Docket

DOCKET / RESEARCH & PRACTICAL GUIDANCE

How Many Records Should You Sample for an Audit?

> Build a more reviewable acquisition pipeline with Docket. Our deal-origination software and managed research help private equity teams screen companies against a mandate and understand the evidence behind each finding.

Published by Docket · 2026-09-22 · audit sample size · acceptance sampling · quality assurance sampling · research data quality · target research · private equity origination · exact binomial · vendor acceptance criteria · docket

Original article: https://docket.capital/articles/target-audit-sampling-worksheet



In this article

1. Executive Summary
2. Introduction and Background
3. What Audit Sampling Means for Target Research
4. Define the Lot and Defects Before Sampling
5. Choose a Representative Sample
6. Build the Acceptance Plan and Worksheet
7. Implement the Review and Escalation Process
8. Data Analysis and Evidence
9. Implications and Future Directions
10. Frequently Asked Questions (FAQs)
11. Conclusion

Executive Summary

The practical answer to **how many records should you sample for an audit** is not a universal percentage or a fixed ten-row check. A defensible number follows from a written decision rule. First define one comparable batch, the defect types, a tolerable quality level, a clearly unacceptable quality level, and the maximum risks of rejecting a good batch and accepting a bad one. Then solve for the sample size **n** and acceptance number **c**. PCAOB guidance states the general statistical tradeoff directly: smaller samples carry greater sampling risk (pcaobus.org). NIST describes acceptance sampling as a lot decision, explicitly “not to estimate the quality of the lot” (www.itl.nist.gov). That distinction matters: acceptance sampling answers whether the evidence is strong enough to accept a research batch under a pre-agreed rule. It does not prove the batch's exact error rate.

For a simple **zero-defect demonstration**, the exact one-sided 95% upper confidence bound after observing zero defects is `1 - 0.05^(1/n)`. The familiar rule of three, approximately `3/n`, is a large-sample shortcut, with one university source suggesting **n at least 100** for that final approximation (faculty.washington.edu). Zero defects in **20** records leaves an exact upper bound of **13.91%**; zero in **59** lowers it to **4.95%**; zero in **299** lowers it to **1.00%**. Therefore, a team that wants zero observed errors to support a 95% upper bound below 5% needs **59 randomly selected records**, while a target below 1% needs **299**. Those are confidence-bound calculations, not necessarily economical acceptance plans.

An acceptance plan can allow some observed defects while controlling two decision errors. NIST defines the producer-side benchmark as an acceptable quality level and the buyer-side benchmark as a high defect level unacceptable to the consumer (www.itl.nist.gov). For an illustrative binomial plan in which **1%** is good, **5%** is bad, and each error risk is capped at **5%**, the smallest integer plan is **n = 181, c = 4**. It accepts with probability **96.3670%** at a 1% defect rate and **4.9163%** at a 5% defect rate. A zero-acceptance plan of **n = 60, c = 0** controls buyer risk at 5% quality but rejects a good 1% batch **45.2843%** of the time. This is why sample size and accept number must be designed together.

Private-equity target research also needs **field-specific controls**. Wrong entity, wrong ownership, or a missed mandate disqualifier can change which company enters the funnel, so the buyer may require 100% review of those fields. Selecting the plan by defect consequence is a transferable risk principle (www.fda.gov). PCAOB's financial-statement audit guidance is not a PE requirement, but its transferable discipline is useful: every population item should have a selection opportunity (pcaobus.org), and full examination is appropriate where accepting sampling risk is unjustified (pcaobus.org). Lower-risk descriptive fields can use a random or stratified acceptance sample. Missing evidence must remain an explicit result, not disappear from the denominator. The worksheet below turns those principles into an accept, expand, remediate, or reject decision.

