Annals of Science and Inquiry
Vol. 2 No. 6 (2026)

Retained Answer Timelines Behind a Mature Acceptance Snapshot

Arthur Bennett[1] · Timothy Young[2] · Gavin Bell[3]

[1] Belgium · [2] Czechia · [3] Poland

Published: June 17, 2026 | Article 847

Abstract

A current acceptance marker cannot date the response history behind it. We retrieved 500 mature Super User questions and all 435 answers publicly retained in a bounded API capture. A retained answer had been created within twenty-four hours for 239 threads, while 121 currently exposed an accepted-answer identifier. The two indicators differed on 174 threads, or 34.8%. Among currently accepted threads, fifteen selected answers were created after the first retained answer; their paired creation gap nevertheless had median zero across all 121 cases. These quantities describe different recorded outcomes. Answer creation dates do not reveal when selection occurred, and deleted or hidden answers may change the historical first response. The analysis therefore characterizes retained-answer chronology without reconstructing acceptance events or actual problem resolution.

Retained Answer Timelines Behind a Mature Acceptance Snapshot

Abstract

A current acceptance marker cannot date the response history behind it. We retrieved 500 mature Super User questions and all 435 answers publicly retained in a bounded API capture. A retained answer had been created within twenty-four hours for 239 threads, while 121 currently exposed an accepted-answer identifier. The two indicators differed on 174 threads, or 34.8%. Among currently accepted threads, fifteen selected answers were created after the first retained answer; their paired creation gap nevertheless had median zero across all 121 cases. These quantities describe different recorded outcomes. Answer creation dates do not reveal when selection occurred, and deleted or hidden answers may change the historical first response. The analysis therefore characterizes retained-answer chronology without reconstructing acceptance events or actual problem resolution.

1. A mature status leaves an undated history behind it

A question may receive a retained answer quickly without currently displaying an accepted-answer identifier. Another may have a selected answer even though its earliest retained answer was created much later. Both situations fit a mature question-and-answer archive. They matter whenever an analyst replaces a timeline with a single terminal-looking status and then interprets that status as evidence of a prompt response.

We ask what current acceptance preserves about the creation timing of retained answers. The primary comparison is between a current accepted identifier and the existence of a retained answer created within twenty-four hours of the question. A second comparison follows the answer currently selected, distinguishing its creation from the earliest retained answer's creation. Neither comparison dates the act of acceptance. The available fields describe a current selection and created records, rather than a sequence of selection events.

Prior research on rapid technical question answering makes response speed a legitimate research object (Mamykina et al., 2011). Work on unanswered questions also treats absence of responses as an issue that warrants its own definition and evidence (Asaduzzaman et al., 2013). Our source is Super User, not a replication of either study's platform conditions or models. In particular, a count of currently retained answers is narrower than proof of everything that happened after a question was posted.

The temporal distinction also matters beyond support forums. A community-thread working-paper abstract combines weak reports with administrator and exit feedback and evaluates early warning relative to later credible feedback (Yuan et al., 2026). That method requires distinguishing an outcome's time from a weak observation's current state. Our public metadata do not reproduce its risk task, private interaction data, label fusion or privacy evaluation. They provide an independent example of why a snapshot selection cannot substitute for a dated feedback event.

2. Response, selection and usefulness require different evidence

Community question answering is a social and technical system in which asking, responding and evaluating coexist (Gazan, 2011). Knowledge-sharing studies investigate participation in answer communities (Adamic et al., 2008), while research on focused sites examines value associated with community activity (Anderson et al., 2012). Those concerns give a response history meaning, but none makes the first created answer identical to the answer ultimately selected or useful to every reader.

Answer-quality prediction and evaluation require their own criteria (Shah and Pomerantz, 2010; Harper et al., 2008). Research on information-seeker satisfaction likewise examines a construct that cannot be read directly from a creation timestamp (Liu et al., 2008). We do not obtain ratings, correctness judgments or a survey of whether the questioner's problem was resolved. The API's current selection is therefore neither a quality gold standard nor an independent evaluation of the first answer.

