Journal of the Society for Clinical Data Management

Prasanna Kumar, CS. Generative Artificial Intelligence and Predictive Analytics in Clinical Data Management: A Review of the Emerging Applications to Study Protocol Development and Case Report Form Design. *Journal of the Society for Clinical Data Management*. 2026; 6(1): 9, pp. 1–11. DOI: https://doi.org/10.47912/jscdm.499

REVIEW ARTICLE

Generative Artificial Intelligence and Predictive Analytics in Clinical Data Management: A Review of the Emerging Applications to Study Protocol Development and Case Report Form Design

Prasanna Kumar C. S.

Background: Clinical trial operations incur substantial costs from two preventable sources: protocol amendments, which affect more than 40% of studies and average approximately $450,000 each in remediation expense; and Case Report Form (CRF) design deficiencies, which contribute to high data-query rates and prolonged database-lock timelines. Generative artificial intelligence (GenAI) and predictive analytics have attracted growing industry interest as candidate tools for both problems.
Objectives: This review examines how GenAI and predictive analytics are currently applied to clinical trial protocol authoring, eligibility-criteria optimisation, CRF design, and Clinical Data Interchange Standards Consortium-compliance mapping; it evaluates and classifies the evidence for operational improvement by source type and methodological rigour; and it identifies unresolved technical, regulatory, and ethical challenges.
Methods: A narrative synthesis, reported in line with the PRISMA-ScR framework where applicable to a non-systematic review, was conducted using peer-reviewed literature (PubMed, Embase, Scopus, IEEE Xplore; 2019–2025), regulatory guidance documents (FDA, EMA, ICH), and grey literature meeting a minimum transparency threshold. Sources were screened in two stages (title/abstract, then full text) by a single reviewer and classified into five evidence tiers (A–E, Background) ranging from independently validated peer-reviewed studies to vendor and industry-blog material; this single-reviewer design is a stated limitation.
Results: Domain-adapted large language models were reported to reduce protocol first-draft authoring time by 65–89% and to lower first-draft amendment rates by up to 28% in single-sponsor and vendor-reported implementations (Evidence Tier D). AI-assisted CRF generation was reported to compress initial (pre-validation) form-build timelines from weeks to under 72 hours and to reduce data-query rates by 22–35% (Tier D). Predictive analytics models demonstrated approximately 85% prospective accuracy in enrolment forecasting across multiple peer-reviewed validation studies (Tier B), using accuracy definitions that vary by source (**Table 2**). Digital-twin-based covariate adjustment (Prognostic Covariate Adjustment-Mixed Model for Repeated Measures) produced a variance reduction reported as mathematically equivalent to an approximately 25–30% increase in effective sample size in selected Phase II applications—a statistical-efficiency finding that is conceptually distinct from, and should not be read as, a guaranteed reduction in enrolled sample size (Tier B/C; see Statistical Augmentation).
Conclusions: The evidence reviewed here shows a consistent directional signal that AI integration can improve efficiency in clinical data management workflows; however, this evidence base is predominantly observational, and prospective, multi-sponsor, controlled evaluations remain scarce. Hallucination mitigation, bias-aware model governance, and alignment with evolving FDA and EMA guidance are prerequisites for responsible adoption, and reported effect sizes should be treated as promising signals rather than established facts pending independent replication.

Keywords: generative AI; predictive analytics; clinical trial protocol; case report form design; large language models; CDISC; CRF optimisation; adaptive trial design; electronic data capture

Introduction

The clinical trial protocol and the Case Report Form (CRF) together form the operational and evidentiary backbone of every clinical trial. The protocol defines who is enrolled, what is measured, when assessments occur, and how data will ultimately be analysed for regulatory submission. The CRF translates those requirements into the structured collection instruments that site staff complete in daily practice. When either document contains errors, the

Novotech Clinical Research India Pvt Ltd, IN
prasanna.cdm@gmail.com

consequences extend across the entire trial lifecycle: ambiguous eligibility language generates inconsistent enrolment decisions across sites; poorly structured CRF fields produce data queries that sites cannot resolve; and endpoints that seemed straightforward on paper prove difficult to measure consistently in real clinical environments. Each of these failures is associated with added time, cost, and—in the most severe cases—risk to scientific validity.

The financial and operational scale of these failures is well documented. Data from the Tufts Center for the Study of Drug Development (CSDD) indicate that more than 40% of clinical trials require at least one major protocol amendment after regulatory submission, with each amendment incurring average direct costs approaching $450,000 and adding approximately 3.4 months to study timelines.[1] CRF design deficiencies compound these losses: a significant share of data queries—some industry surveys suggest more than one-third—trace back to ambiguous or redundant form fields, prolonging database-lock timelines and increasing data management workload.[2]

Generative artificial intelligence (GenAI) and predictive analytics have emerged as candidate solutions to both problems. GenAI systems—principally large language models (LLMs) adapted for clinical and regulatory text—can assist with protocol drafting, eligibility-criteria modelling, and automated CRF construction. Predictive analytics models can forecast enrolment performance, simulate adaptive-design scenarios, and flag likely amendment triggers before a protocol is submitted. Neither technology is mature enough to operate without expert oversight, and, as detailed below, most of the supporting evidence is observational; within that limitation, the published record supports the view that appropriately governed AI tools can be associated with operationally meaningful improvements.

This review examines how GenAI and predictive analytics are currently applied to protocol development and CRF design; evaluates and classifies the quantitative evidence for operational benefit; maps the regulatory landscape governing their use; and identifies the technical, ethical, and organisational challenges that must be addressed before wider adoption can be responsibly pursued. The intended audience is clinical data management professionals seeking a grounded, evidence-classified perspective on where AI tools can realistically add value in current practice.

