Checking 213 OEIS recurrence conjectures against their b-files finds one failure, already known

A preregistered mechanical sweep of conjectured linear recurrences in the On-Line Encyclopedia of Integer Sequences, with planted errors and a cross-entry control

Zain Dana Harper

Version 1.0, 9 November 2026. Finding F308 in the findings ledger. Honest null (tier D).

Abstract

Many entries in the On-Line Encyclopedia of Integer Sequences carry a line such as "Conjecture: a(n) = 3*a(n-1) - 3*a(n-2) + a(n-3)". Most such lines were found by fitting the terms known at the time, and the b-file stored beside the entry often lists hundreds or thousands more. We asked how often these conjectures fail on their own b-files and whether the failures are data errors or counterexamples. From the public OEIS data, 23,883 entries mention a conjecture and 1,179 have a line our parser accepts; we drew a random 500 and tested every line that its b-file extends by at least 30 terms. One of 213 tested conjecture lines failed on its b-file, 0.47 percent, with a 95 percent Wilson interval of 0.08 to 2.6 percent. The one failure, A031740, holds for 64 indices and fails at n = 65, and the entry has recorded the conjecture as false since 2016, so the sweep found no new counterexample. Planted errors were caught in 212 of 212 passing lines, and equations applied to other entries' terms passed at a mean rate of 0.005. The sampling frame changed after the preregistered route failed, and the change is recorded. The result says the check is cheap and, on this kind of line, low in yield.

1 What this means for a reader

When an OEIS entry says "Conjecture" in front of a simple recurrence, the line has usually held far past the terms it was fitted on. In our sample, 212 of 213 such lines held for a median of 967 further terms, and the one that failed had already been marked false by a later contributor. Read such a line as a well-tested guess. It still has no proof behind it, and holding for a thousand more terms does not supply one.

2 Introduction

The OEIS is a public database of integer sequences, online since 1995 and holding over 300,000 entries when Sloane described its recent history in 2018 [2]. Each entry lists the first terms of a sequence in a short data section, with formulas, comments and references contributed by many people. Beside most entries sits a b-file, a plain list of index and term pairs that often runs to thousands of terms.

Some entries carry a conjecture line: an equation for the sequence, offered without proof. Many are a linear recurrence with integer coefficients, found by fitting the terms then listed. Fitting of this kind is a standard method in computer algebra. Kauers and Koutschan [3] describe guessing a recurrence from finitely many terms by linear algebra, note that the overdetermination of the linear system is one source of trust in a guess, and show how to guess from fewer terms by restricting the coefficients. In a later search across the OEIS [4] they found conjectured recurrences for sequences not previously known to satisfy one, proved some, and showed that a recurrence exists for others without proving the guessed one.

Our question was smaller. The b-files already hold the data a mechanical check needs. We asked what share of conjecture lines fail on their own b-files, and whether each failure is a data error or a true counterexample. The

decision it informed was whether such lines are worth a routine check as a cheap source of counterexamples, and how to read one when it appears in our own work. Our searches found the guessing literature above and no published sweep of this exact kind; that is what the searches returned, and it does not show that none exists.

3 Preregistration and design

The preregistration was written and hash-locked on 31 October 2026, before any entry was sampled beyond one probe request that returned an unrelated entry. It names the sampling route, the parser rules, the definition of tested and of failed, the planted errors, the control, the forecasts and the kill conditions.

Kill conditions, without discretion:

- The run is void if fewer than 99 percent of planted errors are caught, if fewer than 100 lines are tested, or if the real pass rate is at or below the control's 95th percentile.
- A failure counts only after a second check that the failing term is not a known b-file error.
- Nothing here proves the passing recurrences; a passing line stays a conjecture with more range.

Each kill condition and forecast became a rule set whose pass or fail is a verdict read off by code.

The locked plan sampled entries through the OEIS search pages at random offsets. That route returned only 10 entries, because the site asks for a sign-in past the first 100 search results, and we did not work around it. The first run tested 2 lines and was void under the first kill condition. Having seen only those 2 lines, both passing, we changed the sampling frame: the public OEIS data repository, every entry with a parsable conjecture line, then a seeded random 500. The change is recorded as a deviation with its reason and what had been seen. The new frame is wider than the locked one, and the second run used the same locked plan, rules and forecasts.

