Global Health & Medicine. 2026; 8(3):182-192. REVIEW
DOI: 10.35772/ghm.2026.01051

Artificial intelligence (AI)-aided clinical data management: Applications, human-in-the-loop workflows, and regulatory considerations

Saya Ohi[1], Tomoko Iwamoto[2], Daiki Ikeda[1], Yuichi Kawanishi[1], Koji Kitajima[2], Hajime Ohyanagi[3,*]

[1] Office of Bioinformatics, Department of Joint Center for Researchers, Associates and Clinicians (JCRAC), Center for Clinical Sciences, Japan Institute for Health Security, Tokyo, Japan;
[2] Office of Clinical Data Management, Department of Joint Center for Researchers, Associates and Clinicians (JCRAC), Center for Clinical Sciences, Japan Institute for Health Security, Tokyo, Japan;
[3] Department of Joint Center for Researchers, Associates and Clinicians (JCRAC), Center for Clinical Sciences, Japan Institute for Health Security, Tokyo, Japan.

Abstract: Clinical data management (CDM) is central to the quality of clinical research. In Japan, CDM faces a shortage of qualified personnel, particularly in academic research organizations (AROs), as well as increasing data volume and complexity. Rapid advances in artificial intelligence (AI), especially large language models, have therefore attracted attention as a way to support CDM. This review summarizes domestic and international examples of AI utilization in CDM-related tasks, including data cleaning, medical coding, and query generation. Across the cases reviewed, a common implementation principle emerged: a human-in-the-loop design in which AI performs initial processing or detection, while final judgment remains with human personnel. This design is especially relevant to AROs, where high data quality must be maintained with limited CDM human resources. Regulatory frameworks, including ICH E6 (R3) and the FDA-EMA Guiding Principles, are beginning to address AI use, but how AI-aided processes should be handled under Good Clinical Practice remains under discussion. Comprehensive risk mitigation is therefore essential. AI and data are interdependent: better data improve AI performance, and better AI can further improve data quality. The shift from manual processes to human-AI collaborative workflows is likely to accelerate, and CDM must develop the technical, regulatory, and risk-management frameworks needed to support that transition.

***Keywords*:** artificial intelligence (AI), clinical data management (CDM), human-in-the-loop (HITL), AI-related regulations and guidelines

1. Introduction

Clinical research is human-subjects research that evaluates health, disease, and medical interventions to generate evidence for patient care, regulatory decision-making, and medical practice (*1*). Clinical data management (CDM) underpins the reliability of clinical research data by supporting the design, collection, quality control, and preparation of data for analysis (*2*). Its responsibilities extend from the study planning stage to the finalization of analysis-ready datasets. The International Council for Harmonisation (ICH) E6 (R3) Good Clinical Practice guideline, an internationally harmonized standard for clinical trials, identifies data integrity, traceability, confidentiality, reliability, and fitness for purpose as core principles of data governance (*3*) (Figure 1, bottom line). Because errors introduced at the data level cannot be corrected by subsequent statistical analysis or interpretation, CDM is a foundational function for ensuring the quality of clinical research.

The environment surrounding CDM has changed substantially in recent years. Clinical trials now use diverse data sources beyond conventional case report forms (CRFs), including electronic health records (EHRs), wearable devices, patient-reported outcomes (PROs), and medical images. As a result, data volume has increased, and data structures have become more complex (*4*). These changes have increased the workload of CDM and have made it increasingly necessary to maintain data quality efficiently. At the same time, qualified data managers remain in short supply (*5*),

particularly in organizations without dedicated CDM departments. In this context, rapid advances in artificial intelligence (AI), especially large language models, have attracted attention as a potential means of supporting CDM tasks such as data cleaning, medical coding, and query generation (*6*).

This review summarizes examples of AI use in key CDM-related tasks, including data cleaning, medical coding, and query generation, and discusses the regulatory requirements and risk-management considerations necessary for the responsible implementation of AI in clinical research. A common feature of the cases reviewed is that AI supports initial processing, anomaly detection, coding, or text generation, whereas final judgment remains with qualified human personnel. This operational model is known as human-in-the-loop (HITL) and is based on collaboration between humans and AI rather than full automation (*7*). The HITL concept is particularly important for clinical research organizations that must maintain high data quality with limited CDM resources (*8*). In this review, HITL is positioned as the central framework for discussing AI applications, relevant regulations, and risk management in CDM.

