Volume 7 • Issue 2 • 2025

Research Article

Journal of Pharmacy and Drug Development

Quantifying Structural Similarity and Informational Diversity in Informed Consent Forms: A Statistical Analysis of Terminated versus Completed Clinical Trials

Author (Escientific Publishers)*

Institutional Research Limited

linguistic complexity, over-standardisation, and poor communication are pervasive (Grant, 2021; Kadam, 2017). Multiple studies have noted that ICFs often exceed recommended readability levels, employ technical jargon, and prioritise legal protection over of the process (Ssali et al., 2017; Bhupathi and Ravi, 2017). Notably, the focus on procedural adherence rather than communicative clarity has resulted in consent documents that may be signed without

Citation: Escientific Publishers. (2025). Quantifying Structural Similarity and Informational Diversity in Informed Consent Forms: A Statistical Analysis of Terminated versus Completed Clinical Trials. *Journal of Pharmacy and Drug Development* 7(2).

genuine informed engagement. Furthermore, while ethical guidelines such as the Declaration of Helsinki and the Belmont Report emphasise participant understanding as a cornerstone of research ethics, practical application remains inconsistent and often unveri-

document length and readability metrics.

By systematically interrogating the linguistic fabric of consent documentation, this study advances the ethical discourse on informed consent from descriptive critique to quantitative evaluation, providing actionable insights for improving both ethical standards and operational trial outcomes.

Methods

Study Design and Data Source

A retrospective cross-sectional study design was employed. Data were extracted from ClinicalTrials.gov, a publicly accessible regis-

Statistical Analysis

Descriptive statistics were calculated for word counts, similarity scores, and ethical clause counts. Mann-Whitney U tests compared terminated and completed trials across linguistic features. Logistic regression models were constructed to predict trial termination, with model evaluation based on Akaike Information Criterion (AIC), McFadden's R^2, and the Hosmer-Lemeshow goodness-of-fit test.

Software and Tools

Data extraction, text processing, and document similarity analyses were performed using Python (version 3.12.1). Statistical analyses, including Mann-Whitney U tests and logistic regression modelling

Visualisations indicated greater clause inclusion frequency among completed trials, supporting the hypothesis that richer ethical disclosures may enhance trial success. This difference was statistically

The mean cosine similarity for completed trials was higher (Mean = 0.607) than for terminated trials (Mean = 0.539), suggesting greater internal linguistic coherence among ICFs from successfully completed studies. Boxplots illustrated an upward shift in similarity distribution for completed trials relative to terminated trials. A Mann-Whitney U test confirmed that the difference was statistically significant (p = 0.0000173).

Ethical Clause Coverage

Completed trials exhibited a higher mean number of ethical clauses identified within their ICFs compared to terminated trials.

				higher semantic similarity
H2	group_mean_jaccard	0.0000000598	Significant (****)	Completed trials have significantly higher ethical clause overlap
H3	group_mean_edit_dist	0.165	Not significant (ns)	No significant difference in structural edit distance

Table 2: *Statistical Test Results for Group Differences in Linguistic Characteristics.*

Note: *Statistical significance thresholds were defined as p < 0.05 (), p < 0.01 (), p < 0.001 (), and p < 0.0001 (**).*

Regression Results

A logistic regression analysis was conducted to examine the asso-

with rising group mean cosine similarity, following a sigmoidal logistic pattern with an inflection around 0.55–0.60.

Higher internal linguistic coherence across informed consent forms (ICFs) was associated with substantially greater trial completion probabilities, supporting Hypothesis 1.

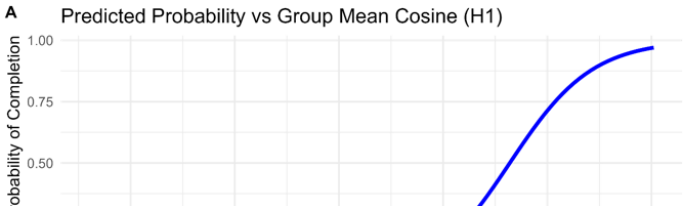
non-significant ($\chi^2(8) = 5.16$, $p = 0.741$), indicating no evidence of model miscalibration.

As shown in Table 3, group mean cosine similarity was a significant positive predictor of trial completion, supporting Hypothesis 1

H3: Group Mean Edit Distance (Structural Consistency)

In contrast, group mean edit distance demonstrated only a modest, near-linear relationship with trial outcome, with minimal impact on completion probability.

