Supplementary Information

Candidate availability and uncertainty-aware selection shape the reliability of dairy rations under variable feed composition

Wei Liu and Seojin Jang

DeepFarm Inc., Seoul, Republic of Korea

Navigation and protocol scope

The current methods and results (S29–S38) are presented first. Section identifiers are retained across revisions and therefore need not follow the PDF page order. Sections S1–S27 document earlier methods and results and are included in the accompanying historical archive. S28, a separate R5 version-and-provenance archive, is not required for current results and is available from the corresponding author. The historical files retain the original assumptions and negative results and document prior-result inspection and random-stream reuse. Numbers labelled R5 belong to that version, not to the current main figures or tables.

The main manuscript examines candidate availability, full-pool selection, measurement uncertainty, transfer and high-risk cases. Each protocol has its own menus, scores, populations and inference families. The current protocols comprise 78 test cells, 498 risk records, 576 prespecified pairs and 96 finite-menu records, with the six later S36 exploratory pairs and the 30 exploratory S37 LP-control records counted separately. These counts refer to simulation and policy records; they are not biological sample sizes.

Historical evidence index

Archive	Contents and identifiers
Historical methods/results	S1–S5 inputs, model and original strategies; S6–S20 original diagnostics, sensitivity, prior access and case-history outcomes; S21 restricted-input status; S22 first extension; S23–S27 second extension, integration, all results and limitations
Historical R5 narrative	S28; complete earlier main narrative, ten R5 figures and eight R5 tables, with version-specific numbering
Current supplement	S29 components; S30 reproduction scope; S31 candidate contrasts; S32 shortage definitions/history; S33 information/transfer; S34 numerical scope; S35 full pools/measurement; S36 continuous candidates; S37 observed-composition LP re-solve control; S38 released rations and derived requirement amounts

Historical references to main Table 5 and Figs. 2–9 point to R5; S28 retains the complete objects. Figures and comparisons from the earlier seven-rule experiment, including former main Fig. 2 and Table 2, are retained but do not substitute for the later S35 tail analysis.

Supplementary Methods: Nutritional target classification and component execution (S29)

Four execution-driver contexts were restored in full, without additional sampling. The historical problem objects and the constraints applied during execution were examined separately because the driver adds the planned starch-source premise of Table 5-1. R0 and A had 21 constraints

during execution, and B and C had 22; each context included all nine main-event rows. CP retained its original definition as an illustrative intake target. It was not reinterpreted as a measure of MP/EAA/RDP adequacy or reduced after the risk results were known. Ca/P requirements, all thresholds, the planned-DM equality, fixed fMCP, support/domain semantics and numerical tolerances were unchanged.

The three lactation demonstrations packaged with the public `nasem-dairy` 1.0.2 release [1] formed the predefined component-comparison set. A schema-only API call preceded the numerical protocol. Unit tolerances were kg/day 10^{-10}, g/day 10^{-8}, Mcal/day 10^{-8}, Mcal/kg 10^{-10}, percentages 10^{-10} and fractions 10^{-12}. The adapter kept external DM-basis `kg_user` distinct from project as-fed q and handled g versus kg and absorption fractions versus percentages explicitly. Initial adapter/unit failures are retained internally; neither source implementation nor frozen research model was changed.

Of 93 comparisons, 68 formula-and-unit checks passed, four showed expected differences in the open-cow pregnancy branch, and 21 could not be assessed because inputs were missing for used feeds. Missing nutrient values in used ingredients were not imputed, and unused ingredients retained their canonical zero contribution. The comparison also retained 15 full-versus-selected energy records. Supplemental NPN and an out-of-range predicted RUP placed the early-lactation demo outside the selected chain; it was reported as non-comparable, with the full feed list retained. The two comparable external cases showed finite supply differences between the full and selected chains that could not be attributed to fixed fMCP alone. The reported policy risks were unchanged.

A separately written selected-chain calculation passed 102 arithmetic checks on complete, explicitly synthetic fixtures. Expected intermediate values were obtained without the original project helpers. Seventeen boundary checks covered missing used inputs, zero and very small positive total DM, non-clipping of lignin above NDF, negative-ROM diagnostic status, both sides of the original energy tolerance and mutual exclusivity of violated and undefined verdicts. The published release's urinary-energy coefficient differed from the retained book equation; the version-specific equations were documented without forcing numerical agreement.

These checks support the consistency of the component implementations. The formal case policies were not re-certified against a complete external NASEM event. Independent biological validation of the illustrative CP target, voluntary intake and response prediction remains unresolved. Exact restored animal and configuration values, the printed-table requirement values (the illustrative CP target and Table 5-1 bounds), third-party assets and their technical exports remain restricted; the equation-derived requirement amounts are released in S38.

Numerical size of component and structural differences

For an open cow, the project returns zero pregnancy Ca and P, whereas the literal external branch gives 0.00130358 and 0.00144588 g/day, respectively; factorial totals inherit these differences. `R7_component_comparison.csv` holds the 93 component rows and 15 structural rows with both values, units, differences, applicable tolerances, comparability and implementation versions.

Executed observation, energy and constraint dependencies

The seven observed fractions were CP, NDF, starch, EE, ash, Ca and P (kg/kg DM); actual DM was distinct from these observations and hidden from nutrient-informed selection. Smooth energy calculations used sampled CP, NDF, starch and ash as percentages of DM, fatty acids from the separate fixed TFA/Fd_FA input, and fixed lignin. No sampled EE-to-FA conversion was executed, so EE had no smooth nonlinear energy derivative or direct primary delta/UA variance contribution.

Supplementary Table S1: Executed component and reproducibility checks and their limitations. Software and fixture cases do not represent new animals.

Check	Observed outcome	Interpretation
Published implementation, three lactation demos	93 items: 68 pass, 4 expected version differences, 21 not assessable	Unit/formula consistency; used-feed missing cells not zero-filled
Selected versus full energy chain	15 structural-comparison records retained	Non-equivalent modules/domains; no full-event policy re-certification
Independent selected-chain arithmetic	102 checks passed	Explicitly synthetic complete fixtures; not biological validation
Canonical energy boundaries	17 checks passed	Missing used inputs/zero DM/energy tolerances and verdict semantics preserved
Clean-environment saved replay	3 existing cells; 420,000 assigned labels, 336 independent choices, 3,072 hypothetical comparisons	No new draws or full-study regeneration; external replay requires the restricted inputs, available from the authors under licence conditions
Whole saved-policy shortage analysis	504 records, original loss/unknown arrays match	All nine row masks independently reconstructed; descriptive, post hoc

Component comparison used `nasem-dairy` 1.0.2 with predefined unit tolerances; the four expected differences concerned the open-cow pregnancy branch. All 21 unavailable comparisons and structural differences remain in Supplementary S29. Policy-state positions repeat saved states across rules and are not independent observations.

Supplementary Table S2: Public demonstration energy values (Mcal/day). Differences are selected minus complete implementation; relative differences use the complete value. Comparisons are structural, not equality tests. Dashes denote the early-lactation example, which could not be assessed with the selected chain because of supplemental NPN, out-of-range RUP and missing inputs. Sensitivity of policy rankings or risk targets to the complete implementation was not tested.

Public case	Component	Full model	Selected chain	Difference	Relative (%)
Test	DE	72.899305	72.921115	0.021810	0.0299
Test	GasE	7.023147	7.023147	0.000000	0.0000
Test	UE	5.900245	6.972107	1.071862	18.1664
Test	ME	59.975914	58.925861	-1.050052	-1.7508
Test	NEL	39.584103	38.891069	-0.693034	-1.7508
Early lactation	DE	74.647611	–	–	–
Early lactation	GasE	4.819062	–	–	–
Early lactation	UE	6.277105	–	–	–
Early lactation	ME	63.551444	–	–	–
Early lactation	NEL	41.943953	–	–	–
Other	DE	78.059005	78.109067	0.050061	0.0641
Other	GasE	5.202967	5.202967	0.000000	0.0000
Other	UE	3.295126	4.178363	0.883238	26.8044
Other	ME	69.560913	68.727737	-0.833176	-1.1978
Other	NEL	45.910203	45.360306	-0.549896	-1.1978

Supplementary Table S3: Actual variable dependencies. "Primary" is the nine-row event; all missing and point-mass semantics are in the full variable-path table.

Input	Status	Executed primary-event role
CP	Sampled fraction	Protein supply and smooth energy; nutrient posterior derivative
NDF	Sampled fraction	T1/T2/T3/T5 and smooth energy; nutrient posterior derivative
Starch	Sampled fraction	T4/T5 and smooth energy; nutrient posterior derivative
EE	Sampled fraction	Discrete energy support only outside $[0, 1]$; no smooth score direction
Ash	Sampled fraction	Smooth energy; nutrient posterior derivative
Ca	Sampled fraction	Absorbed calcium; nutrient posterior derivative
P	Sampled fraction	Absorbed phosphorus; nutrient posterior derivative
DM	Sampled separately	Supply/concentration/energy; hidden-DM variance in nutrient-informed scores
FA	Fixed % DM	TFA/Fd_FA energy input; no sampled EE conversion
Lignin	Fixed % DM	Energy and support; not clipped to sampled NDF

EE was included in the companion upper-EE row, the historical nominal planning row and the analytical composition-overlap diagnostic. It also entered the discrete support check: an EE value outside $[0, 1]$ in a used ingredient invalidated energy support, whereas the declared distributions and censored observations remained within $[0, 1]$. Missing EE was retained without imputation. The delta and UA scores have no numerical dependence on EE for the primary event; other observation interfaces, such as the conditional-DM interface described below, can still use it.

The generic conditional-DM implementation can use EE through a declared copula. All 24 executing-host C0/C1/C2 exports, however, had exactly zero EE off-diagonal correlations and conditional-DM Schur weights, so they supplied no indirect EE information about DM. The conditional interface still rejected nonfinite stochastic EE and checked marginal supports at zero EE weight. The measurement rules instead used independent field posteriors and unchanged hidden-DM moments. These findings apply to the implemented model and the exported matrices; they do not exclude a predictive relationship between EE and DM in other settings.

Centred finite differences on six representative saved A/C states gave 42 energy and 42 nonenergy derivative comparisons, with a largest scaled energy discrepancy of 3.97×10^{-9}. Thirty EE path checks and 14 compiled EE-row checks passed. These limited diagnostics are not global derivative bounds or new risk estimates. Full source-derived paths are in `R7_variable_path.csv`, `R7_EE_companion_and_planning_rows.csv` and `R7_EE_conditional_information_path.csv`.

Supplementary Methods: Reproduction scope and saved-state replay (S30)

The R5 aggregate archive allows counts, rates and paired intervals to be recalculated but omits complete numerical inputs and executable dependencies, so aggregate verification is not full simulation reproduction. Additional internal source files, data and saved states enabled a replay using restricted inputs.

A read-only execution trace, with sampling disabled, covered four cases, TAB/NARROW contexts and loading of the complete original and price pools. It recorded 145 repository files opened during execution, two helper files and 60 parent candidate files, with SHA hashes computed from the file contents. A later clean-copy failure identified six additional dependencies accessed only by file-stat or existence checks. These were documented in an existence-only technical supplement after the new attribution protocol was frozen and execution had begun; the 145-file scientific content map, registered protocol and model values were unchanged.

The internal minimal bundle contains 315 dependency files: frozen original source/helpers, real restricted inputs, three completed root0 C0/TAB cells for R0/A/C and exact execution-host fit metadata. Its dedicated clean environment used Python 3.14.7, numpy 2.4.6, scipy 1.17.1 and PyYAML 6.0.3; historical environment locks were kept separately and the original research environment was not modified. Restricted directories and files use modes 0700 and 0600. Byte comparisons of the source, helpers, protocol and delivered-R5 files passed before and after preparation.

With all sampling entrances disabled, the canonical original reference checker matched, in the three existing cells, the 420,000 assigned policy-state labels, unknown flags, counts, statistical endpoints and 27 paired comparisons (pointwise and adjusted). Independently written selection logic matched 16 fixed saved states per cell for seven rules (336 policy-state choices) and 3,072 limited hypothetical chain comparisons; selection finished before actual saved DM was loaded. The replay did not repeat the complete 72-cell independent policy selection, optimisation or source fitting, or regenerate the original random streams. Technical packaging failures were logged as preparation records.

A separate nonrestricted arithmetic demo, built from invented dimensionless values and standard-library code, illustrates the input, selection and evaluation interfaces. It does not execute the nutritional model and therefore cannot reproduce the reported risks or validate model components. Access to the restricted true inputs is governed by the source-table licence and is arranged through the corresponding author; the released code version and the Supplementary Data are identified in the Data and Code availability statements. The locally prepared controlled-review material also contains readable exact-input tables for all four reference cases, original targets and recipe vectors, and ten frozen-ration illustrations, whose saved test-state index was fixed at zero before extraction (no new draws or deployment-rule choice). These restricted technical materials are separate from the public aggregate tables.

Implementation-check coverage across protocols

The final independent saved-array audit covered 78 formal cells (498 risk records, 576 prespecified paired comparisons and 96 menu-floor records) and reconstructed 9,960,000 assigned policy–state labels with their statistics. A separate clean environment canonically re-evaluated one complete saved cell from each of the six new protocol families (740,000 policy–state outcomes; 151 actual/preflight input files verified in a 315-file dependency closure); this six-cell check did not extend nutritional re-evaluation or selector reconstruction to all 9.96 million positions. The full-pool protocol also underwent complete selected-action and unpruned-score replays, with the larger protocol-specific scope described below. Overlapping audit scopes are reported separately. A later complete replay, performed in a newly prepared clean environment using restricted technical-review inputs, evaluated the saved assigned actions on the saved states in all 78 formal cells (9,960,000 policy–state positions; 151 runtime input files hash-verified before and after execution); no new draws, retraining, source fitting or policy selection occurred. The original evaluator reproduced all saved known and unknown outcomes. Independent redevelopment of all nutritional equations, selectors and random-stream factories, and repetition by an external researcher, lie outside this replay.

Supplementary Methods: Frozen matched attribution experiment (S31)

The matched-attribution protocol was fixed before formal sampling (SHA-256 `c78e651a2a7511d14a94d333bbd73c34d97368ccaf96d367d42f95c0164a26e8`). It retained the original

main-reference plan-domain failure event, source, targets, planned supply and delta information interface. Each of its 18 cells used 4,000 training and 20,000 fresh test states, and all formal executions exited with status zero. Outputs comprise 180 principal menu-risk records, the 12 actually available terminal SLP records, 252 prespecified same-state pairs and 90 finite-menu floors; principal and SLP rows remain distinguishable.

Menus Q_0–Q_4 are defined in main Section 4.3. Q_0 and Q_1 had 332/515 A actions and 353/894 C actions; physical mixtures enlarged these to 599 A and 978 C actions. On the three TAB replicates, compression gave 18–24 A and 11–15 C actions, and uncapped restoration 347–349 A and 363–367 C actions. Because the earlier price pool combines several algorithmic sources, improvements following pool expansion cannot be attributed to a particular generator or source-specific objective.

All six A/C TAB Q_1 versus Q_0 contrasts improved for both fixed and delta rules. Physical mixing gave no adjusted significant fixed-rule increment; delta improved in two A replicates only. Compression preserved the fixed-rule action in every cell but worsened five of six A/C delta comparisons. Every Q_3 replicate in A and C still met the respective 5%/2% criteria. For C on Q_2, adjusted 1% qualification passed in replicates 0 and 2 only (upper endpoints 0.955724%, 1.109024% and 0.997209%). Integer binomial tails independently verified these three boundary decisions and the six principal Q_3 target decisions. Replicates were not pooled, and no new rule was selected from the test results.

The complete copied aggregates are `R6_attribution_risks.csv`, `R6_attribution_pairs.csv`, `R6_attribution_fixed_menu_lower.csv`, `R6_attribution_integer_primary_goals.csv` and `R6_attribution_SLP_slots.csv`. They retain R0/B high-risk controls, NARROW failures and unresolved pair comparisons. Delta Q_3 risk was 90.445–90.770% in R0 and 93.000–93.480% in B. In the six A/C NARROW cells, original delta had zero observed failures, whereas new compressed menus had 3–22 failures per 20,000; zero observations do not certify zero risk.

The independent audit of the saved arrays reconstructed all 3,840,000 principal-plus-SLP assigned positions with their nine-member main event, all 192 summaries, 252 pairs and 90 floors, with beta endpoints computed without the original statistics helper. Independently written code reconstructed fixed-rule training ties and all compression steps, and 16 deterministic saved states per cell yielded 2,880 action/fallback comparisons. Because frozen scoring primitives were reused, this is a limited independent choice check, not a full rewrite of scoring or every physical mixture operation. Test candidate flags were bit-packed. In states labelled as all-action failures, every menu action was confirmed to fail by the canonical evaluator; successful states had a canonically verified successful action. An all-candidate spot check at fixed states supplemented these checks, but not every candidate–test pair was independently recalculated.

The oracle uses actual DM only for diagnosis. Its statewise residual denotes a selected-action failure despite the availability of a successful full-state option. Because the implemented rule cannot observe DM, the residual combines observation limits with excess risk over the best observable rule (three-term population-risk identity in main Methods) and is not scoring error alone. For example, two equiprobable hidden states with identical observations may require opposite recipes for success. The full-state floor is then zero, but every observable rule has risk at least one half. The retained original SLP (cost objective, 128-state prefix, three rounds, per-round budget) differs from the 4,000-state risk-first fixed selector and remains continuous search after pool-based initialisation. Neither that comparison nor a finite-menu floor rules out new continuous recipes. Rate, paired and oracle families were each assigned their own significance budget of 0.05/10,000, so each 95% statement applies within one family.

Compression reduced full-state coverage in all six pools. The residual fell in two comparisons and rose in four, so compression cannot be described as uniformly increasing scoring mistakes. Values were reconstructed independently from saved candidate failure flags and selected actions alone. Full counts are in `R7_2_compression_full_information_decomposition_6.csv`.

Supplementary Table S4: Compression minus full-pool failure counts in the six attribution cells (20,000 states each). The residual is total failures minus full-information all-action failures.

Case	Replicate	Total	All-action failure	Residual
A	0	73	79	-6
A	1	83	85	-2
A	2	133	91	42
C	0	85	63	22
C	1	125	77	48
C	2	85	76	9

Supplementary Methods: Post-hoc shortage magnitude, unknown and fixed-fMCP sensitivity (S32)

This analysis was specified after the second-extension risk results were known. It reused all saved states and seven frozen policies, without new sampling, retraining, changes to targets or test-based policy selection. It covered 10,080,000 assigned policy-state positions and 4,536 row summaries. Historical M0 recipes were independently recovered by `method_id=M0_nominal`, source MAIN9 cell and evaluation world, and source-replicate recipes were deduplicated by exact bytes. The additional 1,440,000 M0 positions and 648 row summaries form a separate descriptive addendum and are excluded from the original seven-label agreement claim.

Independently implemented constraint formulas and the selected energy chain reproduced every saved loss and unknown label, along with all defined and violation masks. Maximal margin discrepancies were 8.53×10^{-14} for the original policies and 6.22×10^{-14} for M0, in their stated units. Output, saved-input and source-table hash readbacks passed. A hash-namespace adapter failed before any canonical state replay; its script and log are retained internally. This floating-point verification is not a formal proof covering all machine and libm operations.

Shortage definitions and normalisation are in main Section 4.10. Unknown outcomes remain part of the original failure event and are excluded from magnitude summaries only where the magnitude is undefined. CP denominators are illustrative case targets rather than validated protein requirements; Table 5-1 rows are normalised by their absolute constraint bounds, not by nutrient requirements. Norm-only aggregate rows are in `R6_candidate_normalized_severity.csv` and the distinct `R6_M0_candidate_normalized_severity.csv`. These severity summaries exclude exact physical shortages, concentration values, targets, state-dependent requirements and recipes.

Expanded delta lowered energy CVaR95 relative to old static and old delta in every A/C TAB cell. Across structures and replicates, its target-normalised ranges were 0.3389–0.5570% in A and 0.1183–0.2105% in C. Historical M0 ranges were 4.8860–5.0358% and 4.8465–4.9973%. Old static ranges were 1.7378–2.6600% and 2.2393–2.4578%. Old delta ranges were 0.5449–0.8126% and 0.6972–0.9290%. These are descriptive tail comparisons on the same states, without a new multiplicity-adjusted hypothesis test of severity.

Phosphorus CVaR95 was higher than under old static in two A cells and three C cells, and CP CVaR95 was higher in one C cell. Calcium CVaR95 was higher than under old delta in one C cell. A's rare calcium violations had conditional mean shortages up to 30.48% of the relevant model requirement, at rates of only 0.005–0.020%. Complete counts, conditional means/medians, P95 and tail means are in the norm-only summaries. These counterexamples preclude a claim that every nutrient shortage was reduced.

Among the original seven rules, 454 event-unknown positions were exclusively energy-support violations. There were 1,338 support-reason occurrences across all states, of which 884 were

already classified as known violations through another row; cause flags can overlap. Historical M0 had three additional unknown positions. Negative-ROM and related anomalies remained diagnostics. The norm-only reason and overlap tables retain their denominators, while restricted physical-input details remain internal. These repeated policy-state counts are neither a pooled failure rate nor a count of animals harmed.

For fixed chosen actions, the selected-chain derivative is

$$\frac{\partial \mathrm{NEL}}{\partial f_{\mathrm{MCP}}} = 0.66(0.0146 \times 160 - 5.65)D/1000 = -0.00218724D.$$

Units are Mcal/day per g/kg DMI. Applying ±10% and ±20% changes to original fMCP, with undefined masks and all other rows retained, produced 2,520 descriptive sensitivity records. Expanded delta TAB rates at 1.2× fMCP were 3.020–4.070% in A and 1.495–1.875% in C. These hypothetical saved-state rates describe sensitivity to the retained selected-chain coefficient; the original certificates are unchanged, and no new family-adjusted target certificate was generated.

Supplementary Methods: Information noise and two directional frozen-policy transfers (S33)

The forward information experiment comprised 18 cells, 162 risk records and 198 paired comparisons, and a separately frozen reverse experiment added 12 cells, 24 risk records and 12 pairs. The aggregate tables `R6_information_*` and `R6_reverse_*` retain all 186 rates and 210 within-target pairs. Replicates were analysed separately, each rule retained its source-population training, and every paired test compared outcomes within one target population. The saved data were read and aggregated without modification.

For the forward rules, training under C0 fixed the compressed menu and 64 empirical joint nuisance node pairs. I0 used these node pairs with no observation. I1 combined the actual observed DM with the same nutrient nodes, and I3 combined ideal observed non-DM nutrient concentrations with the same DM nodes. All three used the same complete joint-event mean-loss score, with frozen cost/hash tie-breaks and fallback. For the C1/C2 tests, the empirical construction remained based on C0; it was not a target-updated conditional model. The nine compared rules also included DM and nutrient observation noise at 0.1 and 0.5 times the field-specific training SD, the original 4,000-state risk-selected fixed rule and the same-menu delta rule.

Observation noise was independent Gaussian at the declared field scale. Only observations were censored to the physical fraction $[0, 1]$; the true simulated inputs, fixed energy and rumen parameters, unknown inputs and marginal draws entered the evaluation unchanged. The noisy plug-in rule treated each censored observation as exact, without a censored likelihood or posterior variance. A diagnostic comparing 16 with 64 nodes on fixed saved states measured disagreement in score and choice; it neither certifies integration error nor defines a new risk-tested policy. The node diagnostic and clipping counts are reported in separate tables.

Under the shared finite-node score, ideal-nutrient I3 improved on I0 in all 18 adjusted pairs and true-DM I1 did so in 15, with three I1 pairs unresolved. DM-only observation was worse than nutrient-only observation in 12 cells and unresolved in six. With ideal nutrients and with 0.1-SD noisy nutrients, a 5% adjusted upper criterion was met in all 18 cells; the maximum upper confidence bounds were 4.382562% and 4.452426%, respectively. Nutrient noise of 0.1 SD worsened two pairs and left 16 unresolved, while every interval comparing DM noise of 0.1 SD with true DM contained zero. At 0.5 SD, nutrient noise worsened all 18 comparisons with ideal observation and DM noise worsened 12, leaving six unresolved. High versus low noise likewise worsened 18 nutrient and 12 DM cells. Risk under 0.5-SD nutrient noise was 3.745–9.850%, and only two cells passed the adjusted 5% criterion.

The scoring controls also showed differences between rules. I3 was worse than the same-menu ideal delta rule in 14 adjusted pairs, with four unresolved, and I0 was worse than the 4,000-state fixed selector (static4000) in 12 pairs, with six unresolved. I0 and the fixed selector differ in node count and frozen rules, so they are two distinct no-information selectors. All of these contrasts describe finite-node policies rather than the global Bayes value of an information sigma-field.

Frozen at C0, the delta rule retained case A's 5% target in every C0, C1 and C2 target replicate. Case C retained 2% in all C1 and C2 replicates but in only two C0 replicates; the failing record, C0 replicate 1, had observed risk 1.670% and upper bound 2.105826%. The observed C2 rate of 0.890% left the 1% target uncertified. Each target population was tested on its own replicates, so the directional rates are descriptive and unpaired across distributions.

The reverse experiment read the original second-extension C2 frozen recipe, price/key, static-baseline and achieved-DM-moment files byte for byte, and tested six historical policies in both the current local C2 and C0 source-factory populations. The local fits and correlations differed in their last bits from the original Linux model exports; these populations are therefore not byte-identical reconstructions. Expanded delta improved on the same-menu frozen fixed rule in all 12 adjusted pairs. For case A, risk was 1.840–2.275% in matched C2 and 2.490–3.085% after transfer to C0, and every A record retained 5%. For case C, the corresponding ranges were 0.865–1.230% and 0.875–1.310%; all C records retained 2% and none certified 1%. The forward menus and training construction differ from these historical C2 menus, so the gap between the two experiments cannot be attributed solely to the source of dependence.

Every assigned outcome was evaluated against the unchanged canonical main event (known violation OR unknown) under the original supplied-DM-consumed, CP-target, fixed-fMCP and Table 5-1 premises. Rate and paired families each allocated 0.05/10,000 per claim. The directional tests do not establish field robustness, calibrated assay accuracy, complete NASEM validity or an overall joint 95% statement. The later full-pool and measurement protocols used their own frozen rules and tests and are reported separately in S35. An independent audit of the saved arrays across the 30 forward and reverse cells reproduced all 3,720,000 saved assigned labels, the selected-action and cost mappings, the 186 risk summaries and the 210 within-target pairs. A separate canonical spot check at fixed indices matched 2,976 policy-state outcomes and 26,784 individual row labels, with a maximal margin difference of 4.62×10^{-14}; it was not a full independent canonical replay of all 3.72 million positions, nor a complete selector rerun. The first adapter failure, in direction naming, was retained and corrected without changing any scientific output.

Supplementary Methods: Limited high-precision numerical audit (S34)

An exploratory numerical protocol defined 96 one-dimensional slices before evaluating any error: 16 deterministic saved observations and three deduplicated frozen actions in each of the A/C TAB/C0 replicate-0 cells. Slice selection did not use observed risk or score error. Observed nutrients were held fixed, and the non-pivot DM coordinates took the original hypothetical training means. The audit read no saved actual test DM and repeated no source fitting, training or random draw.

An independent `mpmath` 1.3.0 implementation evaluated the retained Chapter 3 selected chain and the raw constraint definitions at 80 decimal digits. Multiplying the energy margin by the square of total DM gives a polynomial, whose coefficients were determined from five positive-domain points. Derivative critical points, interval bisection and real roots then partitioned the pass set. A separately implemented truncated-normal CDF integrated the intervals from the exact exported binary64 parameters. Because C0 has independent DM and nutrient blocks, the

audit covers C0 only and leaves the conditional CDFs of C1/C2 outside its scope. High precision gives a numerical reference, not an interval-arithmetic proof.

All 672 ordinary positive-domain classification checks agreed with the canonical event. The largest normalised energy-chain errors were 6.25×10^{-16} for A and 9.57×10^{-16} for C, and the corresponding quartic errors were 4.19×10^{-16} and 3.54×10^{-16}. All 96 slice probabilities fell within the declared 10^{-8} check tolerance, with a maximum absolute difference of 2.301×10^{-14} and no ambiguity on any actual slice. The checks cover finite slices and do not amount to a uniform guarantee on score error.

The 672 near-root probes deliberately included adjacent floating-point values, rounded roots and small physical-DM offsets. Canonical verdicts differed from the 80-digit reference at 128 probes and from the binary64 quartic at 134, and every difference lay inside the predefined 10^{-10} Mcal/day threshold band. All three verdict series are retained in the internal numerical record. The continuous slice distributions assign zero probability to the endpoints, and the integrated probabilities agreed. Neither finding establishes identical floating-point classifications across the full state space. The original tolerance rules and reported risks were left unchanged.

Six distinct, nontrivial slices were also integrated directly against the same truncated-normal density. Direct density integrals and interval-CDF calculations agreed at working precision, with a maximum discrepancy of 1.424×10^{-80} after coefficient and endpoint serialisation. The comparison covers six cases and is not a mathematical claim of zero integration error. Nine deterministic synthetic polynomial cases covered disconnected pass sets, repeated and clustered roots, near-real complex roots, small leading coefficients, constant polynomials and nonfinite coefficients. Four partitions were accepted, four finite cases were flagged explicitly as ambiguous, and the nonfinite case remained NOT_VERIFIED and used the registered fallback. One disconnected example had four real roots and three pass intervals, so pass probability cannot be assumed monotone in the pivot.

Fourteen source-boundary fixtures covered used and unused missing inputs, lignin above NDF, zero and near-zero DM, and nonfinite or negative DM; their complete known-violation and unknown flags matched the canonical checker. In a local forced-ambiguity injection, the partly analytic integration score (RB16) fell back to exactly the same arrays as direct joint-indicator averaging on the identical frozen nodes. The injection tested the fallback implementation, not population reliability or the acceptance of missing farm observations.

An isolated local timing check used 16 observations and three actions with the frozen designs of RB16 and the Sobol-node joint-event score (Joint64). Measured as single-process score time, excluding menu construction, shortlisting, loading, canonical evaluation and I/O, RB16 took 0.03761 s for A and 0.04600 s for C, and Joint64 took 0.00752 s and 0.00861 s. The original runs logged whole-cell times only, so isolated score times cannot be recovered from them. These local timings and the one-dimensional high-precision work support neither a general multidimensional runtime comparison nor a general accuracy comparison.

The released aggregate numerical audit summary is `R6_candidate_numerical_summary.json`; exact physical roots, private recipes, thresholds and detailed source configurations remain restricted. Adapter failures and an initially incorrect interpretation of empty intervals were documented and corrected before preparation of this account; the original numerical rows were not rewritten. Uniform certification of RB16/Joint64 integration error and ranking regret, and global floating-point certification, remain NOT_VERIFIED.

Supplementary Methods: Full candidate pools and measurement uncertainty (S35)

Full-pool coverage with partial-score pruning

The coverage protocol was specified after compression loss had been observed and frozen before its formal test draws (SHA-256 `e0bd7726500625d4f631012aca06469c0c121a4c4ac867891d10203a18e622c1`). It reused the six previously frozen C0-trained pools for cases A and C and tested each in C0, C1 and C2 on independent fresh streams of 20,000 target states. Within each cell, full Q_2 and compressed Q_3 were evaluated on the same states; no pair crossed populations or protocols. All 18 formal cells completed with exit status zero. The original source, nutritional requirements, planned DM, fixed fMCP, structural and domain predicates and the known-violation OR unknown event were unchanged.

Pruning rests on a simple bound. The original proxy is $S(q,s) = \sum_{j=1}^{9} p_j(q,s)$, a sum of nonnegative normal-tail terms. Let $B(q,s)$ be the sum of the eight non-energy terms, so that $B \le S$. Once a complete score u is available, $B > u$ excludes q from the proxy minimum; candidates at equality are kept because cost and hash tie-breaks still apply. All full-pool rations remain available for selection. Compressed-pool rations serve only to initialise complete scores, and their membership places no restriction on the eventual winner. Nonfinite scores and the original fixed-rule fallback retain their original semantics. To guard against rounding, the implementation subtracts 128ϵ from the partial sum, applies a downward `nextafter`, preserves the original nine-term summation order and checks the bound whenever a full value is computed. These guards support finite-state numerical verification; they do not prove correctness under every platform or libm arithmetic, nor optimality for the original event risk.

All 18 adjusted paired intervals favoured full Q_2; none indicated worsening. Observed risks were 1.765–2.155% in A and 0.520–0.825% in C, with maximum adjusted upper confidence bounds of 2.6436630% and 1.1441189%, respectively. Every A and C test cell met its 5% or 2% criterion. In C/C2, the three replicate risks of 0.665%, 0.640% and 0.520% had adjusted upper confidence bounds of 0.9557236%, 0.9259850% and 0.7818149%, all below 1%. C/C0 and C/C1 each passed 1% in two replicates. Direct integer binomial tails verified all principal targets and every C 1% decision. Independently trained replicates were kept as separate policies, and the rate and four-tail paired families each allocated 0.05/10,000 per claim.

Independently computed statistics reproduced all 36 risk records and 18 pairs. Canonical replay of all 720,000 assigned policy–state positions reproduced their nine margins, known and unknown flags and joint outcomes; the replay covered every selected action but used the original evaluator rather than an independent rewrite of each nutritional equation. A separate full, unpruned original-proxy calculation, run without access to actual DM, covered all 360,000 saved nutrient observations and all 283,860,000 Q_2 candidate–state positions. It found zero action or fallback differences, and the largest selected-score discrepancy was 5.829×10^{-16}. The audit involved no new sampling or policy modification.

During formal execution, energy was computed for 47.7303–50.1590% of candidate–state positions and skipped for approximately 49.8410–52.2697%. Qualification runs showed only a small reduction in overall wall time. Full replay and formal execution ran under different load, so their timings cannot support a speed ratio. The first full-choice audit mistakenly passed an array containing derived energy through the nutrient-only interface, which rejected it before complete proxy replay. The failed scripts and logs were retained, and the corrected audit used the original observation constructor without changing draws or scientific outputs.

The aggregate outputs are `R6_coverage_risks.csv`, `R6_coverage_pairs.csv`, `R6_coverage_integer_primary_goals.csv`, `R6_coverage_integer_C_onepercent.csv`, `R6_coverage_independent_readback.json` and `R6_coverage_full_choice_receipt.json`. Complete state

arrays, recipes and protocol inputs remain restricted. The comparison shows improvement over this specific compressed rule; it does not show that enlarging every action space is beneficial.

Censored nutrient observations and uncertainty-aware selection

The measurement protocol was likewise frozen before its formal draws (SHA-256 `92a47d96586ca8f4c08d45a2b36150229f5ea96806459d9788f801afab8cfd19`). It reused the six C0/TAB parents, each with 4,000 saved training states, together with the same Q2 and Q3 actions, requirements and planned DM. Each parent received a fresh 20,000-state test and an independent noise stream. Ten predefined rules and eleven within-cell comparisons yielded 60 risks and 66 pairs. No candidates, training states or test-selected noise levels were added, and no pair crossed protocols.

For each finite model-input nutrient fraction, the observation was $Y = \min(1, \max(0, X + \lambda s Z))$ with $\lambda \in \{0.1, 0.5\}$, where s is the frozen training-sample SD. Both noise levels and all rules shared the same saved standard normals within each cell, and only the observations were censored. The seven observed columns were CP, NDF, starch, EE, ash, Ca and P, in kg/kg DM. Derived mineral or forage columns were observed as model inputs; this does not imply that a direct assay is available. Actual test DM was hidden. Missing values were preserved. Fields with exactly identical finite training values were treated as point masses, without applying a variance threshold. The earlier information protocol assigned rounding-scale pseudo-SDs to three fixed Ca/P fields, which remain part of its historical results; this separately frozen experiment corrected the point-field treatment, so the two protocols cannot be treated as paired replicates.

Each working prior was an independent Gaussian distribution with frozen empirical location and scale, truncated to $[0, 1]$. These working priors differed from the TN/BETA priors that generated the true states. Interior observations used Gaussian conjugacy followed by bounded-normal moment evaluation. Censored endpoints used the likelihoods $\Phi(-x/\sigma)$ and $\Phi((x-1)/\sigma)$, and the deterministic moment quadrature was cached before testing. Parent Gaussian parameters and the actual bounded-prior moments were stored separately. The posterior-mean rule inserted these means into the original delta score. The uncertainty-aware rule added the diagonal posterior nutrient-gradient variance to the existing hidden-DM variance, as defined in the main Methods, while retaining the original row senses, unit factors, concentration denominators and tolerances. Whole-reference energy gradients included every mapped stochastic CP, NDF, starch and ash direction; fatty acid (TFA/Fd_FA) and lignin stayed at their verified fixed inputs. Sampled EE was not converted to fatty acid, and S29 distinguishes its discrete support and companion-row roles. Covariances, higher-order products and nonlinear corrections to the predictive mean were omitted. All Q2 actions were scored, and Q3 used its original subset of the same score matrix.

The nine-term marginal-tail sum is a ranking score: even exact marginal failure probabilities add up to the expected number of violated rows, while risk is the probability of their union. Two coincident 4% failures, for example, give a sum of 8% but a union risk of 4%, whereas two disjoint 3% failures give a sum and a union risk of 6%, so the two measures order these examples oppositely. The normal-tail variance derivative in the main Methods provides another reason for disagreement: added variance can raise or lower a candidate's tail term. Either effect can change the ranking, but neither has been established as the mechanism responsible for the observed energy trade-off.

At 0.5 SD, using posterior means alone worsened selection relative to the raw-observation rule in all six adjusted pairs, by 2.400–4.280 percentage points. Adding posterior nutrient variance improved on that posterior-mean rule in all six (4.025–6.445 points) and on the raw-observation rule in all six (1.625–2.530 points). At 0.1 SD, all three corresponding Q3 contrast groups had intervals containing zero. Full Q2 improved on Q3 in five of six pairs under each of ideal delta, uncertainty-aware 0.1 and uncertainty-aware 0.5, with one unresolved interval in each group.

Each control identifies only its specified contrast: all raw, mean and variance comparisons use Q3, whereas Q2 versus Q3 compares the same uncertainty-aware score. Because the design does not cross every scorer with both pools, it cannot identify a Q2 uncertainty-aware-versus-raw contrast or exclude pool-by-score interactions.

At 0.5 SD, uncertainty-aware risks in A were 4.160–4.200% on Q3 (worst upper bound 4.860%) and 3.495–3.655% on Q2 (worst upper bound 4.275%). All three separately frozen replicates met 5%. In C, Q2 risks were 1.610–1.685% with upper confidence bounds of 2.039–2.123%, so no replicate met 2%, although all met 5%. No pooled certificate was used. All rates used one-sided exact binomial bounds at 0.05/10,000, and paired intervals allocated four tails at $0.05/(4\times10{,}000)$. Because these families are separate, they give no overall joint 95% statement.

Zero-noise qualification reproduced the original full-Q2 scores exactly, with identical choices on the saved qualification states and on 96 fixed formal diagnostic states. Analytic energy derivatives matched 1,472 canonical finite-difference checks (maximum scaled relative discrepancy 1.375×10^{-8}), and the non-energy variance matched explicit Jacobians built with the original sense, unit and concentration factors. All 1,092 precomputed endpoint integrations reported success, with a maximum recorded relative-vector error estimate of 1.580×10^{-13}; this is a diagnostic, not a formal global quadrature certificate. An independent audit of the saved arrays reconstructed all 1,200,000 assigned-event positions and reproduced all 60 risks and 66 pairs, including unknown and fallback outcomes, with a maximum interval discrepancy of 1.477×10^{-14}. It also reconstructed censoring counts from the saved truth, standard normals and frozen scales without new draws. This audit checks the saved labels and statistics; it does not independently re-evaluate the complete nutrition model or fully reconstruct the selector.

The aggregate outputs are `R6_measurement_risks.csv`, `R6_measurement_pairs.csv`, `R6_measurement_case_policy_ranges.csv`, `R6_measurement_paired_directions.csv`, `R6_measurement_noise_clipping.csv`, `R6_measurement_endpoint_diagnostics.csv` and `R6_measurement_readback_scope.json`. Absolute costs, exact scales, recipes, states and random-stream seeds remain restricted. All adverse posterior-mean results and unknown or fallback outcomes are retained. The results describe sensitivity within the model; it does not validate a farm assay, actual intake or full nutritional adequacy.

Shortage magnitude in full-pool and measurement policies

All saved test states of the full-pool and measurement protocols were analysed after their risk results were known: 96 policy–cell records and 1,920,000 policy–state positions. The post-hoc analysis involved no new draw, retraining, target change or test-based rule choice. It applied the S32 shortage definition to all nine constraints, kept the original row-classification tolerance and counted zero shortage for defined states with nonnegative signed margins. The upper-tail mean used the largest $\lceil 0.05 n_{\mathrm{defined}} \rceil$ shortages. In all 864 original-coefficient rows, fewer than 5% of defined states had positive shortage and the all-defined P95 was zero. The reported top-5% mean is therefore exactly $(n_{\mathrm{defined}}/\lceil 0.05 n_{\mathrm{defined}} \rceil)$ times the all-defined mean. Because this finite-sample identity includes zero padding, the top-5% mean carries no independent information about tail shape in these records. Because the multiplier is the same positive constant for both policies up to the at most three-state difference in defined counts, every direction count reported below for the top-5% mean is also the direction count for the mean shortage over all defined states, which is the quantity the main text and main Fig. 4 report. Undefined rows stayed in joint failure and were excluded only from that row's magnitude denominator. The CP normalisation uses the illustrative target, not an independently validated protein requirement.