2. Background: CDM challenges, policy background, and AI utilization

2.1. Roles and workflow of CDM

High-quality data are essential for valid analysis and reliable conclusions in clinical research (*3,9*). Even well-designed studies cannot generate robust evidence unless protocol-specified data are collected accurately and completely (*3,10*). Data managers are involved from the planning stage to ensure that primary endpoint data are captured appropriately (Figure 1, Clarification of Endpoints, Protocol Review). Building an electronic data capture (EDC) system is one of their main responsibilities (Figure 1, EDC Construction). Because EDC is only the data-entry infrastructure, data checks, reporting functions, and entry-support must also be implemented to enable accurate and timely data collection (*3*). Accuracy in CDM goes beyond confirming consistency between entered data and source documents; it also requires confirming that protocol-required data have been collected and are available in an analysis-ready format (*3,11*) (Figure 1, Creating a Collection Plan). For example, primary endpoint data should be collected without missing values, and visit-based data should be entered at time points specified in the protocol schedule (*3,10*). Data cleaning systematically refines collected data and is a core CDM activity for ensuring data quality (Figure 1, Data Cleaning). Appropriate recording and management of audit trails for data modifications and updates are also mandatory in clinical research (*3,12,13*) (Figure 1, Recording of Audit Trails).

Data managers support clinical research by building databases that faithfully capture protocol-defined endpoints and by maintaining data-entry quality. Their responsibilities span EDC construction, system operation, data cleaning, and data delivery, making them essential to research quality. In this article, "clinical research" refers broadly to clinical trials in general and is not limited to pharmaceutical trials.

2.2. Current challenges in CDM

The CDM field faces a serious shortage of qualified personnel, particularly in Japanese Academic Research

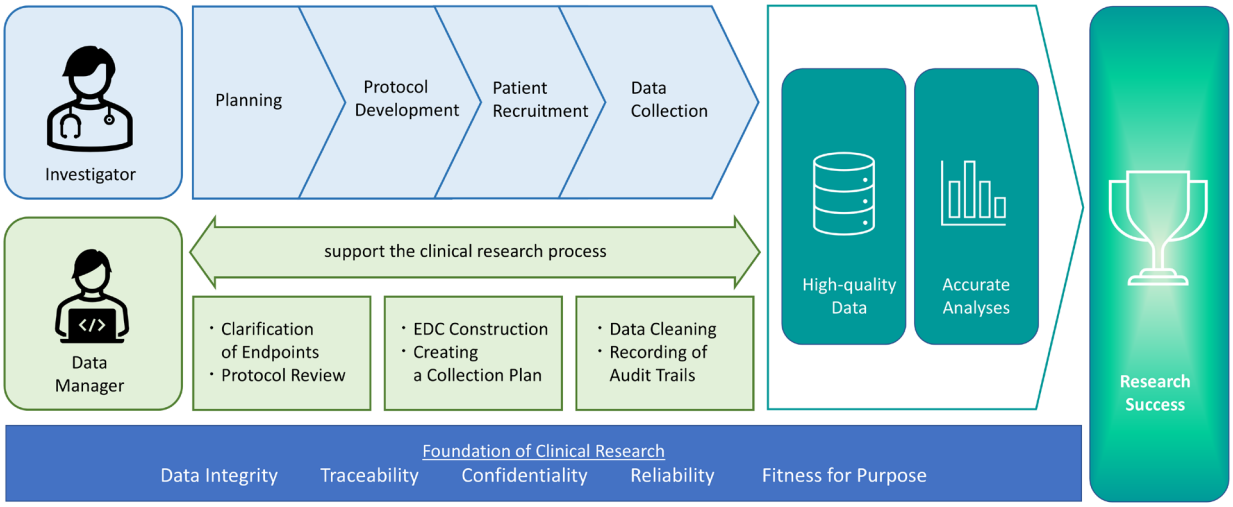

Figure 1. Clinical research process and data management. This simplified schematic highlights clinical research stages and data manager activities relevant to this review. Data managers support investigators across planning, protocol development, patient recruitment, and data collection phases, contributing to high-quality data and accurate analyses that underpin research success. The five principles at the bottom—data integrity, traceability, confidentiality, reliability, and fitness for purpose—represent the foundational CDM requirements specified in ICH E6 (R3).

Organizations (AROs). Unlike pharmaceutical companies, which often have dedicated CDM departments, many AROs rely on a small number of data managers to oversee multiple trials simultaneously. In a report by the All-Japan Conference of Deans of Medical Schools and Hospital Directors, only 16 of 80 universities (20%) reported having three or more full-time data managers (*14*). Including the seven institutions that were developing such capacity, 64 universities had insufficient CDM human resources. Thus, a limited workforce must manage multiple concurrent trials. At the same time, the volume and diversity of data sources are increasing (*15*). Data are becoming more diverse, complex, and voluminous, and decentralized clinical trials (DCTs) have widened the range of sources. Unstructured data, including medical images, are also expected to be incorporated more often (*16,17*). Maintaining data quality under these conditions will become increasingly challenging.

