How Well Does Arrhenius Extrapolation Predict Food Shelf Life? A Hold-Out Test Across 113 Studies

Bryan Quoc Le, PhD · Mendocino Food Consulting · Nov 2, 2026 · @Bryan Quoc Le

Abstract

This report evaluates the predictive validity of Arrhenius extrapolation as applied in accelerated shelf-life testing (ASLT) of foods. Particular focus is placed on the extrapolation from elevated test temperatures to the intended storage temperature. That step underpins most shelf-life claims, yet it is rarely verified at the storage temperature itself.

A retrospective hold-out design was applied to published storage studies reporting quality-loss rates at three or more isothermal temperatures. For each series, the coolest temperature was withheld, an Arrhenius model was fitted to the remaining temperatures, and the extrapolated rate was compared with the measured rate. The dataset comprises 409 series from 113 studies identified through OpenAlex.

The analysis demonstrates that Arrhenius extrapolation predicted the withheld rate within a factor of 2 in 76% of series. The median absolute error was a factor of 1.33, and no systematic bias was detected (study-weighted shelf-life ratio 1.08, 95% CI 0.96–1.25). Accuracy increased to 86% for series without data-quality flags and to 90% for series at low risk of bias.

Predictive error depended chiefly on extrapolation distance. When the withheld temperature lay more than one fitted range below the test temperatures, shelf life was over-predicted by a factor of 1.39 (95% CI 1.08–1.90, high risk-of-bias series excluded). Predictions within that distance were unbiased (1.02, 0.91–1.15). This contrast was identified post hoc and is presented as exploratory. Of 14 studies reporting their own validation, nearly all validated within the fitted temperature range.

Overall, the results indicate that Arrhenius-based ASLT provides reliable estimates near the tested temperature range. The principal risk lies in long extrapolations to refrigerated or ambient storage, a regime that published validations seldom examine.

Keywords: Accelerated shelf-life testing; Arrhenius equation; extrapolation; hold-out validation; food quality kinetics; meta-research

1. Introduction

Shelf-life claims for most packaged foods are derived from prediction rather than direct observation. In accelerated shelf-life testing (ASLT), a product is stored at elevated temperatures, and the rate of change in a limiting quality attribute is measured at each temperature. That rate is then extrapolated to the intended storage temperature by means of the Arrhenius relationship [1, 2]:

```
k = A \exp\left(-\frac{E_a}{RT}\right)
```

The validity of this procedure rests on a single structural assumption: That the same mechanism, with the same activation energy, governs deterioration across the entire temperature span. Phase transitions, glass transitions, moisture redistribution, chilling injury and shifts in the dominant spoilage organism may each violate this assumption. Labuza and co-workers cautioned that extrapolation across such transitions is unreliable [2, 3]. Nevertheless, ASLT predictions are rarely verified against a rate measured at the storage temperature itself.

Two recent reports address related questions. Gorbachev et al. [4] re-analyzed oxidation data for ten vegetable oils and concluded that very high test temperatures (70–120 °C) under-estimate shelf life. Antonov [5] showed by simulation that the conventional regression-based estimator used in pharmaceutical stability testing systematically over-estimates shelf life, and that hierarchical models correct this upward bias. Neither study, however, tests extrapolation performance across the breadth of published food storage data.

Many published storage studies report rates at three or more isothermal temperatures, which permits a retrospective test. The coolest temperature is withheld, the Arrhenius model is fitted to the remaining temperatures, and the extrapolated rate is compared with the measured rate.

This report applies that hold-out design across the food literature. It addresses three questions:

1. How accurate is Arrhenius extrapolation from higher to lower storage temperatures, and is it biased toward over- or under-prediction of shelf life?
2. Which design features (extrapolation distance, number of fitted temperatures, accelerated-only designs) and which deterioration mechanisms are associated with larger errors?
3. Do published validations examine the extrapolation for which ASLT is used in practice?

2. Methods: Search, screening and retrieval

2.1 Information sources and search

OpenAlex was searched on 28 October 2026; Scopus and Web of Science were not accessible. An initial query centered on the term "Arrhenius" retrieved only one of seven studies known to meet the inclusion criteria, since many storage studies do not use that term. The final search therefore combined that query with a broadened query, applied to titles and abstracts from 2000 onward:

```
("shelf life" OR "storage temperature") AND (Arrhenius OR "activation energy"
OR Q10 OR accelerated OR kinetics)
```

The broadened query retrieved six of the seven known studies. Study flow is reported in Section 5.

2.2 Screening

Screening proceeded in two stages:

- **Stage 1:** Records were classified as likely include, possible include or exclude by rules calibrated against the known studies.
- **Stage 2:** Likely includes were screened manually on title and abstract.
- **Second-screener check:** The author independently assessed a sample of 34 stage-2 decisions. Agreement was 97% (33 of 34). The single disagreement led to respiration rate being accepted as a temperature-dependent rate, and 25 respiration studies were re-included.