Full Q2 lowered the NEL tail metric in all 18 coverage contrasts and in all six measurement contrasts under each of the ideal, 0.1-SD and 0.5-SD observation conditions. Tails of the other constraints moved in mixed directions, as the complete direction tables show. On Q3, the uncertainty-

Supplementary Table S5: NEL violation incidence and conditional shortage on Q3 at high noise, raw-observation → uncertainty-aware selection. Incidence (percent) uses all 20,000 test states; the conditional mean and P95 are percentages of the original NEL target among each policy's own NEL-violating states, so they do not support a matched-state severity comparison.

Case	Replicate	Incidence	Conditional mean	Conditional P95
A	0	2.700 → 3.530	1.138 → 1.099	3.566 → 3.417
A	1	2.740 → 3.430	1.143 → 1.119	3.339 → 3.297
A	2	2.710 → 3.595	1.098 → 1.029	3.378 → 2.947
C	0	1.115 → 1.460	1.222 → 1.109	4.024 → 3.729
C	1	1.410 → 1.700	1.120 → 1.082	3.496 → 3.297
C	2	1.295 → 1.665	1.334 → 1.128	5.254 → 3.821

aware (UA) rule raised NEL tails relative to the raw-observation and posterior-mean rules in all six comparisons at both noise levels. At 0.5 SD, it lowered CP, Ca, P, T1, T3 and T5 tails against each comparator in all six comparisons, while the formal joint-failure pairs favoured it. At 0.1 SD, the corresponding adjusted joint-failure intervals included zero. The posterior-mean rule lowered NEL tails but showed the opposite high-noise pattern on those six constraints.

The higher UA NEL tails persisted on the states with defined energy under both policies; definition status differed at no more than three of 20,000 positions. UA/raw tail ratios were 1.019–1.096 at 0.1 SD and 1.088–1.263 at 0.5 SD. UA/mean ratios were 1.032–1.142 and 1.676–2.499, respectively. At high noise on Q3, UA selection raised NEL violation incidence relative to raw-observation selection in all six comparisons but lowered the violation-conditional mean and P95 in all six; maxima moved in mixed directions. Each conditional distribution uses that policy's own violating states, so these summaries neither show smaller deficits on every matched state nor identify a causal mechanism. In all 96 records, the all-defined mean shortage equalled the positive-shortage frequency times the positive-shortage conditional mean. Positive shortages remain distinct from violations beyond numerical tolerance. Every NEL positive-shortage frequency in these records was below 5%, so the top-5% mean is approximately 20 times the overall mean, and a rise in it cannot by itself be read as more severe individual shortages.

All 1,920,000 original joint labels and row classifications were reconstructed. Five fixed state indices per record, 480 positions in total, were checked with the original evaluator and with separate selected-chain and mass-balance arithmetic. This audit adds limited arithmetic checks but does not independently reimplement the complete nutrition model or the selector. There were 484 undefined NEL rows: 478 event-unknown positions and six rows overridden by an existing violation. Every unknown arose from an energy support violation. The historical seven-policy tails in S32 come from different saved states and rules.

The 864 complete nine-row summaries, all 648 descriptive constraint differences, all 90 grouped direction rows and all 96 NEL denominator and identity records are supplied as `R7_candidate_normalized_latest_severity.csv`, `R7_descriptive_same_protocol_differences.csv`, `R7_tail_direction_summary.csv` and `R7_NEL_all96_rule_records_safe.csv`. Further range and common-defined diagnostics retain all specified cases and rules. These shortage summaries do not estimate harm to animals.

Conditional NEL maxima decreased in one of these six contrasts, were unchanged in three and increased in two. For A replicate 0, the maximum rose from 7.536% to 8.567% of the target. Neither the lower conditional P95 nor the lower joint-failure rate therefore shows that every severe deficit became smaller. The exact top-5% identity held in all 864 rows (maximum absolute discrepancy 3.47×10^{-18} in normalised units). In each record, the positive-shortage count equalled the tolerance-defined violation count; this equality was observed, not imposed by the metric

Supplementary Table S6: Tail-shortage directions for the candidate-pool contrasts. Each entry counts lower/equal/higher empirical upper-tail mean shortages for policy versus comparator. Coverage has 18 comparisons and every other column six. All columns compare full Q2 with compressed Q3; the column headings identify the ideal and uncertainty-aware measurement conditions. UA, uncertainty-aware selection; mean, posterior-mean selection; raw, raw-observation selection. The counts are descriptive and carry no tail significance test.

Constraint	Coverage	Ideal	UA 0.1	UA 0.5
NEL	18/0/0	6/0/0	6/0/0	6/0/0
Illustrative CP	10/0/8	3/0/3	3/0/3	6/0/0
Ca	7/4/7	2/1/3	3/1/2	5/0/1
P	11/0/7	5/0/1	3/0/3	6/0/0
T1	17/0/1	5/0/1	6/0/0	5/0/1
T2	0/15/3	0/6/0	0/6/0	0/6/0
T3	16/1/1	6/0/0	6/0/0	5/0/1
T4	1/14/3	2/4/0	2/4/0	2/4/0
T5	14/0/4	4/0/2	6/0/0	5/0/1

Supplementary Table S7: Tail-shortage directions for the measurement-score contrasts on the compressed pool. Each entry counts lower/equal/higher empirical upper-tail mean shortages for policy versus comparator. Every column contains six comparisons on the same Q3 pool. UA, uncertainty-aware selection; mean, posterior-mean selection; raw, raw-observation selection. The counts are descriptive and carry no tail significance test.

Constraint	UA/ raw 0.1	UA/ mean 0.1	UA/ raw 0.5	UA/ mean 0.5	Mean/ raw 0.1	Mean/ raw 0.5
NEL	0/0/6	0/0/6	0/0/6	0/0/6	6/0/0	6/0/0
Illustrative CP	5/0/1	5/0/1	6/0/0	6/0/0	2/0/4	0/0/6
Ca	6/0/0	6/0/0	6/0/0	6/0/0	1/1/4	0/0/6
P	6/0/0	6/0/0	6/0/0	6/0/0	1/0/5	0/0/6
T1	6/0/0	6/0/0	6/0/0	6/0/0	1/0/5	0/0/6
T2	0/6/0	0/6/0	1/5/0	2/4/0	0/6/0	0/4/2
T3	4/1/1	4/1/1	6/0/0	6/0/0	0/3/3	0/0/6
T4	1/5/0	1/5/0	4/1/1	3/2/1	0/6/0	4/2/0
T5	5/1/0	5/1/0	6/0/0	6/0/0	0/4/2	0/0/6

definition. Across the 96 records, NEL positive-shortage frequency among defined states was 0.475–3.596%.

The complete reanalysis is supplied in `R7_2_top5_identity_and_occurrence_864.csv` and `R7_2_high_noise_energy_frequency_and_conditional_severity_6.csv`.

Fixed-action sensitivity of the latest policies to the microbial-protein coefficient

This post-hoc analysis applied the hypothetical fMCP multipliers used previously, $0.8, 0.9, 1.0, 1.1, 1.2$, to all 96 full-pool and measurement policy–cell records: 18 coverage cells with two rules and six measurement cells with ten rules, each on 20,000 saved states. The execution plan was fixed after inspection of the earlier results and before these calculations; this was not a prospective experiment registration. All original actions, test states, targets, tolerances and missing-input masks were held fixed. No new sampling, training or policy selection took place.

Supplementary Table S8: Fixed-action fMCP sensitivity of the final full-pool policies. Entries are empirical failure-rate ranges (percent) across separately trained and tested cells, not confidence intervals. Ideal coverage ranges span nine cells per case; high-noise ranges span three C0 cells and use full-Q2 uncertainty-aware selection at noise 0.5 SD. Coefficient multipliers are hypothetical.

Multiplier	Ideal A	Ideal C	High-noise A	High-noise C
0.8	1.160–1.485	0.315–0.545	2.510–2.605	1.140–1.210
0.9	1.410–1.775	0.385–0.665	2.930–3.145	1.360–1.390
1.0	1.765–2.155	0.520–0.825	3.495–3.655	1.610–1.685
1.1	2.190–2.635	0.660–1.040	4.195–4.300	1.980–2.050
1.2	2.700–3.175	0.895–1.435	5.060–5.105	2.345–2.580

The selected-chain margin transforms exactly as

$$m_{\mathrm{NEL}}(f) = m_{\mathrm{NEL}}(f_0) - 0.00218724D(f - f_0),$$

with D in kg/day and f in g/kg DMI.

All other constraint rows and structural-domain definitions remained fixed, and these positive coefficient scenarios leave the selected chain's structural undefined causes unchanged. The analysis did not apply external demonstration offsets to simulation states, fit a microbial-protein distribution or evaluate the complete external NASEM model.

The 480 records comprise 9,600,000 transformed policy–state evaluations of the same 1,920,000 saved positions, so they are not additional independent samples. The 96 baseline records reproduced the original failure labels, counts and upper bounds. Canonical checks at deterministic evenly spaced, boundary-nearest, undefined and DM-extreme positions covered 43,200 policy–state evaluations across all levels. Labels and unknown masks matched, the maximum energy-margin discrepancy was 2.67×10^{-14} Mcal/day, and no evaluated position lay within 10^{-10} of the energy decision boundary. Input hashes were unchanged. The checks reuse the original evaluator rather than independently reimplementing the whole nutrition model.

Each sensitivity rate has a one-sided exact binomial upper bound in a separately allocated diagnostic family with $\alpha = 0.05$ and 10,000 potential claims (480 reported bounds). The 360 within-state comparisons report discordant counts, empirical differences and directions; like the bounds, they describe post-hoc hypothetical cases, keep replicates and source structures separate, and add no prospective certificate, paired interval or tail significance test.

At +20%, ideal full-pool diagnostic upper bounds stayed below 5% in all nine A cells and below 2% in all nine C cells (maxima 3.757% and 1.842%). In the high-noise full-pool test, empirical risks in A reached 5.060–5.105%, with diagnostic upper bounds of 5.777–5.825%. Empirical risks in C were 2.345–2.580%, with diagnostic upper bounds of 2.853–3.110%. All six high-noise point estimates therefore exceeded their original case targets, while C's diagnostic upper bounds stayed below the common 5% reference. The coefficient thus shifts absolute target attainment without invalidating the formal results obtained at the original coefficients.

Full Q2 kept lower empirical failure than Q3 in all 18 coverage comparisons and in all six measurement comparisons under each of the ideal, 0.1-SD and 0.5-SD conditions, at all five multipliers. At high noise, Q3 uncertainty-aware selection kept lower empirical failure than raw-observation and posterior-mean selection in all six comparisons at every multiplier. Low-noise differences were less stable: at +20%, three of six UA-versus-raw differences and one of six UA-versus-mean differences reversed direction relative to baseline, and the corresponding baseline adjusted intervals had already included zero. All comparison directions and adverse shortage summaries are reported.

The sparse-positive condition required for the top-5% rescaling held in 478 of the 480 sensitivity NEL records. It failed in two perturbed records, where positive-shortage prevalence exceeded 5% (maximum 5.371%), so the 864-row identity above describes the original coefficient only and is not a general identity under perturbation. On Q3 at high noise, UA selection kept higher NEL incidence and higher all-defined mean shortage than raw-observation selection in all six comparisons at every level. The baseline decrease in violation-conditional mean did not hold universally: one comparison reversed at +20%. Conditional P95 decreased in four of six comparisons at −20% and in all six at the remaining levels. The complete outputs include these adverse conditional-severity cases.

The complete normalised outputs are `R7_2_final_fMCP_sensitivity_480.csv`, `R7_2_final_fMCP_descriptive_pairs_360.csv`, `R7_2_final_fMCP_descriptive_ranges.csv` and `R7_2_final_fMCP_pair_directions.csv`. Because the check perturbs one component with actions held fixed, it does not identify a measured coefficient range, a robust optimum, the response of retrained policies or full-model adequacy.

Supplementary Methods: Training-generated candidate extension for R0 and B (S36)

The fresh-test protocol was frozen separately (SHA-256 `a573d157cb194dd986230b5a2eabbaaddaa964ec8dbb839c98ad1c31ecf7b6b1`). Training-only development, solver failures and qualification adapter corrections are kept as separate records and are not independent risk certificates. All six formal fresh-test cells completed with exit status zero, without retries, reseeds or changes to their scientific files.

The comparator, parent-pool delta, applied the delta rule to the Q3 pool frozen in `RRS-ATTRIBUTION-20261002-V1` for the same case, training replicate and C0 structure. It was neither the original-library Q0 delta rule nor a direct load from the second-extension pool. The R0 parents differed from both the Q0 and the second-extension pools. The B parents happened to match the second-extension recipe vectors, but they were still loaded from the attribution-stage Q3 pool. Internal keys such as `oldQ3_delta` and `legaloldQ3_delta` are kept unchanged for traceability; they correspond to parent-pool delta and legal-parent delta in this account.

Supplementary Table S9: Verified lineage of the S36 parent pools. Each parent is the attribution-stage C0 Q3 pool, and every parent vector is retained byte-for-byte in the new pool. Q0 intersections are exact recipe-vector counts, not action-name matches.

Case	Replicate	Parent Q3	Original Q0	Shared with Q0	New pool	Legal parent
R0	0	6	8	4	14	6
R0	1	8	8	4	12	8
R0	2	9	8	4	15	9
B	0	5	1	1	13	5
B	1	5	1	1	13	5
B	2	5	1	1	13	5

Exact comparisons confirmed this lineage in all six tests. Parent-pool and legal-parent action, proxy, fallback and outcome arrays agreed in all 120,000 saved states, and no parent action was removed. The comparison is therefore with the specified parent pool rather than a direct experiment against Q0. `R7_baseline_lineage_safe.csv` supplies all six lineage records without private recipe coefficients.

Candidate generation combined two max-slack linear starts, built from the mean and from all 4,000 saved training rows of the eight non-energy inequalities, with local SLSQP searches of smoothed full-reference deficits. The searches used the original structural constraints, requirements and planned supply. The resulting candidates were checked against the original canonical event and frozen before fresh testing. The new fixed rule chose the ration that minimised the training failure count over all 4,000 states, with deterministic feed-price and action-hash tie-breaks. The new delta rule kept the original training DM moments and ideal nutrient observations, with actual test DM hidden. Because the search is local and non-convex, it does not establish a globally optimal continuous formulation.

Six registered cells (R0 and B, each with three independently trained replicates) used 20,000 fresh test states per cell. The four frozen policies produced 24 risk records and 30 prespecified paired comparisons. Rates used separately allocated one-sided adjusted Clopper–Pearson upper bounds, and paired differences used conservative four-tail intervals, each with family alpha 0.05 and bound 10,000. Replicates were not pooled, and no comparison was paired across experiments. Six further fixed-versus-legal-parent comparisons were computed afterwards as exploratory saved-state analyses in a separate confidence family; its bound of 10,000 conservatively covers all 72 possible ordered pairs of the four frozen policies across the six cells.

In R0, the new fixed formulations reduced observed failure from 90.255–90.635% to 71.750–72.860%. In B, they reduced it from 93.175–93.415% to 64.125–64.380%. Their adjusted upper bounds were 73.144–74.236% in R0 and 65.616–65.868% in B. New delta risks were 72.050–72.745% in R0 and 64.560–64.835% in B. Their adjusted upper bounds were 73.439–74.123% in R0 and 66.046–66.319% in B. Both new policies improved on parent-pool delta in all six prespecified paired comparisons. The fixed-rule reductions were 17.775–18.825 percentage points in R0 and 28.795–29.290 points in B. None of the six new-delta-versus-new-fixed intervals excluded zero, so an additional benefit from this nutrient-informed rule was not established.

The legal-parent control retained every parent action; parent/new action counts were 6/14, 8/12 and 9/15 for R0 and 5/13 in each B replicate. Its selected actions and failure labels matched parent-pool delta on every test state, so legal-domain screening contributed no identifiable improvement in this experiment. The evidence therefore concerns the training-generated candidate extension as a whole, not the isolated contribution of a particular LP or local solver. The six exploratory fixed-versus-legal-parent comparisons also favoured the new fixed rule.

Despite these improvements, none of the 24 registered risk records met the 20% target. Adjusted all-action failure lower bounds for the newly frozen pools were 65.932–67.227% in R0 and 60.177–60.529% in B. Every replicate therefore excluded the registered 1%, 2%, 5%, 10% and 20% targets within its own finite pool, even under full-state selection or randomisation. These floors bound only the finite pools tested and leave other continuous formulations and candidate-generation procedures open. The frozen CP example target, the fixed-fMCP selected energy chain, the Table 5-1 planning conditions and the assumption that supplied DM is consumed remain unchanged, and the experiment does not certify complete NASEM adequacy, observed feed intake or farm outcomes.

Independent saved-label and statistical readback reconstructed all 480,000 assigned policy–state outcomes, the original nine-row verdicts and known/unknown partitions, counts, rate bounds, registered and exploratory paired intervals and the six finite-pool floor counts, without independently reimplementing the selector or all nutritional equations. All draws had already been saved, and the analysis involved no new sampling. A clean-environment full-cell canonical replay of this family forms part of the six-family 740,000-position audit described in S30.

The six aggregate files `R6_continuous_24_registered_risks_safe.csv`, `R6_continuous_30_registered_pairs_safe.csv`, `R6_continuous_6_exploratory_static_vs_legalold_safe.csv`, `R6_continuous_6_fixed_menu_floors_safe.csv`, `R6_continuous_6_domain_and_pool_counts_safe.csv` and `R6_continuous_case_policy_ranges_safe.csv` retain all completed records. Full-precision candidate recipes, printed-table requirement values, DM moments, physical states, prices, absolute costs and random-stream seeds remain restricted; the equation-derived requirement amounts and the one-decimal fixed rations are released in S38. The six exploratory comparisons are excluded from the totals of 498 risk records and 576 registered pairs.

Supplementary Methods: Observed-composition LP re-solve control (S37)

This exploratory, unregistered control evaluates least-cost LP policies that re-formulate the ration once the nutrient composition is observed, as an alternative to selecting from a finite candidate pool. It was added after the matched-pool and measurement protocols had been completed and reported. It uses development random streams rather than the reserved roots of the frozen protocols. It is excluded from the formal comparison sets and provides no prospective certificate.

Zero-margin policy. For every development state $s = (\theta_s, d_s)$ of cases A and C under TAB with independent cells (C0), the nominal least-cost LP (the nine-row main event, arm MAIN9)

was re-solved with θ_s in place of the nominal composition in every nutrient row. These rows were CP supply, absorbed Ca and P, the five Table 5-1 rows and the fixed-DMI energy row, whose per-ingredient coefficients were the fixed-DMI linearisation of the original problem, evaluated at θ_s at the planned DM supply. The DM fractions were held at their planned values $\hat{d}$, every structural row (planned-DM equality, inclusion caps, the starch-source premise row and the planned CP and EE limits) was unchanged, and the as-fed cost was minimised. For this zero-margin variant the LP was optimal in every state, so the fallback to M0 used for the margin variants below was never needed. The re-solved ration was scored on the realised state (θ_s, d_s) through the canonical reference event (failure = known violation or unresolved verdict), and M0, solved once at the nominal composition, was scored on the same states. Three development streams of 20,000 states were used per case (the same three roots for both cases; the C states differ because of the cottonseed cells). Rates carry one-sided Clopper–Pearson upper bounds with per-claim error 0.05/10,000; paired counts n_{10} (re-solve fails, M0 passes) and n_{01} are reported with the paired difference $(n_{10} - n_{01})/n$. The script (`experiments/E4_resolve_lp_control/resolve_lp_control.py`) records the git revision, the problem fingerprints and, for every run, that re-solving at the nominal composition reproduces the M0 ration used in the main analysis and that scoring one state reproduces the batch verdicts.

Supplementary Table S10: Observed-composition LP re-solve control under TAB/C0: failure of the re-solved ration and of the nominal ration M0 on the same 20,000 development states per cell, with one-sided upper confidence bounds (per-claim error 0.05/10,000), paired counts and the mean as-fed cost of the re-solved rations relative to M0. Only counts, rates, bounds and ratios are released.

Case	Stream	Fail k	Rate (%)	Upper (%)	M0 rate (%)	n_{10}	n_{01}	Diff. (pp)	Cost ratio
A	1	11,792	58.960	60.49	87.185	795	6,440	−28.2	1.0065
A	2	11,854	59.270	60.80	87.310	790	6,398	−28.0	1.0069
A	3	11,816	59.080	60.61	87.310	851	6,497	−28.2	1.0060
C	1	12,005	60.025	61.55	87.880	786	6,357	−27.9	1.0065
C	2	11,797	58.985	60.52	87.190	865	6,506	−28.2	1.0063
C	3	11,801	59.005	60.54	87.150	843	6,472	−28.1	1.0065

Outcome. The re-solved rations failed in 58.960–60.025% of states, against 87.150–87.880% for M0 on the same states: a paired reduction of 27.9–28.2 percentage points at a mean cost 0.60–0.69% above M0 (per-state cost ratios ranged from about 0.89 to 1.16). Almost every failure was a known violation (0–2 unresolved states per cell), and every re-solved ration stayed inside the plan-level Table 5-1 domain. Violations of T2 and T3 almost disappeared; violations of the absorbed Ca and P rows and of the reference-chain energy verdict fell by about 14–26%; and violations of T1, T4 and T5 became more frequent. The crude-protein and linear-energy rows were violated in about half of the states, about as often as under M0. The M0 rates on these development streams (87.150–87.880%) agree with the official v3 TAB rates of the frozen study (86.98–87.69% across the three reserved roots).

DM-margin re-solves. A further variant, added after the zero-margin result had been seen, re-solved the LP with a margin for uncertainty in hidden DM. The original study's safety-margin method (M1) was called on a copy of the problem whose nominal composition was the observed θ_s, through its explicit-spread interface with the composition spread set to zero and the DM spread equal to $\hat{d}$, which is the DM part of M1's registered relative setting. In every nutrient row and for every ingredient, the DM fraction became $\hat{d}_i(1 + \kappa\,\mathrm{sign}(s_k c_{ki}))$ (clipped to $[10^{-6}, 1]$), where $s_k = -1$ for $\geq$ rows and $+1$ for $\leq$ rows and c_{ki} is the row coefficient (centred on the bound for the concentration rows). Structural rows were unchanged and the ration was executed as $q = x/\hat{d}$. Margins $\kappa \in \{0.025, 0.05, 0.10, 0.15\}$ were run on the same three streams and states as the zero-margin control; in any state whose margin LP was infeasible, M0 was executed and the state was

counted as a fallback. At $\kappa = 0$ the M1 call reproduced the zero-margin re-solve exactly. The all-coefficient M1 margin, which also shifts the observed concentrations, was infeasible in most states at $\kappa = 0.05$ in a 500-state check (462 and 412 of 500 in A and C). At $\kappa = 0.025$ it was feasible in 491 and 498 of the 500 states and the policy failed in 9.2% (A) and 6.4% (C) of them, close to the DM-only 5% margin on the same states (9.0% and 6.6%). It was not run at full size because, after observation, DM is the only remaining uncertainty.

Supplementary Table S11: DM-margin re-solves under TAB/C0: ranges over the three development streams of 20,000 states (failure rate of the executed policy including fallback states, one-sided upper bound at per-claim error 0.05/10,000, paired difference against M0, number of states whose margin LP was infeasible, and mean cost of the executed rations relative to M0).

Case	κ	Rate (%)	Upper (%)	Diff. (pp)	Infeasible per stream	Cost ratio
A	0	58.960–59.270	60.49–60.80	−28.2 to −28.0	0 / 0 / 0	1.0060–1.0069
A	0.025	21.175–21.420	22.47–22.72	−66.1 to −65.8	0 / 0 / 0	1.0532–1.0541
A	0.05	9.565–9.830	10.51–10.79	−77.7 to −77.4	409 / 405 / 373	1.1190–1.1195
A	0.10	79.065–79.600	80.32–80.84	−8.1 to −7.7	17,921 / 17,961 / 18,037	1.0258–1.0273
A	0.15	= M0	88.21–88.33	0	20,000 / 20,000 / 20,000	1.0000
C	0	58.985–60.025	60.52–61.55	−28.2 to −27.9	0 / 0 / 0	1.0063–1.0065
C	0.025	21.405–21.930	22.71–23.24	−66.5 to −65.2	1 / 0 / 0	1.0533–1.0536
C	0.05	8.075–8.595	8.96–9.50	−79.8 to −78.6	76 / 103 / 112	1.1201–1.1206
C	0.10	72.250–73.445	73.64–74.81	−14.9 to −14.4	16,321 / 16,259 / 16,210	1.0509–1.0526
C	0.15	= M0	88.17–88.88	0	20,000 / 20,000 / 20,000	1.0000

The 5% DM margin gave the lowest failure in this grid: 9.565–9.830% in A and 8.075–8.595% in C, at 1.119–1.121 times the cost of M0. The 2.5% margin gave 21.175–21.420% in A and 21.405–21.930% in C; given the coarse grid, 5% is the best of the tested values only. Relative to M0, the 5% margin reduced the frequency of energy-row violations about five- to sevenfold and CP, Ca and P violations about 18- to 39-fold, while violations of T1, T4 and T5 became more frequent. M0-fallback states contributed 1.620–1.820 percentage points of the policy failure in A and 0.335–0.515 points in C, so most failures at 5% occurred in states where the margin LP was solved. At $\kappa = 0.10$ the margin LP was infeasible in 81–90% of states, and these fallback states accounted for 98.03–99.05% of all policy failures. The higher failure rates at large margins therefore reflect M0 execution following LP infeasibility; among the solved states (10–19% of states, a selected subset), failure was about 8% at a cost ratio of about 1.26–1.28, a descriptive figure and not a policy rate. At $\kappa = 0.15$ the margin LP was infeasible in every state, so the policy coincided with M0.

Interpretation and limits. A least-cost solution places its binding rows at zero nominal margin. With the DM fractions unobserved, the crude-protein and linear-energy rows of the zero-margin re-solve were violated in 49–51% of states, as under M0. The linear energy row is a formulation row; the failure event judges energy through the nonlinear reference chain, whose energy verdict was violated in 26–27% of states for the zero-margin re-solve and 35–36% for M0. Neither the zero-margin nor the DM-margin re-solve reached the failure rates of the pool-based nutrient-informed rules (0.5–2.2% on the full pools in the matched protocols, which tested C0, C1 and C2). Besides the pool itself, the LP differed in minimising cost under a uniform DM margin with a linear energy row and in executing M0 when infeasible, while the pool-based rules chose among candidates from several generators by a score over the nine rows of the failure event. The control does not identify which difference accounts for the risk gap. Similar mean cost ratios do not establish a cost-matched comparison. Nor does it separate hidden DM from linearisation (a counterfactual at $d = \hat{d}$ was not run). It uses independent C0 development streams rather than the protocols' saved states, so no paired contrast with the pool-based rules is formed, and it does not estimate outcomes in animals. The public outputs are `resolve_lp_control/results_public.csv` and

`RESULT_20261003.md` in the Supplementary Data; rations, absolute costs and per-state verdicts remain in the restricted archive.

Supplementary Data: released rations and derived requirement amounts (S38)

Two Supplementary Data files provide the frozen ration vectors and derived requirement amounts without redistributing source-table values.

`fixed_rations_released.csv` lists the as-fed ration of every fixed rule in the matched-pool experiment (Q_0–Q_4, cases A and C, three training replicates each) and the new fixed rules of the generated-ration experiment (cases R0 and B, three replicates each): 36 vectors, 16 of them distinct, in kg as-fed per head per day rounded to one decimal, with the as-fed total. Within each A replicate the fixed rations of Q_1–Q_4 coincide, and within each C replicate those of Q_2–Q_4 coincide, as the main text notes; canola meal is unused in every released R0, B and C ration. These vectors were frozen before testing; at full precision, they match their stored hashes.

`derived_requirements_released.csv` lists the requirement amounts and row bounds that the evaluator computes from the NASEM equations for the reference cow (net energy for lactation in Mcal/d and MJ/d, the linearised energy-row intercept and bound, absorbed calcium and phosphorus requirements and their realised-DM row bounds and coefficients) for the four cases. Values that are printed in NASEM tables rather than computed from equations (the illustrative CP target and the five Table 5-1 bounds) are not released. The planned DM supply is not listed separately. However, it is equation-derived and can be inferred from the calcium and phosphorus rows (requirement, realised-DM bound and coefficient), so these files effectively release it.

Identifiability was checked before release. A released ration vector contains no composition value, but every fixed ration satisfies the planned-DM equality exactly, so a set of rations together with the planned supply would constrain the ingredient DM fractions. Within any one case the released vectors have rank at most six against eight or nine ingredients; pooled across cases, the matrix has full algebraic rank at both full precision and one-decimal rounding, but only six of nine directions in the rounded matrix exceed the worst-case rounding perturbation. With the rounding tolerance applied, a linear-programming interval check over all 16 distinct vectors leaves every ingredient DM fraction free within an interval wider than 42 percentage points when the planned supply is treated as unknown, and wider than 24 points when it is taken as known; canola meal and the two mineral sources are unconstrained. Under this interval check, no source composition value is uniquely recoverable from the released files. The check and its script are described in `fixed_rations_identifiability.json`. A third file, `attribution_cost_ratios_to_M0.csv`, gives the mean feed cost of every matched-pool rule relative to the nominal least-cost ration M0 (and to the Q_0 fixed ration), computed from the frozen rations, the per-state assignments and the scenario prices; main Table 3 reports these ratios.

References

[1] Centre for Nutrition Modelling, University of Guelph. NASEM Model Python, version 1.0.2. https://github.com/CNM-University-of-Guelph/NASEM-Model-Python/releases/tag/v1.0.2 (n.d.). Pinned software release; accessed 2 October 2026.

Historical methods and results S1–S27

Version archive. This file preserves historical methods, assumptions, results and negative outcomes. Internal main-section/figure/table references identify R5, not the current article. Later evidence is in the current S29–S38 supplement. **R7 interpretation note.** Historical references to sampled EE as an energy fatty-acid proxy or to stochastic fatty acids do not describe the executed energy input path. The verified code used sampled CP/NDF/starch/ash, fixed TFA/Fd_FA and fixed lignin. EE retained a companion constraint and discrete support check. Current variable paths and the distinct conditional-DM interface are documented in S29; historical wording below is retained as a version record rather than silently changing the frozen model.

Supplementary Methods: Sources and inputs by pointer (archive S1)

Scope of this supplement. S1–S20 preserve the original reported run summaries and source pointers. S21 records the derived-input release that was unresolved at the time of this archive (fixed-rule rations and equation-derived requirement amounts have since been released, current S38); S22–S27 report the separately registered research extensions and their aggregate outputs. The accompanying CSVs support checks of exported counts, rates and intervals. This archive was written before the code was released. Its references to repository scripts or records describe the research project at that time; the Data and Code availability statements of the current article give the released version and the conditions of access to the restricted inputs. Restrictions described here record the authors' release decision at the time, not an independent legal determination about every derived number.

All evaluation states in this study are simulated from summary-based distributions. Neither the tabular row count, the number of random roots nor metadata distinguishing this case from a software demonstration denotes biological sample size.

S1.1 Sources used by the official run

Current source-date verification (2 October 2026). The original price inputs remain frozen. The conventional-cottonseed report has been checked against the official Volume 17 (August 2026) PDF. The live FEEDVAL page has subsequently changed; the historical 17 September date and the exact dated versions of the three other cited AMS reports were not independently retrieved for this archive (NOT_VERIFIED). FEEDVAL comparative values were used as scenario assumptions rather than actual procurement quotations. The originally supplied 2018 year for the Beuerlein/Lindsey source is likewise not independently verified. These provenance gaps do not change the retained numerical price scenario; they preclude claiming complete dated-source verification. The full R5 narrative below remains an explicitly historical archive.

Source	Used for	Evidence level and provenance cluster	Access and redistribution
NASEM (2021), *Nutrient Requirements of Dairy Cattle*, 8th rev. ed., Chapter 19, Table 19-1 (feed composition means, SDs, sample numbers)	means and SDs of the 48 random cells (six stochastic ingredients × DM, CP, NDF, starch, EE, ash, Ca, P) and the point values of the remaining components	published summary of client samples submitted to four US commercial laboratories (2011–2015) after statistical screening; laboratory cluster	read from the publisher edition available through the NCBI Bookshelf (DOI 10.17226/25806); copyrighted; pointers only

Source	Used for	Evidence level and provenance cluster	Access and redistribution
NASEM (2021), Table 19-3 (mineral sources: element contents and absorption coefficients)	ground limestone and dicalcium phosphate (point values); absorption coefficients of the mineral rows	published table; model cluster	as above
NASEM (2021), Chapters 3, 5, 7 and 21 (energy chain equations, carbohydrate guidelines of Table 5-1, mineral requirements, reference-cow inputs of Table 21-1)	reference evaluator rows and requirement side	published model; model cluster	as above; equations cited by number
`nasem_dairy` (open Python implementation of the NASEM 2021 model), fixed commit, feed library file	absorption-coefficient inputs, phosphorus fractions, forage/concentrate flag of the ingredients	derived from the same upstream data as Table 19-1	public repository; version pinned in the run record; values not redistributed here
St-Pierre and Weiss (2015), *J. Dairy Sci.* 98:5004–5015	numerators of the SD ratios that narrow the corn-silage cells in the SRC12, SRC14, MO12 and NARROW settings	separate study cluster (commercial dairy farms); overlap with the laboratory cluster not excluded	published article; the ratios are not printed (they would reveal the table SDs)
Yoder et al. (2014), *J. Dairy Sci.* 97:5645–5656	population-level correlation structure considered for a dependence scenario (not used in the official run)	laboratory cluster	published article
USDA AMS Market News reports AMS 3511 (18 Sep 2026), AMS 2850 (23 Sep 2026), AMS 2929 (week ending 18 Sep 2026)	dated quotes for four ingredients	public federal reports	public; quotes and dates in the price registry of the repository
University of Missouri Extension, FEEDVAL II page (historical citation reports 17 Sep 2026; archived date NOT_VERIFIED)	comparative values for the two mineral sources used as research-scenario price assumptions, not actual purchase quotes; upstream source and moisture basis were not stated	extension reference page	public page; two values used with attribution
Iowa State University Extension, Ag Decision Maker A1-65 (Edwards and Hart, 2018)	pricing rule for corn silage relative to the corn price and its moisture adjustment	extension fact sheet	public
Ohio State University CFAES fact sheet AGF-503 (Beuerlein and Lindsey, 2018)	bushel weight and moisture basis of corn for the price conversion	extension fact sheet	public

S1.2 Inputs by pointer

Input	Pointer	Status
Composition mean and SD of each random cell	NASEM (2021) Table 19-1, the row of the ingredient and the column of the component; feed identifiers as printed	restricted (copyright); rebuilt by the repository scripts from a licensed copy
Sample numbers per cell	same table	restricted; used only to document that no within-feed correlation can be derived
Mineral-source contents and absorption coefficients	NASEM (2021) Table 19-3	restricted
Reference-cow inputs (body weight, milk yield and composition, days in milk, parity, BCS)	NASEM (2021) Table 21-1, multiparous Holstein column; BCS 3.0 is a research assumption	reference-cow values are held in the restricted case directory; the planned DMI derived from them is treated as restricted
Carbohydrate guideline rows and the starch-source premise	NASEM (2021) Table 5-1 and its text; $\tau = 0.50$ is a research assumption (the source gives no number for "predominant")	thresholds restricted; τ public
Mineral requirement rows	NASEM (2021) Chapter 7 equations for the reference cow	restricted thresholds
Energy requirement	NASEM (2021) Chapter 3 equations (Supplementary S3)	restricted threshold value; equations public
Prices (USD per unit as fed, dated)	S1.1 price sources	specified in the project registry (not supplied in this archive); absolute ration costs are nevertheless withheld because, with public prices, they could be combined with other public outputs to recover restricted inputs
SD ratios of the four narrowed settings	St-Pierre and Weiss (2015) Tables 3 and 4 divided by the corresponding Table 19-1 SDs	restricted (one-step recovery of the table SDs); structure described in Methods 2.3

S1.3 Sources surveyed but not used

Forty further sources (national feed tables, laboratory libraries, on-line databases and datasets of several countries) were surveyed during the data-feasibility audit and not used, either because record-level data were not available for research use, because redistribution terms were unclear, or because they did not cover the reference case; the authors report a project registry of access and licence status; the full registry is not included in this submission archive.

Supplementary Methods: Constraint registry (reference problem v2) (archive S2)

All rows of the frozen registry `configs/dev_case_v1/reference_constraints_v2.csv`. Threshold values are not printed (they are NASEM (2021) values or quantities derived from them and are identified by pointer in the repository); the table gives what each row constrains, its unit and sense, its role, how it is judged and where it applies. Roles: *main reference* = one of the nine rows of the main event; *planning* = imposed on every strategy at decision time on the planned dry matter ($\hat{d}$) and never judged on states; *diagnostic* = scored for reporting, never part of the event or of any comparison; *arm-specific* = judged in states only for the historical training targets. Verdict models: *reference chain* = the Chapter 3 energy chain evaluated per state (Supplementary S3); *linear evaluator* = the row evaluated on the realised composition and DM of the state; *structural check* = satisfied at plan level by construction.

Row	What it constrains	Unit	Sense	Role	Verdict model	Threshold source	In reference problem
SH-DM-PLAN	Planned DM supply equals the reference cow's predicted intake (decision-time $\hat{d}$)	kg d^{-1}	=	planning	structural check	NASEM Eq. 2-1 (reference cow, Table 21-1)	yes
SH-INCL-LIMESTONE-MAX	Planned DM share of ground limestone	% DM	≤	planning	structural check	derived from the Ca maximum tolerable level	yes
SH-INCL-DCP-MAX	Planned DM share of dicalcium phosphate	% DM	≤	planning	structural check	derived from the Ca and P maximum tolerable levels	yes
SH-INCL-SOYHULLS-MAX	Planned DM share of soybean hulls	% DM	≤	planning	structural check	research assumption	yes
SH-NONNEG	Ingredient amounts non-negative	kg d^{-1}	≥	planning	variable bound	not applicable	no
SH-PLAN-T51-DGC-SHARE	Starch from dry ground corn as a share of planned diet starch (Table 5-1 premise, plan level, $\tau = 0.50$)	kg d^{-1}	≥	planning (v2; all arms and strategies)	structural check	research assumption (source gives no number)	yes
SH-PLAN-NEL-FIXEDDMI, -CP-HI, -EE-HI, -CA-ABS, -P-ABS	Plan-level forms of the corresponding rows at table values and $\hat{d}$	as the row	as the row	planning	structural check	as the row	no

Row	What it constrains	Unit	Sense	Role	Verdict model	Threshold source	In reference problem
PN-NEL-FIXEDDMI	Diet NEL supply ≥ requirement (linear fixed-intake row for formulation; main event judged by the reference chain)	Mcal d^{-1} (reported in MJ d^{-1})	≥	main reference	reference chain	NASEM Chapter 3	yes
PN-CP-SUP	Diet crude-protein supply ≥ case-specific target	kg d^{-1}	≥	main reference	linear evaluator	NASEM Table 21-3 illustrative protein intake	yes
PN-CA-ABS	Absorbed Ca supply ≥ requirement (maintenance faecal Ca on realised DM)	g d^{-1}	≥	main reference	linear evaluator	NASEM Chapter 7	yes
PN-P-ABS	Absorbed P supply ≥ requirement (maintenance faecal P on realised DM)	g d^{-1}	≥	main reference	linear evaluator	NASEM Chapter 7	yes
PN-T1	Forage NDF concentration ≥ minimum	% DM	≥	main reference	linear evaluator	NASEM Table 5-1	yes
PN-T2	Total NDF concentration ≥ minimum	% DM	≥	main reference	linear evaluator	NASEM Table 5-1	yes
PN-T3	Total NDF minimum as a continuous function of forage NDF	% DM	≥	main reference	linear evaluator	NASEM Table 5-1	yes
PN-T4	Diet starch concentration ≤ maximum	% DM	≤	main reference	linear evaluator	NASEM Table 5-1	yes
PN-T5	Starch maximum as a continuous function of forage NDF	% DM	≤	main reference	linear evaluator	NASEM Table 5-1	yes
PN-CP-HI	Diet CP concentration ≤ upper limit	% DM	≤	arm-specific (research assumption)	linear evaluator	research assumption	yes
PN-EE-HI	Diet ether-extract concentration ≤ upper limit	% DM	≤	arm-specific (research assumption)	linear evaluator	derived from a related published quantity	yes

Row	What it constrains	Unit	Sense	Role	Verdict model	Threshold source	In reference problem
DIAG-DM-REAL-LO / -HI	Realised DM supply within a band around the plan	kg d^{-1}	$\geq$ / $\leq$	diagnostic	linear evaluator	research assumption	yes
DIAG-CA-MTL, DIAG-P-MTL	Diet Ca and P concentrations $\leq$ maximum tolerable levels	% DM	$\leq$	diagnostic	linear evaluator	NASEM	yes
DIAG-T51-DGC-STARCH-SHARE	Starch from dry ground corn as a share of realised diet starch (per-state form of the premise)	kg d^{-1}	$\geq$	diagnostic	linear evaluator	research assumption	yes
DIAG-NEL-LINEARISATION-RESIDUAL	Difference between the linear NEL row and the Chapter 3 chain, per state	Mcal d^{-1}	report only	diagnostic	post-hoc report	not applicable	no

The main event comprises the nine rows marked *main reference* (PN-NEL-FIXEDDMI by the reference chain, PN-CP-SUP, PN-CA-ABS, PN-P-ABS, PN-T1 to PN-T5); the Table 5-1 rows are judged only when the ration meets the plan-level premise (Methods 2.2 and 2.6). The historical training targets FULL11 (eleven rows: the nine plus PN-CP-HI and PN-EE-HI) and PART6P5 (six rows) are sensitivity arms on the first reserved root.

