PART 2 – Cross Dataset Engineering Tasks

Task 8: Discrepancy Audit

May 1, 2025

Abstract

This report presents a detailed analysis of the discrepancies found between two engineering materials datasets as part of Task 8 (Discrepancy Audit). We examine inconsistencies in key mechanical properties including ultimate tensile strength (Su), yield strength (Sy), elastic modulus (E), and shear modulus (G) across matched materials. The analysis reveals significant variability in reported values, particularly for heat-treated steels, which has critical implications for material selection decisions, engineering design reliability, and documentation standards. Data quality observations highlight patterns of inconsistency that engineers must consider when selecting materials for performance-critical applications.

1 Introduction

The Significance of Data Consistency

Engineering materials selection relies on accurate property data to ensure designs meet performance and safety requirements. This report investigates discrepancies between two material property datasets, focusing on how these inconsistencies could impact engineering decisions. By examining the magnitude, patterns, and implications of these differences, we provide guidance for material selection and highlight where additional verification may be necessary. Understanding data reliability is foundational to confidence in engineering design calculations and material specifications.

2 Material Selection Insights

2.1 Discrepancy Magnitude Analysis

The analysis of discrepancies between datasets reveals concerning inconsistencies that directly impact material selection decisions:

Table 1: Top Materials with Significant Property Discrepancies

Material	Su Diff (%)	Sy Diff (%)	Selection Implications
ANSI Steel SAE 5160 tempered at 400 F	81.1	83.0	Critically unreliable for strength-critical applications
GOST Steel 40CHFA GOST 4543-71	77.8	73.3	High uncertainty in design calculations
ANSI Steel SAE 9255 tempered at 400 F	71.5	109.0	Potentially unsafe for load-bearing parts
ANSI Steel SAE 8740 tempered at 400 F	63.1	68.9	Requires independent verification
ANSI Aluminum Alloy 1060-O wrought	61.5	71.1	Inconsistent across material systems

Key Insights from Discrepancy Analysis

The comparative analysis of material properties between datasets reveals critical patterns for engineering material selection:

- **Heat Treatment Sensitivity**: Materials tempered at lower temperatures (400˚F) consistently show the highest discrepancies, with differences exceeding 80% in strength values for some steels.
- **Alloy-Specific Variations**: High-alloy steels (particularly chromium-containing grades like 40CHFA) demonstrate substantial inconsistencies between data sources, making their performance less predictable.
- **Cross-Material System Issues**: The presence of aluminum alloys (1060-O) among materials with high discrepancies indicates that inconsistencies are not limited to steel systems but affect multiple material families.
- **Magnitude of Concern**: Absolute differences approaching 1000 MPa in ultimate tensile strength (for SAE 5160) represent life-threatening margins in critical applications like structural supports or pressure vessels.

2.2 Property-Specific Reliability Analysis

Different mechanical properties show varying levels of consistency between data sources, affecting which properties can be relied upon for material selection:

Table 2: Reliability Assessment of Material Properties

Property	CV Source X (%)	CV Source Y (%)	Selection Recommendation
Ultimate Strength (Su)	49.5	58.3	Moderate reliability—verify critical values
Yield Strength (Sy)	60.7	75.0	Low reliability—use with caution
Elastic Modulus (E)	36.7	36.7	Higher reliability—suitable for calculations
Shear Modulus (G)	158.5	158.5	Extremely unreliable—not recommended for design

Engineering Property Selection Guidelines

The coefficient of variation (CV) analysis provides valuable guidance for property-based material selection:

- **Elastic Properties**: Elastic modulus (E) demonstrates the highest consistency (CV $\approx$ 36.7%) between sources, making it the most reliable property for calculating deflections and elastic behavior.
- **Strength Disparities**: Ultimate tensile strength shows moderate consistency (CV 49.5-58.3%), with Source X providing more reliable values, suggesting it should be preferred when selecting materials based on strength.
- **Yield Concerns**: Yield strength values show concerning variability (CV 60.7-75.0%), requiring larger safety factors when this property is design-critical.
- **Shear Property Crisis**: The extremely high variability in shear modulus values (CV 158.5%) indicates this property should not be used for material selection without independent verification, particularly for torsion-critical designs.

