Check for updates

RESEARCH ARTICLE

SPIRIT-based prediction of clinical trial completion in intervertebral disc degeneration

[version 1; peer review: awaiting peer review]

Francis Chemorion[1,2], Jerome Noailly[1], Marc-Antonio Bisotti[2]

[1]Universitat Pompeu Fabra Escola d'Enginyeria, Barcelona, Catalonia, Spain
[2]Research and Development, InSilicoTrials Technologies, Trieste, Friuli-Venezia Giulia, 34137, Italy

v1 **First published:** 17 Sep 2025, **5**:285
https://doi.org/10.12688/openreseurope.21092.1
Latest published: 17 Sep 2025, **5**:285
https://doi.org/10.12688/openreseurope.21092.1

Abstract

Introduction

Clinical trial protocols in intervertebral disc degeneration (IDD) vary in quality, which may influence trial outcomes. While the SPIRIT checklist provides a standard for protocol quality, adherence varies. This study aimed to assess IDD trial protocols' adherence to SPIRIT guidelines and to use machine learning to identify which checklist items predict trial completion, thereby providing evidence-based guidance for improving trial success.

Methods

We conducted a predictive modelling study of 18 randomized controlled IDD trials from ClinicalTrials.gov. Trial success was defined as "completed" (n=11) versus "failed" (n=7). Adherence was assessed using a validated 64-item SPIRIT checklist. Four machine learning algorithms were employed with cross-validation due to the small sample size. Feature selection on the highest performing model identified the SPIRIT items the most predictive of clinical trial completion, with SHAP values used for model interpretability.

Results

Adherence varied widely across protocols (range: 11.1%–100%). The XGBoost model achieved the highest performance (ROC-AUC: 0.839, Average Precision: 0.773). The most predictive SPIRIT items related to statistical planning, safety oversight, and operational procedures, including handling of missing data, data monitoring committee structure, and plans for managing adverse events. These findings

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

suggest that specific protocol components, particularly those related to analysis and safety planning, are associated with trial completion in IDD research. Improving adherence to these areas may help reduce trial failure and enhance research efficiency.

Discussion

Our machine learning analysis successfully identified critical SPIRIT items strongly associated with trial completion in IDD research. The predominance of statistical methodology, safety monitoring, and operational planning items as top predictors provides actionable guidance for protocol development. Prioritizing analytical planning, comprehensive safety oversight, and detailed operational protocols can significantly improve trial success rates and reduce research waste in IDD clinical research.

Keywords
SPIRIT guidelines, protocol standardization, predictive modelling, clinical trial protocol, trial success, SHAP analysis

This article is included in the Marie-Sklodowska-Curie Actions (MSCA) gateway.

This article is included in the Horizon 2020 gateway.

Corresponding author: Francis Chemorion (Francis.chemorion@upf.edu)

Author roles: Chemorion F: Conceptualization, Data Curation, Methodology, Validation, Visualization, Writing – Original Draft Preparation; **Noailly J**: Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Bisotti MA**: Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project was supported by the Marie Skłodowska-Curie Actions through Disc4All - Training network to advance integrated computational simulations in translational medicine, applied to intervertebral disc degeneration under grant agreement number 955735.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Chemorion F *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Chemorion F, Noailly J and Bisotti MA. **SPIRIT-based prediction of clinical trial completion in intervertebral disc degeneration [version 1; peer review: awaiting peer review]** Open Research Europe 2025, **5**:285 https://doi.org/10.12688/openreseurope.21092.1

First published: 17 Sep 2025, **5**:285 https://doi.org/10.12688/openreseurope.21092.1

Introduction

Background

Clinical trial protocols directly facilitate the generation of primary evidence on treatment efficacy, effectiveness, and safety in medical research[1]. These documents serve as the blueprint for conducting clinical trials, outlining the study's objectives, design, methodology, statistical considerations, and organization[2]. The publication and accessibility of these protocols is crucial for mitigation of selective reporting[3], enhancement of internal validity[2], reduction of statistical bias[4] and increased reproducibility[5]. In cases where novel experimental approaches are being employed, the publication of protocols is particularly beneficial, as it allows for public scrutiny and peer review, especially in situations with high potential for bias[3].

Components and importance of clinical trial protocols

Clinical trial protocols include essential elements such as study background, rationale, methodological procedures, relevance to clinical practice, administrative details, and ethical safeguards[6]. Public availability of these protocols enhances scientific integrity by preventing research falsification, increasing the visibility of randomized controlled trials (RCTs), and improving the overall quality and reliability of their results[4]. Transparency also facilitates peer review, leading to better study design and execution[7]. Conversely, low-quality protocols lacking key elements can compromise the integrity of an RCT, resulting in unreliable data and affecting the validity of its conclusions[4,8]. Therefore, creating and disseminating high-quality protocols is crucial for ensuring the accuracy and reproducibility of clinical research.

The SPIRIT guidelines

In response to the need for standardization and quality improvement in clinical trial protocols, the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) guidelines were published in 2013[6]. These guidelines offer a comprehensive 33-item checklist (Table 1). The adoption of SPIRIT guidelines has proven pivotal in addressing critical aspects of trial design, execution, and reporting, thereby facilitating research reproducibility and outcome reliability[5].

Intervertebral Disc Degeneration (IDD) and clinical trials

Intervertebral disc degeneration (IDD) represents a significant health concern with substantial impact on quality of life and healthcare systems globally[9]. As the prevalence of IDD continues to grow, there has been a corresponding increase in clinical trials investigating various therapeutic interventions[10]. These interventions span a wide range, including biological therapies such as stem cell treatments and growth factors, medical devices like artificial discs and minimally invasive surgical tools, pharmacological interventions including anti-inflammatory drugs and pain modulators, and physical therapies and rehabilitation approaches[11,12]. The diversity and complexity of these interventions highlight the critical importance of maintaining high-quality, transparent clinical trial protocols in IDD research. As noted by 13, the rapidly evolving landscape of IDD treatments necessitates thorough methodological standards to ensure the validity and reliability of clinical findings, thereby facilitating evidence-based decision-making in patient care and policy development.