Introduction and Background

A research lead receiving 200, 2,000, or 20,000 target records faces two different questions. The first is operational: **is this batch acceptable under the mandate?** The second is inferential: **what is the batch's true defect rate, with stated precision?** Acceptance sampling is built for the first question. Estimation requires a different design, usually a confidence interval and a sample sized for precision rather than a pass or fail boundary. WHO makes the same distinction in its domain-specific lot-quality assurance sampling guidance: the method “does not estimate an exact prevalence” (www.who.int). GAO similarly evaluates the specific data needed for the stated findings and conclusions, rather than treating an entire database as uniformly reliable (www.gao.gov). This report borrows the statistical discipline, not the health-survey application.

The unit under review is a **record-field result**, not merely a row. One company row might have correct identity and location but unsupported ownership and an unanswered revenue field incorrectly presented as zero. GAO defines reliability through **accuracy, completeness, and applicability to the intended purpose** (www.gao.gov). The UK Government Data Quality Framework reinforces that a complete dataset can still contain incorrect field values (www.gov.uk). A batch-level percentage that merges every field therefore conceals the risk that matters.

This framework applies to work produced by employees, contractors, research vendors, or automation. It is a quality-assurance method, not an audit opinion, a substitute for acquisition diligence, or a benchmark of a particular artificial-intelligence agent. Docket is a direct provider of sourced target research. Its documented process separates Triage identity and fit checks, Scout collection, and Audit review. Its evidence guide also distinguishes verifying that quoted text exists from judging whether it supports the recorded answer. Those are useful review layers, but the buyer still needs an acceptance rule suited to its own mandate and consequences.

What Audit Sampling Means for Target Research

Acceptance sampling versus error-rate estimation

An **attribute acceptance plan** classifies each reviewed item as conforming or defective. A single plan has two numbers: sample size **n** and acceptance number **c**. NIST states that the lot is rejected when observed defects exceed **c** (www.itl.nist.gov). The output is a disposition, not a claim that the observed sample percentage equals the true population percentage.

An **estimation study** instead asks for a confidence interval around a defect rate, perhaps within plus or minus two percentage points. Its sample size depends on the target precision, confidence level, likely defect rate, population size, and sampling design. The Census Bureau states that design and size should reflect both the required output detail and **the precision required of key estimates** (www.census.gov). A plan optimized for low-cost accept or reject decisions is not automatically adequate for estimation.

Four review modes

Table 1 compares the principal control modes. Because Docket directly provides target research and a separate review layer, it is included as a row, without treating its service description as a statistical guarantee.

Review mode	Selection and scope	Decision produced	Best use and limitation
Producer self-check	Producer reviews its own output, ideally throughout the lifecycle. Government guidance says assurance should occur at each lifecycle stage (www.gov.uk)	Correct or release records	Fast process control, but reviewer independence is limited.
Random acceptance sample	Every in-scope item has a known selection opportunity. Simple random sampling gives each item an equal chance (www.abs.gov.au)	Accept, expand, remediate, or reject the lot	Efficient for batch disposition; does not determine exact quality.
Docket separate evidence review	Triage checks identity and fit, Scout collects answers, and Audit reviews retained evidence. Unanswered questions remain recorded	A reviewed research file with visible evidence status	Provides a review workflow, but the buyer must still set tolerances and acceptance rules.
Targeted high-risk review	Review every critical field or deliberately select unusual, high-value, stale, or ambiguous records	Resolve specified risks	Necessary for severe consequences, but targeted findings cannot be projected to the whole batch.
100% field review	Examine every identity, ownership, disqualifier, or other designated item	Direct disposition for that field	Removes sampling risk for reviewed items, but still depends on clear definitions and competent review.

The modes are complementary. Producer controls prevent defects, random sampling supports a population-level decision, targeted review addresses known risk, and full review covers fields where a missed defect is unacceptable. NIST describes data-quality assessment as “not a single process” (nvlpubs.nist.gov). A mature acceptance design therefore combines the modes instead of forcing every question into one sample percentage.

