

Queueing-Informed Adaptive Alarm Policies for Continuous Glucose Monitoring

Yupeng Chen MS,¹ Naman Gupta BS,¹ David Scheinker PhD^{1,2}

¹ Department of Management Science and Engineering · ² Department of Pediatrics · Stanford University, Stanford, CA



BACKGROUND

CGM generates frequent threshold alerts. In remote patient monitoring, a subset is forwarded to a reviewing clinician as work items. Guidelines now add time in tight range (70–140 mg/dL) to the standard 70–180 mg/dL target,⁴ raising event volume while clinician capacity stays fixed.

We previously proposed suppressing low-urgency alerts by time of day and rate of change, untested against episode-level event construction or an explicit review model.

Patient-facing CGM alarms are unmodified throughout. Policies govern only which episodes enter the between-visit review queue, not the periodic summary review at a clinic visit. Whoever staffs that queue — a monitoring nurse, an educator, or the patient's caregiver — results are reported per session and per patient-day.

OBJECTIVE

To evaluate review policies for the between-visit remote-monitoring queue on detection of clinically important glycemic events and on the burden they place on whoever reviews them, and to test whether burden can be reduced without loss of detection.

METHODS

Data. DCLP5 (development):¹ 93 children aged 6–13, 3.78M readings, 13,131 patient-days. DCLP3 (external validation):² 168 participants aged ≥14, 7.48M readings, 25,963 patient-days. Dexcom G6, 5-minute sampling; gaps over 15 min treated as breaks in observation.

Episodes. A per-patient state machine produced mutually exclusive glycemic episodes (DCLP5 n = 72,748). An excursion crossing 140 mg/dL that later exceeds 180 is one escalating episode, not two alerts. Episodes were tiered as critical low (<54), low (54–69), tight-range excursion (141–180), level-1 high (181–250) and level-2 high (>250 mg/dL).³

Policies. Twenty-two policies through one interface: four prespecified comparators, including the prior heuristic carried unchanged, and 18 configurations from a grid fixed before any comparison. Every new policy enforces a clinical floor in code: no configuration may withhold an episode below 70 or above 250 mg/dL.

Review model. Review is periodic: the queue is opened on a cadence and whatever accumulated is worked. Session load is counted from the observed task stream, with no service-time, effort-ratio or utilization assumption; cadence is swept hourly to weekly. Arrivals replay bootstrapped patient-days for a 50-patient panel, 200 paired replications of 28 days, common random numbers.

Analysis. All policies compared on identical episodes; 95% CIs from patient-clustered bootstrap (1,000 paired resamples). One configuration selected by nested patient-grouped cross-validation against a prespecified constrained objective.