Knowledge gap and study rationale

Despite the recognized importance of protocol quality and adherence to SPIRIT guidelines in elevating research standards, a significant gap remains in the literature regarding the availability and quality of clinical trial protocols for specific medical conditions, including IDD. Investigations into the accessibility of IDD clinical trial protocols and adherence of these protocols to SPIRIT recommendations have not been comprehensively conducted. This gap limits the ability to critically evaluate and enhance the reporting quality of interventions within this domain[2]. By assessing the availability and quality of IDD clinical trial protocols, we can identify areas of strength and weakness in current practices, paving the way for targeted improvements in trial design and reporting within this critical area of medical research.

Study objectives

The primary objectives of this study are:

1. To evaluate the adherence of available IDD clinical trial protocols to the SPIRIT guidelines, involving a detailed assessment of the protocol structures, contents, and of the presence of essential elements as stipulated by the SPIRIT checklist.

2. To develop and validate machine learning models that predict clinical trial success in intervertebral disc degeneration (IDD) research based on adherence to SPIRIT guideline recommendations, enabling early identification of trials at risk of failure.

3. To identify and rank the most predictive SPIRIT checklist items that influence trial completion outcomes, using feature importance analysis and SHAP (SHapley Additive exPlanations) interpretability methods to determine which specific protocol elements are most critical for trial success.

4. To compare the quality and completeness of protocols between industry-sponsored and non-industry-sponsored IDD clinical trials protocols.

By addressing these objectives, this study aims to provide valuable insights into the current state of IDD clinical trial reporting and identify opportunities for improving the transparency, reproducibility, and overall quality of research in this important field of medicine.

Methods

Study design

In a retrospective predictive modelling study, we utilized machine learning algorithms to identify which SPIRIT checklist items influence clinical trial success within the intervertebral disc degeneration research. Our analysis employed the established ASPIRE methodology[4], adapted for predictive

Table 1. SPIRIT Checklist Key Sections[6].

Section	Description
Administrative information	Title, trial registration, protocol version, funding and roles and responsibilities
Introduction	Scientific background, explanation of rationale and objectives
Methods: Participants, interventions, and outcomes	Study setting, Eligibility criteria, interventions, outcomes, sample size, participant timeline and recruitment
Methods: Assignment of interventions (for controlled trials)	Allocation: sequence generation, allocation concealment mechanism, implementation and blinding
Methods: Data collection, management, and analysis	Data collection methods, data management and statistical methods
Methods: Monitoring	Data monitoring, harms and auditing
Ethics and dissemination	Research ethics approval, protocol amendments, consent or assent, confidentiality, declaration of interests, access to data, ancillary and post-trial care and dissemination policy
Appendices	Informed consent materials and biological specimen

modelling, to assess protocol adherence to SPIRIT recommendations and analyse the relationship between adherence and trial completion outcomes. By combining this protocol evaluation with advanced statistical learning techniques, the study provides domain-specific insights into the most critical elements of protocol quality that affect trial success in this therapeutic area.

Data sources and integration

Data source

This study draws upon a previously published and openly available dataset of intervertebral disc degeneration (IDD) clinical trial protocols curated by Chemorion *et al.*[14]. The dataset includes 18 interventional trials registered on ClinicalTrials.gov (NCT01829997, NCT02338271, NCT02412735, NCT02909689, NCT03092564, NCT03488686, NCT03526356, NCT03617269, NCT03793817, NCT03817606, NCT03955315, NCT04054254, NCT04101534, NCT04115011, NCT04258827, NCT04450214, NCT04594091, NCT04704934), identified between January 2013 and December 2024, and was extracted using a PRISMA-compliant search strategy targeting IDD-specific interventions. The source publication provides a comprehensive analysis of trial design, intervention types, outcomes, and methodological characteristics, making it an appropriate and validated foundation for secondary analyses.

The dataset was compiled through a systematic search conducted in September 2024 and updated until December 2024, using both controlled vocabulary and free-text terms to identify relevant records. Inclusion criteria were restricted to interventional clinical trials focused on IDD, with available protocol information. Trials lacking methodological details or those primarily addressing non-degenerative spinal pathologies were excluded. The final sample includes device-based, pharmacological, biological, and conservative interventions, with details such as trial phase, enrolment, sponsor type, outcome measures, and study design characteristics recorded.

SPIRIT compliance data

SPIRIT adherence was assessed using a systematic 64-item binary checklist based on established methodology from the ASPIRE (Adherence to SPIrit Recommendations) study framework. The 64 items were selected from the complete SPIRIT 2013 checklist, which contains 270 individual components grouped under 33 major headings. The reduction from 270 to 64 components followed a consensus-based approach developed for empirical SPIRIT assessment:

1. PICO-Based Inclusion: All items relating to the formulation of research questions using the PICO structure (Population, Intervention, Comparator, Outcome) were automatically included, as these define the fundamental research question for each trial.

2. Bias Reduction Priority: All items related to sample size calculation, random sequence generation, allocation concealment, and blinding were included due to their critical importance for reducing bias in clinical trials.

3. Core Message Extraction: For remaining SPIRIT headings, two independent investigators reviewed each component and selected those that captured the core message of each heading, avoiding unnecessarily granular details that would complicate empirical analysis.

4. Expert Consensus: All selections were reviewed and consensus achieved through discussion with methodological experts to ensure the final 64-item set maintained the essential elements of SPIRIT compliance assessment.