Supplementary Methods: Energy reference chain and uncertainty model (archive S3)

S3.1 Two energy calculations

1. **Linear formulation row (`PN-NEL-FIXEDDMI`)**: NEL supply = $\Sigma_i\ x_i\ e_i + C_0$, where e_i is a per-ingredient NEL density obtained from the Chapter 3 chain at the planned intake with the feed-mean digestibilities (dNDF from Eq 3-3a at the mean NDF and lignin; concentration and DMI/BW adjustments of Eq 3-9 at the planned intake). It is the row that every strategy optimises against; it is not a state-wise conservative bound.
2. **Reference chain (`NASEM2021_ch3_chain_fixedFMCP_suppliedDMI_v1`)**: the whole-diet Chapter 3 chain of NASEM (2021) evaluated in every test state with the state's own DM supply as intake (intake assumed equal to supply), its own diet starch and DMI/BW, and the drawn NDF and fixed reference lignin of each feed. Main-event energy verdicts use this chain; disagreements with the linear row are reported as false passes and false fails of the linear row (Fig. 3b).

S3.2 Mapping to NASEM (2021) Chapter 3

Step	Equation (page)	Inputs drawn per state	Inputs fixed
Residual organic matter	Eq 3-1 (p. 22)	NDF, starch, FA, CP, ash of each feed	fat factor per feed class
Base NDF digestibility	Eq 3-3a	NDF, lignin	—
Starch and NDF digestibility adjustments	Eq 3-5a / 3-5b (diet starch, DMI/BW)	diet starch, DM supply	base starch digestibility per feed
Digestible energy of the diet	DE from the digestible fractions and the gross-energy coefficients (p. 25–27)	as above	true digestibility of ROM 0.96 (p. 25); energy coefficients 5.65 Mcal kg^{-1} CP and 4.0 Mcal kg^{-1} ROM (p. 27)
Metabolic fecal CP and endogenous fecal ROM	Eq 3-9 and the p. 27 conventions	DMI, diet NDF	undigested microbial CP 16.5 g kg^{-1} DMI (the Table 19-1 convention, p. 27); endogenous fecal ROM 34.3 g kg^{-1} DMI
Urinary energy	Eq 3-10a / 3-10b (p. 28)	digested CP, milk and gain CP	N per CP 1000/6.25
Gaseous energy, ME, NEL	Eq 3-11, 3-12 (p. 28)	as above	—
Requirement side	Eq 3-13 (maintenance), 3-14b (milk), 3-15a to 3-18 (gestation), body-weight change	none (reference cow, fixed)	activity 0 (confinement); BCS 3.0 (research assumption)

Not part of this model: prediction of microbial CP (Eq 3-6b), rumen models, intake prediction (intake equals supply), heat stress, and any animal response. Neither calculation is the complete NASEM model; the chain was not compared with the reference software within that frozen protocol (an open implementation was available on the compute host; the comparison was recorded at the time of this archive as an open item and was later carried out as the component comparison of current S29).

S3.3 Composition uncertainty model

- Random cells: DM, CP, NDF, starch, EE (fatty-acid proxy), ash, Ca and P of the six stochastic ingredients (48 cells); the two mineral ingredients are point values; lignin, RUP

and its digestibility, base digestibilities and other energy-chain inputs are held at table or fixed values.

- Distribution per cell: truncated normal on the physical support whose parent parameters are solved so that the truncated mean and SD equal the targets (`TN_MM`); where no such parent exists, a scaled beta with the same two moments (`BETA_MM`). Targets are never changed to obtain a fit; a cell that no declared family can match stops the run.
- Support and diagnostic checks record lignin/NDF and Eq 3-1 ROM anomalies without clipping. Undefined used calculations follow the canonical unknown flags; an ROM diagnostic anomaly alone does not override that verdict. The extension qualifications retain these original masks rather than adding a new failure definition (S24).
- Dependence in the original phase: cells are independent (assumption). No dependence scenario was included in that original frozen protocol; the outer uncertainty about the composition distribution is declared unidentified and is represented only by the five SD settings (assumption ranges, not confidence intervals).
- SD settings: TAB, SRC12, SRC14, MO12, NARROW as defined in Methods 2.3; the ratios by which cells are narrowed are held in the restricted registry because, together with the published numerators, they would reveal the copyrighted table SDs.

Analytical-component caveat. CP, NDF and ash are not disjoint chemical fractions. A sum above unity is not by itself proof of a physically impossible sample. Independent marginal sampling is nevertheless an assumption, and unusual joint combinations and undefined reference-chain states must be reported rather than silently clipped or renormalised.

Supplementary Methods: Transcription quality assurance of the NASEM (2021) tables (archive S4)

Scope: Table 19-1 (composition means, SDs and sample numbers of the ingredients used) and Table 19-3 (mineral-source contents and absorption coefficients). The values themselves are copyrighted and are not redistributed; this supplement reports the procedure and its counts only.

S4.1 Two transcription routes

Route	Method	Self-check
A	text layer of the PDF (`pdftotext -layout`; italics from `pdftohtml -xml`), then every target block rendered as an image and checked cell by cell	825 Table 19-1 target cells checked visually, 0 differences; 223 cells compared with an earlier extraction from the publisher's HTML version, 0 differences
B	page images rendered at 300 dpi and read cell by cell (no text-layer parsing), then checked against the PDF text-layer coordinates	253 rows checked, 0 differences; the checking tool was first smoke-tested with 5 deliberately altered cells, all 5 detected

The two routes were run by separately run software-assisted procedures that did not read each other's output; they are not independent human double entry (Declaration of generative AI use). A double entry can detect transcription errors but not printing errors in the source.

S4.2 Reconciliation

Item	Count
Route A rows	845 (25 Table 19-1 entries × 33 rows + 19 Table 19-3 rows + 1 premix row); 854 keys after splitting absorption coefficients into rows
Route B rows	283 (23 Table 19-1 entries × 11 rows + 29 Table 19-3 rows + 1 premix row)
Keys covered by both routes	278
Field comparisons on the common keys (mean, SD, n, unit, basis, PDF page, printed page; printed strings compared character by character, trailing zeros kept)	1,930; 0 disagreements
Keys only in route B	5 (elements not listed for a mineral source; adjudicated against page images; all kept as null)
Keys only in route A	576 (components not transcribed by route B); an independent second reading later checked all 576 keys (2,876 field comparisons, 0 disagreements); not part of the core table
Third reading by the adjudicator (not blind)	11 Table 19-1 entries × 11 components (363 mean/SD/n cells) and all Table 19-3 rows: 0 differences
Independent review (blind page-image reading)	54 rows / 162 cells, plus a text-layer check with a different parser of all 253 rows × 3 = 759 cells: 0 differences; the review's checker was mutation-tested with 5 altered cells, all detected

Item	Count
Core table regeneration (version 2)	identical bytes on three consecutive runs; text-layer check of the regenerated table: 759 cells, 0 differences

Adjudication rule: any disagreement or a row present in one route only is decided from the rendered page image; in this transcription there were no value disagreements, so adjudication concerned only the five one-route keys. Sample numbers printed without thousands separators are stored as integers with the printed string kept alongside (three cases, confirmed on page images). Processing variants that share printed statistics were verified against the text layer (three groups).

S4.3 Limits

Both routes read the same printed tables; neither can detect an error in the printed source. The values enter the study only as restricted inputs identified by table, page and row pointers.

Supplementary Methods: Strategy definitions, equivalences and solver settings (frozen protocol) (archive S5)

Notation. $q \geq 0$ is the vector of as-fed ingredient amounts (kg d^{-1}); $\mathcal{P}$ is the planning polytope (planned DM supply, inclusion limits, the starch-source premise row, non-negativity); $g_k(q; \theta) \leq 0$ is row k of the training target in linearised form, with θ the composition state (48 random cells of six stochastic ingredients; two mineral ingredients at point values); $\bar{\theta}$ is the table composition. All strategies use the same $\mathcal{P}$, the same prices and the same training target (the nine-row main event, energy in its linear fixed-intake form).

S5.1 Strategies

Strategy	Program	Tuning parameter and frozen grid	Guarantee type
M0 nominal least-cost	min $c^T q$ s.t. $q \in \mathcal{P}$, $g_k(q; \bar{\theta}) \leq 0$ $\forall$k	none (enters every α)	none
M1 safety margin	as M0 with every random coefficient of row k shifted against the row's limit by a fraction κ of its mean (relative scale; sign chosen so the row tightens); the DM margin applied in the primary entry (`apply_to_dm = true`) and switched off in a sensitivity line	$\kappa \in \{0, 0.025, 0.05, 0.075, 0.10\}$	heuristic
M2 joint chance constraint (SAA)	min $c^T q$ s.t. $q \in \mathcal{P}$, $g_k(q; \theta_s) \leq M_{sk}\ z_s$ $\forall$k, s; $\Sigma_s\ z_s \leq \lfloor \alpha_{train}\ N \rfloor$; $z_s \in \{0,1\}$; one binary per training state shared by all rows	$\alpha_{train} \in \{1, 0.75, 0.5, 0.25\} \times \alpha$; N $\in \{128, 512, 1024\}$ (nested prefixes of the 1024 optimisation draws; 1024 is the headline)	joint, on the training sample
M3a box-robust	max$\theta \in \mathcal{U}_\lambda\ g_k(q; \theta) \leq 0$ $\forall$k, $\mathcal{U}_\lambda$ the product of per-cell intervals mean $\pm\ \lambda$ SD clipped to the physical support (factory moments); exact for a product box	$\lambda \in \{1, 1.5, 2, 2.5, 3\}$ SD	set (all compositions in the box)
M3b budgeted robust	nominal row plus the Bertsimas–Sim protection term with budget Γ per row on the same box	λ as M3a; $\Gamma \in \{0, 1, \ldots, 6\}$, truncated per row to its number of random coefficients; selected over the whole (λ, Γ) grid	set, row-wise; no probability bound claimed

Selection (all tuned strategies): on the validation stream (20,000 states per world), the cheapest grid point whose one-sided 95% Clopper–Pearson upper bound on the training-event rate of violated-or-unknown states is $\leq \alpha$; if no grid point qualifies the entry is "target not met" and no ration is returned. Test states are refused by the selection code.

Interpretation of probabilities and minima. Reported one-sided Clopper–Pearson bounds are pointwise conditional Monte Carlo bounds, not simultaneous coverage across all searches and not uncertainty intervals for farms. The minimum-violation program applies only to its finite training states and linearised constraints. An optimal integer count is not a lower bound for the nonlinear population problem. Held-out diagnostic candidate rates are descriptive and do not

establish a global cost–reliability frontier. Source-file names and field names containing "frontier" or "min_attainable" are retained as identifiers, not adopted as stronger scientific claims.

S5.2 Big-M, polishing and the minimum-violation program

- Big-M (M2): $M_{\mathrm{sk}} = \min(M_{\mathrm{box}}, M_{\mathrm{q}})$, where M_{box} is the maximum of g_{sk} over the structural box of $\mathcal{P}$ (2I linear programs give the per-ingredient bounds) and M_{q} is a quantile-based strengthening over the state differences $(\mathrm{A_s} - \mathrm{A_t})\cdot x$ on the same box; its validity was checked by enumeration on small test instances (unit tests). The minimum-violation diagnostic uses the box value only (`min_violation_big_m = box`).
- Polishing: after the mixed-integer solve, a linear program on the retained states (those with $z_{\mathrm{s}} = 0$) re-optimises cost at the same violation set.
- Minimum-violation program: the smallest number m^* of violated training states over $\mathcal{P}$ (same big-M structure, relative gap 0), reported as proven or as lower/upper bounds at the time limit; the M2 ration at $\alpha_{\mathrm{train}} = m^*/\mathrm{N}$ is scored as a diagnostic frontier point and never enters a comparable set.

S5.3 Pre-declared diagnostics (not strategies)

- DIAG-T45: the nominal program with the two starch-maximum carbohydrate rows removed, solved over a margin grid covering [0, A1] (A1 = the largest nominal margin without those rows); tests whether energy and starch limits conflict.
- DIAG-E1: a chance constraint on the energy row alone (SAA, $\mathrm{N} \in \{128, 512, 1024\}$, $\alpha \in \{0.10, 0.05, 0.01\}$); tests whether energy reliability by itself is attainable.
- Both were run on the first reserved root in the training worlds TAB and NARROW (D-539) and every diagnostic ration was scored on the test streams of all five evaluation worlds.

S5.4 Equivalences (reported once, never counted as separate strategies)

Pair	Condition	What remains different
EQ-M1-M3a	M1 with an SD-scaled margin and DM margin, and M3a on the box mean $\pm$ k SD centred on the nominal table and $\hat{d}$: both are the row-wise worst case over the same box (identical LP)	in the official run M1 uses a relative (mean-fraction) margin with its own energy-column rule, so the pair is not identical there; both rows are reported
EQ-M0-MODES	M0 with mean-of-draws coefficients equals M0 at the nominal point when the draw means equal the nominal products	none under the condition (official run: nominal point)
EQ-M3c-M2-ALPHA0	M2 with $\lfloor \alpha_{\mathrm{train}}\ \mathrm{N} \rfloor = 0$ forces every $z_{\mathrm{s}} = 0$, i.e. the scenario-set robust LP	none under the condition

Also: M1 at $\kappa = 0$ and M3b at $\Gamma = 0$ reproduce M0; a large Γ reproduces M3a.

S5.5 Solver settings (HiGHS through SciPy)

Setting	Value
Primal feasibility tolerance	1×10^{-7}

Setting	Value
MILP relative gap (M2)	1×10^{-4} (minimum-violation diagnostic: 0)
Time limit, method solves and frontier	300 s per LP / MILP main solve (auxiliary LPs for boxes, big-M and polishing: no limit)
Time limit, minimum-violation program	60 s (N = 128), 180 s (N = 512), 360 s (N = 1024)
BLAS / OpenMP threads per job	1 (eight parallel job processes)
Outcome classes	optimal; feasible at the time limit; proven infeasible; no feasible solution found at the limit; numerical error; invalid input
Acceptance	a returned solution is accepted only after an explicit residual check against every row at the declared tolerance; entries without a ration have no cost (never zero) and stay in every denominator

Supplementary Methods: Companion events, endpoint rows and M0 magnitudes (archive S6)

Every public endpoint row of the strategy entries (ration_kind `method`, including entries without a ration) and of the diagnostic rations (`diagnostic_frontier` rows of endpoint_ablation.csv and the declared DIAG-T45 / DIAG-E1 rows of diagnostics_rows.csv), with all main-event, companion-event, energy-check, domain and status columns; the absolute cost, the MIP dual bound (a cost bound) and the ration-hash columns are dropped. The M0 table below reads the companion events of the nominal ration of each SD setting's MAIN9 cell in its own SD setting. S6_magnitudes_M0.csv gives the magnitudes of the M0 ration for the non-concentration rows only, in each row's unit (1 Mcal = 4.184 MJ); concentration rows are withheld (B-451). Run id: `official-20260926T162636Z-e3665bbe`; main event `main_reference_plan_domain`; roots k=0, k=1, k=2 (reported separately, never pooled).

The column inventories below describe the historical schema. The current aggregate CSV copies omit private execution, stream and fingerprint metadata in addition to absolute cost, cost-bound and ration-hash fields; scientific row order, risks, confidence limits and the previously disclosed non-concentration magnitudes are unchanged. Full supply–requirement–margin tables remain withheld.

Root 0

`S6_endpoint_rows_root0.csv`: 458 rows (313 endpoint rows, 145 diagnostic rows), 85 columns in the current aggregate copy; dropped: absolute cost, MIP dual bound (a cost bound), ration-hash columns.

M0 (MAIN9 cell, own SD setting), root 0:

event / check	TAB	SRC12	SRC14	MO12	NARROW
main event, plan-level domain (rate_upper) (`main_reference_plan_domain_rate_upper`)	0.9567	0.9576	0.9729	0.9529	0.968
main event, plan-level domain (one-sided CP upper) (`main_reference_plan_domain_cp_upper`)	0.96	0.9609	0.9755	0.9563	0.9708
main event, per-state domain (rate_upper) (`main_reference_rate_upper`)	0.9591	0.9582	0.973	0.9531	0.968
main event, per-state domain (CP upper) (`main_reference_cp_upper`)	0.9623	0.9614	0.9756	0.9565	0.9708
Table 5-1 verdicts used out of domain (`main_reference_t51_verdicts_used_out_of_domain_rate_upper`)	0.9567	0.9576	0.9729	0.9529	0.968
main + CP maximum row, per-state (`main_reference_plus_cp_hi_rate_upper`)	0.9597	0.9592	0.9739	0.9531	0.968
main + CP maximum row, plan-level (`main_reference_plus_cp_hi_plan_domain_rate_upper`)	0.9573	0.9586	0.9738	0.9529	0.968
main event with the linear energy row (`main_reference_linear_energy_rate_upper`)	0.962	0.9633	0.9757	0.9557	0.9703
H0 eleven rows (`h0_eleven_rate_upper`)	0.9601	0.9635	0.9763	0.9555	0.9703
H0 eleven, reference energy chain (`h0_eleven_reference_energy_rate_upper`)	0.9573	0.9586	0.9738	0.9529	0.968
H0 eleven, domain-conditioned (`h0_eleven_domain_conditioned_rate_upper`)	0.9625	0.9641	0.9764	0.9557	0.9703
S2 declared six rows (`s2_six_rate_upper`)	0.8244	0.798	0.8223	0.8116	0.8285
S2 six, domain-conditioned (`s2_six_domain_conditioned_rate_upper`)	0.8306	0.7992	0.8225	0.8118	0.8285
the five excluded rows (`excluded_five_rate_upper`)	0.848	0.8616	0.868	0.8458	0.8574
energy: linear row violated (`energy_linear_violation_rate`)	0.5017	0.498	0.4956	0.4989	0.4962

event / check	TAB	SRC12	SRC14	MO12	NARROW
energy: reference chain violated (upper) (`energy_reference_violation_rate_upper`)	0.4989	0.4982	0.495	0.4982	0.4969
energy: false pass (count) (`energy_false_pass`)	204	419	286	211	237
energy: false fail (count) (`energy_false_fail`)	234	419	294	218	230
Table 5-1 primary variant: not assessable share (`t51_primary_not_assessable_share`)	0.1514	0.1125	0.0004	0.019	0
analysis-anomaly states (count) (`n_analysis_anomaly_states`)	962	672	304	6	0
support-violation states (count) (`n_support_violation_states`)	3	3	3	0	0

Root 1

`S6_endpoint_rows_root1.csv`: 165 rows (165 endpoint rows, 0 diagnostic rows), 78 columns in the current aggregate copy; dropped: absolute cost, MIP dual bound (a cost bound), ration-hash columns.

M0 (MAIN9 cell, own SD setting), root 1:

event / check	TAB	SRC12	SRC14	MO12	NARROW
main event, plan-level domain (rate_upper) (`main_reference_plan_domain_rate_upper`)	0.9581	0.9617	0.9768	0.9537	0.9694
main event, plan-level domain (one-sided CP upper) (`main_reference_plan_domain_cp_upper`)	0.9613	0.9648	0.9792	0.9571	0.9722
main event, per-state domain (rate_upper) (`main_reference_rate_upper`)	0.9605	0.9621	0.9768	0.9537	0.9694
main event, per-state domain (CP upper) (`main_reference_cp_upper`)	0.9637	0.9652	0.9792	0.9571	0.9722
Table 5-1 verdicts used out of domain (`main_reference_t51_verdicts_used_out_of_domain_rate_upper`)	0.9581	0.9617	0.9768	0.9537	0.9694
main + CP maximum row, per-state (`main_reference_plus_cp_hi_rate_upper`)	0.9607	0.9629	0.9773	0.9537	0.9694
main + CP maximum row, plan-level (`main_reference_plus_cp_hi_plan_domain_rate_upper`)	0.9583	0.9625	0.9773	0.9537	0.9694
main event with the linear energy row (`main_reference_linear_energy_rate_upper`)	0.9632	0.9661	0.9787	0.956	0.971
H0 eleven rows (`h0_eleven_rate_upper`)	0.961	0.9663	0.9792	0.956	0.971
H0 eleven, reference energy chain (`h0_eleven_reference_energy_rate_upper`)	0.9583	0.9625	0.9773	0.9537	0.9694
H0 eleven, domain-conditioned (`h0_eleven_domain_conditioned_rate_upper`)	0.9634	0.9667	0.9792	0.956	0.971
S2 declared six rows (`s2_six_rate_upper`)	0.8175	0.793	0.8264	0.8047	0.8284
S2 six, domain-conditioned (`s2_six_domain_conditioned_rate_upper`)	0.8234	0.7941	0.8264	0.8047	0.8284
the five excluded rows (`excluded_five_rate_upper`)	0.8548	0.8704	0.8733	0.8515	0.8652
energy: linear row violated (`energy_linear_violation_rate`)	0.5087	0.5052	0.5044	0.5072	0.509
energy: reference chain violated (upper) (`energy_reference_violation_rate_upper`)	0.5063	0.5065	0.5042	0.5048	0.5051
energy: false pass (count) (`energy_false_pass`)	220	422	301	219	238
energy: false fail (count) (`energy_false_fail`)	246	412	307	243	277
Table 5-1 primary variant: not assessable share (`t51_primary_not_assessable_share`)	0.1529	0.1117	0.0005	0.0173	0
analysis-anomaly states (count) (`n_analysis_anomaly_states`)	1013	697	278	7	1
support-violation states (count) (`n_support_violation_states`)	5	5	5	0	0

Root 2

`S6_endpoint_rows_root2.csv`: 165 rows (165 endpoint rows, 0 diagnostic rows), 78 columns in the current aggregate copy; dropped: absolute cost, MIP dual bound (a cost bound), ration-hash columns.

M0 (MAIN9 cell, own SD setting), root 2:

event / check	TAB	SRC12	SRC14	MO12	NARROW
main event, plan-level domain (rate_upper) (`main_reference_plan_domain_rate_upper`)	0.9564	0.9603	0.9734	0.9524	0.9687
main event, plan-level domain (one-sided CP upper) (`main_reference_plan_domain_cp_upper`)	0.9597	0.9635	0.976	0.9559	0.9715
main event, per-state domain (rate_upper) (`main_reference_rate_upper`)	0.959	0.9609	0.9735	0.9525	0.9687
main event, per-state domain (CP upper) (`main_reference_cp_upper`)	0.9622	0.964	0.9761	0.956	0.9715
Table 5-1 verdicts used out of domain (`main_reference_t51_verdicts_used_out_of_domain_rate_upper`)	0.9564	0.9603	0.9734	0.9524	0.9687
main + CP maximum row, per-state (`main_reference_plus_cp_hi_rate_upper`)	0.9596	0.9621	0.9742	0.9525	0.9687
main + CP maximum row, plan-level (`main_reference_plus_cp_hi_plan_domain_rate_upper`)	0.957	0.9615	0.9741	0.9524	0.9687
main event with the linear energy row (`main_reference_linear_energy_rate_upper`)	0.9626	0.9655	0.9765	0.9558	0.9718
H0 eleven rows (`h0_eleven_rate_upper`)	0.9602	0.9659	0.9768	0.9557	0.9718
H0 eleven, reference energy chain (`h0_eleven_reference_energy_rate_upper`)	0.957	0.9615	0.9741	0.9524	0.9687
H0 eleven, domain-conditioned (`h0_eleven_domain_conditioned_rate_upper`)	0.9628	0.9665	0.9769	0.9558	0.9718
S2 declared six rows (`s2_six_rate_upper`)	0.8248	0.7987	0.8265	0.8136	0.8348
S2 six, domain-conditioned (`s2_six_domain_conditioned_rate_upper`)	0.8305	0.7998	0.8267	0.8137	0.8348
the five excluded rows (`excluded_five_rate_upper`)	0.8512	0.8656	0.8717	0.8458	0.8617
energy: linear row violated (`energy_linear_violation_rate`)	0.4996	0.5012	0.501	0.4982	0.5013
energy: reference chain violated (upper) (`energy_reference_violation_rate_upper`)	0.4966	0.4982	0.497	0.4961	0.4988
energy: false pass (count) (`energy_false_pass`)	209	392	273	219	214
energy: false fail (count) (`energy_false_fail`)	240	424	315	240	239
Table 5-1 primary variant: not assessable share (`t51_primary_not_assessable_share`)	0.156	0.116	0.0006	0.019	0
analysis-anomaly states (count) (`n_analysis_anomaly_states`)	961	698	315	11	0
support-violation states (count) (`n_support_violation_states`)	3	3	3	0	0

M0 magnitudes (S6_magnitudes_M0.csv)

Rows shown in their unit: DIAG-DM-REAL-HI, DIAG-DM-REAL-LO, PN-CA-ABS, PN-NEL-FIXEDDMI, PN-P-ABS (PN-NEL-FIXEDDMI once per verdict). Withheld (concentration rows): DIAG-CA-MTL, DIAG-P-MTL, DIAG-T51-DGC-STARCH-SHARE, PN-CP-HI, PN-CP-SUP, PN-EE-HI, PN-T1, PN-T2, PN-T3, PN-T4, PN-T5. Columns: n_states, n_violated, n_unknown, rate_lower, rate_upper, cp_upper, mean_deficit_given_violation, max_deficit, margin_q01, margin_q05, margin_q50; deficit = minus the margin in violated states (negative margin = violated), margin quantiles over the defined states (evaluator `_deficit_stats`).

Root 0, TAB:

row	verdict	unit	rate_upper	mean deficit	violated	max deficit	q01	q05
PN-NEL-FIXEDDMI	public_evaluator_linear	Mcal/d	0.5017	2.128	10.25	-6.259	-4.388	-0.01254
PN-CA-ABS	public_evaluator_linear	g/d	0.5118	3.497	17.04	-10.39	-7.241	-0.1122
PN-P-ABS	public_evaluator_linear	g/d	0.5022	3.489	15.23	-9.963	-7.1	-0.01897
DIAG-DM-REAL-LO	public_evaluator_linear	kg/d	0.0456	0.6631	3.529	-1.062	0.08109	2.686
DIAG-DM-REAL-HI	public_evaluator_linear	kg/d	0.0453	0.6489	3.593	-1.002	0.06468	2.667
PN-NEL-FIXEDDMI	reference_chain_ch3	Mcal/d	0.4989	2.02	10.1	-5.919	-4.144	0.009613

Magnitude and export labels. The energy rows labelled `public_evaluator_linear` and `reference_chain_ch3` are different calculations and must not be pooled. Main-text energy magnitudes use the latter and 1 Mcal = 4.184 MJ. The raw CSV labels the withheld CP-supply magnitudes as a "concentration row", although its unit is kg/d; that legacy reason label is inaccurate. The values remain withheld, and the original CSV is preserved unchanged.

Supplementary Methods: Implementation verification (test inventory of the frozen code) (archive S7)

The frozen code carries 1158 automated tests in 69 modules (pytest). On the delivery host all passed; on the compute host, after the restricted inputs built on the delivery host were replayed, 1,155 passed, 1 failed and 2 were skipped: the failure is the byte-identity check between an in-memory rebuild of the case files and the replayed bytes (last-digit floating-point differences of the host's BLAS; the run reads the replayed bytes, whose hashes are pinned in the freeze record), and the two skips need holder-side material absent on that host. These tests verify the implementation, not the biology. Module paths are relative to `tests/`.

Tests	Module	What the module checks (first line of its docstring)
24	integration/test_cross_method_ regression.py	Cross-method regression: M2 joint-chance SAA with alpha_train = 0 equals M3c scenario-set robust (B-133).
12	integration/test_end_to_end_ synthetic.py	Synthetic end-to-end integration test (synthetic_test_only; software check, not a research run).
14	integration/test_endpoint_ ablation_dryrun.py	2 x 2 (+ MAIN9) endpoint ablation driver (review round 3, instruction C; R3C; FIX3_BC): wiring, guards, –dry-run
3	integration/test_official_plan_ dryrun.py	`run_endpoint_ablation.py --official-plan` (official-run plan batch 3): the official cell / job matrix and compute
1	integration/test_official_ rehearsal.py	`run_endpoint_ablation.py --official-rehearsal` (official-run plan batch 4, §3 test 13): holder side only.
1	integration/test_toy_pipeline_ synthetic.py	End-to-end pipeline on the synthetic toy problem (synthetic_test_only; not a research run):
4	leakage/test_dm_hidden_value_ leakage.py	Contract section 8 required test 8 (+ T1, T2.1, C05/C06): hidden true DM must not be used before
32	numerical/test_chance_saa.py	M2 joint chance-constrained SAA and M2b marginal/Bonferroni (contract T4 M2, section 8 item 10).
25	numerical/test_energy_ linearization.py	Fixed-DMI NEL linearisation (NASEM 2021 ch.3) and the configured requirement layer – hand-computed checks.
13	numerical/test_energy_reference.py	Reference energy check (review round 3, F3 / F4; next-round instruction D) – hand-computed synthetic checks.
2	numerical/test_evaluator_analytic_ normal.py	Evaluator probability estimates vs a closed form (independent check of evaluator + sampling).

Tests	Module	What the module checks (first line of its docstring)
3	numerical/test_evaluator_consistency.py	Contract section 8 required test 7: the same known q gives the same public-evaluator output no
9	numerical/test_infeasible_and_status.py	Contract section 8 required test 4: deliberately infeasible problems return `proven_infeasible`
19	numerical/test_m0_small_lp.py	Contract section 8 required test 3: 2-3 ingredient LPs with analytic / enumerated optima.
2	numerical/test_monotone_tightening.py	Contract section 8 required test 6: tightening a constraint (others fixed) never lowers the
51	numerical/test_nutrition_requirements.py	NASEM 2021 DMI and factorial Ca/P requirement code vs the hand calculations of `reports/model_audit.md`.
4	numerical/test_permutation_invariance.py	Contract section 8 required test 9: reordering ingredients / nutrients / constraints / scenarios
4	numerical/test_supply_hand_calc.py	Contract section 8 required test 2: mass and nutrient supply against hand calculations.
4	numerical/test_zero_uncertainty.py	Contract section 8 required test 5: zero uncertainty degenerates to the nominal model; restoring
14	unit/test_build_dev_case.py	dev_case build logic moved out of the restricted directory (review round 3 F6, instruction F-2; R3F).
21	unit/test_config_validator_t9.py	Contract T9 validator: rejects undefined units, missing sources, unclear licences, pending values
17	unit/test_constraint_compile.py	Constraint grammar, unit/dimension checks and class rules (contract 6.2, T2).
9	unit/test_domain_applicability.py	Table 5-1 applicability domain (review round 3, F3 section 5.2; next-round instruction D5-D6).
7	unit/test_domain_premise_row.py	Reference problem v2 (D-533; official-run plan batch 1): the Table 5-1 starch-source premise as a planning row.
12	unit/test_domain_rule_cross_consistency.py	B-442: cross-consistency of the two Table 5-1 domain rules (reference problem v2, official-run plan batch 1).
3	unit/test_endpoint_ablation_dual_bound.py	FIX_PKG (round-3 red team D-3): the endpoint-ablation driver records the MILP dual bound.
9	unit/test_endpoint_reporting_official.py	Official-run plan batch 2, tests 9 and 11b: membership by p-bar, coverage per distinct ration, the per-method report
17	unit/test_evaluate_reference.py	Unified reference evaluation `evaluate_reference` and the reference constraint table (review round 3, C; R3C).
29	unit/test_fix_a_value_and_guard.py	Red-team lens A regressions (round 2, FIX_A). ALL NUMBERS ARE SYNTHETIC unless stated.

Tests	Module	What the module checks (first line of its docstring)
5	unit/test_fixb_dev_driver_streams.py	Review round 2, red team B #4 and #1 (FIX_B): stream namespaces of the dev_case_v1 driver and the declared
5	unit/test_fixb_energy_box_closure.py	Review round 2, red team B (FIX_B): M3 boxes with a signed energy column, the energy box of a factory
12	unit/test_garbling_counterexample.py	Third-review garbling counterexample (R3B; review §3.2), rebuilt independently. ALL NUMBERS ARE SYNTHETIC.
21	unit/test_information_analytic.py	Hand-calculated information values on a solvable toy problem (synthetic numbers only).
4	unit/test_information_economics.py	H x T batch conversion, break-even cost, net value and inventory coverage (synthetic numbers).
12	unit/test_information_library_frozen.py	Candidate library, risk table, frozen policies and approximation checks (synthetic continuous prior).
62	unit/test_information_policy.py	Policy-model properties on random small instances (synthetic arrays; no nutrition data).
15	unit/test_information_signal.py	Signal model, double-count guard, binning, likelihoods and prior-state leakage (synthetic values).
6	unit/test_information_strategies.py	Assay-selection strategies at equal budget on a constructed synthetic case.
11	unit/test_methods_config_matches_driver.py	Official-run plan batch 2, test 4 (BLOCKERS B-453): `configs/methods.yaml` states what the code executes.
11	unit/test_money_needs_a_complete_decision_problem.py	Round-3 red team B-1 (FIX3_BC): no money value without a complete decision problem. ALL NUMBERS ARE SYNTHETIC.
4	unit/test_official_screening_rule_spy.py	Official-run plan batch 2, test 8: the validation screening rule reaches every selection of the method block.
25	unit/test_official_v2_batch3.py	Official-run plan batch 3 (`docs/official_run_v2_plan_20260926.md` §1, §2 "B3", §3 tests 5 / 10 / 11a, §6 items 1-5):
32	unit/test_official_v2_batch4.py	Official-run plan batch 4 (`docs/official_run_v2_plan_20260926.md` §0 F3 / F5, §2 "B4", §3 tests 6 / 7 / 12, §7):
7	unit/test_package_manifests.py	Source manifest, packaged manifest and per-file transformations of a package (review round 3 F6, F-3; R3F).
87	unit/test_package_release.py	scripts/package_release.py: explicit manifests and release gates (FIX_C, second-round red team lens C).
17	unit/test_preflight.py	Preflight of the dev_case inputs (review round 3, instruction F-4; R3F).
16	unit/test_prior_truth_signal_guard.py	Prior / truth / signal guard (second review R2). All numbers are synthetic unless stated.

Tests	Module	What the module checks (first line of its docstring)
37	unit/test_redteam_engine_fixes.py	Regression tests for the red-team findings on the engine (2026-09-24).
15	unit/test_reference_constraints_v2.py	Reference constraint table v2 (`configs/dev_case_v1/reference_constraints_v2.csv`; D-533; batch 1).
11	unit/test_replay_build_outputs.py	scripts/replay_build_outputs.py: the scripted form of the 2026-09-25 Mac-bytes replay (round-3 red team D-10; FIX2).
3	unit/test_replay_dev_case_eval.py	CLOSE1: helpers and guard of the evaluation-only replay `experiments/E0_verification/replay_dev_case_eval.py`.
4	unit/test_report_generators.py	Tracked generators of the round-3 diagnostic reports (FIX3_DEF; review round 3 red team F6, E, D6/C).
50	unit/test_review20260926_numeric_guards.py	Synthetic regression cases for the 2026-09-26 repair; no feed data are read.
16	unit/test_review20260926_tools.py	Review utilities: synthetic examples and read-only stored-result checks.
43	unit/test_robust.py	P4c tests for the M3 robust methods and the mean/nominal equivalence audit.
24	unit/test_run_identity.py	Run identity, code content manifest and environment lock (second review R6, 2026-09-25).
4	unit/test_run_record_and_hashing.py	Run records carry every field of evidence_contract.yaml `run_record_required`.
37	unit/test_safety_margin.py	M1 calibrated safety margin (contract T4 M1, section 9 M1): margin object, direction, analytic
17	unit/test_smoke_loader_guards.py	Loader and builder guards (B-133).
7	unit/test_stats_intervals.py	Monte Carlo interval helpers (T8.2) checked against independent closed forms.
7	unit/test_streams_models.py	Random-stream isolation (T5/T6) and reference uncertainty models.
19	unit/test_uncertainty_correlation.py	Correlation structures: PSD assertion tests (B-133), repair logging, candidates and the copula.
38	unit/test_uncertainty_distributions.py	Moment-matched marginal families (B-137; configs/uncertainty.yaml b_distribution_family).
53	unit/test_uncertainty_factory.py	Uncertainty model factory (review R3). All numbers are synthetic test values.
21	unit/test_uncertainty_scaling.py	SD scaling, mean-shift scenarios and the H0/H1 information states (synthetic values).
11	unit/test_units_conversion.py	Contract section 8 required test 1: unit / proportion / price-basis conversions (both directions).
3	unit/test_value_callsite_scan.py	Call-site and wording scan of the information-value readings (round-3 red team B-2; FIX3_BC).

Tests	Module	What the module checks (first line of its docstring)
11	unit/test_value_definition_counterexample.py	Second-review counterexample for the information-value definitions (R1). ALL NUMBERS ARE SYNTHETIC.
10	unit/test_verify_restricted_inputs.py	scripts/verify_restricted_inputs.py: the FIX_C changes (second-round red team, lens C), on synthetic fixtures.

Numerical tolerances: LP/MILP acceptance uses the solver feasibility tolerance 1e-7 with an explicit residual check on every row; deterministic rebuild checks compare bytes; floating-point comparisons in unit tests use the tolerance stated in each test (typically 1e-9 to 1e-12 relative).

Supplementary Methods: Run records (archive S8)

Public identity, gating, environment, host and timing facts of the run, taken from OFFICIAL_RUN.json, run_record.json and RUN_JOBS.json. Hashes are shown as 12-character prefixes (full values in those files); reserved roots appear only by alias, never by value; the host name is not printed. Run id: `official-20260926T162636Z-e3665bbe`; main event `main_reference_plan_domain`; roots k=0, k=1, k=2 (reported separately, never pooled).

Identity and gating

field	value
run id	official-20260926T162636Z-e3665bbe
mode / output label	official / official
run type (run_record)	official
started / completed (UTC)	2026-09-26T16:26:36+00:00 / 2026-09-26T17:28:38+00:00
code commit	d7f3c97599d7…
code manifest sha256	e9fc3564a5d3…
code manifest complete	True
protocol sha256	c0263b54bc44…
freeze record sha256	a874d17f5dc2…
official config sha256	0a13d58f27c0…
specification digest at start	2cbf4d847463…
config hash	20deb5cb6a0e…
data manifest hash	b2dd10fc3c53…
run identity sha256	8d933aa65273…
gate at start	ok=True, status=protocol_frozen_anchored
import-time sources	consistent
environment lock	matches_lock (linux-x86_64-cpython3.12.3-20260925; host alias redacted)
is synthetic	False

Host and environment

field	value
CPU count	32
workers	8
machine / OS	x86_64 / Linux-6.17.0-1032-oem-x86_64-with-glibc2.39
Python	3.12.3
packages	numpy 2.4.6, pandas 2.2.3, scipy 1.17.1, yaml 6.0.2
solver	HiGHS 1.12.0 (git 4f96ee8) via scipy 1.17.1
solver settings	dual_feasibility_tolerance=1e-07, mip_rel_gap=0.0001, presolve=True, primal_feasibility_tolerance=1e-07, residual_check_rel_tol=1e-06, time_limit_s=300.0
thread environment	MKL_NUM_THREADS=1, OMP_NUM_THREADS=1, OPENBLAS_NUM_THREADS=1, PYTHONDONTWRITEBYTECODE=1

Sizes (draws per stream and world)

field	value
opt	1024

field	value
N_ladder	[128, 512, 1024]
N_headline	1024
N_nesting	N < 1024 uses the first N opt draws (prefix)
validation	20000
test	10000
adaptive_doubling	False

End-of-run identity check

check	value
unchanged	True
every job saw the start manifest	True
code manifest at start = at end	True
code manifest complete at end	True
specification digest at start = at end	True
freeze record at end	a874d17f5dc2…
freeze check at end (other fields)	freeze_record_sha256_at_start=a874d17f5dc2…; reasons=[]

Run-level flags

flag	value
pooled across roots	False
world fingerprints equal across jobs	True
SD registry check ok (merge)	True
SD registry check ok in every job	True
SD registry cells checked	48
SD registry file sha256	f808110afb7b…
SD range rule	sd_range_low = SD-S2 and sd_range_high = 1 in every cell; every point inside
merge wall time (s)	0.9867
run wall time (s)	3721
failed jobs	0

Root 0 (alias `formal-k0`)

field	value
jobs	10
cells	9
endpoint rows	313
candidate rows	2527
comparable sets	15
diagnostic rows / jobs	145 / 1
convergence verdicts	status_equal_no_selection: 27
training-event consistency	True

job	exit status	wall time (s)
root0___SD12MO_MAIN9	0	1214
root0___SDH0_FULL11	0	1253
root0___SDH0_MAIN9	0	1250

job	exit status	wall time (s)
root0___SDH0_PART6P5	0	899.2
root0___SDS2_FULL11	0	1312
root0___SDS2_MAIN9	0	1257
root0___SDS2_PART6P5	0	1149
root0___SDSRC12_MAIN9	0	1230
root0___SDSRC14_MAIN9	0	1340
root0___diagnostics	0	117.6

Root 1 (alias `formal-k1`)

field	value
jobs	5
cells	5
endpoint rows	165
candidate rows	1375
comparable sets	15
diagnostic rows / jobs	0 / 0
convergence verdicts	status_equal_no_selection: 15
training-event consistency	True

job	exit status	wall time (s)
root1___SD12MO_MAIN9	0	1209
root1___SDH0_MAIN9	0	1316
root1___SDS2_MAIN9	0	1223
root1___SDSRC12_MAIN9	0	1261
root1___SDSRC14_MAIN9	0	1416

Root 2 (alias `formal-k2`)

field	value
jobs	5
cells	5
endpoint rows	165
candidate rows	1375
comparable sets	15
diagnostic rows / jobs	0 / 0
convergence verdicts	status_equal_no_selection: 15
training-event consistency	True

job	exit status	wall time (s)
root2___SD12MO_MAIN9	0	1216
root2___SDH0_MAIN9	0	1218
root2___SDS2_MAIN9	0	1205
root2___SDSRC12_MAIN9	0	1248
root2___SDSRC14_MAIN9	0	1255

Supplementary Methods: Prior-access disclosure (archive S9)

What was seen before the protocol was frozen, and which decisions were taken after development results had been seen. Generated from the freeze record and the protocol file; the official run on the reserved roots was executed only after the freeze.

