

Personalized pacing with Juno

A statistical method for energy planning and delayed exertional patterns

Marshall Gould, Juno

Technical whitepaper · 21 September 2026 · Version 1.0

Abstract

We describe a pacing method that converts self-reported energy and activity records into a personal planning budget, estimates activity costs from recorded history, and examines whether exertion is followed by lower reported energy. The method combines a bounded budget coefficient, Bayesian updating of activity costs, a regularized regression with one-day and two-day delays, and descriptive detection of low-energy episodes. Optional wearable signals adjust initial estimates through explicit rules. A schedule-based balance shows when planned activities would exhaust the estimated budget.

The contribution is a reproducible combination of these components within Juno. The statistical techniques themselves are established methods. The implementation examined here is a development prototype; the work does not establish clinical effectiveness, a physiological measure of energy, or reliable prediction of post-exertional malaise. We specify the calculations, describe their limitations and set out the evaluation needed before making those claims.

Why this method

A fixed activity catalogue treats the same task as equally demanding for everyone. A fixed daily allowance cannot reflect a person's changing experience. Our earlier product rule assigned 2.4 spoons to each point on a ten-point energy scale: an energy rating of 5 produced a budget of 12. The new method retains this familiar starting point while using a person's own records to inform selected estimates. "Spoons" are bookkeeping units for perceived demand, not calories or a measurement of biological capacity.

Pacing must also account for delay. NICE describes post-exertional malaise as symptom worsening that can begin hours or days after activity. CDC guidance describes a typical delay of 12 to 48 hours. These observations motivate examining previous days, but they do not validate our model or its thresholds. [1,2]

Intended use

The intended use is to support a person reviewing activities, rest and patterns with appropriate clinical support. NICE recommends a flexible, person-led approach that includes physical, cognitive, emotional and social demands; activity should never be increased automatically. A higher numerical budget is therefore not an instruction to do more, and unused spoons are not a target to spend. [1]

1 Observations and the daily budget

The model uses dated activity records and the latest available self-reported energy rating for each day. Energy ratings are normalized to half-point increments from 1 to 10. Completed activities contribute their recorded actual cost when available and otherwise their stored cost. Positive values represent demand; negative values represent an assigned recovery credit. Deleted records are excluded. The daily series contains zero expenditure where no activity is recorded, which must not be interpreted as proof that no activity occurred.

$$S_t = \max(0, D_t - R_t) \quad B_t = \text{round}(E_t \times k)$$

Here E_t is reported energy, D_t is completed positive demand, R_t is completed recovery credit, S_t is net recorded expenditure, and B_t is the planning budget. The coefficient k starts at 2.4 and is constrained to 1.2–4.0. Budgets are whole spoons. Those constants are product choices rather than clinically established values. Rest credits are an accounting convention; they do not establish that exertional effects have been reversed.

Fitting the personal coefficient

We compare adjacent calendar days with energy recorded on both days and positive net expenditure on the earlier day. Let $r_t = S_t / E_t$. A pair enters the sustainable set if next-day energy is no more than one point lower. It enters the drop-associated set if energy falls by at least two points. A 1.5-point fall enters neither set. “Sustainable” is the implementation’s label for this narrow rule; a stable next-day rating does not exclude delayed symptoms.

With fewer than three sustainable pairs, the method returns the default coefficient, even if several drop-associated pairs exist. Otherwise q is the 80th percentile of sustainable ratios. The evidence value a starts at $\max(q, 2.4)$. If there are at least three drop-associated pairs, a is capped at their median ratio and these pairs are added to the evidence count n . Without that cap, n is the number of sustainable pairs.

$$k = \text{bound}_{[1.2, 4.0]} \left(\frac{10 \times 2.4 + n \times a}{10 + n} \right)$$

The result is rounded to two decimals. Ten pseudo-observations pull the estimate toward the initial value when evidence is sparse. Low expenditure alone cannot lower the coefficient. The method is recalculated against the same default rather than progressively updating from the previously stored coefficient. Quantiles use linear interpolation. [3]

Worked example

Suppose ten sustainable pairs give $q = 3.0$ and there are too few drop-associated pairs to apply the cap. The fitted coefficient is $(24 + 30) / 20 = 2.70$. With reported energy 6, the budget becomes $\text{round}(6 \times 2.70) = 16$ spoons. These are illustrative inputs, not observed patient results. The apparent extra allowance requires cautious interpretation because symptoms outside the next-day energy rating are not represented.