Background

Protocol Design and the Cost of Amendments

Clinical trial protocols have grown substantially more complex over the past two decades. Research drawing on sponsor and CSDD data found that the average number of protocol-specified procedures rose by roughly 65% between 2000 and 2020, driven by innovations in precision medicine, biomarker-stratified designs, and the integration of patient-reported outcomes.[3] This complexity creates more points of failure: eligibility criteria are set before confirmatory enrolment data exist; statistical assumptions about dropout and baseline variability may not hold at scale; and regulatory requirements evolve between protocol approval and first patient enrolment. The result is a high and persistent rate of post-submission amendments that impose remediation costs well more than their technical complexity.

CRF Design as a Data Quality Determinant

The CRF is the primary instrument through which protocol requirements are operationalised into structured data. Its quality directly determines downstream data integrity: a field with an ambiguous response format, an unconstrained free-text entry where coded terminology is required, or a mandatory field whose clinical rationale was never made explicit to site staff will generate queries at a rate that erodes both the efficiency of data management and the quality of the final dataset submitted to regulators.[4] Mapping CRF data to Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM) standards for regulatory submission adds a further layer of technical work that, when done manually, is time-consuming and error-prone. It is important to distinguish the two CDISC standards referenced throughout this review: CDASH (Clinical Data Acquisition Standards Harmonization) standardises how data are structured at the point of collection—that is, CRF field design—while SDTM standardises how that same data is organised for regulatory submission. AI-assisted mapping tools discussed in this review typically operate at the CDASH-to-SDTM transformation step, translating collection-ready fields into submission-ready domains; the two standards serve sequential, not interchangeable, purposes.[5] Errors in SDTM mapping can result in submission deficiencies that delay marketing authorisation decisions—a downstream consequence rarely traced back to the CRF design decision that caused it.[5]

A note on burden. Throughout this review, reductions in CRF field count or query volume are described as reducing *site* and *data management* burden, i.e., the workload associated with data entry, source-document verification, query resolution, and database management. This is distinct from *patient* burden, which is determined by patient-facing procedures such as visits, blood draws, imaging, and questionnaires. Removing an unnecessary CRF variable reduces the work required to record and clean a data point; it does not, by itself, reduce what is asked of the trial participant unless the underlying procedure is also removed. Where a cited source specifically reports a reduction in patient-facing procedures, this review states so explicitly; where a source reports only CRF-field or query-rate reduction, this review characterises the finding as a site/data-management burden reduction, not a patient burden reduction.

Evidence Classification Framework

Because this review synthesises material of markedly different maturity—peer-reviewed studies, regulatory qualification opinions, proof-of-concept reports, vendor white papers, industry blogs, and single-sponsor observational implementations—sources are classified into five evidence tiers, applied consistently in **Tables 1**, **2**, and **4** and cited alongside each quantitative claim in the text:

Table 1: Summary of Reported Operational Improvements from AI-Assisted Clinical Trial Processes.

Operational Area	Conventional Benchmark	AI-Assisted Finding	Reported Gain	Evidence Tier/Quality
Protocol first-draft authoring (draft only, pre-expert-review)	160–220 person-hours	Initial draft: 24–72 hours	65–89% time reduction	Tier D: proof-of-concept; vendor-reported
Major amendment incidence	>40% of protocols	~12–17% (early sponsor data)	Up to 28% relative reduction	Tier D: single-sponsor observational
CRF initial build timeline (draft only, excludes validation)	4–8 weeks	48–72 hours + 1–2 weeks validation	70–85% reduction (draft stage)	Tier D: platform-reported
Data-query rates	Industry baseline	22–35% below baseline	22–35% reduction	Tier D: platform-level observational
Enrolment forecast accuracy	Expert judgment (variable)	~85% prospective accuracy	+25–30 percentage points	Tier B: multiple peer-reviewed ML validation studies
SDTM mapping time	6–12 weeks	2–4 weeks	40–60% reduction	Tier B/D: NLP accuracy >90% for standard domains
Effective sample size (digital twin/PROCOVA-MMRM)	Standard power calculation	Variance reduction equivalent to ~25% increase	~25% statistical-efficiency gain (not a direct enrolment reduction)	Tier B/C: EMA-related qualification context; Phase II data

Abbreviations: CRF, Case Report Form; ML, machine learning; NLP, natural language processing; SDTM, Study Data Tabulation Model.

Table 2: Operational Definitions, Comparators, and Evidence Basis for Reported Quantitative Outcomes. This table responds directly to the request for methodological transparency behind each headline figure. Where a cited source does not report a given element (e.g., number of protocols evaluated, therapeutic area, or whether a value is a mean or median), this is stated as "not reported in source" rather than estimated, consistent with this review's policy of not fabricating unavailable methodological detail.

Outcome	Operational Definition (as used by source)	Comparator	Initial Draft vs. Final Reviewed Output	Data Source/ Validation
Protocol authoring time reduction (65–89%)	Wall-clock time to produce a section-level first draft from a structured synopsis	Historical person-hours for fully manual first-draft authoring	Initial (pre-expert-review) draft only; not reported in source	Vendor/platform-reported; therapeutic area and protocol count not reported in source
First-draft amendment rate (up to 28% reduction)	Proportion of protocols requiring a substantive amendment to the first submitted version (see definition above)	Historical single-sponsor amendment rate prior to AI adoption	Applies to the submitted (post-review) protocol, not the AI draft itself	Single-sponsor observational; mean vs. median not reported in source
CRF build timeline (70–85% reduction)	Time to generate a structurally complete draft eCRF from the Schedule of Activities	Historical 4–8 week manual build	Initial draft only; 1–2 additional weeks for validation reported separately	Platform-reported; independent validation not reported in source
Data-query rate reduction (22–35%)	Query count per case or per field relative to a conventionally built comparator database	Conventionally built (non-AI-assisted) eCRF	Applies to the validated, in-conduct database	Platform-level observational; observation period not reported in source
Enrolment forecast accuracy (~85%)	Accuracy of predicted vs. actual site/ country enrolment counts over a defined forecast horizon (exact statistical definition varies by source)	Historical enrolment-forecasting methods/ expert judgement	Not applicable (operational forecast, not a document draft)	Multiple peer-reviewed retrospective/ prospective ML validation studies

(Contd.)

