© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723143

A Reproducible Data-Quality and Exception-Management Framework for Resource-Constrained Organizations

ALLEN TEERAHUMBA[1], EMMANUEL HAGAN[2], FLORA PHIRI[3], TREVOR KAUYU[4], MUNASHE NAPHTALI MUPA[5]

[1]*University of Louisville, ORCiD: 0009-0001-7353-9205*
[2]*Park University, ORCiD: 0009-0000-9652-472X*
[3]*George Washington University, ORCiD: 0009-0007-6433-3428*
[4]*Arizona State University, ORCiD: 0009-0001-6019-2177*
[5]*Hult International Business School*

Abstract- ***Resource-constrained small and medium-sized enterprises (SMEs) and nonprofit organizations require reliable operational reporting but often lack dedicated data-quality teams, enterprise observability platforms and continuous-audit capacity. This study develops and validates a Reproducible Data-Quality and Exception-Management Framework (RDEMF) that converts a small control library into a governed sequence of detection, risk scoring, root-cause coding, ownership, remediation and closure verification. The empirical demonstration uses the Brazilian E-Commerce Public Dataset by Olist, distributed through Kaggle, comprising 99,441 orders, 112,650 order lines, 103,886 payments, 32,951 products, 99,441 customers, 3,095 sellers and 99,224 reviews. Twelve order-level controls and complementary table-level tests assess completeness, uniqueness, validity, referential integrity, temporal consistency, reconciliation, timeliness and robust statistical anomalies. Overall, 21,715 orders (21.84%) triggered at least one rule; 6,530 (6.57%) accumulated a risk score of four or more. Late delivery affected 7,826 delivered orders (8.11%); 1,359 orders (1.37%) recorded carrier hand-off before approval; 775 orders (0.78%) had no item record; and 381 (0.38%) had an absolute payment-to-item reconciliation difference above R$0.01. Robust outlier rules flagged unusual values or durations but were treated as review candidates rather than errors. A transparent queue simulation, based on stated service-time assumptions rather than observed case handling, reduced mean completion time for high-severity exceptions from 5,275.0 to 381.4 hours under risk-priority sequencing, while total workload remained unchanged. The study contributes an auditable rule schema, entity-by-dimension and co-occurrence heat maps, an exception register, a scoring model, a minimum viable dashboard and a closure playbook. Findings demonstrate that low-cost controls can reveal concentrated reliability risks, but causal claims about remediation performance require prospective organizational pilots.***

Keywords: data quality; exception management; SMEs; nonprofit organizations; reconciliation; anomaly detection; continuous auditing; data governance; operational resilience

I. INTRODUCTION

Organizations increasingly make operational, financial and service decisions through data assembled across sales, accounting, inventory, payment, customer and reporting systems. For a small enterprise or nonprofit, the same employee may collect, reconcile and report those data. That concentration of duties does not make reliable information less important; it makes preventive controls, transparent exception queues and reproducible review procedures more important. A missing relationship, impossible timestamp sequence or unreconciled total can propagate into cash forecasts, donor reports, supplier decisions, customer communications and management dashboards before anyone sees the underlying defect.

Data quality is conventionally defined in relation to fitness for use. Wang and Strong (1996) showed that users judge quality through multiple dimensions rather than accuracy alone, while Pipino, Lee and Wang (2002) translated those dimensions into assessment methods. Contemporary reviews continue to identify completeness, accuracy, consistency and timeliness as recurring core dimensions, but also document inconsistent terminology and a gap

between abstract dimensions and executable checks (Wang et al., 2024; Ridzuan and Zainon, 2024; Miller, 2024; Papastergios, Ehrlinger and Gounaris, 2025). The implementation problem is therefore not solved by adopting a vocabulary. Organizations need rules with explicit denominators, thresholds, severity, owners and evidence of closure.

Exception detection also creates its own governance problem. A rule engine can produce more alerts than staff can investigate, particularly when unusual records are conflated with confirmed errors. Research on continuous auditing has long argued for timely automated tests, but recent work identifies exception proliferation and class imbalance as central barriers to usable systems (Alles, Kogan and Vasarhelyi, 2002; Jans, Alles and Vasarhelyi, 2014; Svanberg et al., 2025). Resource-constrained organizations cannot solve this by adding analysts indefinitely. They need a small, high-yield rule library, a defensible risk score and a workflow that separates deterministic failures from probabilistic anomalies.

The present study addresses that need through a reproducible empirical demonstration. It tests which control families expose reliability risks in a public, linked transactional dataset and converts the results into an implementation framework suitable for an SME or nonprofit using open-source software, SQL, a spreadsheet or a basic business-intelligence tool. The paper does not claim that the public dataset represents every small organization, nor that every anomaly is misconduct. Its contribution is a transparent bridge from data-quality theory to testable controls and accountable closure.

1.1 Research questions and contributions

RQ1: Which deterministic and statistical controls identify the largest concentrations of reporting and process exceptions in a linked transactional dataset?

RQ2: How do exception families overlap, and what does that overlap imply for root-cause investigation and remediation design?

RQ3: Can a transparent severity-first queue improve the modeled time to close high-risk cases without increasing total analytical workload?

The study makes four contributions. First, it operationalizes eight data-quality dimensions as executable and auditable tests. Second, it separates validation failures from robust statistical outliers. Third, it combines entity-by-dimension and exception co-occurrence heat maps to support root-cause reasoning. Fourth, it specifies a minimum viable operating model rule registry, scorecard, exception register, service levels, escalation and closure evidence that can be implemented without a specialist platform.

