Why FDA Withholds Drug Approval: Evidence from 443 Publicly Released Complete Response Letters[1]

Alex Tabarrok
Department of Economics, George Mason University, Fairfax, VA 22030

Revised draft, November 2026

Abstract

Background: The FDA began systematically releasing Complete Response Letters in 2025, providing a broader public record of why drug applications are not approved. This study examines the reasons for initial rejection, the probability of another rejection, and whether the same reason recurs.

Methods: I analyzed 443 released letters covering 333 applications. A language model classified the main and additional reasons for nonapproval after standard template language was removed. Each application's first observed letter defines its initial reason. I measured another rejection within two years among 250 applications with sufficient follow-up and compared reasons across the first two letters of 81 applications with repeated rejections.

Results: Facility inspection was the leading initial reason, accounting for 36.3 percent of applications overall and 42.2 percent of those with a first observed rejection from 2024 onward. Clinical efficacy accounted for 18.9 percent, chemistry and manufacturing controls for 17.7 percent, and clinical safety for 11.4 percent. Another rejection was observed within two years for 28.6 percent of applications initially rejected for facility inspection, compared with 14.5 percent for other reasons. Among applications with a second rejection after an initial facility rejection, facility inspection remained the main reason in 28 of 33 cases (84.8 percent).

Conclusion: Facility inspection is the leading initial reason for nonapproval, is associated with a higher observed rate of another rejection, and usually remains the reason when rejection repeats. Facility readiness is a central obstacle to approval in the released record.

Keywords: Food and Drug Administration; Complete Response Letters; drug approval; facility inspections; manufacturing quality.

1. Background

A Complete Response Letter (CRL) communicates an FDA nonapproval action on an application in its present form. The letter states, in numbered sections, what must be fixed before the application can be approved. Until recently these letters were confidential. Fragments appeared in approval packages for drugs that eventually got through, which made the public record a record of survivors (Lurie et al., 2015).

In July 2025 the agency released more than 200 letters associated with applications that were later approved. In September it expanded disclosure to pending or withdrawn applications and announced that it would release new letters promptly (U.S. Food and Drug Administration, 2025a,b). That release makes a basic descriptive question answerable for the first time. What does the FDA say when it turns a drug down?

Earlier studies answered narrower versions of this question. Sacks et al. (2014) used internal FDA records to study first-time applications for new molecular entities submitted to the Center for Drug Evaluation

[1]The author used ChatGPT Sol extensively in collating and analyzing the data.

and Research (CDER) from 2000 through 2012. That study placed inspection findings inside chemistry, manufacturing, and controls (CMC) and counted CMC as the reason for nonapproval only when safety and efficacy were otherwise satisfactory. Lurie et al. (2015) analyzed 61 nonpublic letters issued from 2008 through 2013 and showed that sponsor announcements omitted most of the FDA's concerns. More recent work also finds manufacturing problems in selected samples. Dilek et al. (2026) reports facility deficiencies in 65 percent of 43 novel therapeutics that were approved from 2020 through 2024 after receiving a letter, while McCarthy and O'Boyle (2026) finds that CMC issues were the most common deficiencies among first-time small-molecule applications submitted from 2013 through 2023.

This paper follows applications from their first observed rejection through subsequent rejections. Throughout, initial rejection means the first observed CRL, which need not be the first ever issued for that application. It asks three questions: What is the leading initial reason for nonapproval? Which initial reasons are associated with another rejection? When rejection repeats, does the reason stay the same? I distinguish facility inspection from chemistry and manufacturing controls (CMC): facility inspection concerns compliance at a plant, while CMC concerns the product- or process-quality information in the application.

The corpus contains 443 Complete Response Letters covering 333 applications and 311 companies, from May 2009 through August 2026. I classify every letter, but count each application once when describing initial reasons for nonapproval. Facility inspection leads at 36.3 percent, followed by clinical efficacy at 18.9 percent, CMC at 17.7 percent, and clinical safety at 11.4 percent. The 2024-onward cohort covers newer letters released more systematically by FDA (see further below). Among applications whose first observed rejection falls in this period, the facility share is 42.2 percent.

Facility inspection is also associated with a higher observed rate of another rejection. Among applications with two years of potential follow-up, 28.6 percent of those initially rejected for facility inspection have another rejection within two years, compared with 14.5 percent of those rejected for other reasons. When a second rejection follows an initial facility rejection, facility inspection remains the main reason in 84.8 percent of cases. These findings identify facility problems as prominent at initial rejection and persistent across review cycles.

FDA is moving toward a complete public list of rejection letters, but we do not yet know the full set, and some information in the released letters is redacted.

2. Methods

2.1 The data

I downloaded 458 records from the openFDA transparency endpoint on August 28, 2026; the newest letter is dated August 7, 2026. The endpoint returns structured records. Each carries the letter date, the sponsor name, the application number, the reviewing division, and the full text of the letter as extracted from a scanned PDF. Bulk PDF archives exist separately and I did not need them.

The available letters are unevenly distributed across years. The corpus holds 27 letters from 2009 through 2014, then 121 from 2015 through 2019, 137 from 2020 through 2023, and 158 from 2024 through the August 2026 cutoff. The tilt toward recent years reflects FDA's more systematic release of newer letters.

FDA does not publish annual counts of Complete Response Letters, so the share of all letters covered by the dataset is unknown. I report applications whose first observed rejection dates from 2024 onward separately because newer letters entered the public record through a more systematic release process. This recent cohort contains 128 applications and is a subset of the full sample.

Drugs@FDA provides additional information, but only for drugs that were eventually approved or tentatively approved. Among the 443 study letters, 320 match Drugs@FDA (72.2 percent), including 44 of the 158 letters from 2024 onward (27.8 percent). Recent rejected applications are often missing because they have not

subsequently been approved. Whether a letter matches therefore depends on the outcome being studied. For that reason, this paper does not analyze approval outcomes or time to approval.

The matching procedure and which applications it identifies are described in Appendix A.10. The primary estimates use all strict CRLs, regardless of whether an application matches Drugs@FDA.

The study includes letters whose source letter type is exactly COMPLETE RESPONSE, after removing one duplicate and one incomplete-response acknowledgement. The study covers New Drug Applications (NDAs) and Biologics License Applications (BLAs), including supplements and combination products. Each application contributes its first observed CRL to the initial-rejection analysis; subsequent letters are used to study repeated rejection. Letters referring to multiple applications are assigned to the first application listed in the FDA record; the dataset retains the other application numbers.[2] Appendix A.1 explains which records were included and excluded.

Figure 1 shows how the study sample was constructed.