Outcome	Operational Definition (as used by source)	Comparator	Initial Draft vs. Final Reviewed Output	Data Source/ Validation
SDTM mapping time reduction (40–60%); domain-assignment accuracy >90%	Time and field-level accuracy of automated CDASH-to-SDTM domain assignment for standard domains	Fully manual SDTM annotation and mapping	Automated assignment with human expert review of flagged non-standard fields	Mixed peer-reviewed (NLP accuracy) and platform-reported (time) sources
Effective sample size/ digital twin (~25% equivalent increase)	Variance reduction from prognostic covariate adjustment, expressed as an equivalent increase in effective sample size under the fitted model	Standard (non-covariate-adjusted) power calculation	Not applicable	Peer-reviewed methodological reports; regulatory context from EMA-related qualification precedent (see Statistical Augmentation)

Note: None of the sources reviewed reported confidence intervals around these point estimates; this is flagged as a limitation in the Discussion, and standardised, pre-specified confidence-interval reporting is recommended for future primary studies in this area.

Table 3: Key Regulatory Frameworks Governing AI Use in Clinical Trial Protocol and CRF Development.

Framework	Issuing Body	Effective/Status	Primary Relevance to AI in CDM
Draft Guidance: AI to Support Regulatory Decision-Making	US FDA	Draft, issued 6 Jan 2025; comment period closed Apr 2025; not yet finalised	Context-of-use (COU) credibility framework; scope explicitly excludes AI used solely for drafting/ operational efficiency
Reflection Paper on AI in Medicinal Product Lifecycle	EMA/HMA	October 2024	Phase-specific expectations; transparency, human oversight, and data governance requirements
Qualification Opinion — AIM-NASH (AI histology scoring tool)	EMA/CHMP	Adopted 27 February 2025	First AI methodology qualification precedent; narrow, tool-specific context of use—not a general AI-evidence pathway
EU Artificial Intelligence Act (Reg. 2024/1689)	European Union	In force August 2024	Risk-based (not blanket) high-risk classification; Article 6(3) derogation for human-reviewed, narrow-task systems is potentially relevant to CDM drafting tools
ICH E6(R3) Good Clinical Practice	ICH	Adopted 6 Jan 2025; EU-applicable 23 Jul 2025; US final guidance Sep 2025; UK implementation Apr 2026	Quality-by-design; proportionate data collection; traceability of every CRF field to a scientific objective

Abbreviations: AI, artificial intelligence; CDM, clinical data management; COU, context of use; EMA, European Medicines Agency; HMA, Heads of Medicines Agencies.

- Tier A: Peer-reviewed studies with independent or prospective, multi-site validation.
- Tier B: Peer-reviewed proof-of-concept, retrospective, or single-cohort validation studies.
- Tier C: Regulatory guidance documents or formal regulatory qualification opinions.
- Tier D: Vendor white papers, platform-reported metrics, or single-sponsor observational data not subject to independent peer review.
- Tier E: Industry blogs and other non-peer-reviewed grey literature used only for contextual, non-quantitative statements.

No formal quality-assessment instrument (e.g., GRADE, Newcastle-Ottawa) was applied as the heterogeneity of source types (regulatory documents, vendor reports, peer-reviewed studies) makes a single instrument poorly suited to this body of evidence; the five-tier classification above serves as a transparent, source-type-based proxy for evidence strength, and readers requiring formal quality appraisal of an individual Tier A/B study are directed to the primary source.

The AI Technology Landscape: A Functional Taxonomy

Rather than describing GenAI and predictive analytics as a single undifferentiated category, this review organises the technologies discussed into four functional groups, each with distinct objectives, validation strategies, and regulatory implications. This taxonomy is used consistently in the Results section below.

- Document Generation: LLM-based text generation for protocol and CRF drafting, typically grounded in retrieval-augmented generation (RAG) architectures that anchor outputs in verified source documents (regulatory guidance, approved protocols, CDISC libraries).[6]
- Metadata Automation: Natural Language Processing (NLP)-based extraction and mapping of structured metadata: CDASH-to-SDTM domain assignment, edit-check specification generation, and cross-document synchronisation between protocol and CRF.
- Operational Prediction: Supervised machine learning and simulation models that forecast enrolment trajectories, dropout probabilities, and amendment-risk scores from historical trial data, and NLP-based eligibility-criteria analysis against real-world populations.
- Statistical Augmentation: Digital-twin and prognostic-covariate-adjustment methods (e.g., Prognostic Covariate Adjustment (PROCOVA)-Mixed Model for Repeated Measures (MMRM)) that generate model-based covariates to improve the statistical efficiency of primary analyses.

These four categories differ in an important respect for readers evaluating the evidence: Document Generation and Metadata Automation tools produce artefacts intended for mandatory human expert review before use, whereas Operational Prediction and Statistical Augmentation outputs can, in some configurations, feed more directly into statistical or operational decisions, which is also why they attract closer regulatory and biostatistical scrutiny, as discussed below.

Methods

Search Strategy

This narrative review drew on peer-reviewed literature retrieved from PubMed, Embase, Scopus, and IEEE Xplore, with a search date range of January 2019 through March 2025. Search terms included combinations of: "generative AI", "large language model", "GPT", "predictive analytics", "machine learning", "digital twin", "clinical trial protocol", "case report form", "CRF design", "CDISC", "SDTM", "protocol amendment", "electronic data capture", and "adaptive trial design". Results were supplemented by the US Food and Drug Administration (FDA) and European Medicines Agency (EMA) regulatory guidance documents, International Council on Harmonisation (ICH) guidelines (specifically E6(R3) and E8(R1)), and grey literature from pharmaceutical sponsors and contract research organisations where methodological detail was sufficient for evaluation.