Define the Lot and Defects Before Sampling

Make one lot genuinely comparable

A **lot** should contain records produced under one mandate version, schema, producer or production process, and reasonably bounded collection window. Do not combine a legacy spreadsheet with a newly automated file, or mix two analyst teams whose instructions and source access differ. If batches differ materially, treat them as separate lots or as explicit strata. NIST defines a sampling plan as “a detailed outline of which measurements will be taken” (www.itl.nist.gov). The outline should identify the population frame, unit, fields, reviewer, timing, and decision rule.

The frame must also be current. Statistics Canada advises teams to ensure the frame is as up to date as possible (www150.statcan.gc.ca). For target research, freeze a batch manifest before drawing the sample. The manifest should include a stable record identifier, producer, collection timestamp, mandate version, and hash or version for the delivered file. Web-record integrity means the retained material is complete and unaltered (www.archives.gov). Additions after selection either form a new lot or require a documented redraw.

Define defects at the field level

Defect definitions should be observable and repeatable. Recommended categories are:

- **Wrong entity:** the record, website, or evidence belongs to another company. Statistics Canada describes record linkage as identifying records associated with the same person or entity (www150.statcan.gc.ca).
- **Incorrect ownership:** the parent, subsidiary, sponsor, or independence status is wrong under the mandate's stated date and definition.
- **Incorrect screen:** the answer to a stated inclusion or disqualification criterion does not follow from the evidence or rule.
- **Unsupported claim:** the cited text exists but does not support the answer. A citation is not itself proof of attribution.
- **Stale source:** the evidence falls outside the allowed recency window or has been superseded.
- **Missing required answer:** a required field is blank without the permitted status and search record.
- **Provenance failure:** the value lacks a source, supporting excerpt, collection date, or retained evidence required by the schema. NIST defines provenance as a “historical, attributed, and documented record of a data asset” (nvlpubs.nist.gov).
- **Unrecorded not-found:** the producer checked defined sources but converted absence of evidence into a blank or unsupported negative. UK guidance recommends reporting missing data and its causes (www.gov.uk).

Each definition needs a severity level and counting rule. Decide whether two defective fields in one row count as two field defects, one defective record, or both in separate metrics. Decide how reviewer disagreement is resolved. Statistics Canada recommends documentation that clearly defines every variable

(www150.statcan.gc.ca). Without a data dictionary, apparent error rates may measure reviewer interpretation rather than producer quality.

Separate critical from descriptive fields

Assign tolerances by consequence. **Critical fields** usually include entity identity, ownership, mandate disqualifiers, and any answer that directly determines outreach or exclusion. The buyer may set zero tolerance and 100% review for them. **Major fields** might include service category, geography, and evidence-backed size indicators. **Minor fields** might include formatting or a non-decision-useful descriptor. No public authority prescribes PE-specific cutoffs; the buyer must document them.

This treatment follows a broader risk principle. FDA guidance in another domain says sampling plans should be selected according to the risk of the particular defect type (www.fda.gov). The analogy supports differentiated controls, not importing pharmaceutical rules into acquisition research.

Choose a Representative Sample

Random selection is the baseline

Generate the random sample from the frozen manifest, preserve the seed, and keep replacement selections if records become unavailable. NIST says representative data are best obtained by random sampling (www.itl.nist.gov). The reviewer should not choose rows that look interesting, easy, familiar, or suspicious if the results will support a claim about the lot.

Convenience and purposive samples can identify defects but cannot automatically support population claims. A peer-reviewed summary states that findings do not generalize “to the entire population” (pubmed.ncbi.nlm.nih.gov). EPA similarly cautions that judgmentally selected samples cannot support scientifically defensible probabilistic statements about the population (nepis.epa.gov). Keep targeted review results in a separate column and do not quietly merge them into the random-sample estimate.