2.3 Prioritization and retrieval

Candidate records were assigned to priority tiers for full-text extraction:

- **P1:** An accelerated prediction and a measured real-time shelf life were both reported.
- **P2:** At least four storage temperatures, including one at or below 25 °C.
- **P3:** Either of the P2 features.
- **R:** All remaining records.

The present analysis comprises tiers P1 and P2. Tier P3, a random sample of tier R and the stage-1 possible includes were not extracted.

Full texts were obtained automatically where open access permitted, and otherwise by the author through free and institutional library access. Records for which no full text could be obtained are reported as not retrieved.

3. Methods: Data extraction and quality control

3.1 Unit of analysis and eligibility

The unit of analysis is the series, defined as one food product × one quality response with a rate measure at three or more isothermal storage temperatures.

Three rate measures were accepted:

- A rate constant of any reaction order, including the maximum specific microbial growth rate (μmax) and the respiration rate
- A half-life, converted as $k = \ln 2 / t_{½}$
- A measured shelf life or time to a quality limit, converted as $k = 1 / t$

Series were excluded under any of the following conditions:

- The temperatures were processing rather than storage temperatures (drying, pasteurization, cooking, retorting).
- Rates were reported only as Arrhenius-model output, or only graphically.
- The material was not a food, or was a model system (buffers, pure solutions, pigment extracts in buffer).
- The temperature range crossed 0 °C.
- The response showed no significant change during storage, so no rate could be estimated.
- The response changed in opposite directions at different temperatures. Dropping the sign in that case would have converted a gain into a loss.

3.2 Data extraction

All qualifying series in a study were extracted, irrespective of whether they appeared consistent with Arrhenius behavior. Where only time-course data were tabulated, zero- or first-order rate constants were estimated by least squares, and the order and data points used were recorded.

Extraction was performed with the assistance of large language model agents (Claude, Anthropic). Agents worked in batches of 9–12 studies, under written extraction rules and a fixed record format. Each batch was subjected to two checks. An automated validator tested format, identifier uniqueness, the three-temperature minimum, non-positive rates and ranges crossing 0 °C. Records were then reviewed for consistency, and all review decisions were logged.

3.3 Quality flags

Each series was assigned one data-quality flag:

- `ok`: No issue identified.
- `non_monotonic`: The rate did not increase with temperature at one or more steps.
- `poor_fit`: The reported coefficient of determination was below 0.6, or values were reported to one significant figure.
- `inconsistent_tables`: Reported values conflicted between tables or with the text.
- `equilibrium_term`: The kinetic model included an equilibrium or plateau term.

3.4 Verification of extraction

Twelve studies were selected at random, and up to three series per study (28 series in total) were verified against the source documents by the author. All 28 transcribed values were correct. Two series, however, contained a sign-handling error introduced by an earlier version of the extraction rules. A search of the extraction records identified two further instances. All four series were excluded, and the rule was revised (Section 3.1, final criterion).

4. Methods: Hold-out analysis and statistics

4.1 Hold-out procedure and effect size

For each series, the coolest temperature was withheld. The linearized Arrhenius model, ln k = ln A − Ea/RT, was fitted by ordinary least squares to the remaining temperatures, and the fitted line was extrapolated to the withheld temperature. The effect size is the base-10 logarithm of the ratio of predicted to observed rate:

```
e = \log_{10}\left(\frac{k_{pred}}{k_{obs}}\right), \qquad \frac{SL_{pred}}
{SL_{obs}} = 10^{-e}
```

A positive e denotes over-prediction of the rate and hence under-prediction of shelf life (the conservative direction). A negative e denotes over-prediction of shelf life (the non-conservative direction).

4.2 Summary measures

Accuracy was summarized by three measures:

- The median absolute error, expressed as a fold-error (10^|e|)
- The proportion of series with |e| ⩽ log10 2 (within a factor of 2)
- The proportion of series in which shelf life was over-predicted

Bias was estimated as the study-weighted mean of e. Series were first averaged within each study, so that studies contributing many series did not dominate the estimate. Confidence intervals (95%) were obtained by a non-parametric bootstrap that resampled studies, with 5,000 resamples for the main analysis and 2,000 for subgroups. Bias is reported as the geometric-mean shelf-life ratio, 10^(−ē).

4.3 Extrapolation distance and subgroups

Extrapolation distance was defined in reciprocal-temperature space, relative to the width of the fitted range:

```
r = \frac{1/T_{hold} - 1/T_{fit,\,min}}{1/T_{fit,\,min} - 1/T_{fit,\,max}}
```

Here T_fit,min and T_fit,max are the lowest and highest fitted temperatures. A value of r ⩽ 1 indicates that the prediction extends no farther below the fitted range than the range is wide. Series were further grouped by withheld temperature: chilled (⩽ 10 °C), ambient (< 30 °C) and accelerated-only (all temperatures ⩾ 30 °C).

