

The Surveillance Protocol

Defensible Evidence Is Built Before Litigation, Not After

Karen Pendergrass

Standards Architect, the Paleo Foundation, Heavy Metal Tested & Certified (HMTc) Program

June 2026

Two heavy-metal certifications can carry the same mark and mean opposite things in a deposition. One is a photograph. The other is a record. This briefing concerns the difference between them as evidence: why a one-time certificate invites the attacks a competent plaintiff's expert is trained to make, why a surveillance-based certificate resists them, and why the question is settled — for the defendant or against it — years before any complaint is filed.

Two certifications behind one mark

Certification is not, for a litigator's purposes, one thing. It is two, and the two behave in opposite ways once the testing behind the mark is placed under oath.

Call them snapshot certification and surveillance certification. A snapshot certifies a product against a limit on the strength of testing performed once, or performed occasionally and at the manufacturer's discretion: a sample is drawn, a laboratory reports a number, the number falls below the limit, the certificate issues. A surveillance certification rests on a continuing protocol: the product is enrolled, tested on a fixed schedule by a sampler independent of the manufacturer, under documented chain of custody, against published decision rules, with every result — favorable or not — recorded in a single dataset. On the package the two are indistinguishable. As evidence they are nearly opposites.

The question that separates them is not the one a brand prepares for. It is not whether the product is certified; certification is conceded. It is the next question: show me the testing behind the certificate. Everything turns on what the company can produce when that question is asked, and on when the answer was created.

Two rulings have made the question more pressing. In December 2025, the court overseeing the federal baby-food multidistrict litigation^[1] excluded five of the plaintiffs' six general-causation experts under Rule 702^[2]. In February 2026, a unanimous Supreme Court vacated a manufacturer's defense verdict and returned the case to state court^[3]. The first ruling makes the plaintiffs' science harder to admit; the second

makes the defendant's preferred forum harder to keep. Together they move the contest away from the expert battle over general causation and toward a narrower question a state jury can decide without a toxicologist: what does each side's own documented testing record show? That contest is not won at trial. It is won, or lost, by the record the company built before the complaint.

What a snapshot invites

Begin with the weaker record, because it is the common one. A snapshot is not worthless. A brand-commissioned certificate of analysis will often clear the low bar of admissibility. But admissibility is not the contest. The contest is weight — what the jury is persuaded the document proves — and on weight a snapshot is exposed to four attacks, each of which a competent plaintiff's expert will make in turn.

The first is representativeness. One passing result describes one sample from one lot. It does not describe the lot the child consumed, and it does not describe the population of lots the company shipped across the years at issue. A single favorable number is not a distribution; it is a point, and a point can always be called an outlier.

The second is selection. When the manufacturer decides which lots to test and which results to retain, the testing file is a product of the manufacturer's own choices, and every choice is a question on cross: which lots went untested, which results went unreported, who made the call.

The third is timing. Testing that begins after an inquiry — after a regulator's letter, after a demand, after a complaint — carries the mark of its origin. Evidence created with litigation in view is evidence built to persuade, and the law has distrusted it since long before the Federal Rules of Evidence. The accident report prepared for the case, not in the ordinary course of the business, is the classic example^[4].

The fourth is independence. When the manufacturer draws its own samples and pays the laboratory that reports the numbers, the plaintiff's expert need not prove the result wrong. He need only observe who produced it.

Consider the cross-examination a snapshot invites.

Q. Your company chose which lots to send to the laboratory?

A. Yes.

Q. And which of the laboratory's reports to keep in the certification file?

A. Yes.

Q. The lot my client's son consumed — was it among the lots you tested?

A. I don't know.

Nothing in that exchange requires an expert. Each answer is true, and each is damaging, because the snapshot cannot supply the missing terms: the lots not chosen, the results not kept, the lot actually eaten.

What a record establishes

The surveillance record answers the same four attacks, and it answers them not by argument but by structure. The advocate does not have to out-argue the plaintiff's expert. The protocol has already done the work.

Representativeness. A surveillance record is a series, not a point. The product is tested on a fixed schedule, lot after lot, across time, so the lot the child consumed sits inside a demonstrated distribution of results

rather than beside a single fortunate one. The defense expert opens with a dataset; the plaintiff's expert is left to argue about the intervals between scheduled tests rather than about the absence of testing.

Selection. The schedule is fixed by the protocol, not chosen by the brand; the sampler is independent of the production team; every result is recorded, in full, in machine-readable form; and the protocol forbids what the defense bar will recognize as the engine of the selection attack — “testing into compliance,” the undocumented resampling and laboratory-shopping by which a clean number is manufactured. A record that could not have been curated cannot be impeached as curated.