II. LITERATURE REVIEW

2.1 Data quality as contextual fitness for use

The influential distinction between intrinsic, contextual, representational and accessibility quality established that the usefulness of data depends on the decision and process it serves (Wang and Strong, 1996). A field can be technically non-null yet operationally incomplete; an optional review comment should not be treated like a missing delivery date on a completed order. Pipino, Lee and Wang (2002) accordingly recommend combining objective metrics with task-sensitive interpretation. ISO/IEC 25012 extends this view through a formal data-quality model, while ISO 8000 emphasizes exchange and master-data quality. The practical implication is that every rule must define its eligible population and business condition before calculating a pass rate.

Recent syntheses show both continuity and fragmentation. Wang et al. (2024) identify completeness, accuracy, timeliness and consistency among the most frequently cited dimensions. Ridzuan and Zainon (2024) find those dimensions remain central in big-data settings, whereas Miller (2024) highlights continuing terminological inconsistency. Papastergios, Ehrlinger and Gounaris (2025) map open-source tool functions to dimensions and show a many-to-many relationship: a null check may inform completeness and validity, while a relationship test may inform consistency and integrity. The RDEMF therefore reports rule-level results before aggregating them into a scorecard.

2.2 Reconciliation, event logic and continuous control

Relational business data contain structural claims. Every order item should point to an order and product; payment totals should be explainable against billed values; and event timestamps should respect the process sequence unless a documented business exception applies. Continuous-audit research proposes embedding such tests close to transaction processing so that deviations are reviewed before they accumulate (Alles, Kogan and Vasarhelyi, 2002; Vasarhelyi, Alles and Kogan, 2004). Process-mining research similarly uses event logs to compare actual and expected process paths, revealing control deviations that conventional sampling may miss (Jans, Alles and Vasarhelyi, 2014).

The accounting analytics literature also warns that access to larger datasets does not automatically improve assurance. Audit evidence must remain interpretable, governed and connected to professional judgment (Appelbaum, Kogan and Vasarhelyi, 2017). Mupa-linked scholarship has recently emphasized this operational connection. Taanisa et al. (2026), including Lucy Ganyani and Munashe Naphtali Mupa, propose a risk-based audit-analytics framework for early irregularity identification in SMEs. Chingezi et al. (2026), also including Ganyani and Mupa, connect repeat findings to remediation ownership, loss prevention and governance. These works support the use of risk-ranked exceptions and closure evidence, although the present study independently tests a defined control library rather than treating a conceptual framework as validation.

2.3 Anomaly detection is triage, not proof

Statistical anomaly detection seeks observations that differ materially from an expected pattern. Survey literature distinguishes point, contextual and collective anomalies and cautions that rarity alone does not establish error or fraud (Chandola, Banerjee and Kumar, 2009). In financial contexts, unusual values can reflect misconduct, capture error, a rare but legitimate transaction or a changing business distribution (Bolton and Hand, 2002). Robust median absolute deviation (MAD) methods are attractive for small organizations because they are transparent and less distorted by extreme observations than the mean and standard deviation (Leys et al., 2013). In the RDEMF, anomaly rules therefore create medium-severity review candidates; they never overwrite source records or assert wrongdoing.

2.4 Resource constraints, resilience and transferability

A technically optimal control may fail when its maintenance cost exceeds organizational capacity. The framework therefore favours deterministic rules, modest data requirements and explicit human review. Cross-domain work from the requested profiles illustrates the broader governance pattern. Okebu and Mupa (2026) link early detection to corrective action and competence verification in aviation training. Machekera et al. (2026) frame infrastructure reliability as a lifecycle of stage-specific controls. Mudzingwa et al. (2026), including Admore Mungwadzi and Mupa, emphasize prioritized threat analytics and resilience. These studies are not evidence about e-commerce data quality; they are used only to support the transferable design principle that detection must be connected to accountable action and verification.

III. METHODOLOGY

3.1 Design, dataset and reproducibility

The study used a quantitative secondary-data design followed by design-science synthesis. The empirical source was version 7 of the Brazilian E-Commerce Public Dataset by Olist, obtained from Kaggle. Kaggle describes the resource as approximately 100,000 Brazilian marketplace orders from 2016 to 2018, with linked customer, order, payment, product, seller, review and geolocation records (Olist, 2018). The dataset is appropriate because it represents a realistic relational workflow and includes timestamps and monetary fields that permit completeness, sequence, reconciliation, timeliness and anomaly tests. It does not identify a specific SME or nonprofit and is not used to infer fraud prevalence.

Data source: Kaggle — Brazilian E-Commerce Public Dataset by Olist

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723143

Table 1. Analytical data inventory

Table	Records	Fields	Primary analytical purpose
Orders	99,441	8	Status and event sequence
Order items	112,650	7	Product, seller, price and freight
Payments	103,886	5	Payment reconciliation
Products	32,951	9	Attribute completeness and plausibility
Customers	99,441	5	Customer-key integrity
Sellers	3,095	4	Seller-key integrity
Reviews	99,224	7	Contextual service evidence

Source: Authors' analysis of Olist (2018). Geolocation and category-translation tables were retained for provenance but excluded from the principal control scorecard because duplicated geographic coordinates and translation coverage require separate business semantics.

3.2 Control taxonomy and eligible populations

Twelve order-level controls formed the primary rule library. Tests were specified before interpretation and preserved the row-level flag, dimension, severity and weight. Deterministic failures received high or critical severity when they broke a mandatory relationship, monetary rule or event sequence. Timeliness and statistical anomalies received medium severity because they may indicate service risk or legitimate variation. A rule rate was calculated as flagged eligible records divided by eligible records; for example, a missing customer-delivery timestamp was assessed only among orders whose status was "delivered."

