Vol. 1 — How Unsold grades

Methodology of record: GRADING.md v3.1.2 · unsoldlife.com · typeset from source, nothing paraphrased

Methodology v3.1.2 — 2026-08-30

This is the whole method, in public. Unsold penalizes proprietary blends; it would be strange to run on one. Every number below lives in `src/engine/score.ts` and `src/engine/ingredients.ts`, and every grade in the catalog follows from them: the same label data always produces the same grade, and the rules that get it there are the ones written down here.

What a grade is, and what it is not

Unsold grades the label's best case. Every stamp assumes the Supplement Facts panel is telling the truth — the doses as printed, the forms as named. Most products fail anyway.

Independent laboratory testing answers a different question: is the label true? That question matters, and the answer can move a grade in only one direction. **An F on paper cannot be redeemed by a lab. An A on paper still has to survive one.**

The office grades the promise. It does not test the powder, and it does not pretend to.

What the letters mean

Grade	Plain language
A+	Top-of-range clinical doses, best-in-class forms, nothing hidden. Rare on purpose.
A / A-	Dosed like they actually read the research. Buy with confidence.
B+ / B / B-	Mostly honest and mostly effective, with a soft spot or two — a light dose, a cheaper form, one thing undisclosed.
C+ / C / C-	Middling. The right ingredients shy of the doses that make them work, or fine on paper and cheap in the details. Not a scam, not a standout.
D+ / D / D-	Sprinkles of the good stuff priced like the full dose. Mostly filler and fairy dust.
F	Marketing with a Supplement Facts panel attached.

The cutoffs are the familiar academic ones (97/93/90 for A+/A/A-, and so on down). A grade is a compressed summary — the per-ingredient breakdown is the actual argument, and it is always one tap away.

The pipeline

Each product runs through the same seven steps, in order, every time:

1. **Dose normalization.** Every labeled dose is converted into the reference ingredient's canonical unit (mg ↔ mcg, IU → mcg for vitamin D). A dose we cannot honestly convert is not guessed at — it is set aside.
2. **Basis matching.** A range stated per *day* is compared against a day's dose; a range stated per *serving* is compared against one serving. Where components are only meaningful together — EPA and DHA — they are summed first and graded once. See *Comparing like with like* below.
3. **Clinical-range banding.** The normalized dose is compared against the clinically studied range in the reference table: effective (interpolated 90–100 by position inside the range), above-range (86), excessive (68), slightly under (62), underdosed (42), trace (24), meaningless trace (12). Concealed doses score 34. A row explicitly marked *unsupported* has no published generic range and does not contribute to the grade.
4. **Form-quality modulation.** Ingredients are not interchangeable chemicals. Magnesium glycinate absorbs; magnesium oxide mostly passes through. Each form carries a 0–1 quality factor, and a cheap form can dock up to 40% of the band score.
5. **Weighted mean.** Ingredient scores roll up into one 0–100 number, weighted by how much each ingredient matters (hero actives up to 2.0, incidental fairy dust as low as 0.6). Rows marked *incidental* — a processing aid the product makes no claim on — are left out here, and EPA and DHA have already been summed into the single row their evidence is stated on.
6. **Blend-concealment penalties.** On top of each hidden ingredient's low score, that mean is then shaved in proportion to how much of the formula is hidden inside proprietary blends (up to 40% of the grade), plus a flat extra penalty when a hero active — the ingredient you are actually buying the product for — is the thing being hidden. The order matters and it is this way round: you cannot shave a grade before you have computed one.
7. **Letter cutoffs + deterministic flags.** The number maps to a letter, and pure rules over the computed scores emit the flag chips: *proprietary blend hides the doses*, *underdosed hero ingredient*, *cheap ingredient form*, *most of the formula is hidden*, *every dose disclosed*, and so on. The one-line verdict comes from a fixed template table. No step anywhere in this pipeline involves randomness, the clock, the network, or a model generating text.

Where the numbers come from

A grade is only as good as the range it is measured against. For two years this document described those ranges in the aggregate — "commonly cited clinical literature and monographs" — which sounded reasonable and told you nothing about any particular number.

So in July 2026 every one of the 69 reference ingredients was audited against the literature, one row at a time, and every citation was re-checked by a second reader working independently of the one that found it. Both readers were AI agents, kept separate on purpose; the audit does not claim they were equivalent, and two machines agreeing is weaker evidence than two experts agreeing. The result is uncomfortable, and it is published here rather than smoothed over:

3 of the 69 ranges are anchored to a source that states their numbers. Three more carry a real citation that fixes at most one end of the band. The other 63 are not anchored at all.