Stratify when risks or producers differ

Stratification divides a population into non-overlapping groups and samples within each. Useful strata include:

- **Producer:** analyst, contractor, vendor workflow, or automation version.
- **Collection period:** especially before and after a mandate, tool, or source change.
- **Field risk:** critical, major, and minor.
- **Company type:** platform, add-on, independent, subsidiary, or unclear ownership.
- **Evidence status:** answered, not found, conflicting, or exception-flagged.

Statistics Canada describes proper strata as preferably homogeneous, mutually exclusive, and exhaustive (www150.statcan.gc.ca). When units within strata are similar, stratification can reduce standard errors (www.ons.gov.uk). In quality acceptance, its greater practical value is visibility: a pooled result cannot let 500 strong records hide 50 weak records from a new producer.

Allocate a minimum sample to every material stratum, then distribute the remainder proportionally or by risk. Record both stratum-level and overall decisions. The importance assigned to each quality characteristic depends on user needs and priorities (one.oecd.org). If one producer breaches the rule, remediate that producer's records rather than rejecting unrelated work automatically.

Account for finite populations

The binomial model treats observations as independent trials with a constant defect probability. Sampling records without replacement from a finite batch follows a hypergeometric model; NIST notes that this model applies to the count of defectives when sampling without replacement (www.itl.nist.gov). The difference is small when the sample is a small fraction of the lot, but material when review covers a large share.

For example, if a batch of **500** has exactly **25** defective records and **100** are sampled without replacement, the zero-defect probability is **0.3233%** under the hypergeometric model, versus **0.5921%** under a binomial approximation. Use an exact finite-population calculator when the sampling fraction is large. The assumptions, not just the displayed sample size, belong in the worksheet.

Build the Acceptance Plan and Worksheet

Set quality points and decision risks

An operating characteristic curve shows the probability of accepting a lot at each true defect rate (www.itl.nist.gov). Specify four inputs before inspecting the sample:

- **Good-quality point:** a defect rate the buyer intends usually to accept, such as 1%.
- **Bad-quality point:** a rate the buyer intends rarely to accept, such as 5%.
- **Producer's risk, alpha:** probability of rejecting at the good-quality point.
- **Buyer's risk, beta:** probability of accepting at the bad-quality point.

NIST reports typical alpha values from **0.20 to 0.01**, but this is a general engineering range, not a prescribed PE standard (www.itl.nist.gov). WHO's LQAS guidance likewise requires acceptable error probabilities to be specified in advance (www.who.int). An analyst should not tune the rule after seeing defects.

Use the audit workbook

Table 2 shows why “zero defects found” is not proof of zero errors. The exact one-sided 95% upper bound is calculated as `1 - 0.05^(1/n)`. The final column assumes a true 5% defect rate and calculates `(0.95)^n`, using NIST's binomial defect-count model (www.itl.nist.gov).

Random sample n	Observed defects	Exact 95% upper bound	Approximate rule of three	Chance of zero if true rate is 5%
20	0	**13.9108%**	15.0000%	**35.8486%**
30	0	**9.5034%**	10.0000%	**21.4639%**

Random sample n	Observed defects	Exact 95% upper bound	Approximate rule of three	Chance of zero if true rate is 5%
59	0	**4.9508%**	5.0847%	**4.8495%**
60	0	**4.8703%**	5.0000%	**4.6070%**
95	0	**3.1042%**	3.1579%	**0.7651%**
100	0	**2.9513%**	3.0000%	**0.5921%**
150	0	**1.9773%**	2.0000%	**0.0456%**
299	0	**0.9969%**	1.0033%	**0.000022%**
300	0	**0.9936%**	1.0000%	**0.000021%**

The table supports a quick calculator: for a desired upper bound **p** after zero defects, compute `n = ceiling[ln(0.05) / ln(1-p)]`. This yields **29** for 10%, **59** for 5%, **99** for 3%, **149** for 2%, and **299** for 1%. Exact binomial intervals are preferable when the sample or failure count is small (www.itl.nist.gov). These calculations assume independent, representative trials and a fixed defect definition. A statistical reviewer should approve a production calculator, especially for nonzero defects, finite populations, sequential expansion, or complex strata.