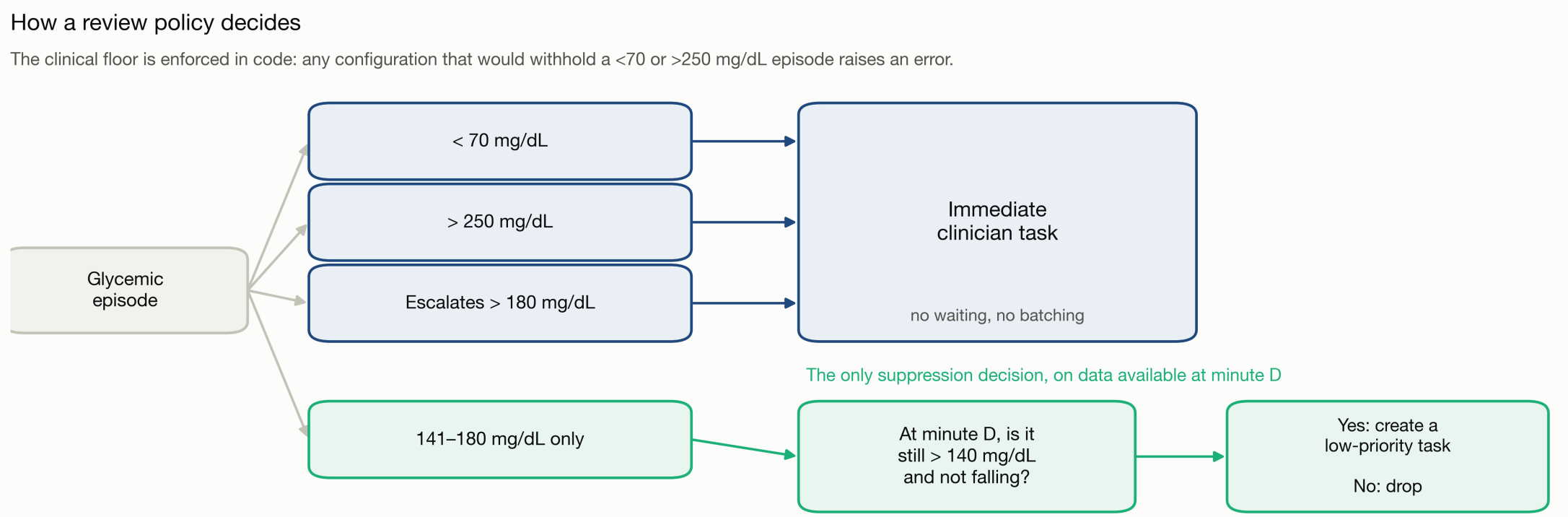


Figure 1. Decision path from a glycemic episode to a clinician work item. Hypoglycemia, level-2 hyperglycemia and any escalation above 180 mg/dL bypass every gate.

REFERENCES

- Breton MD, et al. A randomized trial of closed-loop control in children with type 1 diabetes. *N Engl J Med.* 2020;383(9):836–845.
- Brown SA, et al. Six-month randomized, multicenter trial of closed-loop control in type 1 diabetes. *N Engl J Med.* 2019;381(18):1707–1717.
- Battelino T, et al. Clinical targets for continuous glucose monitoring data interpretation. *Diabetes Care.* 2019;42(8):1593–1603.
- Battelino T, et al. Continuous glucose monitoring and metrics for clinical trials: an international consensus statement. *Lancet Diabetes Endocrinol.* 2023;11(1):42–57.

Data and code. Analyses used the publicly available DCLP5 and DCLP3 datasets (JAEB Center for Health Research). Analysis code and complete result tables are available from the authors.

Disclosures. The authors declare no competing interests.

RESULTS: BURDEN AND DETECTION

Review burden. Extending review from >180 to >140 mg/dL raised clinician tasks from 3.58 to 5.54 per patient-day (+55%), adding 1.96 (95% CI 1.83–2.10) low-priority tasks per patient-day (Figure 2).

Prior heuristic. It reduced low-priority flags by 0.9% while adding 5,079 unrealized precursor alerts at high priority: 5.4% more review items and 18% more night-time interruptions. It also withheld 3.1% of >250 mg/dL episodes.