That headline used to read "6 of the 69", and correcting it is the point of v3.0.3. Six rows carry a genuine, resolvable citation — but a citation is not automatically a warrant for both ends of a range. The NIH fixes magnesium's 350 mg ceiling and says nothing about where an effect begins. The taurine meta-analysis fixes a 1 g floor and runs all the way to 6 g, while this table stops at 2,000 mg by convention. The theanine trial administered a single 200 mg dose, which fixes neither end of 100–400. Counting those three as ranges a source *states* was the same overclaim, one level down, that this whole audit existed to remove — so they are now reported separately instead of folded in to make the number look better.

Every row says which it is, in the table itself. A row either carries a **source** — title, issuing body, year, a resolvable DOI/PMID/URL, the dose sentence quoted from it, and whether that sentence covers the whole range or only part of it — or it carries an explicit **no-source declaration** naming what *was* found and why that cannot govern the number. There is no third option, and there is no blank. A partial source that does not say which end it fixes will not pass the checks, and a row that leaves the field out will not compile, and both `npm run verify` and `npm test` refuse a declaration filled with an empty string that merely looks like one. CI runs the compiler, the checks and the tests on every change.

"No source" is a first-class answer here, and it is always better than a plausible-looking one. The five reasons a row can give:

Declaration	What it means
practitioner convention	A band the industry uses. Widely used is not established.
intake reference	Built on an RDA, AI, or UL. An adequacy floor or a safety ceiling is not an effective dose.
off-basis	Real evidence exists, but doses in mg/kg, or mg/litre, or a branded extract — so it cannot set a fixed milligram band.
null at the coded dose	Someone tested doses in this range and found nothing.
implementation sentinel	The number is a deliberate engine device, not a dose. (Apoaequorin only.)

No-source and unsupported are separate decisions. A no-source declaration preserves what was found about a coded range; it does not silently decide every product-grade consequence. An *unsupported* row is a narrower, explicit policy decision: no defensible generic range is published or used for that row's coded use, the ingredient remains visible on the label, and it is left out of the weighted mean. In v3.1.0 that decision applies only to **melatonin**, **L-tyrosine**, and **arginine**. Their no-source findings remain printed on the row; no other ingredient's policy changed. An unsupported ingredient hidden in a proprietary blend still counts as hidden mass — refusing a range does not launder concealment into transparency.

The 63 do not divide the way the comfortable version of this sentence would have it, so here is the count. **Intake reference is the largest group at 23** — rows built on an RDA, AI, or UL, which the table directly above says are not effective doses. **Practitioner convention is second at 18.** Twelve are **off-basis**, nine are **null**

at the coded dose — someone tested that range and found nothing — and one is the sentinel. So a plurality of our unsourced ranges rest on adequacy and safety references rather than on formulator practice, and nine of them sit on top of a null result. *Widely used*, *adequacy floor*, and *tested and found nothing* are three different claims, and you are entitled to tell them apart.

Where the line between "partial source" and "convention" currently sits, and why it is drawn conservatively. A row is recorded as a *partial source* when a resolvable paper states a dose for the same compound on the same basis that lands on an endpoint of the coded range. It stays *practitioner convention* when the located evidence is for a branded or standardized extract that a generic milligram row cannot inherit, when the dose sits inside the range without fixing either edge, or when the coded number came from formulator practice and the citation merely sits nearby.

Applied strictly, that rule would probably move more rows out of *convention* and into *partial* than the three now recorded: sixteen of the eighteen convention rows cite a resolvable paper, and several name a dose at or very near an endpoint. We have not moved them, and the reason is directional. Every such reclassification takes a row **out** of the 63 and puts it in the partial column; none of them can raise the headline 3. So the number this document leads with is a floor, not an estimate, and the error runs against us. Deciding each of those sixteen is an evidence judgement that deserves its own pass rather than a batch edit made while tidying, and until it happens this paragraph is the honest description of the boundary.

The two are independent, on purpose. Magnesium carries a real source — the NIH's 350 mg ceiling for supplemental magnesium — and still sits at the lowest certainty tier, because a safety limit is not evidence that a dose works. A citation is not a compliment.

A range is never tuned to protect a product. The ranges are the same for every product on the site, which is the entire reason the grades are reproducible. When a range moves, it moves because the evidence moved, and every product it touches moves with it — including the ones that were flattering us.