Table 3 is a combined worksheet, field-by-producer defect heat map, and decision log. Its figures are a **hypothetical example**, not Docket or client performance. Color labels are categorical text so the record remains accessible.

Lot / field / producer	Population and stratum	Defect definition	Good / bad rates	Risks and plan	Observed result	Heat and logged decision
L-26-09 / identity / Alpha (Hypothetical Example)	N = 800; established producer	Wrong company, domain, or legal entity	0% / any occurrence	100% review; no sampling risk accepted	800 reviewed; 1 corrected	**Red severity, green residual:** remediate correction and inspect root cause
L-26-09 / ownership / Beta (Hypothetical Example)	N = 240; new producer	Incorrect parent, sponsor, subsidiary, or independence status	0% / any occurrence	100% review	240 reviewed; 0 defects	**Red severity, green result:** accept field, retain log
L-26-09 / service category / Alpha (Hypothetical Example)	N = 800; random stratum	Unsupported category under mandate definition	1% / 5%	alpha 5%, beta 5%; n = 181, c = 4	3 defects	**Amber:** accept under numeric rule; remediate the 3 records and causes
L-26-09 / evidence provenance / Beta (Hypothetical Example)	N = 240; random stratum	Missing URL, excerpt, collection date, or retained page	1% / 5%	alpha 5%, beta 5%; n = 181, c = 4	6 defects	**Red:** expand review and remediate; numeric rule breached

Lot / field / producer	Population and stratum	Defect definition	Good / bad rates	Risks and plan	Observed result	Heat and logged decision
L-26-09 / descriptive text / all producers (Hypothetical Example)	N = 1,040; pooled only after stratum checks	Materially inaccurate description	2% / 8%	plan selected with approved exact calculator	Pending	**Grey:** no decision until sample and stratum results complete

The illustrative **n = 181, c = 4** plan has a **96.3670%** acceptance probability at 1% defects and **4.9163%** at 5% defects. A simpler **n = 60, c = 0** plan accepts a 1% defect batch only **54.7157%** of the time, imposing **45.2843%** producer risk, even though its buyer risk at 5% is **4.6070%**. The larger plan is more balanced. It also shows why “review 60 and allow no errors” can be operationally wasteful when small, correctable defect rates are acceptable.

Implement the Review and Escalation Process

Review evidence, not just values

For every sampled field, the reviewer should test the full chain:

1. **Identity:** Is the evidence about the intended company and time period?
2. **Source fit:** Is the source type appropriate for this field?
3. **Quotation:** Does the cited excerpt appear in the retained page?
4. **Entailment:** Does that excerpt support the exact answer?
5. **Mandate logic:** Was the definition applied consistently?
6. **Freshness:** Is the date within the field's recency rule?
7. **Completeness:** Is the required status populated, including not-found or conflict?
8. **Provenance:** Can the reviewer trace producer, activity, page, excerpt, and collection time?

Attribution of a source does not remove the need to assess reliability (www.gao.gov). The National Archives defines integrity for a web record as being complete and unaltered (www.archives.gov). Retaining page content, collection time, and the exact excerpt makes later review possible when live pages change.

Treat missing evidence as a result

Use explicit statuses such as **answered**, **not found after defined checks**, **conflicting**, **not applicable**, and **review required**. A blank is not a neutral outcome. Docket's public evidence framework says a not-found should be a first-class record, and its homepage states that it retains source page, excerpt, and collection date for researched answers. For acceptance testing, missing required evidence counts according to the prewritten defect definition rather than being removed from the sample denominator.