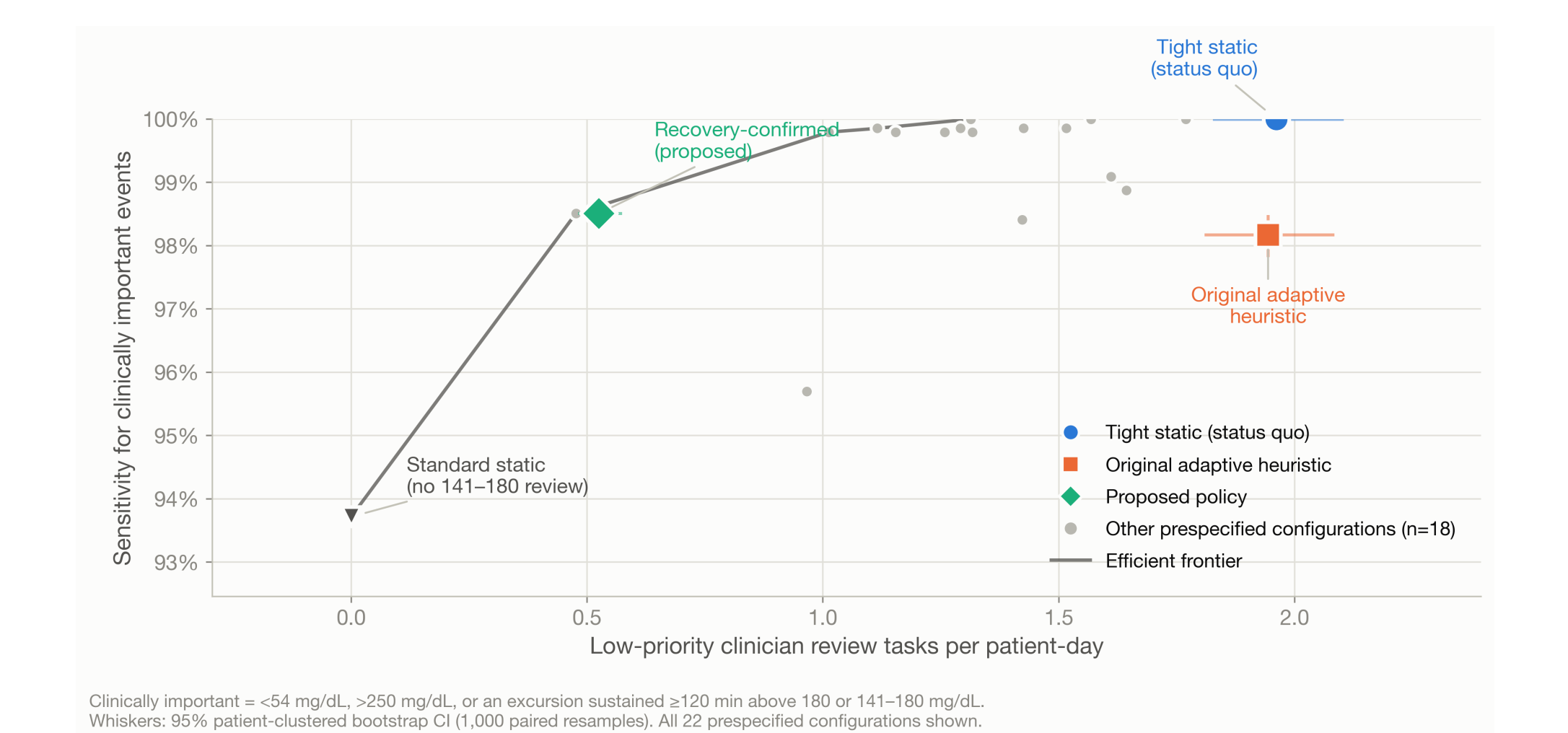


Figure 2. Detection and burden for all 22 prespecified configurations. Whiskers are 95% patient-clustered bootstrap intervals.

Recovery-confirmed review. For excursions confined to 141–180 mg/dL no task is created at onset; at 30 minutes the episode is checked once and forwarded only if still above 140 mg/dL and not falling. Low-priority tasks fell to 0.53 (0.48–0.57) per patient-day, a **73% reduction** (paired difference –1.44, 95% CI –1.53 to –1.33), retaining 100% of <70, >250 and sustained >180 mg/dL episodes.

Batching adds nothing. Batching the surviving tasks into 8-hour windows withheld the identical episodes and cut tasks a further 1.2%, but raised 95th-percentile onset-to-review time from 78 to 328 minutes at unchanged capacity. Not adopted.

Persistence rules. Comparable burden reduction (51%) but lower sensitivity for sustained >180 mg/dL excursions, 94.5% (93.8–95.2%); 31% of what they withheld later exceeded 180 mg/dL (Figure 3).