S9.1 Freeze record

- Frozen (UTC): 2026-09-26T16:17:00+00:00
- Parent commit of the freeze commit: `ee3607f50cff`; specification digest `2cbf4d847463ff08...`; protocol file SHA-256 `c0263b54bc44827b...`
- Declared matrix: SD settings ['SD-H0', 'SD-S2', 'SD-SRC14', 'SD-12MO', 'SD-SRC12']; cells ['SDH0_MAIN9', 'SDS2_MAIN9', 'SDSRC14_MAIN9', 'SD12MO_MAIN9', 'SDSRC12_MAIN9', 'SDH0_FULL11', 'SDS2_FULL11', 'SDH0_PART6P5', 'SDS2_PART6P5']; reserved root indices [0, 1, 2]; 20 jobs; diagnostics ['DIAG-T45', 'DIAG-E1'] on root index [0] in the training worlds ['SD-H0', 'SD-S2']
- Reserved roots drawn before the freeze: False (no official output directory exists at the freeze (checked by write_protocol_freeze); RandomStreams refuses a reserved root without a token of require_protocol_frozen (streams.py))

S9.2 Development material

- Development material: 3 named development root seeds and 1 declared seed range, all of their streams and sub-keys (seed values are listed in the repository's streams policy, never in the article; the three reserved official roots are distinct from all of them by construction)

Development runs whose results were inspected (development roots only; none is presented as held out):

- pilot-20260924T190634Z-8faf25eb
- pilot-20260924T192615Z-1fc52728
- pilot-20260924T205621Z-93d8654c
- pilot-20260925T055921Z-fb4f75af (endpoint ablation on the compute host, reference problem v1)
- smoke runs under results/smoke/ and results/pilot/smoke-* (pipeline checks)
- debug/tiny runs under data/restricted_local/debug/ (pipeline checks, toy size)

Run directories present at the freeze: results/pilot: 6; results/smoke: 12; compute-host run logs: 1

S9.3 Decisions taken after development results had been seen

Identifiers refer to the project decision log, which is held by the authors; the reproduction repository summarises the decisions that affect the reported analysis.

- D-453: the first development run was marked superseded (rounding defect in the training violation budget; a public residual table not fully redacted)
- D-466: status wording corrected: the specification and the development runs were executed, but the nutritional and statistical interpretation under the table-SD setting was not closed

- D-478: the second development run was marked superseded; the third became the current development run
- D-489: rules for calling a task item closed, partially closed or blocked, and their application to the development items
- D-499: the evaluation replay of the development case was recorded as a smoke run (evaluation only; no solving)
- D-513: the information-value quantity min(deterministic, randomised) was withdrawn as an executable basis and renamed a heuristic; monetary values only from a complete decision problem, and not in the main results
- D-514: the main text answers the cost-reliability questions only; information value is an exploratory, non-monetised appendix with the two policy classes reported separately; when no policy is feasible the value is undefined, not zero
- D-515: reference problem v1 was frozen for development: nine main-event rows (adjudication rules A1-A7); Table 5-1 primary domain variant tau = 0.50 (a research assumption); the 2 x 3 ablation with the MAIN9 arm; any change registers a v2
- D-516: the absence of a target-meeting ration under the table-SD setting was reclassified as a feasibility-boundary result; no SD, constraint or alpha was changed to obtain feasibility
- D-519: the full 2 x 3 development ablation was executed once on the compute host; the authorisation record was transcribed from the research lead's written instructions (disclosed deviation from the rule that the research lead writes the file)
- D-521: closing conditions of the development round: comparable sets were empty in both worlds and at all three alphas, so no cost comparison was made
- D-525: interpretation narrowed to the evidence: 'reliability cannot be bought' limited to the two development SD settings, N = 128 and the 26 rations of that run; no new solves or diagnostics to gather evidence
- D-528: eight nominal-precheck ration amounts moved from a tracked report to the restricted directory
- D-530: two residual gaps of the tenth packaging gate fixed (full-precision and non-UTF-8 scans)
- D-531: the earlier research question was set aside and the Methods were rewritten from the evidence
- D-532: story line and contribution order of the R5 article fixed (three questions; information value secondary)
- D-533: the decision package for reference problem v2 and the official protocol adopted (domain premise as a planning row; two pre-declared diagnostics; five declared SD settings; primary evaluation world TAB and primary arm MAIN9; lowest observed candidate risk when a set is empty; one-sided Clopper-Pearson screening and membership; coverage by entry and by distinct ration; DM uncertain; N ladder 128/512/1024 with a pre-declared convergence criterion; three reserved roots reported separately)
- D-535: the plan-level reading of the Table 5-1 domain adopted for the main event; the per-state reading kept as a companion event; tau = 0.50 unchanged, sensitivity variants reported

Protocol note: D-532, D-533, D-535 (story line, decision package, plan-level domain reading) and adjudication A2 were taken after the development results had been seen (disclosed in DECISIONS.md and docs/reference_problem_v2.md)

S9.4 What the official run did not change

- No alpha, threshold, tau, SD setting, grid, row or tolerance was changed after the freeze; the freeze commit is the run's commit and the work tree was clean before and after the run (run record).

- The NARROW setting was chosen during development because it made a target attainable; it is a feasibility-selected calibration point (Methods 2.3).
- Formal results are reported whatever their direction; the convergence criterion and the definitions of 'a margin suffices' and of the plotted frontier were fixed before testing (the archived protocol in S8; revised Methods 2.5–2.6).

Supplementary Methods: Exploratory value of feed information (archive S10)

Status of this specific deferred module in the original official run: **not run**. The frozen protocol registers the information-value step as a deviation ("RQ3 not run on the current code; information value exploratory/secondary only", D-532), so that original module supplies no result on those reserved roots. The later risk-only information policies and recourse are separately registered extensions (S22–S27), not a retrospective completion of this deferred module. This appendix records what the deferred analysis would compute, so that a reader can judge the deferred plan separately from the reported results. No monetary value of information, assay fee, willingness to pay or noisy-assay benefit is inferred from the new risk-only results.

S10.1 Decision model (as implemented in the research software, development status)

- Library: a finite set of candidate rations (the declared grid candidates of every strategy, distinct by ration).
- Prior: the declared composition distribution of the SD setting under study.
- Signals: knowledge of one ingredient–component cell (for example corn-silage DM or NDF), represented by a finite number of signal bins; measurement error, if declared, as a garbling of the perfect signal.
- Policies: a mapping from signal bin to a library ration. Two policy classes are evaluated separately: deterministic (one ration per bin) and randomised (a lottery over rations per bin, which contains the constant policy).
- Quantity: the minimum attainable ex-ante joint main-event risk without information and with each declared signal; reported as risks only. When no policy of a class meets the target the value is undefined, never zero.
- Not computed: expected value of perfect or sample information in monetary units, willingness to pay, testing priorities, break-even assay costs (they require assay cost, validity period, animals and days covered and a loss function, none of which was specified; Section 4.4).

S10.2 What would be reported if run

Per SD setting and α: the minimum attainable ex-ante risk without information; the same with each declared signal; the ingredient–component pairs, if any, whose signal lowers the attainable risk below the target when no conservative strategy does; the sensitivity to declared measurement-error scenarios.

S10.3 Why it was deferred

The compute budget of the official run was spent on the three questions of the article on three reserved roots; the information step multiplies the library scoring by the number of signals and bins. Running it later on the same frozen code and a further reserved root would be registered as a separate official run.

Supplementary Methods: Failure records per strategy entry (archive S11)

One record per strategy entry (and the M1 sensitivity line without the DM margin) of every cell: whether the entry returned a ration, the selection and solver status, the selected parameter, the statuses of the declared grid candidates and the scope of an infeasibility (full text; a grid without a solution is not a statement about the whole decision space). The CSV has one row per evaluation world and alpha (per_method_report.csv); the tables below show each entry once, in its own SD setting (grid counts are of that world), and M0 once for all alphas. Run id: `official-20260926T162636Z-e3665bbe`; main event `main_reference_plan_domain`; roots k=0, k=1, k=2 (reported separately, never pooled).

All three roots

The full table is provided as a CSV file in the supplementary data (S11_failure_records.csv).

Supplementary Methods: Sensitivity analyses (archive S12)

(a) the own-setting strategy results by SD setting and α for every root (Table S12-1; the main text's Table 5 is the cross-evaluation table; the tables under (b)–(e) are unnumbered panels of this section); (b) the Table 5-1 tau variants of the per-state main event for the M0 ration; (c) the historical training targets FULL11 and PART6P5 next to MAIN9 on root 0; (d) the rations of the MAIN9 cells of the other SD settings scored in TAB (the own-world value and the TAB value form an assumption range, not a confidence interval); (e) the M2 convergence verdicts. Costs appear only as premiums over M0 of the same cell and evaluation world. Run id: `official-20260926T162636Z-e3665bbe`; main event `main_reference_plan_domain`; roots k=0, k=1, k=2 (reported separately, never pooled).

(a) Table S12-1, every root

Table S12-1. Own-setting strategy results by SD setting and α (each ration scored in its own SD setting; event `main_reference_plan_domain`; membership: one-sided CP upper bound $\leq \alpha$). Cost premium relative to M0 of the same cell and evaluation world. Roots are reported separately, never pooled.

Root 0

SD	α	M0 rate [lo, up] ($\bar{p}$)	members / entries	members / distinct rations	meeting α	cheapest member (premium)	M1 (DM on)	M1 (DM off, sens.)	min attainable $\bar{r}$ ($\bar{p}$) if empty	M2 convergence
TAB	0.1	[0.957, 0.957] (0.960)	0/63	0/2	none	–	no ration	no ration	0.907 (0.911)	status_ equal_no_ selection
TAB	0.05	[0.957, 0.957] (0.960)	0/63	0/2	none	–	no ration	no ration	0.929 (0.933)	status_ equal_no_ selection
TAB	0.01	[0.957, 0.957] (0.960)	0/63	0/2	none	–	no ration	no ration	0.929 (0.933)	status_ equal_no_ selection
SRC12	0.1	[0.958, 0.958] (0.961)	0/63	0/2	none	–	no ration	no ration	0.892 (0.897)	status_ equal_no_ selection
SRC12	0.05	[0.958, 0.958] (0.961)	0/63	0/2	none	–	no ration	no ration	0.921 (0.926)	status_ equal_no_ selection
SRC12	0.01	[0.958, 0.958] (0.961)	0/63	0/2	none	–	no ration	no ration	0.921 (0.926)	status_ equal_no_ selection

SD	α	M0 rate [lo, up] ($\bar{p}$)	members / entries	members / distinct rations	meeting α	cheapest member (premium)	M1 (DM on)	M1 (DM off, sens.)	min attainable $\bar{r}$ ($\bar{p}$) if empty	M2 convergence
SRC14	0.1	[0.973, 0.973] (0.976)	0/63	0/2	none	–	no ration	no ration	0.861 (0.867)	status_ equal_no_ selection
SRC14	0.05	[0.973, 0.973] (0.976)	0/63	0/2	none	–	no ration	no ration	0.910 (0.915)	status_ equal_no_ selection
SRC14	0.01	[0.973, 0.973] (0.976)	0/63	0/2	none	–	no ration	no ration	0.910 (0.915)	status_ equal_no_ selection
MO12	0.1	[0.953, 0.953] (0.956)	0/63	0/2	none	–	no ration	no ration	0.844 (0.850)	status_ equal_no_ selection
MO12	0.05	[0.953, 0.953] (0.956)	0/63	0/2	none	–	no ration	no ration	0.897 (0.902)	status_ equal_no_ selection
MO12	0.01	[0.953, 0.953] (0.956)	0/63	0/2	none	–	no ration	no ration	0.897 (0.902)	status_ equal_no_ selection
NARROW	0.1	[0.968, 0.968] (0.971)	0/63	0/2	none	–	no ration	no ration	0.801 (0.808)	status_ equal_no_ selection
NARROW	0.05	[0.968, 0.968] (0.971)	0/63	0/2	none	–	no ration	no ration	0.877 (0.883)	status_ equal_no_ selection
NARROW	0.01	[0.968, 0.968] (0.971)	0/63	0/2	none	–	no ration	no ration	0.877 (0.883)	status_ equal_no_ selection

Root 1

SD	α	M0 rate [lo, up] ($\bar{p}$)	members / entries	members / distinct rations	meeting α	cheapest member (premium)	M1 (DM on)	M1 (DM off, sens.)	min attainable $\bar{r}$ ($\bar{p}$) if empty	M2 convergence
TAB	0.1	[0.958, 0.958] (0.961)	0/35	0/1	none	–	no ration	no ration	0.958 (0.961)	status_ equal_no_ selection

SD	α	M0 rate [lo, up] ($\bar{p}$)	members / entries	members / distinct rations	meeting α	cheapest member (premium)	M1 (DM on)	M1 (DM off, sens.)	min attainable $\bar{r}$ ($\bar{p}$) if empty	M2 convergence
TAB	0.05	[0.958, 0.958] (0.961)	0/35	0/1	none	–	no ration	no ration	0.958 (0.961)	status_ equal_no_ selection
TAB	0.01	[0.958, 0.958] (0.961)	0/35	0/1	none	–	no ration	no ration	0.958 (0.961)	status_ equal_no_ selection
SRC12	0.1	[0.962, 0.962] (0.965)	0/35	0/1	none	–	no ration	no ration	0.962 (0.965)	status_ equal_no_ selection
SRC12	0.05	[0.962, 0.962] (0.965)	0/35	0/1	none	–	no ration	no ration	0.962 (0.965)	status_ equal_no_ selection
SRC12	0.01	[0.962, 0.962] (0.965)	0/35	0/1	none	–	no ration	no ration	0.962 (0.965)	status_ equal_no_ selection
SRC14	0.1	[0.977, 0.977] (0.979)	0/35	0/1	none	–	no ration	no ration	0.977 (0.979)	status_ equal_no_ selection
SRC14	0.05	[0.977, 0.977] (0.979)	0/35	0/1	none	–	no ration	no ration	0.977 (0.979)	status_ equal_no_ selection
SRC14	0.01	[0.977, 0.977] (0.979)	0/35	0/1	none	–	no ration	no ration	0.977 (0.979)	status_ equal_no_ selection
MO12	0.1	[0.954, 0.954] (0.957)	0/35	0/1	none	–	no ration	no ration	0.954 (0.957)	status_ equal_no_ selection
MO12	0.05	[0.954, 0.954] (0.957)	0/35	0/1	none	–	no ration	no ration	0.954 (0.957)	status_ equal_no_ selection
MO12	0.01	[0.954, 0.954] (0.957)	0/35	0/1	none	–	no ration	no ration	0.954 (0.957)	status_ equal_no_ selection
NARROW	0.1	[0.969, 0.969] (0.972)	0/35	0/1	none	–	no ration	no ration	0.969 (0.972)	status_ equal_no_ selection

SD	α	M0 rate [lo, up] ($\bar{p}$)	members / entries	members / distinct rations	meeting α	cheapest member (premium)	M1 (DM on)	M1 (DM off, sens.)	min attainable $\bar{r}$ ($\bar{p}$) if empty	M2 convergence
NARROW	0.05	[0.969, 0.969] (0.972)	0/35	0/1	none	–	no ration	no ration	0.969 (0.972)	status_ equal_no_ selection
NARROW	0.01	[0.969, 0.969] (0.972)	0/35	0/1	none	–	no ration	no ration	0.969 (0.972)	status_ equal_no_ selection

Root 2

SD	α	M0 rate [lo, up] ($\bar{p}$)	members / entries	members / distinct rations	meeting α	cheapest member (premium)	M1 (DM on)	M1 (DM off, sens.)	min attainable $\bar{r}$ ($\bar{p}$) if empty	M2 convergence
TAB	0.1	[0.956, 0.956] (0.960)	0/35	0/1	none	–	no ration	no ration	0.956 (0.960)	status_ equal_no_ selection
TAB	0.05	[0.956, 0.956] (0.960)	0/35	0/1	none	–	no ration	no ration	0.956 (0.960)	status_ equal_no_ selection
TAB	0.01	[0.956, 0.956] (0.960)	0/35	0/1	none	–	no ration	no ration	0.956 (0.960)	status_ equal_no_ selection
SRC12	0.1	[0.960, 0.960] (0.963)	0/35	0/1	none	–	no ration	no ration	0.960 (0.963)	status_ equal_no_ selection
SRC12	0.05	[0.960, 0.960] (0.963)	0/35	0/1	none	–	no ration	no ration	0.960 (0.963)	status_ equal_no_ selection
SRC12	0.01	[0.960, 0.960] (0.963)	0/35	0/1	none	–	no ration	no ration	0.960 (0.963)	status_ equal_no_ selection
SRC14	0.1	[0.973, 0.973] (0.976)	0/35	0/1	none	–	no ration	no ration	0.973 (0.976)	status_ equal_no_ selection
SRC14	0.05	[0.973, 0.973] (0.976)	0/35	0/1	none	–	no ration	no ration	0.973 (0.976)	status_ equal_no_ selection

SD	α	M0 rate [lo, up] ($\bar{p}$)	members / entries	members / distinct rations	meeting α	cheapest member (premium)	M1 (DM on)	M1 (DM off, sens.)	min attainable $\bar{r}$ ($\bar{p}$) if empty	M2 convergence
SRC14	0.01	[0.973, 0.973] (0.976)	0/35	0/1	none	–	no ration	no ration	0.973 (0.976)	status_ equal_no_ selection
MO12	0.1	[0.952, 0.952] (0.956)	0/35	0/1	none	–	no ration	no ration	0.952 (0.956)	status_ equal_no_ selection
MO12	0.05	[0.952, 0.952] (0.956)	0/35	0/1	none	–	no ration	no ration	0.952 (0.956)	status_ equal_no_ selection
MO12	0.01	[0.952, 0.952] (0.956)	0/35	0/1	none	–	no ration	no ration	0.952 (0.956)	status_ equal_no_ selection
NARROW	0.1	[0.969, 0.969] (0.972)	0/35	0/1	none	–	no ration	no ration	0.969 (0.972)	status_ equal_no_ selection
NARROW	0.05	[0.969, 0.969] (0.972)	0/35	0/1	none	–	no ration	no ration	0.969 (0.972)	status_ equal_no_ selection
NARROW	0.01	[0.969, 0.969] (0.972)	0/35	0/1	none	–	no ration	no ration	0.969 (0.972)	status_ equal_no_ selection

Entries attempted count every declared method entry of every cell scored in that world (M1 without the DM margin is a sensitivity line, not an entry). Minimum attainable risk: lowest main-event rate_upper over the declared grids of the set's entries (per_method_report.csv grid_min_ rate_upper), its CP bound in brackets; descriptive, not a comparable cost.

(b) Table 5-1 tau variants, M0 (per-state reading, rate_upper)

root	SD	plan-level	per-state (primary)	primary variant	T51-DGC-0.50	T51-DGC-0.33	T51-DGC-0.67	T51-CORNKERNEL-0.50
0	TAB	0.9567	0.9591	T51-DGC-0.50	0.9591	0.9567	0.9989	0.9567
0	SRC12	0.9576	0.9582	T51-DGC-0.50	0.9582	0.9576	0.9982	0.9576
0	SRC14	0.9729	0.973	T51-DGC-0.50	0.973	0.9729	1	0.9729

root	SD	plan-level	per-state (primary)	primary variant	T51-DGC-0.50	T51-DGC-0.33	T51-DGC-0.67	T51-CORNKERNEL-0.50
0	MO12	0.9529	0.9531	T51-DGC-0.50	0.9531	0.9529	1	0.9529
0	NARROW	0.968	0.968	T51-DGC-0.50	0.968	0.968	1	0.968
1	TAB	0.9581	0.9605	T51-DGC-0.50	0.9605	0.9581	0.9991	0.9581
1	SRC12	0.9617	0.9621	T51-DGC-0.50	0.9621	0.9617	0.9982	0.9617
1	SRC14	0.9768	0.9768	T51-DGC-0.50	0.9768	0.9768	1	0.9768
1	MO12	0.9537	0.9537	T51-DGC-0.50	0.9537	0.9537	1	0.9537
1	NARROW	0.9694	0.9694	T51-DGC-0.50	0.9694	0.9694	1	0.9694
2	TAB	0.9564	0.959	T51-DGC-0.50	0.959	0.9564	0.9985	0.9564
2	SRC12	0.9603	0.9609	T51-DGC-0.50	0.9609	0.9603	0.9979	0.9603
2	SRC14	0.9734	0.9735	T51-DGC-0.50	0.9735	0.9734	1	0.9734
2	MO12	0.9524	0.9525	T51-DGC-0.50	0.9525	0.9524	1	0.9524
2	NARROW	0.9687	0.9687	T51-DGC-0.50	0.9687	0.9687	1	0.9687

(c) Historical training targets vs MAIN9, root 0 (own SD setting)

SD	training target	cell	α	M0 rate [lo, up]	M0 $\bar{p}$	labels with a ration	set members from this cell	min attainable $\bar{r}$ (cell)	its $\bar{p}$
TAB	MAIN9	SDH0_MAIN9	0.1	[0.957, 0.957]	0.96	1/22	0	0.9567	0.96
TAB	MAIN9	SDH0_MAIN9	0.05	[0.957, 0.957]	0.96	1/22	0	0.9567	0.96
TAB	MAIN9	SDH0_MAIN9	0.01	[0.957, 0.957]	0.96	1/22	0	0.9567	0.96
TAB	FULL11	SDH0_FULL11	0.1	[0.957, 0.957]	0.96	1/22	0	0.9567	0.96
TAB	FULL11	SDH0_FULL11	0.05	[0.957, 0.957]	0.96	1/22	0	0.9567	0.96
TAB	FULL11	SDH0_FULL11	0.01	[0.957, 0.957]	0.96	1/22	0	0.9567	0.96
TAB	PART6P5	SDH0_ PART6P5	0.1	[0.957, 0.957]	0.96	1/22	0	0.9567	0.96
TAB	PART6P5	SDH0_ PART6P5	0.05	[0.957, 0.957]	0.96	1/22	0	0.9567	0.96
TAB	PART6P5	SDH0_ PART6P5	0.01	[0.957, 0.957]	0.96	1/22	0	0.9567	0.96
NARROW	MAIN9	SDS2_MAIN9	0.1	[0.968, 0.968]	0.9708	1/22	0	0.968	0.9708
NARROW	MAIN9	SDS2_MAIN9	0.05	[0.968, 0.968]	0.9708	1/22	0	0.968	0.9708
NARROW	MAIN9	SDS2_MAIN9	0.01	[0.968, 0.968]	0.9708	1/22	0	0.968	0.9708
NARROW	FULL11	SDS2_FULL11	0.1	[0.968, 0.968]	0.9708	1/22	0	0.968	0.9708
NARROW	FULL11	SDS2_FULL11	0.05	[0.968, 0.968]	0.9708	1/22	0	0.968	0.9708
NARROW	FULL11	SDS2_FULL11	0.01	[0.968, 0.968]	0.9708	1/22	0	0.968	0.9708

SD	training target	cell	α	M0 rate [lo, up]	M0 $\bar{p}$	labels with a ration	set members from this cell	min attainable $\bar{r}$ (cell)	its $\bar{p}$
NARROW	PART6P5	SDS2_ PART6P5	0.1	[0.968, 0.968]	0.9708	1/22	0	0.8015	0.808
NARROW	PART6P5	SDS2_ PART6P5	0.05	[0.968, 0.968]	0.9708	1/22	0	0.8775	0.8829
NARROW	PART6P5	SDS2_ PART6P5	0.01	[0.968, 0.968]	0.9708	1/22	0	0.8775	0.8829

(d) Rations of the other SD settings scored in TAB

root	trained in	entry	α	own-world $\bar{r}$	own-world $\bar{p}$	TAB $\bar{r}$	TAB $\bar{p}$	TAB set	premium (members only)
0	SRC12	M0	all	0.9576	0.9609	0.9567	0.96	not a member	–
0	SRC12	(entries without a ration)	–	–	–	–	–	21	–
0	SRC14	M0	all	0.9729	0.9755	0.9567	0.96	not a member	–
0	SRC14	(entries without a ration)	–	–	–	–	–	21	–
0	MO12	M0	all	0.9529	0.9563	0.9567	0.96	not a member	–
0	MO12	(entries without a ration)	–	–	–	–	–	21	–
0	NARROW	M0	all	0.968	0.9708	0.9567	0.96	not a member	–
0	NARROW	(entries without a ration)	–	–	–	–	–	21	–
1	SRC12	M0	all	0.9617	0.9648	0.9581	0.9613	not a member	–
1	SRC12	(entries without a ration)	–	–	–	–	–	21	–
1	SRC14	M0	all	0.9768	0.9792	0.9581	0.9613	not a member	–
1	SRC14	(entries without a ration)	–	–	–	–	–	21	–
1	MO12	M0	all	0.9537	0.9571	0.9581	0.9613	not a member	–
1	MO12	(entries without a ration)	–	–	–	–	–	21	–
1	NARROW	M0	all	0.9694	0.9722	0.9581	0.9613	not a member	–
1	NARROW	(entries without a ration)	–	–	–	–	–	21	–
2	SRC12	M0	all	0.9603	0.9635	0.9564	0.9597	not a member	–

root	trained in	entry	α	own-world $\bar{r}$	own-world $\bar{p}$	TAB $\bar{r}$	TAB $\bar{p}$	TAB set	premium (members only)
2	SRC12	(entries without a ration)	–	–	–	–	–	21	–
2	SRC14	M0	all	0.9734	0.976	0.9564	0.9597	not a member	–
2	SRC14	(entries without a ration)	–	–	–	–	–	21	–
2	MO12	M0	all	0.9524	0.9559	0.9564	0.9597	not a member	–
2	MO12	(entries without a ration)	–	–	–	–	–	21	–
2	NARROW	M0	all	0.9687	0.9715	0.9564	0.9597	not a member	–
2	NARROW	(entries without a ration)	–	–	–	–	–	21	–

(e) M2 convergence (N = 512 vs 1024)

root	status_equal_no_selection
0	27
1	15
2	15

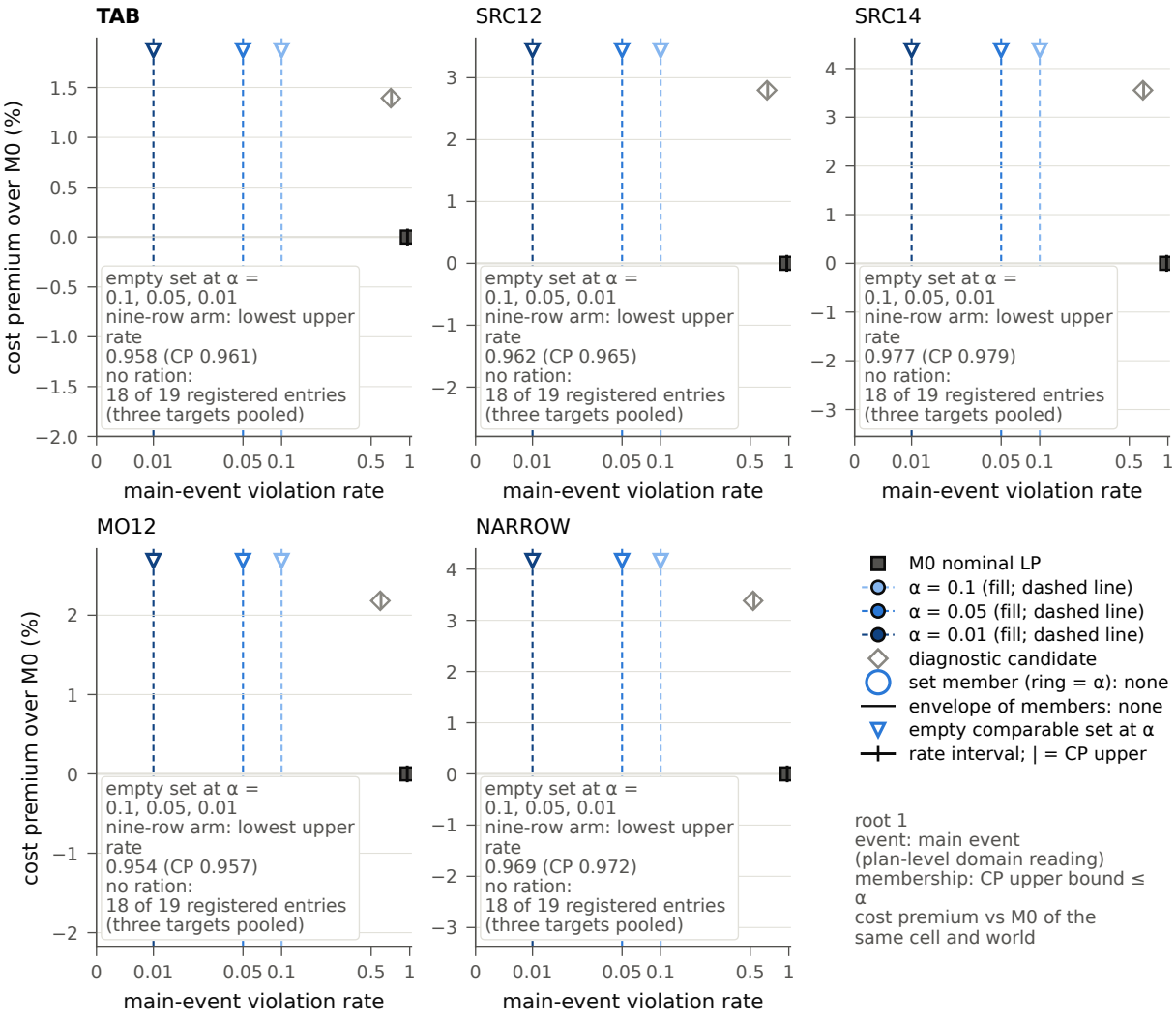
Reasons: no selected value at either N; the rate and cost criteria do not apply (57)

Supplementary Methods: Supplementary figures (archive S13)

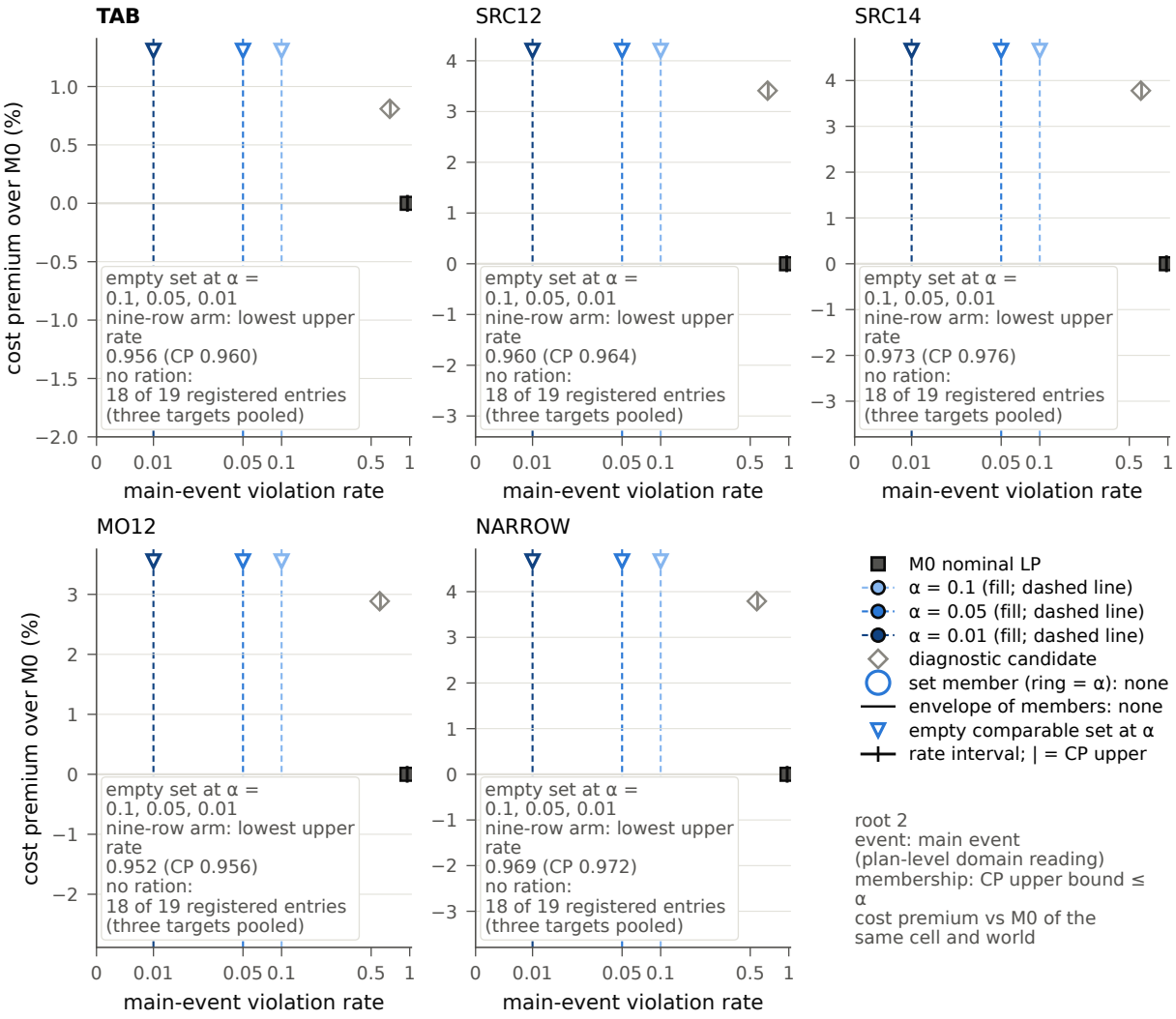
Figures drawn from the same official run by the figure scripts of the earlier version. At the time of this archive each figure existed as a PDF (vector), a PNG (300 dpi) and a CSV with the plotted numbers; those separate files are held in the study archive. The figures are embedded below as Fig. S1 to Fig. S14; the map from figure number to file stem is:

Figure	File stem
Fig. S1	`Fig2_cost_reliability_frontier_root1`
Fig. S2	`Fig2_cost_reliability_frontier_root2`
Fig. S3	`Fig3_constraint_failure_root1`
Fig. S4	`Fig3_constraint_failure_root2`
Fig. S5	`Fig4_strategy_map_root1`
Fig. S6	`Fig4_strategy_map_root2`
Fig. S7	`Fig5_attainable_risk_root1`
Fig. S8	`Fig5_attainable_risk_root2`
Fig. S9	`FigS2_cost_reliability_frontier_in_TAB_root0`
Fig. S10	`FigS2_cost_reliability_frontier_in_TAB_root1`
Fig. S11	`FigS2_cost_reliability_frontier_in_TAB_root2`
Fig. S12	`FigS3_strategy_map_in_TAB_root0`
Fig. S13	`FigS3_strategy_map_in_TAB_root1`
Fig. S14	`FigS3_strategy_map_in_TAB_root2`

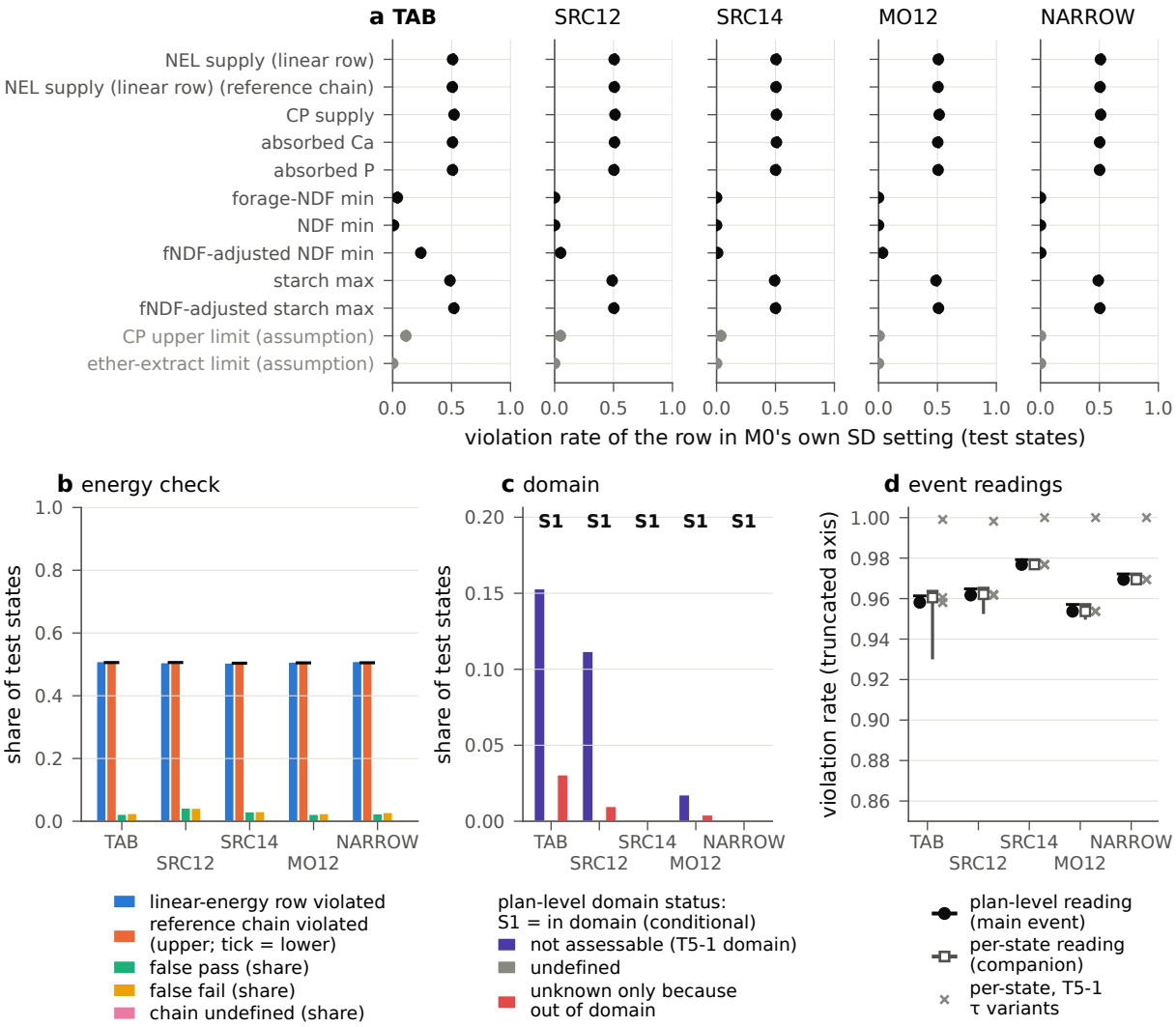
The supplementary figures are embedded below (Fig. S1 to Fig. S14); their separate files are held in the study archive.