Comparing like with like

The per-day problem. A Supplement Facts panel states a dose *per serving*. Almost all of the governing evidence — RDIs, loading protocols, nearly every clinical trial — is a *daily total*. Comparing one directly against the other is a category error, and it quietly mis-grades every product meant to be taken more than once a day.

Each reference range now declares which basis it is on, and each product may declare how many servings a day its label's **directions** state. Where both are known, the engine compares a day against a day.

Where the directions are *not* on file, it does what it has always done — treats one serving as the day's dose, the most charitable reading available — but it now **says so on the row** instead of burying the assumption in the arithmetic:

> *That clinical range is a daily total, and this label's directions are not on file — so one serving was judged as one day's dose. If the label tells you to take more than one serving a day, the real daily amount is higher than what is measured here.*

Being straight about the current state: **no product in the catalog records its label directions yet**, so that notice appears on every per-day row today. Transcribing directions is label-reading work, product by product, and it

has not been done. The mechanism is built, tested, and honest about its own gap — which is better than an engine that silently assumed the answer, and better than a number invented to fill the field. Not one grade moved because of this change.

What the range is measured in. Every row also declares its *mass basis*, because milligrams are not interchangeable. Elemental magnesium is not magnesium salt. Folate is dosed in DFE, vitamin A in RAE, niacin in NE — conversions most labels never print. Curcuminoids are not turmeric. Citrulline malate is about two-thirds citrulline. Alpha-GPC raw material is often half alpha-GPC. Probiotics are counted, not weighed. Where a row's number means something narrower than the word on the label, the row says so.

EPA and DHA. A fish oil can print one number ("EPA/DHA 600 mg") or two ("EPA 360 mg, DHA 240 mg"). Until v3.0.0 those were graded by different rules: the combined row cleared its floor at 250 mg, while the split rows each demanded 250 mg — so an identical oil needed *twice as much* to pass, purely for printing more detail. It was also weighted twice. Punishing a label for itemizing is exactly backwards for this project.

The NIH records that no EPA-specific or DHA-specific intake recommendation was ever set, and the American Heart Association's language is combined. There is no authority for a 250 mg EPA-alone floor, so the engine no longer enforces one: **EPA and DHA are summed and graded once, against the combined total.** Both rows stay on the panel exactly as printed, marked as counted in the total above.

Two limits are deliberate. A component hidden inside a proprietary blend is never absorbed into a total — that would launder concealment into a clean number. And if a label already discloses the combined figure itself, nothing is combined, so no dose is ever counted twice.

The stance

Why blends are penalized. A hidden dose is worse than a known-bad dose. A known-bad dose you can at least see; a hidden one asks you to pay for a promise. "Proprietary blend" is not a formulation secret — the FDA already requires every active in the blend to be named, just not quantified. The only thing concealed is the part that lets you judge value. So concealment is graded as what it is: a choice to keep you from checking.

Why some rows do not count. A Supplement Facts panel lists what is in the product, which is not the same as what the product is offering. A creatine gummy sets with pectin and a sodium-citrate buffer, so its panel prints a few milligrams of sodium. Nobody buys a creatine gummy for sodium.

A row is marked **incidental** when both are true: the nutrient appears only as an artifact of manufacture, and the product makes no claim on it. An incidental row is transcribed and shown exactly as printed — the panel is never edited — but it does not enter the weighted mean, because this engine grades toward the ingredients you actually bought the product for, and grading a creatine gummy against its buffer salt contradicts that.

The rule is deliberately narrow. A nutrient the product's name, category, marketing, or panel positioning presents as something it delivers is **not** incidental, however small the dose: an electrolyte powder with 40 mg of sodium is underdosed, not incidental, and a multivitamin's 20 mg of magnesium is a claim it chose to make. Where the evidence does not clearly show a processing aid, the row stays scored. The conservative answer is always to leave it scored.

This marker can only ever improve a grade, so every use of it is published in the product data itself — visible on the product's own result page, not hidden in an internal note.

Why forms matter. Two labels can print "400 mg magnesium" and deliver wildly different amounts of usable magnesium. Grading the molecule and ignoring the form would grade the marketing, not the product.

Why unscored rows exist. If an ingredient is not in the clinical reference table, we show it and score *nothing*. We never invent a range for an ingredient we have not reviewed. (Note what that does and does not promise: the 69 rows we *do* score against are declared one by one above, and 63 of them carry no governing source. Not inventing a range is a lower bar than citing one, and this sentence only claims the lower bar.) An unrecognized ingredient is neutral — it neither helps nor hurts the grade — because "we haven't reviewed it" is not evidence of anything.

