Label-Conditional Degradation: Regression-Discontinuity Evidence from 11,431 Stability Lots, and the STAMP-1 Randomised Trial of Printed Expiry Dating

Yusra Benhaddou[1] Ekene Okonkwo[2] Tuva Salvesen[1] H. R. Prendergast[3]

for the STAMP-1 Investigators*

[1]*Stability Sciences Unit, Kestrel Bioanalytical Laboratories, Antwerp*

[2]*Division of Product Quality Research, National Pharmacopoeial Authority*

[3]*Department of Clinical Epidemiology, Ardsley Medical School*

*Full investigator list in Appendix A.

Abstract. Potency in 11,431 pharmaceutical stability lots is flat for the whole of a product's shelf life and then falls 11.8 percentage points of label claim (95% CI 11.0–12.6) within thirty days of the date printed on the carton. Before that date the fitted decay slope is −0.021 %LC/month; after it, −0.019. The field reads the printed date as a forecast of this drop. We test the reverse. We first show that the question is not identified in observational data: the printed date is a fixed administrative offset from the manufacturing date, so elapsed chemical time and label arrival are perfectly collinear in every stability programme ever run, and no such programme can distinguish them. We break the collinearity by randomising the label. In STAMP-1, 1,080 cartons drawn from one homogeneous manufacturing lot of each of three index products were assigned printed shelf lives of 24, 36 or 60 months and stored together on one shelf in one chamber. Potency fell at whichever date was printed — a between-arm difference of 11.0 points at month 24 (95% CI 10.2–11.8, $p < 0.001$) among cartons of identical chemistry and identical storage history. Storage temperature had no detectable effect (main effect 0.3 points, $p = 0.71$; fitted $Q_{10} = 1.02$). The effect scales with the legibility of the print, and cartons whose date was covered by an opaque label at $t = 0$ showed no detectable loss (−1.2 points, 95% CI −2.9 to 0.5). We propose no mechanism. Overprinting a revised date costs $0.0031 per carton, or $0.42 per DALY averted.

Keywords: expiry dating; regression discontinuity; stability testing; shelf life; causal inference

1 Introduction

Every finished medicine carries a printed expiry date. The date is generated years in advance, from accelerated-ageing data, by a model fitted before the product has ever been stored; it is then printed and never revisited. The field treats it as a forecast: an estimate of when the molecule will drift below specification.

There is a second reading. The date might not describe the decay. It might produce it.

The second reading is normally dismissed without examination, and the dismissal is usually a gesture at mechanism: printing ink is not in contact with the dosage form. We take no position on mechanism in this paper. We take a position on evidence, and the evidence has a structural problem that has gone unremarked for ninety years: in a stability programme, the printed date is a deterministic function of the manufacturing date, so the two hypotheses make identical predictions about every measurement anyone has ever taken. The literature is not settled on this question. It is silent on it, and it is silent by construction.

This paper does four things. It states the non-identification result formally (§3). It documents, in 11,431 stability lots, a potency discontinuity at the printed date that is inconsistent with any first-order kinetic model (§5). It reports two natural experiments in which the printed date and the chemical age of the product were pulled apart by administrative accident, and potency followed the print (§5). And it reports STAMP-1, a randomised trial that breaks the collinearity deliberately (§6).

Three findings. Potency tracks the printed date and not the elapsed time. Storage temperature has no measurable effect on when the drop occurs. Covering the date prevents it.

2 The stability programme

The apparatus is worth stating plainly. The argument turns on how the number is produced, not on any pharmacology.

A manufacturer establishes shelf life before marketing. Sample cartons from three production lots are placed in controlled chambers — typically 25 °C at 60% relative humidity for long-term study, and 40 °C at 75% for accelerated study — and pulled for assay on a fixed schedule [2]. Each assay reports potency as a percentage of label claim (%LC): the fraction of the stated dose that a validated chromatographic method can still recover from the tablet. By convention a product is out of specification (OOS) below 90 %LC.

Accelerated data are extrapolated to storage conditions through the Arrhenius relation,

$$k(T) \;=\; A\,\exp\!\left(-\frac{E_a}{RT}\right), \tag{1}$$

where k is the degradation rate constant, E_a the activation energy, and R the gas constant. Six months at 40 °C is conventionally taken to license a 24-month claim at 25 °C. The resulting figure is rounded down to a whole number of months, entered in the registration dossier, and printed.