Timing. This is the decisive property. A surveillance record is made in the regular course of the company's quality operation, on a schedule set before any dispute, and it is therefore the kind of document the business-records exception was written to admit. Federal Rule of Evidence 803(6), and its state analogues such as California Evidence Code section 1271, reach records made and kept as a regular practice, and they extend a presumption of trustworthiness for the very reason that such records were not made for the courtroom^[5], [6]. The same principle, traceable to *Palmer v. Hoffman*, withholds that trust from the document prepared once litigation is in view^[4]. A snapshot and a surveillance record can sit in the same file and draw opposite treatment, for the single reason that one was built before the question and the other after it.

Independence and authentication. Samples drawn by a party independent of the brand, analyzed by a laboratory accredited to ISO/IEC 17025, and moved under a documented chain of custody in which a break invalidates the result, supply the authentication a snapshot leaves to argument^[7]. The chain-of-custody record is not a formality. It is the foundation that forecloses the handling-and-tampering cross before it can begin.

The decision rule a hostile expert cannot move

One feature of a serious surveillance protocol deserves separate notice, because it is the feature a hostile expert is least able to disturb. A surveillance program judges each result against the limit using the laboratory's expanded measurement uncertainty: a lot passes only if the result plus its uncertainty is at or below the limit, and fails only if the result minus its uncertainty exceeds it; a result in between is borderline and requires confirmatory testing before any status issues^[8]. The rule is not tilted toward the brand or toward the regulator. It is tilted toward analytical accuracy. A decision rule biased only toward accuracy is one a cross-examiner cannot recharacterize as a thumb on the scale; the expert who tries finds himself arguing against measurement science, before a jury, on the defendant's chosen ground.

Two files, side by side

It helps to set the two files on the table, because the difference an advocate cares about is visible at a glance.

The snapshot file is a certificate and a handful of laboratory reports. It is thin, and its thinness is the problem: it shows a result, not a practice. Where it is thick, it is often thick in the wrong way — assembled, organized, and dated in a manner that reveals when the company began to pay attention, which is frequently after the letter arrived.

The surveillance file is a narrative of attention. It opens with a baseline — three production lots tested before the product is certified at all — and a status assigned from the worst of the three. It continues as a scheduled series of independent samples, each with its chain-of-custody record and its laboratory report stating the analyte, its form, the limit, and the measurement uncertainty. It includes reflex tests triggered not by the calendar but by defined events — a new supplier, a new water source, a packaging change —

which show a company that tests when the risk changes, not only when the schedule comes due. Where a result drifted upward, it includes the corrective action that followed. The file is not a document a company assembles. It is the residue of something a company did, continuously, while no one was watching. That is precisely why it persuades.

“They didn’t test” is the sentence that opened this litigation wave, drawn from a congressional finding that one major manufacturer had not tested its finished products for heavy metals until 2019^[9]. It is a sentence a jury understands without help, and no cross-examination makes it go away after the fact. The surveillance file is the only complete answer to it, because the answer cannot be a document. It has to be a history.

The advantage, stated precisely

State the advantage exactly, and do not overstate it. A surveillance record does not make a product safe, and it does not establish that any exposure did or did not cause any injury. Those are different questions, governed by the science of general causation, and they belong where they are decided. What the record governs is narrower and, after the forum shift, more often dispositive: the testing history, and what a jury makes of it. On that question the record changes the contest from one the brand tends to lose into one it can win. A snapshot puts the brand’s credibility in issue — who chose, who paid, when did you start — and credibility is the contest a litigation-built file loses. A record puts data in issue, and a brand whose product is in fact compliant wins a contest about data.

Candor runs the other way as well, and it belongs in the analysis. A surveillance record is discoverable. It is generated continuously, it is held by a third party, and a brand should assume the plaintiff will obtain it. For a product that is genuinely contaminated, the protocol will document that too; it is not a device for concealing results, and a program worth advising toward will say as much. The honest claim is the narrower one. For the ordinary brand — whose product is compliant but whose proof of compliance is a drawer of occasional certificates — a surveillance protocol converts a weak and impeachable file into a strong and self-authenticating one, and it does so in the only window when the conversion is possible, which is before the dispute.

The architecture to advise toward

For counsel asked which certification to steer a client toward, the analysis resolves into a checklist, because each evidentiary function corresponds to a specific protocol feature. The Heavy Metal Tested & Certified program was built to this architecture, and it serves as a usable reference for what to require of any program^[10]:

- A baseline established before certification — here, three production lots — so the record carries a starting line that predates the relationship.
- Testing on a fixed, risk-based schedule rather than at the brand’s discretion — monthly for the highest-exposure categories, such as infant formula and infant cereal — which supplies the continuity that answers the representativeness attack.
- Sampling by a party independent of the brand’s production team, which meets the independence and selection attacks at their source.
- A documented chain of custody in which a break invalidates the result, which supplies authentication instead of leaving it to argument.
- A decision rule that applies the laboratory’s measurement uncertainty, which a hostile expert cannot recharacterize as bias.