Scoring System: Each protocol element was evaluated using a standardized binary approach: 0 (not compliant/absent), 1 (fully compliant/present). For specific trial designs, some SPIRIT items are not mandatory SPIRIT, but whenever fulfilled, these items were treated as compliant (score = 1). This binary encoding ensures clear, unambiguous assessment and optimal compatibility with machine learning algorithms. An illustration of the data collection table is shown in Table 2.

Outcome variable definitions

Trial success was defined using a composite measure based on trial completion status, where successful trials were those with "Completed" status, whereas failed trials included studies with "Terminated," "Suspended," or "Withdrawn" status. Additional outcome variables were derived to capture specific failure mechanisms and quality indicators. Termination-related outcomes included safety-related termination for trials stopped due to safety concerns or adverse events, futility-related termination for trials stopped due to lack of efficacy or futility, and recruitment-related termination for trials stopped due to enrollment or recruitment difficulties. Quality indicators encompassed results reporting, defined as whether trial results were publicly reported, and whether serious adverse events emerged during the trial. This comprehensive outcome framework enabled analysis of both overall trial success and specific mechanisms underlying trial failure, providing insights into which SPIRIT protocol elements were most predictive of different types of unsuccessful outcomes.

Feature engineering

The initial 64 SPIRIT checklist items were already in a binary format, with '1' representing adherence and '0' representing non-adherence. For missing data, items deemed 'Not applicable' to a specific trial's design were coded as compliant ('1'), following the established ASPIRE methodology. Given the class imbalance between successful (n=11) and failed (n=7) trials, we implemented class balancing by adjusting class weights within the machine learning algorithms, which gives a higher penalty to misclassifications of the minority 'failed' class. The most critical step was feature selection to reduce the high dimensionality (64 features for 18 samples) and prevent model overfitting. This was achieved by first performing a

Table 2. Illustration of Data Extraction according to the SPIRIT checklist.

Variable	Spirit Item Number	NCT03817606
Title: Basic study design, patient population, and intervention provided in study title	1	1
Trial Registration: Registry name and trial identifier provided	2	1
Protocol version, number and date	3	1
Funding sources of financial and non-financial support declared	4	1
Names of protocol contributors/ authors	5a	1
Name and contact details of sponsor	5b	1
...		
Sample size: rationale for sample size if not derived statistically	14	0
Location of participant recruitment	15	1
Person(s) who will recruit participants	15	0
Expected recruitment rate	15	1
Method for generation of random sequence	16a	1
Allocation concealment mechanism	16b	
Person who will enrol/assign participants	16c	1
...		
Blinding status of care providers	17a	0
Details of specimen collection	33	0

univariate filter, identifying SPIRIT items with the largest difference in adherence rates between the completed and failed trial groups, and then applying clinical relevance filtering to ensure the final feature set was statistically significant for predicting trial completion.

Machine learning methodology

Model selection

Three classification algorithms were implemented to predict trial outcomes, selected for their complementary strengths with binary features and small datasets:

- Logistic Regression with L1 and L2 penalization reported to provide automatic feature selection and interpretable coefficients inherently suited for binary data[15].
- Random Forest Classifier utilized ensemble methods with natural binary feature handling, built-in importance ranking, reported to reduce overfitting through bootstrap aggregation with limited tree depth[16].
- Gradient Boosting (XGBoost) employed sequential ensemble optimization with automatic feature interaction detection and conservative hyperparameters, reported as optimized for small dataset performance with regularization to prevent overfitting[17].

Performance evaluation and feature importance analysis

Performance evaluation utilized Area Under the Receiver Operating Characteristic Curve (ROC-AUC) as the primary metric, supplemented by F1-score for balanced precision-recall assessment and precision and recall for clinical interpretation. The validation framework employed Leave-One-Out Cross-Validation for primary performance estimates, Repeated Stratified K-Fold (3-fold, 10 repeats) for stability assessment, and bootstrap validation (1000 iterations) for confidence interval estimation. Feature importance was calculated using multiple complementary approaches: Gini importance and permutation importance for tree-based models, absolute coefficient values for logistic regression, and SHAP values for model-agnostic explanations of individual predictions.

Clinical Interpretation: The most predictive SPIRIT items were evaluated for clinical plausibility and practical implementation feasibility, ensuring that model insights translated to actionable protocol development guidance. The entire methodology was summarized in Figure 1.

Results

Study characteristics

From the available clinical trial protocols, a wide range of interventions were utilized to address various spinal conditions, particularly degenerative disc disease. A total of 18 clinical trials were retrieved, out of which 11 were completed. Drug-based interventions were featured in 4 studies (24%). Biological interventions, including cell-based therapies and bone grafts, were examined in 5 studies (29%). Device-based interventions, such as artificial discs, spinal cages, and fusion systems, were the most commonly studied, comprising 9 studies (53%). Additionally, 2 studies (12%) investigated alternative interventions, including behavioural and rehabilitative approaches.

Overall adherence

Figure 2 further provides a detailed visualization of adherence patterns across various clinical trial protocols, with dark shaded cells indicating adherence and light shaded cells representing non-adherence. This visualization allows for a quick identification of consistent adherence or non-adherence within and across different studies. There was substantial heterogeneity in adherence rates, certain studies or protocols exhibiting high levels of adherence, while others showed frequent non-adherence. One clinical trial (NCT01829997) showed adherence to all items but 33. In contrast, most of the other studies show more mixed patterns, with clusters of non-adherence.

Degrees of adhesion

While the data does not show 100% adherence across all items, the components such as “Inclusion and exclusion criteria for trial participants described”, “Primary Outcome: specific measurement variable”, “Primary outcomes: time point of measurement”, and “Who has authority to stop the trial” are among the most adhered to, demonstrating high compliance across various types of sponsorships.

Following closely are items such as “Personnel who will collect data” and “Strategies to promote participant retention and complete follow-up”, show partial adherence, especially in industry-sponsored trials.