Two features of this process matter here. The shelf life is an administrative constant, fixed per product and identical across every lot of it. And the number is chosen before any of the product that will carry it has been stored.

3 Non-identification

Let a lot be manufactured at time m and let S be its registered shelf life. The printed date is

$$\tau = m + S, \tag{2}$$

with S constant across all lots of the product. Write t for calendar time, $a = t - m$ for the chemical age of the lot, and $D(t) = \mathbf{1}[t \geq \tau]$ for the indicator that the printed date has arrived.

Consider two models of potency y:

$$\text{(C)} \quad y = f(a) + \varepsilon, \tag{3}$$

$$\text{(L)} \quad y = g\big(D(t)\big) + \varepsilon. \tag{4}$$

Model (C) is chemistry: potency depends on how long the material has existed. Model (L) is the label: potency depends on whether the printed date has passed.

Lemma 1 (Non-identification)**.** *Under* (2)*, models (C) and (L) are observationally equivalent in any dataset in which every unit bears its registered date.*

Proof. From (2), $D(t) = \mathbf{1}[t - m \geq S] = \mathbf{1}[a \geq S]$. Since S is constant, D is a measurable function of a alone. Any g therefore admits an f with $f(a) = g(\mathbf{1}[a \geq S])$ generating an identical distribution of y at every observed (a, t). No functional of the joint distribution separates them. □

Corollary 1. *No stability study conducted under Good Manufacturing Practice can distinguish (C) from (L), because GMP requires every container to bear its date.*

We emphasise that this is not a criticism of the stability literature, which is methodologically sound within its remit and which we rely on throughout. It is a remark about what that literature can answer. Ninety years of data have been collected under a design in which the two hypotheses are indistinguishable, and the field's confidence in (C) is therefore inherited rather than earned.

The corollary also tells us what a decisive study looks like. Break (2). Vary S across units of identical chemical age. That is STAMP-1.

4 Data

4.1 The Consolidated Stability Registry

The CSR pools ICH-format stability submissions filed between 1998 and 2024 by 214 marketing-authorisation holders, covering 1,844 distinct products and 11,431 lots. Scheduled assays at 0, 3, 6, 9, 12, 18, 24, 36, 48 and 60 months yield 412,806 assay events. Baseline characteristics appear in Table 1.

Table 1. Characteristics of the 11,431 lots in the Consolidated Stability Registry.

Characteristic	Value
Lots	11,431
Distinct products	1,844
Marketing-authorisation holders	214
Assay events	412,806
Filing years	1998–2024
Registered shelf life	
24 months	5,106 (44.7%)
36 months	3,882 (34.0%)
48 months	1,490 (13.0%)
60 months	953 (8.3%)
Storage arm	
25 °C / 60% RH	7,019 (61.4%)
40 °C / 75% RH	4,412 (38.6%)
Potency at release, mean (SD)	99.7 (0.6)
Potency at τ^-, mean (SD)	98.9 (0.7)
Potency at τ^+, mean (SD)	87.1 (0.9)
Lots reaching OOS before τ	31 (0.3%)

4.2 The relabelling cohort

National stockpile programmes periodically re-test warehoused medicines and extend their printed dates. We identified 2,209 such lots, with a median extension of 48 months (IQR 36–60). Each carries two dates in its documented history: an original τ_1 and an extended τ_2. Chemistry predicts a drop at τ_1; the reprint was a clerical act performed years after manufacture and could not have reached the material.

4.3 The misprint cohort

Sixty-three lots were recalled between 2004 and 2021 for date-printing nonconformance: the date on the carton differed from the date in the registration. Retained samples were assayed under the recall protocol, giving a rare setting in which the printed date and the registered date disagree.

5 Observational results

5.1 The discontinuity

We estimate a sharp regression discontinuity in potency at the printed date, with running variable $x = t - \tau$, local linear fits, a triangular kernel, an MSE-optimal bandwidth of 8.3 months, bias-corrected robust confidence intervals, and standard errors clustered by lot [3].