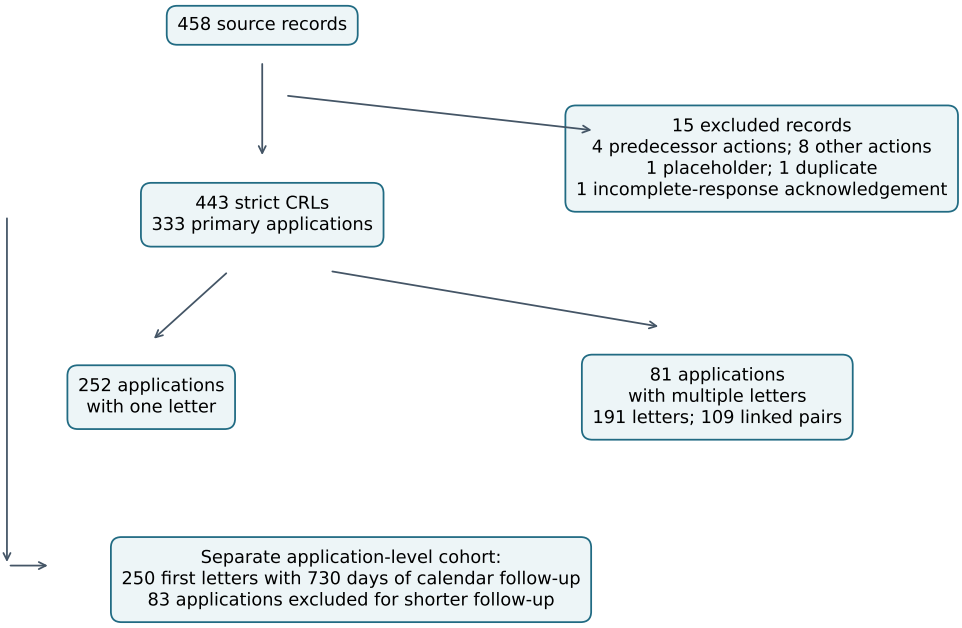


Figure 1: Sample construction. The two-year sample was created to avoid truncation from insufficient follow-up, as described in Section 2.3. Further details on which records were included and excluded are provided in Appendix A.1.

2.2 Coding the letters

Complete Response Letters contain routine instructions about labeling and resubmission that do not necessarily explain why approval was withheld. In an initial analysis of the full text, for example, the model incorrectly identified labeling as a reason for nonapproval in 99 percent of letters. To avoid this problem, I used a two-step procedure to prevent standard template language from being mistaken for reasons for nonapproval.

[2]Only one letter in the dataset refers to multiple distinct applications.

The two-step procedure first used text rules to remove introductory and standard closing language, retaining the section that explains why the application could not be approved. I then asked a language model to identify the main reason for nonapproval and any additional reasons from that retained text, using ten categories: clinical efficacy, clinical safety, trial design, chemistry and manufacturing controls, facility inspection, stability, bioequivalence, nonclinical toxicology, labeling, and administrative deficiencies.

The coding rules define the main reason (the primary category) as the issue most directly responsible for nonapproval, judged from FDA's decision language and required remedy. Other substantive reasons are recorded as secondary categories. Facility inspection covers inspection findings, current good manufacturing practice (CGMP) compliance, or facility acceptability; Form FDA 483 records inspectional observations. CMC covers deficient product- or process-quality information. Both categories are assigned only when a distinct application/product-quality deficiency accompanies the facility issue. The coding rules require one primary category and supply no mechanical tie-breaker for otherwise coequal deficiencies; the stored rationale records the model's judgment. Appendix A.3 gives examples and the overlap analysis tests sensitivity to this boundary.

For letters classified as facility-inspection cases, I defined four mutually exclusive inspection categories: an inspection occurred and produced adverse findings, FDA could not complete the inspection, the inspection was pending or conditional, or the outcome was unclear. I then used an LLM to assign each letter to one of these categories based on its extracted inspection language. Separately, the model flagged references to contract manufacturers or suppliers.

I checked whether the text-extraction rules retained substantive reasons for nonapproval and excluded standard language. An LLM-assisted review compared the extracted sections with the full text of 30 letters to check for omissions. A separate facility-inspection review checked whether routine inspection language had incorrectly caused letters to be classified as facility-inspection cases.

To check classification agreement, a second LLM classified a 50-letter sample without seeing the original model's answers. The sample was selected before the main classification was run. Appendices A.2, A.5, and A.7 explain these checks and report the detailed comparisons.

Redaction limits the coding in 48 letters and OCR damage limits it in 26. Section 3.5 takes up both.

2.3 Statistical analysis

The analysis follows three steps. First, I identify the main reason in each application's first observed CRL. Table 1 compares all 333 applications with the 128 whose first observed letter dates from 2024 onward. Second, I measure whether another rejection appears within two years among the 250 applications with sufficient follow-up. Third, among applications with at least two letters, I compare the main reason in the first rejection with the main reason in the second. Each application contributes one comparison. Submission identifiers and references to earlier FDA decisions were checked to establish review sequences; Appendix A.1 describes the check, resulting record corrections, ordering rules, and follow-up dates.

The two-year outcome is any subsequent rejection observed in the released record, regardless of its reason. An application without another letter may have been approved, withdrawn, or not resubmitted, or may have another letter that was not released. The analysis therefore measures observed repeat rejection rather than eventual approval. Differences in repeat rejection may also reflect whether and how quickly sponsors resubmit, which these data do not systematically observe.

I report 95 percent confidence intervals for proportions and differences between application groups.[3] To test whether the initial category ranking depends on selecting a single main reason, I also count every

[3]Intervals for proportions use the Wilson score calculation with $z = 1.96$. Differences use Newcombe's unpooled Wilson-score method without continuity correction. Appendix A.11 gives the calculation.

substantive category in each application's first observed letter. Supporting analyses use all 443 letters and all 109 consecutive-letter pairs. A further check combines facility inspection and CMC into manufacturing quality.

3. Results

3.1 Reasons for initial rejection

Table 1 shows the main reason for each application's first observed rejection. Facility inspection leads, accounting for 121 of 333 applications (36.3 percent), compared with 18.9 percent for clinical efficacy, 17.7 percent for CMC, and 11.4 percent for clinical safety. Facility inspection and CMC together account for 54.1 percent of initial rejections.

Table 1. Main reasons for initial rejection (application-level)

Main reason	Full N	Full share	2024+ N	2024+ share
Facility inspection	121	36.3%	54	42.2%
Chemistry and manufacturing controls	59	17.7%	13	10.2%
Clinical efficacy	63	18.9%	30	23.4%
Clinical safety	38	11.4%	11	8.6%
Nonclinical toxicology	17	5.1%	5	3.9%
Bioequivalence	18	5.4%	7	5.5%
Trial design and conduct	7	2.1%	3	2.3%
Administrative	5	1.5%	2	1.6%
Labeling	5	1.5%	3	2.3%
Stability	0	0.0%	0	0.0%
Total applications	333	100%	128	100%

Each application is classified using its first observed Complete Response Letter. The 2024-onward cohort includes applications whose first observed letter is dated January 1, 2024 or later.

Facility inspection also leads when every substantive reason in the first letter is counted, appearing in 186 of 333 applications (55.9 percent), compared with 127 (38.1 percent) for CMC. Thus, its prominence does not depend on selecting a single main reason (Appendix Table A9).

The recent cohort checks whether the main finding depends on older, less systematically released records. Facility inspection remains the leading initial reason among the 128 applications first observed from 2024 onward, accounting for 42.2 percent. Clinical efficacy accounts for 23.4 percent, CMC for 10.2 percent, and clinical safety for 8.6 percent.

Facility inspection accounts for 36.3 percent of applications' first observed rejections and 41.5 percent of all rejection letters, reflecting its greater representation in subsequent rejections. Appendix A.11 reports the letter-level distributions and comparisons by application type and period.

3.2 Another rejection within two years

Is an initial facility rejection associated with another rejection? Among the 250 applications with two years of potential follow-up, another rejection is observed within two years for 24 of 84 applications initially rejected primarily for facility inspection (28.6 percent), compared with 24 of 166 applications initially rejected for other reasons (14.5 percent). The difference is 14.1 percentage points, with a 95 percent confidence interval from 3.6 to 25.5 points. This comparison includes applications with no subsequent letter and counts any second rejection, regardless of its reason. The pattern holds within both application types: 31.5 versus 15.7

percent for NDAs and 23.3 versus 9.4 percent for BLAs (Appendix Table A13), suggesting that the pooled difference is not explained by the NDA/BLA mix.

