2026 Volume: 02 Issue: 11 www.ijrpa.com ISSN 2456-9995 Review Article

International Journal Research Publication Analysis

Page: 01-08

COUNTERFACTUAL DATA-QUALITY IMPACT AUDITING IN TABLEAU DASHBOARD A CONCEPTUAL FRAMEWORK FOR KPI RELIABILITY AND BUSINESS DECISION TRANSPARENCY

***Shubhangi Balaji and Dr. Kiran Tikar**

Sadhu Vaswani Institute of Management Studies for girls, Pune – 41100.

Article Received: 22 September 2026
Article Revised: 12 October 2026
Published on: 01 November 2026

***Corresponding Author: Shubhangi Balaji**
Sadhu Vaswani Institute of Management Studies for girls, Pune – 41100.

DOI: https://doi-doi.org/101555/ijrpa.2738

ABSTRACT

Business intelligence dashboards help organizations make sense of sales, customer activity, inventory, and other areas of performance. However, a dashboard can present a precise-looking number even when the data behind it contain duplicate transactions, missing values, inconsistent labels, or outdated records. When these problems are not visible, decision-makers may rely on indicators that do not fully represent the situation they are trying to understand. This conceptual paper proposes Counterfactual Data-Quality Impact Auditing (CDQIA), an approach for examining how a key performance indicator (KPI) may change when a documented data-quality treatment is applied. The proposed framework connects six activities: identifying a possible data issue, tracing it to affected KPIs, recording the treatment rule, recalculating the KPI under that rule, interpreting the size and significance of the change, and communicating the result to dashboard users. The framework also considers interactive filters, since a data issue may affect one dashboard view differently from another. CDQIA does not assume that a corrected value is automatically the truth; its usefulness depends on the quality of the evidence and the transparency of the assumptions. The paper develops conceptual propositions and identifies questions for future empirical research. As a theoretical study, it does not report experiments or claim measured improvements. Its main contribution is a decision-oriented way to make the possible effects of data-quality problems easier to examine and discuss.

KEYWORDS: Data quality; Tableau; business intelligence; counterfactual analysis; KPI reliability; decision transparency.

1. INTRODUCTION

Organizations use business intelligence (BI) dashboards to follow performance and support everyday decisions. A manager may use a sales dashboard to compare regions, a marketing team may monitor customer conversion, and an operations team may review inventory levels. Tools such as Tableau make it possible to explore these indicators through charts, filters, and interactive views. Yet a clear chart does not necessarily mean that the information behind it is reliable.

Consider a dashboard that reports monthly revenue. If some transactions have been recorded twice, the total may be higher than the actual revenue. Missing records can have the opposite effect, while inconsistent product names can divide one product category into several groups. A poorly defined join may also multiply rows and distort totals. In each case, the dashboard may calculate the KPI correctly from the data it receives, while the resulting number still gives a misleading picture of business activity.

Finding a data-quality problem is therefore only one part of the task. Users also need to know which indicators might be affected, what assumptions are used to handle the problem, and whether the resulting change could influence a decision. For example, a small change in a KPI may not affect a decision, while a modest change near a sales target or risk threshold may matter a great deal.

This paper introduces Counterfactual Data-Quality Impact Auditing (CDQIA) as a conceptual approach to this issue. In this paper, counterfactual means comparing the KPI calculated from the data currently available with the KPI calculated under a clearly stated alternative treatment of the data. The comparison does not prove causality and does not guarantee that the alternative value is correct. Instead, it helps make the consequences of a particular assumption visible.

1.1 Problem Statement

Data-quality checks often identify missing, duplicated, invalid, or inconsistent records, but an alert alone may not explain how the problem affects a specific KPI or a business decision. Dashboard users may see the reported number without seeing the assumptions, uncertainty, or potential impact associated with the underlying data. The problem addressed in this paper is the lack of a clearly structured connection between a data-quality issue, its treatment, the resulting KPI change, and the interpretation presented to decision-makers.

1.2 Aim and Objectives

The aim is to develop a conceptual framework for explaining how data-quality issues may influence dashboard KPIs under explicit data-treatment assumptions. The objectives are to: (i) connect data-quality problems with the KPIs they may affect; (ii) distinguish issue detection from the decision about how to treat the issue; (iii) describe a transparent comparison between reported and counterfactual KPI values; and (iv) identify propositions that can guide future evaluation of the framework.

1.3 Research Questions

RQ1: How can data-quality issues be linked to the KPIs they may affect?

RQ2: What assumptions should be disclosed when a KPI is recalculated under an alternative data-treatment policy? RQ3: How can uncertainty and traceability be communicated to dashboard users? RQ4: What questions should future studies investigate to assess the usefulness of CDQIA?

2. LITERATURE REVIEW AND CONCEPTUAL BACKGROUND

Data quality has long been treated as more than simple correctness. Wang and Strong (1996) showed that data users value several dimensions, including accuracy, relevance, completeness, and accessibility. The importance of a particular dimension depends on the intended use of the data. For a sales dashboard, for example, duplicate transactions may directly affect revenue, while missing customer identifiers may primarily affect customer-level analysis.

