \sup_\Pi |v - v_0| \le k\varepsilon$. Applying this to the adaptive family and to the fixed schedules separately and subtracting,

$$\big|\Delta_k(themodel) - \Delta_k(theunconditionalmodel)\big| \;\le\; 2k\varepsilon.$$

Proposition 3 gives $\Delta_k(theunconditionalmodel) = 0$, which yields the claim. □

**A closer tool exists, for a different criterion.** Demirci et al. (2025) give explicit value-difference bounds for POMDPs under exactly this kind of observation-kernel perturbation, in total variation and Wasserstein distance, and their filter-kernel sensitivity is sharper than what a generic simulation lemma provides. Their results are stated for discounted and average-cost criteria; ours is finite-horizon and undiscounted, which is why we use the simulation lemma instead. Adapting their bounds to a finite horizon would likely tighten the constant.

**What the argument uses and what it does not.** It bounds the *path measure* and never decomposes the reward, which is why submodular coverage raises no difficulty — the counting argument fails precisely because it does decompose. It also does not need the deviations to be independent, small in any average sense, or structured: $\varepsilon$ is a uniform bound over states.

### H.11 THE ADAPTIVITY GAP UNDER CORRELATION

Asadpour & Nazerzadeh (2016) prove that for a stochastic monotone submodular objective under a matroid constraint, with each element active *independently*, the adaptive-to-non-adaptive ratio is exactly $e/(e-1) \approx 1.58$. Two features of our problem match: coverage is monotone submodular (Proposition 2) and a cardinality budget is a uniform matroid. One does not. A single excursion spans several consecutive slots, so whether slot $t$ is covered and whether slot $t'$ is covered are dependent, and the independence hypothesis fails.

**That difference is not cosmetic.** Gupta et al. (2026) study probing with correlated items and show the adaptivity gap becomes $\Theta(d_{\max} S)$, with a matching lower bound — correlation removes the constant. Their $d_{\max}$ is the maximum degree of the correlation graph and $S$ the probing budget, the number of queries allowed; here the analogue of $d_{\max}$ is how many slots a single excursion couples, its duration, and the analogue of $S$ is the measurement budget $k$. So the right expectation is that adaptivity can be worth *more* under correlation than the independent bound suggests, not less.

Across our exactly solved instances the observed ratio stayed below $1.40$ and our matched-count result is $64.5/55.9 = 1.154$. **We report these as measurements, not as evidence that a constant bound holds**, since Gupta et al. (2026) show none is available in general.

A second strand quantifies *how much* adaptivity is needed rather than what it is worth. Esfandiari et al. (2021) show that $O(\log n \log k)$ adaptive rounds suffice for a $1 - 1/e - \epsilon$ guarantee against a fully sequential optimum, and that no constant factor is achievable with $o(\log n)$ rounds; Ghuge et al. (2024) and Cui & Nagarajan (2025) give the corresponding picture for adaptive submodular cover. Our solver is fully sequential over $24$ slots, so it sits far above the round threshold these results identify — which is consistent with a positive gap, though it does not predict its size.

Golovin & Krause (2011) give the adaptive-submodularity framework under which dependence of the kind we have can be handled; establishing that our objective is adaptive submodular, and hence importing a gap bound properly, is open.

### H.12 THE COMPARATOR LADDER AND WHAT SELECTION COSTS

Proposition 2 gives a greedy schedule where enumeration is infeasible, but the more useful thing to measure is what *choosing* a schedule from data costs on held-out patients. Across twenty patient-level folds and budgets $k \leq 4$, each rule is fitted on the training half and scored on the other:

| rule | train | held out | optimizm |
|---|---|---|---|
| lattice, best phase | 51.0 | 51.0 | $-0.0$ |
| greedy submodular | 63.9 | 52.4 | $+11.5$ |
| exhaustive over the class | 64.7 | 54.3 | $+10.4$ |

The lattice selects nothing from the data — its phase is fixed by the budget — and its train-test gap is zero to the precision we can measure, which is what a rule that fits nothing should give. Exhaustive selection over the class buys $10.4$ points of optimizm, and greedy slightly more at $11.5$ despite searching far less, while ending $1.8$ points lower on held-out patients. **Greedy is therefore not the safer choice one might expect**: it is cheaper to compute and no less prone to selecting on noise.