3.3 Does the same reason recur?

Eighty-one applications have at least two rejection letters. Comparing each application's first rejection with its second shows whether the main reason changes. Among the 33 applications whose first letter cites facility inspection as the main reason, the second cites it again in 28 cases (84.8 percent).[4]

The result is similar when all consecutive letters are included. The 81 applications contribute 109 pairs, and the same main reason repeats in 63 pairs (57.8 percent). Facility inspection repeats in 42 of 50 pairs beginning with it (84.0 percent), compared with 3 of 15 pairs beginning with CMC, 7 of 14 beginning with clinical efficacy, and 2 of 8 beginning with clinical safety. The median gap between letters in these facility pairs is 359 days.

Facility problems stand out in both analyses. Applications initially rejected for facility inspection are more likely to receive another rejection within two years. Among applications with repeated rejections, facility inspection usually remains the main reason.

3.4 What the facility problems involve

One possible explanation for the large facility share is that FDA lacked the capacity to complete required inspections, especially at foreign plants and during the pandemic. The letters instead predominantly describe inspections that occurred and found problems. Across all 260 letters citing facility inspection as a main or additional reason, 239 (91.9 percent) describe adverse findings from a completed inspection. An uncompleted inspection is mentioned in 32 letters (12.3 percent), including five from 2024 onward. Twelve of these letters also report adverse findings from a completed inspection. The principal issue is the condition of the plant, rather than the agency's inability to inspect it.

Facility inspection and CMC can describe related parts of a manufacturing problem. In 9 of the 15 consecutive pairs beginning with CMC, facility inspection becomes the main reason in the next letter. A review found that five of these pairs concern the same site, process step, or quality attribute. Combining the two categories, manufacturing quality remains the main reason in 41 of 45 first-pair comparisons (91.1 percent). The appendix reports the underlying reviews and the full transition matrix.

The higher repeat-rejection rate is particularly associated with facility inspection. For applications initially rejected for CMC, the corresponding rate is 8 of 54 (14.8 percent). Combining facility inspection and CMC gives a repeat-rejection rate of 23.2 percent, versus 14.3 percent for other applications. The difference for this broader category is less precisely established (8.9 percentage points; 95 percent confidence interval, -1.0 to 18.3).

In 259 of 443 letters (58.5 percent), FDA cites facility or manufacturing problems with no clinical deficiency identified in the available text, where clinical includes efficacy, safety, trial design, and bioequivalence (Appendix Table A16). This suggests that many rejections at this stage concern manufacturing quality rather than contested judgments about clinical benefits and risks.[5] Patients may disagree about acceptable clinical risks while sharing an interest in reliably manufactured products.

3.5 Validation and robustness

The review of 30 letters found no stated reasons for nonapproval missing from the retained text. The two LLM passes agreed on the main reason in 45 of 50 letters (90 percent; kappa 0.862). The blinded human

[4] A source-text review identified a common plant in ten pairs and narrative evidence of continuity in two more. The remaining pairs could not be matched to a specific plant, mainly because of redaction (Appendix A.11).

[5] Absence of a stated clinical deficiency does not establish that all clinical requirements were satisfied.

comparison agreed with the main LLM in 41 of 50 letters (82 percent; kappa 0.763). Appendix A.5 reports disagreements, often due to redacted text, in more detail.

The application-level finding is stable when uncertain classifications are excluded. Facility inspection accounts for 36.3 percent of first observed rejections overall, 35.9 percent among those without classification flags, and 34.2 percent after excluding OCR-uncertain cases. Excluding redaction-uncertain cases raises the share to 38.0 percent. Facility inspection remains the leading category in each comparison. In the recent cohort, the corresponding shares are 42.2, 48.0, 41.7, and 47.7 percent.

A separate review found no facility classification based solely on routine wording (Appendix Section A.7). A blinded second LLM agreed with all 19 facility assignments among the 26 OCR-uncertain letters. Appendix Section A.6 reports this review and the text-quality checks (Tables A7 and A8).

4. Discussion

The findings trace a consistent pattern across applications and review cycles. Facility inspection is the leading initial reason for nonapproval, is associated with a higher observed rate of another rejection, and usually remains the reason when rejection repeats. Counting applications rather than letters shows that its prominence is not simply a consequence of repeated rejections at a small number of applications.

The importance of facility problems is also consistent with FDA's analysis of nearly 10,000 original applications and supplements received in fiscal years 2024 and 2025 that required a facility assessment. Among submissions receiving a Complete Response Letter, 43 percent had a facility withhold (U.S. Food and Drug Administration, Office of Pharmaceutical Quality, 2026). That analysis covers a different population, including generic applications and supplements, but provides independent context for the role of facility readiness.

The letters cannot show why facility problems recur. Some plants may have difficulty correcting the conditions found during an inspection. In other cases, the problem may be the time and work required to demonstrate that the corrections are adequate. Distinguishing these explanations would require inspection records, including the underlying Form 483 observations, subsequent inspection classifications, and evidence of how and when each problem was resolved.

Related work reaches a similar conclusion at a later stage of the product life cycle. Manufacturing-quality problems are a major source of drug shortages, and weak market rewards for quality can reduce investment in reliable production (Woodcock and Wosinska, 2013; U.S. Food and Drug Administration, Drug Shortages Task Force, 2019). Those studies concern the supply of drugs already on the market, whereas this paper concerns entry. Together, they show that facility quality can constrain both approval and continued availability.

At initial rejection, clinical efficacy and clinical safety are each less often the main obstacle than facility inspection. These findings suggest that manufacturing-site readiness is not a peripheral compliance issue but a central determinant of whether FDA can approve an application. Sponsors may therefore benefit from treating inspection readiness, including oversight of contract facilities, as part of development strategy rather than as a late-stage operational task. More consistent disclosure of facility identity and deficiency type would allow researchers and policymakers to distinguish sponsor-controlled problems from failures in outsourced production.

4.1 Limitations

Redaction prevents a reliable comparison of sponsor-owned plants and contract manufacturers. FDA often withholds plant identities, making it impossible to determine whether outsourced production is especially prone to initial or repeated rejection. More consistent disclosure would enable research on this question.

The public record also remains incomplete. FDA does not publish the total number of Complete Response Letters issued, so the share covered by this study is unknown. Differences between the older and recent cohorts may partly reflect release practices.

Finally, the data support an analysis of repeated rejection, rather than eventual approval or the time needed to resolve deficiencies. Drugs@FDA mainly identifies applications that were ultimately approved, while resubmission dates are available only selectively.

5. Conclusion

Facility inspection is the leading initial reason for nonapproval, accounting for 36.3 percent of applications overall and 42.2 percent of applications first observed from 2024 onward. Most facility findings concern inspections that took place and found problems.

An initial facility rejection is also associated with a higher observed rate of another rejection: 28.6 percent within two years, compared with 14.5 percent for other initial reasons. When a second rejection follows an initial facility rejection, facility inspection remains the main reason in 84.8 percent of cases. Facility readiness is therefore central both to the initial approval decision and to repeated rejection.