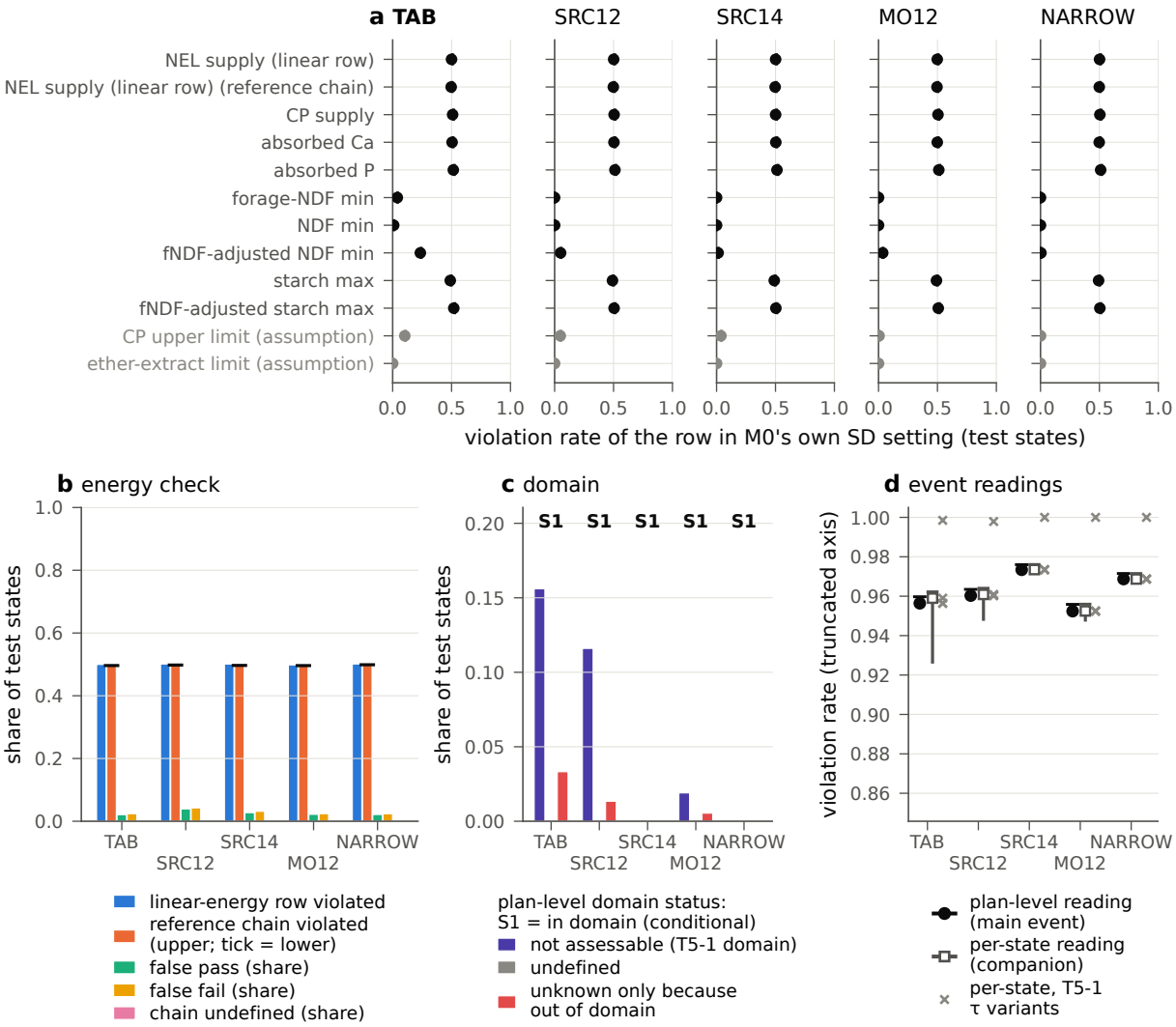
Supplementary Fig. S1. Own-setting cost and reference-risk outcomes (the R5 main Fig. 2 now uses common TAB evaluation) for the second reserved root (root 1): selected MAIN9 rations in their own SD setting, main-event rate interval with the one-sided CP bound, cost premium over M0; empty comparable sets marked with the lowest observed candidate risk. File: `Fig2_cost_reliability_frontier_root1`.



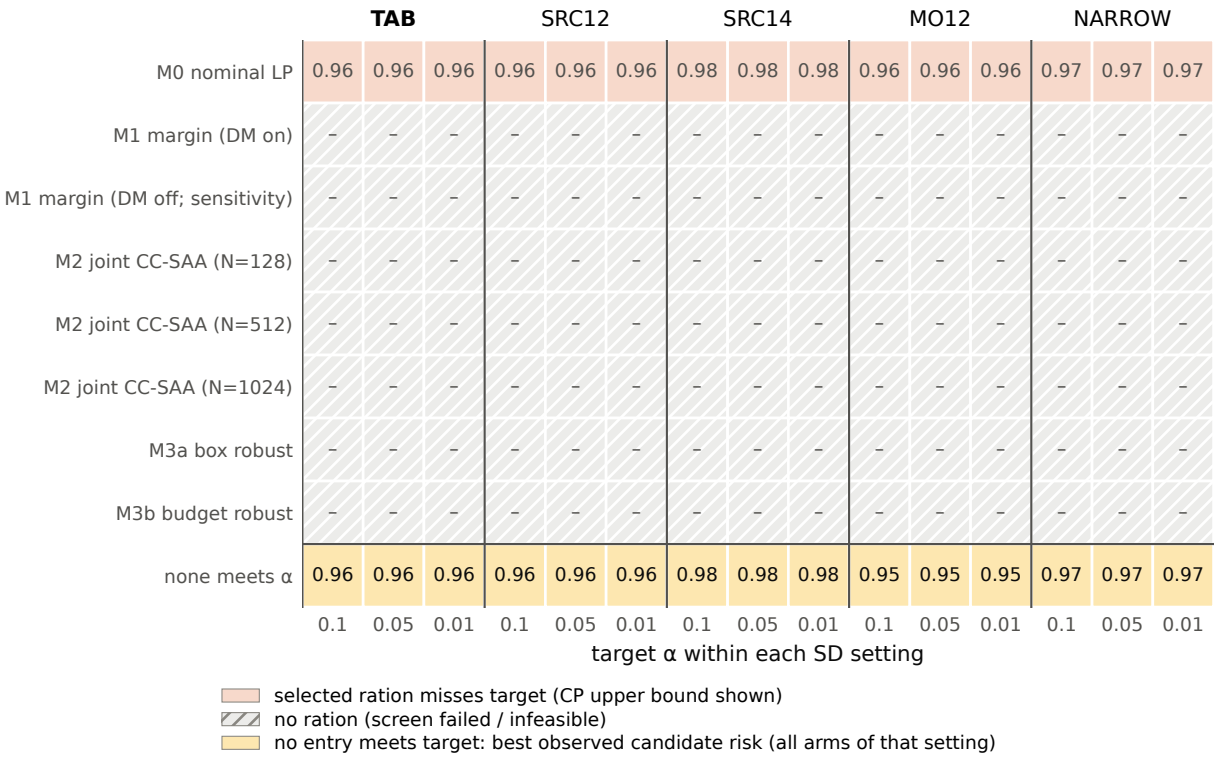
Supplementary Fig. S2. Own-setting cost and reference-risk outcomes (the R5 main Fig. 2 now uses common TAB evaluation) for the third reserved root (root 2): selected MAIN9 rations in their own SD setting, main-event rate interval with the one-sided CP bound, cost premium over M0; empty comparable sets marked with the lowest observed candidate risk. File: `Fig2_cost_reliability_frontier_root2`.



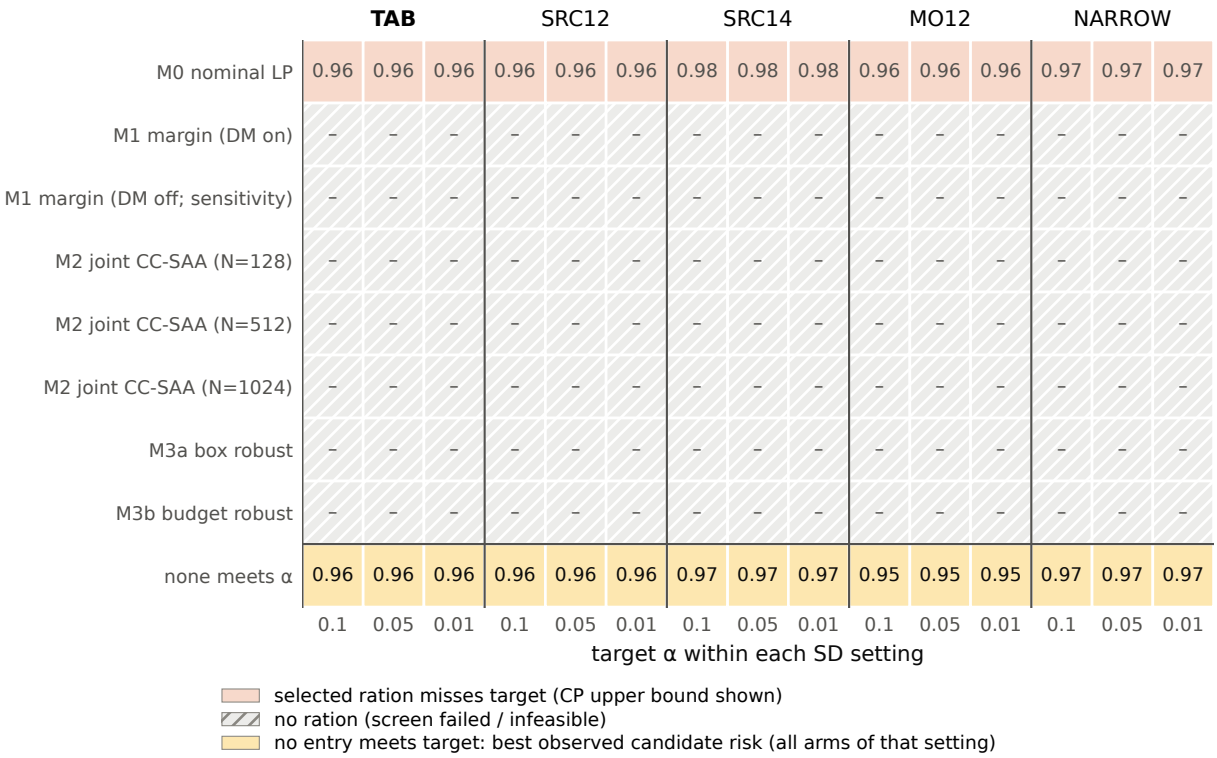
Supplementary Fig. S3. Fig. 3 of the main text for the second reserved root (root 1). File: `Fig3_constraint_failure_root1`.



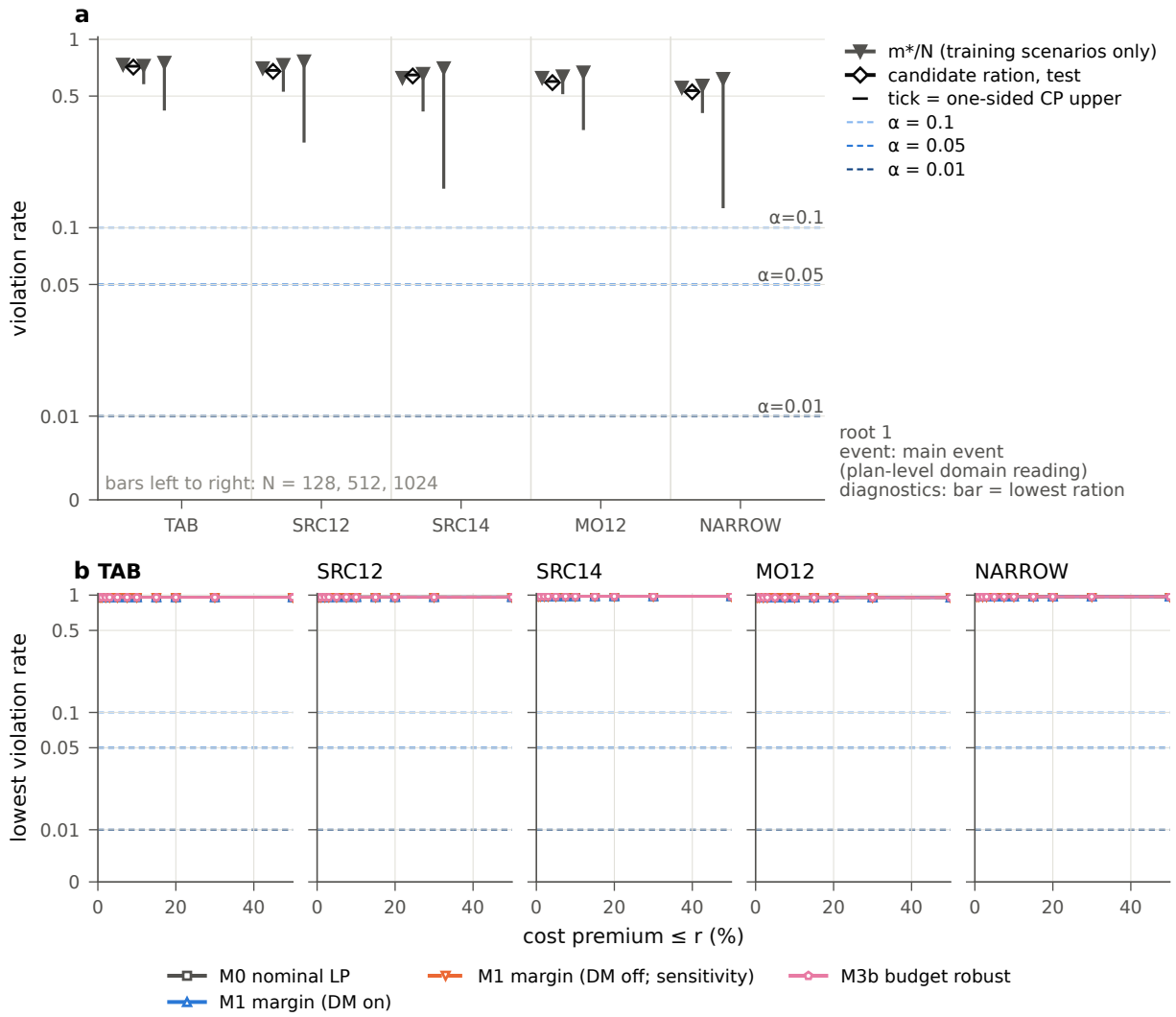
Supplementary Fig. S4. Fig. 3 of the main text for the third reserved root (root 2). File: `Fig3_constraint_failure_root2`.



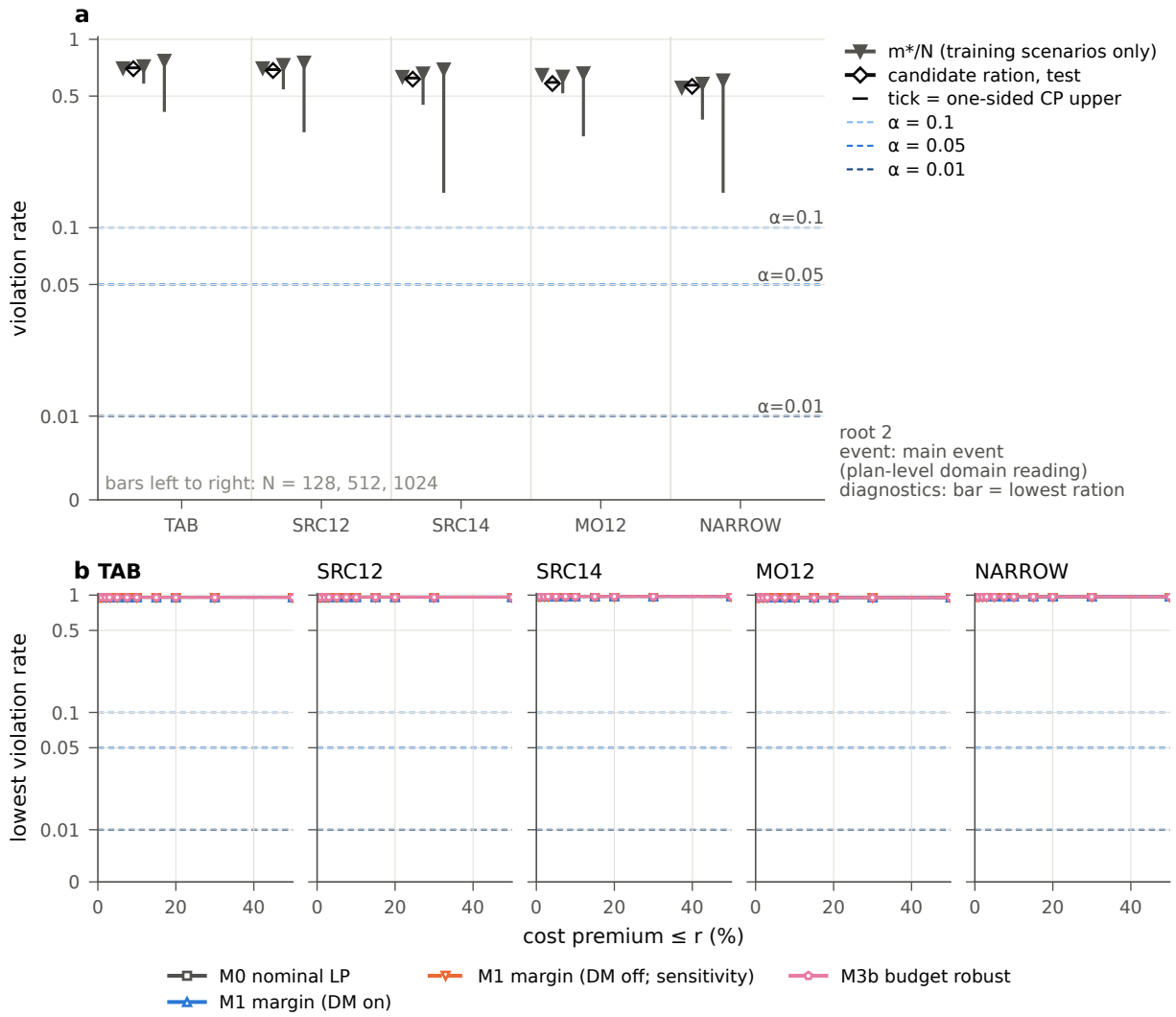
Supplementary Fig. S5. Own-setting target-attainment map (the R5 main Fig. 4 now uses common TAB evaluation) for the second reserved root (root 1): entry state per SD setting and target α (hatched: no ration; shaded: selected ration misses the target, one-sided 95% CP upper bound shown); last row: lowest observed candidate upper bound when no entry meets the target. Roots 1 and 2 carry the MAIN9 arms only, so this row equals the lowest MAIN9 value. File: `Fig4_strategy_map_root1`.



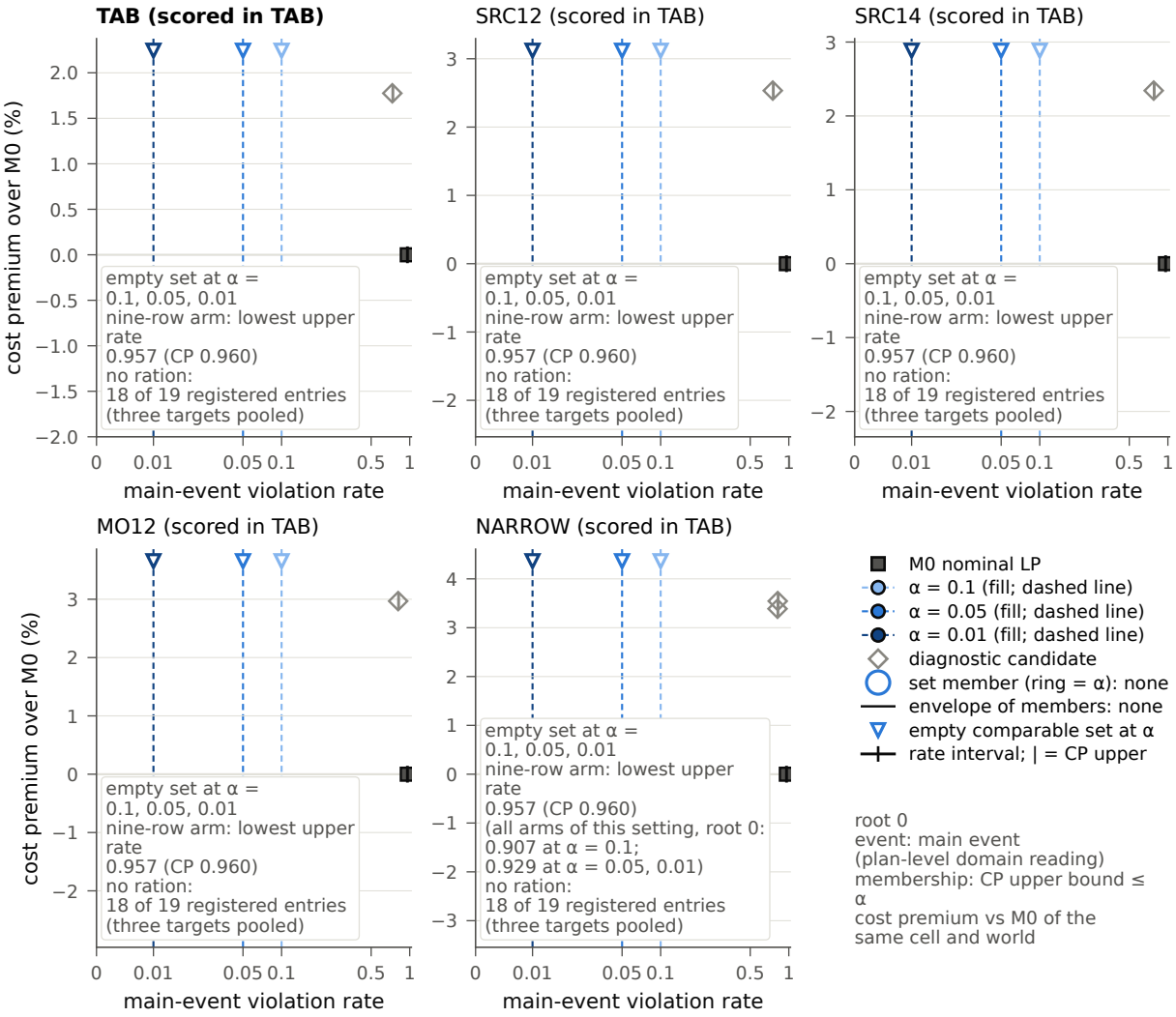
Supplementary Fig. S6. Own-setting target-attainment map (the R5 main Fig. 4 now uses common TAB evaluation) for the third reserved root (root 2): entry state per SD setting and target α (hatched: no ration; shaded: selected ration misses the target, one-sided 95% CP upper bound shown); last row: lowest observed candidate upper bound when no entry meets the target. Roots 1 and 2 carry the MAIN9 arms only, so this row equals the lowest MAIN9 value. File: `Fig4_strategy_map_root2`.



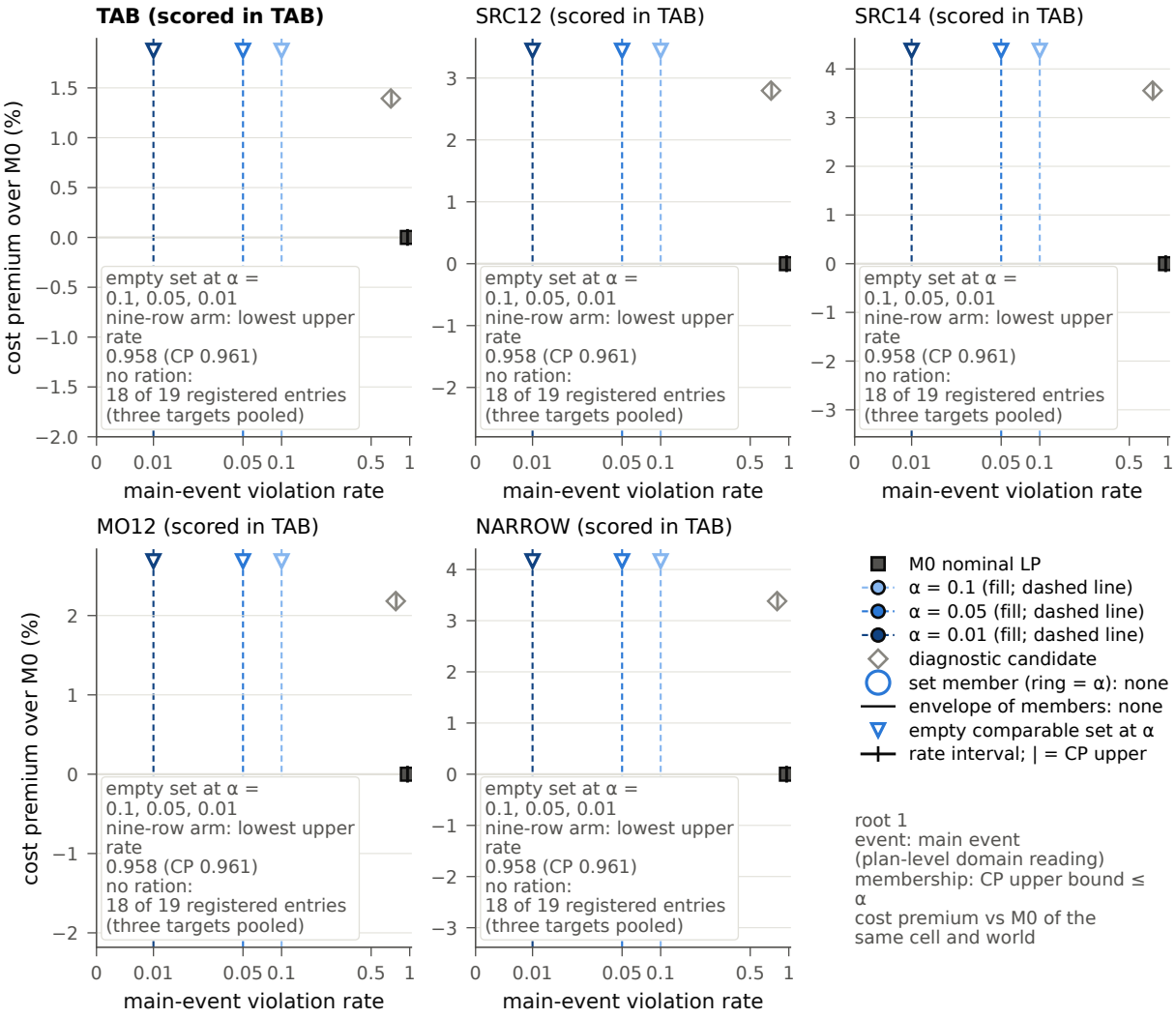
Supplementary Fig. S7. Fig. 5 of the main text for the second reserved root (root 1); the energy-only and starch-relaxation diagnostics were run on the first root only and are not shown; this panel shows the training minima and the held-out rates of the diagnostic (minimum-violation) rations. File: `Fig5_attainable_risk_root1`.



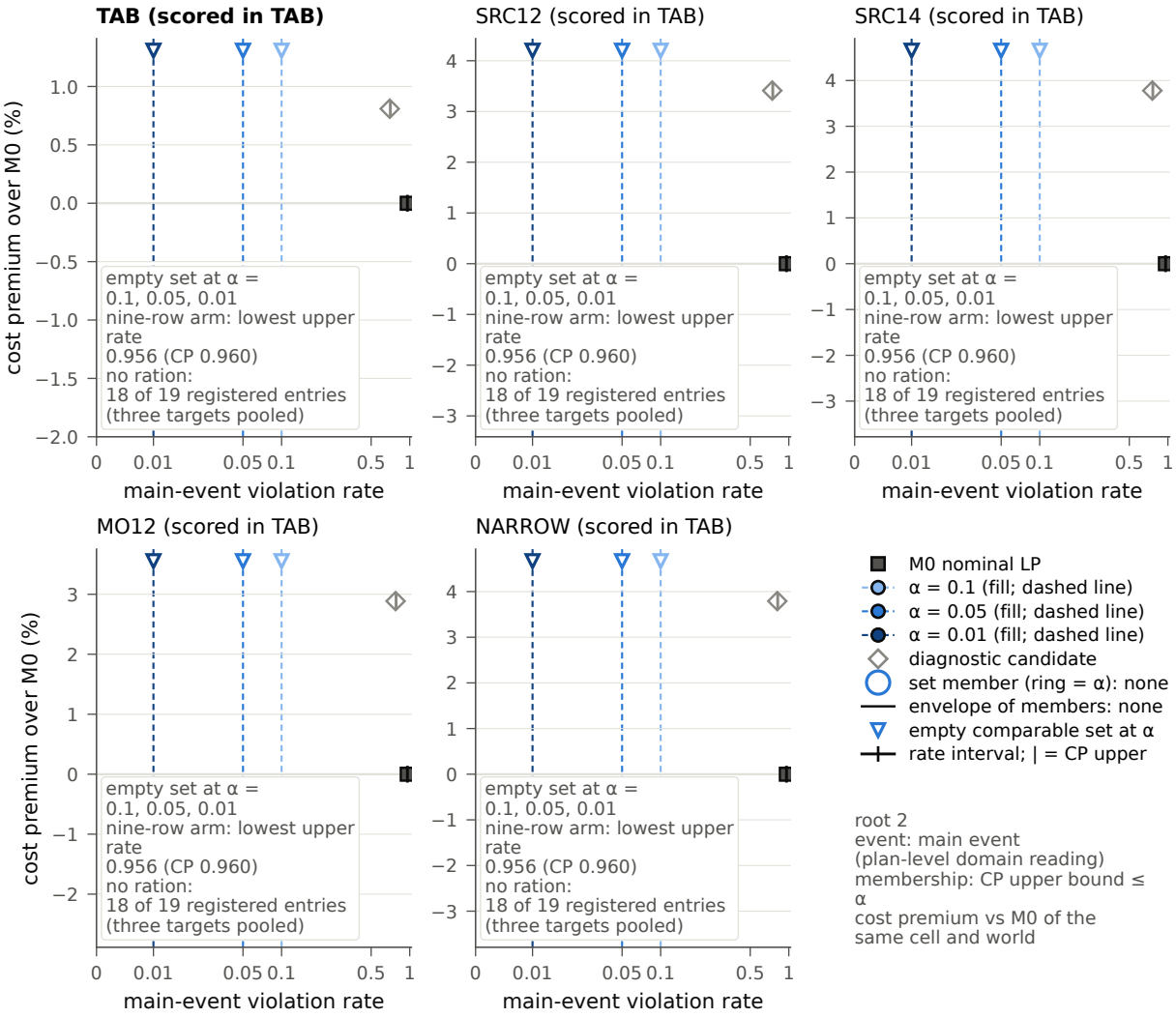
Supplementary Fig. S8. Fig. 5 of the main text for the third reserved root (root 2); the energy-only and starch-relaxation diagnostics were run on the first root only and are not shown; this panel shows the training minima and the held-out rates of the diagnostic (minimum-violation) rations. File: `Fig5_attainable_risk_root2`.



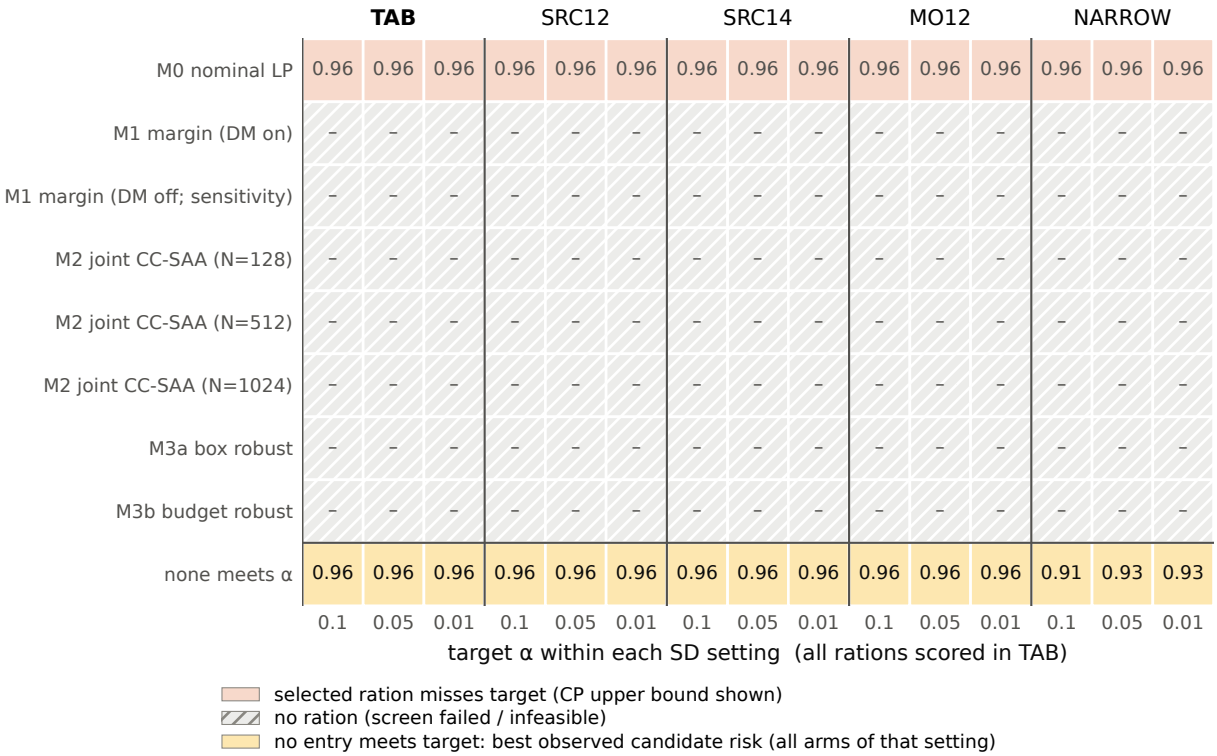
Supplementary Fig. S9. Main Fig. 2, root 0: all candidates scored in TAB. Rate brackets denote unresolved classification; upper confidence-bound markers are separate. File: `FigS2_cost_reliability_frontier_in_TAB_root0`.



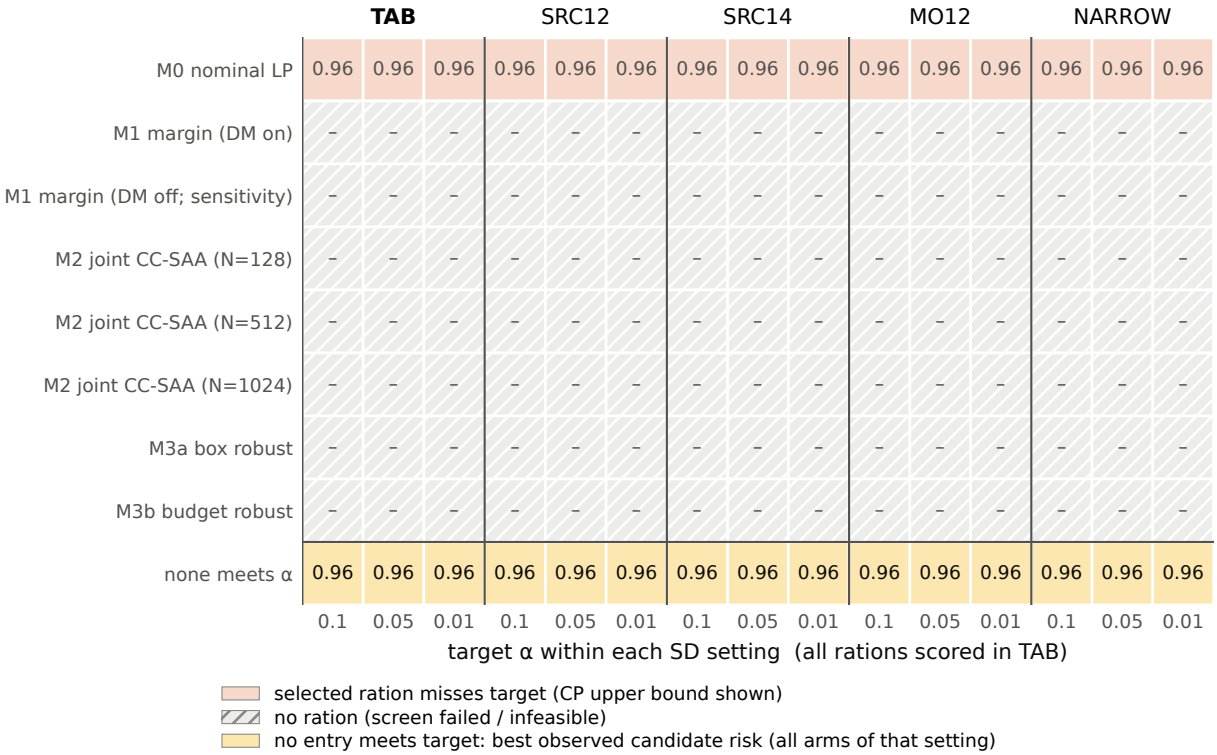
Supplementary Fig. S10. Companion of R5 main Fig. 2, root 1: all candidates scored in TAB. Rate brackets denote unresolved classification; upper confidence-bound markers are separate. File: `FigS2_cost_reliability_frontier_in_TAB_root1`.



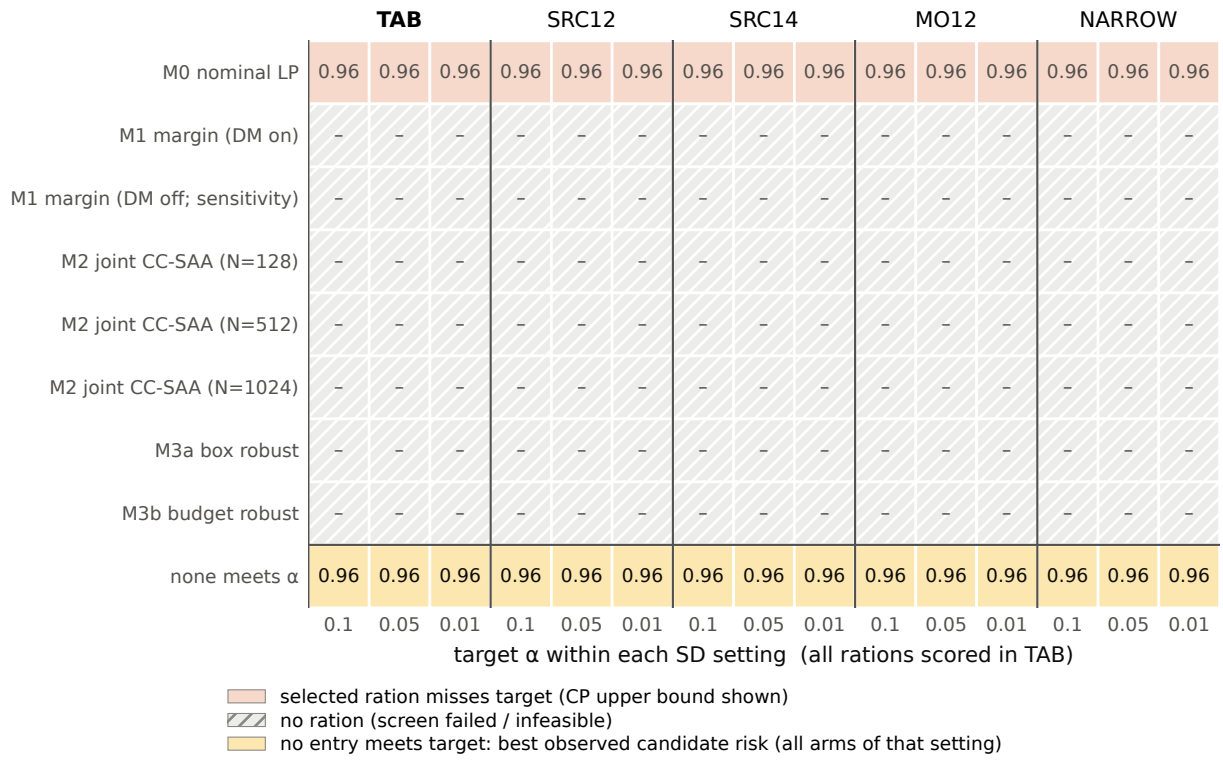
Supplementary Fig. S11. Companion of R5 main Fig. 2, root 2: all candidates scored in TAB. Rate brackets denote unresolved classification; upper confidence-bound markers are separate. File: `FigS2_cost_reliability_frontier_in_TAB_root2`.



Supplementary Fig. S12. Main Fig. 4, root 0: the same entries scored in TAB. File: `FigS3_strategy_map_in_TAB_root0`.



Supplementary Fig. S13. Companion of Fig. 4, root 1: the same entries scored in TAB. File: `FigS3_strategy_map_in_TAB_root1`.



Supplementary Fig. S14. Companion of Fig. 4, root 2: the same entries scored in TAB. File: `FigS3_strategy_map_in_TAB_root2`.

Supplementary Methods: Paired feasibility (archive S14)

For every root, evaluation world and alpha: per strategy (M0, M1 with the DM margin, M2 at each N, M3a, M3b) the number of cells whose entry returned a ration and the solve-success rate over the declared grid (candidates with a ration / declared candidates, summed over cells); per pair the feasible counts of both sides, the common feasible set (cells where both returned a ration) and the difference in main-event rate_upper (a minus b) over that set, 'not comparable' when it is empty. Cells: every cell scored in that world (root 0 includes the FULL11 / PART6P5 cells). Full tables: S14_paired_feasibility_strategies.csv and S14_paired_feasibility.csv. Run id: `official-20260926T162636Z-e3665bbe`; main event `main_reference_plan_domain`; roots k=0, k=1, k=2 (reported separately, never pooled).

All three roots

Feasible cells / cells (grid solve-success rate):

The full table is provided as a CSV file in the supplementary data (S14_paired_feasibility.csv).