2 Learning the cost of an activity

For repeated activities, we estimate a personal mean cost using a normal prior and a normal observation model. This is a Bayesian update: a starting estimate is combined with recorded values, with the relative influence determined by their assumed uncertainty. The default prior standard deviation is 1.5 spoons and the observation standard deviation is 1 spoon. Both are fixed assumptions in this prototype.

$$v_n = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \quad \mu_n = v_n \times \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_i}{\sigma^2} \right)$$

$$90 \text{ percent interval for the mean} = \mu_n \pm 1.645 \times \sqrt{v_n}$$

μ_0 is the initial estimated cost, $\sigma_0 = 1.5$, $\sigma = 1$, and x_i is a recorded cost. The interval concerns the activity's estimated mean under these assumptions. It is not a prediction interval for the next occurrence and does not guarantee the user's next experience will lie inside it. [3]

Worked example

Input or output	Spoons
Initial estimate	3.00
Three recorded costs	4.00, 5.00, 4.00
Updated mean	4.16
Posterior standard deviation of the mean	0.54
90 percent interval for the mean	3.27 to 5.05

The output moves toward the person's recorded experience. With only three records, it remains partly influenced by the starting estimate. More records narrow the interval under the model, but more records do not repair systematic bias or incorrect input labels.

How records are grouped

Activity names are lowercased, trimmed and have repeated spaces collapsed. At least two records are needed for a displayed estimate. The earliest record's stored estimate supplies the prior; if it is missing, its recorded cost supplies the prior. The most recent name supplies the display label. Results are ranked by the absolute change from the prior, and the insights response returns at most 40 activity estimates.

This grouping does not adjust for duration, intensity, posture or context. A five-minute shower and a twenty-minute shower with the same name are pooled. More importantly, an "observation" falls back to the stored estimated cost when no actual cost was entered. Repeated uncorrected estimates can therefore look like evidence and produce unjustified precision. Before interpreting these estimates as learning from patient feedback, an evaluation must separate explicitly reported costs from these fallback values.

The prototype displays learned estimates but does not yet feed them back into the quick-pick activity catalogue or calendar estimation. Completing that feedback path is a separate implementation step.

[4]

3 Estimating delayed associations

A task can appear manageable at the time and be followed by worsening symptoms later. We model daily energy as a function of previous-day energy and expenditure over the previous two days. The purpose is to describe a recurring association in an individual's records. The model does not determine whether an activity caused a crash and does not diagnose post-exertional malaise.

$$E_t = \alpha + \phi E_{t-1} + \beta_1 S_{t-1} + \beta_2 S_{t-2} + \varepsilon_t$$

The intercept α represents the fitted baseline; ϕ captures persistence in energy ratings. β_1 and β_2 represent conditional associations with expenditure one and two days earlier. Coefficients are fitted by ridge regression, which penalizes large slope coefficients. The penalty is 0.5 and the intercept is not penalized. Inputs are not standardized, so this penalty depends on the measurement scales. [3]

Eligibility and reported outputs

Each training row requires a three-day sequence of consecutive dates, with energy observed on the current and previous day. The oldest day contributes expenditure but does not need an energy rating. At least 21 qualifying rows are required. An uninterrupted series therefore needs at least 23 dated records, rather than 21 calendar days. Qualifying rows can also come from separate stretches of history. Missing activity logs may still appear as zero expenditure.

The response includes the number of fitted rows, the coefficients and an in-sample R-squared value, clipped below at zero. In-sample fit describes the history used for estimation; it is not evidence of future predictive accuracy. A failed fit or insufficient data returns an explicit insufficient-data status.

$$L = 10 \times (\beta_1 + \beta_2)$$

Combined lag index L	Prototype label
$L \leq -1.0$	Strong
$-1.0 < L \leq -0.5$	Weak
$L > -0.5$	None

This index corresponds to adding ten spoons on each of the two preceding days while holding the other input fixed. It is not the effect of ten total spoons distributed across both days. For example, $\beta_1 = -0.04$ and $\beta_2 = -0.08$ produce $L = -1.20$ and the label "strong." The labels are heuristic categories, not calibrated probabilities or clinical severity grades.

What the model leaves unresolved

Sleep, infection, medication changes, baseline illness severity and the reasons someone chose to rest may affect both activity and energy. A single energy rating also misses pain, cognitive symptoms, orthostatic symptoms and other features of PEM. Effects beginning later than two days are not explicitly modeled. The current budget-fitting rule does not use this lag model, so a delayed

association does not automatically lower the daily allowance. These limits constrain the conclusions that can be drawn from the output.