What "unscorable" means. If a product contains *nothing* we can score — an all-novel formula — it does not get a fabricated grade. It gets an explicit "nothing recognizable to grade" result. A grade we cannot support is worse than no grade.

A worked example of the philosophy: the concealed-blend fix

For a while, the concealment math had a hole. The blend penalty counted hidden rows only when the hidden ingredient was *also* in the reference table. A row hidden in a blend **and** unrecognized escaped entirely — it didn't count as hidden mass, and in products where it was the *only* blend row, the product didn't even register as having a blend.

The predictable result: a multi-collagen product hid its entire active formula behind one row labeled "Proprietary Blend" and graded **B+** on its two disclosed token vitamins. A pre-workout parked a 13.5-gram stimulant matrix in unrecognized blend rows and graded **B** on the handful of ingredients it chose to disclose.

That is backwards. A label does not get *less* suspicious because the thing it hides is obscure — the less we can identify the hidden mass, the more the concealment *is* the story. The fix: concealment now counts **every** blend-hidden row, recognized or not. Unrecognized hidden rows still get no per-ingredient score (we never invent a range), but their mass counts toward the "most of the formula is hidden" penalties and flags. Those two products dropped to **C** and **C-**, and 47 other products with the same pattern moved down with them. No product was special-cased; the rule changed, the math followed.

Certainty

A grade says how honest and well-dosed a label is. It does not say how strong the science behind the ingredients is — creatine and coffee-fruit extract can both be dosed perfectly, and they are not equally proven. So every result also carries a small `evidence:` mark next to the grade: **high**, **solid**, **emerging**, or **early**. Each reference ingredient is tiered by the body of evidence located for that ingredient (A: meta-analysis, position stand, or multiple RCTs; B: two or more controlled trials; C: a single trial or clearly limited human evidence; D: observational, traditional-use, or mechanism-only), and the mark reports which tier dominates the scored formula, weighted by how much each ingredient matters. It rides alongside the grade and never changes it — an F built on high-certainty evidence is still an F, and an A built on early evidence is still an A.

v3.0.0 lowered 46 of the 69 tiers. The rule is that a tier grades the evidence located for the *ingredient* — how much good research says it does anything at all — judged on efficacy findings only. Under that rule most of the vitamin and mineral rows fell to D: they are built on RDAs and ULs, which are adequacy and safety references rather than efficacy findings. **No tier was raised.**

A tier is not a claim about our numbers. Whether a source states the range we coded is a separate field, recorded separately, and the two do not move together. Caffeine is tier A carrying no source: the literature is deep, and no single paper prescribes 100–400 mg. Magnesium is the mirror image — a real citation at the lowest tier, because a safety ceiling is not evidence that a dose works. Read the tier as *how well studied is this ingredient*, and the range's own source line as *who says these numbers*. Six ingredients have the second. Sixty-nine have the first.

Where a clinical guideline recommends *against* an ingredient — the American Academy of Sleep Medicine on melatonin, systematic reviews on valerian and chamomile — that lowers the certainty mark and is written into the row's caveat, where a reader will see it. It does **not** delete the range and it is **never** smuggled in as a dose penalty. This engine grades whether a label matches the doses that were studied. It does not decide for you whether the ingredient is worth taking; it tells you what was found and lets you weigh it.

Manufacturing Assurance

The grade and the certainty mark answer two questions about the *label*: is it dosed well, and how well studied are the ingredients. Neither one asks whether the tub in your hand actually contains what the label says. Starting with a small pilot, some result pages also carry a third, separate line: **Manufacturing Assurance** — what publicly verifiable quality-control evidence exists for *this product*, found in manufacturer statements, certifier registries, and facility audit records.

What this is not. Unsold did not open a single tub. It did not run a single test. Manufacturing Assurance reports what is publicly documented about *someone else's* testing — the manufacturer's own claim, an accredited lab's registry entry, a facility auditor's certificate — never Unsold's own verification of purity, potency, contents, or safety. A high mark here is not Unsold vouching for the product; it is Unsold pointing at what is checkable and naming exactly who checked it.