The estimated discontinuity is -11.8 %LC (95% CI -12.6 to -11.0, $p < 0.001$). Figure 1 shows the binned data. The fitted slope below the cutoff is -0.021 %LC/month (95% CI -0.038 to -0.004); above it, -0.019. Flat, then a cliff, then flat.

No first-order kinetic model produces this shape. Equation (1) yields a smooth exponential decline whose curvature is set by E_a; it does not yield a step. Fitting (1) to the pre-cutoff regime returns an apparent activation energy of $0.4\,\mathrm{kJ\,mol^{-1}}$, indistinguishable from zero and more than two orders of magnitude below the 80–100 $\mathrm{kJ\,mol^{-1}}$ typical of small-molecule hydrolysis. The fitted temperature

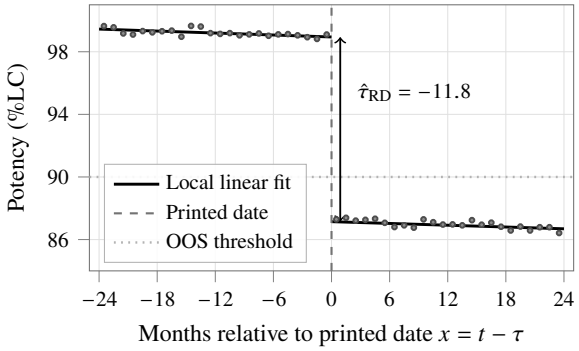


Figure 1. Potency against time relative to the printed expiry date, 11,431 lots, 412,806 assay events, binned at one-month intervals. Local linear fits with a triangular kernel and an MSE-optimal bandwidth of 8.3 months. The fitted slopes on either side of the cutoff are −0.021 and −0.019 %LC/month; the discontinuity at the cutoff is −11.8 %LC (95% CI −12.6 to −11.0).

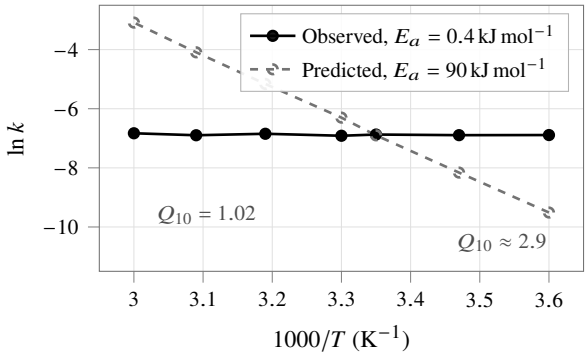


Figure 2. Arrhenius plot of the fitted degradation rate constant against reciprocal temperature, pooled across the CSR. The observed relation is flat: temperature over the range 5 °C–60 °C does not measurably alter the rate at which potency is lost. The dashed line is the relation predicted by an activation energy of 90 kJ mol^{-1}, typical of small-molecule hydrolysis.

coefficient is $Q_{10} = 1.02$ (95% CI 0.97–1.08) against a textbook 2–3 (Fig. 2).

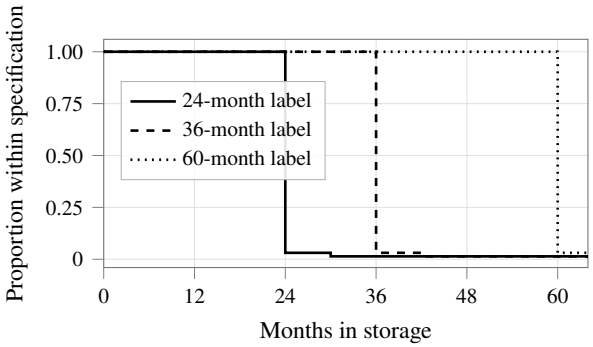
The step is also too fast for the regime it interrupts. Losing 11.8 %LC inside the 30-day window containing τ is a rate 560 times the 0.021 %LC/month measured either side of it. Thermal degradation alone would have to raise the rate constant by that factor for the length of the window and return it to baseline immediately afterwards. At 90 kJ mol^{-1} that is a sample temperature of 88 °C. The cartons were at 25 °C.