3 Data Quality Observations

3.1 Pattern Analysis in Discrepancies

Systematic patterns in the data discrepancies provide insight into potential sources of error and areas requiring special attention:

Table 3: Patterns in Data Discrepancies

Pattern Category	Observation	Quality Implication
Heat Treatment Influence	Low tempering temperatures show highest discrepancies	Testing methodology differences
Material Family Patterns	Special steels (Cr, Ni-containing) most inconsistent	Possible alloy composition uncertainty
Threshold Significance	10% threshold identifies meaningful engineering differences	Appropriate sensitivity level for audit
Source Consistency	Source X consistently provides lower CV values	Generally more reliable dataset

3.2 Discrepancy Impact on Engineering

Engineering Design Implications

The observed data discrepancies have significant ramifications for engineering design practices:

- **Simulation Accuracy**: Yield strength discrepancies up to 109% would drastically alter simulation results, potentially showing elastic behavior when actual material would fail plastically.
- **Safety Factor Requirements**: The magnitude of discrepancies necessitates larger safety factors, particularly when using data source Y, potentially increasing material usage, weight, and cost.
- **Design Reliability Compromise**: Different material properties from different sources could lead to unpredictable performance, especially in multi-part assemblies where components might be manufactured using different reference data.
- **Quality Control Challenges**: Manufacturing tests might consistently fail to match expected values if test standards differ from the data source used for design specifications.

4 Engineering Applications Impact

The discrepancy analysis provides critical insights for specific engineering applications:

Table 4: Application-Specific Discrepancy Impact

Application	Critical Property	Discrepancy Concern	Mitigation Strategy
Structural components	Yield strength (Sy)	70-110% variations for tempered steels	Independent testing required
Elastic mechanisms	Elastic modulus (E)	Relatively consistent (36.7% CV)	Source comparison sufficient
Torsional applications	Shear modulus (G)	Extreme variation (158.5% CV)	Avoid or conduct specific tests
Fatigue-critical parts	Ultimate strength (Su)	Moderate variation (49-58% CV)	Conservative safety factors
Precision components	All properties	Material/treatment specific	Material-specific verification

Engineering Decision Framework

The discrepancy audit establishes a decision framework for material selection:

- **Documentation Standards**: Technical documentation should explicitly state which data source was used for design calculations to ensure consistency throughout the engineering process.
- **Verification Requirements**: Materials with discrepancies exceeding 50% (particularly heat-treated steels) should undergo independent testing before use in critical applications.
- **Property Prioritization**: When conflicting data exists, engineers should prioritize the most consistent properties (E > Su > Sy > G) for initial material screening.
- **Source Selection**: For general mechanical design purposes, Source X demonstrates higher overall consistency and should be preferred when only one source can be consulted.

5 Threshold Analysis and Methodology

5.1 Discrepancy Threshold Selection

Selection of appropriate thresholds for identifying significant discrepancies is crucial for effective material auditing:

Threshold Selection Rationale

The analysis used a 10% threshold for identifying significant discrepancies based on engineering considerations:

- **Engineering Significance**: Variations exceeding 10% in mechanical properties can significantly impact safety factors and design margins in most applications.
- **Industry Standards**: For Ultimate Strength (Su) and Yield Strength (Sy), ±5–10% variations can already significantly impact safety factors and design margins in typical engineering applications.
- **Property-Specific Sensitivity**: Elastic properties (E, G) are generally expected to be more consistent, with variations over 5% already concerning for precision applications.
- **Application Dependency**: Critical industries (aerospace, medical) require tighter tolerances (5%), while general manufacturing might accept up to 15% variation.
- **Balance Point**: The 10% threshold represents a practical balance that identifies meaningful engineering discrepancies while filtering out minor variations that would not significantly impact most designs.