Data availability. The complete data, code, prompts, stored model outputs, and validation materials are available in a replication ZIP at alextabarrok.com. Readers can reproduce the results offline or redo the classifications using the supplied letter text.

Author Contributions

Alex Tabarrok is the sole author and conceived and designed the study and interpreted the data. Tabarrok thanks ChatGPT Astra and Claude Fable for error checking, productive comments, assistance with analysis, and editorial summarizing.

References

Sophie Dilek, Richard H. Woods, Jeromie Ballreich, Thomas J. Moore, and G. Caleb Alexander. Deficiencies delaying prescription drug approvals by the U.S. Food and Drug Administration, 2020–2024. *Therapeutic Innovation & Regulatory Science*, 60(3):837–846, 2026. doi: 10.1007/s43441-026-00921-3.

Peter Lurie, Harpreet Singh Chahal, Daniel W. Sigelman, Sylvie Stacy, and Joshua Sclar. Comparison of content of FDA letters not approving applications for new drugs and associated public announcements from sponsors: Cross sectional study. *BMJ*, 350:h2758, 2015. doi: 10.1136/bmj.h2758.

Gemma McCarthy and Niamh M. O'Boyle. Analysis of FDA approvals, delays and denials of new drug applications, 2013–2023. *Drug Discovery Today*, 31(4):104739, 2026. doi: 10.1016/j.drudis.2026.104739.

Leonard V. Sacks, Hala H. Shamsuddin, Yuliya I. Yasinskaya, Khaled Bouri, Michael L. Lanthier, and Rachel E. Sherman. Scientific and regulatory reasons for delay and denial of FDA approval of initial applications for new drugs, 2000–2012. *JAMA*, 311(4):378–384, 2014. doi: 10.1001/jama.2013.282542.

U.S. Food and Drug Administration. FDA embraces radical transparency by publishing complete response letters. FDA press announcement, July 2025a. July 10, 2025.

U.S. Food and Drug Administration. FDA announces real-time release of complete response letters, posts previously unpublished batch of 89. FDA press announcement, September 2025b. September 4, 2025.

U.S. Food and Drug Administration, Drug Shortages Task Force. Drug shortages: Root causes and potential solutions. Technical report, U.S. Food and Drug Administration, 2019. Updated February 2020; available from FDA.

U.S. Food and Drug Administration, Office of Pharmaceutical Quality. Report on the state of pharmaceutical quality, FY2025. Technical report, U.S. Food and Drug Administration, 2026. Available from FDA.

Janet Woodcock and Marta Wosinska. Economic and technological drivers of generic sterile injectable drug shortages. *Clinical Pharmacology & Therapeutics*, 93(2):170–176, 2013. doi: 10.1038/clpt.2012.220.

Appendix: Data, Code and Methods

Appendix Section A.1 Corpus

I downloaded 458 records from the openFDA transparency endpoint at `https://api.fda.gov/transparency/crl.json` on August 28, 2026. The newest letter is dated August 7, 2026. Each record carries the letter date, sponsor name, application number, reviewing division, approver, and the full text of the letter as extracted from a scanned PDF.

Of the 458 records, 443 are distinct Complete Response Letters included in the analysis. The exclusions are four predecessor not-approvable actions, eight other action types, one placeholder without a usable application identifier, one duplicate, and one acknowledgement of an incomplete response. The 443 letters cover 333 applications and 311 companies.

Table A1. Analysis samples and units of observation

Analysis	N	What is counted
Initial rejection	333	One first observed CRL per application
Recent initial rejection	128	Applications with a first observed CRL from 2024 onward
Another rejection within two years	250	Applications with at least 730 days of potential follow-up
First-to-second rejection	81	One comparison per application with at least two letters
All consecutive rejections	109	Every consecutive pair, including later pairs for the same application
Supporting letter analyses	443	All CRLs, including repeated rejections

Within each application, letters are ordered by date and then letter identifier. The first letter defines the initial reason for nonapproval. The recent sample includes only applications whose first observed letter is dated January 1, 2024 or later. The two-year sample requires at least 730 days between the first letter and August 28, 2026; the outcome is another observed CRL within those 730 days. The first-to-second comparison has no two-year restriction.

Checking submission linkage. A review of all 82 applications initially identified as having repeated letters checked submission identifiers, opening paragraphs, and references to earlier FDA actions. Of the original 112 pairs, 85 explicitly named the earlier CRL date and 17 acknowledged a resubmission with the relevant dates obscured. The remaining cases received individual review. This evidence supports treating the retained pairs as successive decisions on the same submission, including cases where redaction prevents exact-date confirmation.

The review identified three corrections. For BLA 761215, two records contained the same letter, so one was removed. For NDA 204017, the July 17, 2013 document acknowledged an incomplete response rather than issuing a new CRL; it was removed, leaving the February 2013 and December 2017 CRLs as the observed pair. For BLA 761303, two same-day letters concerned separate original submissions; both remain in the letter counts, but only the Original 1 letter is linked to the later Original 1 rejection. After these corrections, 81 applications contribute 191 letters and 109 linked consecutive pairs. Each contributes its earliest linked pair to the first-pair analysis.

None of these corrections changes a facility-inspection first-pair comparison. As an additional check, restricting those comparisons to letters explicitly naming the earlier CRL date gives recurrence in 20 of 24 pairs (83.3 percent), compared with 28 of 33 (84.8 percent) in the full first-pair sample. The pair-level evidence and correction decisions are retained in the replication files.

The sample includes supplemental applications, which seek changes to an already approved drug or biologic, such as a change in labeling or manufacturing. The extracted submission references identify four letters concerning supplemental applications and twelve concerning original submissions; the remaining 427 do not identify the submission type. The sample also includes combination products, such as a drug paired with a delivery device.

One letter refers to multiple distinct applications and is assigned to the first application listed in the source record. All associated identifiers remain in the linkage file.

The script `code/19_application_summary.py` reproduces these application summaries and classification sensitivities.

Appendix Section A.2 Separating reasons for nonapproval from standard text

A letter contains an opening paragraph, a body of numbered deficiencies, and closing sections on labeling, resubmission procedure, and dispute resolution. The body explains the reasons for nonapproval. Using whole letters led the model to mistake standard labeling instructions for reasons for nonapproval, as described in Section 2.2 of the paper.

Text rules identify where the explanation of nonapproval begins and where standard closing language starts. The beginning is usually marked by a sentence introducing the list of deficiencies. If no such phrase is found, the rules use the decision sentence or, failing that, the first section heading.

The rules check the following headings and phrases, in this order, to identify where to stop:

```
^OTHER REQUEST(S) AND COMMENT(S)$
^SAFETY UPDATE$
^ADDITIONAL COMMENT(S)$        (conditional; see below)
^Within [number] days after [the] date of this letter
^Within one year after [the] date of this letter
^A resubmission must fully address
^If you decide to resubmit
^You are required to resubmit
```

The late fallback expressions are `^You may request a meeting`, `^If you have any questions`, and `^Sincerely`.

Three headings do not always mark the end of the relevant text. The rules check the text after `SAFETY UPDATE`, `OTHER REQUESTS AND COMMENTS`, and `ADDITIONAL COMMENTS` and retain it when it states a condition that must be met for approval.

The original extraction rules omitted three substantive inspection conditions appearing after the `SAFETY UPDATE` heading. The corrected rules recovered these findings. They increase the number of letters with a main or additional facility finding from 257 to 260, and the number classified as 5b (an inspection could not be completed, with no adverse inspection findings) from 17 to 20. No primary category or 2024-onward count changes. Original classifications are preserved, with the corrections recorded separately in `segmentation_corrections_v20260807.csv` and incorporated in the supporting secondary-category counts.