Table 2. Order-level rule library

ID	Dimension	Test	Severity	Weight
E01	Referential	No order-item record	Critical	5
E02	Referential	No payment record	Critical	5
E03	Completeness	Delivered order lacks delivery timestamp	High	4
E04	Consistency	Approval precedes purchase	High	4
E05	Consistency	Carrier hand-off precedes approval	High	4
E06	Consistency	Customer delivery precedes carrier hand-off	Critical	5
E07	Reconciliation	$\lvert\text{payments} - (\text{items} + \text{freight})\rvert > R\$0.01$	High	4
E08	Validity	Non-positive invoice or payment value	Critical	5
E09	Timeliness	Delivered after estimated date	Medium	2
E10	Anomaly	Invoice value robust $\lvert z\rvert > 3.5$	Medium	2
E11	Anomaly	Freight-to-item ratio robust $\lvert z\rvert > 3.5$	Medium	2

ID	Dimension	Test	Severity	Weight
E12	Anomaly	Delivery duration robust |z| > 3.5	Medium	2

Source: Authors' specification. Robust z = 0.6745(x − median)/MAD. Statistical anomaly flags identify review candidates and are not classified as confirmed data errors.

3.3 Reconciliation and anomaly methods

Order-item value was the sum of line price plus freight. Payment value was the sum of all payment records for the order. An exception was raised when the absolute difference exceeded R$0.01. This is a strict technical reconciliation; a production implementation should document legitimate causes such as vouchers, credits, refunds or marketplace adjustments before assigning culpability. Temporal controls compared purchase, approval, carrier hand-off and customer-delivery timestamps. Delivery timeliness compared the customer-delivery date with the estimated-delivery date.

For invoice value, freight ratio and delivery duration, a two-sided robust z-score greater than 3.5 was used. The threshold follows common robust-screening practice and reduces sensitivity to skewed distributions, but it does not establish an error (Leys et al., 2013). Records with missing denominators were excluded from the relevant anomaly test and handled by deterministic completeness or relationship rules.

3.4 Risk scoring, heat maps and closure simulation

Each flagged rule contributed its severity weight to an order-level score. Scores of zero, 1–2, 3–5, 6–9 and 10 or more were labelled none, low, moderate, high and critical. This additive score is intentionally interpretable; it is a triage device, not a calibrated probability model. The entity-by-dimension heat map reports exception rates across orders, items, payments, products, customers and sellers. The co-occurrence heat map reports Jaccard overlap between exception-family sets: the number of orders flagged by both families divided by the number flagged by either family.

To examine issue-resolution sequencing, each exception event was assigned a transparent service-time assumption: 10 minutes for completeness or validity, 12 for timeliness, 15 for referential integrity, 20 for temporal consistency, 25 for reconciliation and 30 for anomaly review. A fixed random seed generated a reproducible first-in-first-out (FIFO) arrival order; this was compared with severity-first processing. Completion time was cumulative analyst minutes. The scenario isolates queue order and does not represent observed staff performance, parallel work, service-level interruptions or true remediation duration.

3.5 Ethical and analytical safeguards

The dataset is public and de-identified. The analysis did not attempt re-identification. Exact source values were not altered; all flags were stored separately. The study distinguishes detection from adjudication, avoids labeling unusual transactions as fraudulent and reports scenario assumptions alongside results. Code, thresholds, eligible populations and aggregate outputs are structured to permit independent reproduction. The principal limitation is external validity: an e-commerce marketplace has different workflows from a charity, community organization or small manufacturer.

IV. RESULTS

4.1 Completeness and structural integrity

The linked dataset contained 99,441 orders and 112,650 order lines. Mandatory-field completeness was generally high when eligibility was defined by process state. Only eight delivered orders lacked a customer-delivery timestamp. Product master data were weaker: category was missing for 610 of 32,951 products (1.85%), while 611 products (1.85%) lacked at least one physical-dimension or weight field. Customer and seller state fields were complete. Structural tests found 775 orders (0.78%) without an item record and one order without a payment record. Those relationship gaps warrant investigation because they can exclude transactions from revenue,

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723143

margin, fulfilment or cash analyses depending on join direction.

Table 3. Selected mandatory-field completeness results

Field and eligible population	Missing	Eligible	Completeness
Delivered-order customer delivery timestamp	8	96,478	99.992%
Order-item product key	0	112,650	100.000%
Order-item seller key	0	112,650	100.000%
Product category	610	32,951	98.149%
Product weight	2	32,951	99.994%
Customer state	0	99,441	100.000%
Seller state	0	3,095	100.000%

Source: Authors' analysis. Optional review title and comment fields are excluded because absence is not, by itself, a quality defect

4.2 Rule yield and concentration

Across the twelve order-level controls, 21,715 orders (21.84%) triggered at least one rule. This does not mean that one fifth of orders were erroneous: the numerator includes service lateness and robust outlier candidates. The risk score identified 6,530 orders (6.57%) with scores of at least four, a narrower pool for control-owner review. The highest-volume rule was late delivery, affecting 7,826 of 96,478 delivered orders (8.11%). Robust invoice-value, freight-ratio and duration tests flagged 6.86%, 5.55% and 3.88% of orders, respectively.

Table 4. Order-level control results

Control	Family	Severity	Exceptions	Rate	Pass rate
E09 late delivery	Timeliness	Medium	7,826	7.870%	92.130%
E10 value outlier	Anomaly	Medium	6,819	6.857%	93.143%
E11 freight ratio outlier	Anomaly	Medium	5,518	5.549%	94.451%
E12 delivery duration outlier	Anomaly	Medium	3,862	3.884%	96.116%
E05 carrier before approval	Temporal consistency	High	1,359	1.367%	98.633%
E01 missing item	Referential integrity	Critical	775	0.779%	99.221%
E07 payment reconciliation	Reconciliation	High	381	0.383%	99.617%
E06 customer before carrier	Temporal consistency	Critical	23	0.023%	99.977%
E03 missing delivery date	Completeness	High	8	0.008%	99.992%
E08 nonpositive value	Validity	Critical	3	0.003%	99.997%

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723143

Control	Family	Severity	Exceptions	Rate	Pass rate
E02 missing payment	Referential integrity	Critical	1	0.001%	99.999%
E04 approval before purchase	Temporal consistency	High	0	0.000%	100.000%

Source: Authors' analysis. Rates use orders as the common denominator except the reported delivered-order lateness rate in the text.