4 Data and provenance

The data are the OEIS entries from the public oeis/oeisdata repository and the b-files from oeis.org [1], retrieved on 31 October 2026. OEIS content is licensed CC BY-SA 4.0 under the OEIS End-User License Agreement. We used it as data for checks; this paper reports counts and quotes one entry's conjecture and comment lines with attribution.

Of all entries, 23,883 mention a conjecture and 1,179 have a conjecture line our parser accepts. The seeded random 500 entries held 531 lines that mention a conjecture and a term a(n), and 512 of those parsed. Only about 5 percent of the entries that mention a conjecture had a line our parser accepts: 1,179 of 23,883. The result therefore describes simple linear recurrences, not OEIS conjectures in general.

5 Methods

Parsing

A line was parsed only if it states an equation in n and earlier terms $a(n-k)$ with integer coefficients. A pattern match finds the candidate equation, and a parser whitelist accepts it only if every part of its syntax tree is an integer, n, a term $a(n-k)$, addition, subtraction, multiplication or a power. Anything else was skipped and counted.

Testing

For each sampled entry, the first conjecture line that parsed and qualified was tested; at most one line per entry was tested. A line qualified when the b-file holds at least 30 more terms than the entry's data section and the equation can be evaluated at 10 or more indices. The equation was evaluated at every index in the b-file's range where all the terms it needs exist. A line fails if its two sides differ at any such index; the first failing index

and the number of indices it held for were recorded. This left 213 tested lines, with a median of 967 b-file terms beyond the data section.

The failure rate

With k failures among N tested lines, the failure rate is $f = k/N$. Its Wilson interval [5] at $z = 1.96$ has centre c and half-width w:

$$D = 1 + \frac{z^2}{N} \tag{1}$$

$$c = \frac{1}{D}\left(f + \frac{z^2}{2N}\right) \tag{2}$$

$$w = \frac{z}{D}\sqrt{\frac{f(1-f)}{N} + \frac{z^2}{4N^2}} \tag{3}$$

6 Controls and planted checks

The planted check raised the first integer coefficient of each passing equation by one and ran the check again. A planted line should fail. Of the 213 tested lines, 212 passed, so 212 planted versions were built. All 212 were caught.

The cross-entry control applied each equation to a different, randomly chosen entry's terms, up to index 200. One draw takes 30 equations at random, applies each to another entry's terms, and records the share that pass. This share is the pass rate of an equation on a sequence it was not written for, the chance level of the test. Over 200 draws the mean share was 0.0048, and the 95th percentile of this null distribution was 0.033, one equation in 30.

The real pass rate, 0.995, was above every draw. The p value from draws is

$$p = \frac{k+1}{n+1} \tag{4}$$

with k the number of draws at or above the real pass rate and n the number of draws. Here $k = 0$ and $n = 200$, so $p = 0.005$, the smallest value 200 draws can give.

7 Results

Figure 1 follows the counts from the whole corpus to the one failure.

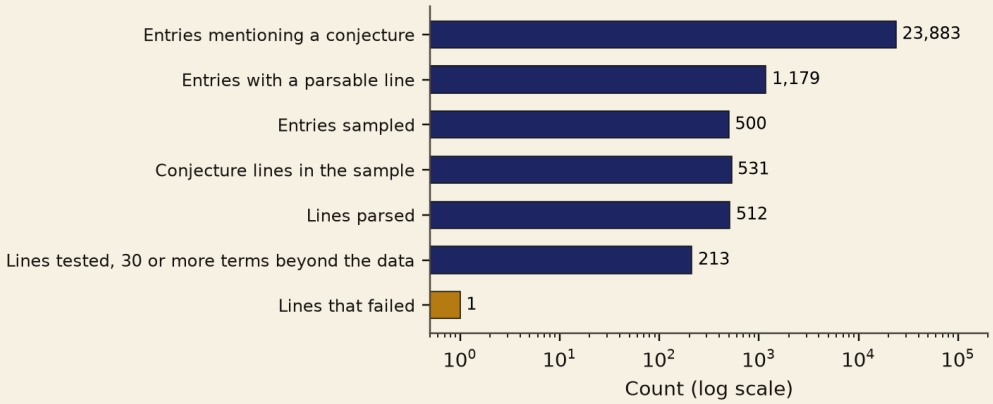


Figure 1. From the whole OEIS corpus to the one failing line. The first two counts are from the frame count file; the rest are from the run's result file.