The lowest adherence rates were observed for advanced governance and transparency components. Three SPIRIT items showed particularly poor compliance at only 11.1% adherence: steering committee general membership and role descriptions (SPIRIT 5d), authorship eligibility criteria (SPIRIT 31b), and plans for granting access to full trial protocols (SPIRIT 31c). Only two trials out of 18 included these governance specifications. Slightly higher but still poor adherence (16.7%) was observed for two additional components: definition of subgroup categories for analysis (SPIRIT 20b) and details of specimen collection procedures (SPIRIT 33). These elements were present in only three trials each.

Industry vs non-industry adherence

The analysis of compliance percentages across different intervention categories furthermore revealed distinct patterns between industry-sponsored and non-industry-sponsored studies (Figure 3). Industry-sponsored studies exhibited a higher average compliance in categories like ‘DEVICE’ (62.50%) and ‘DRUG’ (79.69%), compared to non-industry-sponsored studies, which had compliance rates of 34.38% and 46.35% for these categories, respectively. This discrepancy was particularly pronounced in the ‘DRUG’ category, where the industry compliance rate was significantly higher. Conversely, for

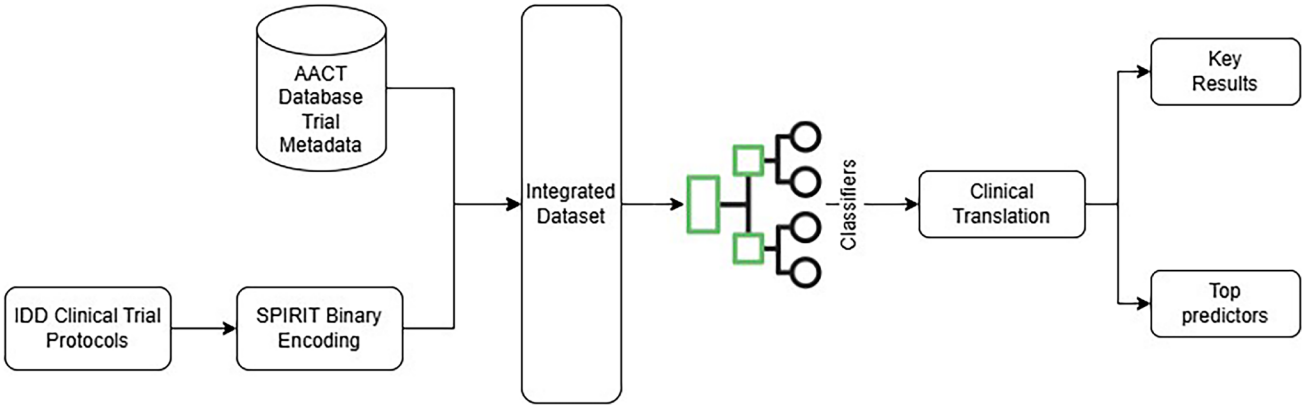


Figure 1. **Methodology.**



Figure 2. **A Heatmap of SPIRIT Adherence.**

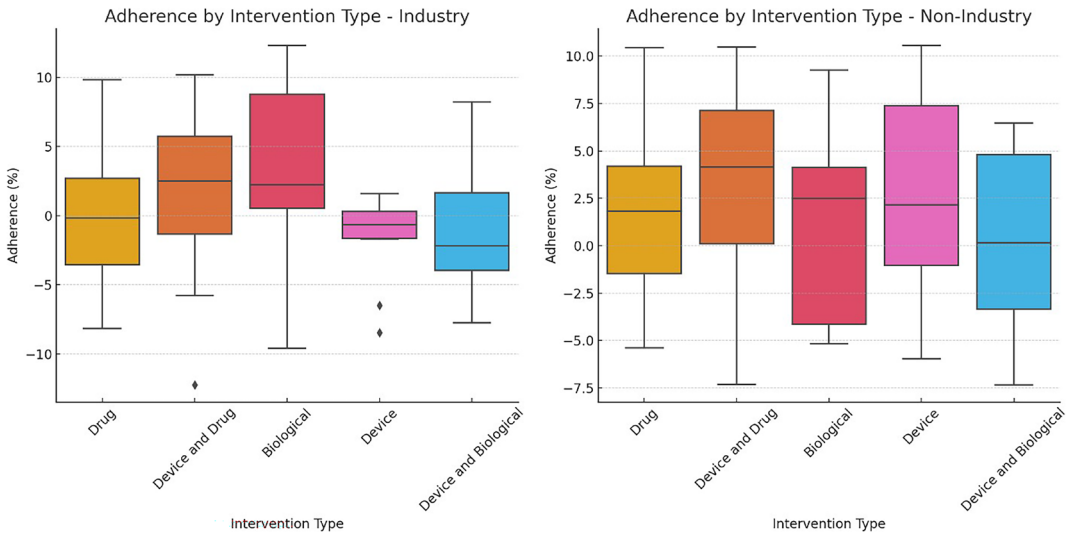


Figure 3. Average Adherence by Intervention Category in Industry vs Non-Industry.

'BIOLOGICAL' interventions, non-industry sponsors showed slightly higher compliance (50.00%) compared to industry sponsors (44.53%).

Adherence by SPIRIT categories

Analyzing the boxplots for both "Industry" and "Non-Industry" funders across various SPIRIT categories provides several key insights into adherence behaviors in clinical trials (Figure 4). First, the variation within each category, as indicated by the spread of the boxplots, suggests differing levels of consistency in adherence practices among trials funded by both industry and non-industry sources. For instance, categories such as "Methods: Data collection, management, and analysis" and "Ethics and dissemination" often showed wider interquartile ranges, indicating a variability on how these SPIRIT items are implemented or reported across different studies.