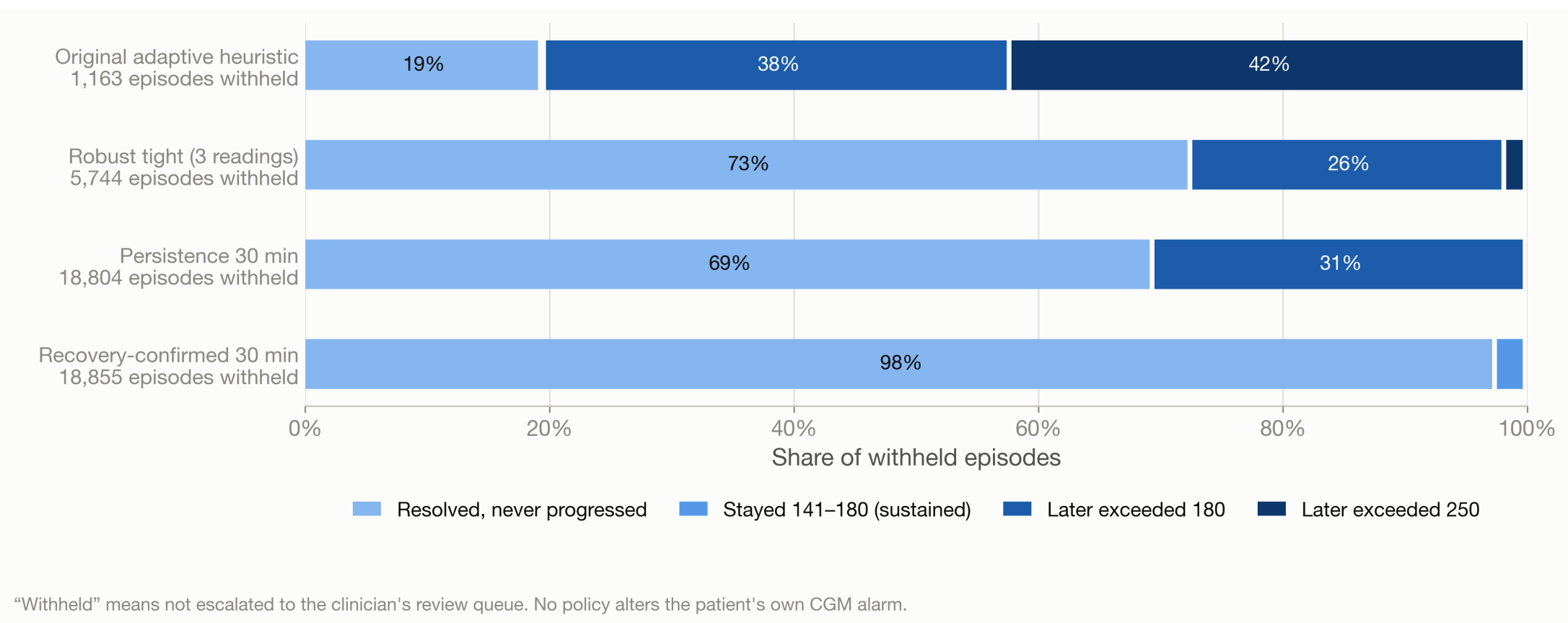


Figure 3. Outcome of the episodes each policy withheld from clinician review, as a percentage of that policy’s withheld episodes. Shading denotes the most severe glycemic level subsequently reached. Row labels give the policy and the number of episodes withheld.

LIMITATIONS

Detection targets are CGM-derived proxies, not demonstrated clinical actionability. The composite score depends on the sustained-event definition, though retention of <70, >250 and sustained >180 mg/dL is 100% at every definition tested. Who staffs the review queue is not observed, so results are given per session and per patient-day rather than in staff hours. Both cohorts are closed-loop trials from one network.

RESULTS: REVIEWING THE QUEUE

Session load. For a 50-patient panel reviewed once daily the proposed policy leaves 190 items per session rather than 257, a **26% cut** that holds at every cadence from hourly to twice weekly (Figure 4A). It does not reduce how many patients a session opens, but what there is to read per patient: 5.2 items to 3.9.

Delay is set by the cadence. Not by the policy: at once-daily review the 95th-percentile onset-to-review time differs by 38 minutes out of 23 hours.

Night-time interruptions. When the reviewer is the patient or a parent, the burden that matters is being woken. Items generated 23:00–07:00 fall from 1.12 to 0.67 per patient-day, **–40.1%** (95% CI –42.6% to –37.4%); the prior heuristic raises them 18.2% (+15.6% to +20.8%) (Figure 4B).