An LLM-assisted review compared the retained text with the full text of 30 letters: three before 2008, fourteen from 2008 through 2023, and thirteen from 2024 onward. This extraction check includes four predecessor actions excluded from the 443-letter study sample. It found no omitted reasons for nonapproval under the corrected rules.

Appendix Section A.2.1 Examples of standard text removed

This is the standard labeling language that appears in nearly every letter, from NDA 210649 (AB Pharmaceuticals, September 12, 2018):

> PRESCRIBING INFORMATION
>
> 2. We reserve comment on the proposed labeling until the application is otherwise adequate. We encourage you to review the labeling review resources on the PLR Requirements for Prescribing Information and Pregnancy and Lactation Labeling Final Rule websites, including regulations and related guidance documents and the Selected Requirements for Prescribing Information (SRPI) — a checklist of important format items from labeling regulations and guidances.

Salamandra's letter for NDA 208647, issued two years earlier, contains the same paragraph with the same wording. A keyword counter reads both as labeling deficiencies. Neither is one.

The resubmission language is equally standard, from the same AB Pharmaceuticals letter:

> Within one year after the date of this letter, you are required to resubmit or take other actions available under 21 CFR 314.110. If you do not take one of these actions, we may consider your lack of response a request to withdraw the application under 21 CFR 314.65.

Appendix Section A.3 Category scheme

The main LLM classification assigns the retained text from each letter one primary category and any number of secondary categories:

Table A2. Deficiency category definitions

Code	Category	Scope
1	Clinical efficacy	Evidence does not establish effectiveness
2	Clinical safety	Specific safety signal or unfavorable risk-benefit
3	Trial design and conduct	New or different trial required, data integrity, GCP
4	Chemistry and manufacturing controls	Formulation, specification, process validation
5	Facility inspection	Inspection findings, CGMP status, site compliance
6	Stability and shelf life	Evidence supporting storage conditions and shelf life
7	Bioequivalence and bioavailability	Drug exposure, equivalence, or pharmacokinetic bridging
8	Nonclinical toxicology	Nonclinical safety studies and toxicology findings
9	Labeling and proprietary name	Specific concerns, beyond standard reserved comments
10	Administrative and procedural	Incomplete submission, missing data, REMS

Letters with a main or additional facility finding receive one of four inspection-status categories. A separate flag records references to third-party manufacturers or suppliers:

Table A3. Facility-inspection categories

Code	Subcategory	Letters (of 260)
5a	Inspection occurred, adverse findings	239
5b	Inspection could not be completed	20
5c	Pending or conditional statement only	0
5e	Unclear	1
5d	Contract manufacturer or supplier mentioned (separate flag)	27

Each letter receives one inspection category. If it reports both adverse inspection findings and an inspection that could not be completed, it is classified as 5a. Thus, 32 of the 260 letters (12.3 percent) mention an uncompleted inspection, but only 20 are classified as 5b; the other 12 also report adverse findings. Five of the 32 letters date from 2024 onward, including two classified as 5b. Third-party involvement is recorded separately and can accompany any inspection category. Counts include the extraction correction described in Appendix A.2.

The model also records whether a new clinical trial is explicitly required, whether the deficiency lies with a contract manufacturer, and whether the letter states the application is otherwise approvable.

Appendix Section A.4 Model configuration

The main model classified each letter using its retained deficiency text and the category definitions above. It returned one primary category, any secondary categories, and a rationale identifying the language supporting its judgment. The stored run records identify model alias `gpt-5.6-terra`, low reasoning effort, and calls made on August 28, 2026; temperature was not explicitly set.

The model was required to select a primary category even when the text was uncertain. It also marked classifications made uncertain by redaction or scanning errors. Appendix Section A.6 examines these cases.

The replication archive contains the complete prompt, coding guide, output format, model responses, timestamps, model identifiers, and input hashes. The prompt and coding guide are in `code/04_classify_deficiencies.py` and `notes/phase3_codebook.md`.

Appendix Section A.5 Validation

A 50-letter random sample was drawn before the main classification ran. A second LLM classified those letters without access to the main LLM output. Twenty-one of the 50 are dated 2024 or later. Agreement between the two LLM classifications on the primary category is 45 of 50, or 90 percent (Cohen's kappa 0.862).

A human reviewer independently classified the main reason in each letter using the same blinded 50-letter packet. The reviewer agreed with the main model on 41 letters (82 percent; Cohen's kappa 0.763). The review covered primary categories only.

Table A4. Agreement with independent classifications

Comparison	Agreements	Agreement	Cohen's kappa
Main LLM versus blinded second LLM	45/50	90%	0.862
Main LLM versus blinded human primary-only review	41/50	82%	0.763

Table A5 counts letters assigned to each category by the main model, the second model, and both models.

Table A5. LLM agreement by primary category

Category	Main model	Second model	Both models
Clinical efficacy	6	5	5
Clinical safety	6	4	4
Trial design and conduct	1	1	1
Chemistry and manufacturing controls	10	11	10
Facility inspection	22	23	22
Bioequivalence	0	1	0
Nonclinical toxicology	3	4	3
Labeling	1	1	0
Administrative	1	0	0

The second LLM agrees with all 22 facility assignments made by the main LLM and identifies one additional facility case. The two LLMs also agree on all ten main-LLM chemistry assignments, while the second LLM identifies one additional chemistry case.

The sample is too small to assess labeling, bioequivalence, or administrative classifications reliably. In the administrative disagreement, the product-quality section is fully redacted and the main model relied on surviving procedural text.

The human reviewer agrees with 17 of the main LLM's 22 facility assignments and assigns the other five to CMC, labeling, or clinical safety. In a post hoc review, the author examined all nine human–LLM disagreements, supporting the LLM classification in six, the human classification in two, and finding one case ambiguous. The analysis retains the original LLM classifications.

Model-to-model agreement measures consistency. The blinded human comparison provides a separate check, and the full texts, prompts, outputs, human coding workbook, and disagreement records are available for independent assessment.

Appendix Section A.5.1 The five LLM-to-LLM disagreements

Redaction affects four of the five disagreements.

Table A6. LLM classification disagreements

Application	Date	Second LLM	Main LLM	What happened
BLA 761326	2024-07-10	Facility	Clinical safety	Both reasons present; letter covers two submissions with different binding issues
NDA 217433	2024-07-15	Labeling	Clinical safety	Device human-factors boundary; some product-quality text is redacted
NDA 219357	2024-12-23	CMC	Administrative	Product quality body fully redacted; only category label survives
NDA 209531	2025-09-22	Nonclinical	Labeling	Nonclinical and product quality detail redacted; responsive labeling text remains visible
NDA 220138	2026-06-15	Bioequivalence	Clinical efficacy	PK bridging boundary; CMC and nonclinical details are redacted

Two of the five LLM-to-LLM disagreements are boundary disagreements between defensible readings rather than errors.

Appendix Section A.6 Text quality and classification sensitivity

Each letter has a status indicating whether redaction or scanning errors make its classification uncertain. The following application-level checks use each application's first observed letter, matching Table 1 of the paper.