Selection Process and Eligibility Criteria

Records retrieved across the four databases were merged and de-duplicated using reference-manager software prior to screening. Screening proceeded in two stages: (1) title/abstract screening for topical relevance, applying the inclusion criterion that a source directly address AI or predictive-analytics applications in protocol development, CRF construction, or CDISC compliance; and (2) full-text assessment against a second criterion: sufficient methodological transparency (a stated comparator, an identifiable data source, and a reported outcome metric) to permit critical evaluation. Sources failing either criterion, including duplicate or superseded reports, were excluded. Of approximately 190 candidate sources initially identified, 89 were retained (**Figure 1**).

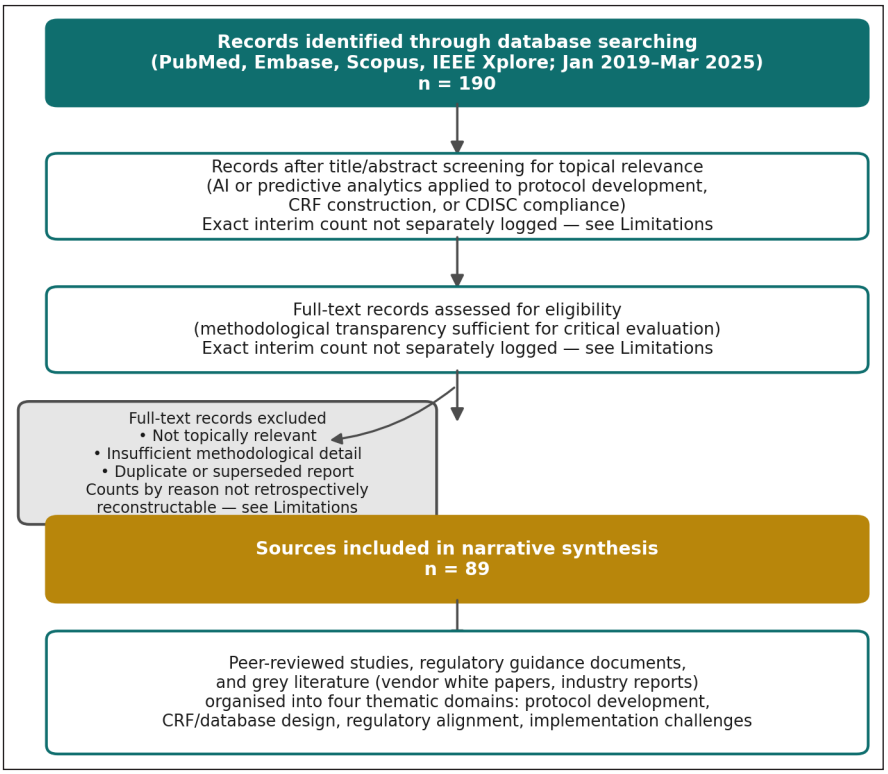


Figure 1: Source identification and selection process, adapted from the PRISMA-ScR reporting framework.[26]

Screening and eligibility assessment were performed by a single reviewer (the author). This is a material limitation: no second independent screener was used, formal inter-rater agreement statistics (e.g., Cohen's kappa) are not available, and exact record counts at each intermediate screening stage were not contemporaneously logged in a form that permits retrospective reconstruction of a complete PRISMA-ScR count-by-count flow diagram. **Figure 1** therefore reports only the two counts that can be stated with confidence—the initial candidate pool (n = 190) and the final included set (n = 89)—and this gap is disclosed explicitly rather than backfilled with reconstructed figures.

Grey literature (vendor white papers, platform blogs, sponsor-issued reports) was eligible for inclusion where it met the same transparency criterion; where a grey-literature source reported a quantitative claim without a stated comparator or observation period, the claim is retained here only if it is explicitly labelled Tier D or E and is not used to anchor a headline conclusion. No formal critical-appraisal checklist (e.g., AMSTAR, GRADE) was applied for the reason given in the Evidence Classification Framework above; each finding is instead tagged with its evidence tier at first mention and in **Tables 1, 2,** and **4**.

Synthesis

Narrative synthesis was used rather than meta-analysis, given the heterogeneity of study designs, outcome definitions, and the predominance of proof-of-concept and observational reports in the current evidence base. Findings were organised into the four functional technology categories defined in the Background (Document Generation, Metadata Automation, Operational Prediction, Statistical Augmentation), followed by regulatory alignment and implementation challenges.

Results

Document Generation: LLM-Based Protocol and CRF Drafting

Several commercial platforms have demonstrated the feasibility of converting a structured study synopsis into an initial full protocol draft within hours rather than weeks. Systems such as Protocol AI (Risklick) and AuroraPrime RMA (AlphaLife Sciences) couple LLMs with RAG knowledge bases drawn from regulatory guidance documents, CDISC standards, and sponsor-specific historical protocols, generating section-level draft text organised around protocol-type templates encoding ICH E6(R3) and ICH E8(R1) structural requirements (Tier D:

Table 4: Key Challenges in AI Adoption for Protocol and CRF Development, with Mitigation Strategies.

Challenge	Nature of Risk	Mitigation Approach	Regulatory Consideration
LLM hallucination	Factually incorrect protocol content may pass without expert detection	RAG architectures grounded in verified sources; mandatory expert review of all AI-generated text	FDA COU credibility framework requires evidence of output accuracy in intended use context (where in scope)
Training data bias	Models trained on non-representative populations perpetuate exclusions	Fairness-constrained learning; stratified subgroup performance evaluation; diversity-aware objective functions	FDA and EMA diversity guidance; EU AI Act bias-auditing requirements (where high-risk criteria are met)
Data privacy constraints	Patient-level training data subject to GDPR/HIPAA; limits cross-sponsor data sharing	Federated learning; synthetic data generation using GANs or diffusion models	HIPAA authorisation provisions; GDPR data-processing agreements with AI vendors
Lack of prospective validation	Published gains may reflect publication bias, vendor-reporting bias, and single-sponsor optimism; CIs rarely reported	Pre-registered prospective multi-sponsor studies; standardised, consistently defined performance metrics with reported CIs	ICH E6(R3) quality-by-design; FDA credibility assessment requires fit-for-purpose evidence
Workforce readiness	CDM professionals lack hybrid skills for AI governance in regulated environments	Targeted training; human-in-the-loop governance frameworks; SCDM certification pathway updates	GCP requires qualified personnel to oversee all regulated processes
Audit trail, model monitoring, and change management	Undocumented model updates or silent performance drift can invalidate prior validation without detection	Version-controlled model registries; scheduled performance re-validation; documented change-control procedures for any model, prompt, or RAG-corpus update	ICH E6(R3) traceability principles; FDA COU framework expectations for maintained model credibility over the product lifecycle