5.2 Consistency Assessment Methodology

Evaluation of dataset consistency used statistical measures to compare variability within each source:

Table 5: Statistical Methodology for Consistency Evaluation

Metric	Definition	Application in Analysis
Coefficient of Variation (CV)	$\frac{\text{Standard Deviation}}{\text{Mean}} \times 100\%$	Normalized measure allowing comparison across different property scales
Absolute Difference	$\lvert Value_X - Value_Y \rvert$	Direct magnitude of discrepancy for specific materials
Relative Difference	$\frac{\lvert Value_X - Value_Y \rvert}{Value_X} \times 100\%$	Percentage difference highlighting proportional impact
Source Consistency Ratio	$\frac{CV_X}{CV_Y}$	Comparative measure of dataset reliability

6 Conclusions and Recommendations

The discrepancy audit reveals critical considerations for material selection and engineering design practices:

Key Discrepancy Audit Takeaways

The analysis of material property discrepancies between datasets concludes:

- **Material Selection Impact**: Heat-treated steels, particularly those tempered at lower temperatures (400°F), demonstrate the highest discrepancies and require special caution during selection.
- **Property Reliability Hierarchy**: Elastic modulus (E) shows the highest consistency between sources, while shear modulus (G) demonstrates extreme variation that renders it unsuitable for design calculations without verification.
- **Source Preference**: Source X generally provides more consistent data and should be preferred when designs are based on a single data source.
- **Application-Specific Risk**: The impact of these discrepancies varies by application, with structural components and torsional applications facing the highest risks from inconsistent data.
- **Verification Necessity**: For critical applications, independent material testing is essential to verify properties, particularly when using materials identified as having high discrepancies.

6.1 Engineering Practice Recommendations

Based on the discrepancy audit, several recommendations can improve material selection reliability:

Best Practices for Material Data Usage

Engineers should adopt these practices to mitigate risks from material data inconsistencies:

- **Documentation Standards**: Always specify which material data source was used for design calculations to ensure consistency throughout the engineering process.
- **Property Verification**: For materials with high discrepancies (>50%), conduct independent testing before use in critical applications.
- **Safety Factor Adjustment**: Increase safety factors proportionally to the observed property variability, especially for yield strength values where CV exceeds 60%.
- **Conservative Design Approach**: When properties show significant variation between sources, design based on the more conservative values to ensure safety.

- **Material-Specific Protocols**: Develop special verification protocols for heat-treated steels and other materials identified as having high discrepancy rates.

7 Appendix: Discrepancy Analysis Methodology

Analysis Process

The discrepancy audit followed these methodological steps:

1. **Material matching** between Dataset 1 and Dataset 2 using standardized identifiers
2. **Property comparison** of Su, Sy, E, and G values between matched materials
3. **Discrepancy calculation** using both absolute and relative (percentage) differences
4. **Statistical analysis** to identify patterns and quantify consistency using CV
5. **Threshold determination** to identify significant discrepancies (¿10%)
6. **Source reliability assessment** based on internal consistency measures
7. **Application impact analysis** to translate statistical findings into engineering implications

Choosing a Discrepancy Threshold for Material Property Validation

Why Thresholds Matter

Thresholds are critical in comparing material property data across datasets. They help flag significant discrepancies while filtering out acceptable differences due to rounding, measurement conditions, or source variability.