4 Recognizing low energy episodes

A separate rule looks for unusual declines relative to a person's recent energy history. It uses the preceding 14 energy-logged days, with at least seven required. These are logged observations rather than necessarily 14 consecutive calendar days. Let m be their mean energy and s their sample standard deviation. To prevent tiny variation from producing very large standardized differences, the effective standard deviation has a floor of 0.75.

$$z = \frac{m - E_t}{\max(s, 0.75)}$$

A candidate low-energy episode is recorded when z is at least 1.5 and the absolute fall below m is at least 1.5 energy points. The implementation calls this a “crash.” That label identifies an energy-pattern rule, not a confirmed clinical event. Severity is computed as $\text{round}(3z)$, constrained to 1–10; it is a display index rather than a validated severity scale.

Expenditure before an episode

The detector examines the preceding two energy-logged observations that fall within two calendar days. If the highest net expenditure among them is at least 1.5 expenditure standard deviations above the baseline mean, it labels the episode “boom-bust.” The expenditure standard deviation has a minimum of 1 spoon. A high-expenditure day without an energy record is excluded from this trigger search. The selected date is an association marker and should not be presented as the proven cause.

Once an episode starts, the detector groups subsequent low readings until the first recorded energy value within 0.5 points of the original baseline. Recovery duration is the calendar-day distance to that observation. If no such reading exists, duration is unknown. Gaps in logging make the true recovery time uncertain; the first observed return is not necessarily the day recovery occurred.

Worked example

With baseline energy 6.0, baseline standard deviation 0.6 and today's energy 4.0, the effective deviation is 0.75. The standardized drop is $2 / 0.75 = 2.67$, which meets both thresholds. The display severity is 8. If the first later recorded energy of at least 5.5 occurs three calendar days later, the recorded recovery duration is three days. No clinical outcome is implied by this arithmetic.

Describing consistency

Over the most recent seven logged days, we calculate the coefficient of variation of net expenditure: its sample standard deviation divided by its mean. A display score is $100 \times (1 - \text{coefficient of variation})$, clipped to 0–100. It is undefined when mean expenditure is zero. These seven logged days may span more than a week. The model also reports the fraction of energy-logged days whose expenditure was within the fitted budget.

Consistency and staying within budget describe the diary. Neither proves successful pacing: consistently excessive activity could receive a high consistency score, while variable activity may be an appropriate response to fluctuating illness. Zero-filled missing activity data can also make adherence look better than it is. These measures should support interpretation rather than reward activity or judge a person's effort.

5 Incorporating optional wearable signals

Wearable information is optional. We apply explicit multipliers to positive initial activity-cost estimates using available sleep, heart-rate variability and resting heart-rate data. Heart-rate variability is expressed as a ratio to the individual's stored baseline. Resting heart rate is compared with its stored baseline. These rules reproduce product assumptions in executable arithmetic; the coefficients have not been calibrated against clinical outcomes.

Signal	Value	Cost multiplier
Sleep duration	Below 5 hours	1.30
	5 to under 6 hours	1.25
	6 to under 7 hours	1.10
	7 to 8 hours inclusive	1.00
	Above 8 hours	0.95
HRV relative to baseline	Below 0.65	1.30
	0.65 to under 0.80	1.20
	0.80 to under 0.95	1.05
	0.95 or above	1.00
Resting heart rate	More than 10 bpm above baseline	1.15

Applicable multipliers are multiplied together, rounded to two decimals, and applied to positive model-estimated costs. Final positive costs are rounded to the nearest quarter spoon. Nonpositive costs and entries marked as holidays bypass this multiplication. A missing signal contributes no adjustment. The function imposes no additional cap on the combined multiplier. [3,5]

For 5.5 hours of sleep, HRV at 75 percent of baseline and resting heart rate 12 bpm above baseline, the product is $1.25 \times 1.20 \times 1.15 = 1.725$, rounded to 1.73. A positive base estimate of 3 spoons becomes 5.25 after quarter-spoon rounding. This is a reproducible example of the rule, not evidence that the activity is physiologically 73 percent harder.

Suggesting an energy rating

A separate heuristic converts available signals into scores from zero to one and combines them with weights of 40 percent for sleep, 35 percent for HRV and 25 percent for resting heart rate. Weights are renormalized when signals are missing. The combined score maps to energy 1–10. A one-point penalty applies when yesterday's recorded expenditure exceeded 120 percent of its budget; the result is normalized to half steps. With no wearable inputs, the function returns no suggestion. Appendix A specifies the score transformations.

Wearable quality, illness, medications and baseline selection can affect these inputs. Longer sleep does not establish recovery, and low HRV is not specific to PEM. A suggested rating should remain

distinguishable from what a person reports about how they feel.