Five tiers, low to high:

Tier	What it means
No assurance evidence found	Checked; no public quality-control claim located beyond a bare legal baseline (every U.S. facility must already be cGMP-compliant — that alone is not evidence of anything more).
Manufacturer-declared testing	The manufacturer states testing publicly, but no independent registry confirms it. A real certification program *named* by the manufacturer is still a manufacturer claim until the certifier's own listing corroborates it.
Independently audited facility	A named independent body currently audits the manufacturing *facility* — not this exact product.
Exact-product independent certification	This exact SKU is listed by name in an official independent certifier's own registry.

Current lot / every-batch verification	An independent source names covered lots, or states every batch is tested before release, for this exact product.

Scope is not optional reading. Every entry also states what the evidence actually *covers* — the finished product, the facility that makes it, or (for a branded raw ingredient inside the formula, like a chelated mineral) only that ingredient's own supply chain. A branded ingredient's authenticity program is real evidence about *that ingredient*; it never certifies the finished, blended, encapsulated product around it, and this page never lets it try.

"Current" is bounded, not permanent. Every record is explicitly marked as tier support, context, or history, and every supporting record must match the entry's stated scope. Currency is evaluated against the frozen evidence date printed on the disclosure. A certificate must remain valid on that date; a live registry with no published expiration records when it was observed and the date through which that observation may support current wording. Moving the evidence date forward requires a fresh registry check. An expired or malformed supporting record fails even when another valid source sits beside it; historical evidence can remain visible, but cannot support the tier.

The v1 pilot is 11 catalog products, hand-researched against primary sources, deliberately spanning the full tier range above. Every other product on the site shows **"Not yet reviewed"** — an honest gap, never confused with a checked-and-found-nothing "no assurance evidence found." The pilot does not move, rank, or filter anything else on the site: it is absent from the score, the grade distribution, category rankings, better-graded alternatives, and the certainty mark, by construction.

Reproducibility

The grading engine is a pure function. Same label data in, same grade out — no randomness, clock, network, or LLM. Manufacturing Assurance likewise never reads the machine clock: it validates its frozen evidence ledger against the explicit evidence date stored with that ledger.

```bash npm run verify # 41 checks: worked examples, determinism, flag rules,

npm test # unit tests + a full-catalog snapshot test ```

The snapshot test (`src/engine/snapshot.test.ts`) records every product's grade, numeric score, concealment fraction, and flags. Any tuning change — a constant, a reference range, a data edit — shows up in the pull request as a complete, line-by-line diff of its consequences. CI runs verify, tests, and a build on every PR and posts the grade distribution as a comment. If a grade in the app ever disagrees with what `npm run verify` prints on the same commit, that is a bug, and we want to hear about it.

## What grades are NOT

- **Not a safety assessment.** An F does not mean a product is dangerous and an A+ does not mean it is safe *for you*. Interactions, contraindications, allergies, pregnancy, medication — none of that is in scope.
- **Not medical advice.** Unsold grades whether a label delivers the doses and forms the clinical literature studied. Whether you should take anything is a conversation for a clinician who knows your history.
```

- **Not a review of the brand.** We grade formulas, not companies. A brand can earn an A on one SKU and an F on the next; several do.
- **Not permanent.** Labels reformulate. Each grade is a dated snapshot of a specific panel from the source cited on its result page. If the tub in your hand prints a different panel, the label wins — confirm against the physical product before you buy.

Changelog

- **v3.1.2 — 2026-08-30.** Two NIH DSLD records were corrected against their structured labels. Gorilla Mode now records both the printed 190 mg elemental sodium and the separate 500 mg Pink Himalayan Sea Salt mass; the salt row is shown but not scored, so sodium is not counted twice. Its amended filing moves from C / 74.1 to C- / 72.7. Redcon1 Total War now records its omitted 5 mg Cocoa Seed Extract row and the label's printed botanical forms and standardizations; those rows have no matching repository reference and remain visible and unscored, so its B / 84.6 result is unchanged.

No scoring rule, range, weight, cutoff, or evidence policy changed. One product changed numeric score and letter grade; the raw grade distribution remains A:62 B:61 C:28 D:31 F:97. Counted as verdicts, the catalog still has 263 graded products with **81** real F grades and 16 refusals standing outside the distribution entirely.

- **v3.1.1 — 2026-08-28.** Four records were refreshed against current manufacturer or NIH DSLD evidence. Transparent Labs BULK now records its 20.8 g serving, disclosed sodium, corrected potassium, and the printed B6/B12 forms; its amended filing moves from B+ to A-. Garden of Life Vitamin Code Men now keeps K1 120 mcg and K2 (MK-4) 20 mcg as separate rows. Hiya now records calcium at 25 mg and its disclosed 25 mg fruit-and-vegetable blend. Seed DS-01 now records the manufacturer's current four blend totals; its stored 400 mg prebiotic row remains explicitly unconfirmed by that page.