The official help center describes acceptance as an option exercised by a question's author, rather than a definitive claim of perfect resolution (Stack Exchange Inc., n.d.-b). It also acknowledges that selection need not track every later answer. We cannot determine why any particular identifier is absent or whether a previous selection was replaced. These possibilities delimit the observation; they are not explanations assigned to individual threads.

Measurement analysis requires a connection between a released field and the construct attached to it (Jacobs and Wallach, 2021). A correctly returned state can still be unsuitable for a timing question. Data quality depends on the intended use (Wang and Strong, 1996), and assessment must specify the information needed for that use (Pipino et al., 2002). Here, creation metadata support elapsed creation lags, while an accepted identifier without an event date cannot support elapsed time to acceptance.

3. A fixed queue with complete retained-answer pagination

Before requesting the new cohort, we recorded the selection rule: the first 500 public Super User questions returned in descending creation order, created strictly before 31 October 2025 at 00:00 UTC. Five pages of one hundred questions supplied this fixed convenience queue. No titles, bodies, owner fields, profiles, tags or links were requested. The cohort is not a random sample or a complete history of the site, and current deletions or API ordering ties can affect a later repetition of the query.

The questions were created between 25 September and 30 October 2025. Their ages at capture ranged from 381.729 to 417.245 days. Thus every question had enough calendar age for the declared one-hour, twenty-four-hour and seven-day horizons. Maturity removes a simple administrative-age obstacle to those windows; it does not restore answers that were subsequently deleted, hidden or otherwise unavailable.

We retrieved answers for the five batches of question identifiers, following every answer page until has_more was false. The acquisition required twelve anonymous API calls: one minimal filter, five question pages and six answer pages. The current records contain 500 unique question identifiers and 435 unique answer identifiers. All question answer counts matched the retrieved counts, and all current accepted identifiers matched visible answers marked is_accepted. There were no negative answer-creation lags.

The API documentation defines integer Unix dates in UTC and distinguishes public answer fields from private information (Stack Exchange Inc., n.d.-a). Question and answer calls spanned approximately twelve seconds on 16 November 2026. This is a bounded capture interval, not an atomic database transaction. Agreement of the observed counts and identifiers is a coherence check within that interval, rather than a guarantee that the site could not change.

All 935 question and answer records returned CC BY-SA 4.0 as their content license. Source payload hashes, per-page metadata, licensing documentation and the request record are retained. We did not need contributed text, user identification or another data query to reconstruct the recorded creation order. Reputation studies examine contributors and incentives (Movshovitz-Attias et al., 2013); the identity fields needed for that different question were deliberately outside this collection.

4. Reconstructing the observable clocks

For each question, we take the minimum creation date across every retained answer returned for it. Subtracting the question creation date gives the first-retained creation lag. A thread without a retained answer has no observed lag; it is not assigned zero or infinity. The fixed-horizon indicator instead means that at least one currently retained public answer has a creation date within the stated elapsed interval. Its negative state does not prove that no historical answer existed then.

For the 121 questions with a current accepted identifier, we separately look up the selected answer's creation date. Its lag is measured from the question creation date. The paired difference subtracts the earliest retained creation date from that selected creation date. By construction it cannot be negative. That structural restriction is not an empirical finding that acceptance takes time, because neither timestamp measures acceptance.

We preserve the distinction between creation and public availability. A currently visible post's creation date is not a separately observed history of when it became visible, was noticed or was selected. Likewise, its current accepted status cannot reveal earlier selections that may have been replaced. More generally, unavailable historical quantities cannot be recovered from a present state without additional assumptions (Manski, 2003).