Moreover, comparing the median adherence levels between industry and non-industry funders, we can observe that some categories might be more consistently adhered to by one group over the other. For example, "Industry" funders might show higher adherence in technical and regulatory aspects such as "Data collection methods" and "Research ethics approval", which are critical to ensure regulatory compliance and drug approval processes. Conversely, "Non-Industry" funders, which often include academic and non-profit research organizations, exhibited higher adherence in areas related to "Participant engagement and outcomes", reflecting perhaps a greater focus on patient-centered research approaches.

Machine learning model performance and predictive SPIRIT items

Model performance

All three models demonstrated comparable and robust performance for the analytics of adherence data and study completion across the different clinical trials: XGBoost emerged as the top performer with a ROC-AUC of 0.839 and Average Precision (AP) of 0.773, followed closely by Random Forest (ROC-AUC: 0.833, AP: 0.757) and Logistic Regression (ROC-AUC: 0.831, AP: 0.751) (Figure 5). The precision-recall analysis revealed that all models maintained strong precision across varying recall levels, with XGBoost showing the most stable performance curve (Figure 5)

Most predictive SPIRIT items

SHAP (SHapley Additive exPlanations) analysis on the best performing model; XGBoost, identified the five most influential SPIRIT checklist items to discriminate trial completion (Figure 6). Top 5 of the most important features include:

a) Item 20c: Definition of analysis population relating to protocol non-adherence (e.g., as randomised analysis), and any statistical methods to handle missing data (e.g., multiple imputation)

b) Item 21a: Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed

c) Item 11b: Criteria for discontinuing or modifying allocated interventions for a given trial participant (e.g., drug dose change in response to harms, participant request, or improving/worsening disease)

d) Item 22: Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct

e) Item 11d: Relevant concomitant care and interventions that are permitted or prohibited during the trial

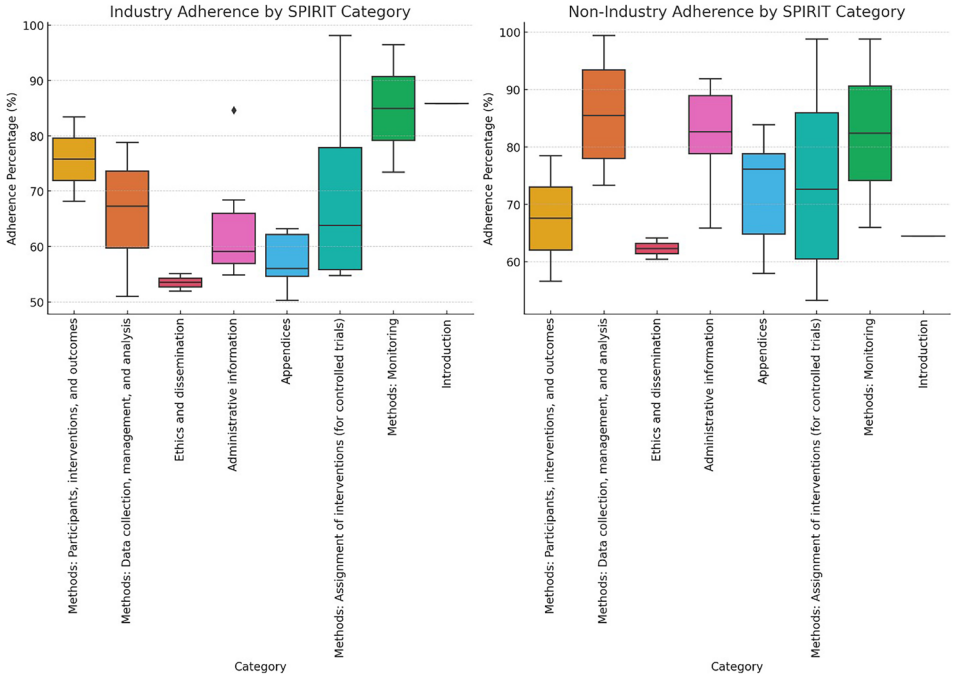


Figure 4. Adherence by SPIRIT Item Category.

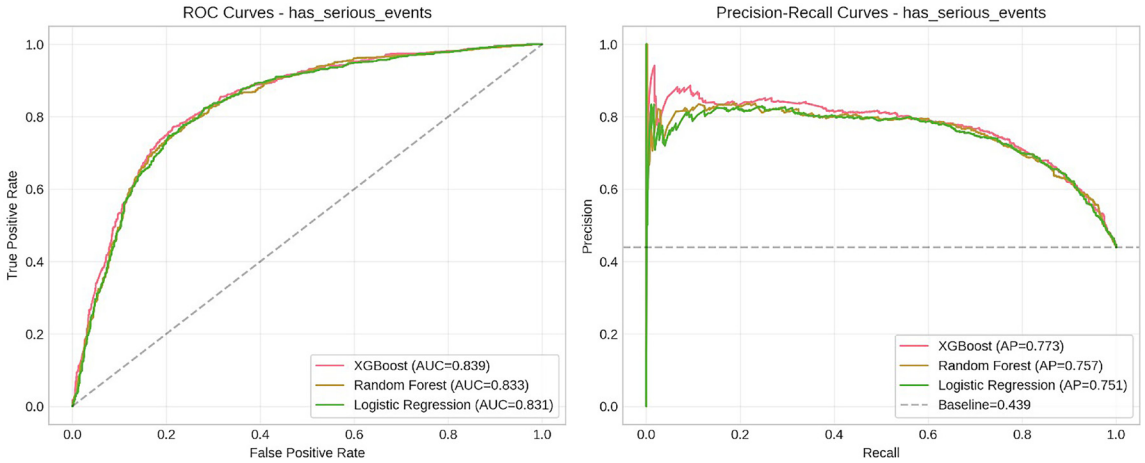


Figure 5. Model performance.