platform-reported).[8–9] A published proof-of-concept using GPT-4 with domain-specific fine-tuning demonstrated cosine similarity scores of up to 0.80 and Recall-Oriented Understudy for Gisting Evaluation (ROUGE) precision of up to 0.70 between AI-generated and expert-authored introduction sections (Tier B). These metrics indicate substantial lexical and semantic overlap with an expert reference text, though neither metric assesses clinical judgement, factual accuracy, or regulatory adequacy directly.[10]

On evaluation methodology beyond ROUGE and cosine similarity: both metrics quantify surface-level textual similarity to a reference document and are known to correlate imperfectly with expert judgements of factual correctness or clinical appropriateness in health-text generation tasks.[21] Several of the sources reviewed here supplement these metrics with structured expert-rater review (e.g., completeness against a section checklist, or a binary/Likert acceptability rating from a clinical reviewer), but a standardised, publicly validated rubric for evaluating AI-generated protocol or CRF text—covering completeness, factual accuracy, and regulatory-compliance dimensions separately—does not yet exist across the sources reviewed; this is identified as a priority methodological gap (**Table 4**).

Amendment prevention represents a particularly valuable application within this category. By analysing patterns in historical amendment data, AI tools can flag protocol sections associated with high revision risk: eligibility thresholds inconsistent with disease epidemiology, visit schedules infeasible at typical site staffing levels, and statistical assumptions at odds with published natural history data. A synthesis from the Duke Clinical Research Institute estimated that systematic adoption of AI-based protocol simulation could reduce major amendment incidence by 20–30% across Phase II and III programmes (Tier D/E: think-tank synthesis).[12] Early sponsor-reported data suggest first-draft amendment rates declining by up to 28% in AI-assisted workflows versus historical comparators (Tier D: single-sponsor observational), though independent confirmation is lacking.[8] For clarity, "first-draft amendment rate" as used by the cited sources denotes the proportion of protocols requiring a substantive amendment measured against the first submitted (pre-regulatory-review) version, as distinct from amendment rates measured after internal review board/ethics committee or regulatory review; sources do not consistently specify which definition they use, which limits cross-study comparability (**Table 2**).

Metadata Automation: CRF Design and CDISC Compliance

AI-assisted CRF generation treats the Schedule of Activities (SoA) table as a structured data input. NLP pipelines parse the SoA and surrounding protocol text to identify data collection requirements at each visit, then map those requirements to CDASH-standard domain templates. Systems with cross-transclusion features maintain synchronisation across related documents when SoA parameters change, reducing version inconsistencies that arise when protocol and CRF documents are managed separately (Tier D).[9,14] Initial (draft) form-build timelines have been reported to compress from four-to-eight weeks to under 72 hours; sources are explicit that this figure describes the draft build only, with one-to-two additional weeks required for review and regulatory-compliance validation before the database is production-ready, a distinction this review preserves throughout **Table 1** rather than presenting the 72-hour figure as the total time to a validated database.[14]

Machine learning models trained on historical CRF data and query repositories can generate edit-check recommendations ranked by predicted query frequency and clinical significance. Adaptive eCRF systems incorporating this logic report data-query rate reductions of 22–35% relative to conventionally built comparators (Tier D).[2,15] For SDTM mapping, NLP tools that assign CRF fields to SDTM domains by analogy with structurally similar historical fields have achieved domain-assignment accuracy exceeding 90% for standard domains, with more complex or non-standard fields flagged for human expert review (Tier B/D). Across published evaluations, SDTM mapping and annotated CRF generation time has been reported to reduce by 40–60% compared with fully manual processes.[5,16] As noted in the Background, these tools operationalise the CDASH-to-SDTM transformation specifically; they do not replace CDASH-level CRF design decisions, which remain a separate, upstream step.

Reductions in CRF field count achieved through this category of tool should be read, per the burden distinction introduced in the Background, as reductions in site workload, data entry, source-document verification, and data-management effort—not, without further evidence, as reductions in patient burden. None of the sources reviewed in this subsection reported a corresponding reduction in patient-facing procedures (visits, assessments, or sample collection); where such a reduction is claimed elsewhere in the literature, it should be substantiated by a change to the visit schedule or assessment battery itself, not inferred from CRF field count alone.

Operational Prediction: Enrolment Forecasting and Eligibility Optimisation

NLP-based protocol analysers can cross-reference proposed eligibility criteria against natural history datasets and real-world EHR populations to identify criteria that may render target populations implausibly small. The Stanford Trial Pathfinder system demonstrated retrospectively that targeted relaxation of specific haematological and medication-based exclusion criteria could double the size of eligible patient populations in some therapeutic areas without measurable change in safety profile. For enrolment forecasting, machine learning models trained on historical site- and country-level recruitment data achieved prospective accuracy rates of approximately 85%, with reported deployment reductions in recruitment delay of

around 25% (Tier B).[11] As noted in **Table 2**, the precise statistical definition of "accuracy" (e.g., mean absolute percentage error against a defined forecast horizon versus a categorical on-time/delayed classification) is not uniformly reported across the underlying studies, and this heterogeneity should be considered when comparing figures across sources.

Statistical Augmentation: Digital Twins and PROCOVA-MMRM

A note on scope and terminology. This subsection touches on prognostic-covariate-adjustment statistics that fall outside clinical data management's core disciplinary boundary; the author is not a biostatistician, and the terminology below follows exactly what each cited source reports without extrapolation. Readers intending to apply these methods to trial design should consult a qualified biostatistician, particularly regarding the regulatory acceptability of a specific covariate-adjusted analysis for a given protocol.