The analysis plan was frozen after complete source acquisition and marginal support inspection, before the joint status-timing and paired selected-answer calculations. It is a secondary study, not a preregistered or outcome-blinded analysis. Twenty-four hours is primary; one hour and seven days are preset secondary horizons. We save the full cumulative retained-creation coverage curve rather than choose a favorable cutoff after observing its shape.

Proposition P1 concerns the difference between current acceptance and twenty-four-hour retained-creation coverage on the same 500 threads. Proposition P2 concerns first versus currently selected answer creation among the 121 current selections. Figure 1 connects their different denominators and clocks.

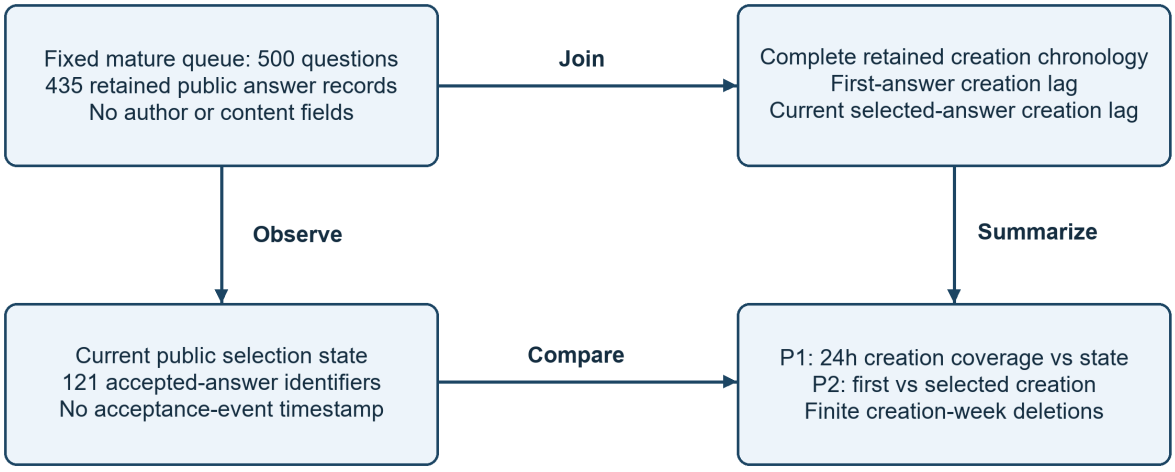

Figure 1. Research model. *All retained answer creation records determine the first-answer lag and the currently selected answer's creation lag. Current selection supplies a separate snapshot state. Creation timestamps do not provide an acceptance-event clock, and neither proposition measures verified resolution.*

5. Current selection does not summarize early creation coverage

There are 317 questions with a retained answer and 183 without one. The API answered flag is true for 278, while 121 expose a current accepted identifier. We retain those source counts but do not repeat a three-flag agreement analysis. The substantive addition is the complete answer chronology used in Table 1 and the distribution of its first creation times.

Table 1. Retained-answer creation coverage at the fixed horizons.

Creation horizon	First retained answer: count / 500	Share of 500	First retained: count among 121 selected threads	Currently selected answer created: count among 121
1h	94	0.188	42	37
24h	239	0.478	93	84
7d	287	0.574	110	104

The second and third columns use all 500 questions. The last two columns use the same 121 currently selected threads. They count answer creation by the horizon, not acceptance by that time. A selected answer created by a horizon necessarily implies a retained answer created by that horizon; this nesting follows from the definition.

At twenty-four hours, the two principal indicators are not nested. Ninety-three questions have both current acceptance and early retained creation; 233 have neither. Another 146 have early retained creation without current acceptance, and 28 have current acceptance without a retained answer created during the first day. These four groups sum to 500. Calling one indicator an error against the other would introduce a gold standard that the source does not provide.

Table 2. Finite timing-state contrasts and creation-week deletion sensitivity.