Decide, expand, remediate, or reject

Apply the rule exactly after review:

- **Accept:** observed defects do not exceed **c**, no critical field rule was breached, and each stratum passes its required check.
- **Expand:** the initial result is near the boundary, a qualitative pattern suggests clustering, reviewer disagreement is material, or the approved plan specifies a second stage. Expansion rules must be set before results are seen.
- **Remediate:** correct sampled defects and the underlying process, then determine whether affected unsampled records need targeted or full review. UK guidance says to fix quality problems as close to the source as possible (www.gov.uk).
- **Reject:** defects exceed **c**, a zero-tolerance critical rule fails, required provenance is absent at scale, or remediation cannot bound the affected population.

Always record qualitative causes even when the numeric plan passes. Three unrelated typographical defects do not imply the same remediation as three ownership errors created by one broken entity-matching rule. Qualified reviewers outside the author's unit add independence when available (www.census.gov). Census quality standards also call for another review after recommended revisions are addressed (www.census.gov). The same discipline supports a focused recheck after research-batch remediation.

Data Analysis and Evidence

The quantitative answer depends on the decision the buyer wants to defend. For zero observed defects and a one-sided 95% bound, the exact sample sizes are especially transparent: **59 records** support an upper bound below 5%, **99** below 3%, **149** below 2%, and **299** below 1%. Zero defects in **20** only excludes rates above roughly **13.91%** at that confidence level. Therefore, checking 20 rows can be a useful process smoke test, but it is weak evidence for a low-error population claim.

The more flexible acceptance design is an operating-characteristic plan. A plan with **n = 181 and c = 4**, good quality at 1%, bad quality at 5%, and both decision risks below 5% is a useful worked example. It does not become a universal recommendation merely because the mathematics balance. A 181-record review may be excessive for a 200-record lot, insufficient for a critical zero-tolerance field, or misdirected if defects are clustered by producer. The plan's quality points must express the buyer's real consequences.

Population size matters mainly when the sample is a large share. With random sampling without replacement, an exact hypergeometric calculation accounts for the shrinking remainder. The binomial approximation is generally conservative or close when the sample fraction is modest, but the workbook should record the chosen model. EPA notes that determining a minimum sample size relies on an estimate of total variability (nepis.epa.gov). Historical defect data can inform the expected rate, but should not silently replace the pre-agreed good and bad quality points.

Review cost can be forecast directly. Multiply sample records by fields per record, expected minutes per field, and reviewer cost, then add time for conflicts and remediation. If 181 records contain 12 sampled fields, the plan contains **2,172 field decisions** before exceptions. This workload is one reason to separate 100% critical-

field validation, representative record sampling, and automated structural checks. Automated checks can cover missing URLs, malformed dates, duplicate identifiers, and schema conformity across the full batch. Human review can then focus on identity, support, meaning, and mandate logic.

The quality dashboard should retain at least five measures:

- **Record defect rate:** sampled records with one or more defined defects.
- **Field defect rate:** defective field results divided by field results reviewed.
- **Critical defect count:** reported separately, never diluted by minor fields.
- **Provenance completeness:** sampled answers with all required evidence attributes.
- **Not-found integrity:** not-found results with the required sources-checked record.

These measures answer different questions. OECD guidance notes that the importance of quality characteristics depends on user needs and priorities (one.oecd.org). A sourcing team may prioritize correct identity and mandate fit, while a later diligence team may require much deeper verification. The acceptance record should state its intended use so that downstream users do not mistake screening quality for diligence completion.

Implications and Future Directions

Sampling programs improve when they accumulate structured evidence about causes, not only pass rates. Over successive batches, the team can estimate defect patterns by producer, source type, field, mandate version, and age. Those data support better stratification and targeted controls. They also reveal whether a tolerance remains appropriate. Any change to thresholds should be prospective and approved, not retrofitted to rescue a current batch.