This is the quantity that makes the asymmetry between the arms concrete. The adaptive policy is fitted with a single price and a restricted model, so it has less scope to select on noise than a search over thousands of schedules does; how much less we do not measure, and the direction of the net effect stays open.

### H.13 WHERE THE PROTOCOL SITS

Rerunning selection inside every bootstrap replicate is the bootstrap-after-model-selection construction of Tibshirani et al. (2018), applied to a combinatorial selection rather than a regression path. The wider programme (Taylor & Tibshirani, 2015) conditions on the selection event to restore valid inference; we do not condition, because our estimand is defined at the level of the *procedure* — fit, select, calibrate — rather than at the level of whichever schedule that procedure happens to return.

That choice is what makes the resampling legitimate: each replicate reruns the procedure end to end, so the selection variability it induces is inside the interval rather than conditioned away.

The price is that our interval covers a procedure-level functional, not the performance of any one schedule. A reader wanting an interval for a specific selected schedule would need the conditional machinery instead, and that is a different question from the one we ask.

### H.14 Why the selected schedule moves between folds

Appendix B.1 notes that the exhaustively selected schedule changes across folds, which is why selection is rerun inside every bootstrap replicate rather than conditioned away. The reason is visible in the structure of the coverage function.

For a monotone objective the schedules attaining a given coverage level form an up-set, and its minimal elements are the *lower boundary points* of that level — the object that multistate network reliability calls $d$-minimal paths and uses to decompose a demand event into a union of upsets (Forghani-Elahabad & Bonani, 2017; Zuo et al., 2007). Enumerating them on our cohort:

Counting exact optima understates the problem, so we follow the alternate-optima literature in reporting the $\varepsilon$-optimal set — every schedule within $\varepsilon$ of the best — rather than the ties alone (Petit & Trapp, 2019; Ahanor et al., 2024):

| | | schedules within $\varepsilon$ of the best | | | |
|---|---|---|---|---|---|
| $k$ | best coverage | $\varepsilon = 0$ | 0.5 pp | 1 pp | 2 pp |
| 2 | 57.9% | 2 | 2 | 2 | 6 |
| 3 | 69.4% | 2 | 2 | 6 | 21 |
| 4 | 79.2% | 2 | 2 | 7 | 31 |
| 5 | 85.2% | 1 | 1 | 10 | 51 |

**Exact optima are almost unique — one or two at every budget — but within two percentage points there are up to 51 schedules.** Detection on a resampled cohort moves by far more than two points, so the argmax is selected from a set of near-equals and a single excursion changing hands is enough to reorder them. The schedule that wins is unstable while what winning is worth is not. That is the mechanism behind the instability, and it is why an interval conditioned on one selected schedule would understate uncertainty.

The reliability literature's machinery does not transfer: it computes the probability that a *random* state vector dominates some boundary point, while our schedule is chosen rather than drawn, so there is no union probability to evaluate. What transfers is the object — the boundary set at each level — and counting it is what makes the instability quantitative rather than anecdotal.

**Effective sample size of the nocturnal comparison.** The headline comparison is scored on 646 held-out nights containing 93 hypoglycemic excursions. At the design effect of 5.27 measured on the full cohort, those 93 carry the information of roughly 18 independent events. **That is the sample the interval of Section 4 rests on**, and it is why we describe the magnitude as unresolved rather than quoting either endpoint as settled.