4.4 Risk of bias

Each series was rated on four domains as low risk, some concerns or high risk:

Domain	Low risk	Some concerns	High risk
D1 Rate derivation	Reported rate constant, half-life, or authors′ raw data	Rate fitted in this study, or inverse shelf life	—

Domain	Low risk	Some concerns	High risk
D2 Data quality	Flag `ok`	`non_monotonic`	`poor_fit`, `inconsistent_tables`, `equilibrium_term`
D3 Fit design	Three or more fitted temperatures	Two fitted temperatures (linearity untestable)	—
D4 Reporting	Complete	Unit or order assumed, illegible source, censored values	—

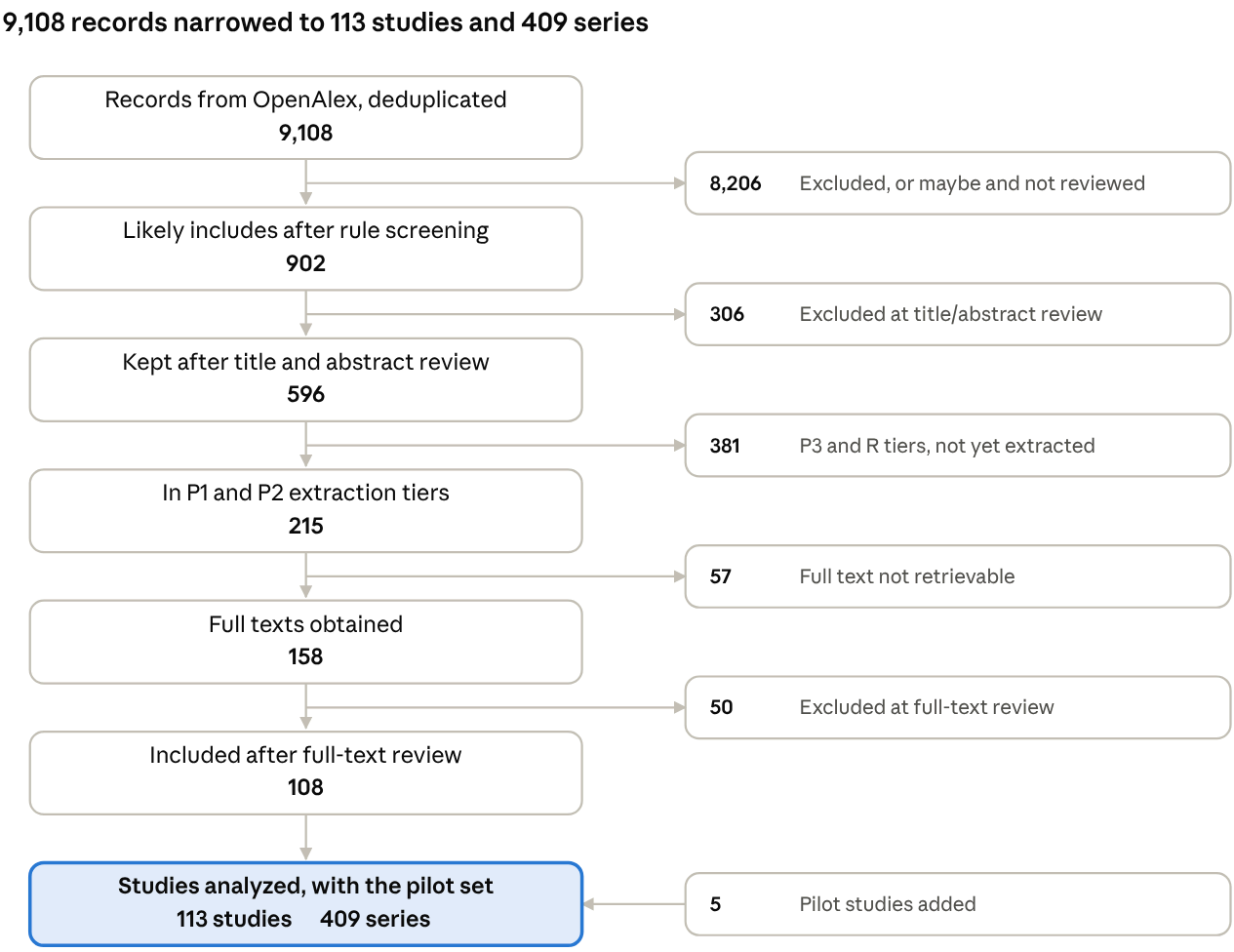
The overall rating was high if any domain was rated high, some concerns if any domain raised concerns, and low otherwise. Ratings were assigned algorithmically from the extraction fields and notes. Sensitivity analyses were performed on the low-risk subset and on the subset excluding high-risk series, together with a leave-one-study-out analysis.