Finite contrast	Estimate	Leave-creation-week range	Denominator
Any retained answer minus first created by 24h	0.156000	[0.145408, 0.161836]	500
First created by 24h minus current acceptance	0.236000	[0.221258, 0.250000]	500
Current acceptance differs from first created by 24h	0.348000	[0.336788, 0.357143]	500
First minus selected created by 24h, among current acceptance	0.074380	[0.053763, 0.089109]	121

Each contrast is computed on the same declared threads. Deletion removes one of six observed UTC question-creation weeks and recomputes the ratio. The first three rows use all retained questions; the final row conditions on current acceptance and recomputes that denominator after deletion. These deterministic composition ranges are not confidence intervals, p-values or independent replications.

The primary coverage-minus-acceptance difference is 0.236, while the two indicators disagree on 0.348 of questions. Removing one creation week leaves disagreement between 0.336788 and 0.357143. The numerical difference therefore survives those particular changes of cohort composition. It does not estimate how accepting an answer would change response speed or what would occur on another site.

The any-retained-answer indicator exceeds first-day creation coverage by 78 questions, or 0.156 of the cohort. Nonnegativity is guaranteed because first-day retained answers are a subset of all retained answers. The empirical contribution is the size of that late-created group. Similarly, no greater calendar maturity can turn an undated current selection into an acceptance event observed during the first day.

6. First and selected creation lags have different tails

Thirty of the 317 questions with retained answers have their earliest retained answer created after seven days. The median first-retained lag is 4.6492 hours, but the maximum is 248.0216 days. These observations characterize retained posts. They do not establish that those questions truly waited that long before any public or useful response, because earlier deleted answers and resolution outside the site are unobserved.

Table 3. Descriptive creation lags, with explicit conditioning.

Creation-lag distribution	Threads	Median hours	95th percentile days	Maximum days
First retained, all with retained answers	317	4.6492	22.9373	248.0216
First retained, currently selected threads	121	3.2414	23.0670	174.5757
Currently selected answer, selected threads	121	4.9053	64.5714	264.1918
Selected minus first, paired creation gap	121	0.0000	19.3319	222.6915

All lags use recorded creation dates. The final row is a within-thread difference between two answer creation times. Quantiles use linear interpolation as a stated sample-summary convention, consistent with the need to distinguish quantile definitions (Hyndman and Fan, 1996). The displayed percentiles describe observed values rather than inferential uncertainty.

The currently selected answer is also the first-created retained answer in 106 of the 121 selected threads. In the other fifteen, the selected answer was created later. Across all 121 paired differences, the median is zero. Among the fifteen positive differences, the median is about 8.35 days and the maximum about 222.69 days. A long tail therefore coexists with a zero paired median; a larger marginal median must not be described as a typical wait to acceptance.

At the primary horizon, 93 selected threads have a retained answer created within twenty-four hours, whereas only 84 have their currently selected answer created within that interval. The nine-thread gap is 0.074380 among current selections. Its direction is guaranteed by the minimum operation, but its magnitude comes from the source. Even those 84 cases cannot be called accepted within twenty-four hours. Their selection could have occurred later, and earlier selected answers are not reconstructed by the snapshot.

7. Temporal feedback requires a dated outcome

The creation curve and current selection answer different questions even when both are accurately recorded. Label-noise analysis presumes a target against which an observation could be wrong (Frenay and Verleysen, 2014); here neither current acceptance nor early creation is designated truth for the other. Broader data-quality frameworks distinguish the usefulness of a record for particular analytical purposes (Batini and Scannapieco, 2016). For an early-warning evaluation, the missing selection-event date is material, not a nuisance that can be replaced by the answer's creation date.

A prospective benchmark also needs a defensible information boundary. Leakage analysis emphasizes what information a model could actually access at a claimed prediction time (Kapoor and Narayanan, 2023). We fit no model, but the same issue arises if a present acceptance marker is attached to an early

thread state without acknowledging that its selection time is unknown. The present analysis does not establish whether any particular early predictor would leak information or succeed operationally.