6 Turning a plan into a balance

The within-day calculation orders scheduled activities by their end times and applies each signed cost at completion. Already-completed items enter the opening balance. Recovery credits increase the balance at their scheduled end. An item with no start time contributes to a separate unscheduled total and is omitted from the timed trajectory. Missing or invalid durations in the app adapter default to 30 minutes.

$$\text{Remaining}(t) = \text{daily budget} - \text{completed net cost} - \text{scheduled net cost by } t$$

The first point at or below a reserve is reported as the depletion time; the default reserve is zero. The response also includes the next scheduled point and the balance after scheduled items. This is a scenario calculation conditional on the plan and its estimates, not a measured rate of energy use. Completed activities are placed at a computational time zero, so the result is not a reconstruction of the person's actual earlier energy curve.

Worked day

For an illustrative budget of 12 spoons and 3 already used, the opening balance is 9. The remaining schedule gives the following results.

Point in the plan	Change in balance	Remaining
Opening balance after completed items	-3 from the budget	9
12 noon: activity ends	-4	5
2 pm: rest period ends	+1 assigned credit	6
5 pm: activity ends	-7	-1
Unscheduled activity	-2 held separately	Excluded above

The schedule reaches zero or below at 5 pm and finishes at -1. If the unscheduled activity also occurs, the full accounting balance would be -3, but the model cannot place that cost on the time axis. Moving or reducing an activity allows the person to compare plans. The calculation does not itself establish which plan is safe.

Where the language model fits

The calendar-estimation path uses a language model to interpret activity descriptions and return initial, unadjusted costs. The surrounding code applies wearable multipliers and rounding and filters returned activity identifiers against the submitted events. Statistical fitting, event detection and schedule arithmetic run in a shared deterministic module. Given identical numeric inputs and the same implementation, those calculations produce the same outputs. Initial language-model estimates can still be uncertain or variable. [3,5]

The insights endpoint assembles the history, estimates and event records. Its default daily lookback is 90 days, with a maximum requested range of 365 days. Activity-cost observations use a separate 180-day window and a 5,000-record limit. It can return results without persistence, or store the fitted

coefficient and detected events. A recalculated coefficient may change a future displayed budget automatically; any policy requiring human review of increases is an additional safeguard, not a control demonstrated by this prototype. [4]

7 Evidence and evaluation

The reviewed implementation contains the numerical estimators, data queries, an insights endpoint, an insights screen and a one-line schedule forecast. Its development checklist leaves deployment, migration application and device verification open. The learned-cost feedback into initial estimates and a full intraday chart are also unfinished. This paper describes that development snapshot; it makes no claim that these components are live for users. [4,6]

What has been checked for this paper

We inspected the source calculations and data-selection rules and directly executed the cost-update, biometric, energy-suggestion, daily-budget and schedule examples. The implementation also contains automated tests for statistics, payload assembly and database behavior. We did not rerun the full historical test suite or conduct a clinical study for this paper. Correct arithmetic is necessary, but it does not show that the resulting advice improves health.

A proposed evaluation sequence

First, validate the records. Retain whether each cost was explicitly reported or inherited from an estimate, distinguish missing activity from confirmed inactivity, and capture the timing of energy check-ins. Assess how duration, intensity, posture and activity type affect a repeated activity's meaning. Compare the default coefficient and all heuristic thresholds across realistic missing-data patterns.

Second, evaluate prediction on later observations that were not used to fit the model. Use chronological training and test splits within each person and report performance across people. Compare energy predictions with a last-observation baseline and a model without expenditure terms. Report error in energy points, uncertainty calibration and subgroup variation. For activity costs, assess both mean estimates and separately constructed prediction intervals using explicit feedback only.

Third, compare detected episodes with independent, time-stamped symptom reports and clinically reviewed definitions. Measure missed episodes, false alerts, alert frequency and recognition delay. A “no signal” output must be assessed for false reassurance. Broader symptom measures and longer delays should be studied because low energy alone is an incomplete representation of PEM.

Finally, test usability and outcomes with patient and clinical input before using the output to direct behavior. Measure whether people understand uncertainty, feel pressured to use a budget, or change activity because a number rises. A prospective study should specify outcomes, comparison groups and analysis in advance, including symptom worsening, daily function, burden of recording and adverse effects. Eligibility, sample size and safeguards need a study protocol; they are not established here.

Interpretation for use

The most important unresolved risk is treating a fitted allowance as physiological capacity. The present method can raise an allowance after apparently stable next-day readings while missing delayed or non-energy symptoms. Rest credits and biometric adjustments add further assumptions. A release design should preserve user control, avoid automatic activity escalation, communicate insufficient evidence and allow estimates to be revised downward when symptoms worsen. These are

requirements for responsible use, not claims of completed validation. NICE's person-led energy-management principles provide the clinical context. [1]

Appendix A Reproducible parameters

These values describe the inspected implementation. They are engineering parameters and should not be interpreted as clinical thresholds.