4.5 Validation set

Fourteen included studies reported a validation at a real-time or out-of-fit temperature. For these studies, the authors' own predicted-versus-observed pairs at temperatures outside the fitted range were extracted and expressed on the same shelf-life ratio scale.

5. Results: The dataset

The final dataset comprises 409 series from 113 studies, totaling 1,493 rate measurements. The principal reduction occurred at rule-based screening, and 57 prioritized records could not be retrieved in full text (Figure 1).



Study flow · OpenAlex search Oct 28, 2026, extraction through Nov 2, 2026 · title/abstract exclusions are net of 25 respiration studies re-included

Figure 1. Study flow.

The series span a broad range of matrices, mechanisms and temperatures:

- **Matrices:** Fresh produce (61 series), dry powders (50), prepared dishes (36), juices and beverages (28), fresh fish (25), fresh-cut produce (22) and confectionery (21), with 13 further matrix classes.
- **Mechanisms:** Sensory acceptance (70), microbial growth (64), lipid oxidation (63), color change (43), texture (28), moisture loss (27), anthocyanin degradation (23) and phenolic oxidation (22).
- **Withheld temperatures:** A median of 10 °C, ranging from −20 °C (frozen series) to 62 °C. Of these, 202 were chilled, 114 ambient and 93 accelerated-only hold-outs.
- **Fit design:** 229 hold-outs relied on a two-point fit and 180 on three to seven points.

- **Data quality:** 245 series (60%) were flagged `ok`. The remainder were flagged `non_monotonic` (67), `poor_fit` (49), `inconsistent_tables` (35) or `equilibrium_term` (13).

6. Results: Overall accuracy and bias

Arrhenius extrapolation was generally accurate and showed no overall bias. The median absolute error was a factor of 1.33, and errors in the conservative and non-conservative directions occurred with similar frequency (Figure 2, Table 1).

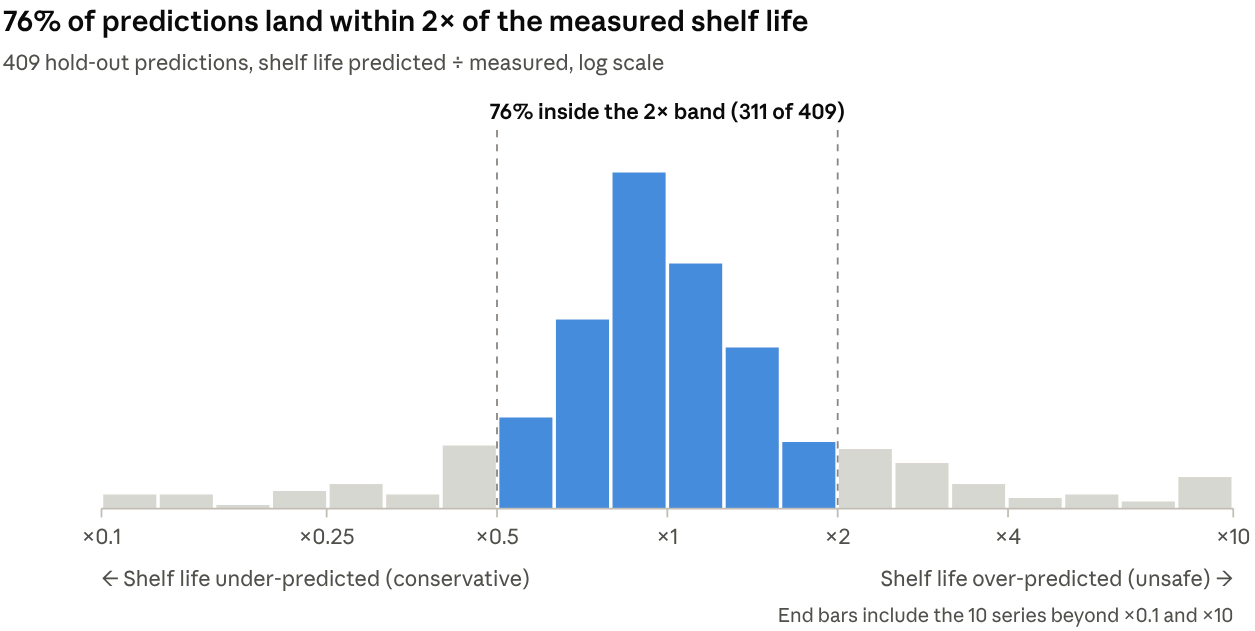


Figure 2. Distribution of hold-out prediction errors. Table 1. Accuracy and bias by data subset.

Subset	Series	Studies	Median miss	Within 2×	Shelf-life ratio (95% CI)
All series	409	113	×1.33	76%	1.08 (0.96–1.25)
Quality flag `ok`	245	89	×1.29	86%	1.07 (0.97–1.19)
Low risk of bias	94	31	×1.28	90%	1.04 (0.94–1.15)

The shelf-life ratio is the study-weighted geometric mean of predicted to observed shelf life; values above 1 denote over-prediction. In each subset, the 95% confidence interval includes unity.