One of 213 tested conjecture lines failed on its b-file, 0.47 percent, with a 95 percent Wilson interval of 0.08 to 2.6 percent. Figure 2 places the rate against the forecast of 4 percent and the preregistered 1 percent rule.

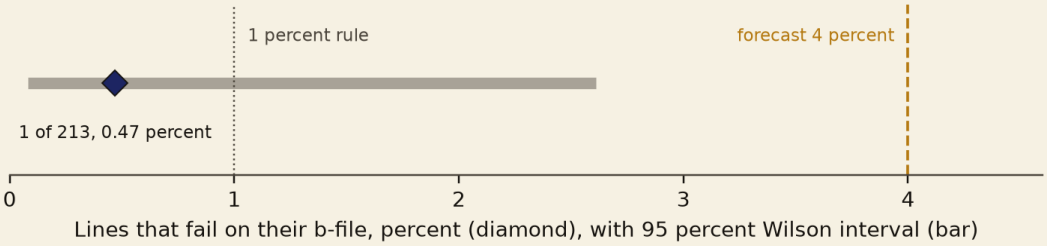


Figure 2. Failure rate with its 95 percent Wilson interval, against the 1 percent rule and the preregistered forecast of 4 percent. Drawn from the run's result file and the preregistration.

The failing line is in entry A031740. Its conjecture, contributed by Colin Barker on 30 September 2012, reads:

$$a(n) = n(961n + 1) \tag{5}$$

$$a(n) = 3a(n - 1) - 3a(n - 2) + a(n - 3) \tag{6}$$

The entry's data section has 32 terms and its b-file 10,000. The one failure, A031740, holds for 64 indices and fails at n = 65, and the entry has recorded the conjecture as false since 2016, so the sweep found no new counterexample. The stored copy of the entry carries a comment by Chai Wah Wu dated 19 June 2016: the conjecture is false, and 3940288 is a term not of the form $m(961m + 1)$. The preregistration asked for a second code path before a failure counts; the result file records no such rerun. The check by hand was reading this comment, which confirms the failure is a true counterexample and not a b-file error.

Table 1 gives the counts and rates.

Table 1. Counts and rates from the run.

Quantity	Value
Entries mentioning a conjecture	23,883
Entries with a parsable conjecture line	1,179
Entries sampled	500
Conjecture lines in the sample, parsed	531, 512
Lines tested	213
Median b-file terms beyond the data section	967
Lines that failed	1 (A031740, first failure at n = 65)
Failure rate, 95% Wilson interval	0.47 percent, 0.08 to 2.6 percent
Planted errors caught	212 of 212
Cross-entry pass rate, mean and 95th percentile	0.0048, 0.033
Real pass rate, p from 200 draws	0.995, 0.005

Table 2 sets the forecasts written before the run against the outcomes.

8 Discussion

The sweep shows that a mechanical check of OEIS conjecture lines is cheap: the analysis took about 6 seconds once the data were stored. On simple linear recurrences it is also low in yield. The lines held at a rate of 99.5 percent, the failure rate came in an order of magnitude under the forecast, and the single failure had been caught by a contributor four years after it was posted.

Two features of the lines explain much of this. They were fitted to data, so passing on further terms is what a

Table 2. Forecasts fixed in the preregistration against what the run gave.

Quantity	Forecast	Outcome
Lines parsed	300	512
Lines tested	150	213
Failure rate	0.04	0.0047
Pass rate	0.96	0.995
Planted errors caught	1.0	1.0
Cross-entry pass rate	0.001	mean 0.0048, 95th percentile 0.033
Share of failures from data errors	0.5	0 of 1
At least one failure not from a data error	probability 0.5	one, already recorded in the entry; none new

correct fit does. And a recurrence of order d fitted to a few more than d terms is close to determined by them, so the guess is either right or wrong in a way that shows quickly; A031740's conjecture failed just past twice the length of its data section. Kauers and Koutschan [3] point to the same property, overdetermination, as a source of trust in a guessed recurrence. Our pass rate is a measurement of how well that trust held on one class of OEIS line.

The more interesting conjectures are the ones our parser skipped: 95 percent of the entries that mention a conjecture state something our whitelist cannot read; we did not classify what those lines say. A sweep of those would need a different checker for each kind of line.