This finding corroborates the earlier statistical result, providing no strong support for Hypothesis 3.



a significant positive effect (H2). In contrast, group mean edit distance, despite yielding a large odds ratio (OR ≈ 1.50×10^{22}), exhibited a wide and unstable confidence interval crossing 1, indicating no statistically reliable association with trial outcome (H3).

Predictor	Odds Ratio	95% CI	95% CI

larity metrics, clause count, and trial outcome (completed versus terminated).

Group mean cosine similarity demonstrated an extremely large odds ratio (OR ≈ 9.67×10^{10}), with a confidence interval entirely above 1, confirming a strong and statistically significant association with trial completion (H1). Similarly, group mean Jaccard similarity showed a very large odds ratio (OR ≈ 1.16×10^{15}), also supporting

This study demonstrated that internal linguistic consistency and ethical clause richness within informed consent forms (ICFs) are significantly associated with clinical trial success. Trials with higher semantic similarity and greater ethical disclosure were more likely to reach completion, while structural formatting consistency alone showed no predictive value. Additionally, greater document complexity, reflected in increased clause counts, was associated with reduced likelihood of trial completion.

Interpretation

These findings suggest that excessive standardisation and structural uniformity may undermine the communicative purpose of

Kadam, 2017	Informed consent compre-	Highlight- ed barriers in low-	No structural document	Intro- duced edit distance

		autonomy		
Trung et al., 2021	Underre- porting of ethics and incentives	Major gaps in ethics reporting in trials	No con- sent text feature analysis	Applied struc- tural and semantic measures to success prediction

Strengths and Limitations

Strengths

This study represents the first large-scale, systematic application of natural language processing (NLP) techniques to informed consent forms (ICFs), quantitatively linking linguistic characteristics to clinical trial outcomes. By combining advanced semantic embedding models (Sentence-BERT) with manual ethical clause detection, it offers a novel integration of machine learning and bioethical analysis, contributing uniquely to both clinical trial operations and research ethics literature. The use of a real-world, publicly

available dataset enhances transparency and reproducibility. Furthermore, rigorous statistical methodologies, including Mann-Whitney U tests, logistic regression modelling, and model calibra-

development, beyond mere compliance with regulatory templates. Ethics committees should evaluate not only content completeness but also linguistic coherence. Regulators could consider guidance

corporated to investigate mediating pathways between document characteristics and trial outcomes. Stratified analyses by therapeutic area are warranted, as differing disease contexts may moderate the relationship between consent quality and trial success. Lastly, examining the impact of culturally tailored ethical clauses on participant engagement and retention would provide valuable insights for global clinical research practices.

Implications for Practice

The findings highlight the need for sponsors and investigators to prioritise semantic clarity and ethical transparency in ICF

Access is unrestricted and provided to promote transparency, reproducibility, and further research development.

References

1. Adrian A. M., Norwood S. H., Mask P. L. (2005). Producers' perceptions and attitudes toward precision agriculture technologies. Computers and Electronics in Agriculture. 48(3): 256–271.
2. Ssali, A., Poland, F., & Seeley, J. (2017). Exploring informed consent in HIV clinical trials: A case study in Uganda. Heliyon, 3(2): e00196.

3. Rebers, S., Aaronson, N. K., van Leeuwen, F. E., & Schmidt, M. K. (2016). Exceptions to the rule of informed consent for research with an intervention. BMC Medical Ethics, 17(9).
4. Association of Clinical Research Professionals (ACRP). (2013). The Process of Informed Consent. ACRP White Paper.
5. Manti, S., & Licari, A. (2018). How to obtain informed consent
16. Enpr-EMA Working Group on Ethics. (2021). Assent / Informed Consent Guidance for Paediatric Clinical Trials with Medicinal Products in Europe. European Medicines Agency (EMA). EMA/671028/2019.
17. Kadam, R. A. (2017). Informed consent process: A step further towards making it meaningful. Perspectives in Clinical Re-

search Involvement and Engagement, 7, 51.
13. Health Research Authority (HRA). (2018). Applying a proportionate approach to the process of seeking consent: HRA guidance. Version 1.02.
14. Schwarz G. (2025). Informed Consent Considerations. ACT-EU Workshop, European Medicines Agency.
15. Geetter J. S., Siegfried S. (2023). FDA Issues Final Guidance on Informed Consent in Clinical Investigations. McDermott Will & Emery.

Each ICF was scanned for the presence of these clauses using case-insensitive regex matching. Detection was considered binary (presence/absence) for each clause per document.