Research on data-quality management has also examined how organizations can assess, measure, and improve information quality. Pipino, Lee, and Wang (2002) discussed approaches to measuring data quality, while Batini et al. (2009) reviewed methods for data-quality assessment and improvement. This work provides a basis for identifying and managing quality problems. However, a quality score or a list of failed checks may not by itself show a business user how a particular problem changes a KPI.

A second relevant area is business intelligence and performance measurement. KPIs are not just numbers: their meaning depends on how they are defined, which records are included, the time period, the filters applied, and the aggregation rules used. If these choices are unclear, users may interpret two apparently similar measures differently. A reliable audit should therefore document the KPI definition as well as the data treatment.

A third area is counterfactual reasoning. In general terms, counterfactual analysis asks what an outcome would look like under a different condition or assumption. In the context of data quality, this can be used to compare a KPI calculated from the observed dataset with the KPI calculated after a specified treatment, such as removing confirmed duplicates or excluding records that violate a documented validity rule. Such a comparison is only as defensible as the treatment rule. A missing value, for instance, should not automatically be replaced with an assumed value unless the basis for that replacement is explained.

The present paper brings these areas together around a practical question: after identifying a data issue, how can a dashboard user understand its possible effect on a decision-relevant indicator? The proposed contribution is a conceptual integration, not a claim that prior research has ignored data quality or counterfactual reasoning entirely. A full journal submission should expand this review through a documented search of relevant databases and should compare CDQIA carefully with the closest existing approaches.

3. MATERIALS AND METHODS

3.1 Research Design

This study uses a conceptual research design. It does not collect survey responses, conduct an experiment, train a machine-learning model, or evaluate a deployed Tableau system. Its purpose is to define the proposed concept, connect relevant research themes, develop a coherent sequence of activities, and formulate propositions for later testing.

3.2 Literature Synthesis Approach

The conceptual discussion draws on four connected themes: data-quality dimensions and assessment; KPI definition and interpretation; counterfactual comparison; and the communication of evidence in interactive dashboards. These themes were selected because the proposed framework depends on all four. The synthesis is narrative and selective rather than a systematic review. It should not be described as exhaustive unless a reproducible database search, search strings, date range, and screening procedure are added.

3.3 Key Terms

A data-quality issue is a record or dataset condition that conflicts with a requirement for a particular use. A KPI is a defined measure used to monitor performance. A data-treatment policy is an explicit rule for retaining, correcting, excluding, imputing, or flagging records. The observed KPI is calculated from the dataset as currently supplied. The counterfactual KPI is calculated under a stated alternative treatment policy. The KPI impact is the difference

between those two values, interpreted in the context of the measure and the assumptions used. An audit trail records the issue, policy, calculations, and unresolved uncertainties.

4. PROPOSED CDQIA FRAMEWORK

CDQIA is organised into six stages. The stages are intended to make the reasoning behind a KPI comparison visible and repeatable, rather than to prescribe one universal cleaning rule.

Stage 1: Identify the data-quality issue

The process begins by describing the suspected issue, such as duplicate transaction IDs, missing values, invalid quantities, inconsistent category names, or records that are no longer current. The issue should be supported by a rule or evidence, not merely by an assumption that an unusual value must be wrong.

Stage 2: Link the issue to affected KPIs

The next step is to identify which KPIs depend on the affected records or fields. Duplicate sales records may influence total revenue, order counts, and average order value. Missing customer IDs may affect customer counts or segmentation, even if the revenue amount remains usable. This mapping helps focus the audit on the measures that could actually change.

Stage 3: Document the treatment policy

A rule is then selected and documented. Depending on the evidence, a record may be retained, corrected, excluded, imputed, or flagged for review. The choice should include a reason and acknowledge uncertainty where necessary. For example, two records with the same transaction ID may be duplicates, but the same customer making two purchases on the same day is not enough evidence to remove either record.

Stage 4: Recalculate the KPI

The KPI is calculated under the existing data and then recalculated after applying the documented policy. The two values must use the same KPI definition, time window, and relevant filter context so that the comparison is meaningful. The paper refers to the second value as counterfactual because it represents an alternative treatment, not a guaranteed ground truth.

Stage 5: Interpret the impact and uncertainty

The numerical difference should be interpreted in context. A change can be expressed as an absolute difference and, where meaningful, a percentage difference. The audit should also consider whether the change crosses a target, threshold, ranking, or other decision boundary.

If the treatment policy is uncertain, the result should be presented as conditional on that policy rather than as a definitive correction.

Stage 6: Communicate the result in Tableau

A dashboard could present the original KPI, the counterfactual KPI, the difference, the reason for the treatment, and a short note about uncertainty. Users should be able to trace the comparison to the issue and policy that produced it. Where feasible, the display should preserve the selected filters and time period so that the audit remains relevant to the view being examined.