The error distribution is heavy-tailed. Ten series exhibited errors exceeding a factor of 10. Nine of these carried a data-quality flag (`poor_fit` or `non_monotonic`). Several, including the largest (a powder redispersibility series), are attributable to probable unit or transcription inconsistencies in the source data rather than to kinetic failure. The single unflagged case is an anthocyanin series with a questionable 55 °C value. The median error and the within-2× proportion are robust to these outliers, and most are removed in the quality-flag subset.

7. Results: What drives the misses

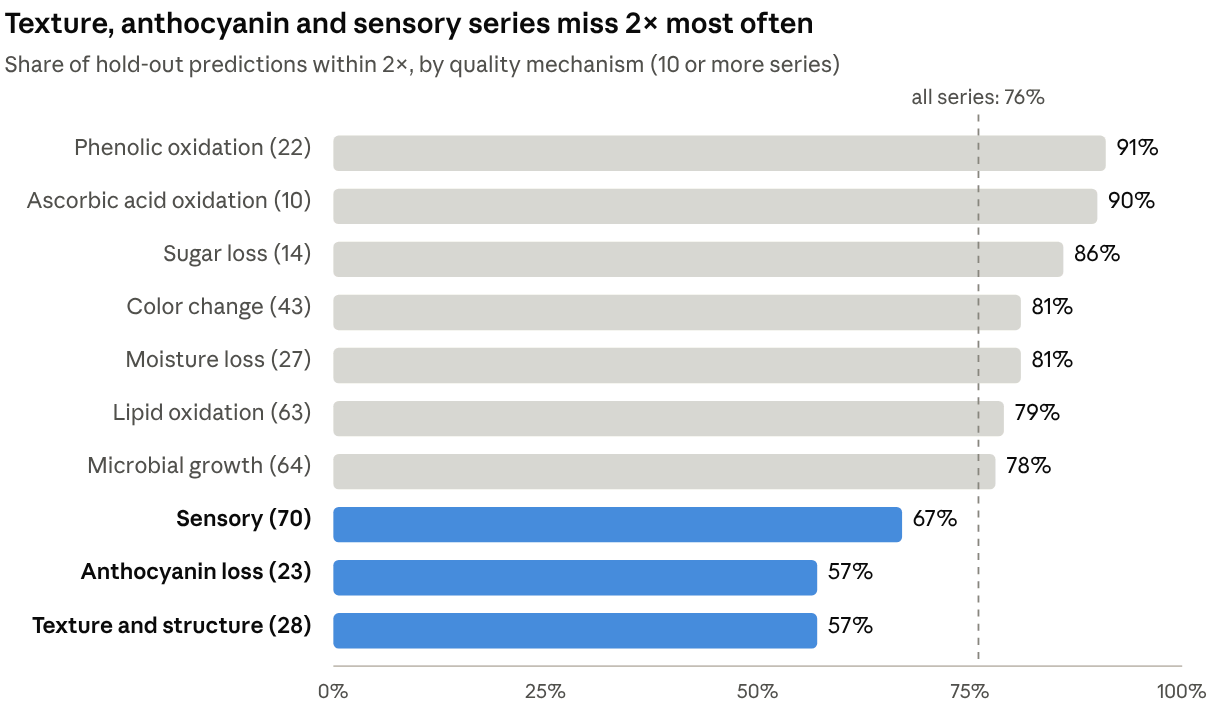
Extrapolation distance was a stronger determinant of error than the number of fitted temperatures (Table 2). Accuracy also varied with the deterioration mechanism: Physical and sensory responses extrapolated less reliably than single chemical reactions (Figure 3).

Table 2. Accuracy by design feature.

Design feature	Series	Studies	Median miss	Within 2×	Shelf life over-predicted
Extrapolation ≤ 1× the fitted range	198	58	×1.29	83%	50%
Extrapolation > 1× the fitted range	211	58	×1.44	70%	43%
2 fitted temperatures	229	64	×1.36	72%	42%
3 fitted temperatures	122	36	×1.30	80%	52%
Accelerated-only (all ≥ 30 °C)	93	28	×1.30	81%	49%

Number of fitted temperatures. The within-2× proportion increased from 72% with two fitted temperatures to 80% with three. Series with four fitted temperatures reached 94% (35 series, 11 studies), whereas series with five or more reached 60% (20 series, 4 studies). Given the small and clustered samples at the upper end, no monotonic trend can be inferred. An earlier analysis of a smaller subset had suggested that three or more temperatures ensured accuracy; that pattern was not reproduced in the full dataset.