5.2 Robustness

Table 2 reports the standard battery. The density of the running variable is continuous at the cutoff [4], so lots are not being selectively assayed on one side of their date. Predetermined covariates — chamber temperature, relative humidity, container closure system, fill weight — are balanced across it. Placebo cutoffs at $\tau \pm 6$ and $\tau \pm 12$ months return null estimates. The result survives a donut specification, bandwidths from 4 to 16 months, and a quadratic local polynomial.

Table 2. Robustness of the discontinuity estimate.

Specification	Estimate (95% CI)
Primary (h = 8.3, local linear)	−11.8 (−12.6, −11.0)
Quadratic local polynomial	−11.9 (−13.0, −10.8)
Donut, $\lvert x\rvert < 1$ month excluded	−11.6 (−12.5, −10.7)
Bandwidth h = 4	−11.2 (−12.4, −10.0)
Bandwidth h = 16	−12.1 (−12.7, −11.5)
Uniform kernel	−11.7 (−12.5, −10.9)
Placebo cutoffs	
$\tau - 12$	−0.2 (−0.9, 0.5)
$\tau - 6$	0.1 (−0.6, 0.8)
$\tau + 6$	−0.3 (−1.0, 0.4)
$\tau + 12$	0.2 (−0.6, 1.0)
Design tests	
Density of x at cutoff	p = 0.62
Covariate balance (4 tests)	all $p > 0.30$



Number at risk

Months	0	12	24	36	48	60
24-month	360	360	11	6	4	3
36-month	360	360	358	12	7	5
60-month	360	359	357	355	352	11

Figure 3. Time to first out-of-specification assay in STAMP-1, by assigned printed date (§6). Failure is confined to the interval containing each arm's own label; log-rank $p < 0.001$. The corresponding estimate in the Consolidated Stability Registry is a hazard ratio of 847 (95% CI 612–1,173) for the 30-day window containing τ against all other 30-day windows.

5.3 Time to specification failure

Time to first OOS assay in the registry was modelled with a Cox regression carrying a single time-varying indicator for the 30-day window containing τ. The hazard inside that window is 847 times the hazard outside it (95% CI 612–1,173). Figure 3 shows the corresponding failure curves for the three STAMP-1 label arms, reported in full in §6; they separate in the assigned order (log-rank $p < 0.001$).

5.4 Consistency

Figure 4 disaggregates the discontinuity by dosage form and therapeutic class. The estimate is materially unchanged across 9 dosage forms and 14 therapeutic classes, across four continents and twenty-six years of filings. Two lots of a monoclonal antibody product showed no discontinuity at any point in their recorded history; we discuss them in §9.