Table A7. Application-level sensitivity to text quality

	All applications		2024 onward	
Sample	Facility / N	Share	Facility / N	Share
All	121/333	36.3%	54/128	42.2%
No classification flags	99/276	35.9%	47/98	48.0%
Exclude scanning-uncertain	107/313	34.2%	50/120	41.7%
Exclude redaction-uncertain	113/297	38.0%	51/107	47.7%

Facility inspection remains the leading initial reason under each treatment of uncertain classifications.

Supporting letter-level checks. Across all 443 letters, 367 have no classification flag, 48 are uncertain because of redaction, 26 because of scanning errors, and two are flagged as failed. Facility inspection is the main reason in 40.9 percent of unflagged letters, 31.3 percent of redaction-uncertain letters, and 73.1 percent of scanning-uncertain letters.

Table A8. Letter-level sensitivity to text quality

Sample	Facility share, all letters	Facility share, 2024 onward
All letters	41.5%	46.2%
No classification flags	40.9%	50.8%
Exclude scanning-uncertain	39.6%	44.9%

The high facility share among scanning-uncertain letters prompted a blinded second-LLM review of all 26. It agreed with all 19 facility assignments and with 25 classifications overall. The remaining letter, NDA 219846, did not contain enough readable text for the second model to classify. Plant names and inspection details may scan poorly even when the sentence requiring correction before approval remains readable.

Appendix Section A.7 Checking routine inspection language

This check asked whether routine inspection wording had incorrectly produced facility classifications. All five letters in which inspection terms appeared in the full text but not in the initially extracted section received a separate LLM-assisted review. Four contained substantive inspection conditions; one was too heavily redacted to determine the nature of the omitted material. The extraction correction arising from this review is described in Appendix A.2.

In these five reviewed letters, no facility classification was attributable solely to routine inspection language.

Appendix Section A.8 Worked examples

Appendix Section A.8.1 A failed inspection (5a), NDA 205920, Armstrong Pharmaceuticals, May 22, 2014

> During a recent inspection of the manufacturing facility for this application, our field investigator conveyed current Good Manufacturing Practices (cGMP) deficiencies to the representative of the

> facility. [...] Satisfactory resolution of these cGMP deficiencies is required before this application may be approved.

The stored main reason is trial design and conduct, reflecting the required label-comprehension, human-factors, and actual-use studies. Facility inspection is an additional reason, classified as 5a because the inspection produced adverse findings.

Appendix Section A.8.2 An inspection that could not happen (5b), BLA 761192, MediWound, June 25, 2021

> An inspection of the MediWound Ltd. (FEI: 3003889199), Drug Substance, Drug Product, and Gel Vehicle manufacturing facility is required before this application can be approved. FDA must assess the ability of that facility to conduct the listed manufacturing operations in compliance with CGMP. Due to restrictions on travel, we were unable to conduct an inspection during the current review cycle for your application. [...] However, even if these deficiencies are addressed, the application cannot be approved until the required FDA inspection is conducted and any findings are assessed with regard to your application.

Main reason: facility inspection, 5b. The letter describes required inspections of the MediWound and Challenge Bioproducts facilities that could not be conducted because of travel restrictions. CMC and trial-design deficiencies are additional reasons. This is one of the 20 letters classified as 5b, which requires that no adverse completed facility finding is stated. By contrast, Kashiv's August 2, 2021 letter (BLA 761082) describes both an uncompleted inspection and adverse inspection findings; it is classified as 5a under the coding hierarchy.

Appendix Section A.8.3 A chemistry deficiency (category 4), NDA 205394, RedHill Biopharma, January 31, 2014

> The proposed acceptance criterion for the disintegration is overly broad. Tighten the specification for disintegration to a limit that is justified by data and patient requirements.
>
> The acceptance criterion for the dissolution in the drug product specification was revised [...] Update the drug product release and stability specifications to reflect this change and provide dissolution data for the last time point on the stability samples with this method.

Main reason: chemistry and manufacturing controls (CMC). This excerpt concerns product-quality specifications. The letter also cites facility-inspection deficiencies as an additional reason, illustrating how CMC and facility findings can coexist even when CMC is primary.

Appendix Section A.8.4 An efficacy deficiency (category 1), BL 125827, Replimune, April 10, 2026

> [...] subject matter experts from Oncology Center of Excellence unanimously determined data presented are insufficient to conclude substantial evidence of effectiveness of vusolimogene oderparepvec for treatment of unresectable advanced cutaneous melanoma. [...] The deficiencies for each of the studies are outlined here. For RPL-001-16: Inability to isolate the contribution of vusolimogene oderparepvec when administered in combination with nivolumab

Main reason: clinical efficacy. FDA could not determine the drug's contribution to the effectiveness of the combination treatment.

Appendix Section A.8.5 Redaction plus scanning damage, NDA 218317, NeuroDerm, May 24, 2024

> (b) (4)

FACILITY INSPECTIONS

(9)

(10

Following pre-approval inspection of om

, listed in the application, FDA conveyed deficiencies to the representative of the facility. The facility should provide satisfactory responses to these deficiencies to the FDA office indicated on the form FDA-483 prior to your complete response to your application.

This one letter shows both problems at once. The `(b) (4)` marker removes preceding material. The facility name has been redacted and what survives the scan is the fragment `om`. The stray `(9)` and `(10` are misread section numbers. The model can still see that a pre-approval inspection produced deficiencies conveyed on a form 483, which is enough to assign 5a, and it cannot see whose plant it was.

Plant identities are redacted in 128 of the 257 letters reviewed for ownership, limiting comparisons between sponsor-owned and third-party facilities.

Appendix Section A.9 Interpreting the categories

Facility inspection and CMC can describe different aspects of the same manufacturing problem. A change between these categories does not necessarily mean that the underlying problem changed. The main analysis keeps them separate; the combined-category results and case review below show how this distinction affects recurrence.

The categories record the kind of deficiency, not its severity. A letter requesting one specification revision and a letter listing several manufacturing violations each count once.

Appendix Section A.10 Matching applications to Drugs@FDA

Applications remain in the study whether or not they match Drugs@FDA. Matching is less complete for BLAs partly because Drugs@FDA generally excludes products reviewed by the Center for Biologics Evaluation and Research (CBER), including vaccines, blood products, and cell and gene therapies. It includes therapeutic biologics reviewed by CDER. A CBER-regulated BLA may therefore be absent even if it was eventually approved.

Facility inspection is the main reason for nonapproval in 25 of 42 unmatched BLA letters (59.5 percent), compared with 40 of 74 matched BLA letters (54.1 percent). Restricting the sample to matched letters would therefore modestly understate the facility share among BLAs. CBER coverage explains only part of the matching gap: of the 33 unmatched BLA applications, 12 were reviewed by CBER, 16 by CDER, and five could not be assigned to a center.

Appendix Section A.11 Supporting results and sensitivity analyses

Initial reasons without a primary-category ranking. Table A9 counts every substantive category in each application's first observed CRL, using primary and secondary codes and the extraction corrections described in Appendix A.2. The applications and first letters are the same as in Table 1. Each application counts at most once per category, but can appear in several categories; percentages therefore need not sum to 100.