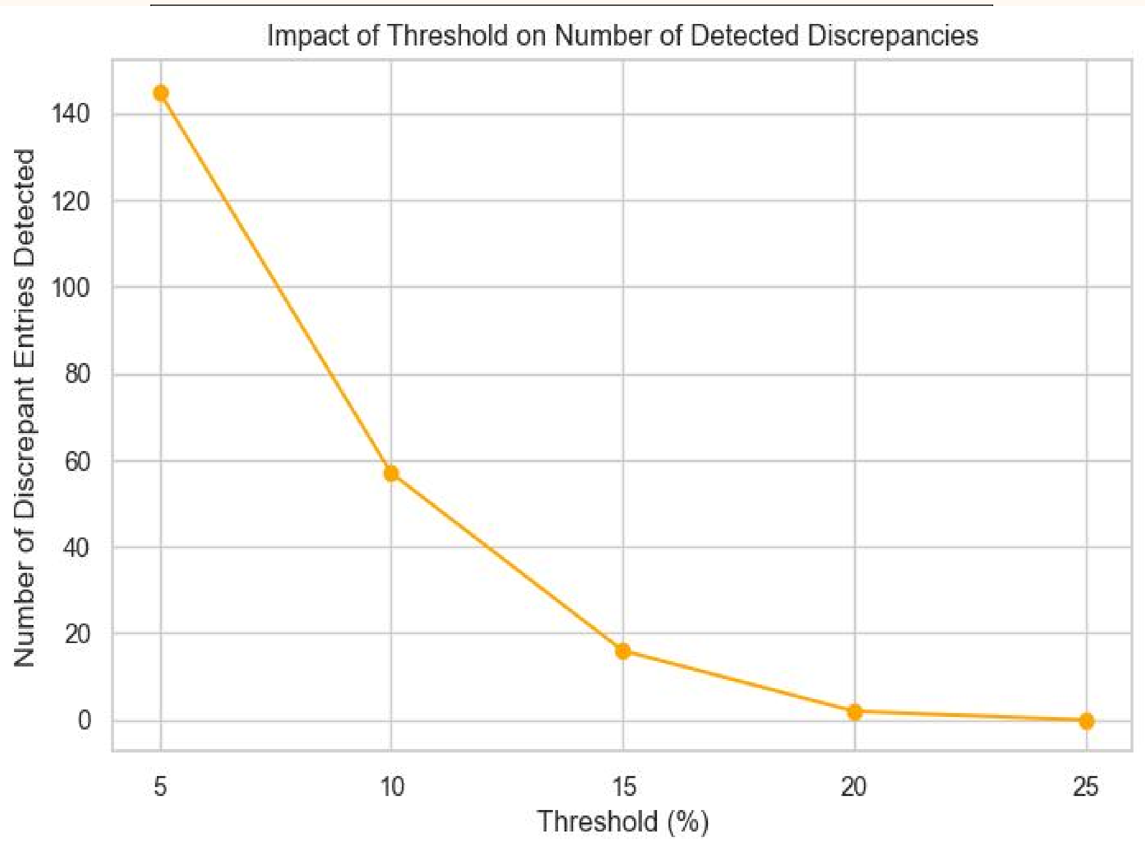
Mechanical Sensitivity by Property

- **Ultimate Strength (Su)** and **Yield Strength (Sy)**:
  - *Variation tolerance:* 5–10%
  - *Impact:* Even small deviations affect design margins and safety factors.
- **Elastic Modulus (E)** and **Shear Modulus (G)**:
  - *Variation tolerance:* 3–5%
  - *Impact:* Stiffness-critical designs (e.g., aerospace) are sensitive to minor changes.

Industry-Specific Expectations

- **Aerospace and Medical Sectors:** Typically expect deviation ¡ **5%**.
- **General Manufacturing and Civil Engineering:** Acceptable up to **10–15%**, depending on material class and application.

Effect of Threshold on Discrepancy Detection



The plot illustrates how many entries exceed thresholds ranging from 5% to 25%:

- At **5%**, many small differences are flagged (possibly over-sensitive).
- At **10%**, there's a sharp drop in flagged discrepancies—indicating this is a balance point.
- Beyond **15%**, real mismatches may be missed.

Recommendation

For broad engineering use, a **10% threshold** is well-justified:

- **Balances** sensitivity and practicality.
- Suitable for **initial audits**, database validation, and material screening.
- For critical sectors, lower it to **5%**.