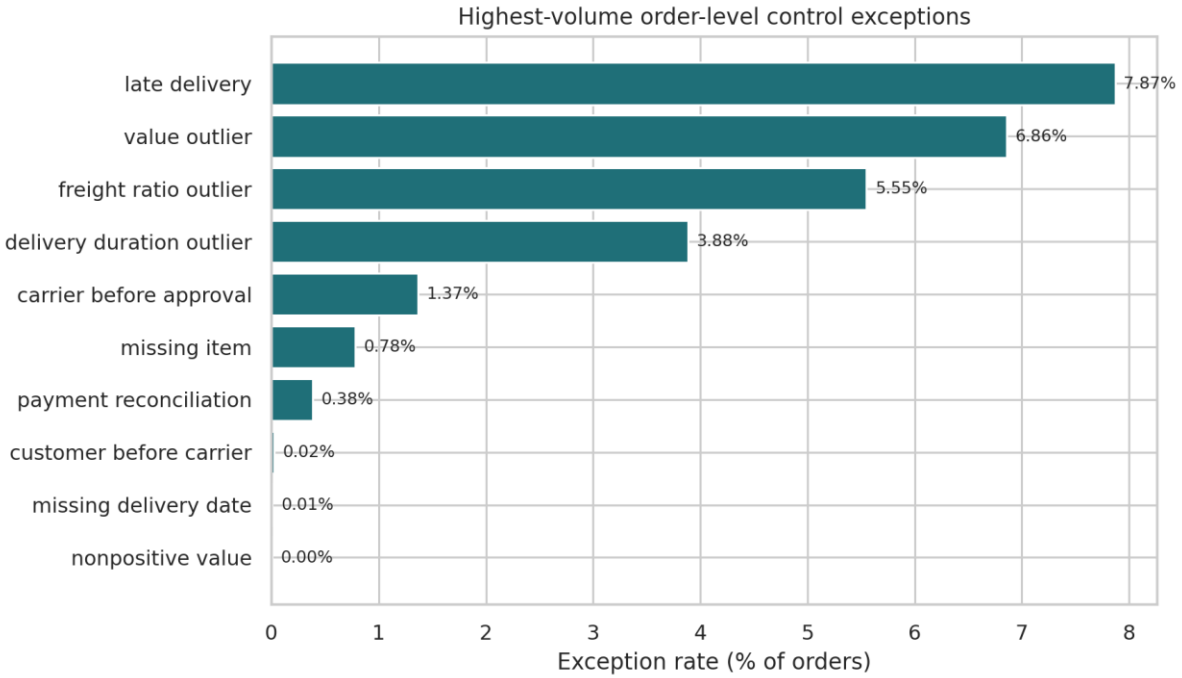


Figure 1. Highest-volume order-level control exceptions

Source: Authors' analysis of Olist (2018). Anomaly bars are screening results, not confirmed error rates.

4.3 Reconciliation and temporal consistency

Payment-to-item reconciliation identified 381 orders (0.383%) with an absolute difference above R$0.01. Among flagged orders, the median absolute difference was R$3.77, and the aggregate absolute difference was R$3,270.00. These values indicate technical non-equality, not necessarily misstatement, because the public files do not fully encode all discount, refund or adjustment semantics. The appropriate control response is to assign a reason code and obtain supporting evidence, not to auto-correct either side.

Temporal tests found 1,359 carrier hand-offs recorded before approval (1.37%) and 23 customer deliveries preceding carrier hand-off (0.023%). No approvals preceded purchase. The asymmetric result is informative: a broad timestamp-conversion failure would be expected to affect several sequence rules, whereas concentration in a particular transition may point to batching, asynchronous source updates or different event definitions. Root-cause review should therefore begin with process semantics and system lineage.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723143