Discussion

Principal findings

This study represents the first application of machine learning methodology to identify SPIRIT guideline items associated with clinical trial completion in intervertebral disc degeneration research. Despite the inherent limitations of our small sample size (n=18), our exploratory analysis demonstrated that quantifiable differences in protocol quality contain meaningful predictive signals for trial outcomes, with XGBoost achieving strong cross-validated performance (ROC-AUC: 0.839, AP: 0.773). This proof-of-concept analysis provides preliminary evidence that computational approaches can augment traditional protocol assessment methods, extending beyond established frameworks that have primarily focused on descriptive adherence evaluation[18].

While previous studies have applied predictive modeling to clinical trial recruitment (Weng *et al.*, 2019) and patient outcomes[19], our work extends this paradigm to protocol quality assessment—a critical but understudied determinant of trial success. This approach aligns with emerging trends in clinical trial optimization using artificial intelligence, though most applications have focused on patient selection and outcome prediction rather than protocol design[20,21].

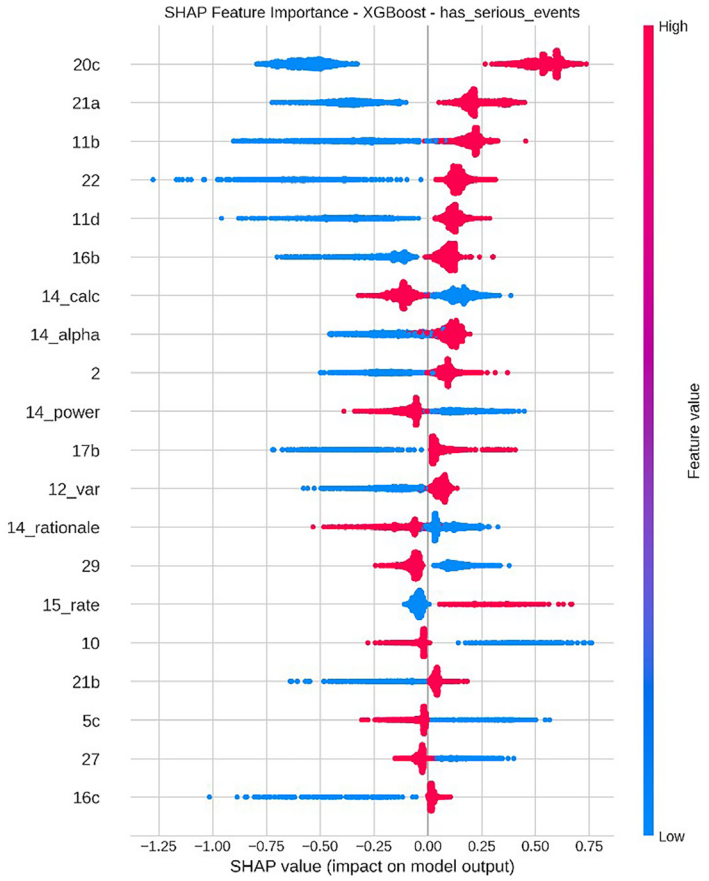


Figure 6. SHAP Analysis.

SPIRIT adherence patterns

Our systematic assessment revealed substantial heterogeneity in SPIRIT adherence across IDD trials, with adherence rates ranging from 11.1% to 100% across different checklist items. This variability markedly exceeds that reported in broader assessments of clinical trial protocols. 22 found adherence rates typically ranging from 30–90% in their comprehensive review of randomized controlled trials, while[23] reported more consistent adherence patterns in oncology trials, with most items achieving 40–80% compliance.

In contrast to these findings, our IDD-specific analysis revealed more extreme variability, with particularly poor adherence to governance and transparency items. Only 11.1% of IDD trials included steering committee descriptions (SPIRIT 5d), authorship criteria (SPIRIT 31b), and protocol access plans (SPIRIT 31c), compared to 25–35% reported in the broader ASPIRE framework across multiple therapeutic areas[18]. This suggests that IDD research may lag behind other fields in implementing comprehensive governance structures.

However, our findings showed excellent adherence to fundamental design elements, with 100% compliance for basic study design components and primary outcome specifications. This contrasts with 24, who found that even basic protocol elements showed only 70–85% adherence across therapeutic areas, suggesting that IDD researchers excel in core methodological planning while struggling with advanced governance requirements.

Industry vs. academic sponsorship: divergent quality patterns

Our analysis revealed distinct patterns between industry-sponsored and non-industry-sponsored trials that both align with and diverge from established literature. Industry-sponsored trials demonstrated higher adherence in regulatory and technical domains, particularly for device (62.50% vs. 34.38%) and drug interventions (79.69% vs. 46.35%). These findings strongly support previous observations by Als-25 and 26, who documented superior methodological quality in industry trials, attributing this to stringent regulatory requirements and standardized operational procedures.

However, our results challenge some assumptions about academic research quality. For biological interventions, non-industry sponsors showed slightly higher compliance (50.00% vs.

44.53%), contrasting with 27, who generally found lower methodological quality in academic trials. This divergence may reflect the specialized nature of IDD biological interventions, where academic investigators often possess deeper scientific expertise in regenerative medicine approaches than industry sponsors.

28 reported that industry trials typically achieve 85–95% adherence to regulatory requirements, while our IDD industry trials showed more modest performance (62–80% across categories). This suggests that IDD research, even within industry settings, faces unique challenges that may not be captured by general clinical trial quality assessments.

Predictive SPIRIT items

The identification of statistical methodology items as top predictors provides strong empirical support for fundamental principles established in clinical research methodology literature. Our finding that analysis population definition and missing data handling (Item 20c) emerged as the strongest predictor directly validates extensive evidence from 29 and 30, who demonstrated that inadequate statistical planning is a primary cause of trial failure. However, while these authors focused on post-hoc analyses of failed trials, our predictive approach demonstrates that these deficiencies can be identified prospectively through protocol assessment.

31 found that trials with comprehensive statistical analysis plans had 23% higher completion rates and significantly more reliable results. Our machine learning analysis not only confirms this relationship but quantifies the relative importance of specific statistical planning elements, with Item 20c showing the highest SHAP value among all predictive features.