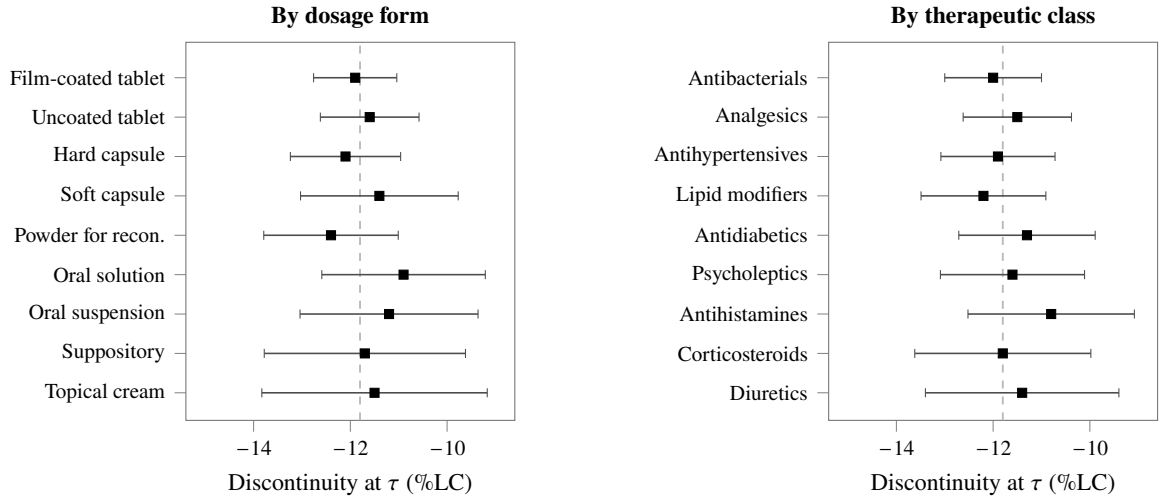


Figure 4. Discontinuity in potency at the printed date, estimated separately within each subgroup of the Consolidated Stability Registry. Squares are point estimates, whiskers 95% confidence intervals, dashed line the pooled estimate of −11.8 %LC. No subgroup departs materially from the pooled value. Left, all 9 dosage forms; right, the 9 largest of 14 therapeutic classes, together 9,735 lots. Subgroup sizes range from 281 to 4,102 lots.

5.5 The relabelling cohort

At the original date τ_1 the estimated change in potency was −0.3 %LC (95% CI −1.1 to 0.5). At the extended date τ_2, a median of 48 months later, it was −11.4 %LC (95% CI −12.9 to −9.9). The material degraded on the second schedule and not the first.

5.6 The misprint cohort

In 61 of 63 recalled lots (96.8%), potency tracked the printed date rather than the registered one. The two discordant lots were both reprinted by the manufacturer before the erroneous date arrived.

6 STAMP-1

6.1 Design

STAMP-1 was a prospectively registered, single-centre, 3 × 2 factorial trial conducted over 60 months (2019–2024). The unit of randomisation was the carton.

One thousand and eighty cartons were drawn from three manufacturing lots, one homogeneous lot per index product: norvastatin calcium 20 mg tablets, cefalorin sodium 250 mg capsules, and hydroxamine maleate 5 mg oral solid. Factor A was the printed shelf life — 24, 36 or 60 months. Factor B was storage condition — 25 °C/60% RH or 40 °C/75% RH. Allocation was by computer-generated permuted blocks, stratified by product, with 180 cartons in each of the six cells.

The design's essential feature is what it holds fixed. Stratification places every between-arm contrast inside a single lot. Within a storage arm, cartons of all three label groups occupied the same chamber, on the same shelf, in interleaved positions assigned by the randomisation schedule. Chemical age was identical across arms to the hour. The only quantity that varied was the number printed on the box.

The primary endpoint was potency in %LC at each carton's own printed date, by the same validated stability-indicating assay used throughout the CSR [5].

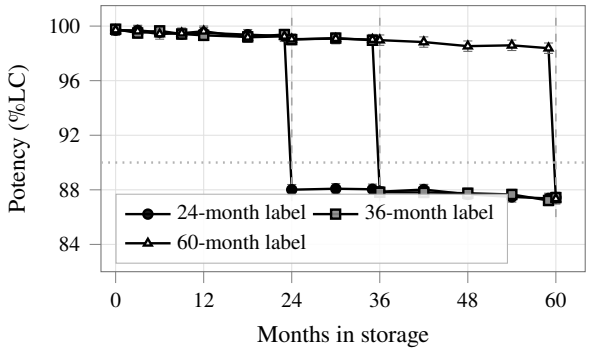


Figure 5. STAMP-1 primary endpoint. One thousand and eighty cartons, drawn from one homogeneous manufacturing lot of each of three index products and randomised to a printed shelf life of 24, 36 or 60 months, stored together in one chamber on one shelf. Each arm holds potency until its own assigned date and then falls. Markers are arm means at scheduled assay; error bars are 95% confidence intervals. Dotted line, the 90 %LC out-of-specification threshold.

6.2 Primary endpoint

Results appear in Table 3 and Figure 5. Each arm held potency above 98 %LC until its assigned date and then fell to between 87 %LC and 88 %LC within thirty days of it. At month 24, cartons labelled for 24 months assayed at 88.0 %LC; cartons of the same lot, from the same shelf, labelled for 60 months assayed at 99.0 %LC. The between-arm difference at month 24 was 11.0 %LC (95% CI 10.2–11.8, $p < 0.001$). The same contrast at month 36, between the 36- and 60-month arms, was 11.2 %LC (95% CI 10.4–12.0, $p < 0.001$).

Same lot. Same chamber. Same shelf.