Accelerated-only designs. Where all temperatures were 30 °C or above, shelf life was over-predicted by a factor of 1.15 on average (95% CI 0.97–1.38). The direction is non-conservative, but the effect is not statistically distinguishable from zero. A significant effect of 1.23 observed in the smaller subset was likewise not sustained.



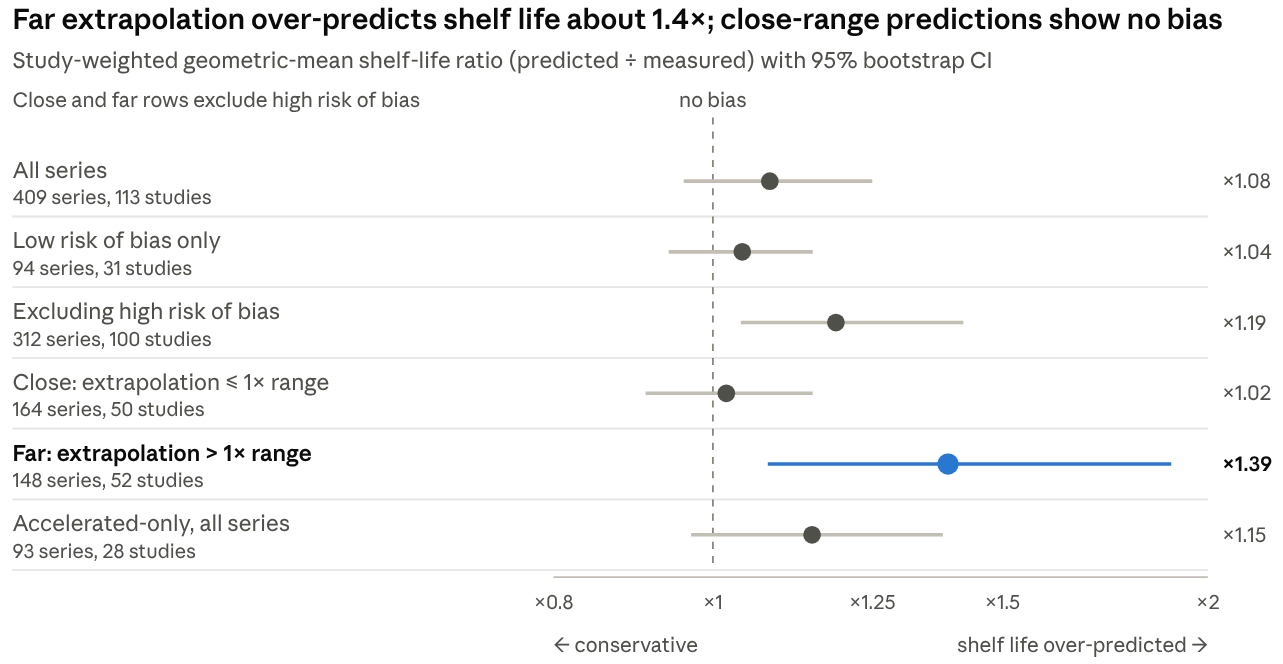
breakdown.py · 409 series · mechanisms with fewer than 10 series omitted

Figure 3. Proportion of predictions within 2×, by deterioration mechanism.

Deterioration mechanisms. Well-defined single reactions were extrapolated most reliably: phenolic oxidation (91% within 2×) and ascorbic acid degradation (90%). Texture, anthocyanin degradation and sensory acceptance were extrapolated least reliably (57–67%). Each of these responses is plausibly subject to a temperature-dependent change in mechanism. For texture, these include fat melting and protein or starch transitions; for anthocyanins, pH- and oxygen-dependent degradation pathways; and for sensory scores, the integration of several attributes. Respiration and browning reached 100%, but each was represented by only eight series.

8. Results: Risk of bias and the validation set

Restriction to higher-quality data sharpened rather than altered the principal findings. Low-risk series were accurate and unbiased, and the non-conservative signal was concentrated among predictions extending far beyond the fitted range (Figure 4).



bias_scoring.py · 2,000 study-level bootstrap resamples · log scale

Figure 4. Shelf-life ratio by subset, with 95% bootstrap confidence intervals.

8.1 Risk-of-bias sensitivity analyses

- Of 409 series, 94 (31 studies) were rated low risk, 218 some concerns and 97 high risk.
- Low-risk series reached 90% within 2×, with no detectable bias.
- Exclusion of high-risk series yielded an overall shelf-life ratio of 1.19 (1.04–1.42). This effect was attributable to long extrapolations (1.39, 1.08–1.90), whereas short extrapolations were unbiased (1.02, 0.91–1.15).
- Series flagged `non_monotonic` showed the largest over-prediction of shelf life (1.72). The effect is partly structural: A rate that does not decline at the withheld temperature necessarily produces an over-prediction. It nonetheless reflects a genuine, non-conservative behavior of food systems.
- Only 9 low-risk accelerated-only series (3 studies) were available, which precludes inference for that subgroup.
- In the leave-one-study-out analysis, the overall ratio ranged from 1.03 to 1.10, indicating that no single study determined the result.