The critical importance of data monitoring committee specifications (Item 21a) as our second-strongest predictor aligns with established regulatory guidance but provides new quantitative evidence of its predictive value. 32 and 33 emphasized the necessity of independent safety monitoring, particularly for trials involving surgical procedures or investigational biologics. However, their recommendations were based on qualitative assessments and regulatory requirements rather than predictive modeling. Our analysis provides the first quantitative evidence that the quality of Data Monitoring Committee specifications directly correlates with trial completion probability.

Interestingly, our findings partially contradict some established assumptions about protocol complexity. While conventional wisdom suggests that overly detailed protocols may indicate anticipated difficulties[34], our analysis showed mixed directional effects for intervention modification criteria (Item 11b). This nuanced finding suggests that moderate complexity in discontinuation protocols may actually predict success, challenging the linear relationship between protocol detail and trial difficulty assumed in previous literature.

Clinical interpretation and evidence-based protocol optimization

Our identification of three critical domains: statistical reliability, safety oversight, and operational planning—provides empirical validation for theoretical frameworks proposed in clinical trial methodology literature. 35 estimated that 85% of biomedical research represents waste, with inadequate design and poor conduct representing major contributors. Our findings suggest that focusing protocol development resources on the specific SPIRIT items we identified could substantially reduce this waste in IDD research.

The predictive value of operational planning items (concomitant care protocols, adverse event management) supports observations by 36 and 37 that inadequate operational planning leads to protocol deviations and premature termination. However, while these studies identified operational challenges through qualitative interviews and retrospective analyses, our machine learning approach provides quantitative rankings of which operational elements are most critical for success.

Notably, our findings diverge from some conventional prioritization frameworks. While regulatory guidance typically emphasizes safety reporting and data management equally[38,39], our predictive analysis suggests that specific safety oversight elements (DMC composition, adverse event collection plans) are substantially more predictive of success than general data management procedures. This suggests that current guidance may not adequately differentiate between protocol elements of varying predictive importance.

The practical implications of our findings contrast with existing protocol development practices. Most institutional templates and regulatory guidance provide equal emphasis to all SPIRIT checklist items[40]. However, our analysis suggests that protocol developers should prioritize the five most predictive items we identified, potentially allowing more efficient resource allocation during protocol development while maintaining or improving success rates.

Ethics and consent

Ethical approval and consent were not required for this study, as all analyses were conducted on previously published, publicly available protocol documents and did not involve any direct experimentation on human participants or access to confidential data.

Data availability

The datasets generated and analysed during the current study are openly available in the Zenodo repository. The complete dataset, titled “, SPIRIT Evaluations of IDD clinical Trial Protocols” can be accessed at https://doi.org/10.5281/zenodo.16605984[41].