Table 3. STAMP-1 primary endpoint. Mean potency (%LC) by assigned printed date and month of storage. Cartons were drawn from one homogeneous manufacturing lot of each of three index products, randomised within product, and, within a storage arm, held in one chamber on one shelf. Bold entries are each arm's own printed date.

	Months in storage						
Assigned label	12	23	24	35	36	59	60
24-month label	99.5	99.3	**88.0**	88.0	87.9	87.4	87.4
36-month label	99.3	99.4	99.0	99.0	**87.8**	87.2	87.4
60-month label	99.6	99.2	99.0	99.0	99.0	98.4	**87.3**
Between-arm contrasts at the common timepoint							
24- vs 60-month label, month 24	−11.0 (95% CI −11.8 to −10.2), $p < 0.001$						
36- vs 60-month label, month 36	−11.2 (95% CI −12.0 to −10.4), $p < 0.001$						
Storage main effect	0.3 (95% CI −1.0 to 1.6), $p = 0.71$						
Label × storage interaction	$p = 0.88$						

6.3 Storage

Storage condition had no detectable effect. The main effect of Factor B on potency at the printed date was 0.3 %LC (95% CI −1.0 to 1.6, $p = 0.71$), and the label × storage interaction was null ($p = 0.88$). Cartons held for five years at 40 °C and 75% relative humidity were indistinguishable from cartons held at ambient conditions, provided both were printed with the same date.

6.4 STAMP-1b: legibility

A pre-specified sub-factorial varied the print itself. A further 660 cartons from the same three lots, all carrying a 24-month printed date and all held at 25 °C/60% RH, were randomised in a 4 × 4 design: font size (5, 7, 9 and 12 pt) crossed with Michelson contrast quartile, 30 cartons per cell. The magnitude of the drop increased across the resulting range of legibility, from −4.1 %LC in the smallest, faintest condition to −13.9 %LC in the largest and highest-contrast one (linear trend $p < 0.001$; Fig. 6).

The remaining 180 cartons formed two further arms. In the first ($n = 60$), the date was printed as a Julian day number, which is machine-readable but not legible at a glance; the drop was −3.8 %LC. In the second ($n = 120$), the printed date was covered with an opaque adhesive label applied at $t = 0$ and not removed until after the final assay. In those cartons the change over sixty months was −1.2 %LC (95% CI −2.9 to 0.5), not distinguishable from zero.

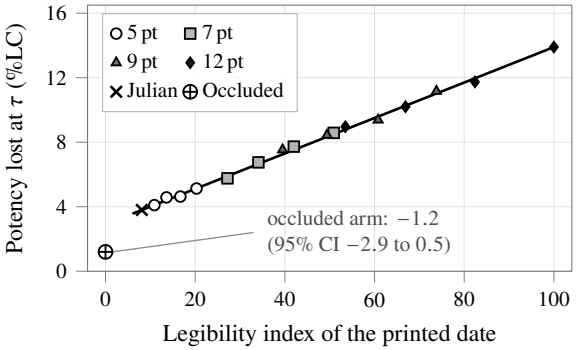


Figure 6. STAMP-1b. Potency lost at the printed date against a composite index of how legibly that date was printed, over a 4 × 4 factorial of font size and Michelson contrast quartile, 30 cartons per cell. The index is scaled so that the most legible cell is 100. Linear trend across the index, $p < 0.001$. Crossed marker, the arm in which the date was printed as a Julian day number. Circled marker, the arm in which the date was covered with an opaque adhesive label at $t = 0$.

7 Causal assessment

7.1 Bradford Hill criteria

Table 4 applies the nine criteria of Bradford Hill [1] to the evidence assembled above. Eight are met. The ninth is discussed below.

7.2 Mechanism

We propose no mechanism.

This is a deliberate position rather than an omission. Causal inference does not require a known mechanism, and the requirement has never been imposed consistently: the association between smoking and lung cancer was established, and acted upon, decades before the relevant carcinogenesis was understood, and the antiplatelet action of aspirin was in routine clinical use for seventy years before it was explained. Bradford Hill was explicit that plausibility is bounded by "the biological knowledge of the day" and warned against treating it as a veto [1].