Table A9. All substantive reasons in the first observed CRL

Substantive reason	Full N	Full share	2024+ N	2024+ share
Facility inspection	186	55.9%	71	55.5%
Chemistry and manufacturing controls	127	38.1%	20	15.6%
Clinical efficacy	76	22.8%	33	25.8%
Clinical safety	70	21.0%	23	18.0%
Nonclinical toxicology	39	11.7%	9	7.0%
Bioequivalence	40	12.0%	14	10.9%
Trial design and conduct	84	25.2%	31	24.2%
Administrative	21	6.3%	5	3.9%
Labeling	42	12.6%	12	9.4%
Stability	55	16.5%	2	1.6%

The denominators are 333 applications overall and 128 from 2024 onward. Facility inspection is the most frequent category in both samples: 186 of 333 (55.9 percent) overall and 71 of 128 (55.5 percent) in the recent cohort. Its leading position therefore does not depend on assigning a single primary reason. This check uses the existing substantive codes without changing their assignments or creating a new coequal-reasons category.

Initial reasons and the distribution across all letters. Table 1 in the paper counts applications using their first observed rejection. Table A10 below counts all 443 letters, including subsequent rejections; its recent column includes all 158 letters dated from 2024 onward. Facility inspection's share rises from 36.3 percent of initial rejections to 41.5 percent of all letters.

Table A10. Main reasons for nonapproval (letter-level)

Main reason	Full N	Full share	2024+ N	2024+ share
Facility inspection	184	41.5%	73	46.2%
Chemistry and manufacturing controls	73	16.5%	14	8.9%
Clinical efficacy	72	16.3%	34	21.5%
Clinical safety	45	10.2%	12	7.6%
Nonclinical toxicology	27	6.1%	6	3.8%
Bioequivalence	18	4.1%	7	4.4%
Trial design and conduct	10	2.3%	4	2.5%
Administrative	8	1.8%	5	3.2%
Labeling	5	1.1%	3	1.9%
Stability	1	0.2%	0	0.0%

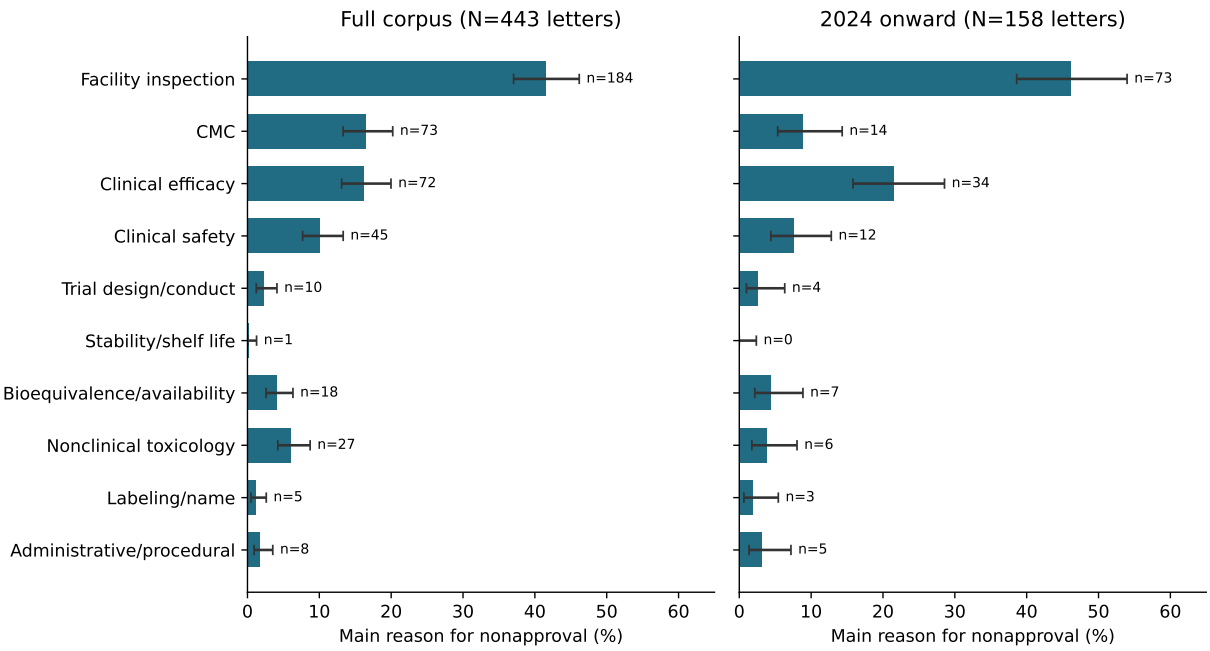


Figure 2: Main reasons for nonapproval across all letters. Bars show proportions, labels show counts, and whiskers show 95 percent confidence intervals. The 2024-onward letters are a subset of the full corpus.

Table A11. Facility inspection as the main reason for rejection, by letter date and application type

For each period and application type, the table reports the total number of letters, the number citing facility inspection as the main reason, and the corresponding percentage. Across all periods, facility inspection is the main reason in 56.0% of BLA letters and 36.4% of NDA letters. Periods refer to when FDA issued the letters, not when it released them publicly. Small early samples and differences in public disclosure limit comparisons over time.

Period	Type	Letters	Facility	%
2009–2014	NDA	27	3	11.1
2015–2019	BLA	15	3	20.0
2015–2019	NDA	106	41	38.7
2020–2023	BLA	40	24	60.0
2020–2023	NDA	97	40	41.2
2024–2026	BLA	61	38	62.3
2024–2026	NDA	97	35	36.1

Another rejection within two years. Table A12 includes applications with and without a subsequent letter. It measures any second rejection within two years, regardless of the reason cited.

Table A12. Another released CRL within 730 days (application-level)

Applications are grouped by the main reason in their first observed rejection letter. Each of the 250 applications received that letter at least 730 days before the data cutoff of August 28, 2026. The table shows how many received another observed rejection within those 730 days. Intervals are 95 percent confidence intervals. The facility versus other difference is 14.1 percentage points (95 percent confidence interval 3.6–25.5); the manufacturing-quality versus other difference is 8.9 points (-1.0 to 18.3).

First-letter category	Applications	Another CRL	%	95% CI
Facility	84	24	28.6	20.0–39.0
Other than facility	166	24	14.5	9.9–20.6
Manufacturing quality	138	32	23.2	16.9–30.9
Other than manufacturing quality	112	16	14.3	9.0–22.0

Table A13. Another rejection within two years, by application type

This comparison uses the same 250 applications and two-year outcome as Table A12. Cells show repeat rejections divided by eligible applications, followed by percentages. Differences compare initial facility-inspection reasons with other initial reasons within each application type; intervals are 95 percent confidence intervals.

Type	Initial facility reason	Other initial reason	Difference (points)	95% CI (points)
NDA	17/54 (31.5%)	21/134 (15.7%)	15.8	2.9 to 30.1
BLA	7/30 (23.3%)	3/32 (9.4%)	14.0	-4.8 to 32.6

The facility–other differences are similar for NDAs (15.8 percentage points) and BLAs (14.0 points), indicating that the pooled association is not explained by the application-type mix. The BLA estimate is less precise: its confidence interval includes zero, reflecting the smaller sample of 62 applications, compared with 188 NDAs.

Repetition of the reason. Tables A14 and A15 compare the reasons in successive letters. The first-pair analysis gives each application one comparison; the all-pair analysis also includes later compar-

isons for applications with three or more letters. These comparisons are not restricted to two years.

Table A14. Consecutive observed category transitions (pair-level)

Read each row across: it shows the main reason in one rejection letter and the main reason in the next. Cells report the number of pairs and the percentage of the row total; N is that total. Fac = facility inspection; Eff = efficacy; Saf = safety; Trial = trial design/conduct; Stab = stability; Bio = bioequivalence/availability; Noncl = nonclinical; Label = labeling/name; Admin = administrative. Dashes mean no pairs began with that category. Percentages are rounded to one decimal place; the replication files contain the unrounded values.