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

References

1. Butcher NJ, Mew EJ, Monsour A, *et al.*: **Outcome reporting recommendations for clinical trial protocols and reports: a scoping review.** *Trials.* 2020; **21**(1): 620.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
2. Chan AW, Hróbjartsson H: **Promoting public access to clinical trial protocols: challenges and recommendations.** *Trials.* 2018; **19**(1): 116.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
3. Whittington RA: **The pros of publishing standalone clinical trial protocols in anesthesiology journals.** *Anesth Analg.* 2018; **126**(5): 1441–1443.
PubMed Abstract | **Publisher Full Text**
4. Gryaznov D, Odutayo A, von Niederhäusern B, *et al.*: **Rationale and design of repeated cross-sectional studies to evaluate the reporting quality of trial protocols: the Adherence to SPIrit REcommendations (ASPIRE) study and associated projects.** *Trials.* 2020; **21**(1): 896.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
5. Cuschieri S: **The CONSORT statement.** *Saudi J Anaesth.* 2019; **13**(Suppl 1): S27–S30.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
6. Chan AW, Tetzlaff JM, Altman DG, *et al.*: **SPIRIT 2013 statement: defining standard protocol items for clinical trials.** *Ann Intern Med.* 2013; **158**(3): 200–7.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
7. Taichman DB, Sahni P, Pinborg A, *et al.*: **Data sharing statements for clinical trials.** *BMJ.* 2017; **357**: j2372.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
8. Tetzlaff JM, Chan AW, Kitchen J, *et al.*: **Guidelines for Randomized Clinical Trial protocol content: a systematic review.** *Syst Rev.* 2012; **1**: 43.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
9. Hartvigsen J, Hancock MJ, Kongsted A, *et al.*: **What low back pain is and why we need to pay attention.** *Lancet.* 2018; **391**(10137): 2356–2367.
PubMed Abstract | **Publisher Full Text**
10. Maher C, Underwood M, Buchbinder R: **Non-specific low back pain.** *Lancet.* 2017; **389**(10070): 736–747.
PubMed Abstract | **Publisher Full Text**
11. Guterl CC, See EY, Blanquer SBG, *et al.*: **Challenges and strategies in the repair of ruptured annulus fibrosus.** *Eur Cell Mater.* 2013; **25**: 1–21.
PubMed Abstract | **Free Full Text**
12. Samartzis D, Karppinen J, Cheung JPY, *et al.*: **Disk degeneration and low back pain: are they fat-related conditions?** *Global Spine J.* 2013; **3**(3): 133–143.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
13. Yim RLH, Lee JTY, Bow CH, *et al.*: **A systematic review of the safety and efficacy of mesenchymal stem cells for disc degeneration: insights and future directions for regenerative therapeutics.** *Stem Cells Dev.* 2014; **23**(21): 2553–2567.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
14. Chemorion F, Bisotti MA: **Assessing the adherence of intervertebral disc degeneration clinical trial protocols to SPIRIT guidelines.** 2024.
Publisher Full Text
15. Greenwood CJ, Youssef GJ, Letcher P, *et al.*: **A comparison of penalised regression methods for informing the selection of predictive markers.** *PLoS One.* 2020; **15**(11): e0242730.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
16. Salman HA, Kalakech A, Steiti A: **Random forest algorithm overview.** *Babylonian Journal of Machine Learning.* 2024; **2024**: 69–79.
Publisher Full Text
17. Langat KK, Waititu AG, Ngare PO: **Modified XGBoost hyper-parameter tuning using adaptive particle swarm optimization for credit score classification.** *Machine Learning Research.* 2024; **9**(2): 64–74.
Publisher Full Text
18. Gryaznov D, Odutayo A, von Niederhäusern B, *et al.*: **Rationale and design of repeated cross-sectional studies to evaluate the reporting quality of trial protocols: the Adherence to SPIrit REcommendations (ASPIRE) study and associated projects.** *Trials.* 2020; **21**(1): 896.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
19. Goldstein BA, Navar AM, Carter RE: **Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges.** *Eur Heart J.* 2017; **38**(23): 1805–1814.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
20. Harrer S, Shah P, Antony B, *et al.*: **Artificial Intelligence for clinical trial design.** *Trends Pharmacol Sci.* 2019; **40**(8): 577–591.
PubMed Abstract | **Publisher Full Text**
21. Wouters OJ, McKee M, Luyten J: **Estimated research and development investment needed to bring a new medicine to market, 2009–2018.** *JAMA.* 2020; **323**(9): 844–853.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
22. Turner L, Shamseer L, Altman DG, *et al.*: **Consolidated Standards of Reporting Trials (CONSORT) and the completeness of reporting of Randomised Controlled Trials (RCTs) published in medical journals.** *Cochrane Database Syst Rev.* 2012; **11**(11): MR000030.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
23. Speich B, von Niederhäusern B, Schur N, *et al.*: **Systematic review on costs and resource use of randomized clinical trials shows a lack of transparent and comprehensive data.** *J Clin Epidemiol.* 2018; **96**: 1–11.
PubMed Abstract | **Publisher Full Text**
24. Butcher NJ, Mew EJ, Monsour A, *et al.*: **Outcome reporting recommendations for clinical trial protocols and reports: a scoping review.** *Trials.* 2020; **21**(1): 620.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
25. Als-Nielsen B, Chen W, Gluud C, *et al.*: **Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events?** *JAMA.* 2003; **290**(7): 921–8.
PubMed Abstract | **Publisher Full Text**
26. Lundh A, Lexchin J, Mintzes B, *et al.*: **Industry sponsorship and research outcome.** *Cochrane Database Syst Rev.* 2017; **2**(2): MR000033.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
27. Bekelman JE, Li Y, Gross CP: **Scope and impact of financial conflicts of interest in biomedical research: a systematic review.** *JAMA.* 2003; **289**(4): 454–65.
PubMed Abstract | **Publisher Full Text**
28. Califf RM, Zarin DA, Kramer JM, *et al.*: **Characteristics of clinical trials registered in ClinicalTrials.gov, 2007–2010.** *JAMA.* 2012; **307**(17): 1838–47.
PubMed Abstract | **Publisher Full Text**
29. Ioannidis JPA: **Why most published research findings are false.** *PLoS Med.* 2005; **2**(8): e124.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
30. Button KS, Ioannidis JPA, Mokrysz C, *et al.*: **Power failure: why small sample size undermines the reliability of neuroscience.** *Nat Rev Neurosci.* 2013; **14**(5): 365–76.
PubMed Abstract | **Publisher Full Text**
31. Gamble C, Krishan A, Stocken D, *et al.*: **Guidelines for the content of statistical analysis plans in clinical trials.** *JAMA.* 2017; **318**(23): 2337–2343.
PubMed Abstract | **Publisher Full Text**
32. DeMets DL, Ellenberg SS: **Data monitoring committees — expect the unexpected.** *N Engl J Med.* 2016; **375**(14): 1365–1371.
PubMed Abstract | **Publisher Full Text**
33. Fleming TR, DeMets DL, Roe MT, *et al.*: **Data monitoring committees: promoting best practices to address emerging challenges.** *Clin Trials.* 2017; **14**(2): 115–123.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
34. Treweek S, Bevan S, Bower P, *et al.*: **Trial forge guidance 1: what is a Study Within A Trial (SWAT)?** *Trials.* 2018; **19**(1): 139.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
35. Chalmers I, Glasziou P: **Avoidable waste in the production and reporting of research evidence.** *Obstet Gynecol.* 2009; **114**(6): 1341–1345.
PubMed Abstract | **Publisher Full Text**
36. Sullivan KJ, Wayne C, Patey AM, *et al.*: **Barriers and facilitators to the implementation of evidence-based practice by pediatric surgeons.** *J Pediatr Surg.* 2017; **52**(10): 1666–1673.
PubMed Abstract | **Publisher Full Text**
37. Walters SJ, Dos Anjos Henriques-Cadby IB, Bortolami O, *et al.*: **Recruitment and retention of participants in randomised controlled trials: a review of trials funded and published by the United Kingdom health technology assessment programme.** *BMJ Open.* 2017; **7**(3): e015276.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
38. **Data monitoring committees - Scientific guideline | European Medicines Agency (EMA)**. Accessed: July 18, 2025.
Reference Source
39. Center for Biologics Evaluation and Research: **Establishment and operation of clinical trial data monitoring committees**. Accessed: July 18, 2025.
Reference Source
40. Chan AW, Tetzlaff JM, Gøtzsche PC, *et al.*: **SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials.** *BMJ.* 2013; **346**: e7586.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
41. Chemorion F: **SPIRIT evaluations of IDD clinical trial protocols.** Zenodo. July 30, 2025; Accessed: July 30, 2025.
http://www.doi.org/10.5281/zenodo.16605984