4.4 Entity-by-dimension heat map

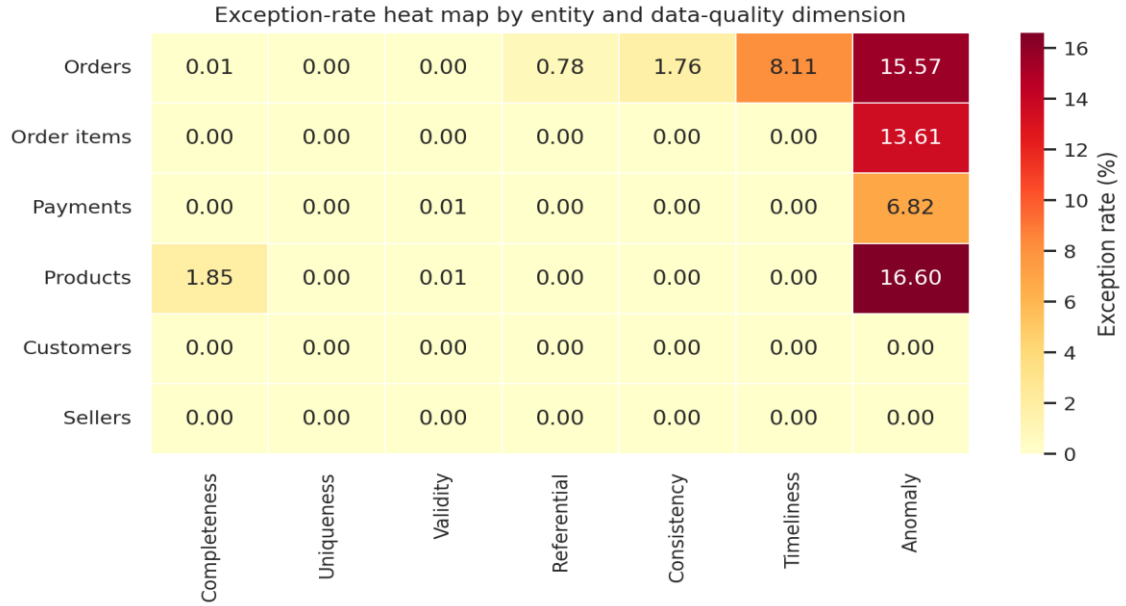
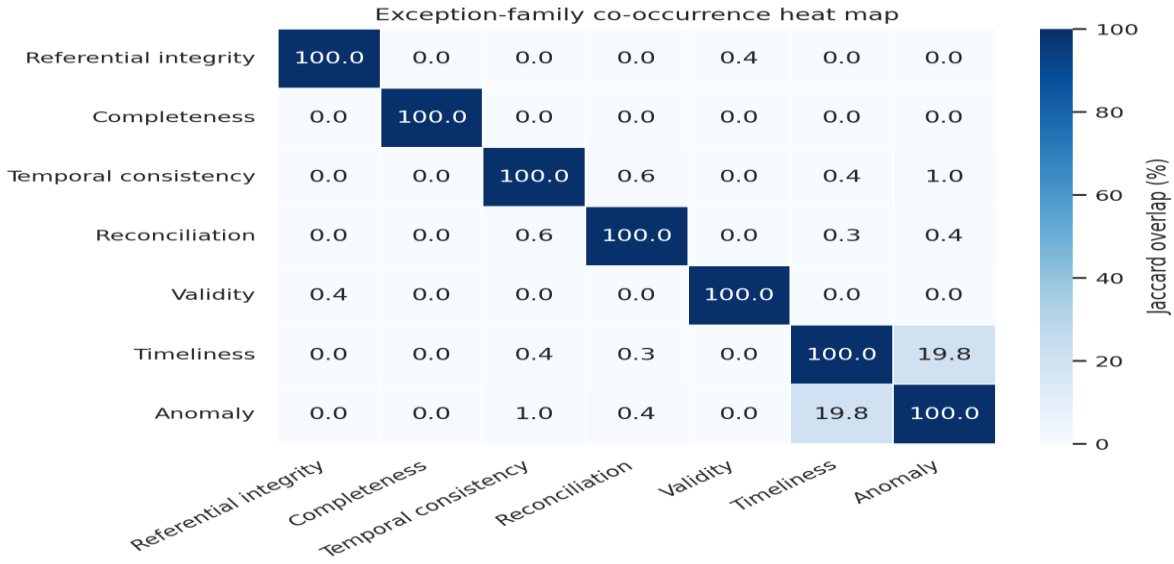


Figure 2. Exception-rate heat map by entity and data-quality dimension

Source: Authors' analysis. Cells show flagged records as a percentage of the relevant entity table; rates are diagnostic, not directly additive across dimensions.

The heat map shows why a single enterprise-wide quality percentage is inadequate. Orders carried most temporal and timeliness risk; product data carried the largest mandatory-attribute completeness issue; payments were structurally clean but contained unusual values; and order items exposed value and relationship checks. A resource-constrained organization can use this pattern to place controls at the point of origin instead of centralizing every remediation task in finance or information technology.

4.5 Exception-family co-occurrence



© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723143

Figure 3. Exception-family co-occurrence heat map
Source: Authors' analysis. Values are Jaccard percentages; the diagonal is 100%. Low overlap indicates that control families identify different record populations.

Most off-diagonal overlaps were low, indicating complementary rather than redundant tests. Timeliness did not strongly subsume value anomalies, and relationship failures did not routinely co-occur with temporal inconsistencies. This matters operationally: removing a control because another rule "usually catches the same problem" would leave distinct failure modes undetected. Where overlap is material, the exception register should retain one case with multiple rule tags rather than generating duplicate tickets.

4.6 Risk bands and closure-sequencing scenario

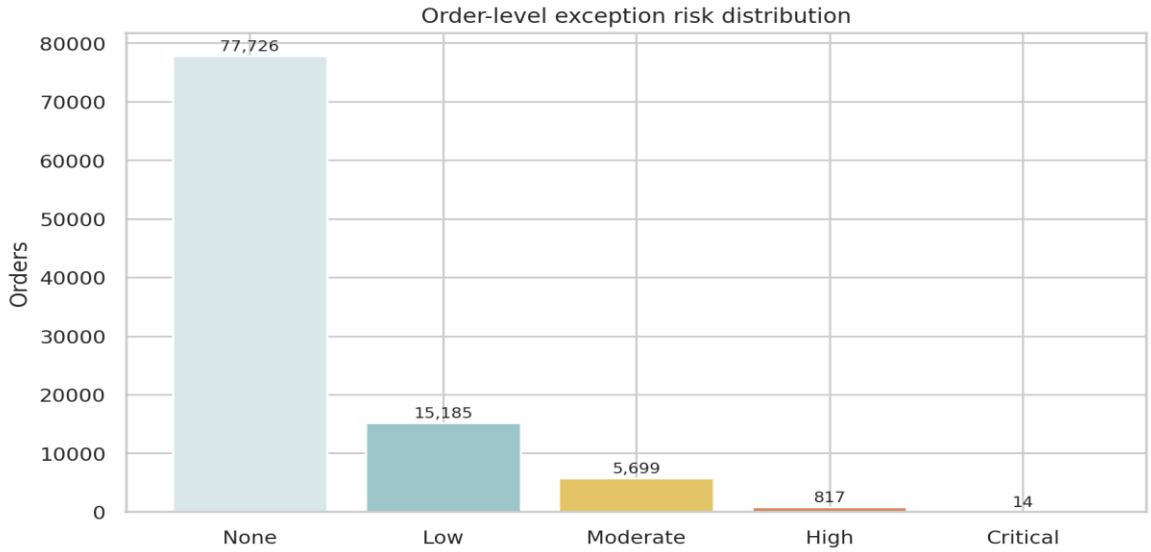


Figure 4. Order-level exception-risk distribution

Source: Authors' analysis. The score is a transparent triage index, not an estimated probability of error or fraud.