At the frontier of AI integration, digital-twin approaches provide prognostic covariates for adaptive statistical models. The PROCOVA-MMRM framework, which incorporates AI-generated digital-twin predictions into mixed-model repeated-measures analyses, has been reported to produce power gains described by the source as equivalent to a 25–30% increase in effective sample size, while maintaining Type I error control acceptable to regulatory reviewers in the cited applications (Tier B/C).[7,13] Four related but distinct concepts appear across this literature, and this review treats them as non-interchangeable:

- Variance reduction: the direct statistical consequence of adjusting for a prognostic covariate; this is the primitive quantity reported by the underlying statistical model.
- Effective sample-size increase: a re-expression of that variance reduction in sample-size-equivalent terms, valid only under the assumptions of the fitted model.
- Projected sample-size reduction: a design-stage inference about how many fewer participants a future trial could enrol for the same target power; this additionally depends on the effect-size assumption and regulatory acceptance of the covariate-adjusted analysis for that specific programme, and is not automatically implied by the first two quantities.
- Observed power improvement: a post hoc or simulation-based estimate of realised power gain in a specific dataset, which may differ from the design-stage projection.

Its first application in Alzheimer's disease trials projected sample-size reductions sufficient to shorten development timelines by 12–18 months per programme, per the cited source's own projection rather than an independently confirmed outcome (Tier B).[13] Readers are directed to the primary methodological references for the full model specification before applying any of these figures to a specific trial design.[7,13,22]

Regulatory Landscape

The regulatory environment for AI in drug development has matured considerably since 2022, though several frameworks discussed here remain in draft or early-implementation form; specific effective dates and scope limitations are stated explicitly below because earlier drafts of this review understated them.

The FDA's January 2025 draft guidance on using AI to support regulatory decision-making introduces a context-of-use (COU) credibility framework that calibrates validation requirements to the risk associated with a specific intended use.[17] As of this review, the guidance remains in draft form (public comment closed April 2025; not yet finalised), and its stated scope applies specifically to AI models used to produce data supporting a regulatory decision about a drug or biological product's safety, effectiveness, or quality—the guidance explicitly does not extend to AI used for drug discovery or for operational efficiencies such as internal workflow support or drafting a regulatory submission.[24] This is a materially important scope boundary for this review: most of the Document Generation and Metadata Automation tools discussed above fall outside the COU framework's current stated scope, whereas Operational Prediction and Statistical Augmentation tools that generate evidence feeding directly into a regulatory decision are more likely to fall within it. By late 2024, the FDA had processed several hundred submissions containing AI components in some form, reflecting substantial sponsor adoption.[17]

The EMA's October 2024 Reflection Paper on AI in the medicinal product lifecycle adopts a compatible risk-based orientation, with expectations organised by development phase.[18] Separately, the EMA's Committee for Medicinal Products for Human Use (CHMP) adopted its first qualification opinion for an AI methodology on 27 February 2025: AIM-NASH, an AI-based tool that assists pathologists in scoring liver biopsies for disease activity in MASH/NASH clinical trials.[24] This is a narrower and more specific precedent than a general endorsement of "AI-generated evidence" in marketing-authorisation submissions; it qualifies one context-of-use for one histological scoring tool, and should not be read as establishing a blanket qualification pathway for AI protocol- or CRF-authoring tools.

The EU AI Act (Regulation 2024/1689) entered into force in August 2024; it establishes a risk-based framework in which "high-risk" status is not a blanket classification for all AI used in healthcare, but depends on specific criteria under Article 6: an AI system is high-risk either as a safety component of a regulated product (e.g., certain medical devices) or where it falls within an Annex III use case that poses a significant risk to health, safety, or fundamental rights.[25] Article 6(3) further provides a derogation for systems that perform a narrow procedural task, improve the result of a previously completed human activity, or require proper human review before their output is acted upon; criteria that may be relevant to the mandatory-expert-review design of the Document Generation and Metadata Automation tools discussed in this review, though the applicability of this derogation

to any specific tool is a case-by-case legal determination outside the scope of this review and evolving via Commission guidance.[25]

The revised ICH E6(R3) Good Clinical Practice guideline provides an additional normative anchor. Contrary to earlier drafts of this review, E6(R3) was not finalised in 2023: the ICH Assembly adopted the final guideline on 6 January 2025; it became applicable in the European Union on 23 July 2025; the US FDA issued its final E6(R3) GCP guidance in September 2025; and UK implementation occurred on 28 April 2026,illustrating that regional adoption timelines differ materially and should be checked against the applicable jurisdiction before citing a single "effective" date.[20] Its quality-by-design mandate—that every collected data element be traceable to a specific scientific or safety objective—creates a direct rationale for AI-assisted CRF rationalisation, since systematic traceability of this kind is difficult to achieve consistently through manual review alone (**Table 3**).

Discussion

The body of evidence reviewed here is directionally consistent: AI-assisted workflows for protocol authoring and CRF design have been associated, across the Tier B–D sources cited, with faster outputs, fewer downstream errors, and more reliable CDISC compliance relative to purely manual processes. Where the underlying evidence is strong enough to support it, this association is potentially large enough to be clinically and operationally meaningful; where it is not—most notably for the Tier D, single-sponsor, and platform-reported figures that dominate **Table 1**—it should be read as a promising signal rather than a demonstrated causal effect. For clinical data management professionals, the practical implication is the same either way: earlier validated databases, shorter query-resolution cycles, and more reliable SDTM mapping represent plausible, evidence-supported directions for efficiency gains, to be confirmed rather than assumed at the level of an individual programme.

Several important qualifications apply, consolidated here for emphasis even though most are introduced at first mention above.

Evidence maturity. The evidence base is predominantly observational (Tier D). Most available data come from single-sponsor deployments or vendor-reported benchmarks rather than prospective controlled evaluations, none of the reviewed sources reported confidence intervals, and outcome definitions—particularly "accuracy," "consistency," and "first-draft amendment rate"—are not uniformly specified across sources (**Table 2**). Publication and vendor-reporting bias almost certainly inflate reported performance figures. The field needs prospective, multi-sponsor studies that compare AI-assisted and conventional workflows using pre-specified, consistently defined outcome metrics, with reported confidence intervals, across representative ranges of therapeutic areas, protocol types, and organisational settings.