9 Limitations

- The lines were fitted to data, so passing is what a correct fit does, and a high-order fit can pass by construction when the b-file is short relative to its order. A passing line stays a conjecture with more range.
- Only about 5 percent of conjecture entries parse under our rules, so the rate describes simple linear recurrences with integer coefficients.
- Many lines come from a few contributors, so the rate also describes their habits.
- The sampling frame changed after the locked route failed. A later review of our ledger marked this as narrowing the claim. The new frame was chosen after seeing 2 lines, both passing.
- The preregistered second code path for a failure was not recorded; the single failure was confirmed by the entry's own comment.
- Scoring was by our own code and was not blind.

10 Verified

Verified: on 9 November 2026 the frame counts (23,883, 1,179 and 500) were re-read from the frame count file, and the 531 lines, 512 parsed and 213 tested, the median of 967 terms, the single failure with its first failing index of 65 after 64 indices, the data and b-file lengths, the Wilson interval, the 212 of 212 planted catches, the cross-entry mean, percentile and p value, and the three verdicts were re-read from the run's result file. They match the findings ledger. The 2012 conjecture and the 2016 comment marking it false were read in the stored copy of entry A031740. The preregistration's SHA-256 matches the hash recorded in the result file and in the lock file. Both figures were drawn by a script that reads only the result file, the frame count file and the preregistration. The study was not re-run for this paper. The literature was found by searches run on 9 November 2026 and read at the level of abstracts, arXiv pages and search summaries; the full texts were not read.

11 Corrections

- 31 October 2026: a review of our ledger recorded that the sampling changed after the search route refused requests past 100 results without sign-in, and that fitted lines pass by construction while only 5 percent of conjecture entries parse. Both narrow the claim and are stated here.
- 9 November 2026: first published version. The cross-entry control is described here as the code runs it, 30 equations per draw, each applied to another entry's terms; an earlier summary of ours described it as 30 terms at a time.

12 Reproduction

- Data: the public oeis/oeisdata repository and the b-files from oeis.org (CC BY-SA 4.0).
- Find every entry with a comment or formula line that contains "Conjecture" and a term a(n); keep entries with a line that the whitelist accepts; draw a seeded random 500.
- For each sampled entry, test the first qualifying line: at least 30 b-file terms beyond the data section and at least 10 indices where the equation can be evaluated. Record the first failing index.
- Compute the Wilson interval, eqs. 1 to 3. Raise the first integer coefficient of each passing equation by one and rerun. Build the cross-entry control from 200 draws of 30 equations and compute eq. 4.
- Preregistration SHA-256: 9736ef8a01a18ddaa94949f06de0b127eb9bdc9085bf9265fbb1d33a061f9d37. Plan SHA-256: f8fbd127e431e832302cdd548c974029630a778f7c14125e20711b5a5ffb045f. Run 20261031T223144Z; the earlier void run is 20261031T220825Z.
- Cost: the analysis took about 6 s and the control under 1 s on one desktop CPU, after the data were downloaded.

13 Appendix A. The preregistered verdicts

Each verdict is a set of rules written before the run. A verdict passes only when every rule holds. Values are read from the run's result file. The first three rules are shared by all three verdicts.

Table 3. Preregistered verdicts and the rules that decided each one.

Verdict	Deciding rule	Value	Holds
Instrument works (not void)	planted errors caught at least 0.99; at least 100 lines tested; pass rate above the control's 95th percentile (0.033)	1.0; 213; 0.995	yes
Any failure	the three rules above, and at least one failure	1	yes
Failure rate above 1 percent	the three rules above, and failure rate above 0.01	0.0047	no
First run, search-page frame	the instrument rules	2 lines tested	no (void)

14 Glossary

Each term is defined where it is first used in the text. The same definitions, with the field each comes from and what each excludes, are collected here.

- **OEIS**. The On-Line Encyclopedia of Integer Sequences, a public database of number sequences in which each entry carries the first terms, formulas, comments and links contributed by many people. Field: mathematics (reference databases). Excludes a peer-reviewed journal; a line in an entry is a contribution, and a conjecture line in it carries no proof.
- **b-file**. A text file stored beside an OEIS entry that lists the sequence's index and term pairs much further than the entry's own data section, often to thousands of terms. Field: the OEIS. Excludes the entry's data section, which holds only the first few dozen terms.