First / next	Fac	CMC	Eff	Saf	Trial	Stab	Bio	Noncl	Label	Admin	N
Facility inspection	42 (84.0)	6 (12.0)	1 (2.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (2.0)	50
CMC	9 (60.0)	3 (20.0)	0 (0.0)	1 (6.7)	1 (6.7)	1 (6.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	15
Clinical efficacy	1 (7.1)	1 (7.1)	7 (50.0)	3 (21.4)	0 (0.0)	0 (0.0)	0 (0.0)	2 (14.3)	0 (0.0)	0 (0.0)	14
Clinical safety	3 (37.5)	2 (25.0)	1 (12.5)	2 (25.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	8
Trial design/conduct	2 (66.7)	0 (0.0)	0 (0.0)	1 (33.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3
Stability/shelf life	—	—	—	—	—	—	—	—	—	—	0
Bioequivalence/availability	1 (50.0)	1 (50.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2
Nonclinical toxicology	4 (28.6)	1 (7.1)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	8 (57.1)	0 (0.0)	1 (7.1)	14
Labeling/name	—	—	—	—	—	—	—	—	—	—	0
Administrative/procedural	1 (33.3)	0 (0.0)	0 (0.0)	0 (0.0)	1 (33.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (33.3)	3

Table A15. Recurrence sensitivities (pair-level)

Cells show how often the same main reason appears in successive rejection letters, as a count out of the total and a percentage. Manufacturing quality combines facility inspection and CMC. The first-pair column counts each application once; the all-pairs column also includes later rejection cycles. For first pairs, the 95 percent confidence intervals are 44.7–65.9 percent for any same-category recurrence, 69.1–93.3 percent for facility recurrence, and 79.3–96.5 percent for manufacturing-quality recurrence. All-pair estimates are descriptive because several pairs can come from the same application.

Initial main reason	First pair per application	All consecutive pairs
All categories	45/81 (55.6%)	63/109 (57.8%)
Facility inspection	28/33 (84.8%)	42/50 (84.0%)
CMC	3/12 (25.0%)	3/15 (20.0%)
Manufacturing quality (facility or CMC)	41/45 (91.1%)	60/65 (92.3%)

The linkage corrections described in Appendix A.1 remove both originally identified same-date pairs; no retained pair has tied dates.

Do repeated facility rejections concern the same plant? An LLM-assisted source-text review examined all 28 first-pair comparisons in which facility inspection remained the main reason. Ten pairs identify at least one common plant through matching FDA establishment identifiers or a specific plant name and location. Two additional pairs support continuity through references to re-inspection or previous inspections, although the plant identity is redacted. Sixteen pairs remain indeterminate: fifteen obscure the relevant plant identities, and one names the same manufacturer without identifying its plant in both letters. No pair explicitly identifies replacement of the earlier plant by a different plant as the explanation for recurrence.

Same-plant recurrence does not establish that every deficiency remained unchanged. For NDA 214520, the same Rovi facility appears in both letters, but the second also cites another facility. For BLA 761147,

the second letter explicitly states that the facility's responses provided insufficient evidence that deficiencies from previous inspections had been resolved. These findings support recurring obstacles at the same facilities where identity can be established; the reported recurrence rate measures repetition of the facility-inspection category. Pair-level evidence and assessments are supplied in the replication files.

Facility inspection and CMC. In nine of the fifteen consecutive pairs beginning with CMC, facility inspection becomes the main reason in the next letter. A review found that five pairs concern the same site, process step, or quality attribute, two concern different problems, and two cannot be assessed because of redaction. In all five same-problem pairs, the first letter already cited facility inspection as an additional reason.

Combining facility inspection and CMC into manufacturing quality therefore checks sensitivity to the category boundary. Manufacturing quality repeats in 60 of 65 consecutive pairs and 41 of 45 first pairs (Table A15). Its two-year repeat-rejection difference is smaller and less precise than the facility-only difference (Table A12).

Table A16. Manufacturing and clinical findings (letter-level)

Presence includes primary and secondary categories, using the corrected extraction described in Appendix A.2. Manufacturing means facility inspection or CMC. The broad clinical definition includes efficacy, safety, trial design/conduct, and bioequivalence/bioavailability. Cells report counts and percentages of all letters in the column: 443 overall and 367 with classification status `ok`. The latter excludes redaction, OCR, and failed-classification flags.

Coded findings	All letters	Unflagged
Manufacturing without clinical	259 (58.5)	208 (56.7)
Both	81 (18.3)	71 (19.3)
Clinical without manufacturing	86 (19.4)	76 (20.7)
Neither	17 (3.8)	12 (3.3)

Manufacturing findings without a coded clinical deficiency appear in 259 of 443 letters (58.5 percent), and in 208 of 367 unflagged letters (56.7 percent). Excluding bioequivalence/bioavailability from the clinical definition gives 275 of 443 letters (62.1 percent). These groups can include other findings, such as nonclinical or labeling deficiencies; absence of a coded clinical finding does not establish that all clinical requirements were satisfied. Letter-level indicators and both clinical definitions are supplied in the replication archive.

Table A17. Primary and secondary categories (letter-level)

"Primary" counts letters in which the category is the main reason for rejection; "Secondary" counts letters in which it is an additional reason. "Any" is their sum. Percentages use all 443 letters. Because a letter can cite several reasons, percentages across categories need not sum to 100. Counts include the extraction correction described in Appendix A.2.

Category	Primary	Secondary	Any	Any (%)
Facility inspection	184	76	260	58.7
CMC	73	81	154	34.8
Clinical efficacy	72	15	87	19.6
Clinical safety	45	39	84	19.0
Trial design/conduct	10	87	97	21.9
Stability/shelf life	1	65	66	14.9
Bioequivalence/availability	18	26	44	9.9
Nonclinical toxicology	27	25	52	11.7
Labeling/name	5	57	62	14.0
Administrative/procedural	8	22	30	6.8

Facility inspection and CMC co-occur in 74 of 443 letters (16.7 percent). Together they are the main reason in 257 letters (58.0 percent), including 87 of 158 letters from 2024 onward (55.1 percent). Application-level and letter-level sensitivity to text-quality flags is reported together in Appendix A.6.

Confidence interval calculations. Intervals for single proportions use the Wilson score method with $z = 1.96$. Differences between independent application-level proportions use Newcombe's unpooled Wilson-score method without continuity correction. For proportions p_1, p_0 with Wilson bounds $(L_1, U_1), (L_0, U_0)$, the difference $d = p_1 - p_0$ has bounds $d - \sqrt{(p_1 - L_1)^2 + (U_0 - p_0)^2}$ and $d + \sqrt{(U_1 - p_1)^2 + (p_0 - L_0)^2}$.

Facility ownership and redaction. Ownership was examined in the 257 letters identified as facility cases before the extraction correction in Appendix A.2. Plant identities were redacted in 128 letters, accounting for most unclear ownership assignments.

Table A18. Facility ownership and redaction

Ownership assignment	Letters
Sponsor-owned plant	65
Third-party plant	27
Unclear	165
Of which plant identity redacted	128

Ownership is identifiable in 19 of the 42 consecutive pairs where facility inspection remains the main reason, but only one of the eight pairs where the reason changes. These gaps prevent a reliable comparison of recurrence at sponsor-owned plants and contract manufacturers.