No scoring rule, range, weight, cutoff, or evidence policy changed. Three products changed numeric score, one changed letter grade, and Hiya's score stayed fixed. Two third-party source records now bind to manufacturer pages, reducing the unverified-source count from five to three. Raw engine distribution is A:62 B:61 C:28 D:31 F:97. Counted as verdicts, the catalog has 263 graded products with **81** real F grades and 16 refusals standing outside the distribution entirely.

- **v3.1.0 — 2026-08-06.** Three owner-decided unsupported evidence findings now change the engine rather than sitting beside it: melatonin, L-tyrosine, and arginine have no generic clinical range to publish or score for their coded uses. Their labels and no-source declarations remain visible, but their former bands no longer supply a grade. This is not a narrower substitute range, and no other ingredient was reclassified.

Seventeen products contain a target row; sixteen numeric scores changed. Seven products changed letter grades, GNC Men's Nitric Oxide Maximizer became *not graded* rather than carrying the cutoff artifact F, and eight changed score without a letter or refusal change. The eight amended filings preserve their former stamps; Nutra Innovations Epitome also retains its earlier B → C amendment. Built By Nature Nitric Oxide remains F / 21.4. The full deterministic inventory is produced from the catalog by the verification tests and build artifacts.

Raw engine distribution is A:61 B:62 C:28 D:31 F:97. Counted as verdicts, the catalog has 263 graded products with **81** real F grades and 16 refusals standing outside the distribution entirely. The remaining 16 carry no grade.

Unsupported rows render an explicit refusal on the product page, score no dose band, and leave a hidden blend's opacity penalty intact. The one refusal follows the same public/API rule as every other ungraded product: no letter grade is published in its title, share card, certificate, docket, structured data, or generated API metadata.

- **v3.0.4 — 2026-07-29.** Fifteen products were being published as **F** for a reason that is not their fault: nothing on their panel is in this reference table. Greens powders, digestive enzymes, colostrum, fruit-and-vegetable blends. The engine scores them 0, the letter cutoffs turn 0 into F, and the page title, the share description and the grade card all went on to state it.

F is a verdict — the table above defines it as "Marketing with a Supplement Facts panel attached." Publishing it for a formula we simply have no reference data for asserts a judgement the record does not support, about a named company, in exactly the places that travel: an indexable `<title>`, an `og:title` unfurled by every messenger, and a downloadable card. The structured data on those pages had been honest all along, emitting no rating at all; the human-readable half had not.

No grade moved and nothing was re-scored. The numeric is still 0, the internal letter is still F, the snapshot is untouched. What changed is what gets *published*: those pages now read "not graded", the card presses NOT GRADED in caption ink rather than a red F, the stamp on the page does the same, and the score line says nothing on this panel is in the reference table. The distribution is unchanged — the halls, the homepage and `llms.txt` had already excluded these fifteen, which is precisely why the grade surfaces and the counting surfaces disagreed.

The predicate for "the engine refused to grade this" had been copy-pasted into six files. It is now one exported function, which is the actual fix: the duplication is why the surfaces drifted apart in the first place.

Also in this release: the boundary between *partial source* and *practitioner convention* is now written down under *Where the numbers come from*, including the admission that applied strictly it would likely move more rows into the partial column — never into the headline 3, so the published number is a floor.

- **v3.0.3 — 2026-07-29.** The headline number was too flattering, and this corrects it downward. **3 of the 69 ranges are anchored to a source that states their numbers. Three more are anchored at one end at most. The other 63 are not.** Between them sit three rows that carry a real, resolvable citation covering at most one end of the coded band — magnesium's ceiling, taurine's floor, and a single theanine dose that fixes neither end of its range.

Nothing was found to be wrong with those three citations; they are the same sources, still quoted, still resolvable. What changed is that "a source states this range" now means the source states *the range*, not that a source exists somewhere near it. Every source row declares its coverage, a partial one must name which end it fixes and which end is convention, and a build check refuses a partial that does not.

The reader-facing effect is on the row itself: the breakdown now prints the gap, so a magnesium range of 200–350 mg sits directly above a quote that mentions only 350. The same check that guards the split in this document now also guards it in `llms.txt`, where the sourcing breakdown had never been verified — the catalog numbers one line above it were.