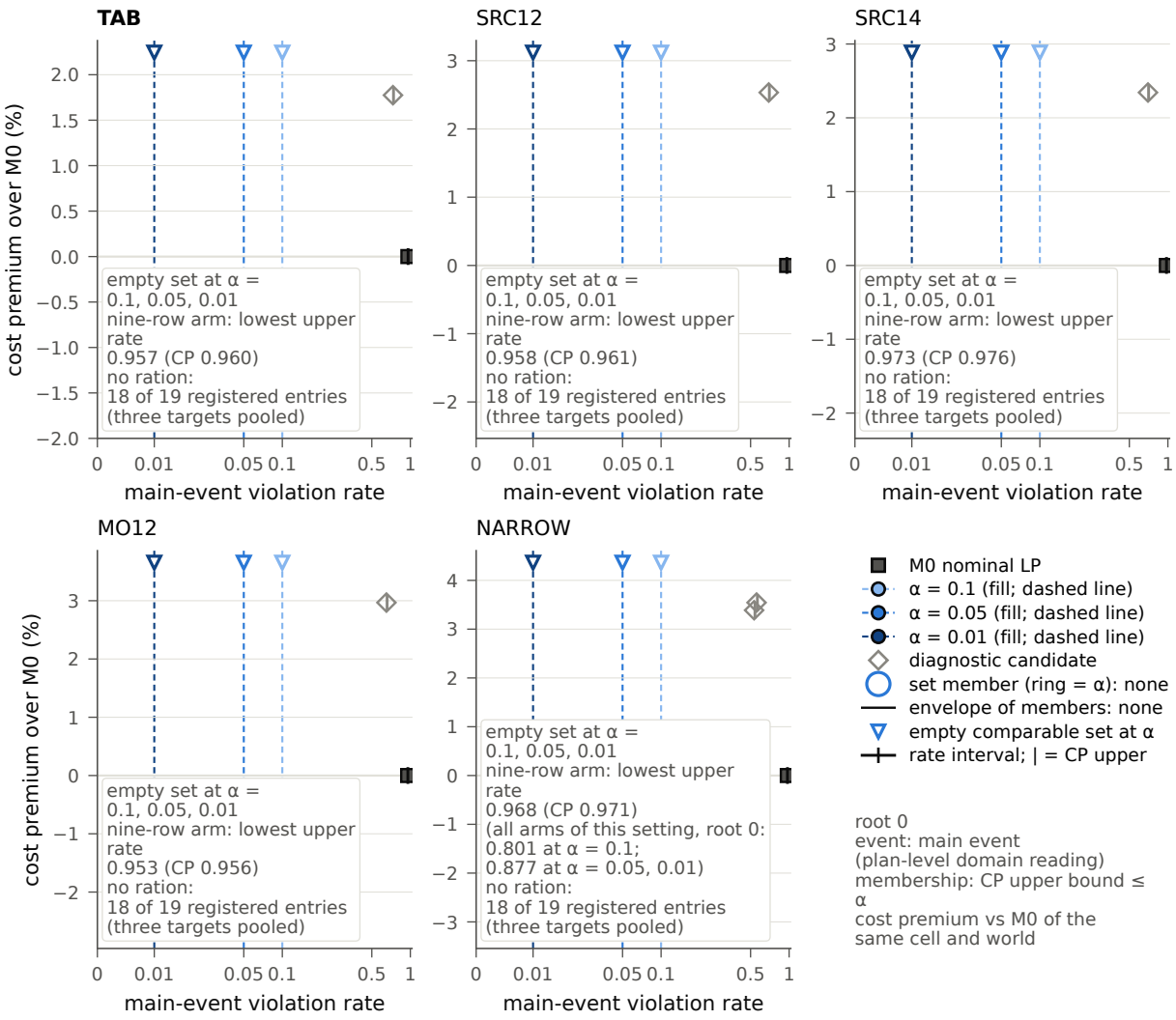
Supplementary Methods: Cross-evaluation re-tabulation (R5 main-text Table 5) and relocated displays (archive S15)

The revised main text foregrounds the prespecified TAB evaluation. Figures 2 and 4 now use the existing TAB panels from S13, without altering their plotted results. The earlier own-setting panels are retained below. In own-setting comparisons both formulation assumptions and evaluation distributions vary; these are not fixed-condition comparisons of candidate quality.

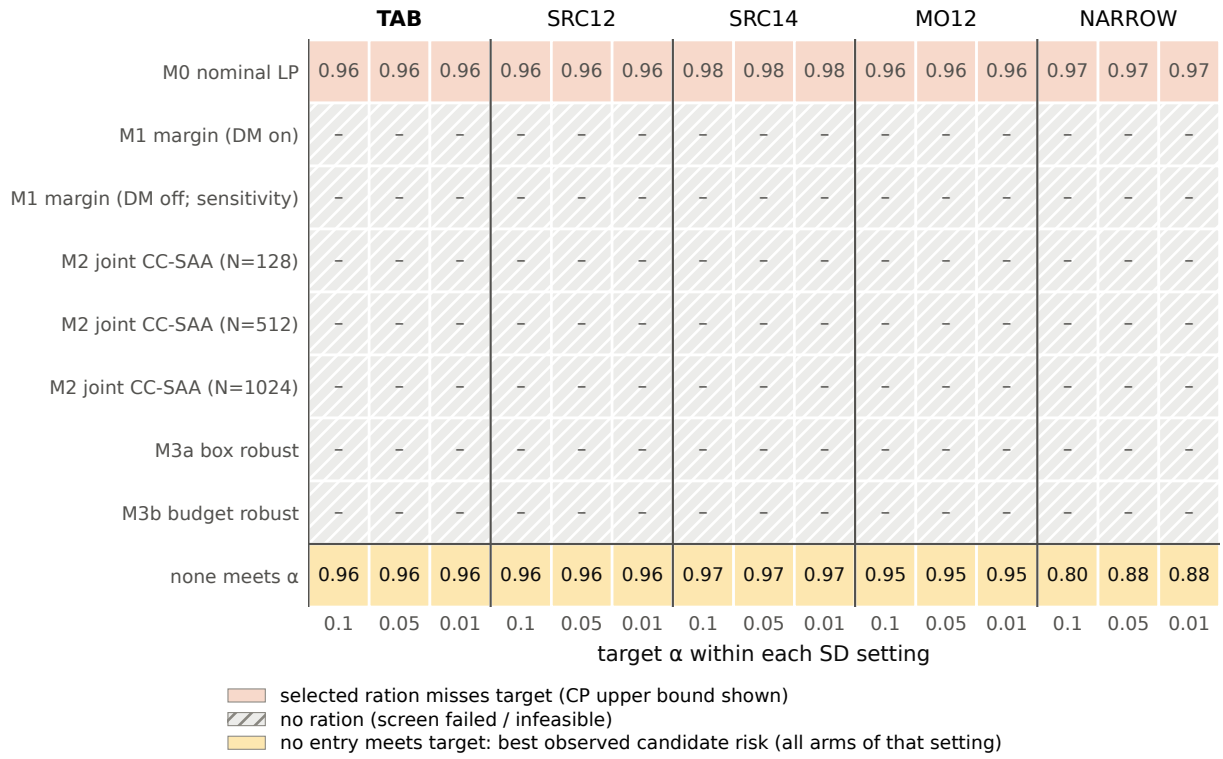
Main-text Table 5 (reference case) was constructed during manuscript revision from the archived endpoint CSVs for roots 0, 1 and 2. Rows were retained when `ration_kind` was `diagnostic_frontier`, `objective_arm` was `MAIN9`, the candidate label contained `[N=128,`, and an explicit evaluation world was present. This gave 75 records: 15 candidate identifiers evaluated in five worlds. Each record had 10,000 test states. Reported rates were checked against the sum of known-violation and unresolved counts divided by the test denominator. Own-setting and TAB records were matched by root and candidate identifier. Their paired differences, not differences of range endpoints, define the within-root difference column of Table 5. The maximum-regret columns of Table 5 and the break-even prior in its note are computed from the same 75 records; their definitions and the full regret matrices of all three roots are given in S18.

This is a descriptive re-tabulation of previously generated results. It is not a new experiment, newly reserved test, or external validation. No paired biological replication or inferential test is claimed. Exact ration-vector identity could not be independently checked because the release does not include those vectors; matching uses the recorded identifiers.

The two added CSVs are `S15_diagnostic_cross_evaluation_75_rows.csv` and `S15_diagnostic_cross_evaluation_summary.csv`. The scientific row order and retained aggregate values, including the own-setting summary that formed Table 4 of the earlier version, are unchanged; the current CSV copies omit private execution metadata. At the time of this archive, a script (`reproduce_table4.py`, named after that earlier table number) regenerated the failure-rate columns of R5 Table 5 directly from the three metadata-projected endpoint CSVs; it is held in the study archive and is not part of the current Supplementary Data.



Supplementary Fig. S15. Original own-setting cost and risk display, first reserved root. Each panel evaluates rations under that panel's training assumption. Diagnostic diamonds are separate from the selected-strategy pool and from the candidate-pool minima printed in the boxes. For the same rations evaluated in the primary TAB setting, see R5 main Fig. 2.



Supplementary Fig. S16. Original own-setting target-attainment display, first reserved root. Settings change across columns. Hatching means that no ration was selected, not necessarily that the complete nonlinear problem was proven infeasible. The bottom row summarises a different candidate pool from the diagnostic rations. See R5 main Fig. 4 for the common-TAB comparison.

Supplementary Methods: Reference problem v3: case family, registration and freeze (archive S16)

The three arms of the case family are separately registered reference problems (`reference_problem_v3a`, `v3b` and `v3c`; decision D-550), not replacements of the reference case. The reference case (R0) is reference problem v2 (`dev_case_v1` with the starch-source premise row) as evaluated in the official run `official-20260926T162636Z-e3665bbe`; its results are unchanged and were not rerun. Each arm changes only the items listed in S16.1–S16.3 and Table S16-1. The cow, the nine-row main event, the reference evaluator, the five SD settings, the strategies and their grids, the validation screen, the N ladder and the validation and test sizes are those of R0 (S5, S8). As for R0, numerical inputs and thresholds of the arms are withheld and identified by pointer and hash (S1, S2); costs are expressed relative to each arm's own nominal ration (M0). Internal identifiers are mapped to the descriptive names of the main text in S20.

S16.1 Arm A: planned intake +10%

The case directory `configs/dev_case_v3a/` is a verbatim copy of `configs/dev_case_v1/` except for the items below.

- Planned DM supply. The planned-DM setpoint (`SH-DM-PLAN`) is the Eq. 2-1 intake prediction multiplied by 1.1, the registered sensitivity scenario `DMI_plus10` of that row. It is a setpoint, not an intake band (D-551). The multiplier enters in three files only: the animal file (planned DM supply), the energy file (the intake at which the energy row is linearised) and the constraint file (the bound of the planned-DM equality).
- Energy row. The linear energy row is re-derived by the case builder at the scenario intake; the NEL requirement is unchanged.
- CP supply target. Carried over unchanged under the declared transfer rule, which transfers the Table 21-3 amount across intakes (research scenario assumption). The registered alternative that scales the CP target with intake remained an optional sensitivity and is not part of arm A.
- Absorbed calcium and phosphorus. Maintenance faecal losses are computed on the realised DM supply (the adopted option), so the requirements contain no fixed-intake constant to revise; the row bounds are unchanged.
- Reference table. `reference_constraints_v3a.csv` equals the R0 table (`reference_constraints_v2.csv`, 28 rows) except the planned-DM row, whose setpoint, source pointer, status (research assumption), rationale and after-seeing-results flag state the $\times 1.1$ scenario. Its `table_version` field keeps the label `reference_problem_v2`, the name of the rule set that the evaluator checks; the identity of the arm is carried by the file name, the case directory and the pin.
- Unchanged. Inventory (eight ingredients), prices, carbohydrate rules, inclusion caps, the premise planning row, the SD registry (`reports/sd_scaling_sources.csv`, 48 cells) and the restricted feed-library extract and restricted values are those of R0; the restricted constants file differs from that of R0 only in the case identifier.

The 10% increase is an intake-assumption lever (intake-prediction uncertainty or a higher-intake cow), not a claim that the reference cow eats more. An exploratory development probe motivated the arm; none of its numbers is used in the manuscript.

S16.2 Arm B: whole cottonseed added

The case directory `configs/dev_case_v3b/` differs from `dev_case_v1` in the inventory, price and constraint files; the animal and energy files are copies with the case identifier changed.

- Ingredient. Whole linted cottonseed (`cottonseed_whole_linted`; NASEM "Cottonseed, whole"; feed-library entry `NRC16F56`) is added as a ninth, stochastic ingredient (inventory `DEV-INV-V3B-9`). The eight R0 ingredients keep their order and wording.
- Composition. Means, SDs and record counts of eight random cells (DM, CP, NDF, starch, EE, ash, Ca and P) are taken from NASEM (2021) Table 19-1 (p. 374) and were checked by the double-entry route of S4 (ten core rows present and agreed, including lignin and total fatty acids); RUP and its digestibility agree with the feed library within print rounding. Lignin and total fatty acids are held at single values, as for the R0 ingredients; total fatty acids enter only the energy chain. One cell, starch, has a table SD at least as large as its mean. It is flagged in the build record and handled by the existing fallback rule (S3.3): no truncated-normal parent reproduces the two target moments on the physical support, so the cell takes the moment-matched scaled beta, with the targets unchanged. The same fallback applies to one cell of the R0 inventory.
- Digestibility and absorption inputs (feed library). Starch digestibility at the Table 3-1 default for feeds not listed there (checked equal to the library value); calcium absorption coefficient at the library class value for other plant concentrates; phosphorus absorption by the NASEM rule on p. 113 with the library inorganic and organic phosphorus fractions (checked: the fractions sum to 100% and the rule reproduces the library value).
- Fibre class. The NDF of cottonseed counts as non-forage NDF: forage weight 0, obtained from the library concentrate share by Eq. 20-29 (decision D-07).
- Cap. A structural row (`SH-INCL-COTTONSEED-MAX`) limits cottonseed to 10% of planned DM (decision-time DM; not a member of any event). Basis: NASEM (2021) printed p. 340 reports that diets with 10% and 15% of DM as whole cottonseed supplied 817 and 1,295 mg free gossypol per kg DM, that free gossypol up to 1,050 mg per kg DM had no detrimental effect on DM intake or milk production, and that erythrocyte fragility increased at 1,050 mg per kg DM. A 10% share therefore kept free gossypol below that concentration for the cottonseed of those observations. Because the book reports observations rather than a limit, and gossypol content varies among cottonseed sources, the cap is a research assumption; effects of gossypol and of fat on the rumen are not modelled.
- Fat loading. The ether-extract limit stays a companion row in evaluation (D-552; the D-550 wording that promoted it to a structural row is superseded). In formulation, CP and fat loading are bounded by the plan-level CP-upper and EE rows of the primary nine-row arm (`SH-PLAN-CP-HI`, `SH-PLAN-EE-HI`; nominal composition, decision-time DM; carried over from R0) and by the 10% cap.
- Price. Scenario price = midpoint of the conventional-cottonseed price range of 215–350 USD per ton (2025 crop) in the USDA Agricultural Marketing Service Organic Cotton Market Summary, Vol. 17 (August 2026), accessed 28 September 2026 (source identifier `SRC-P-USDA-AMS-COTTON`; dated snapshot held in the project's source-snapshot directory), taken per short ton as fed. Status: research scenario assumption (a range midpoint, not a dated Midwest quote; freight excluded). The range endpoints (215 and 350 USD per ton) are registered as the sensitivity band. The band was not run in the original case-family phase; its completed separately registered price/reference-screen extension is reported in S22. In the original phase, at the midpoint price cottonseed did not enter the nominal ration of arm B or arm C. The other eight prices are those of R0 (price scenario `DEV-PRICE-USMW-2026W38-V3B`).
- Reference table. `reference_constraints_v3b.csv` consists of the 28 rows of the R0 table, byte-identical, plus the cap row (29 rows), which follows the soybean-hull cap (planning

and optimisation only; structural check on decision-time DM).

- Build. 149 of 149 build checks passed (133 of R0, 16 new). The restricted feed-library extract has 13 rows (the 12 R0 rows, byte-identical, and the cottonseed row from the pinned full library version); the restricted constants file is that of R0 with the case identifier and layout changed; no new restricted value was needed.
- SD registry extension. `reports/sd_scaling_sources_dev_case_v3b.csv` has 56 cells: the 48 R0 cells, unchanged, and the eight cottonseed cells. No ratio of any basis is registered for cottonseed, so its cells follow the rule for cells without a sourced ratio: ratio 1 (the table SD) in TAB, SRC12 and SRC14, and the borrowed corn-silage 12-month DM ratio in MO12 and NARROW. The borrowed value is not a source value for these cells, its 12-month time scale differs from the declared 14-day batch horizon, and its evidence status is "assumption only". Ratio values are not printed (S3.3). The registry is shared by arms B and C.

S16.3 Arm C: arms A and B combined

The case directory `configs/dev_case_v3c/` combines the arm-B inventory, prices and cottonseed cap with the arm-A planned-DM scenario: planned DM supply at the Eq. 2-1 prediction $\times$ 1.1 in the animal, energy and constraint files; the energy row re-derived at that intake; the CP target carried over unchanged; calcium and phosphorus maintenance on realised DM; other row bounds unchanged. `reference_constraints_v3c.csv` is the arm-B table with the planned-DM row at the $\times$1.1 setpoint (29 rows). The SD registry is the arm-B file (56 cells). The restricted constants file is that of arm B with the case identifier changed (13-row feed-library extract; restricted values as R0); 149 build checks.

Table S16-1. Case family. Hashes are 12-character SHA-256 prefixes of the tracked files (full values in `configs/protocol_v3_freeze.json` and the pins). No numerical input, threshold or price is shown.

Case	Inventory	Planned DM supply	Inclusion-cap rows	Reference table	SD registry	Pin
R0: reference problem v2 (`dev_case_v1` + premise row)	Eight ingredients: six stochastic, two point-valued minerals; 48 random cells	Eq. 2-1 prediction	Limestone, dicalcium phosphate, soybean hulls	`reference_constraints_v2.csv` (28 rows; `795427bfa707...`)	`sd_scaling_sources.csv` (48 cells; `f808110afb7b...`)	`reference_problem_v2_freeze.json` (digest `2cbf4d847463...`)
Arm A: `reference_problem_v3a` (`dev_case_v3a`)	As R0; 48 random cells	Eq. 2-1 prediction $\times$ 1.1 (`DMI_plus10`)	As R0	`reference_constraints_v3a.csv` (28 rows; planned-DM row changed; `cff9e6d4c2e1...`)	As R0 (same file)	`reference_problem_v3a_freeze.json` (file `be7baa152aba...`; digest `b18224346d85...`)
Arm B: `reference_problem_v3b` (`dev_case_v3b`)	Nine: R0 plus whole linted cottonseed (stochastic); 56 random cells	As R0	As R0 plus cottonseed $\leq$ 10% of planned DM	`reference_constraints_v3b.csv` (29 rows: the R0 rows plus the cap row; `c0f8f717ea65...`)	`sd_scaling_sources_dev_case_v3b.csv` (56 cells; `d217519df784...`)	`reference_problem_v3b_freeze.json` (file `aa1e9bae5fca...`; digest `df6d19bc35e4...`)
Arm C: `reference_problem_v3c` (`dev_case_v3c`)	As arm B; 56 random cells	As arm A	As arm B	`reference_constraints_v3c.csv` (29 rows: arm-B table with the planned-DM row at $\times$1.1; `eefb94cf46f6...`)	As arm B (same file)	`reference_problem_v3c_freeze.json` (file `230760931622...`; digest `e8f6dddb167d...`)

S16.4 Official configuration per arm

Each arm runs the primary nine-row training arm (MAIN9) in the five SD settings on each of the three reserved roots (15 jobs) and one diagnostics job on root 0 (the starch-maxima removal and energy-only diagnostics, trained in TAB and NARROW): 16 jobs per arm, 48 in total. The historical eleven-row and six-row training targets of R0 are not repeated (registered deviation). Strategies and grids (S5), the validation screen (one-sided Clopper–Pearson bound $\leq \alpha$, confidence 0.95), the evaluator (reference energy chain; plan-level Table 5-1 reading), the N ladder (128, 512, 1,024; prefix nesting), 20,000 validation and 10,000 test states per evaluation setting, the targets ($\alpha = 0.10$, 0.05, 0.01; TAB and 0.05 primary) and eight workers are those of R0. The v3 driver adds only the case switch and the arm registry. The primary assumption identifier of each arm is `<case>|SD-H0|TN_MM+BETA_MM|C0|DM-uncertain|plan-domain`. The declared matrix, streams and official-configuration hash of each arm are recorded in `configs/protocol_v3_freeze.json`.

S16.5 Registration, freeze and prior access

- Decisions. D-550 (28 September 2026) registered the case family after the R0 results and exploratory development results had been seen; D-551 fixed the arm-A setpoint and classified constraint pricing as exploratory (S17); D-552 settled the pre-freeze items (ether-extract limit kept as a companion row, cottonseed cap, price source, calcium absorption input from the feed library, the `table_version` label, transcription of the compute authorisation); D-553 froze the family. The freeze record also lists D-545, D-546, D-548 and D-549 (reporting and manuscript revision of the reference case) among the decisions taken after results had been seen.
- Freeze commit `2779a056c3e2...` (parent `d423ccd6d11c...`). It contains exactly five files: `configs/protocol_v3.yaml` with `is_frozen: true` (freeze timestamp 2026-09-27T18:24:13Z; SHA-256 `d432c9c8e22b...`), the freeze record `configs/protocol_v3_freeze.json` (written 2026-09-27T18:24:15Z; SHA-256 `fdec09351435...`) and the three pins `experiments/E1_cost_reliability/reference_problem_v3{a,b,c}_freeze.json` (Table S16-1). Code-manifest subset at the freeze: 208 files (the freeze record and the pins excluded), subset SHA-256 `9198d2807c33...`, complete; code manifest SHA-256 `95d5a316f260...`.
- Run-time gate. For each arm the gate checks that HEAD is the freeze commit, that the work tree is clean and that the protocol, pin, specification and data digests match the freeze record.
- Pre-freeze checks on the development machine. 1,205 automated tests passed (three expectations updated to the four-case and authorised-host semantics); a rehearsal of arm B (six jobs) passed; the official dry-run of each arm matched the frozen, anchored specification and was stopped only by the host-authorisation check, as intended.

Reuse of the reserved roots (disclosed limitation). The three reserved roots of the arms are the three roots of R0. They had already been drawn by the R0 run `official-20260926T162636Z-e3665bbe`, and the R0 results, together with exploratory results on development streams (probes of the planned-intake lever and the constraint-pricing analysis of S17), had been seen before the arms were designed and frozen; the freeze record states that the reserved roots were drawn before the freeze. The v3 arms are therefore new problems evaluated on the same reserved streams, not fresh roots: their results are a second look at those streams under changed problems, not an independent held-out test. The roots were not changed to select a result (streams policy), and the stream generator still refuses a reserved root without a token of the frozen-protocol check. The arm-A setpoint (rather than an intake band) and the choice of whole cottonseed as an energy source supplying fat rather than starch were made after the energy–starch restriction of the reference case had been seen (D-550, D-551).

Documentation edited after the freeze. The case README of each arm and the design note of the case family are among the files whose hashes enter each arm's specification fingerprint. Their wording was corrected after the freeze for documentation only; the four files were then restored to the frozen bytes and the corrected text kept in a note outside the fingerprint scope (`docs/v3_post_freeze_doc_corrections_20260928.md`), so the working tree matches the three pins again. The official runs execute the freeze commit; no specification file changed.

S16.6 Compute-host preparation

The freeze was transferred to the compute host as a git bundle and checked out at the freeze commit; an untracked copy of the R0 result directory was first moved aside (its content was identical to the tracked copy). The restricted build inputs of the three arms were transferred as the bytes built on the development machine, together with the R0 side files and the compute-authorisation record. Byte reproduction of the restricted build records differs across platforms (last-digit floating-point differences in check fields of the build record; open item B-445). As for R0 (S7), the compute host therefore replays the development-machine bytes, whose hashes are pinned (Table S16-1; S19.1). The automated preparation chain stopped at its test step with four failures of this class (comparisons of an in-memory rebuild on the compute host with the replayed bytes; one of them compares only echoed field names). The remaining preparation steps were executed manually and passed: the official preflight of each arm; the official dry-run of each arm (matches the frozen, anchored specification; host authorised; no draws); the environment check against the lock; and a clean work tree within the checked scope. The test step had reported 1,191 passed, 4 failed and 3 skipped tests for each arm. The compute-authorisation record (scope: the official v3 runs) was transcribed from the research lead's written instructions, the same disclosed deviation as for R0 (S9.3, D-519). The arms were then run serially (A, B, C) with eight workers at reduced priority. Run identifiers: arm A `official-v3a-20260927T184000Z-8d7138b2` (wall time 9,623 s), arm B `official-v3b-20260927T212028Z-a3536fbd` (2,566 s) and arm C `official-v3c-20260927T220316Z-5d86b41a` (10,194 s); each completed all 16 jobs with exit status 0 and passed the six checks of the output verification script (S19).

S16.7 Displays of the arm results

The arm results are displayed in the main text: the attainability ladder (Fig. 6), the price of reliability in all five evaluation settings (Fig. 7, which replaces the earlier TAB-only display and the planned all-settings supplementary figure), the regret matrices of the first root (Fig. 8) and the linearisation asymmetry (Fig. 9). No supplementary figure was added for the arms; roots 1 and 2 and the underlying records are tabulated in S18, and the plotted values are in the CSV files `Fig6_attainability_ladder.csv`, `Fig7_price_of_reliability.csv`, `Fig8_regret_matrix.csv`, `Fig8_regret_summary.csv`, `Fig9_linearisation_asymmetry.csv` and `v3_attainability_by_world.csv`.

Supplementary Methods: Constraint pricing at the nominal solution (exploratory) (archive S17)

Status. Exploratory analysis, registered as v2.1 after the reference-case results had been seen (D-550 item 1; D-551). It uses development streams only, is never a member of any comparable set, is not merged into the official tables and replaces no official number. Numbers from this section should be quoted with this label.

S17.1 Problem, streams and run

- Problem: the reference case (reference problem v2) with the primary nine-row arm, built by the official driver exactly as in the official run, including the starch-source premise row and the plan-level CP-upper and EE rows.
- Setting and streams: TAB (`SD-H0`); one development root of its own (not a reserved root; its seed is not printed, S9.2) with 128 optimisation, 2,000 validation and 2,000 test states.
- Code and run: `experiments/E3_constraint_pricing/constraint_pricing.py` (SHA-256 `a6cc365b2bc9...`), code commit `a291fd2462...`; run identifier `dev-20260927T172445Z-...` (the suffix carries the development seed and is withheld); run time 6 s.
- At the time of this archive, dual prices in USD, margin SDs in natural units, rations, requirements and absolute costs were held in the restricted directory and only relative quantities, signs, counts, κ values and rates were shown; the equation-derived requirement amounts have since been released (current S38).

S17.2 Method

- Dual prices. The M0 program was re-solved as the same linear program with HiGHS through SciPy to obtain the dual prices of the inequality and equality rows. At M0, 7 of the 8 ingredient variables were positive (canola meal at zero), 6 inequality rows and the planned-DM equality were binding, and the 8 active constraints matched the 8 variables: the vertex was non-degenerate and the duals unique. Finite-difference derivatives with right-hand-side steps of $\pm 10^{-4}$ of each row's activity scale agreed with the solver duals for every binding row, in both directions.
- Margin SD. For each probabilistic row k, s_k is the SD of the row margin at the fixed M0 ration over the 2,000 TAB test states of the development stream, in the units of the linear row (the units of its dual); for the energy row also through the reference chain.
- Price. $P_k = 100\,|y_k|\,s_k / C(q^{(0)})$, in % of the M0 cost per SD (first order).
- Exact tightening. M0 re-solved with only row k's right-hand side tightened by 0.5, 1 and 2 SD; the result is a cost premium over M0 in %, or *infeasible* (no ration satisfies all other rows with row k so tightened).
- Head-room. Largest feasible relative margin κ (the M1 relative, coefficient-directional margin with the DM margin applied) for each row alone, each row group alone, each pair of groups and each pair of rows (common κ), and for all nine rows (the uniform M1), by bisection on $[0, 1]$ to a tolerance of 10^{-4}. The largest tightening of the energy row alone, in SD of its margin, was found by bisection to a tolerance of 10^{-3}.
- Classes: free = not binding; cheap = binding, 2 SD alone feasible and κ alone ≥ 0.10; limited = binding, 1 SD alone feasible but not both of those conditions; locked = binding and 1 SD alone infeasible.
- Row removal. The problem was re-solved without the named row or rows, and the energy row was then tightened alone. This is a counterfactual on the problem, not a proposed formulation.

- Box-robust head-room. The largest feasible half-width k of the engine's M3a (box of mean $\pm k$ SD on the primitive inputs, exact energy image, all nine rows), by bisection; and, as a per-row proxy that is not the engine's M3a, the largest feasible k of the Soyster counterpart of each row alone, with spreads equal to the sample SDs of the 128 optimisation draws.
- Directed safety margins (M1-T). Four row groups: minerals (absorbed calcium and phosphorus), protein (CP supply), energy (NEL) and carbohydrate (the five Table 5-1 rules), each with its own $\kappa_g \in \{0, 0.025, 0.05, 0.10\}$: $4^4 = 256$ patterns. Each group's rows take the M1 rows at the group's κ; because the M1 construction is row-wise, a pattern is an exact recombination of M1 rows and the engine code is unchanged. The M1 grid value 0.075 is not part of M1-T. Every feasible pattern was screened on the validation stream (one-sided Clopper–Pearson upper bound $\leq \alpha$ at confidence 0.95; $\alpha \in \{0.10, 0.05\}$) and scored on the test stream against the main event (reference energy chain, plan-level Table 5-1 reading). As a floor, the exact minimum number of violated training states was computed at $N = 128$ and its frontier ration scored on the same test stream.
- Regression checks. The re-solved program reproduced the engine's M0 (objective difference $\leq 10^{-9}$; KKT identity residual $< 10^{-9}$). Uniform M1-T patterns reproduced the engine's M1 at $\kappa = 0$ (optimal) and 0.025, 0.05 and 0.10 (proven infeasible). The linear margin of the reference energy check equalled the LP energy-row margin, so the energy dual prices both. On the development test stream the M0 main-event rate was 0.9570 (official root 0: 0.9567) and the linear training-event rate 0.9585; the ration lay in the plan-level Table 5-1 domain.

S17.3 Row pricing

Table S17-1. Price of one SD of margin for each probabilistic row at the nominal ration (development streams, TAB; one root; test stream of 2,000 states).

Row	Group	Binding at M0	Dual sign	Price, first order (% of M0 cost per SD)	Exact 0.5 SD premium (%)	Exact 1 SD premium (%)	Exact 2 SD premium (%)	Failure share at M0 (test)	Largest κ alone	Class
Energy (NEL), linear row (`PN-NEL-FIXEDDMI`)	energy	yes	−	14.38	infeasible	infeasible	infeasible	0.4970	0.0044	locked
CP supply (`PN-CP-SUP`)	protein	yes	−	1.56	0.78	1.57	infeasible	0.5085	0.0289	limited
Absorbed calcium (`PN-CA-ABS`)	minerals	yes	−	0.33	0.17	0.33	0.66	0.4985	0.1988	cheap
Absorbed phosphorus (`PN-P-ABS`)	minerals	yes	−	0.19	0.09	0.19	0.38	0.5215	0.3422	cheap
Forage-NDF minimum (`PN-T1`)	carbohydrate	no	0	0.00	0.00	0.00	0.13	0.0340	0.1883	free
Total-NDF minimum (`PN-T2`)	carbohydrate	no	0	0.00	0.00	0.00	0.00	0.0070	0.1376	free
Forage-NDF-dependent NDF minimum (`PN-T3`)	carbohydrate	no	0	0.00	0.00	0.12	infeasible	0.2180	0.0724	free
Starch maximum (`PN-T4`)	carbohydrate	yes	−	1.74	2.00	infeasible	infeasible	0.4995	0.0346	locked
Forage-NDF-dependent starch maximum (`PN-T5`)	carbohydrate	yes	−	0.34	0.17	2.50	infeasible	0.5095	0.0496	limited
Energy (NEL), reference chain	energy	—	—	13.67	—	—	—	0.4965	—	—

Notes. Dual sign: sign of the change in cost per unit change of the row's right-hand side as reported by the solver; 0 = not binding. First-order price = |dual| × margin SD as % of the M0 cost; finite-difference prices (tightening and loosening) equal the dual price in every binding row. Largest κ alone: largest feasible relative margin on that row alone. Structural rows at M0: only the planned-DM equality was binding (sign −); the three inclusion caps, the starch-source premise and the plan-level CP-upper and EE rows were slack. Energy row: largest feasible tightening alone about 0.15 SD of its M0 margin (bisection bracket 0.1465–0.1471 SD); the margins of the linear row and of the reference chain correlated 0.989 across test states; joint nine-row training event at M0 0.9585 on the test stream. Signs of dual × coefficient at M0: dry ground corn was credited by the energy, CP-supply and absorbed-phosphorus rows and debited by the absorbed-calcium row, both starch maxima and the planned-DM equality.

S17.4 Group and pair head-room

Table S17-2. Largest feasible common relative margin κ for row groups, group pairs, all nine rows and row pairs (bisection; DM margin applied). Rows alone are in Table S17-1.

Scope	Largest feasible common κ
Row groups alone	
Minerals (absorbed Ca and P)	0.1815
Protein (CP supply)	0.0289
Energy (NEL)	0.0044
Carbohydrate (five Table 5-1 rules)	0.0209
Row-group pairs (common κ)	
Minerals + protein	0.0289
Minerals + energy	0.0043
Minerals + carbohydrate	0.0193
Protein + energy	0.0044
Protein + carbohydrate	0.0209
Energy + carbohydrate	0.0036
All nine rows	
All nine rows (common κ = uniform M1)	0.0035
Row pairs with a common κ below 0.10	
Energy + Starch max.	0.0039
Energy + fNDF-dep. starch max.	0.0040
Ca + Energy	0.0043
P + Energy	0.0044
CP supply + Energy	0.0044
Energy + fNDF min.	0.0044
Energy + NDF min.	0.0044
Energy + fNDF-dep. NDF min.	0.0044
Starch max. + fNDF-dep. starch max.	0.0209
Ca + CP supply	0.0289
P + CP supply	0.0289
CP supply + fNDF min.	0.0289
CP supply + NDF min.	0.0289
CP supply + fNDF-dep. NDF min.	0.0289
CP supply + Starch max.	0.0289
CP supply + fNDF-dep. starch max.	0.0289
Ca + Starch max.	0.0306
P + Starch max.	0.0341
fNDF-dep. NDF min. + Starch max.	0.0344
fNDF min. + Starch max.	0.0346
NDF min. + Starch max.	0.0346
Ca + fNDF-dep. starch max.	0.0421
P + fNDF-dep. starch max.	0.0489
fNDF min. + fNDF-dep. starch max.	0.0496
NDF min. + fNDF-dep. starch max.	0.0496
fNDF-dep. NDF min. + fNDF-dep. starch max.	0.0496
Ca + fNDF-dep. NDF min.	0.0621
P + fNDF-dep. NDF min.	0.0712
fNDF min. + fNDF-dep. NDF min.	0.0724
NDF min. + fNDF-dep. NDF min.	0.0724

Notes. Abbreviations: Ca, P = absorbed calcium, absorbed phosphorus; fNDF min. = forage-NDF minimum; NDF min. = total-NDF minimum; fNDF-dep. NDF min. = forage-NDF-dependent NDF minimum; Starch max. = starch maximum; fNDF-dep. starch max. = forage-NDF-dependent starch maximum (identifiers in S20). Row pairs not listed had a common κ of at least 0.10. The nine pairs that cannot jointly take $\kappa = 0.025$ are the eight pairs that include the energy row and the pair of starch maxima; the uniform M1 at 0.025 therefore fails because of the energy row alone (0.0044), the carbohydrate group (0.0209), the starch-maxima pair (0.0209) and all nine rows together (0.0035).

Box-robust head-room. The engine's M3a was infeasible at $k = 1$; its largest feasible half-width was $k = 0.0566$. Per-row box proxy (largest feasible k with that row alone): absorbed calcium and absorbed phosphorus ≥ 3; forage-NDF minimum 1.7561; total-NDF minimum 1.4241; forage-NDF-dependent NDF minimum 0.6954; CP supply 0.5824; forage-NDF-dependent starch maximum 0.5025; starch maximum 0.4253; energy 0.0819; all nine rows 0.0598. The rows that alone could not take a 1 SD box were CP supply, energy, the forage-NDF-dependent NDF minimum and both starch maxima.

S17.5 Row-removal experiment on the energy row

Table S17-3. What locks the energy row: the reference problem re-solved without the named row(s), then the energy row tightened alone. Counterfactual on the problem, not a formulation.

Row(s) removed	Nominal cost change of the relaxed problem (%)	Largest κ on the energy row alone	Energy row +0.5 SD	Energy row +1 SD	Energy row +2 SD	Premium of the 1 SD tightening over the reference-problem M0 (%)
none (reference problem)	0.00	0.0044	infeasible	infeasible	infeasible	—
Limestone cap	0.00	0.0044	infeasible	infeasible	infeasible	—
Dicalcium phosphate cap	0.00	0.0044	infeasible	infeasible	infeasible	—
Soybean-hull cap	0.00	0.0044	infeasible	infeasible	infeasible	—
CP supply	−2.82	0.0044	infeasible	infeasible	infeasible	—
Absorbed calcium	−1.87	0.0084	infeasible	infeasible	infeasible	—
Absorbed phosphorus	−0.06	0.0044	infeasible	infeasible	infeasible	—
Forage-NDF minimum	0.00	0.0044	infeasible	infeasible	infeasible	—
Total-NDF minimum	0.00	0.0044	infeasible	infeasible	infeasible	—
Forage-NDF-dependent NDF minimum	0.00	0.0044	infeasible	infeasible	infeasible	—
Starch maximum	−2.05	0.0141	infeasible	infeasible	infeasible	—
Forage-NDF-dependent starch maximum	−0.31	0.0064	infeasible	infeasible	infeasible	—
Starch-source premise	0.00	0.0056	infeasible	infeasible	infeasible	—
Plan-level CP upper row	0.00	0.0095	infeasible	infeasible	infeasible	—
Plan-level EE row	0.00	0.0044	infeasible	infeasible	infeasible	—
Both starch maxima	−2.56	0.0193	feasible	infeasible	infeasible	—
All five carbohydrate rules	−4.27	0.0673	feasible	feasible	feasible	3.81
Both starch maxima and CP supply	−23.38	0.0373	feasible	feasible	infeasible	−19.22

Notes. Nominal cost change: cost of the relaxed problem's nominal solution relative to the reference-problem M0. Energy tightening in SD of its M0 margin; "infeasible" = no ration meets the remaining rows with the energy row so tightened. The premium of the 1 SD tightening is relative to the reference-problem M0 and is negative when the relaxed problem is cheaper; "—" = not computed because the 1 SD tightening was infeasible.

S17.6 Directed safety margins (M1-T)

Of the 256 group patterns, 8 were feasible, and feasibility was monotone in κ (no monotonicity violation). The minimal infeasible patterns (every one-step reduction feasible) were a carbohydrate margin of 0.025 alone, an energy margin of 0.025 alone and a protein margin of 0.05 alone, the other groups at 0. The only maximal feasible pattern was minerals 0.10 with protein 0.025. The eight feasible patterns are therefore the four mineral levels combined with a protein margin of 0 or 0.025 and no energy or carbohydrate margin; the all-zero pattern is M0.

Table S17-4. M1-T feasibility map and test-stream scores (development streams, TAB). (a) Feasibility of minerals × protein margins with energy and carbohydrate margins at 0 (F = feasible, x = infeasible). (b) Number of feasible (minerals, protein) combinations, out of 16, for each (energy, carbohydrate) pair. (c) Test-stream scores of the named patterns.

(a)

Minerals κ \ protein κ	0	0.025	0.05	0.10
0	F	F	x	x
0.025	F	F	x	x
0.05	F	F	x	x
0.10	F	F	x	x

(b)

Energy κ \ carbohydrate κ	0	0.025	0.05	0.10
0	8	0	0	0
0.025	0	0	0	0
0.05	0	0	0	0
0.10	0	0	0	0

Table S17-4 (continued). (c) Test-stream scores of the named patterns.

Pattern κ (minerals / protein / energy / carbohydrate)	Role	Cost premium over M0 (%)	Main event rate (test)	Linear training-event rate (test)	Validation rate	Energy row, reference chain (test)	Worst non-energy row (rate)
0 / 0 / 0 / 0	nominal ration (M0)	0.00	0.9570	0.9585	0.9575	0.4965	P (0.5215)
0.10 / 0.025 / 0 / 0	lowest main-event rate on test; lowest linear-event rate on test	3.69	0.8085	0.8245	0.8275	0.4995	fNDF-dep. starch max. (0.5110)
0.025 / 0 / 0 / 0	cheapest pattern below M0; cheapest pattern more than 2 MC SE below M0	0.38	0.9090	0.9125	0.9115	0.4985	fNDF-dep. starch max. (0.5095)
0.10 / 0 / 0 / 0	minerals alone at the largest grid value	1.71	0.8535	0.8615	0.8595	0.5030	fNDF-dep. starch max. (0.5105)
0 / 0.025 / 0 / 0	protein alone at the largest grid value	2.01	0.9340	0.9390	0.9375	0.4950	Starch max. (0.5005)

Notes. Main event: reference energy chain, plan-level Table 5-1 reading. Linear training event: nine rows with the linear energy row. Cost premium relative to M0. Validation screen (one-sided Clopper–Pearson upper bound $\leq \alpha$, confidence 0.95, 2,000 validation states): not met at $\alpha = 0.10$ or 0.05; no pattern met it, and the lowest validation upper bound was 0.8413. Floor (development): exact minimum violations at $N = 128 = 85$ of 128 (proven), training rate 0.6641; frontier ration on the same test stream: main event 0.6965, linear event 0.7705, cost premium -0.73%. Official orientation (root 0, TAB, nine-row arm): frontier main event 0.7315; minimum-violation training fractions 0.742, 0.715 and 0.736 at $N = 128$, 512 and 1,024. The best pattern (0.8085) lay 0.1120 above the development floor and 0.0770 above the official root-0 frontier, and 0.1485 below M0. Monte Carlo SE of a rate with 2,000 test states: about 0.005 near 0.95 and 0.01 near 0.75; the ranking of close candidates is not resolved.