The contrast between long and short extrapolations was identified during subgroup exploration and was not specified in advance. Although consistent with mechanistic expectation, it should be regarded as a hypothesis requiring confirmation in an

independent dataset.

8.2 Validation set

Fourteen included studies reported a validation of their own predictions.

- In the present hold-out test, series from these studies performed no better than the remainder: 73% within 2× (60 series) versus 77% (349 series).
- The authors' own predicted-versus-observed pairs appeared more favorable. From 27 pairs in six independent studies, the median error was a factor of 1.13, 85% were within 2×, and the shelf-life ratio was 0.90 (mildly conservative).
- Nearly all of these validations were interpolations, with the validation temperature within the fitted range. Only one study, on vacuum-packaged scallops fitted at 20–40 °C, examined a true downward extrapolation (to 4 and 15 °C); its predictions were within 4% of the observed shelf life.
- A further four studies described as validated re-examined only the temperatures used for fitting.

The published literature therefore seldom tests the long extrapolation for which ASLT is employed in practice.

9. Discussion and practical guidance

The results support the use of Arrhenius extrapolation in food development, within defined limits. Across 113 studies, the withheld storage rate was predicted within a factor of 2 in three of four series, and in nine of ten where data quality was high. No systematic tendency toward over- or under-prediction of shelf life was detected.

The principal weakness arises in the regime where the method is most frequently applied: Long extrapolation from elevated test temperatures to refrigerated or ambient storage. Where the prediction extended more than one fitted range below the test temperatures, shelf life was over-predicted by a factor of approximately 1.4 on average. This is the non-conservative direction.

This pattern is consistent with mechanistic expectation. Greater extrapolation distance increases the opportunity for the governing chemistry or physics to change. Relevant transitions include lipid crystallization, glass transitions, moisture redistribution, the emergence of psychrotrophic organisms and chilling injury. Each alters the temperature dependence in a manner that cannot be detected from high-temperature data alone. The

poorer performance of physical and sensory responses (texture, sensory acceptance, anthocyanin color) admits the same interpretation, since these responses integrate several processes whose relative contributions shift with temperature.

These findings complement two recent reports. Gorbachev et al. [4] concluded that very high test temperatures (70–120 °C) lead to under-estimation of shelf life in vegetable oils. The present dataset was drawn predominantly from storage-type tests between 30 and 60 °C, a distinct regime, and the two results are therefore not in conflict. Antonov [5] demonstrated by simulation that conventional regression-based estimators over-estimate shelf life; the empirical over-prediction observed here at long extrapolation distances is consistent with that bias.

The validation findings are of comparable importance. Most published validations were conducted within the fitted temperature range, which represents the least demanding case. A statement that an ASLT model has been validated therefore provides limited assurance about performance at the storage temperature on which the label date depends.

9.1 Implications for practice

1. **Extrapolation distance should be assessed alongside the number of test temperatures.** Where the gap between the lowest test temperature and the storage temperature exceeds the width of the tested range, the predicted shelf life should be regarded as optimistic.
2. **The lowest test temperature should approach the storage temperature.** A 25 °C condition for ambient products, or a 10–15 °C condition for chilled products, reduces extrapolation distance more effectively than an additional elevated temperature.
3. **Real-time storage at the target temperature should run in parallel.** Early comparison with real-time data tests the prediction in the relevant regime and can reveal model failure well before the claimed shelf life elapses.
4. **Particular caution is warranted for texture, sensory and anthocyanin endpoints**, for which departures from Arrhenius behavior were most frequent.
5. **Non-monotonic temperature dependence should be treated as a warning rather than noise.** A deterioration rate that does not decline at lower temperatures is precisely the behavior that leads ASLT to over-predict shelf life.
6. **The scope of a reported validation should be established.** Validation at a temperature within the tested range does not confirm a shelf-life claim at the storage temperature.

10. Limitations

The principal limitations concern search coverage, the use of AI-assisted extraction, and the post hoc identification of the extrapolation-distance effect.