Three developments are likely to make acceptance programs more effective:

- **Full-population structural testing:** schemas, duplicate keys, required statuses, date formats, and evidence links can be checked on every row before statistical sampling.
- **Risk-weighted review:** identity, ownership, and disqualifiers receive full or heavier review, while lower-consequence descriptions use acceptance sampling.
- **Versioned provenance:** each answer retains its producer, source, excerpt, collection time, mandate version, and derivation activity. That makes affected records discoverable when a process changes.

Independent review should remain visible. The Census Bureau's information-quality standard prefers reviewers outside the author's organizational unit when qualified reviewers are available (www.census.gov). For PE research operations, independence can mean a separate analyst, a buyer-side audit sample, or a distinct evidence-review stage. It does not require identical staffing for every field, but the reviewer should not merely confirm their own prior judgment.

Finally, teams should publish the plan beside the result. “181 randomly selected records, accept at four or fewer major defects, 1% good and 5% bad quality points, both risks below 5%” is interpretable. “We checked a sample” is not. Metadata should give reviewers enough context to understand how data was collected (www.gov.uk).

Frequently Asked Questions (FAQs)

How many records should be checked if no errors are allowed in the sample?

Choose the maximum defect rate the team wants the zero-defect result to exclude at 95% one-sided confidence. The exact minimums are **29 for 10%**, **59 for 5%**, **99 for 3%**, **149 for 2%**, and **299 for 1%**. These answers assume random, independent classification under a fixed definition. They do not mean the population has zero defects.

Is 10% of the batch a good audit sample?

Not by itself. Ten percent is 20 records in a lot of 200 but 2,000 in a lot of 20,000. Neither figure states the tolerable rate, bad-quality rate, confidence target, or acceptance number. Design to decision risk, then apply a finite-population model if the sample is a large fraction of the lot.

What is the audit sample size calculator input list?

At minimum, enter population size, sampling model, strata, good-quality defect rate, bad-quality defect rate, producer risk, buyer risk, and acceptance number constraints. For confidence-bound estimation, enter confidence level, desired upper bound or margin of error, expected rate, and finite population size. WHO's domain-specific guidance makes the transferable point that sample size is based on defined threshold values (www.who.int).

Should every field use the same tolerable defect rate?

No. Consequences differ. Entity identity, ownership, and mandate disqualifiers can justify zero tolerance and full review. Descriptive fields can use a nonzero acceptance rule. The quality standard must say which field families affect inclusion, outreach, prioritization, or merely description.

What are sensible vendor research deliverable acceptance criteria?

Criteria should include a frozen lot definition, schema and variable dictionary, defect taxonomy, field severities, randomization method, strata, sample size, accept number, evidence-retention requirements, not-found handling, escalation rule, remediation scope, and re-review requirement. GAO's framework is risk based (www.gao.gov), which supports tailoring depth to the intended use rather than applying one checklist mechanically.

When should the team expand instead of reject?

Expand only under a prewritten second-stage rule or when the initial review reveals a bounded stratum that can be isolated. Reject or require full remediation when a critical zero-tolerance rule fails or when the affected population cannot be bounded. Never keep adding records until a preferred answer appears.

Conclusion

The defensible answer is a method, not a percentage. Define a homogeneous batch, write field-level defect rules, separate critical fields from descriptive ones, and specify good quality, bad quality, producer risk, and buyer risk before drawing records. Use random selection for population-level decisions, stratify when producers or risks differ, and keep targeted review analytically separate.

For a zero-defect 95% demonstration, the memorable benchmarks are **59 records for an upper bound below 5%**, **149 for below 2%**, and **299 for below 1%**. The rule-of-three shortcut is intended for larger samples, with **n at least 100** suggested for the final approximation (faculty.washington.edu). For a balanced worked acceptance plan, **n = 181, c = 4** separates 1% good quality from 5% bad quality with each decision risk below 5%. Neither set of numbers is a universal default. The correct plan reflects consequences, field severity, lot size, expected defects, and the cost of both wrong decisions.