Severity-first sequencing changed when high-severity events were completed but did not reduce total assumed workload. Under FIFO, mean completion time for the 2,550 high-severity events was 5,275.0 hours; under risk priority it was 381.4 hours, a modeled reduction of 92.8%. The 90th percentile fell from 9,417.8 to 726.6 hours. Total serial effort remained 10,480.0 hours in both cases. The result demonstrates a queueing principle, not an empirical estimate of organizational savings.

Table 5. Transparent exception-queue scenario

Strategy	Events	High-severity events	Mean high closure (h)	P90 high closure (h)	Total effort (h)
FIFO	26,575	2,550	5,275.0	9,417.8	10,480.0
Risk priority	26,575	2,550	381.4	726.6	10,480.0

Source: Authors' scenario analysis. Service times are stated assumptions; single-server cumulative completion times are used only to compare sequencing policies.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723143

V. THE REPRODUCIBLE DATA-QUALITY AND EXCEPTION-MANAGEMENT FRAMEWORK

The RDEMF converts isolated tests into a closed control loop. Its minimum viable implementation contains six artifacts: a rule registry, run log, data-quality scorecard, exception register, ownership matrix and closure-evidence repository. Each artifact can be maintained in a relational database, a governed spreadsheet or an open-source workflow tool. Reproducibility requires versioned rule logic, fixed eligible populations, preserved raw values and a recorded software and data snapshot.

Table 6. RDEMF operating cycle

Stage	Required question	Minimum evidence	Control decision
1. Scope	Which decisions and reports depend on the data?	Critical data elements, owner and intended use	Exclude fields with no defined business requirement.
2. Validate	Which rules failed and on what denominator?	Rule ID, version, timestamp, eligible count, flag count	Retain row-level flags; never overwrite source silently.
3. Prioritize	Which cases threaten reporting, cash, compliance or service?	Severity, materiality, recurrence and dependency score	Route deterministic critical failures before anomalies.
4. Diagnose	What process or system condition produced the exception?	Reason code, lineage, evidence and reviewer note	Distinguish source defect, timing, policy exception and false positive.
5. Remediate	Who will correct or accept the risk, and by when?	Owner, action, SLA, status and approval	Escalate overdue high-risk cases.
6. Verify	Did the fix resolve the case and prevent recurrence?	Re-run result, reconciled evidence and closure approval	Close only after validation passes or documented risk acceptance.
7. Learn	Which rules or processes should change?	Recurrence trend, false-positive rate and control performance	Tune thresholds; add preventive controls at source.

5.1 Rule registry and scorecard

The rule registry is the control system of record. Each rule should contain a stable ID, name, dimension, business rationale, dataset and fields, eligible population, test expression, threshold, severity, owner, run frequency, expected pass rate, effective date and version. A code change must create a new version so that historical rates remain interpretable. The scorecard should show both exception count and eligible denominator; a lower count can result from fewer eligible records rather than improvement.

5.2 Root-cause taxonomy

A compact root-cause taxonomy prevents free-text fragmentation. Recommended codes are: source-entry error; missing mandatory input; broken interface or join; asynchronous event update; master-data defect; unauthorized override; documented policy exception; legitimate rare event; rule-design false positive; and unresolved. The reviewer records one primary cause and optional contributing causes. Recurrence should be measured by rule, source process and cause—not merely by ticket count.

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723143

5.3 Risk scoring and escalation

Risk scores should remain explainable. A practical case score can combine rule severity, monetary materiality, downstream dependency, recurrence and age. The present additive rule score is a starting point. Before production deployment, leaders should define overrides: a low-value referential failure may still be critical if it affects a statutory report, and a high-value anomaly may be legitimate after documentation. Critical cases require immediate owner notification; high cases receive a short service level; medium anomaly candidates are reviewed in batches unless compounded by other rules.

5.4 Exception register and closure evidence

Table 7. Minimum exception-register fields

Field group	Required fields	Purpose
Identity	Case ID, rule ID/version, source key, detection timestamp	Reproduce the finding and prevent duplicate cases.
Risk	Severity, score, materiality, affected report/process	Support transparent prioritization.
Diagnosis	Root-cause code, reviewer, evidence link, false-positive flag	Separate technical detection from adjudication.
Action	Owner, action, target date, status, escalation level	Create accountability and visible ageing.
Closure	Re-run result, reconciled evidence, approver, closure timestamp	Verify correction or document accepted risk.
Learning	Recurrence link, preventive-control action, rule-tuning decision	Reduce repeat findings and alert fatigue.

5.5 Minimum viable dashboard

The dashboard should privilege decisions over decoration. The top line reports eligible records, exception rate, high-risk open cases, overdue cases, median age and closure rate. A second layer shows the entity-by-dimension heat map, weekly rule trend and exception ageing. A third layer permits drill-through to case evidence. False-positive rate and recurrence rate are essential: without them, teams may celebrate detection volume even while control quality deteriorates.