- An express prohibition on testing into compliance — no undocumented resampling, no laboratory-shopping — which is what permits counsel to represent the file as the program’s actual findings.
- Reflex testing on defined change events, which rebuts the “should have known” beat of the plaintiff’s narrative.
- A complete, machine-readable dataset rather than a drawer of unconnected reports, so the record is analyzable — by the defense expert, and admittedly by the plaintiff’s — rather than curated.
- Corrective action on a confirmed exceedance, and published limits that ratchet tighter over time and never loosen, which document a company that responded rather than concealed.
- A mark disciplined against overclaiming — no “heavy-metal free,” no “safe,” a program that states it does not certify safety — which removes the misrepresentation hook instead of supplying a new one.

One structural point completes the picture, because a plaintiff will attack the standard as readily as the testing. A defensible program does not set its own limits to suit itself. HMTc derives its limits from a separately operated synthesis of the published literature and caps each at the lowest applicable government maximum, with the methodology published and reproducible; the evidence base and the certifier are kept architecturally distinct, so the standard is not the certifier’s self-published justification but a reference the certifier happens to operate. A standard that can show its work is a standard that survives being read aloud.

Built before, not during

The practitioner’s conclusion is the subtitle of this briefing, and it is meant literally. Defensible evidence is built before litigation, not during it. Every other exhibit in a heavy-metal case can be assembled after the complaint: the expert reports, the regulatory correspondence, the corporate witnesses. The contemporaneous testing record is the one exhibit that cannot. Its value lies in its date, and its date cannot be moved. A certificate obtained after the complaint is an argument the brand makes about itself; a surveillance record built before it is proof the brand does not have to argue.

That places the decision outside the ordinary sequence of litigation. By the time a matter reaches the defense, the record exists or it does not, and counsel’s options are fixed by a choice the client made years earlier. The useful advice, then, is not advice about how to defend the case. It is advice for the client who has no case yet: the testing record now mandated in California, required in Maryland, and pointed to by the direction of the federal action levels is going to exist regardless^[1], and the only question within the client’s control is whether it is built as a snapshot — ad hoc, discretionary, impeachable — or as a record — scheduled, independent, continuous — in the window that closes the day the complaint is served. Counsel who raise that question early are practicing the only kind of heavy-metal defense that reliably works: the kind conducted before there is anything to defend.

Karen Pendergrass is the Standards Architect of the Heavy Metal Tested & Certified program at the Paleo Foundation. She can be reached at karen@paleofoundation.com. This briefing is informational and is not legal advice; the legal characterizations in it are the Foundation’s and are offered for counsel’s independent evaluation.

References

[1] *In re: Baby Food Products Liability Litigation*, MDL No. 3101, U.S. District Court for the Northern District of California.

[2] Rule 702 / general-causation ruling, *In re: Baby Food Products Liability Litigation* (N.D. Cal., Dec. 2025) (excluding five of the plaintiffs’ six general-causation experts). Counsel should consult the order directly for its exact holdings and any subsequent developments.

[3] *Hain Celestial Group, Inc. v. Palmquist*, 607 U.S. ____ (No. 24-724), decided Feb. 24, 2026 (Sotomayor, J.; unanimous).

[4] *Palmer v. Hoffman*, 318 U.S. 109 (1943) (a report prepared in anticipation of litigation, rather than in the regular course of business, lacks the trustworthiness on which the business-records exception rests); the principle is carried into the trustworthiness clause of Fed. R. Evid. 803(6)(E).

[5] Fed. R. Evid. 803(6) (records of a regularly conducted activity), including the requirement that the record be kept and made as a regular practice and the opponent’s burden under 803(6)(E) to show a lack of trustworthiness.

[6] Cal. Evid. Code § 1271 (business-records exception; California is the forum of the consolidated baby-food litigation and of AB 899).

[7] Fed. R. Evid. 901 (authentication); ISO/IEC 17025:2017 accreditation and documented chain of custody as the means of authenticating an analytical result.

[8] ISO/IEC 17025:2017, *General Requirements for the Competence of Testing and Calibration Laboratories*; EURACHEM/CITAC and ILAC-G8 guidance on decision rules and statements of conformity under measurement uncertainty.

[9] U.S. House of Representatives, Committee on Oversight and Reform, Subcommittee on Economic and Consumer Policy, “Baby Foods Are Tainted with Dangerous Levels of Arsenic, Lead, Cadmium, and Mercury,” Staff Report, Feb. 4, 2021.

[10] K. Pendergrass, *HMTc Infant and Child Foods Program Manual*, 2026 Edition, and the companion *HMTc Lot Testing Schedule*, the Paleo Foundation, 2026 (risk-tiered surveillance frequency, baseline testing, independent sampling, chain-of-custody, reflex triggers, measurement-uncertainty decision rules, and the prohibition on testing into compliance). doi: [10.5281/zenodo.20270512](https://doi.org/10.5281/zenodo.20270512).

[11] California Assembly Bill 899 (2023) (baby-food heavy-metal testing and public disclosure); Maryland “Rudy’s Law,” HB 97 / SB 723 (2024) (testing and labeling); U.S. FDA, “Closer to Zero” action levels for lead and other elements in foods for babies and young children.