Hallucination risk. The hallucination risk associated with LLM-based tools is a genuine concern in a regulated context. An LLM generating a plausible but factually incorrect dosing rationale, or an eligibility-criterion threshold with no regulatory precedent, could introduce protocol errors whose consequences may not be apparent until patient enrolment is under way.[6,21] RAG architectures substantially reduce this risk by anchoring outputs in verified source material but cannot eliminate it. Human expert review of all AI-generated protocol content must be treated as a non-negotiable governance requirement, not an optional quality-assurance step, and this obligation must be formally documented in sponsor quality management systems and resourced accordingly.

Training-data bias. AI models trained on historical clinical trial data inherit the demographic and clinical biases of those data. Clinical research populations have historically under-represented women, older patients, and ethnic and racial minorities; AI tools that optimise eligibility criteria based on historical precedent risk encoding and perpetuating these exclusions.[22] Addressing this requires fairness-constrained model design, stratified performance evaluation across demographic subgroups, and explicit integration of FDA and EMA diversity and inclusion requirements into the objective functions against which AI tools are trained and validated.

Data governance constraints. Patient-level clinical trial data is subject to GDPR, HIPAA, and GCP data protection obligations, limiting what can be shared with third-party AI developers. Federated learning—training across distributed datasets without centralising sensitive data—and synthetic data generation offer technically credible pathways around these constraints, but neither approach has been fully validated for the complex, high-dimensional data types that characterise clinical trial populations.[23]

For the clinical data management profession specifically, AI integration raises questions about role evolution. Data managers who previously spent the majority of their time building CRF databases and writing edit-check specifications may find those activities increasingly automated. The skills most valuable in an AI-augmented environment—critical evaluation of model outputs, governance framework design, regulatory interaction around AI methodologies, and translating AI capabilities into sponsor-specific quality systems—are not yet systematically taught in CDM training programmes. Professional development curricula need to anticipate this transition rather than respond to it reactively.

Conclusions

The evidence reviewed here shows a consistent and, within several operational areas, substantial directional signal that generative AI and predictive analytics can improve the efficiency of clinical trial protocol development and CRF design. That evidence base is predominantly observational (Tier D) and concentrated in single-sponsor or platform-reported deployments; the peer-reviewed, independently validated evidence (Tier A/B) is currently strongest for enrolment-forecasting models and narrowest for protocol- and CRF-drafting tools. For clinical data management professionals, these tools represent a plausible opportunity to reduce preventable rework, improve data quality at source, and contribute

more strategically to trial design—provided that adoption is grounded in rigorous governance, mandatory expert oversight, and honest, tier-explicit acknowledgement of current evidentiary limitations, rather than treatment of vendor-reported figures as established fact.

The regulatory landscape is moving in a direction that is broadly supportive of AI integration while maintaining appropriate expectations around validation, transparency, and accountability, though several of the frameworks discussed here—notably the FDA's COU credibility framework—remain in draft form or apply to a narrower scope than is sometimes assumed in industry commentary. Alignment between the FDA's (draft) COU framework, the EMA's Reflection Paper and AIM-NASH qualification precedent, and ICH E6(R3)'s quality-by-design principles provides a coherent, if still evolving and jurisdictionally staggered, normative foundation for AI governance in clinical research.

What the field most needs now is the evidence infrastructure to substantiate the operational claims that vendor-reported data have suggested: prospective, pre-registered, multi-sponsor studies with pre-specified primary outcomes, consistently defined metrics, reported confidence intervals, diverse therapeutic areas, and robust comparator conditions. Building that evidence base is a shared responsibility of sponsors, academic clinical triallists, professional bodies such as SCDM, and the AI developers whose tools are already in use. Continued adoption on the basis of uncontrolled observational data carries a real risk of a high-profile failure that would set the field back further than a temporary, evidence-gathering delay in adoption.

Competing Interests

The author has no competing interests to declare.