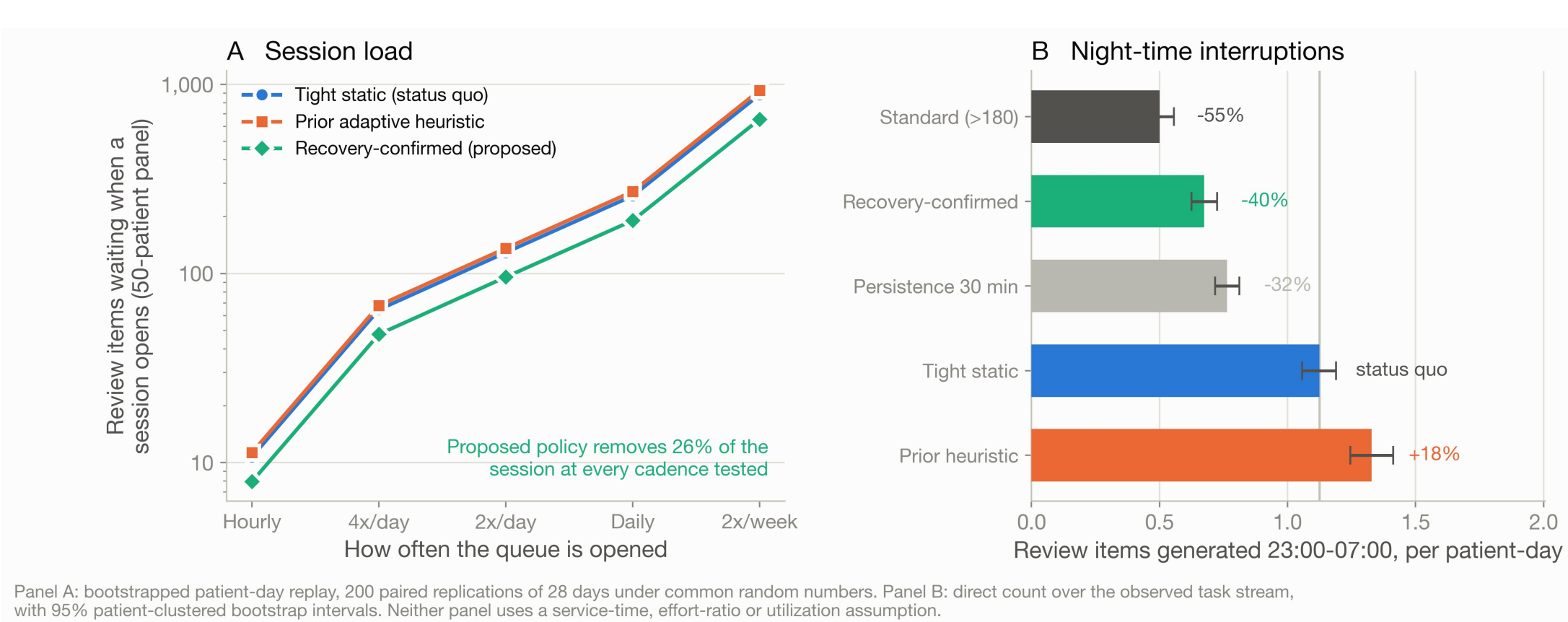


Figure 4. Periodic review. Neither panel rests on a service-time, effort-ratio or utilization assumption.

SENSITIVITY AND EXTERNAL VALIDATION

Policy ranking was invariant to the critical-low and >250 mg/dL thresholds and to the progression horizon (Spearman 1.00), and burden reduction was 73.2% at every event definition tested. The fixed set was then applied to DCLP3 with no refitting (Table 1). A logistic model given the rule’s decision point, features and task budget detected 83.4% of sustained tight-range excursions versus the rule’s 76.3% (+7.1 points, 95% CI +4.4 to +9.8), but unrefitted on DCLP3 that margin fell to +2.6. A single-reviewer queueing model, which does assume an effort ratio, agrees: 25.9% fewer tasks, a panel of 55 rather than 50.

Table 1. Detection and burden by policy, development cohort with the external cohort alongside. Night items are those generated between 23:00 and 07:00.

Policy	Items / pt-day	Low-priority reduction DCLP5	DCLP3	<70 & >250 retained	Sustained >180 sens.	Night items / pt-day
Standard static (>180)	3.58	—	—	100%	100%	0.50
Tight static (>140)	5.54	reference	reference	100%	100%	1.12
Prior adaptive heuristic	5.84	0.9%	0.8%	96.9%	98.4%	1.33
Recovery-confirmed, 30 min (proposed)	4.10	73%	67%	100%	100%	0.67

CONCLUSIONS

- Deferring the review decision by 30 minutes and forwarding only excursions still elevated and not falling removed about three quarters of low-priority review with no measurable loss of hypoglycemia, severe or sustained hyperglycemia, and replicated externally.
- The gain needs no staffing model: a review session holds a quarter fewer items at every cadence, and items arriving 23:00–07:00 fall 40%.
- The prior heuristic failed by routing a low-precision predictor into the top priority class of a shared queue, not through its thresholds; it also raises night-time interruptions 18% and withholds severe events. An exemption for those belongs in the implementation, not in convention.
- Not adopted: batching, which a review cadence already subsumes; precursor alerts, whose 10-minute lead is too short for asynchronous review; and a learned model whose margin over the rule more than halved on transfer.