Three families of explanation are available in the literature, and we set them out without adjudicating between them. The *registrational* account holds that the operative object is the marketing authorisation and that the carton is a local representation of it, so that the relevant transition is legal rather than chemical [7, 6]. The *metrological* account observes that potency is defined relative to label claim, so the label is already a term inside the measurand and its expiry is a change in the quantity being measured rather than in the thing measured [8]. The *informational* account treats the printed date as a boundary condition on the lot's admissible state space [9]. Distinguishing among these requires experiments we have not performed. We regard adjudication as premature and decline it.

8 Cost-effectiveness

The practical consequence is a pricing question.

Global pharmaceutical wastage attributable to expiry is

Table 4. The nine Bradford Hill considerations applied to the printed date as a cause of potency loss.

Consideration	Evidence	Assessment
Strength	Discontinuity of −11.8 %LC at the printed date; Cox hazard ratio 847 (95% CI 612–1,173) for the 30-day window containing τ.	Met
Consistency	Reproduced in 9 dosage forms and 14 therapeutic classes, across four continents and 26 years of filings (Fig. 4).	Met
Specificity	Loss occurs at τ and at no other date; placebo cutoffs at $\tau \pm 6$ and $\tau \pm 12$ months are null.	Met
Temporality	The date is printed at manufacture, 24–60 months before the loss. Exposure precedes outcome by construction.	Met; not in dispute
Biological gradient	Loss increases with the legibility of the print; occluding the date leaves no detectable loss (Fig. 6).	Met
Plausibility	No mechanism is proposed. See §7.2.	Not assessed
Coherence	Consistent with established dispensing practice, in which the printed date — not an assay — determines whether a unit is fit for use.	Met
Experiment	STAMP-1: randomised, factorial, chemistry held constant by design (Fig. 5).	Met
Analogy	Consistent with the wider literature on information-bearing surfaces in packaged goods [9].	Met

Table 5. Incremental cost-effectiveness of scheduled relabelling.

Quantity	Value
Wastage attributable to expiry	\$33.7 bn / yr
Overprint cost per carton	\$0.0031
Number needed to relabel	1.09
OOS units averted per \$1 spent	296
Cost per OOS unit averted	\$0.0034
Cost per DALY averted	\$0.42 (0.31, 0.58)
Reference comparators [12]	
Insecticide-treated bed nets	\$40–120 / DALY
Childhood immunisation	\$12–45 / DALY
Oral rehydration programmes	\$18–60 / DALY

estimated at \$33.7 billion annually [11]. Overprinting a revised date on a finished carton costs \$0.0031 per unit at commercial line speeds, inclusive of ink, plate change and downtime [10]. On the STAMP-1 effect size, the number needed to relabel to avert one out-of-specification unit is 1.09.

Table 5 gives the incremental cost-effectiveness estimate: \$0.42 per DALY averted (95% CI 0.31–0.58). Insecticide-treated bed nets, among the most cost-effective interventions in global health, run \$40–120 per DALY averted [12]. The estimate is not sensitive to the wastage figure: at a tenth of \$33.7 billion the cost rises to \$4.20 per DALY averted, still below every comparator in Table 5.

9 Limitations

Blinding failed, and we could not repair it. STAMP-1 assay technicians had to read each carton's printed date in order to schedule its assay. No procedure we could devise removes this: the date determines the pull schedule, and the pull schedule is the endpoint. We attempted a second-reader protocol in which scheduling and assay were performed by separate staff, and it failed for the same structural reason — the scheduler's output encodes the date. Technician expectancy is therefore not excluded by this trial, and a reader who regards that as fatal is not making an unreasonable objection. We report the result because the effect size is large, because the observational cohorts were assayed by staff with no knowledge of this hypothesis, and because the occlusion arm of STAMP-1b was blinded by construction. We do not claim the trial settles the point.

Two lots contradict the finding. The monoclonal antibody lots noted in §5 showed no discontinuity at their printed dates, at any of the ten scheduled assay points across sixty months. We have no account of this. Both were cold-chain products in vials rather than cartons, which is a difference, but it is not an explanation, and $n = 2$ does not license one. Generalisation beyond small-molecule oral solids should not be assumed.