Supplementary Methods: Training-by-evaluation regret (archive S18)

S18.1 Definitions

For a candidate trained in setting t (the nine-row cell of t) and scored on the test stream of setting e on the same root, $r(t, e)$ is its upper main-event rate r^+ (unresolved states counted as failures; plan-level domain reading; 10,000 test states).

- Regret: $\rho(t, e) = 100\,[\, r(t, e) - \min_{t'} r(t', e)\,]$ percentage points, the minimum running over the training settings that have a candidate of the same type.
- Maximum regret: $\bar{\rho}(t) = \max_e \rho(t, e)$; the minimax training setting minimises $\bar{\rho}$.
- Two-point prior between the extreme settings. If the true setting is TAB with probability π and NARROW otherwise, let $a = r(\text{TAB}, \text{NARROW}) - r(\text{NARROW}, \text{NARROW})$ and $b = r(\text{NARROW}, \text{TAB}) - r(\text{TAB}, \text{TAB})$. With $a, b \geq 0$ and $a + b > 0$, training on TAB has the lower expected failure rate if and only if $\pi > \pi^* = a/(a + b)$. If $a = b = 0$, both choices have equal expected rates at every π. If $a \leq 0 \leq b$ with at least one strict inequality, TAB training is never worse ($\pi^* = 0$); if $b \leq 0 \leq a$, NARROW training is never worse ($\pi^* = 1$); if both are negative, the inequality reverses. Here a is a pairwise difference; it differs from the TAB-trained regret in NARROW when the NARROW-trained candidate is not the best candidate in NARROW (roots 0 and 2 below, where the MO12-trained candidate had the lowest NARROW rate).
- Cost-weighted variant. For each $t \neq$ TAB, $\Delta c = c_{\text{TAB}} - c_t$, where c is the candidate's cost premium over the nominal ration (%, same arm and root, own setting), and $\Delta\bar{\rho} = \bar{\rho}(t) - \bar{\rho}(\text{TAB})$. The ratio $\Delta\bar{\rho}/\Delta c$ is formed only when the two differences have the same sign: both positive means TAB training is dearer with lower maximum regret (percentage points of worst-case regret avoided per percentage point of extra cost); both negative means TAB training is cheaper with higher maximum regret (extra worst-case regret per percentage point saved). Sign caveat: the ratio is positive in both cases, so the direction must be read from the signs of Δc and $\Delta\bar{\rho}$, not from the ratio; when one setting is neither dearer nor higher in maximum regret it dominates and no ratio is formed. Premiums of different candidates are each relative to the same nominal ration, so their difference is in percentage points of its cost; for diagnostic rations they are recorded costs, not matched-reliability costs.

Computation: at the time of this archive, the figure script for the earlier Fig. 8 (`fig8_regret_matrix.py`, held in the study archive) computed these values from the archived endpoint and per-method CSVs. For R0 the script checks that its 75 diagnostic rates equal the 75 rows of `S15_diagnostic_cross_evaluation_75_rows.csv` and stops otherwise. Output files: `Fig8_regret_matrix.csv` (600 rows: 150 per case, 75 diagnostic and 75 qualifying-type rows) and `Fig8_regret_summary.csv` (153 rows: 33 for R0, 42 for arm A, 33 for arm B, 45 for arm C). No optimisation or simulation was repeated.

S18.2 Reference case: 128-state diagnostic rations

Table S18-1. Regret matrices of the 128-state diagnostic rations of the reference case, by root (percentage points). Rows: training setting; columns: evaluation setting (diagonal = own setting). Bold: maximum regret of the minimax training setting. The last row of each root gives the lowest r^+ in each evaluation column (the reference for regret). Cost premium: the candidate's premium over M0 (%).

Training setting t	TAB	SRC12	SRC14	MO12	NARROW	Maximum regret $\bar{\rho}(t)$	Attained in	Cost premium over M0 (%)
Root 0								
TAB	0.00	1.21	4.59	1.06	9.28	9.28	NARROW	1.78
SRC12	1.83	0.00	0.00	0.00	0.81	**1.83**	TAB	2.53
SRC14	4.36	3.97	6.66	7.46	8.71	8.71	NARROW	2.34
MO12	8.10	7.57	8.12	4.03	0.00	8.12	SRC14	2.97
NARROW	8.80	7.82	8.47	5.49	0.60	8.80	TAB	3.54
Column minimum r^+	0.7315	0.6932	0.6326	0.6104	0.551	—	—	—
Root 1								
TAB	0.00	4.15	12.37	5.95	21.16	21.16	NARROW	1.39
SRC12	1.42	0.00	0.00	1.08	4.27	**4.27**	NARROW	2.80
SRC14	4.73	2.87	1.67	4.82	2.52	4.82	MO12	3.55
MO12	6.58	6.94	8.91	0.00	4.71	8.91	SRC14	2.18
NARROW	9.82	8.75	8.02	4.68	0.00	9.82	TAB	3.38
Column minimum r^+	0.7124	0.6765	0.6224	0.5895	0.5263	—	—	—
Root 2								
TAB	0.00	2.19	13.90	5.77	21.65	21.65	NARROW	0.81
SRC12	4.41	0.00	0.00	3.74	2.14	**4.41**	TAB	3.41
SRC14	5.77	0.69	0.99	6.04	1.44	6.04	MO12	3.78
MO12	7.74	5.37	6.41	0.00	0.00	7.74	TAB	2.89
NARROW	11.48	7.78	10.88	8.40	3.63	11.48	TAB	3.79
Column minimum r^+	0.6985	0.6832	0.6053	0.5829	0.5264	—	—	—

Table S18-2. Break-even prior between TAB and NARROW training, reference case, 128-state diagnostic rations.

Root	a (pp)	b (pp)	π^*	Rule
0	8.68	8.80	0.497	train on TAB if $\pi > \pi^*$
1	21.16	9.82	0.683	train on TAB if $\pi > \pi^*$
2	18.02	11.48	0.611	train on TAB if $\pi > \pi^*$

Table S18-3. Cost-weighted comparison of TAB training with each other training setting, reference case, 128-state diagnostic rations.

Root	Setting t	$c_{\text{TAB}} - c_t$ (percentage points of cost)	$\bar{\rho}(t) - \bar{\rho}(\text{TAB})$ (pp)	Ratio (pp of maximum regret per percentage point of cost)	Reading
0	SRC12	−0.76	−7.45	9.8	TAB cheaper, higher maximum regret
0	SRC14	−0.57	−0.57	1.0	TAB cheaper, higher maximum regret
0	MO12	−1.19	−1.16	1.0	TAB cheaper, higher maximum regret
0	NARROW	−1.77	−0.48	0.3	TAB cheaper, higher maximum regret
1	SRC12	−1.40	−16.89	12.0	TAB cheaper, higher maximum regret
1	SRC14	−2.16	−16.34	7.6	TAB cheaper, higher maximum regret
1	MO12	−0.79	−12.25	15.5	TAB cheaper, higher maximum regret
1	NARROW	−1.99	−11.34	5.7	TAB cheaper, higher maximum regret
2	SRC12	−2.60	−17.24	6.6	TAB cheaper, higher maximum regret
2	SRC14	−2.97	−15.61	5.3	TAB cheaper, higher maximum regret
2	MO12	−2.08	−13.91	6.7	TAB cheaper, higher maximum regret
2	NARROW	−2.98	−10.17	3.4	TAB cheaper, higher maximum regret

Notes. Both differences are negative in every row, so TAB-trained candidates were cheaper and had higher maximum regret against every other setting on every root; the ratio is the extra worst-case regret per percentage point of cost saved. The main text (Section 3.5) reports π^* for these rations; the cost-weighted ratios are given here only.

S18.3 Cheapest qualifying rations at $\alpha = 0.05$

For this candidate type the ration of each training setting is its cheapest qualifying ration at $\alpha = 0.05$, with qualification judged in its own setting, then scored in every setting; a training setting without such a ration is left out of the minimum. In R0 no training setting had a qualifying ration on any root (all 75 rows of this type empty), so its regret matrices, π^* and cost-weighted ratios are not computable for R0. For arms A–C they are reported per arm and root where the qualifying sets are non-empty (Table S18-4). Arm B had no qualifying ration in any setting, so, as for R0, its regret matrices of this type are empty. In arm A, TAB- and SRC12-trained entries returned no ration that qualified in its own setting on any root, and in arm C only TAB-trained entries returned none; these settings are left out of the minimum, and π^* and the cost-weighted

comparison with TAB training are not computable. On every root the ration trained under the widest remaining setting (SRC14 in arm A, SRC12 in arm C) had zero regret in all five evaluation settings, while the MO12- and NARROW-trained rations had their maximum regret under SRC14 (32.0–37.8 percentage points in arm A, 33.5–42.3 in arm C). Table S18-5 gives the corresponding summary for the 128-state diagnostic rations of the arms.

Table S18-4. Regret matrices of the cheapest ration of each training setting that qualified at $\alpha = 0.05$ in its own setting, arms A and C, by root (percentage points, one decimal). Rows: training setting; columns: evaluation setting. Bold: maximum regret of the minimax training setting. The last row of each block gives the lowest r^+ in each evaluation column over the training settings with a qualifying ration.

Training setting t	TAB	SRC12	SRC14	MO12	NARROW	Maximum regret $\bar{\rho}(t)$	Attained in	Own-setting r^+	Cost premium over the arm's M0 (%); strategy
Arm A, root 0									
TAB	no ration qualified at $\alpha = 0.05$ in its own setting (left out of the minimum)								
SRC12	no ration qualified at $\alpha = 0.05$ in its own setting (left out of the minimum)								
SRC14	0.0	0.0	0.0	0.0	0.0	**0.0**	—	0.0187	17.1; M2, $N = 1{,}024$
MO12	21.3	33.0	34.9	1.7	2.0	34.9	SRC14	0.0202	11.1; M2, $N = 1{,}024$
NARROW	18.4	29.9	32.0	1.2	1.5	32.0	SRC14	0.0152	11.0; M2, $N = 1{,}024$
Column minimum r^+	0.1676	0.0418	0.0187	0.0027	0.0000	—	—	—	—
Arm A, root 1									
TAB	no ration qualified at $\alpha = 0.05$ in its own setting (left out of the minimum)								
SRC12	no ration qualified at $\alpha = 0.05$ in its own setting (left out of the minimum)								
SRC14	0.0	0.0	0.0	0.0	0.0	**0.0**	—	0.0175	17.2; M2, $N = 1{,}024$
MO12	20.9	33.0	35.7	1.6	2.2	35.7	SRC14	0.0235	10.7; M2, $N = 1{,}024$
NARROW	20.1	30.9	33.3	2.0	2.6	33.3	SRC14	0.0264	10.7; M2, $N = 512$
Column minimum r^+	0.1772	0.0481	0.0175	0.0074	0.0000	—	—	—	—
Arm A, root 2									
TAB	no ration qualified at $\alpha = 0.05$ in its own setting (left out of the minimum)								
SRC12	no ration qualified at $\alpha = 0.05$ in its own setting (left out of the minimum)								
SRC14	0.0	0.0	0.0	0.0	0.0	**0.0**	—	0.0237	17.2; M2, $N = 1{,}024$
MO12	20.5	34.0	37.8	2.6	3.0	37.8	SRC14	0.0304	10.5; M2, $N = 1{,}024$
NARROW	20.3	33.4	37.2	2.9	3.3	37.2	SRC14	0.0325	10.4; M2, $N = 1{,}024$
Column minimum r^+	0.2027	0.0630	0.0237	0.0042	0.0000	—	—	—	—
Arm C, root 0									
TAB	no ration qualified at $\alpha = 0.05$ in its own setting (left out of the minimum)								
SRC12	0.0	0.0	0.0	0.0	0.0	**0.0**	—	0.0254	23.3; M2, $N = 1{,}024$
SRC14	8.9	4.6	2.2	0.6	0.0	8.9	TAB	0.0318	17.0; M2, $N = 512$
MO12	29.6	38.1	39.3	3.1	3.2	39.3	SRC14	0.0321	10.6; M2, $N = 1{,}024$
NARROW	24.5	32.5	33.5	3.1	3.2	33.5	SRC14	0.0318	10.4; M2, $N = 1{,}024$
Column minimum r^+	0.1197	0.0254	0.0102	0.0011	0.0000	—	—	—	—
Arm C, root 1									
TAB	no ration qualified at $\alpha = 0.05$ in its own setting (left out of the minimum)								

Training setting t	TAB	SRC12	SRC14	MO12	NARROW	Maximum regret $\bar{\rho}(t)$	Attained in	Own-setting r^+	Cost premium over the arm's M0 (%); strategy
SRC12	0.0	0.0	0.0	0.0	0.0	**0.0**	—	0.0240	21.7; M2, $N = 1{,}024$
SRC14	2.0	1.5	1.5	0.2	0.0	2.0	TAB	0.0223	17.0; M2, $N = 512$
MO12	30.9	40.8	42.3	2.6	2.5	42.3	SRC14	0.0262	10.9; M2, $N = 1{,}024$
NARROW	27.0	36.9	38.3	2.3	2.2	38.3	SRC14	0.0221	10.7; M2, $N = 1{,}024$
Column minimum r^+	0.1293	0.0240	0.0070	0.0003	0.0000	—	—	—	—
Arm C, root 2									
TAB	no ration qualified at $\alpha = 0.05$ in its own setting (left out of the minimum)								
SRC12	0.0	0.0	0.0	0.0	0.0	**0.0**	—	0.0153	22.8; M2, $N = 1{,}024$
SRC14	8.5	3.2	1.3	0.6	0.0	8.5	TAB	0.0197	17.0; M2, $N = 1{,}024$
MO12	32.6	40.8	41.5	3.3	3.2	41.5	SRC14	0.0330	10.7; M2, $N = 512$
NARROW	29.9	37.5	37.8	2.7	2.7	37.8	SRC14	0.0270	10.4; M2, $N = 1{,}024$
Column minimum r^+	0.1046	0.0153	0.0067	0.0005	0.0000	—	—	—	—

Notes. Source: `Fig8_regret_matrix.csv` and `Fig8_regret_summary.csv` (candidate type `cheapest_qualified_a0.05`). π^* and the cost-weighted comparison with TAB training are not computable in either arm, because no TAB-trained ration qualified in its own setting on any root (rule "not computable (candidate missing in TAB or NARROW)"). Arm B and R0 had no qualifying ration in any setting (all 75 rows of this type empty in each). Costs are relative to each arm's own nominal ration; they are not comparable across arms.

Table S18-5. Maximum regret of the 128-state diagnostic rations of arms A–C by training setting and root (percentage points, one decimal; evaluation setting attaining the maximum in parentheses; bold = minimax), with the break-even prior π^* between TAB and NARROW training (S18.1). The corresponding reference-case values are in Tables S18-1 and S18-2.

Arm	Root	TAB	SRC12	SRC14	MO12	NARROW	Minimax	π^* (a, b in pp)
A	0	**0.0**	2.2 (SRC14)	13.9 (TAB)	47.5 (SRC14)	41.4 (SRC14)	TAB	0.00 (−6.7, 32.9)
A	1	6.9 (SRC14)	**0.4** (TAB)	3.0 (TAB)	45.1 (SRC14)	49.1 (SRC14)	SRC12	0.00 (−4.2, 36.9)
A	2	3.7 (SRC12)	**0.1** (MO12)	8.3 (TAB)	36.3 (SRC14)	32.7 (SRC12)	SRC12	0.00 (−2.5, 25.4)
B	0	13.1 (NARROW)	**2.9** (TAB)	8.6 (MO12)	13.8 (SRC14)	13.4 (TAB)	SRC12	0.43 (10.2, 13.4)
B	1	25.6 (NARROW)	20.8 (NARROW)	**9.4** (TAB)	15.8 (SRC14)	14.2 (TAB)	SRC14	0.62 (23.1, 14.2)
B	2	9.3 (NARROW)	7.3 (NARROW)	**6.9** (TAB)	7.4 (SRC12)	9.8 (TAB)	SRC14	0.55 (9.3, 7.7)
C	0	**3.3** (SRC14)	3.8 (TAB)	3.4 (TAB)	33.6 (SRC12)	32.7 (SRC14)	TAB	0.00 (−5.2, 28.6)
C	1	**0.6** (SRC14)	6.3 (TAB)	4.2 (TAB)	44.3 (SRC14)	42.4 (SRC12)	TAB	0.00 (−4.0, 38.2)
C	2	4.7 (SRC14)	**0.1** (TAB)	8.7 (TAB)	33.0 (SRC14)	34.5 (SRC14)	SRC12	0.00 (−1.9, 25.6)

Notes. Source: `Fig8_regret_summary.csv` (candidate type `diagnostic_N128`); full matrices in `Fig8_regret_matrix.csv`. $\pi^* = 0$ with $a < 0$: TAB-trained candidates were never worse than NARROW-trained ones in the two-point comparison. Diagnostic rations were outside target-qualified selection; their costs are recorded costs, not matched-reliability costs.

S18.4 Linearisation-asymmetry records

Per returned ration, evaluation setting, root and case, `Fig9_linearisation_asymmetry.csv` gives the false passes and false fails of the linear energy row, the energy asymmetry, the linear training-event and main-event rates and their difference, from which the residual δ (Methods 2.9) follows. Table S18-6 summarises the own-setting records by case and ration type. At the higher planned intake of arms A and C the nominal ration had a large energy asymmetry but a much smaller event-rate gap (large negative residual), because most of its false-fail states also failed another row. Table S18-7 lists the strategy entries that the validation screen, which uses the linearised event, rejected although their best grid candidate met the target in the entry's own setting under the main event.

Table S18-6. Linearisation asymmetry of the energy row, each ration scored in its own training setting, all roots (ranges over rations, settings and roots; one decimal). Rows: ration-by-setting-by-root records.

Case	Ration	Rows	False passes (% of states)	False fails (% of states)	Asymmetry (pp)	$r_{\mathrm{lin}} - r_{\mathrm{main}}$ (pp)	Residual δ (pp)
R0	Nominal (M0)	15	2.0–4.2	2.2–4.2	−0.1–0.4	0.2–0.5	−0.2–0.5
R0	Diagnostic, $N = 128$	15	0.1–2.4	3.9–8.7	2.9–8.6	2.9–8.1	−0.5–1.8
R0	Diagnostic, $N = 512$	1	0.4	6.7	6.3	6.2	−0.1
A	Nominal (M0)	15	0.0–0.6	15.4–28.7	15.0–28.7	3.1–7.0	−21.7–−11.6
A	Diagnostic, $N = 128$	15	0.0	0.1–10.6	0.1–10.6	0.0–10.2	−0.4–0.0
A	Diagnostic, $N = 512$	12	0.0	0.0–3.1	0.0–3.1	0.0–3.1	−0.1–0.0
A	Diagnostic, $N = 1{,}024$	9	0.0	0.0–0.4	0.0–0.4	0.0–0.4	−0.1–0.0

Case	Ration	Rows	False passes (% of states)	False fails (% of states)	Asymmetry (pp)	$r_{\text{lin}} - r_{\text{main}}$ (pp)	Residual δ (pp)
A	Returned by a strategy entry	97	0.0	0.0–5.6	0.0–5.6	0.0–5.4	−0.3–0.0
B	Nominal (M0)	15	2.0–4.5	2.0–4.2	−0.3–0.5	0.2–0.6	−0.4–0.8
B	Diagnostic, $N = 128$	15	0.1–1.5	4.1–9.7	3.3–9.6	3.4–9.2	−1.5–0.5
B	Diagnostic, $N = 1{,}024$	1	0.5	6.3	5.8	5.6	−0.2
C	Nominal (M0)	15	0.0–0.5	14.7–28.7	14.3–28.7	3.4–7.1	−21.6–−10.9
C	Diagnostic, $N = 128$	15	0.0	0.3–8.8	0.3–8.8	0.2–8.3	−0.6–0.0
C	Diagnostic, $N = 512$	12	0.0	0.1–2.1	0.1–2.1	0.1–2.0	−0.1–0.0
C	Diagnostic, $N = 1{,}024$	10	0.0	0.1–1.9	0.1–1.9	0.1–1.9	0.0
C	Returned by a strategy entry	110	0.0	0.0–5.3	0.0–5.3	0.0–5.1	−0.3–0.0

Notes. Source: `Fig9_linearisation_asymmetry.csv` (1,860 rows: every nominal, diagnostic and returned ration of the four cases scored in all five evaluation settings on its root; 372 own-setting rows summarised here). Asymmetry = (false fails − false passes)/n; residual δ as in Methods 2.9. Returned rations include the no-DM-margin M1 sensitivity line.

Table S18-7. Strategy entries of arms A and C that returned no ration (validation screen on the linearised event not passed) although the grid candidate with the lowest test-stream main-event rate met the target in the entry's own setting (22 entries in arm A, 17 in arm C; none in R0 or arm B).

Arm	Root	Training (= evaluation) setting	Strategy entry	Target α	Grid candidates with a ration	Best candidate r^+	Its U_{95}
A	0	MO12	M2, $N = 128$	0.10	4	0.0877	0.0925
A	0	MO12	M2, $N = 512$	0.01	4	0.0039	0.0051
A	0	NARROW	M2, $N = 1{,}024$	0.01	4	0.0081	0.0097
A	0	NARROW	M2, $N = 128$	0.10	4	0.0803	0.0849
A	0	SRC12	M2, $N = 128$	0.10	3	0.0897	0.0945
A	0	SRC14	M2, $N = 128$	0.10	4	0.0888	0.0936
A	1	MO12	M2, $N = 128$	0.05	4	0.0426	0.0461
A	1	MO12	M2, $N = 512$	0.01	4	0.0062	0.0077
A	1	NARROW	M2, $N = 128$	0.05	4	0.0463	0.0499
A	1	NARROW	M2, $N = 512$	0.01	4	0.0056	0.0070
A	1	SRC12	M2, $N = 1{,}024$	0.05	1	0.0266	0.0294
A	1	SRC12	M2, $N = 128$	0.05	3	0.0349	0.0381
A	1	SRC12	M2, $N = 512$	0.05	2	0.0287	0.0316
A	1	SRC14	M2, $N = 1{,}024$	0.01	4	0.0079	0.0095
A	1	SRC14	M2, $N = 128$	0.05	4	0.0416	0.0450
A	2	MO12	M2, $N = 128$	0.05	4	0.0256	0.0284
A	2	MO12	M2, $N = 512$	0.01	4	0.0070	0.0085
A	2	NARROW	M2, $N = 128$	0.05	4	0.0440	0.0475
A	2	SRC12	M2, $N = 1{,}024$	0.05	1	0.0223	0.0249
A	2	SRC12	M2, $N = 128$	0.05	4	0.0410	0.0444
A	2	SRC12	M2, $N = 512$	0.05	2	0.0339	0.0370
A	2	SRC14	M2, $N = 1{,}024$	0.01	4	0.0069	0.0084
C	0	MO12	M2, $N = 1{,}024$	0.01	4	0.0057	0.0071
C	0	MO12	M2, $N = 128$	0.05	4	0.0335	0.0366
C	0	MO12	M2, $N = 512$	0.01	4	0.0062	0.0077
C	0	NARROW	M2, $N = 1{,}024$	0.01	4	0.0056	0.0070
C	0	SRC12	M2, $N = 128$	0.10	4	0.0663	0.0705
C	0	SRC12	M2, $N = 512$	0.05	2	0.0269	0.0297
C	1	MO12	M2, $N = 1{,}024$	0.01	4	0.0056	0.0070
C	1	MO12	M2, $N = 128$	0.05	4	0.0449	0.0485
C	1	MO12	M2, $N = 128$	0.10	4	0.0449	0.0485
C	1	NARROW	M2, $N = 1{,}024$	0.01	4	0.0034	0.0045

Arm	Root	Training (= evaluation) setting	Strategy entry	Target α	Grid candidates with a ration	Best candidate r^+	Its U_{95}
C	1	NARROW	M2, $N = 128$	0.05	4	0.0462	0.0498
C	1	NARROW	M2, $N = 128$	0.10	4	0.0462	0.0498
C	1	SRC12	M2, $N = 128$	0.05	4	0.0323	0.0354
C	1	SRC12	M2, $N = 512$	0.05	3	0.0240	0.0267
C	1	SRC14	M2, $N = 128$	0.05	4	0.0201	0.0226
C	1	SRC14	M2, $N = 128$	0.10	4	0.0201	0.0226
C	2	SRC12	M2, $N = 128$	0.05	3	0.0356	0.0388

Notes. Source: `per_method_report.csv` of each root (columns `selection_status`, `grid_min_rate_upper`, `grid_min_rate_upper_cp_upper`). The best candidate is chosen on the test stream, so this comparison is optimistic and descriptive; its cost is not a comparable cost.

Supplementary Methods: Reference problem v3 run records and identity (archive S19)

Public identity, gating, environment, host and timing facts of the three arm runs, taken from each run's `OFFICIAL_RUN.json`, `run_record.json` and `RUN_JOBS.json`, as in S8. Hashes are 12-character prefixes; reserved roots appear only by alias (`formal-k0`, `formal-k1`, `formal-k2`); the compute host appears only by alias ("compute host"), and its name and network facts are withheld. Values fixed at the freeze are filled in (S19.1) so that the run values can be checked against them. In S19.2–S19.7

S19.1 Frozen identity (known before the runs)

Field	Arm A	Arm B	Arm C
Reference problem / case	`reference_problem_v3a` / `dev_case_v3a`	`reference_problem_v3b` / `dev_case_v3b`	`reference_problem_v3c` / `dev_case_v3c`
Freeze commit	`2779a056c3e2...`	`2779a056c3e2...`	`2779a056c3e2...`
Protocol file SHA-256	`d432c9c8e22b...`	`d432c9c8e22b...`	`d432c9c8e22b...`
Freeze record SHA-256	`fdec09351435...`	`fdec09351435...`	`fdec09351435...`
Official configuration SHA-256	`16f68f54426f...`	`dee887ff5483...`	`3f2e5c7c50b6...`
Methods file SHA-256	`483eab3f7f51...`	`483eab3f7f51...`	`483eab3f7f51...`
Restricted build record / build identity (SHA-256 only)	`ab0ce47cc40f...` / `17225066858e...`	`f1f3312dec72...` / `c1292b461dba...`	`347a40a94c80...` / `359ff65ac528...`
Pin file, specification digest, SD registry, reference table	Table S16-1	Table S16-1	Table S16-1
Declared jobs	16 (15 nine-row cells + 1 diagnostics job on root 0)	16	16
Primary world / targets / workers	TAB / $\alpha = 0.05$ (primary), 0.10, 0.01 / 8	as arm A	as arm A

S19.2 Identity and gating

Field	Value (A / B / C)
Run id	`official-v3a-20260927T184000Z-8d7138b2` / `official-v3b-20260927T212028Z-a3536fbd` / `official-v3c-20260927T220316Z-5d86b41a`
Mode / output label	official / official (all three)
Run type (run record)	official (all three)
Started / completed (UTC)	A: 2026-09-27T18:40:00+00:00 / 2026-09-27T21:20:26+00:00; B: 2026-09-27T21:20:28+00:00 / 2026-09-27T22:03:14+00:00; C: 2026-09-27T22:03:16+00:00 / 2026-09-28T00:53:13+00:00
Code commit	`2779a056c3e2...` (all three) (expected: the freeze commit, S19.1)
Code manifest SHA-256	`c3d77c70c60d...` (all three)
Code manifest complete	True / True / True
Protocol SHA-256 at start	`d432c9c8e22b...` (all three)
Freeze record SHA-256 at start	`fdec09351435...` (all three)
Official configuration SHA-256 at start	`16f68f54426f...` / `dee887ff5483...` / `3f2e5c7c50b6...`
Specification digest at start	`b18224346d85...` / `df6d19bc35e4...` / `e8f6dddb167d...` (must equal S19.1)
Configuration hash	`7982a5351cb5...` / `7f0ca7caa2ac...` / `7ede6b4c9491...`
Data manifest hash	`d2bb88548268...` / `5737e97e7e76...` / `6365a00d4948...`
Run identity SHA-256	`e802e7e932d1...` / `5d4175ad2779...` / `a451fe62e22a...`
Gate at start	ok=True, status=protocol_frozen_anchored; 0 draws and 0 solves at the gate (all three)
Import-time sources	consistent (all three)

Field	Value (A / B / C)
Environment lock	matches_lock, all three (linux-x86_64-cpython3.12.3-20260925; host alias redacted)
Is synthetic	False / False / False

S19.3 Host and environment

Field	Value (A / B / C)
Host	compute host (alias; name and network facts withheld)
CPU count	32 (all three)
Workers	8 / 8 / 8 (declared: 8)
Machine / OS	x86_64 / Linux-6.17.0-1032-oem-x86_64-with-glibc2.39 (all three)
Python	3.12.3 (all three)
Packages	numpy 2.4.6, pandas 2.2.3, scipy 1.17.1, yaml 6.0.2 (all three)
Solver	HiGHS 1.12.0 (git 4f96ee8) via scipy 1.17.1 (all three)
Solver settings	dual_feasibility_tolerance=1e-07, mip_rel_gap=0.0001, presolve=True, primal_feasibility_tolerance=1e-07, residual_check_rel_tol=1e-06, time_limit_s=300.0 (all three)
Thread environment	MKL_NUM_THREADS=1, OMP_NUM_THREADS=1, OPENBLAS_NUM_THREADS=1, PYTHONDONTWRITEBYTECODE=1 (all three)
Automated tests on the compute host before the runs	1,191 passed, 4 failed, 3 skipped (each arm) (four failures of the cross-platform byte-identity class; S16.6)

S19.4 Sizes (declared; identical for the three arms)

Field	Value
opt	1,024
N ladder	128, 512, 1,024
N headline	1,024
N nesting	N < 1,024 uses the first N opt draws (prefix)
validation	20,000
test	10,000
adaptive doubling	no
evaluation settings	TAB, NARROW, SRC14, MO12, SRC12 (every ration scored in all five)

S19.5 End-of-run identity check

Check	Value (A / B / C)
Unchanged	True / True / True
Every job saw the start manifest	True / True / True
Code manifest at start = at end	True / True / True
Code manifest complete at end	True / True / True
Specification digest at start = at end	True / True / True
Freeze record at end	`fdec09351435...` (all three)
Freeze check at end (other fields)	freeze_record_sha256_at_start=`fdec09351435...`; reasons=[] (all three)

S19.6 Run-level flags

Flag	Value (A / B / C)
Pooled across roots	False / False / False (expected: no)
World fingerprints equal across jobs	True / True / True
SD registry check ok (merge; every job)	True / True / True (merge); True in every job of each arm
SD registry cells checked	48 / 56 / 56 (expected: 48 / 56 / 56)
SD registry file SHA-256	`f808110afb7b...` / `d217519df784...` / `d217519df784...` (expected: S19.1)
SD range rule	sd_range_low = SD-S2 and sd_range_high = 1 in every cell; every point inside (all three)
Merge wall time (s)	2.3393 / 0.8262 / 2.3179
Run wall time (s)	9,623 / 2,566 / 10,194
Failed jobs	0 / 0 / 0

S19.7 Roots and jobs

Field (A / B / C)	Root 0 (`formal-k0`)	Root 1 (`formal-k1`)	Root 2 (`formal-k2`)
Jobs (declared)	6	5	5
Cells (declared)	5	5	5
Endpoint rows	309 / 165 / 337	329 / 165 / 325	329 / 169 / 361
Candidate rows	2,739 / 1,375 / 2,867	2,755 / 1,375 / 2,883	2,767 / 1,375 / 2,875
Comparable sets	15 / 15 / 15	15 / 15 / 15	15 / 15 / 15
Diagnostic rows / jobs (rows/jobs)	270/1 / 145/1 / 280/1	— (none declared)	— (none declared)
Convergence verdicts	A: demonstrated 3, not demonstrated 5, no selection at either N 7; B: no selection at either N 15; C: not demonstrated 9, no selection at either N 6	A: demonstrated 3, not demonstrated 6, no selection at either N 6; B: no selection at either N 15; C: demonstrated 3, not demonstrated 6, no selection at either N 6	A: not demonstrated 9, no selection at either N 6; B: no selection at either N 15; C: demonstrated 1, not demonstrated 10, no selection at either N 4
Training-event consistency	True / True / True	True / True / True	True / True / True

Per-job exit status and wall time (16 jobs per arm: `root<k>__<cell>` for the five nine-row cells on each root and `root0__diagnostics`; source `RUN_JOBS.json`). All 48 jobs ended with exit status 0 and no job failed.

Job	Arm A: exit status / wall time (s)	Arm B	Arm C
`root0__SDH0_MAIN9`	0 / 1,755.3	0 / 1,243.5	0 / 2,005.5
`root0__SDS2_MAIN9`	0 / 4,300.3	0 / 1,221.3	0 / 4,090.1
`root0__SDSRC14_MAIN9`	0 / 4,287.6	0 / 1,280.9	0 / 4,612.2
`root0__SD12MO_MAIN9`	0 / 4,190.5	0 / 1,225.6	0 / 4,041.5
`root0__SDSRC12_MAIN9`	0 / 4,681.6	0 / 1,298.7	0 / 5,205.4
`root0__diagnostics`	0 / 218.8	0 / 155.8	0 / 200.0
`root1__SDH0_MAIN9`	0 / 1,783.0	0 / 1,261.8	0 / 2,006.4
`root1__SDS2_MAIN9`	0 / 3,992.4	0 / 1,227.2	0 / 3,700.5
`root1__SDSRC14_MAIN9`	0 / 4,642.9	0 / 1,552.8	0 / 4,573.6
`root1__SD12MO_MAIN9`	0 / 3,670.9	0 / 1,237.6	0 / 3,774.3
`root1__SDSRC12_MAIN9`	0 / 4,799.9	0 / 1,265.2	0 / 5,380.8
`root2__SDH0_MAIN9`	0 / 1,697.2	0 / 1,222.1	0 / 1,853.6
`root2__SDS2_MAIN9`	0 / 3,536.1	0 / 1,245.2	0 / 4,155.8
`root2__SDSRC14_MAIN9`	0 / 4,377.4	0 / 1,303.7	0 / 4,515.5
`root2__SD12MO_MAIN9`	0 / 3,795.6	0 / 1,216.3	0 / 4,063.6
`root2__SDSRC12_MAIN9`	0 / 4,941.6	0 / 1,253.5	0 / 5,420.5

S19.8 Output files

SHA-256 prefixes of the run-level files (`OFFICIAL_RUN.json`, `run_record.json`, `RUN_JOBS.json`, `SHA256SUMS`) and, per root, of `endpoint_ablation.csv`, `per_method_report.csv`, `comparable_sets.json`, `convergence_report.json`, `diagnostics_summary.json` and `training_event_consistency.json`. Redacted copies of `OFFICIAL_RUN.json` (host, command line and authoriser replaced by placeholders) do not reproduce the archived digests; the digests refer to the archived originals, which are kept in the study archive. The per-root prefixes equal the entries of each run's `SHA256SUMS` (109 entries per arm), whose own SHA-256 equals the value recorded in `MERGE_DONE`. The diagnostics summary exists on root 0 only. The output check `scripts/verify_official_outputs.py` passed all six checks for each arm (SHA manifests, run record, no pooled n, diagnostics not members, no reserved root value in public files, training-event consistency).

File	Arm A	Arm B	Arm C
`OFFICIAL_RUN.json`	`be5f07ecc97d...`	`e88cec46265d...`	`82dbb03ccc94...`
`run_record.json`	`8cfa0d3b7fac...`	`c6c6b51beab9...`	`2ceb5f7beba6...`
`RUN_JOBS.json`	`67d0a0f03cf9...`	`0c3b82416d7e...`	`2a68df391c3f...`
`SHA256SUMS`	`cf8748ab4989...`	`4ee7ace2a773...`	`a4bdd3985f56...`
`root0/endpoint_ablation.csv`	`71e706743db8...`	`4303ad1dabc7...`	`e407069d8048...`
`root0/per_method_report.csv`	`fd29e514be8c...`	`4f1446f4aff1...`	`d2ef79560c9f...`
`root0/comparable_sets.json`	`5a9329351421...`	`4f6251459b47...`	`02103a5e6a7a...`
`root0/convergence_report.json`	`60bf7b69a4f6...`	`d53bf37024cf...`	`b815fb2a9ecf...`
`root0/diagnostics_summary.json`	`cadf45c792a5...`	`c33c297d86a0...`	`faeeff8a0d63...`
`root0/training_event_consistency.json`	`9f6881c60652...`	`c1e0924c886d...`	`415bee181c61...`
`root1/endpoint_ablation.csv`	`b451b5cde115...`	`b06ac8c5611c...`	`68a326992d2a...`
`root1/per_method_report.csv`	`75779cf7e703...`	`6e4fa7ba7788...`	`0b57d1822733...`
`root1/comparable_sets.json`	`0951aa0b9654...`	`85c4a97cd990...`	`9f5ecdbd7b4e...`
`root1/convergence_report.json`	`8244a2c394c3...`	`7fe95c36f4c9...`	`2c5eabc98b93...`
`root1/training_event_consistency.json`	`94e3e9fe7281...`	`e0060603aa6e...`	`187ed1de25ed...`
`root2/endpoint_ablation.csv`	`029b87661ad7...`	`5fc5ac93921c...`	`999fc084b34c...`
`root2/per_method_report.csv`	`833e16e95f99...`	`7bfa6c2ccd96...`	`e537ec16bd68...`
`root2/comparable_sets.json`	`3d7416b6f7b0...`	`9ea142cb7d18...`	`071d8d6dfa1e...`
`root2/convergence_report.json`	`2ff77ae445f6...`	`cc15d529f06c...`	`e5196336962f...`
`root2/training_event_consistency.json`	`3c3928895c4f...`	`30979ef83eb5...`	`94067272cb7b...`

Supplementary Methods: Internal identifiers used in the archived tables (archive S20)

The main text uses descriptive names; archived CSVs, run records and several sections of this supplement use the internal identifiers below.

Identifier	Name used in the main text (meaning)
Cases, runs and training arms	
`reference_problem_v2` (`dev_case_v1` + premise row); run `official-20260926T162636Z-e3665bbe`	reference case (R0)
`reference_problem_v3a` / `v3b` / `v3c` (`dev_case_v3a` / `v3b` / `v3c`)	arms A / B / C
`DMI_plus10`	planned intake +10% (registered sensitivity scenario)
`cottonseed_whole_linted` (`NRC16F56`)	whole linted cottonseed
`SRC-P-USDA-AMS-COTTON`	cottonseed scenario price source
MAIN9	primary nine-row training arm (the nine-constraint main event as training target)
FULL11	historical eleven-row training target
PART6P5	historical six-row training target
DIAG-T45	starch-maxima removal diagnostic
DIAG-E1	energy-only diagnostic (chance constraint on the energy row alone)
M0; M1; M2; M3a; M3b	nominal; uniform relative safety margin; joint chance-constrained sample approximation; box-robust; budgeted robust (Table 3)
M1-T	directed safety margin (one margin per row group; kept as a defined label)
`min=...|prot=...|en=...|carb=...`	M1-T pattern: margins of the minerals, protein, energy and carbohydrate groups
Variability settings, cells and roots	
SD-H0 / SD-SRC12 / SD-SRC14 / SD-12MO / SD-S2	TAB / SRC12 / SRC14 / MO12 / NARROW
`SDH0_MAIN9`, `SDS2_MAIN9`, ...	nine-row training arm in the named setting (one cell)
`TN_MM` / `BETA_MM`	truncated normal / scaled beta, both moment-matched
`formal-k0` / `formal-k1` / `formal-k2`	reserved roots 0 / 1 / 2 (first, second, third reserved root)
Rows	
`PN-NEL-FIXEDDMI`	energy (NEL) row: linear row in formulation; reference chain in evaluation
`PN-CP-SUP`	CP supply
`PN-CA-ABS` / `PN-P-ABS`	absorbed calcium / absorbed phosphorus
`PN-T1` / `PN-T2` / `PN-T3`	forage-NDF minimum / total-NDF minimum / forage-NDF-dependent NDF minimum
`PN-T4` / `PN-T5`	starch maximum / forage-NDF-dependent starch maximum
`PN-CP-HI`, `SH-PLAN-CP-HI`	project CP upper limit (plan-level row in formulation)
`PN-EE-HI`, `SH-PLAN-EE-HI`	ether-extract limit (plan-level row in formulation; companion outcome)
`SH-DM-PLAN`	planned DM supply
`SH-INCL-LIMESTONE-MAX` / `SH-INCL-DCP-MAX` / `SH-INCL-SOYHULLS-MAX`	inclusion caps (limestone, dicalcium phosphate, soybean hulls)
`SH-INCL-COTTONSEED-MAX`	cottonseed cap ($\leq$ 10% of planned DM)
`SH-PLAN-T51-DGC-SHARE`	starch-source premise
Candidates, events and CSV columns	
diagnostic frontier ration, `diagnostic_N128`	128-state diagnostic ration
`cheapest_qualified_a0.05`	cheapest qualifying ration at $\alpha = 0.05$
`main_reference_plan_domain` (`_rate_upper`)	main event, plan-level domain reading (upper rate r^+)
`main9_training_event`	linear training event
`rate_lower` / `rate_upper`	lower and upper rate r^-, r^+ (unresolved states excluded / counted as failures)
`cp_upper`	one-sided 95% Clopper–Pearson upper bound (U_{95})
`regret_pp`; `max_regret_pp`; `is_minimax`	regret $\rho(t, e)$; maximum regret $\bar{\rho}(t)$; minimax training setting
`min_over_training`	lowest r^+ over training settings in an evaluation setting
`a_pp`, `b_pp`, `pi_star`	a, b and the break-even prior π^* (S18.1)

Identifier	Name used in the main text (meaning)
`cost_premium_pct`	cost premium over the nominal ration of the same case (%)
`extra_cost_premium_pct_TAB_minus_t`; `max_regret_avoided_pp`; `regret_pp_per_pct_cost`	$\Delta c = c_{\mathrm{TAB}} - c_t$; $\Delta\bar{\rho} = \bar{\rho}(t) - \bar{\rho}(\mathrm{TAB})$; cost-weighted ratio (S18.1)

Supplementary Methods: Derived-input release: reserved, not released (archive S21)

Interpretation and access. In these materials, "registered" denotes a study protocol recorded and fixed before its corresponding run; no public preregistration identifier is supplied. Later stages were designed after earlier results had been inspected. The supplement contains aggregate evidence and workflow records, not a complete independently executable release. Verification records below describe the authors' workflow unless another scope is explicitly stated.