### H.15 How much the solved policy pools across nights

The fixed arm spends exactly $k$ measurements on every night. The adaptive arm is solved under a price, which constrains the expected rate and not the count on any one night (Section 2), so it has a second freedom the fixed arm lacks: spending little on a quiet night and a great deal on a volatile one. We measure that freedom rather than argue about it. Replaying the solved policy over all 1,343 nights and counting measurements on each:

| price $\kappa$ | mean per night | sd | min | max | sd/mean |
|---|---|---|---|---|---|
| 0.03 | 10.34 | 9.48 | 0 | 24 | 0.92 |
| 0.08 | 6.70 | 8.51 | 0 | 24 | 1.27 |
| 0.15 | 4.51 | 7.00 | 0 | 24 | 1.55 |
| 0.26 | 3.14 | 6.40 | 0 | 24 | 2.04 |
| 0.40 | 1.34 | 4.10 | 0 | 24 | 3.05 |

**The dispersion is large and it grows as measurement gets dearer.** At every price the policy uses anywhere from nothing to the whole night, and at the sparsest setting the standard deviation is three times the mean. A fixed schedule has standard deviation zero by construction.

**This is a confound we do not separate.** Our reported gap combines two things: conditioning the timing within a night on what has been seen, which is the question the paper asks, and reallocating measurements across nights, which a fixed schedule cannot do at all. **We do not claim the gap is attributable to the first alone.** Separating them needs an adaptive arm under a pathwise budget — a per-night cardinality constraint rather than a price — which is a different solver and which we have not run. Until then the honest reading is that the comparison is between a class that may reallocate and one that may not.

### H.16 Circadian drift and what the stationary fit costs

Section 2 fits one AR(1) to the whole night. Overnight physiology is not stationary, so we measure what that assumption costs rather than defend it.

Fitting $x_{t+1} = ax_t + b$ separately in each third of the night:

| window | $a$ | $b$ | $b/(1-a)$ | mean | below 70 |
|---|---|---|---|---|---|
| 23:00–01:00 | 0.9870 | 0.310 | 23.9 | 130.3 | 3.91% |
| 01:00–03:00 | 0.9829 | 1.363 | 79.7 | 122.3 | 4.93% |
| 03:00–05:00 | 0.9871 | 1.731 | 134.2 | 119.7 | 4.69% |
| whole night | 0.9834 | 2.355 | 141.8 | — | — |

The autoregressive coefficient is stable to three decimals. The intercept is not — it rises more than fivefold — and the implied stationary mean $b/(1-a)$ ranges over a hundred units, which is a sign that the stationary parametrization is being asked for something the data does not identify at this $a$. The measured level falls by ten units across the night and the fraction below 70 mg dL$^{-1}$ rises by a factor of 1.20.

**What matters is whether this changes the decision, so we test that directly.** Solving under each parameter set and replaying every resulting policy on the *real* process, at a measurement rate matched by bisecting the price:

| fitted on | detection | rate |
|---|---|---|
| whole night | 77.0% | 2.49 |
| 23:00–01:00 | 65.6% | 2.48 |
| 03:00–05:00 | 61.7% | 2.44 |

**Fourteen points separate the whole-night fit from a third-of-night fit at the same measurement rate.** This is not a harmless idealisation: the parameter set is a modeling choice with consequences of the same order as the effect we report. The whole-night fit happens to be the strongest of the three, which is the direction that flatters us, and we report it because a reader should know the comparison is sensitive to a choice we made before seeing that table. `src/circadian_drift.py` reproduces both.

**Active measuring, and why its heuristics do not transfer.** Krale et al. (2023) study action-contingent noiselessly observable MDPs, where an action has two components: a control action that moves the environment and a measurement action that reveals it. Their act-then-measure heuristic

chooses the first while ignoring future state uncertainty, and bounds the loss that causes. **The heuristic has nothing to act on here.** Under Assumption 1 our action set is $\{0, 1\}$ and neither element moves the state, so there is no control action to decouple from measurement; this is controlled sensing, not control. What does carry across is their *measuring value*: the quantity that decides whether to spend an observation. Our $A_h(\widetilde{\beta})$ of Equation (3) is that object on the reduced belief, and Proposition 1 is a statement about what it cannot certify.