5. CONCEPTUAL REPRESENTATION AND PROPOSITIONS

Let K(D, F) represent a KPI calculated from dataset D under dashboard filter context F. Let D_0 be the observed dataset and D_1 be the dataset produced after applying a documented data-treatment policy. The observed and counterfactual values can be represented as $K_0 = K(D_0, F)$ and $K_1 = K(D_1, F)$. The impact is $\Delta K = K_1 - K_0$. If the KPI supports a meaningful percentage comparison and K_0 is not zero, a relative change can also be calculated as $100 \times (K_1 - K_0) / K_0$.

This notation is deliberately simple. It clarifies what is being compared but does not determine whether the treatment policy is correct. The same policy should be applied consistently, and the KPI definition and filter context should remain comparable. When the original value is zero, a percentage change may be undefined or misleading; the absolute difference should be reported instead.

The framework leads to six propositions for future investigation:

P1. Linking a data-quality issue to the specific KPIs that depend on it should make the potential business impact easier to interpret than presenting the issue as an isolated warning.

P2. Clearly documenting the data-treatment policy should help users understand why the counterfactual KPI differs from the reported KPI.

P3. Reporting both the KPI difference and its decision context should be more informative than reporting the numerical difference alone.

P4. Preserving the active filters and time period should make the audit more relevant to interactive dashboard analysis.

P5. Showing uncertainty and unresolved records should reduce the risk that users interpret a counterfactual value as unquestionable truth.

P6. Maintaining a traceable record of the issue, policy, and recalculation should support review and accountability.

These are propositions, not findings. They require empirical evaluation before claims can be made about their effects on comprehension, trust, or decision quality.

6. DISCUSSION

The central idea of CDQIA is that data-quality monitoring should not stop at reporting that something is wrong. For a decision-maker, the more useful question is often whether the issue matters to the KPI currently being used and what happens to the interpretation if a defensible treatment is applied. Connecting data issues to KPI changes can help analysts explain why a number may need to be treated cautiously.

The framework also separates a numerical change from a decision change. For example, if a reported KPI is comfortably above a target, a small reduction after removing confirmed duplicates may not change the decision. If the same KPI is close to a target, even a small adjustment could affect whether the target appears to have been achieved. The importance of an impact therefore depends on business context, not only on its size.

Filter awareness is another important consideration. A defect may have little effect on a company-wide total but a substantial effect on a particular region or product category. In an interactive dashboard, the audit should therefore make clear which filters and time period were active when the comparison was calculated. Otherwise, users may compare results that do not refer to the same subset of data.

There are practical limits. Some records cannot be corrected confidently from the available evidence. Different defensible policies may produce different counterfactual values. A dashboard explanation can also become crowded if every issue is shown at once. Implementations would need to balance transparency with usability, perhaps by displaying a concise summary and allowing users to open supporting details when needed.

6.1 Theoretical and Practical Implications

Theoretically, CDQIA offers a way to connect data-quality assessment with KPI interpretation and counterfactual reasoning. Practically, it suggests that dashboard designers and data stewards should document the reason for a treatment, identify affected metrics, preserve the calculation context, and communicate uncertainty. The framework is intended to support informed judgement, not replace data governance or domain expertise.

6.2 Limitations and Future Research

This paper is conceptual and has not validated CDQIA in a real organization or through a user study. The propositions remain untested, and the framework may require changes for different industries, KPI types, or governance environments. Future work could implement a small prototype in Tableau, compare alternative treatment policies on a transparent dataset, and conduct user studies to test whether the explanations improve understanding. Research could also examine how to represent uncertainty without overwhelming dashboard users.

7. CONCLUSION

A dashboard KPI can be calculated correctly and still be misleading when the underlying data contain quality problems. Identifying those problems is important, but decision-makers also need to understand which indicators may be affected, how a stated treatment changes the result, and what uncertainty remains. This paper proposed Counterfactual Data-Quality Impact Auditing as a conceptual framework for connecting those steps.

The six-stage framework moves from issue identification to KPI mapping, policy documentation, recalculation, impact interpretation, and dashboard communication. Its main principle is that a counterfactual KPI should always be presented with the assumptions that produced it. It is not automatically more accurate than the reported value. Since this is a theoretical proposal, its usefulness must be established through implementation and empirical evaluation. Future studies should test whether the framework helps users interpret dashboard evidence more carefully and make better-informed decisions.

8. REFERENCES

1. Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. ACM Computing Surveys, 41(3), Article 16. https://doi.org/10.1145/1541880.1541883
2. Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. Communications of the ACM, 45(4), 211–218. https://doi.org/10.1145/505248.506010
3. Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems, 12(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099
4. Guidotti, R. (2024). Counterfactual explanations and how to find them: Literature review and benchmarking. Data Mining and Knowledge Discovery. https://doi.org/10.1007/s10618-022-00831-6