References

1. **Getz KA, Stergiopoulos S, Kaitin KI.** Quantifying the magnitude and cost of collecting extraneous protocol data. *Am J Ther*. 2015;18(2):e51–e57. DOI: https://doi.org/10.1097/MJT.0b013e31826fc4aa
2. **Rai A.** *eCRFs in Clinical Trials: Understanding Structure, Challenges, and the Role of AI*. Clinion. Published February 28, 2026. Accessed October 5, 2026.
3. **Getz KA, Wenger J, Campo RA, Seguine ES, Kaitin KI.** Assessing the impact of protocol design changes on clinical trial performance. *Am J Ther*. 2008;15(5):450–457. DOI: https://doi.org/10.1097/MJT.0b013e31816b9027
4. **Bellary S, Krishnankutty B, Latha MS.** Basics of case report form designing in clinical research. *Perspect Clin Res*. 2014 Oct;5(4):159–66. PMID: 25276625; PMCID: PMC4170533. DOI: https://doi.org/10.4103/2229-3485.140555
5. **Yang E, Katz L, Shenoy S.** Automated mapping of electronic data capture fields to SDTM. *PLoS One*. 2024 Nov 7;19(11):e0312721. PMID: 39509395; PMCID: PMC11542782. DOI: https://doi.org/10.1371/journal.pone.0312721
6. **Liddicoat JE, Lenarczyk G, Aboy M,** et al. A policy framework for leveraging generative AI to address enduring challenges in clinical trials. *npj Digit Med*. 2025;8:33. DOI: https://doi.org/10.1038/s41746-025-01440-5
7. **Wang D, Florian H, Lynch SY,** et al. Using AI-generated digital twins to boost clinical trial efficiency in Alzheimer's disease. *Alzheimers Dement (N Y)*. 2025;11:e70181. DOI: https://doi.org/10.1002/trc2.70181
8. **Foote HP, Hong C, Anwar M,** et al. *Embracing generative artificial intelligence in clinical research: opportunities, challenges, and solutions*. Think Tank Report. Duke Clinical Research Institute; 2025. DOI: https://doi.org/10.1016/j.jacadv.2025.101593
9. **AlphaLife Sciences.** *Blueprint for breakthrough: AI-driven protocol authoring accelerates Phase III clinical trials*. AlphaLife Sciences; Published November 18, 2025. Accessed October 5, 2026. https://alphalifesci.com/blog/blueprint-for-breakthrough-ai-driven-protocol-authoring-accelerates-phase-iii-clinical-trials
10. **Maleki M.** Clinical Trials Protocol Authoring using LLMs; 2024. DOI: https://doi.org/10.48550/arXiv.2404.05044
11. **Li Y, Zhang W, Chen K,** et al. Advancing clinical trial outcomes using deep learning and predictive modelling. *World J Adv Res Rev*. 2024;24(3):001–025. DOI: https://doi.org/10.30574/wjarr.2024.24.3.3671
12. **Foote H, Hong C, Anwar M,** et al. Embracing Generative Artificial Intelligence in Clinical Research and Beyond: Opportunities, Challenges, and Solutions. *JACC Adv*. 2025 Mar;4(3). DOI: https://doi.org/10.1016/j.jacadv.2025.101593
13. **Vidovszky AA, Fisher CK, Loukianov AD,** et al. Increasing acceptance of AI-generated digital twins through clinical trial applications. *Clin Transl Sci*. 2024;17:e13897. DOI: https://doi.org/10.1111/cts.13897
14. **Vangipurapu M.** *AI and automation in clinical trials*. Clinion. Published May 12, 2022. Accessed October 5, 2026.
15. **Janardhanan A.** *How AI, automation, and machine learning are upgrading clinical trials*. Octalsoft Industry Blog. Published July 10, 2024. Accessed March 2025. https://www.octalsoft.com/how-ai-automation-and-machine-learning-are-upgrading-clinical-trials/
16. **Coherent Solutions.** *Machine learning and AI in clinical trials: use cases 2025*. Coherent Solutions White Paper. Published 2025; updated April, 2026. Accessed October 5, 2026. https://www.coherentsolutions.com/insights/role-of-ml-and-ai-in-clinical-trials-design-use-cases-benefits
17. **US Food and Drug Administration.** Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products: Draft Guidance for Industry and Other Interested Parties. Docket FDA-2024-D-4689. FDA;

Published January 2025. Accessed October 5, 2026. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological

18. **European Medicines Agency.** Reflection paper on the use of artificial intelligence in the medicinal product lifecycle. EMA/CHMP; Published October 2024. Accessed October 5, 2026. https://www.ema.europa.eu/system/files/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle-en.pdf
19. **European Parliament and Council.** Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Off J Eur Union*. Published July 12, 2024. Accessed October 5, 2026. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
20. **International Council for Harmonisation.** E6(R3) Guideline for Good Clinical Practice, Final Version, adopted 6 January 2025. Applicable in the EU from 23 July 2025; US FDA final guidance issued September 2025; UK implementation scheduled 28 April 2026. Accessed October 5, 2026. https://www.ich.org/news/ich-e6r3-good-clinical-practice-annex-2-adopted-and-published
21. **Huang L, Bedi S, Liu Y,** et al. Testing and evaluation of health care applications of large language models: a systematic review. *JAMA*. 2025;333:319–328. DOI: https://doi.org/10.1001/jama.2024.21700
22. **Bordukova M, Makarov N, Rodriguez-Esteban R,** et al. Generative artificial intelligence empowers digital twins in drug discovery and clinical trials. *Expert Opin Drug Discov*. 2024;19(1):33–42. DOI: https://doi.org/10.1080/17460441.2023.2273839
23. **Podichetty JT, Bauer AM, Xu R,** et al. How AI transforms regulatory submission: current clinical implementation and future prospects. *Clin Transl Sci*. 2025;18(12):e70434. DOI: https://doi.org/10.1111/cts.70434
24. **European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP).** Qualification Opinion for Artificial Intelligence-Based Measurement of Non-alcoholic Steatohepatitis Histology in Liver Biopsies to Determine Disease Activity in NASH/MASH Clinical Trials (AIM-NASH). 24 April 2025 EMADOC-1700519818-1761332 Corr.1 https://www.ema.europa.eu/en/documents/other/qualification-opinion-artificial-intelligence-based-measurement-non-alcoholic-steatohepatitis-histology-liver-biopsies-determine-disease-activity-nash-mash-clinical-trials_en.pdf
25. **European Commission, AI Act Explorer.** Article 6: Classification Rules for High-Risk AI Systems; Annex III. Regulation (EU) 2024/1689. Accessed October 5, 2026. https://ai-act-service-desk.ec.europa.eu/en/ai-act-explorer
26. **Tricco AC, Lillie E, Zarin W,** et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. 2018;169(7):467–473. DOI: https://doi.org/10.7326/M18-0850

How to cite this article: Prasanna Kumar, CS. Generative Artificial Intelligence and Predictive Analytics in Clinical Data Management: A Review of the Emerging Applications to Study Protocol Development and Case Report Form Design. *Journal of the Society for Clinical Data Management*. 2026; 6(1): 9, pp. 1–11. DOI: https://doi.org/10.47912/jscdm.499

Submitted: 30 March 2026 **Accepted:** 23 September 2026 **Published:** 26 October 2026

Copyright: © 2026 SCDM publishes JSCDM content in an open access manner under a Attribution-Non-Commercial-ShareAlike (CC BY-NC-SA) license. This license lets others remix, adapt, and build upon the work non-commercially, as long as they credit SCDM and the author and license their new creations under the identical terms. See https://creativecommons.org/licenses/by-nc-sa/4.0/.

Journal of the Society for Clinical Data Management is a peer-reviewed open access journal published by Society for Clinical Data Management.

OPEN ACCESS