**The expectation in Equation (4) covers the cost.** For a fixed schedule $Q_\phi$ is deterministic and the placement of the expectation is immaterial. For an adaptive policy it is not: $Q_\phi$ is a random set determined by the trajectory, so both the detection count and the accumulated price are random and the expectation must cover both. **Writing the cost outside the expectation would define a different functional**, one that charges a non-random price for a random number of measurements, and it is not the quantity our solver optimizes or our protocol estimates.

**Pathwise against expectation budgets.** Proposition 3 assumes a pathwise budget; the solver enforces the budget only in expectation through the price $\kappa$. McMahan & Zhu (2024) call this an *anytime* constraint against an expectation constraint, and three of their results apply.

First, an optimal policy under an anytime constraint is not Markovian in the state: it needs the cumulative cost in its state, which is what a budget-augmented solver carries. Our priced solver has no such coordinate, so it is not solving the constrained problem Proposition 3 describes.

Second, and more sharply, **an optimal policy for the expectation-constrained problem can violate the anytime constraint by an arbitrary amount**. The gap between what we solve and what Proposition 3 assumes is therefore not a technicality to be waved through; in general it is unbounded, and nothing in our setting rules that out a priori.

Third, the obvious repair fails: solving with a smaller budget so that the resulting policy is pathwise-feasible can be arbitrarily suboptimal. Tightening $\kappa$ until the realised count never exceeds $k$ would not recover the right comparison.

The dispersion measured in Appendix H.15 is large but finite on this cohort, which is evidence about these data rather than a bound. The adaptive arm is solved under a weaker constraint than the fixed arm faces, and in general that can matter without limit.

**What the expectation budget is worth.** Adding a remaining-count axis to the solver — same grid, same quadrature, same interpolation, with the measure action unavailable once the count is spent — gives a policy that keeps within-night timing and cannot reallocate across nights. The two arms then differ only in the constraint. On held-out nights:

| $k$ | pathwise budget | expectation budget | difference |
|---|---|---|---|
| 2 | 55.8% | 63.3% | +7.5 |
| 3 | 58.3% | 67.5% | +9.2 |
| 4 | 62.5% | 74.2% | +11.7 |

**Relaxing the budget from pathwise to expectation is worth seven to twelve points**, and the margin grows with the budget. That is the size of the freedom the fixed class does not have, measured on this cohort at one discretization. The solver is sound where the constraint is slack: at $k = 24$ it measures $14.4$ times a night and reaches $97.8\%$.

**This does not decompose the reported gap.** Comparing the pathwise arm against the fixed schedule would need the two fitted alike, and they are not: the schedule is selected on training nights while the kernel is estimated on the cohort. The comparison above escapes that because both arms share a kernel.

### H.17 Why the interval covers zero when the classes are nested

A fixed schedule is an adaptive policy that ignores its observations, so at the population level the adaptive optimum weakly dominates and a negative gap should be impossible. Our interval covers zero. Three explanations are available and the data distinguish them.

**Containment holds in our data.** Evaluated on the same nights that select the schedule and calibrate the price, at a matched realized rate, the solved policy beats the exhaustive fixed optimum at every budget: $+10.9$, $+7.7$ and $+4.4$ points at $k = 2$, 3 and 4. Nesting is not the issue.

**Nor is it the approximation.** The solver works on a reduced Gaussian belief we know to be misspecified, which could in principle cost more than the class inclusion buys. If that were the source, the in-sample comparison above would already show it, and it does not.

**It is sampling variation in the evaluation.** Holding the kernel, the price and both policies fixed — so that nothing is refitted — and resampling only the held-out nights, the gap at $k = 3$ has mean $+7.8$ and standard deviation $4.7$ points, and falls below zero in $4.0\%$ of draws. The full bootstrap, which also reruns selection and calibration, gives $3.6\%$. The two agree, so the negative mass is accounted for by the evaluation sample alone.

The arithmetic is unremarkable once stated: $7.8/4.7$ is $1.7$ standard deviations, and a normal approximation puts $4.9\%$ below zero. **The population quantity is almost certainly positive; what** $93$ **excursions worth some** $18$ **independent events cannot do is establish that from a single cohort.** This is why we report the interval rather than the point estimate, and why the paper claims a direction rather than a magnitude.