- **Search.** Only OpenAlex was searched, as Scopus and Web of Science were not available. The search retrieved six of seven known eligible studies. Abstract coverage varies by publisher, and citation searching was not performed.
- **Coverage.** Only tiers P1 and P2 were extracted. Fifty-seven prioritized records could not be retrieved. Tier P3, the tier R sample and 2,743 stage-1 possible includes were not assessed. Studies reporting rates only graphically were excluded rather than digitized.
- **Extraction.** Extraction was AI-assisted, with a single human reviewer.
  - All 28 transcribed values in the verification sample were correct.
  - Verification identified a rule defect in sign handling that affected four series; this has been corrected.
  - Sign reversals not recorded in the extraction notes cannot be excluded without re-examination of the sources.
  - Inclusion decisions were made by one reviewer, apart from the 34-record second-screener check.
- **Pre-specification.** The analysis was not pre-registered. The extrapolation-distance effect emerged from subgroup exploration and is presented as a hypothesis.
- **Risk-of-bias assessment.** Ratings were assigned algorithmically from extraction fields and notes, rather than by item-level human appraisal.
- **Hold-out design.**
  - Withholding the coolest temperature left 229 of 409 series with a two-point fit, in which linearity cannot be assessed.
  - Shelf lives reported in whole days or at sampling intervals introduce measurement error.
  - Conversion of shelf life to a rate assumes a constant quality limit across temperatures.
- **Heterogeneity.** Studies differ in matrix, packaging, endpoint and kinetic order. The bootstrap treats studies as exchangeable, and a formal multilevel meta-regression was not fitted.
- **Validation set.** Only six independent studies contributed usable predicted-versus-observed pairs, and only one involved true extrapolation.

11. Conclusions

This report examined the predictive validity of Arrhenius extrapolation in accelerated shelf-life testing across 409 series from 113 food storage studies. Extrapolation from elevated to lower storage temperatures predicted the withheld rate within a factor of 2 in 76% of series overall and 90% of low-risk series, with no overall bias.

Predictive error depended primarily on extrapolation distance rather than on the number of test temperatures. Long extrapolations over-predicted shelf life by a factor of approximately 1.4, whereas short extrapolations were unbiased. Texture, sensory and anthocyanin endpoints were least reliable. Published validations rarely tested the long-range extrapolation on which shelf-life claims depend.

Overall, Arrhenius-based ASLT remains a defensible tool when test temperatures are close to the intended storage temperature and when real-time data are collected in parallel. Confirmation of the extrapolation-distance effect in an independent, pre-registered dataset is recommended.

12. Data, code and references

All data and code supporting this report are available from the author on request (bryan@mendocinofoodconsulting.com). The files are listed in Table 3.

Companion resources. A plain-language summary for product developers, *When Can You Trust Your Accelerated Shelf-Life Test?*, is available at tools.mendocinofoodconsulting.com/aslt-check/whitepaper. A free calculator that applies these results to a specific test design, the ASLT Extrapolation Check, is available at tools.mendocinofoodconsulting.com/aslt-check.

Table 3. Data and code files.

File	Contents
`rates_combined.csv`	1,493 rate measurements in 409 series, with study, DOI, matrix, response, mechanism, temperature, rate, unit, source and quality flag
`holdout_results.csv`	One row per series: fit range, held-out temperature, extrapolation ratio, Ea, effect size

File	Contents
`rob_scores.csv`	Risk-of-bias domains and overall rating per series
`validation_pairs.csv`	Authors' own predicted-versus-observed pairs
`p1_fulltext_log.csv`, `p2_fulltext_log.csv`	Every full-text decision with its reason
`spec_*.txt`, `HELPER_INSTRUCTIONS.md`	Raw extraction records and the extraction rules
`holdout_analysis.py`, `breakdown.py`, `bias_scoring.py`, `validation_pairs_analysis.py`	Analysis code (Python, numpy and pandas)
`openalex_search.py`, `screen_stage1.py`, `prioritize.py`, `SEARCH_LOG.md`	Search, screening and prioritization

Use of AI tools. Large language model agents (Claude, Anthropic) were used, under the author's direction, for screening rules, data extraction, analysis code and drafting of this report. The author determined all inclusion criteria, scope and interpretation, performed the second-screener and extraction verification checks, and takes full responsibility for the content.

Funding and competing interests. This work received no external funding. The author is a consultant in food product development and advises clients on shelf-life testing.

References (bibliographic details of [1]–[3] to be verified before deposit)

1. Arrhenius S. Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren. *Zeitschrift für Physikalische Chemie* 4 (1889): 226–248.
2. Labuza TP, Schmidl MK. Accelerated shelf-life testing of foods. *Food Technology* 39(9) (1985): 57–64, 134.
3. Fu B, Labuza TP. Shelf-life testing: Procedures and prediction methods. In: Erickson MC, Hung YC (eds.), *Quality in Frozen Food*. Chapman & Hall, 1997.
4. Gorbachev VV, Belyaeva MA, Smirnov AA. High temperatures when using ASLT underestimate the shelf life of food. *Bulletin of Donetsk National University, Series A: Natural Sciences* 2026; (2): 47–56. doi:10.5281/zenodo.20799672

5. Antonov L. Calibration and diagnostics of a Bayesian model for vaccine shelf life estimation based on a linearized Arrhenius equation. Working paper, Zenodo, 2026. Zenodo record 21931045
6. Included studies: The complete list of 113 studies, with titles, journals, DOIs and series counts, is available as included-studies.csv. Study-level rate data are provided in `rates_combined.csv` and the full-text decision logs.