- **conjecture line**. A comment or formula line in an OEIS entry that states an equation for the sequence and labels it a conjecture; most were found by fitting the terms then known. Field: the OEIS, our usage. Excludes a line marked as proved, and a statement with no equation.
- **linear recurrence**. An equation that gives each term of a sequence as a fixed weighted sum of a set number of earlier terms, possibly with terms in n; here with integer coefficients. Field: mathematics (combinatorics). Excludes an equation with a product of earlier terms, and an equation with coefficients that are not integers.
- **counterexample**. A single case where a general statement is false; one is enough to refute a conjecture. Field: mathematics (logic). Excludes a term that disagrees because the data are wrong.
- **preregistration**. A dated written plan for a study, fixed and hash-locked before any result is seen: the question, the data, the statistic, the controls, the forecasts and the kill conditions. Field: research method (metascience). Excludes a plan revised after looking at the outcome, and a protocol written only for the record after the run.
- **void**. The verdict of a run whose planted check failed or whose sample fell below its stated minimum; it says nothing for or against the claim. Field: research method (our usage). Excludes a null result, which comes from a working instrument.
- **kill condition**. A rule written in the preregistration that, when met, decides against the claim with no room for judgement after the run. Field: research method (our usage, after Popper's falsification). Excludes reasons found after the run for setting a result aside.
- **verdict**. The pass or fail of one preregistered rule set against the run's numbers, read off by code. Field: research method (our usage). Excludes an interpretation written after reading the numbers.
- **sampling frame**. The list from which a sample is drawn; a result describes the frame, so a change of frame changes what the result is about. Field: survey statistics. Excludes the sample itself.
- **deviation**. A dated, recorded change from the preregistered plan, stated with its reason and with what had been seen when it was made. Field: research method (metascience). Excludes a change made silently, and a change made after seeing the outcome it would affect without saying so.
- **parser whitelist**. A rule that turns each line into a syntax tree and accepts it only if every part is on a fixed list: integers, n, the terms a(n-k), addition, subtraction, multiplication and powers. Field: computer science (parsing). Excludes any line that needs division, sums, named functions or words to state.
- **Wilson interval**. A 95 percent interval for a proportion, found by inverting the score test; it stays inside 0 to 1 and behaves well when the count is small or near zero. Field: statistics. Excludes the simple interval of the proportion plus or minus 1.96 standard errors, which fails near zero.
- **planted check**. Data built so that the right answer is known, run through the same analysis before the real data, to show that the analysis can see the effect when it is there. Field: research method; called a positive control in experimental science. Excludes any check made on the real data, which cannot show what the analysis misses.
- **control**. A version of the data with the structure under test removed (for example shuffled or pooled), scored the same way as the real data, so the real score has something to beat. Field: statistics and experimental design. Excludes a comparison with a different dataset or with a published number.
- **null distribution**. The spread of scores a control gives over many random draws; its 95th percentile is the bar a real score must clear. Field: statistics. Excludes a theoretical distribution assumed rather than drawn from the data.
- **p value (from draws)**. The share of control draws scoring at or above the real score, with one added to the count and to the number of draws, so it is never zero. Field: statistics. Excludes a p value from a textbook formula that assumes a distribution.
- **findings ledger**. The append-only, dated list of every result we have recorded, numbered F1, F2 and on; corrections are added as new dated lines and old entries are never rewritten. Field: our records. Excludes

a curated list of successes.

15 References

[1] The OEIS Foundation. The On-Line Encyclopedia of Integer Sequences, entries and b-files, and the oeis/oeisdata repository. Retrieved 31 October 2026. Licence CC BY-SA 4.0. https://oeis.org

[2] Sloane, N. J. A. (2018). The On-Line Encyclopedia of Integer Sequences. Notices of the American Mathematical Society 65(9), 1062-1074. doi:10.1090/noti1734. arXiv:1805.10343

[3] Kauers, M. and Koutschan, C. (2022). Guessing with little data. Proceedings of ISSAC 2022, 83-90. arXiv:2202.07966

[4] Kauers, M. and Koutschan, C. (2023). Some D-finite and some possibly D-finite sequences in the OEIS. Journal of Integer Sequences 26, article 23.4.5. arXiv:2303.02793

[5] Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association 22(158), 209-212. doi:10.1080/01621459.1927.10502953