Missing-data reasoning requires stating the absent information and its assumptions (Little and Rubin, 2019). These data are not a complete time-to-first-response follow-up with only ordinary administrative censoring. Current retention selects which answers remain observable; conditioning on current acceptance selects another subset. Selection can change the population represented by a comparison (Heckman, 1979), and no correction for those processes is claimed.

Digital traces make such bounded studies possible (Salganik, 2017), while collection and retention mechanisms can diverge from the social question of interest (Olteanu et al., 2019). We therefore do not explain individual selection decisions, contributor effort or question difficulty. Without text, tags or identities, those mechanisms remain unmeasured. Causal conclusions would require a different relationship between observations and interventions (Holland, 1986).

8. Reporting the timeline that was actually observed

Timestamp differences and binary states also occupy different measurement roles (Stevens, 1946). The cumulative curve uses all 500 as its denominator and can plateau below one because some threads have no retained answer. Among the answered subset, descriptive lag distributions use 317; paired selected-answer comparisons use 121. These denominators cannot be exchanged merely to make two percentages appear comparable.

Source documentation should record collection limits and intended uses (Gebru et al., 2021). Separating question-level rows from answer-level timestamps keeps the observations inspectable (Wickham, 2014). Computational reproducibility requires saved inputs and transformations (Peng, 2011), alongside intermediate decisions that explain the resulting quantities (Sandve et al., 2013). The retained package includes all minimal pages, complete pagination checks, per-thread timing values, the creation curve and independent recalculations.

For this mature queue, current acceptance omits substantial information about when retained answers were created. The disagreement includes both early-created answers without current acceptance and current selections whose earliest retained answer was created later than a day. Most selected answers were first-created retained answers, yet a smaller group supplies substantial creation gaps. Those findings concern the dated records that survived to capture. They do not date acceptance, recover deleted responses or establish when a problem was resolved.

References

1. Movshovitz-Attias, Dana; Movshovitz-Attias, Yair; Steenkiste, Peter; Faloutsos, Christos. (2013). Analysis of the reputation system and user contributions on a question answering website: StackOverflow. Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. https://doi.org/10.1145/2492517.2500242
2. Little, Roderick; Rubin, Donald. (2019). Statistical Analysis with Missing Data, Third Edition. Wiley Series in Probability and Statistics. https://doi.org/10.1002/9781119482260
3. Stack Exchange Inc. (n.d.-a). Stack Exchange API v2.3: question, answer, date and pagination definitions. Accessed 16 November 2026. https://api.stackexchange.com/docs
4. Anderson, Ashton; Huttenlocher, Daniel; Kleinberg, Jon; Leskovec, Jure. (2012). Discovering value from community activity on focused question answering sites: a case study of stack overflow. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. https://doi.org/10.1145/2339530.2339665
5. Liu, Yandong; Bian, Jiang; Agichtein, Eugene. (2008). Predicting information seeker satisfaction in community question answering. Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. https://doi.org/10.1145/1390334.1390417

6. Adamic, Lada A.; Zhang, Jun; Bakshy, Eytan; Ackerman, Mark S.. (2008). Knowledge sharing and yahoo answers: everyone knows something. Proceedings of the 17th international conference on World Wide Web. https://doi.org/10.1145/1367497.1367587

7. Salganik, Matthew J.. (2017). Bit by Bit: Social Research in the Digital Age. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691158648/bit-by-bit

8. Harper, F. Maxwell; Raban, Daphne; Rafaeli, Sheizaf; Konstan, Joseph A.. (2008). Predictors of answer quality in online Q&A sites. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/1357054.1357191

9. Batini, Carlo; Scannapieco, Monica. (2016). Data and Information Quality: Dimensions, Principles and Techniques. Data-Centric Systems and Applications. https://doi.org/10.1007/978-3-319-24106-7