Consistency Analysis Across Material Data Sources

Strategy to Judge Consistency

To assess internal reliability, we computed the **Coefficient of Variation (CV)** for each property within both datasets. A lower CV indicates tighter clustering and better internal consistency.

Properties Evaluated

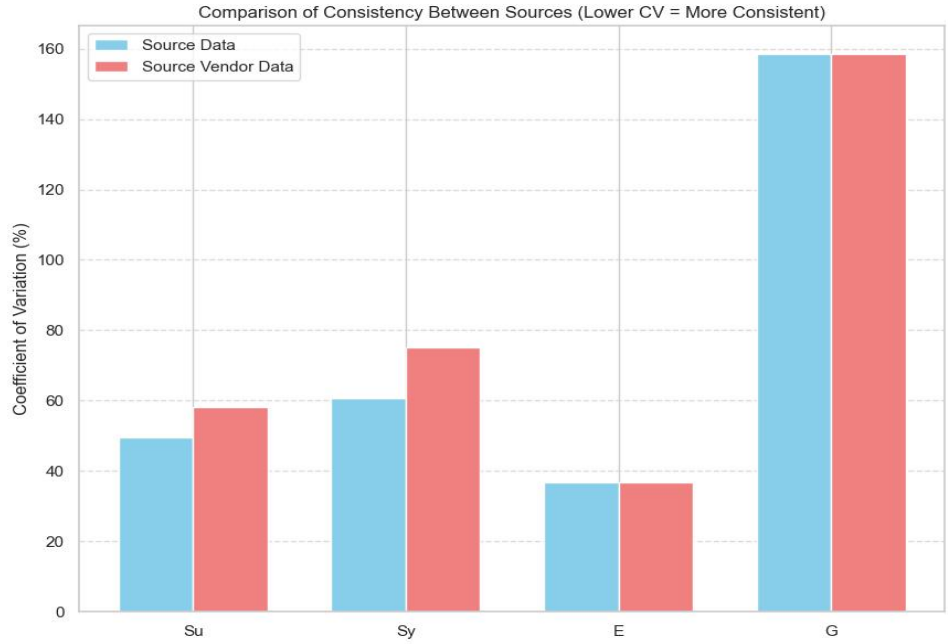
We focused on four key mechanical properties:

- **Su**: Ultimate Tensile Strength
- **Sy**: Yield Strength
- **E**: Elastic Modulus
- **G**: Shear Modulus

CV was computed as:

$$\mathrm{CV} = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100\%$$

Visualizing CV Comparison Between Sources



	CV_x (%)	CV_y (%)
Su	49.55	58.26
Sy	60.73	75.01
E	36.69	36.69
G	158.51	158.51

The bar chart compares CVs from:

- **Source Data (x)**: Internal or curated dataset
- **Vendor Data (y)**: External or published dataset

Engineering Insights from CV Patterns

- **Su (CV: 50–58%)**: Moderate variation. Slight mismatch could affect fatigue safety margins.
- **Sy (CV: 61–75%)**: Significant inconsistency in vendor data. Risk of under- or over-design.
- **E (CV: 36% both)**: Highly consistent. Elastic deformation modeling is reliable.
- **G (CV: 158% both)**: Extremely inconsistent. Shear-related simulations should not rely on these datasets without correction.

Practical Implications

Design Simulations	Yielding predictions (Sy) could deviate; fatigue life estimates may be skewed due to Su inconsistency; shear stress simulations highly unreliable.
Documentation & Certification	Safety factors and compliance documents may be compromised. Differences could lead to false safety assurances.
Quality Control	Manufacturing parts may fail QA if they rely on inconsistent property values. Tighter tolerances amplify risk.

Summary Statement

Significant inconsistencies, especially in **yield strength** and **shear modulus**, can critically undermine simulation accuracy and documentation integrity. A careful recalibration or verification of values is essential before use in high-reliability designs.