Table 8. Dashboard indicators and definitions

Indicator	Definition	Management use	Review cadence
Rule pass rate	1 − flagged eligible records / eligible records	Monitor a specific control without denominator distortion.	Each run
High-risk open cases	Open cases above approved severity threshold	Allocate scarce review capacity.	Daily/weekly
SLA breach rate	Overdue cases / cases due in period	Identify ownership or capacity failure.	Weekly
Median time to triage	Median detection-to-first-review time	Test whether alerts reach a human promptly.	Monthly

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-1723143

Indicator	Definition	Management use	Review cadence
Verified closure rate	Cases passing re-test or approved risk acceptance / cases resolved	Prevent administrative closure without evidence.	Monthly
Recurrence rate	Closed cases with same rule/cause recurring within window	Measure preventive-control effectiveness.	Monthly/quarterly
False-positive rate	Reviewed cases judged non-actionable / reviewed cases	Tune rules and protect analyst attention.	Monthly

5.6 Ninety-day implementation plan

Days 1–15 establish scope: select one critical report or workflow, identify five to ten critical data elements and assign business and technical owners. Days 16–30 profile the data and implement three deterministic controls with explicit denominators. Days 31–45 create the exception register, reason codes and service levels; run in observation mode without auto-correction. Days 46–60 train reviewers, adjudicate false positives and revise thresholds. Days 61–75 add reconciliation and one transparent anomaly rule. Days 76–90 publish the scorecard, review recurrence and decide whether to expand. A nonprofit can substitute grant, donor, beneficiary or program-service keys for commercial order entities without changing the governance sequence.

VI. DISCUSSION

6.1 Which controls added the most practical value?

Volume alone did not determine value. Timeliness and anomaly controls produced the largest queues, but referential, reconciliation and temporal-sequence rules had stronger direct implications for reporting reliability. An order without items can disappear from an inner-joined revenue report; an event sequence can distort cycle-time metrics; and a non-reconciled payment can affect cash or liability analysis. The results support a tiered control portfolio: first protect structural and monetary integrity, then monitor service timeliness, and finally add anomalies where review capacity exists.

The low overlap between families further argues against replacing a multi-dimensional library with a single “data health” metric. Fitness for use is conditional. A dataset can be nearly complete yet temporally inconsistent, or structurally valid yet operationally late. This empirical pattern is consistent with the multidimensional tradition from Wang and Strong (1996) and the tool-to-dimension mapping described by Papastergios, Ehrlinger and Gounaris (2025).

6.2 Exception management under scarcity

The scenario analysis showed that priority sequencing can accelerate attention to severe cases without changing total workload. This is precisely why risk scoring matters for a small team. However, the extreme simulated completion times also expose a second lesson: prioritization cannot cure an oversized rule library. Staff must suppress duplicates, batch medium-severity anomalies, retire low-value rules and invest in preventive fixes. Svanberg et al. (2025) describe exception proliferation as a core continuous-audit problem; the RDEMF answers it with a deliberately small registry, severity differentiation, co-occurrence consolidation and false-positive monitoring.

6.3 Human judgment and accountable automation

Automation should detect and route; people should adjudicate material cases and approve closure. This division is especially important for anomaly rules. A very high-value transaction may be correct, while a small reconciliation difference may indicate a repeatable systems defect. The framework therefore preserves the original record, explanation, evidence and reviewer. It also makes risk acceptance explicit. These safeguards align with accounting-analytics arguments that richer data do not remove the need for traceable professional judgment (Appelbaum, Kogan and Vasarhelyi, 2017).

6.4 Implications for SMEs and nonprofits

For an SME, the framework can begin with sales-to-cash: customer keys, invoice completeness, payment reconciliation and order-cycle timestamps. For a nonprofit, an analogous design can test donor restrictions, grant-period validity, beneficiary identifiers, approved budgets, procurement evidence and program-output reporting. The reusable component is not the Olist business rule itself; it is the rule contract eligible population, expression, denominator, severity, owner, action and verification. That distinction permits adaptation without pretending that one dataset or threshold is universal.

The requested author-profile literature reinforces this implementation emphasis. Taanisa et al. (2026) frame analytics as a way to strengthen SME controls; Chingezi et al. (2026) emphasize repeat-finding reduction through accountable remediation; Okebu and Mupa (2026) link early warning to corrective action; Machekera et al. (2026) place reliability across a lifecycle; and Mudzingwa et al. (2026) emphasize resilience through prioritized risk response. The present paper advances that shared logic with an independently reproducible data test, while avoiding claims that cross-domain studies directly validate the measured e-commerce rates.

VII. LIMITATIONS AND FUTURE RESEARCH

First, the public data describe a Brazilian marketplace between 2016 and 2018, not a representative sample of SMEs or nonprofits. Second, source documentation does not fully specify the semantics of vouchers, refunds, event batching or every missing field; therefore reconciliation and sequence flags require business adjudication. Third, the analysis measures detection yield, not actual reporting error, loss, fraud or customer harm. Fourth, the risk weights are normative and transparent rather than statistically calibrated. Fifth, MAD screens are univariate and may flag legitimate tail behavior or miss contextual anomalies. Sixth, the queue simulation uses stated service-time assumptions, random FIFO arrivals and one serial analyst; it does not estimate causal productivity gains. Seventh, the article validates detection on one snapshot and cannot measure recurrence reduction after remediation.

Future research should run authorized, de-identified pilots in consenting SMEs and nonprofits. A stepped implementation could measure pre/post exception recurrence, mean time to triage, verified closure, report restatement, user trust and staff workload. Rule thresholds should be validated against adjudicated outcomes. Comparative studies should test simple robust rules against isolation forests or explainable gradient boosting, but only where case volume supports model validation. Qualitative interviews should examine whether owners understand scores, whether root-cause codes remain stable and whether dashboard use changes decisions. External replication should publish rule versions and denominators so that performance differences can be separated from changes in scope.

VIII. CONCLUSION

This study demonstrates that a modest, transparent control library can expose distinct concentrations of relational, temporal, monetary, timeliness and anomaly risk in a complex public transaction dataset. Of 99,441 orders, 21.84% triggered at least one rule, but only 6.57% accumulated a score of four or more. The distinction matters: most flags were late-delivery or anomaly candidates, while smaller populations contained the structural and reconciliation issues most directly relevant to reporting reliability. Heat maps showed that defects clustered differently across entities and rarely collapsed into one redundant family.