The operational safeguard is the worksheet: record the population, strata, definitions, tolerances, confidence or risk targets, model, sample size, accept number, observed defects, causes, and disposition. Preserve missing evidence as a defined result. Pair statistical sampling with full review of fields where sampling risk is unacceptable, full-population structural checks, and documented remediation. That combination turns an arbitrary spot check into a repeatable acceptance system for acquisition-target research.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://pcaobus.org/oversight/standards/auditing-standards/details/AS2315
2. https://www.itl.nist.gov/div898/handbook/pmc/section2/pmc21.htm
3. https://faculty.washington.edu/fscholz/DATAFILES/ConfBoundsSlides.pdf
4. https://www.itl.nist.gov/div898/handbook/pmc/section2/pmc22.htm
5. https://www.fda.gov/media/154868/download
6. https://www.who.int/docs/default-source/medicines/norms-and-standards/current-projects/guidelines-on-medicines-quality-surveys-qas15-630-30062015.pdf?sfvrsn=4869e295_2
7. https://www.gao.gov/assets/710/706666.pdf
8. https://www.gao.gov/products/gao-20-283g
9. https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework
10. https://www.census.gov/about/policies/quality/standards/standarda3.html
11. https://www.abs.gov.au/ausstats/abs%40.nsf/0/A493A524D0C5D1A0CA2571FE007D69E2
12. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/1500-18/NIST.SP.1500-18r2.html
13. https://www.itl.nist.gov/div898/handbook/ppc/section3/ppc33.htm
14. https://www150.statcan.gc.ca/n1/pub/12-539-x/2019001/process-concernant-eng.htm
15. https://www.archives.gov/records-mgmt/policy/managing-web-records.html
16. https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework-guidance

17. https://www.itl.nist.gov/div898/handbook/pmd/section2/pmd215.htm
18. https://pubmed.ncbi.nlm.nih.gov/34349313
19. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=900B0B00.TXT
20. https://www150.statcan.gc.ca/n1/pub/12-001-x/2019002/article/00006/02-eng.htm
21. https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/onsmethodologyworkingpaperseriesno9guidetocalculatingstandarderrorsforonssocialsurveys
22. https://one.oecd.org/document/std/qfs%282011%291/en/pdf
23. https://www.itl.nist.gov/div898/handbook/glossary.htm
24. https://www.itl.nist.gov/div898/handbook/pmc/section2/pmc243.htm
25. https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/exacbici.htm
26. https://www.census.gov/about/policies/quality/standards/standarde3.html

THE PUBLISHER

About Docket

Docket provides deal-origination research software and managed research for private equity firms. We help investment teams investigate acquisition targets using structured screening criteria, retained sources and reviewable company evidence. Teams can work through a self-serve platform or use managed research, depending on how they want research delivered.

Research against a defined mandate

A useful target list needs more than company names. Docket focuses on the questions that determine whether a company fits an investment mandate, including the evidence needed to support or qualify each answer. Our research approach makes the connection between screening criteria, source material and conclusions visible to the team reviewing the work.

Triage, Scout and Audit

Docket's three named research agents perform complementary tasks. Triage resolves company identity and screens fit. Scout collects sourced answers against the mandate. Audit checks retained evidence, addresses contradictions and leaves unsupported answers visibly unresolved. This structure helps reviewers distinguish established findings from missing information and questions requiring further investigation.

Evidence that supports investment-team judgment

Our research library covers market mapping, screening criteria, private-company data, succession and ownership, source evaluation and evidence standards. These resources explain the methods and limitations behind origination research. Findings support a team's judgment; they do not establish that a company is for sale or guarantee a transaction or investment outcome.

Work with Docket

Visit Docket to explore the platform, managed research and the current contact path. Read about screening criteria, evidence standards, AI research agents and market mapping.

Public examples are illustrative unless explicitly identified otherwise. Research preparation and authorized outreach are separate activities; confidential target lists and customer outcomes should never be inferred from an educational example.

Website: https://docket.capital

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.