No range, tier, score, letter, or flag changed, and no grade moved.

- **v3.0.2 — 2026-07-29.** The provenance every range carries is now visible on the site, and the facts this document states are now checked against the engine instead of being retyped.

The declarations are published, not just recorded. v3.0.0 made every range declare its origin and then showed it to nobody — the only reader was a build lint. A methodology fact that lives only in the repository is not published, whatever the changelog says. Expand any scored row in the per-ingredient breakdown and it now prints what the range is measured in and where it came from: the source that states its numbers, quoted, or the explicit declaration that no source does and what was found instead. No range, tier, weight or grade changed.

Four surfaces were saying things this document had already retracted. `llms.txt` — the file written for AI assistants to repeat verbatim — called every one of the reference ranges *cited*, which is the exact aggregate claim v3.0.0 was written to withdraw, and every catalog number in it was stale by ten products and two methodology versions. `humans.txt` cited v2.2.0. The downloadable bulletin PDF was pressed from v2.2.0 and offered under a line reading "typeset from source, nothing paraphrased". The type definitions still defined an evidence tier as grading "the citations behind a reference row" — the definition v3.0.1 retired. All four are corrected.

One reassurance was false and is withdrawn. This document said "most of the 63 are practitioner conventions". They are not: convention is the second-largest group at 18. The largest is intake reference at 23 — RDAs and ULs, which this document elsewhere says are not effective doses — and nine rows are null at the coded dose. The composition is now printed and checked.

The pipeline was described in the wrong order. Step 5 shaved the grade and step 6 computed it. The engine has always done it the other way round; the list now matches the code.

The F in the distribution is two things added together, and says so now. A:61 B:62 C:30 D:32 F:94 is the raw engine distribution over all 279 products, in which the 15 the engine refused to grade carry a placeholder F. Counted as verdicts rather than ranks — the way the halls, the category reports and the front page count them — it is 264 graded products and 79 real F grades. Both numbers were already published on different surfaces; neither moved.

Thirteen new checks close the class. A duplicated count, version or definition must now be derived from the engine or asserted equal to it. The sweep fails when this document's sourced/unsourced split, declaration composition, distribution, version, check count, or retired-definition ban stops matching the table; when `llms.txt` or `README.md` state a catalog the engine does not produce; when the breakdown stops rendering the declarations; when the downloadable PDF was pressed from a different GRADING.md; and — for the first time — when the scoring path can see an evidence tier at all. That last one is the rule this document has stated in seven places since v2.2.0 and had never once tested.

No range, tier, weight, hero flag, form quality, score, letter or flag changed, and no grade moved.

- **v3.0.1 — 2026-07-29.** A contradiction introduced by v3.0.0, found on review and corrected before publication. v3.0.0 gave every range a formal source declaration, and 63 of the 69 declared they had none — while the product page went on calling each scored ingredient a "tier-A **source**," and this document defined a tier as grading "the located evidence for the range as coded." Under that definition six tier-A rows could not have kept an A, since no source states their range.

The tiers were right and the words were wrong. A tier grades how well studied the *ingredient* is; provenance records who states our *numbers*. They are separate questions and the repository now says so consistently — the page counts ingredients rather than sources, and the *Certainty* section states the distinction outright. No range, tier, score, letter, or flag changed, and no grade moved.

- **v3.0.0 — 2026-07-29.** The clinical evidence audit of all 69 reference ingredients is closed out, and every range now declares where its numbers came from.

The headline as v3.0.0 published it: six rows carried a citation, and the remaining 63 declared none, each one naming in the table what was actually located and why it cannot govern. (v3.0.3 corrected the first half of that sentence: only three of those six state their range end to end.) Reasonable practitioner convention is the second-largest group, at 18; the largest is intake reference at 23, and nine rows are null at the coded dose. "Widely used" and "established" are different claims and the table no longer blurs them. A build-failing check makes an undeclared range impossible, and a second one fails if a declaration is left empty.