10. Sandve, Geir Kjetil; Nekrutenko, Anton; Taylor, James; Hovig, Eivind. (2013). Ten Simple Rules for Reproducible Computational Research. PLoS Computational Biology. https://doi.org/10.1371/journal.pcbi.1003285

11. Shah, Chirag; Pomerantz, Jefferey. (2010). Evaluating and predicting answer quality in community QA. Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. https://doi.org/10.1145/1835449.1835518

12. Hyndman, Rob J.; Fan, Yanan. (1996). Sample Quantiles in Statistical Packages. The American Statistician. https://doi.org/10.1080/00031305.1996.10473566

13. Kapoor, Sayash; Narayanan, Arvind. (2023). Leakage and the reproducibility crisis in machine-learning-based science. Patterns. https://doi.org/10.1016/j.patter.2023.100804

14. Mamykina, Lena; Manoim, Bella; Mittal, Manas; Hripcsak, George; Hartmann, Björn. (2011). Design lessons from the fastest q&a site in the west. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/1978942.1979366

15. Wang, Richard Y.; Strong, Diane M.. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems. https://doi.org/10.1080/07421222.1996.11518099

16. Pipino, Leo L.; Lee, Yang W.; Wang, Richard Y.. (2002). Data quality assessment. Communications of the ACM. https://doi.org/10.1145/505248.506010

17. Yuan, Yuewei; Yin, Jiayang; Huang, Jialei; Xu, Tianyi. (2026). Weakly Supervised Anomaly Detection and Privacy Risk Scoring in Community Message Threads. SSRN working paper 7393478. https://doi.org/10.2139/ssrn.7393478.

18. Frenay, Benoit; Verleysen, Michel. (2014). Classification in the Presence of Label Noise: A Survey. IEEE Transactions on Neural Networks and Learning Systems. https://doi.org/10.1109/tnnls.2013.2292894

19. Gazan, Rich. (2011). Social Q&A. Journal of the American Society for Information Science and Technology. https://doi.org/10.1002/asi.21562

20. Peng, Roger D.. (2011). Reproducible Research in Computational Science. Science. https://doi.org/10.1126/science.1213847

21. Stack Exchange Inc. (n.d.-b). What does it mean when an answer is accepted?. Accessed 16 November 2026. https://superuser.com/help/accepted-answer

22. Stevens, S. S.. (1946). On the Theory of Scales of Measurement. Science. https://doi.org/10.1126/science.103.2684.677

23. Jacobs, Abigail Z.; Wallach, Hanna. (2021). Measurement and Fairness. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. https://doi.org/10.1145/3442188.3445901

24. Heckman, James J.. (1979). Sample Selection Bias as a Specification Error. Econometrica. https://doi.org/10.2307/1912352

25. Wickham, Hadley. (2014). Tidy Data. Journal of Statistical Software. https://doi.org/10.18637/jss.v059.i10

26. Olteanu, Alexandra; Castillo, Carlos; Diaz, Fernando; Kıcıman, Emre. (2019). Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Frontiers in Big Data. https://doi.org/10.3389/fdata.2019.00013

27. Gebru, Timnit, et al. "Datasheets for datasets" *Communications of the ACM*, 2021. https://doi.org/10.1145/3458723.

28. Asaduzzaman, Muhammad; Mashiyat, Ahmed Shah; Roy, Chanchal K.; Schneider, Kevin A.. (2013). Answering questions about unanswered questions of Stack Overflow. 2013 10th Working Conference on Mining Software Repositories (MSR). https://doi.org/10.1109/msr.2013.6624015

29. Holland, Paul W.. (1986). Statistics and Causal Inference. Journal of the American Statistical Association. https://doi.org/10.1080/01621459.1986.10478354

30. Manski, Charles F.. (2003). Partial Identification of Probability Distributions. Springer Series in Statistics. https://doi.org/10.1007/b97478