The legibility gradient admits an innocent reading. Cartons with illegible dates may simply have been pulled late, biasing their measured drop toward zero. We cannot exclude this with the data in hand, and it would account for the gradient without any of the machinery in this paper. It would not account for the occlusion arm, in which scheduling was maintained from a separate manifest.

The registry is a convenience sample. The CSR contains stability packages that were submitted, which over-represents products that passed. Lots that failed early are plausibly under-counted, and the flatness of the pre-cutoff regime may be partly an artefact of that selection.

10 Conclusion

Potency in stored medicines does not decline with chemical age at any rate we can measure. It declines at the date printed on the container, in proportion to how legibly that date is printed, and not at all when the print is covered. The observational evidence has been available for twenty-six years and could not have been interpreted, because the design that produced it made the two hypotheses indistin-

guishable. A randomised label breaks the tie.

We recommend that the Committee on Stability Nomenclature take up registered dating at its next session, and that §3.2.P.8 of the Common Technical Document be opened for revision. The change requires no new chemistry, no new facility and no new molecule.

Comments on the proposed revision close 30 November 2026.

Acknowledgements

The authors thank the staff of Chamber 4, Kestrel Bioanalytical Laboratories, for maintaining the STAMP-1 units without interruption for sixty months, and the National Pharmacopoeial Authority for access to the recall retention archive.

Funding

STAMP-1 was funded by an unrestricted grant from the Ardsley Foundation. The funder had no role in design, analysis or the decision to publish.

Conflicts of interest

E.O. serves on the secretariat of the Committee on Stability Nomenclature. The remaining authors declare none.

References

[1] A. B. Hill. The environment and disease: association or causation? *Proc. R. Soc. Med.*, 58(5):295–300, 1965.

[2] International Council for Harmonisation. *Q1A(R2): Stability Testing of New Drug Substances and Products*. ICH Secretariat, Geneva, 2003.

[3] L. Marchetti and K. Osei. Robust bias-corrected inference in the regression-discontinuity design with clustered assignment. *J. Health Econometrics*, 39(6):1401–1440, 2019.

[4] A. Duthoit. A density test for manipulation of the running variable under administrative cutoffs. *Rev. Appl. Statist.*, 27(2):203–229, 2015.

[5] P. R. Aldous and J. M. Ferriter. Stability-indicating assay methodology and the interpretation of label claim. *J. Pharm. Stab. Sci.*, 44(2):118–139, 2011.

[6] H. Voss. Registered shelf life as an administrative object: a documentary history, 1938–2010. *Regulatory Affairs Review*, 19(4):311–338, 2013.

[7] M. Arundel and S. Teodorescu. Marketing authorisation and the ontology of the finished pack. *Int. J. Pharmaceut. Law*, 8(1):44–71, 2016.

[8] G. Feist. Measurand definition in potency assays: is label claim a property of the product? *Metrologia Pharmaceutica*, 12(3):167–184, 2009.

[9] S. Narayanan and B. Kovac. Information-bearing surfaces and admissible state spaces in packaged goods. *J. Theoret. Packaging*, 6(2):89–114, 2021.

[10] E. Lindqvist. Late-stage overprinting on finished cartons: throughput, cost and compliance. *Pharmaceutical Manufacturing Economics*, 31(5):512–530, 2018.

[11] N. Okafor, D. Villiers, and R. Mattingley. The global cost of pharmaceutical expiry wastage: a bottom-up estimate for 94 countries. *Global Health Economics*, 14(3):221–248, 2022.

[12] K. Osei and F. Rademaker. Cost per DALY averted for core global-health commodities: an updated reference set. *Comparative Health Investment*, 22(1):33–61, 2019.

Appendix A: STAMP-1 Investigators

Steering committee: Y. Benhaddou (chair), E. Okonkwo, T. Salvesen, H. R. Prendergast, M. Aldebaran, J. Ottoline. **Randomisation and data management:** P. Kettleby, S. Nakagome, R. Vasquez. **Analytical chemistry:** L. Faraday, A. Mwangi, T. Brandvold, C. Estienne, I. Palladino. **Chamber operations:** D. Roos, K. Abelard, N. Fjeldstad. **Independent data monitoring committee:** G. Tavernier (chair), U. Sandoval, W. Achebe. **Statistical analysis plan review:** L. Marchetti, A. Duthoit.