46 of the 69 evidence tiers were lowered. A tier grades the body of evidence behind the *ingredient* — how much good research says it does anything at all. That is a different question from provenance, which asks whether a source states our *range*, and the two are deliberately independent: caffeine is tier A carrying no source (the literature is deep, but no single paper prescribes 100–400 mg), while magnesium carries a real citation at the lowest tier (a safety ceiling is not evidence that a dose works). Under that test most vitamin and mineral rows fell to the lowest tier: they rest on RDAs and ULs, which measure dietary adequacy and safety, not efficacy. No tier was raised. Certainty still rides alongside the grade and still never changes it, so **not one grade moved because of a tier.**

The per-serving/per-day mismatch is fixed. Panels state doses per serving; the evidence is almost always a daily total. Ranges now declare their basis, products may declare the servings per day their directions state, and the engine compares a day against a day. No product records its directions yet, so today every per-day row carries a visible notice that one serving was judged as one day's dose — the assumption is stated instead of hidden. No grade moved.

EPA and DHA are now graded on their combined total. A label printing "EPA 360 mg, DHA 240 mg" used to need 500 mg to clear two separate 250 mg floors, while a label printing "600 mg" cleared one — the same oil, graded worse for itemizing. No authority sets an EPA-alone or DHA-alone floor, so the engine stopped enforcing one. Both rows stay on the panel; the total carries the grade. Hidden components are never absorbed into a total, and a label that discloses the combined figure is never counted twice.

One range moved: beta-alanine, to 4,000–6,000 mg. The ISSN position stand states 4 to 6 g daily; the table had run 3,200–6,400 mg, awarding a full effective score 800 mg below the cited floor. As with magnesium in v2.4.0, the source states a dose on the same basis as the row, so the source sets the number.

Twelve products were restamped, each carrying a Notice of Amended Filing with the previous grade struck through: seven pre-workouts fell as beta-alanine came into line with its source (Legion Pulse and Transparent Labs BULK from A to B+, Ghost Legend and Redcon1 Total War from A to B, and three more), and five fish oils rose once EPA and DHA stopped being measured against a floor nobody set (Pure Encapsulations EPA/DHA Vegetarian from D- to B, four others from C- to B). Twenty-three products moved in total; every one traces to those two changes, and nothing else in the catalog moved at all.

Distribution went from A:65 B:54 C:35 D:31 F:94 to A:61 B:62 C:30 D:32 F:94. Nothing was special-cased. Some grades went up, some went down, and the ones that went down were ours to lose.

Read that F carefully, because it is two things added together. It is the raw engine distribution over all 279 products, and the 15 the engine *refused* to grade carry a placeholder F inside it. Counted as verdicts rather than ranks — which is how the Hall of Shame, the category reports and the front page all count them — the catalog is 264 graded products with **79** real F grades and 15 refusals standing outside the distribution entirely. Refusing to grade is a separate verdict, not a rank; the raw number is printed here because it is what `npm run verify` prints, and a document that quoted a friendlier total than its own harness would be doing the thing this page is about.

- **v2.4.0 — 2026-07-29.** Magnesium's effective range now ends at 350 mg, not 400 mg. The NIH Office of Dietary Supplements sets the Tolerable Upper Intake Level for *supplemental* magnesium at 350 mg for adults, so the previous table awarded its best possible score to a dose above an official safety ceiling. Six products that had been stamped A are restamped B or B-, and each carries a Notice of Amended Filing showing the previous grade struck through. Several mid-range products rose slightly, because a narrower range means a 300 mg dose now sits higher within it. Nothing was special-cased: the range moved, the math followed.

This is the first range corrected as a result of the clinical evidence audit of all 69 reference ingredients. That audit also found that most ranges in this table carry no recorded source, and that several evidence tiers overstate what the located literature supports. Those remain open and are being worked through deliberately rather than quietly.

- **v2.3.0 — 2026-07-29.** Added the incidental-row rule above. The engine weights sodium as a hero ingredient against a 200–1000 mg clinical range, which is correct for an electrolyte powder and wrong for a creatine gummy whose only sodium is its citrate buffer. Force Factor Essentials Creatine Gummies doses creatine at a clinical 5000 mg, conceals nothing, and graded **D** on 5 mg of sodium; it now grades A+. Beast Bites moved the same way. No clinical range, weight, or hero flag was changed — the fix is a per-row marker, applied only where a label names the processing aid. Three rows across the catalog qualified. Effective Nutra carries one and still grades F, because its creatine dose is 1000 mg: the marker corrects a category error, it does not rescue an underdosed product.
- **v2.2.0 — 2026-07-27 (launch).** First versioned public snapshot of the methodology, including the certainty indicator described above. This document lived as unversioned working notes in the repo before launch; v2.2.0 is simply the moment it started keeping count, numbered to match the catalog's `DATA_VERSION` so the doc and the data never drift apart.