The Reproducible Data-Quality and Exception-Management Framework converts those findings into an operating model for resource-constrained organizations: define fitness for use, version executable rules, preserve row-level evidence, prioritize by transparent risk, code root causes, assign actions, verify closure and learn from recurrence. The framework's strongest claim is practical rather than technological. Reliable reporting does not require an opaque platform to begin; it requires a controlled path from an explainable exception to an accountable and evidenced resolution.

REFERENCES

[1] Alles, M.G., Kogan, A. and Vasarhelyi, M.A. (2002) 'Feasibility and economics of continuous assurance', *Auditing: A Journal of Practice & Theory, 21*(1), pp. 125–138. Crossref

[2] Appelbaum, D., Kogan, A. and Vasarhelyi, M.A. (2017) 'Big Data and analytics in the modern audit engagement: Research needs', *Auditing: A Journal of Practice & Theory, 36*(4), pp. 1–27. Crossref

[3] Bolton, R.J. and Hand, D.J. (2002) 'Statistical fraud detection: A review', *Statistical Science, 17*(3), pp. 235–255. Crossref

[4] Chandola, V., Banerjee, A. and Kumar, V. (2009) 'Anomaly detection: A survey', *ACM Computing Surveys, 41*(3), Article 15. ACM

[5] Chingezi, E., Chingezi, L., Ganyani, L., Yelduora, P.G. and Mupa, M.N. (2026) 'Reducing repeat findings and control failures in operationally intensive U.S. enterprises: A data-driven internal audit framework for remediation, loss prevention, and governance improvement', *World Journal of Advanced Research and Reviews, 30*(3), pp. 2188–2199. Crossref

[6] ISO (2009) *ISO/IEC 25012:2008 Software engineering—Software product Quality Requirements and Evaluation (SQuaRE)—Data quality model*. Geneva: International Organization for Standardization. ISO (Accessed: 2 September 2026).

[7] Jans, M., Alles, M.G. and Vasarhelyi, M.A. (2014) 'A field study on the use of process mining of event logs as an analytical procedure in auditing', *The Accounting Review, 89*(5), pp. 1751–1773. Crossref

[8] Leys, C., Ley, C., Klein, O., Bernard, P. and Licata, L. (2013) 'Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median', *Journal of Experimental Social Psychology, 49*(4), pp. 764–766. ScienceDirect

[9] Machekera, G., Mtemeli, P.S., Dongo, M.M. et al. (2026) 'From trench safety to lifecycle reliability: A project-control framework for water and sewer reticulation delivery in U.S. communities', *World Journal of Advanced Research and Reviews*. ResearchGate (Accessed: 2 September 2026).

[10] Miller, R. (2024) 'A framework for current and new data quality dimensions', *Data, 9*(12), 151. MDPI

[11] Mudzingwa, K., Nyamutswa, S.M., Mungwadzi, A.T., Dhegwa, P.Y. and Mupa, M.N. (2026) 'Cybersecurity of critical infrastructures: Challenges, empirical threat analytics and future resilience perspectives', *Iconic Research and Engineering Journals, 10*(2). ResearchGate (Accessed: 2 September 2026).

[12] Okebu, D.T. and Mupa, M.N. (2026) 'Predictive training analytics for competency-based pilot instruction: Early detection of checkride risk, procedural noncompliance and safety-relevant learning gaps', *World Journal of Advanced Research and Reviews*. ResearchGate (Accessed: 2 September 2026).

[13] Olist (2018) *Brazilian E-Commerce Public Dataset by Olist*, version 7. Kaggle. Kaggle (Accessed: 2 September 2026).

[14] Papastergios, V., Ehrlinger, L. and Gounaris, A. (2025) 'Unfolding data quality dimensions in practice: A survey', *arXiv:2507.17507*. arXiv

[15] Pipino, L.L., Lee, Y.W. and Wang, R.Y. (2002) 'Data quality assessment', *Communications of the ACM, 45*(4), pp. 211–218. ACM

[16] Ridzuan, F. and Zainon, W.M.N.W. (2024) 'A review on data quality dimensions for big data', *Procedia Computer Science, 234*, pp. 341–348. ScienceDirect

[17] Svanberg, J., Öhman, P., Samsten, I., Neidermeyer, P.E., Rana, T. and Danielson, M. (2025) 'Addressing the exception prioritization problem in continuous auditing systems with thresholding', *Intelligent Systems in Accounting, Finance and Management, 32*(4), e70022. Wiley

[18] Taanisa, T., Mukwata, N.A., Sydney, J., Ganyani, L., Maturure, R.N., Chingezi, E., Yelduora, P.G. and Mupa, M.N. (2026) 'AI-enabled audit analytics for SME financial reporting and anomaly detection: A risk-based

framework for early irregularity identification and control strengthening', *World Journal of Advanced Research and Reviews, 30*(3), pp. 1113–1126. Crossref

[19] Vasarhelyi, M.A., Alles, M.G. and Kogan, A. (2004) 'Principles of analytic monitoring for continuous assurance', *Journal of Emerging Technologies in Accounting, 1*(1), pp. 1–21. Crossref

[20] Wang, J., Liu, Y., Li, P., Lin, Z., Sindakis, S. and Aggarwal, S. (2024) 'Overview of data quality: Examining the dimensions, antecedents and impacts of data quality', *Journal of the Knowledge Economy, 15*(1), pp. 1159–1178. Springer

[21] Wang, R.Y. and Strong, D.M. (1996) 'Beyond accuracy: What data quality means to data consumers', *Journal of Management Information Systems, 12*(4), pp. 5–33. Taylor & Francis