This number was reserved for the derived-input release proposed at the time of this archive, which had not then been approved or incorporated. In that version, exact animal inputs, requirement amounts, ration vectors, composition means and SDs, SD ratios and absolute planned DM were withheld, and the additional aggregate results below did not change that decision. The current article has since released the fixed-rule rations at one decimal and the equation-derived requirement amounts with an identifiability check (current S38); composition means and SDs, printed table targets and bounds and full-precision candidate rations remain withheld. Source pointers in S1–S5 and S16 remain the input-access route; full independent re-execution is not possible from the aggregate supplementary files alone.

Supplementary Methods: First research extension: design, results and boundaries (archive S22)

S22.1 Relationship to the original comparison

The original comparison in S1–S20 is retained as a completed, frozen experiment. Its empty comparable sets mean that the registered formulation and selection procedures returned no target-qualified ration; they do not establish population infeasibility over all continuous rations or over later decision classes. The first extension was designed after inspecting those results, then registered and frozen before its own formal streams. It used new simulation namespaces and retained the original complete reference event, requirements and inclusion/domain rules. No test-selected ration replaced an original returned ration.

Let θ denote DM-basis nutrient composition and d the ingredient DM fractions. For an executed as-fed ration q, the common loss throughout both extensions is

$$L(q, \theta, d) = \mathbf{1}\{\text{main reference event is violated or unknown}\}.$$

The event uses the whole-diet reference energy chain and the original nine member rows under the original plan-level Table 5-1 premise. Used missing or undefined values are not removed from the denominator or converted into successes. Structural and nominal plan-domain conditions remain outside the risk budget. Supply is still assumed eaten, and the retained fixed-fMCP energy chain is not the complete NASEM model.

S22.2 Completed designs and statistical units

All 588 formal computational tasks completed; these produce 412 merged experimental units, not 588 statistical replicates. The two 180-batch certificate designs are included in the task count. Solver infeasibility, timeout, empty choice and weak bounds were retained as results. Table S22-1 distinguishes the units.

Table S22-1. First-extension designs. A cell identifies the specified case, variability setting and independent root, with any additional design factors shown.

Design	Tasks	Merged units	Scope
PCA information hierarchy and full-state oracle	12	60	Four cases, five settings, three roots
Fixed-marginal PSD factor dependence sensitivity	12	60	Ten dependence scenarios per cell
Cottonseed price and reference-aligned candidate generation	60	60	Registered prices, screens and targets
Six-point planned-DM frontier	12	12	R0/B, TAB/NARROW, three roots
Nonenergy continuous-space lower certificate	180	20	Nine batches per case/setting
Nutrient decision-loss trees	60	60	Two frozen depths per cell
Hidden-DM sensitivity recourse	60	60	Nominal-margin and delta rules
Author-source copula sensitivity	12	60	C0, C1 and C2 per cell
Whole-reference-energy lower certificate	180	20	Independent certificate streams
Total	588	412	Sum of the nine designs

PCA, tree and recourse used 4,000 training and 20,000 independent test states per policy cell. The actually frozen common libraries were identical across their 180 cells: 8, 332, 1 and 353 legal actions in R0, A, B and C, respectively. Equality was verified for this experiment rather than assumed generally. A constant comparator minimized training reference loss, with cost and ration-hash ties; it was not necessarily the original nominal least-cost ration. PCA, tree and recourse used separate new roots, so between-protocol differences are descriptive. Within each protocol, policies and their own constant comparator used the same test states.

The registered tree/recourse one-sided Clopper–Pearson risk upper endpoint allocated 0.05/10,000 per statistic [1]. Their paired-difference intervals combined exact binomial intervals for the two discordant outcomes with the protocol's separate allocation. The earlier PCA design retained its own two-sided risk and empirical-Bernstein paired-reduction procedures. Different risk, paired-difference and protocol families have separate error budgets; they are not advertised as one combined 95% coverage statement. Roots were reported separately, not pooled. Statistical bounds concern conditional simulation uncertainty under the declared model, not uncertainty in the source moments or field performance.

S22.3 Nutrient-only decisions with hidden actual DM

Operational policies observed ideal, noiseless current nutrient composition θ; actual test DM d was hidden. The derived fixed-intake energy column was excluded from observations and recalculated for hypothetical states. Every executed action retained its case's original planned-DM equality and plan-domain constraints; in A and C this is the original +10% arm, not an additional intake increase.

The PCA policies used frozen two-by-two and four-by-four bins. The decision-loss trees used ingredient CP, NDF, starch and EE features, excluding DM and derived energy. Candidate splits were the training quartiles, minimum leaf size was 100, and leaf actions minimized training reference loss. Depths 2 and 4 were both frozen and reported. The greedy tree is not globally optimal. A variance-based partition need not separate states requiring different actions, so explained variance does not establish decision value.

Recourse held observed θ fixed and reconstructed hypothetical DM using training-only means and sample SDs. Its nominal-margin rule maximized the smallest scaled signed reference margin. Its delta rule used exact source-ratio DM derivatives and the analytically differentiated whole-reference energy. For row margin m_j, training moments μ_i, σ_i, and gradient $g_{ji} = \partial m_j / \partial d_i$ at μ,

its ranking proxy was

$$\widetilde{R}(q \mid \theta) = \sum_j \Phi\left(\frac{-t_j - m_j(q,\theta,\mu)}{\sqrt{\sum_i g_{ji}^2 \sigma_i^2}}\right).$$

The source C0 DM model is independent. Zero-variance rows used the original strict verdict, and the sum was not clipped to one. This normal-tail sum ignores nonlinear tail calibration and dependence between row failures: it is an action-ranking proxy, not a proven probability or bound. Invalid used inputs invalidated a score; known unused directions were masked before multiplication. An all-invalid score invoked the frozen constant fallback and was counted. The subsequent original reference evaluation alone determines test reliability.

Table S22-2. First-extension TAB/C0 delta recourse, each root with 20,000 independent test states. The paired interval is delta minus its own constant comparator, in percentage points, under that protocol's paired family.

Case/root	Constant (%)	Delta (%)	Failures	Adjusted upper (%)	Paired interval (pp)
A/0	12.895	4.165	833	4.822487	[-9.8609, -7.5850]
A/1	12.695	4.015	803	4.661743	[-9.7861, -7.5597]
A/2	12.810	3.865	773	4.500766	[-10.0339, -7.8414]
C/0	10.185	3.745	749	4.371810	[-7.4585, -5.4109]
C/1	10.370	3.380	676	3.978526	[-8.0371, -5.9315]
C/2	10.165	3.500	700	4.108006	[-7.7066, -5.6125]

All six delta upper endpoints passed the original 5% target; exact integer/rational binomial tails independently confirmed the decisions. Unknown and all-invalid fallback counts were zero for these six rules. Nominal-margin recourse improved risk but was weaker: A had 10.280–10.685% observed loss and did not qualify at 10%; C had 7.925–8.380% and qualified at 10% on all three roots. Depth-4 trees improved A/C TAB against their own constants, but only two C roots qualified at 10%; the other C upper endpoint was 10.059499%. PCA showed no stable improvement; one C four-by-four rule was significantly worse. R0/B did not attain the targets; B's single-action original library prevented any change of action. All settings, roots and both tree depths remain in the aggregate tables.

S22.4 Static reference alignment, prices and planned-DM witnesses

The price/reference and planned-DM experiments retained 1,920 registered choice slots across 72 slices: 848 returned a ration, 1,072 were empty, and 616 choices qualified at their own targets under a conservative common risk-family allocation of 0.05/10,000. Each selected fixed ration had 10,000 independent test states; no roots were pooled. Candidate origins included historical actions, a price grid and reference-aligned successive linear programming (SLP), sometimes jointly. SLP is a local method, not a global infeasibility test. Selection used validation, not test success.

In A/TAB, the baseline-price reference-screen rule at 10% qualified on all three roots, with adjusted upper endpoints 9.35574, 8.69347 and 9.03556%. This is a new fixed-ration witness in a new experiment, distinct from the conditional policies and the old empty comparison. C/TAB did not show the same all-root result: the low and baseline price rules passed on two roots each and the high-price rule on one. The nine selected C actions in this comparison had exclusive SLP provenance. A fixed action's nutritional risk cannot change with price; only its cost and the selected action can change. Prices were converted from US short tons as-fed with the original unit convention. Finite-candidate affine cost envelopes were checked over whole segments, including exact rational endpoint comparisons, within the frozen tie guard; no global continuous cost-optimality claim follows.

The planned-DM frontier used six registered increases from the R0/B original setpoint: 0, 5, 10, 15, 20 and 25%, holding animal requirements fixed. At the 5% target, the validation-selected

smallest grid point passed test on only 2/3, 1/3, 1/3 and 2/3 roots for R0/TAB, R0/NARROW, B/TAB and B/NARROW, respectively. These minimum-grid choices were not reselected after test. Separately, simultaneous upper endpoints of already frozen and tested choices supply existence witnesses: R0/TAB at +20% and +25%, B/TAB at +25%, R0/NARROW from +10%, and B/NARROW from +5% had a 5%-qualified action in every root. At +15%, both NARROW cases also had three 1% witnesses. These observations neither identify the continuous minimum intake nor establish that the reference animal can consume these supplies. They do not establish monotonic grid feasibility.

S22.5 Dependence sensitivities and continuous-space certificates

The registered PSD factor sensitivity and later author-source copulas preserved declared marginal targets, used all cases/settings/roots and retained both increases and decreases in risk. Factor-sensitivity descriptive extrema included a +7.18 percentage-point change in one C/SRC12 root and -19.445 points in a different C/SRC14 root; they are not a universal dependence direction. C1/C2 are specified below in S26. Their first-extension results concern fixed actions, not qualification of the C0-trained conditional rules under those models.

Two independent certificate designs addressed the minimum population risk of a *single fixed* action over the full common structural continuous action box. Each used four cases, five settings and 2,304 states per final cell (three fresh roots, three 256-state batches per root), with one family-20 lower inversion. Necessary nonenergy rows and, in the second design, state-aware upper relaxations of the whole reference energy supplied LP relaxations. Feasible LP dual multipliers were checked using exact rational arithmetic with residual compensation; a valid integer lower count was obtained from each batch. The fixed-DMI linear energy proxy was excluded, rather than substituted for the reference energy.

If the aggregate certified count $M(S)$ satisfies

$$M(S) \leq \sum_{s \in S} L(q, s) \quad \text{for every structurally legal fixed } q,$$

then it is stochastically dominated by the binomial failure count of any fixed population comparator. Its count need not itself be binomial. Monotone exact-binomial lower inversion therefore gives a conservative lower bound for the population infimum, using a population minimizer or a sequence approaching the infimum; a data-selected comparator is not required. This argument differs from certifying a returned action with an upper endpoint [2, 3].

All 20 nonenergy lower endpoints were zero. The whole-energy design gave positive lower endpoints in 8/20 cells, with maximum 0.12217164%; all were below the smallest registered 1% target. Thus neither design excluded any registered target. A zero lower endpoint is uninformative, not a feasible-ration proof. The two independent count sets were not added. The second frozen raw summary inherited a historical "nonenergy" scope string; its actual relaxation includes whole-reference energy, as clarified here and in the derived aggregate table. Exact LP/tail checks do not constitute a formal verification of all floating-point or transcendental source calculations.

S22.6 Finite-library information boundary and qualification

For the original common library Q, $H_Q(s) = \min_{q \in Q} L(q, s)$ is one if every library action fails. Every deterministic or randomized policy taking values in Q, under any information pattern, has loss at least H_Q pointwise. A canonical-confirmed count no larger than the true all-action-fail count gives a conservative binomial lower bound for $\mathrm{E}H_Q$; the denominator is the full independent test sample, not the selected subset. On the separate first-extension TAB streams, all three A roots excluded 2% and all three C roots excluded 1% *within this original library.* The original

evaluator confirmed every action on the qualifying all-fail subset, comprising 936,553 action–state checks. The result does not constrain later new actions, blends or the continuous space and is compatible with improvement after expanding Q.

The first extension independently checked all completed artifacts and preserved source/code/protocol identities. The parent strict-domain reconstruction covered 1,116 registered streams and 15,655,296 state positions with no unmarked strict-chain gap. All 1,080 available saved stream hashes matched; 36 planned-DM streams lacked a saved fingerprint and could not receive that byte-comparison check. Tree/recourse training fingerprints matched in all 60 cells of each family, but their runners did not save test draw fingerprints. Their saved actual loss evidence and original-evaluator coverage on every assigned test state were checked independently; reconstructed test draws are not advertised as comparisons with nonexistent saved fingerprints. The 60 tree launches that failed at environment setup occurred before sampling or policy freezing; they were retained separately from the corrected formal launches, which completed without dispatcher exit failures.

Supplementary Methods: Second research extension: frozen menus and conditional joint-risk decisions (archive S23)

S23.1 New protocol and fixed information pattern

The second extension was designed after the first extension's findings. Before its formal draws it fixed four cases, two original variability settings (TAB and NARROW), three dependence scenarios (C0/C1/C2) and three new independent roots: 72 cells, each with 4,000 training and 20,000 test states. Code, sources, candidate input identities, integration design and all decision rules were frozen. The new roots excluded earlier formal and development roots. There were 288,000 training and 1,440,000 test state positions; these counts span distinct learned rules and are not pooled certificates for one policy. The protocol fingerprint is retained in the aggregate provenance record.

The operational observation was the same ideal complete θ, with actual test d hidden. Known declared marginals and a declared latent-Gaussian dependence model defined hypothetical DM laws. Every action retained the original case's nutritional requirements, planned DM, inclusion bounds and plan-domain premise. In A/C the original +10% planned-DM equality was maintained. No demand reduction, actual-state ration renormalization, extra planned intake or distribution clipping was introduced. The model dependence and exact observation assumptions are not claims of identified farm distributions or usable assays.

S23.2 Training-only complementary action construction

The input pool combined the original common library with *all* 15 previous raw price-candidate files per case, without filtering old roots, settings or origins by their validation/test outcomes. Original ingredient axes and coefficient identities were checked, then every action was checked against the original structural and nominal plan-domain evaluator. Legal deduplicated pools contained 8, 515, 2 and 894 actions for R0, A, B and C, respectively.

On the 4,000 training states, a greedy procedure added the action with greatest incremental success coverage, using frozen deterministic ties. Finite-set maximum coverage is the established submodular objective [4]; its classical approximation result is not a population-risk or whole-pipeline guarantee here. The procedure was capped at 64 actions. The first at most eight complementary actions supplied pairs for physical mixtures

$$q_{ab,\lambda} = \lambda q_a + (1-\lambda) q_b, \qquad \lambda \in \{0.25, 0.50, 0.75\}.$$

All mixture actions were checked and evaluated through the original reference chain; success was not inferred from their endpoints. Nonlinear energy means the pass set need not be convex. A second training-only compression produced each final menu Q, actually containing 1–25 actions. All unselected mixtures, training labels, selection gains and lineage were archived so each greedy step could be independently replayed. The reference-loss objective, rather than a prediction-error objective, follows the established distinction between prediction and decision quality [5]. Neither maximum-coverage greediness nor menu compression guarantees minimum population risk.

The uniform increase in planned intake within the original A/C arms was preserved by every action and mixture. Maximum relative plan-equality rounding residual over the executed cases was 2.29×10^{-15}. This numerical check addresses the frozen equality, not biological intake prediction. Changing feed allocation and complementary action coverage, rather than further increasing supply, was the new intervention.

S23.3 The seven prespecified rules

Table S23-1. Operational rules. All test decisions exclude actual DM; hypothetical nodes are separate from test DM.

Rule	Frozen definition
`old_static`	Training-minimum-reference-loss constant in the original common library.
`old_delta`	Original-library delta ranking with this cell's training-only DM moments.
`expanded_static`	Training-minimum-reference-loss constant in the new menu; no nutrient observation is needed at execution.
`expanded_delta`	The same delta ranking on the new menu.
`joint64`	Complete joint-loss mean on 64 fixed Sobol hypothetical DM nodes from the declared conditional law; chooses within the frozen delta shortlist.
`rb16`	One physical DM integrated by its conditional CDF over all joint-pass intervals, averaged over 16 nodes for the other DM coordinates; uses the same shortlist.
`joint64_prior`	Joint64 with the same design and menu but the marginal DM prior, withholding the nutrient-to-DM update.

For each observation, the shortlist began with at most eight lowest delta-score actions; if the frozen training-static fallback was absent, it replaced the last action, leaving at most eight total. Scoring and ties used the declared original feed costs and coefficient identities. The 16-node design was the fixed prefix of the conditional 64-node design; nodes were fixed before outer test sampling. The rules were all retained, not selected from seven by the test outcome. A full-state oracle was not a deployable eighth rule.

S23.4 Conditional DM reconstruction

Each observed nutrient value was transformed by its declared marginal CDF to a latent normal coordinate. Partition the registered latent correlation matrix as

$$\begin{pmatrix} R_{DD} & R_{D\theta} \\ R_{\theta D} & R_{\theta\theta} \end{pmatrix}.$$

The declared Gaussian model gives

$$Z_D \mid z_\theta \sim N\left(R_{D\theta}R_{\theta\theta}^{-1}z_\theta, R_{DD} - R_{D\theta}R_{\theta\theta}^{-1}R_{\theta D}\right).$$

Hypothetical latent nodes were transformed through the original fitted physical DM inverse CDFs. Point-valued components stayed point valued. Saved execution-platform fits and latent matrices were used in independent replay; there was no refitting from test outcomes.

The actual C0/C1/C2 conditional DM covariance matrices were diagonal, checked exactly for these models. Thus S24 may use each pivot's conditional CDF given nutrients only. A non-diagonal future model would require its CDF conditional also on the other DM coordinates; treating such coordinates as independent would change the target integral. Endpoint/extreme-tail guards were explicitly counted. Their numerical implementation does not identify a farm posterior or incorporate assay noise.

Supplementary Methods: Whole-reference joint-risk preintegration (archive S24)

S24.1 Linear row intervals with unchanged tolerance

Fix q, observed θ and all DM fractions except $x = d_h$. Let total supplied DM be $A(x) = \sum_i q_i d_i$, to distinguish it from the DM vector. For a compiled nonenergy row, its original numerator $S_j(x)$, including any as-fed terms, is affine in x. Write its raw bound b_j, tolerance $t_j \geq 0$, and sense $s_j = +1$ for a minimum or -1 for a maximum. Supply rows pass if

$$s_j(S_j(x) - b_j) + t_j \geq 0.$$

Concentration rows, on $A(x) > 0$, pass if

$$s_j S_j(x) + (t_j - s_j b_j)A(x) \geq 0.$$

The tolerance is multiplied by $A(x)$; omitting it changes the event. Declared unit factors cancel after the original bound and tolerance have been placed in common units. Intersecting these affine half-lines with physical DM support and positive total DM gives an interval J, possibly empty. Known unused ingredients are masked before arithmetic. Used missing or source-undefined composition yields failure under the original verdict rather than an invented pass interval. A source ROM diagnostic anomaly is not automatically an undefined energy verdict: the canonical reference flags remain authoritative.

S24.2 Quartic factorization of the retained energy chain

At fixed composition, let the following totals be affine in x: total DM A, starch S, NDF N, fatty-acid proxy F, base-digestible NDF B, base-digestible starch H, and the remaining DM-proportional energy/loss term W. Their coefficients retain the original nutrient units, digestibility factors, fixed-fMCP convention and source constants (S3). Without printing any restricted numerical inputs, the original reference chain can be written

$$T = B + aN - bNS/A - cNA, \qquad U = H + a_s S - c_s SA,$$

$$E = \kappa\{(g_N - g_2/A)T + g_S U + W + g_1 F/A + K_0\}.$$

Here the named constants simply collect the original starch/NDF adjustments, gross/gas energy coefficients, unit conversion and protein-output constant; they are not fitted to improve test reliability. For original energy requirement R and tolerance t_E, multiplying $E - R + t_E$ by $A^2 > 0$ gives

$$\begin{aligned} P(x) = \kappa\big[&(g_N A^2 - g_2 A)(B + aN - cNA) \\ &- b(g_N A - g_2)NS + g_S A^2(H + a_s S - c_s SA) \\ &+ A^2 W + g_1 FA + K_0 A^2\big] - (R - t_E)A^2. \end{aligned}$$

Every total has degree at most one, so P has degree at most four. On the original defined domain with $A > 0$, energy passes if and only if $P(x) \geq 0$. This is an identity for the whole retained reference chain, not the fixed-intake linear formulation proxy. The joint-pass set is

$$\mathcal{C} = J \cap \{x : P(x) \geq 0\}.$$

All real roots within the allowed interval partition it into sign intervals. The set may be disconnected; no single-root or monotonic-energy assumption was made. Degenerate constant polynomials and no-continuous-pivot cases retain their original direct verdicts. The continuous stochastic pivot was chosen by the frozen maximum q_i times prior DM SD rule, not by actual test DM. Numerical root ambiguity invoked an explicit original-indicator fallback on hypothetical nodes.

The first extension also differentiated this same identity analytically. At positive d_i, $\partial E/\partial d_i = (q_i/d_i)\partial E/\partial q_i$, with known unused DM exposure set to zero before arithmetic. Zero DM requires direct treatment rather than this division. Qualification covered 64,000 factory action–states and 40 artificial boundaries; definition masks matched the original evaluator, and the maximum checked DM-derivative error was 1.05×10^{-9}. This numerical qualification is not a full-space proof of floating-point equivalence.

As a separate algebraic sensitivity, scaling a fixed-composition ration along tq yields $E(tq) = C + Lt - Qt^2$, $t > 0$. Specifically, $T = tT_1 - t^2T_2$ and $U = tU_1 - t^2U_2$, where $T_2 = cN_0A_0$, $U_2 = c_sS_0A_0$, and $Q = \kappa(g_NT_2 + g_SU_2) \geq 0$ under the declared nonnegative-composition/constants assumptions. Thus increasing uniform supply need not indefinitely increase this retained mathematical energy output. This ray generally leaves the original planned-DM equality; its algebraic turning point is not an optimized ration, an intake recommendation or evidence of an animal's achievable intake.

S24.3 Exact ideal integral and what the implementation does not certify

For a continuous pivot conditional distribution with CDF $F_h(\cdot \mid \theta, d_{-h})$, the ideal joint-pass probability, holding other DM fixed, is

$$p_{\text{pass}}(q, \theta, d_{-h}) = \sum_m \{F_h(b_m \mid \theta, d_{-h}) - F_h(a_m \mid \theta, d_{-h})\}, \quad \mathcal{C} = \bigcup_m [a_m, b_m].$$

Continuous endpoints carry no mass; atoms instead require the original endpoint verdict or direct evaluation. For the verified diagonal conditional models, $F_h(\cdot \mid \theta, d_{-h}) = F_h(\cdot \mid \theta)$. Let $Z = 1 - p_{\text{pass}}$. The tower property and total variance give

$$\mathrm{E}[Z \mid \theta] = \mathrm{E}[L \mid \theta], \qquad \mathrm{Var}(Z \mid \theta) \leq \mathrm{Var}(L \mid \theta).$$

This is the established Rao–Blackwell argument for the ideal conditional integral and ordinary IID single-node variance. Conditional quasi-Monte Carlo and preintegration motivate its implementation [6, 7]. It preserves all row dependence through the complete joint event. It does not prove that the actual fixed 16-node design outperforms a fixed 64-node design, that action rankings are correct, or that the polynomial root solver is formally certified over the entire domain. Active constraints, repeated roots and merging intervals can leave non-smooth integrands; integrating one coordinate does not establish a smooth high-order quasi-Monte Carlo rate [8].

If all numerical action-risk errors on a shortlist are at most ε, its minimizing rule has conditional regret at most 2ε relative to the best shortlisted action. Relative to the full menu, a further nonnegative screening gap must be added. Neither quantity was certified here. A Sobol net's dependent nodes cannot be treated as IID Bernoulli observations for an inner Clopper–Pearson bound. The 64/16 scores are only ranking devices. The independent outer test, after menu and rule freezing, is the source of the reported policy-risk bounds.

Supplementary Methods: Second-extension results, finite-menu bounds and independent checks (archive S25)

S25.1 Static 10% and nutrient-conditional 5%/2% results

The expanded static rule selects one ration using training only and requires no nutrient measurement at execution. In TAB, A and C each qualified at 10% in all three dependence scenarios and all three roots (9/9 cells per case); their largest adjusted upper endpoints were 8.6946% and 6.1597%, respectively. They did not pass 5% in every cell. This is distinct from the ideal-nutrient conditional result.

The prespecified `expanded_delta` rule passed 5% in every A/TAB cell and 2% in every C/TAB cell, across C0/C1/C2 and the three separate roots. Table S25-1 gives observed risk and adjusted upper endpoints. None of these results asserts an all-cell 1% success. Table S25-2 gives the paired C0 comparison against `old_delta`, reconstructed and trained within the same second-extension cell; it is not a paired comparison with the first-extension roots.

Table S25-1. Nutrient-conditional expanded delta in TAB. Each three-number entry is root 0/1/2; all risks use 20,000 test states per root, unknown included.

Case	Dependence model	Observed risk (%)	Adjusted upper (%)
A	C0	2.890 / 2.775 / 2.615	3.4477 / 3.3226 / 3.1482
A	C1	2.615 / 2.405 / 2.530	3.1482 / 2.9184 / 3.0553
A	C2	2.060 / 2.460 / 2.045	2.5389 / 2.9787 / 2.5223
C	C0	0.930 / 1.195 / 1.125	1.2662 / 1.5705 / 1.4906
C	C1	1.035 / 1.010 / 1.065	1.3874 / 1.3586 / 1.4219
C	C2	0.940 / 1.130 / 1.145	1.2778 / 1.4963 / 1.5135

Table S25-2. Same-state C0/TAB comparison. The adjusted paired interval is expanded delta minus old delta in percentage points.

Case/root	Old delta (%)	Expanded delta (%)	Paired interval (pp)
A/0	3.880	2.890	[-1.7459, -0.2323]
A/1	3.965	2.775	[-1.9124, -0.4655]
A/2	3.885	2.615	[-1.9390, -0.5986]
C/0	3.555	0.930	[-3.3238, -1.9213]
C/1	3.720	1.195	[-3.3028, -1.7428]
C/2	3.545	1.125	[-3.1915, -1.6442]

On these C0/TAB tests, A selected original-library, earlier raw-price and new-mixture actions in 31.47, 28.91 and 39.62% of decisions; C selected them in 1.08, 73.705 and 25.215%. This verifies that new actions were used. Without a source-by-source ablation, it does not assign the total reliability gain causally to each source. Same-state mean feed-cost ratios of expanded to old delta were 0.9794–0.9845 in A and 0.9812–0.9833 in C, secondary descriptive values excluding assays, implementation, transport and other costs.

S25.2 Full positive and negative scope

Table S25-3. Observed loss ranges across the nine dependence/root cells of each case/setting, in percent. The ranges are descriptive, not pooled intervals or between-setting paired contrasts. Complete tables retain all seven rules.

Case	Setting	Old delta	Expanded delta	Joint64	RB16
R0	TAB	89.315–92.745	89.305–92.705	89.225–92.720	89.130–92.565

Case	Setting	Old delta	Expanded delta	Joint64	RB16
R0	NARROW	79.605–81.430	79.525–81.420	79.620–81.575	79.345–81.350
A	TAB	3.565–4.305	2.045–2.890	2.365–3.195	1.945–2.855
A	NARROW	0.000–0.000	0.000–0.120	0.000–0.120	0.000–0.120
B	TAB	94.860–97.370	92.090–95.485	91.460–95.205	91.350–95.125
B	NARROW	96.215–97.835	92.635–95.390	92.270–95.075	92.245–94.980
C	TAB	3.045–3.720	0.930–1.195	1.070–1.550	0.880–1.235
C	NARROW	0.000–0.000	0.000–0.035	0.000–0.035	0.000–0.035

R0 obtained no reliability breakthrough. B showed significant relative improvements but still failed in about 91–95% of states under the joint-integration rules, far above even 20%. In A/C NARROW, old delta already had zero observed failures and some expanded rules introduced small failure counts; there is no universal improvement claim. Zero observed failures is not zero population risk.

On the same expanded menu, joint64 improved significantly over expanded delta in B on 8/9 TAB and 2/9 NARROW cells, while rb16 improved on 9/9 and 6/9. In the 18 A/C TAB cells, joint64 was significantly worse on five and rb16 showed no adjusted significant improvement in any. The C2 conditional update also supplied no demonstrated additional gain: all 24 C2 cell-wise paired intervals of joint64 versus joint64-prior included zero. These negative contrasts were prespecified and retained. Complete probability calculations need not improve finite-budget ranking; shortlist error, finite nodes and cost ties are not separately identified by this experiment.

S25.3 Exact outer inference and integer target checks

Conditional on each frozen trained rule, the independent outer test losses are IID Bernoulli under the declared source model. The one-sided risk bound uses the allocation $\delta = 0.05/10{,}000$. For a threshold α, count k and $n = 20{,}000$, qualification is independently confirmed by

$$\Pr\{\mathrm{Bin}(n, \alpha) \le k\} \le \delta,$$

with exact integer recurrence after writing the rational target and error budget in integer form. All 18 A/TAB 5% and C/TAB 2% decisions in S25.1 passed this arithmetic check, separately cross-checked against direct combinatorial sums on small examples. Multiple targets use the same risk endpoint, rather than pretending to create independent statistics. The paired-difference family has its own registered allocation. Rate and paired families and the separate lower-bound addendum are not merged into a single overall 95% claim. No 60,000-state pooled bound is reported for the three differently trained policies.

S25.4 Exploratory bound within each new frozen menu

After primary execution began, a separate analysis addendum used the already saved test states without new sampling, reseeding or menu changes. For its frozen menu Q, the pointwise inequality

$$L(\pi(s), s) \ge H_Q(s) = \min_{q \in Q} L(q, s)$$

holds for any deterministic information rule in Q, and after averaging randomization for any randomized rule. All fast-flagged all-action-fail states were checked for *every menu action* with the original reference evaluator. Their confirmed count m_{cert} is no larger than the true all-fail count. Its lower endpoint uses denominator 20,000 and binomial domination; it does not regard the selected subset as an IID sample. Unknown-needed all-fail states remain failures. Pointwise lower endpoints and a separately adjusted exploratory family-10,000 endpoint were reported for all 72 cells, including zero and weak values.

For A/TAB/C0 the adjusted lower endpoints were 1.0713, 1.0495 and 1.2419%, excluding 1% for any rule restricted to those particular menus. For C/TAB/C0 they were 0.2939, 0.3686 and 0.3646%, neither excluding nor proving attainability of 1%. These floors cover the five expanded-menu policies. The old-library rules may choose outside Q, so they are not covered. Larger source pools, other blends and continuous rations are also outside the claim. The oracle analysis does not construct a test-selected deployable policy and was not prespecified as an eighth primary rule.

S25.5 Independent artifact, action and numerical checks

All 72 formal tasks completed with zero exit failures. The original runner checked every assigned policy action on all 20,000 actual test states with the original reference evaluator: 10,080,000 policy–state comparisons. Saved actual nutrient/DM arrays, loss/unknown masks, selections, frozen menu/parameters and complete draw fingerprints support independent readback. A first independent pass recomputed all 504 risk records and 648 paired contrasts, replayed training menu generation and mixture lineage, and checked plan/structure, original costs and frozen identities. A second selection replay, with sampling disabled, checked 1,152 fixed saved state positions: 8,064 selected actions and their original loss labels, plus 73,728 hypothetical joint indicators, all matched. This second check is a specified subsample, not a second independent full replay of 1.44 million states.

The finite-menu analysis evaluated 10,640,000 action–state pairs with the qualified algebraic chain, then used the original evaluator on 3,673,350 distinct action–state pairs covering 645,453 distinct state positions. All 644,821 all-fail state positions were confirmed without removed mismatches; 112 needed unknown verdicts. Its independent pass checked 936 input/evidence hashes and 144 exact integer lower-tail calculations. It did not run the original evaluator for every action on every one of the 20,000 states.

Across formal execution, 116,773,808 integrated action-nodes were recorded, with zero numerical-ambiguity fallbacks and zero audited clipping. The original persisted node audits cover the conditional64 and prior64 designs; rb16 is their fixed prefix, so the zero-clipping implication for it follows from that prefix relation. An independent rb16 node-audit log did not exist and is not claimed. Neither zero fallback nor zero clipping supplies a 16-node integration-error certificate or a full-domain root-solver guarantee.

The initial cross-platform conditional fingerprint check failed and stopped before scientific replay. Execution-platform fit/achieved-moment/latent-matrix metadata were then exported and matched to every original cell fingerprint, without refitting or redrawing. A greedy replay likewise detected last-bit costs differing between cached per-action dot products and a matrix multiplication. Reconstructing the original matrix shape and exact saved execution-platform costs reproduced all frozen selection keys and gains; the rule and outputs were not modified. These records preserve computational qualifications rather than claim byte-for-byte portability without evidence.

Supplementary Methods: Conditional information and the provenance of dependence assumptions (archive S26)

The first and second extensions retain the original moment-matched physical marginals. C0 is the original independent-cell scenario. C1 uses five matched after-screening nutrient pairs from the published correlation table of [9]: corn-silage CP–NDF and starch–NDF, canola-meal CP–NDF, and whole-cottonseed CP–NDF and starch–NDF. C1 does not link DM to the observed nutrients. C2 replaces the corn-silage pairs with the six-pair DM/CP/NDF/starch block from the authors' associated workbook, while retaining C1's three non-corn pairs (one canola and two

cottonseed nutrient pairs, where applicable). The author supplement is identified by the Ohio State University dairy feeding-and-nutrition resources and the Supplemental Feed Composition Table workbook. The registered analysis used its archived snapshot; source-table and workbook checks were independently recorded on 2 October 2026. This source pointer does not redistribute workbook numerical inputs. Unreported pairs are set to zero; physical correlations are translated to a latent-Gaussian model and the registered PSD/achieved-correlation checks are retained. This specifies a transfer sensitivity, not a newly estimated farm copula. The separately registered factor sensitivity is a different design and is not relabelled as author-source evidence.

These sources are historical laboratory submission populations, with between-farm/batch variation and database screening. The matched columns and ingredient rows were independently checked against the original workbook/table. Neither their correlation coefficients nor the five variability settings identify current within-farm covariance, source variance decomposition or assay noise. Preserving marginal moments does not preserve nonlinear joint failure or nutrient-supply covariance when DM and composition become dependent. Information-based sensor criteria have established uses under explicit models [10]; those assumptions do not supply assay reliability or economic value in this study.

Within the eight executed C2 case/setting models, complete nutrient observation left corn-silage latent-DM conditional variance about 0.693609 (prior one), corresponding to mutual information about 0.182924 nats. In TAB, starch alone left about 0.695307; adding NDF left about 0.693611, then CP about 0.693609. Thus CP's nonzero raw pair correlation provides almost no extra latent DM information once starch and NDF are known under this declared model. This is a conditional redundancy calculation, not a causal nutrient mechanism, instrument-priority recommendation or quantified risk gain. Its expected benefit need not be visible after a finite action menu and screening; S25.2 found no adjusted significant C2 risk improvement from the update.

For a common fixed library and information T, the ideal decision risk is

$$r^*(T) = \mathrm{E}\left[\min_{q \in Q} \mathrm{E}[L(q, S) \mid T]\right].$$

More informative signals cannot increase this *optimal population* risk for the same decision problem, because a rule can ignore extra information. A trained PCA/tree/delta rule with finite data and approximation error need not obey that monotonicity. Signal variance coverage, mutual information and source correlations are therefore not substitutes for independently tested decision risk. A full-state oracle sees actual DM and is not operational under the nutrient-only information pattern.

Supplementary Methods: Interpretation, negative results, costs and release limits (archive S27)

The results support a case- and model-specific progression: the original linear-energy procedures can return empty target-qualified sets; reference-aligned generation supplies new static witnesses; nutrient-only rules exploit complementary actions while hiding actual DM. The second extension's A 5% and C 2% results retain the original plan equality and requirements. They do not retroactively qualify any original empty entry or prove global optimality. The lower certificates distinguish continuous fixed-action risk from finite-menu information limits; neither failed search nor a weak lower bound proves population impossibility.

The practical assumptions remain substantial: one reference cow, restricted inventories, declared source moments/dependence, exact noiseless nutrient observation, supplied DM assumed consumed, fixed fMCP and selected NASEM constraints. No animal outcomes, actual intake, complete NASEM reference-software comparison, measurement-noise study, assay validity period or

field execution was supplied by these experiments. Historical correlation transport is uncertain. R0/B failures, PCA non-improvement, C2 update non-improvement and the A/C finite-budget integration negatives are part of the evidence, not omitted exceptions.

Mean feed-cost changes are secondary descriptions at the declared source prices. A baseline not qualified at the target cannot define a target-matched information value. Detection, implementation, transport and other fees and an economic loss function were not specified; no monetary EVSI/EVPI, net-profit or testing-priority claim follows. S10 records the original deferred information-value plan and remains historical; the extensions report independent risk-policy experiments rather than retrospectively completing a monetary value calculation.

The aggregate CSVs supply the available first-extension result summaries, all 504 second-extension risk records, all 648 same-state paired records, all menu lower endpoints and the applicable statistical families. Restricted raw arrays, full-precision rations, source registries, root seeds and source-table values remain in the controlled study archive, which is neither a submission supplement nor a public dataset. This section records the release status of the earlier version, which released no rations or requirement amounts; the current article releases the fixed-rule rations at one decimal and the equation-derived requirement amounts (current S38).

The application-specific mathematical addition is the one-DM quartic/linear-interval representation of this retained reference event and its qualified conditional-risk implementation, paired with independent action-space and risk comparisons. Conditional decisions, delta propagation, greedy coverage, Gaussian conditioning, Rao–Blackwellization and quasi-Monte Carlo are established methods. Neither these results nor the performed literature search establishes worldwide priority for a new general optimization method.

References

[1] Clopper, C. J. & Pearson, E. S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**, 404–413 (1934). URL http://dx.doi.org/10.1093/biomet/26.4.404.

[2] Luedtke, J. & Ahmed, S. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization* **19**, 674–699 (2008). URL http://dx.doi.org/10.1137/070702928.

[3] Pagnoncelli, B. K., Ahmed, S. & Shapiro, A. Sample Average Approximation method for chance constrained programming: Theory and applications. *Journal of Optimization Theory and Applications* **142**, 399–416 (2009). URL http://dx.doi.org/10.1007/s10957-009-9523-6.

[4] Nemhauser, G. L., Wolsey, L. A. & Fisher, M. L. An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming* **14**, 265–294 (1978). URL https://doi.org/10.1007/BF01588971.

[5] Mandi, J., Bucarey, V., Tchomba, M. M. K. & Guns, T. Decision-focused learning: Through the lens of learning to rank. *Proceedings of Machine Learning Research (ICML 2022)* **162**, 14935–14947 (2022). URL https://proceedings.mlr.press/v162/mandi22a.html.

[6] He, Z. On the error rate of conditional Quasi–Monte Carlo for discontinuous functions. *SIAM Journal on Numerical Analysis* **57**, 854–874 (2019). URL https://doi.org/10.1137/18M118270X.

[7] Liu, S. Conditional Quasi-Monte Carlo with constrained active subspaces. *SIAM Journal on Scientific Computing* **46**, A2999–A3021 (2024). URL https://doi.org/10.1137/23M1548918.

[8] Gilbert, A. D., Kuo, F. Y. & Sloan, I. H. Preintegration is not smoothing when monotonicity fails. *Advances in Modeling and Simulation: Festschrift for Pierre L'Ecuyer (Botev, Z., Keller, A., Lemieux, C. and Tuffin, B., eds; Springer, Cham)* 169–191 (2022). URL https://doi.org/10.1007/978-3-031-10193-9_9.

[9] Yoder, P., St-Pierre, N. & Weiss, W. A statistical filtering procedure to improve the accuracy of estimating population parameters in feed composition databases. *Journal of Dairy Science* **97**, 5645–5656 (2014). URL http://dx.doi.org/10.3168/jds.2013-7724.

[10] Krause, A., Singh, A. & Guestrin, C. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research* **9**, 235–284 (2008). URL https://www.jmlr.org/papers/v9/krause08a.html.