Component	Specified behavior
Energy and budget	Energy 1–10 in half steps; k default 2.4; k bounds 1.2–4.0; budget rounded to whole spoons.
Budget learning	At least 3 sustainable pairs; 80th percentile; drop cap after at least 3 drop-associated pairs; prior weight 10.
Activity learning	At least 2 records; prior SD 1.5; observation SD 1; 90 percent mean interval using 1.645.
Lag regression	At least 21 qualifying rows; 2 expenditure lags; ridge penalty 0.5; unpenalized intercept.
Episode detection	Up to 14 earlier energy records; at least 7; 1.5 SD and 1.5 energy points; energy SD floor 0.75.
Recovery and consistency	Return to within 0.5 points of baseline; consistency over the last 7 logged days.
Numerical precision	Fitted k and cost outputs to 2 decimals; lag coefficients to 3; adjusted positive calendar costs to quarter spoons.

Wearable score transformations

Let $C(x)$ constrain x to the interval from zero to one. Sleep duration h in hours becomes $s = C((h - 4) / 4)$. If sleep efficiency e is available as a fraction, s becomes the average of that duration score and $C((e - 0.70) / 0.25)$. An HRV ratio r becomes $v = C((r - 0.60) / 0.60)$. A resting-heart-rate difference d in bpm becomes $u = C((10 - d) / 20)$.

$$Q = \frac{0.40s + 0.35v + 0.25u}{\text{sum of available weights}}$$

Only available terms enter the numerator and denominator. Suggested energy is $1 + 9Q$ minus the one-point expenditure penalty, then constrained to 1–10 and rounded to a half step. For seven hours of sleep, HRV ratio 0.90 and resting heart rate 2 bpm above baseline, with no efficiency value and no penalty, $Q = 0.575$ and suggested energy is 6.0.

Assumptions that need empirical calibration

The model treats energy-scale steps as numerically comparable, recorded spoon costs as additive, activity labels as stable groupings and fixed observation noise as appropriate across people. None of these assumptions follows from the clinical guidance. The implementation provides bounded calculations, not evidence that the chosen scales, priors, cutoffs or wearable weights are correct for a particular condition or individual.

References and implementation provenance

Clinical references provide context for pacing and delayed symptoms. They do not endorse Juno or validate the numerical method described here. Web sources were checked on 21 September 2026.

Clinical sources

[1] National Institute for Health and Care Excellence. Myalgic encephalomyelitis or encephalopathy and chronic fatigue syndrome diagnosis and management. NICE guideline NG206. Published 29 October 2021. Recommendations 1.11.2-1.11.7 on energy management and 1.11.14 on interventions that should not be offered. <https://www.nice.org.uk/guidance/ng206/chapter/Recommendations>

[2] Centers for Disease Control and Prevention. Strategies to Prevent Worsening of Symptoms. ME/CFS clinical care. 10 May 2024. Sections on post-exertional malaise and activity planning. <https://www.cdc.gov/me-cfs/hcp/clinical-care/treating-the-most-disruptive-symptoms-first-and-preventing-worsening-of-symptoms.html>

Implementation sources

The method was inspected in the Juno development branch `feat/pacing-statistical-model-20260914` at commit `260c4ae18c7634338b1e1c018169c982c2205405`. File paths below are relative to that repository. A separate machine-readable provenance record retains file hashes and numerical example outputs so this document can be reconciled with later revisions.

[3] `supabase/functions/_shared/pacing-statistics.ts`. Numerical definitions for budget fitting, Bayesian cost updates, lag regression, episode detection, consistency, wearable rules and the schedule balance.

[4] `supabase/functions/pacing-insights/insights.ts` and `index.ts`;
`supabase/migrations/20260914120000_pacing_statistical_model.sql`. Payload assembly, lookback windows, record selection and optional persistence.

[5] `supabase/functions/pacing-estimate/index.ts` and `app/lib/pace-burndown.ts`. Initial language-model estimates, deterministic adjustments, rounding and schedule adaptation.

[6] `docs/pacing-statistical-model.md`. Development checklist dated 14 September 2026, including incomplete feedback, visualization, deployment and device-verification work.

Scope of this paper

This paper describes a Juno product method and its development implementation. It is a technical account rather than a report of a clinical trial. Worked examples use illustrative inputs and contain no patient records.