2.3. Policy background

Several policy developments have shaped the CDM environment in Japan. In Japanese clinical research, the Ethical Guidelines for Medical and Biological Research Involving Human Subjects (*18*) and the Enforcement Regulations on the Clinical Research Act (*19*) require research protocols to specify data collection and management methods; monitoring is required under the guidelines, when research involves invasiveness or intervention. These requirements make data quality assurance a planning-stage legal obligation. For pharmaceutical trials, the GCP Ministerial Ordinance establishes equivalent requirements (*20*). In parallel, Japan has adopted the Society 5.0 vision, and the Medical DX Promotion Headquarters, led by the Cabinet Office and the Ministry of Health, Labor and Welfare (MHLW) is accelerating healthcare digitalization (*21*). AI-aided data management is therefore increasingly viewed not as an experimental option, but as a practical necessity. Together with personnel shortages, these policy signals have encouraged active consideration of AI applications in CDM.

2.4. AI in CDM

AI, particularly large language models (LLMs), has advanced rapidly in recent years. Major AI and cloud providers are increasingly developing healthcare-oriented services based on medical data standards such as FHIR and DICOM (Supplementary Table S1, *https://www.globalhealthmedicine.com/site/supplementaldata.html?ID=123*). Among these advances, several features of AI and LLMs are especially relevant to clinical research data (*22*). First, they can process unstructured text, such as physician notes, pathology reports, and adverse event narratives, and extract structured information. This capability addresses an area in which conventional tools have struggled and may create new categories of research data (*23*). Second, LLMs can be adapted to specific domains through fine-tuning on specialized corpora or through careful prompt engineering, reducing the need for large-scale custom development. Third, AI can scale processing volume while improving output consistency. Manual data review is affected by reviewer experience, fatigue, and workload, which can reduce the thoroughness and consistency of checks. AI systems, including LLMs, avoid these sources of variation and can apply uniform criteria repeatedly to large datasets, making them useful complements to quality management (*6*).

Current AI systems also have important limitations and risks, discussed later in this article. The next section uses case examples to examine how AI capabilities can be applied to specific CDM-related tasks.

3. Case examples in AI-aided CDM

A defining feature of AI in professional applications is that non-specialists can operate it through natural language instructions. This has accelerated adoption across many professional domains (*24,25*). The following sections present examples of AI utilization in CDM-related tasks and adjacent data-processing approaches that may be transferable to CDM workflows. Peer-reviewed literature documenting AI applications in routine CDM practice remains limited; therefore, some examples are drawn from publicly accessible online sources. The sources should not be interpreted as official positions of the organizations concerned. The examples should therefore be interpreted according to their maturity: some are already used or evaluated in CDM-related tasks, whereas others illustrate approaches that may require further validation before routine CDM implementation.

3.1. Data cleaning

Among CDM tasks, data cleaning is quality-critical and rule-intensive, making it a strong target for automation (Figure 2A). A notable precedent is the work of Shi *et al.* on Belgian primary care EHR data (*26*). They built a clinical knowledge database (CKD) containing reference ranges, unit-conversion formulas, and outlier-detection criteria for each variable, and implemented an automated cleaning pipeline combining fuzzy matching and outlier detection (*26*). For more than one million records across 52 variables, the pipeline completed in 5.2 minutes a task that would have required 30 to 40 hours of manual work by an experienced statistician. Quality indicators improved for most variables, particularly the proportion of abnormal values caused by digit or unit errors (*26*). A key strength of this approach is that decisions are based on objective clinical knowledge rather than statistical distributions, enabling robust performance even in real-

world datasets with a high proportion of patients with disease.

Bönisch *et al*. proposed a machine learning (ML) approach for predictive data-quality assessment using data from medical data integration centers at German university hospitals (*27*). XGBoost and SVM were applied to echocardiography, laboratory, and medication data, and prediction results were stored as quality metadata in a data warehouse (*27*). Compared with conventional approaches that require item-by-item logical checks, this method may streamline large-scale quality checking while maintaining comparable rigor.

Although these studies were not conducted in routine clinical trial CDM settings, they illustrate data-quality assessment approaches that may be transferable to CDM workflows involving large-scale, heterogeneous clinical data.

We explored AI-aided data checking (Iwamoto *et al.*, The potential of data checking using AI and RPA. In: The 17th Annual Meeting of the Japan Society of Clinical Trials and Research. 2026). The workflow was designed to read raw CSV data downloaded from an EDC system and extract records that met predefined error conditions (Figure 2A). Generative AI produced check scripts from natural language descriptions of error conditions, and robotic process automation (RPA) executed the scripts to extract error data (Figure 2B). More than 94% of the anticipated errors embedded in test data were detected. The study also showed that prompt rules can improve reproducibility in data checking.

Together, these international precedents and the domestic initiative share a common implementation principle: AI performs the initial processing, while humans make the final judgment. This hybrid structure reflects a HITL (see Section 6.1) perspective rather than full automation. Streamlining data checking is central to CDM quality and efficiency, and AI may have a particularly large impact in AROs with lean CDM teams.

3.2. Medical coding

Evidence supporting AI-aided medical coding continues to accumulate. WHODrug Koda, developed by the Uppsala Monitoring Centre (UMC), combines text-processing algorithms, coding rules, and ML to convert free-text drug names in adverse event reports into standard WHODrug Global codes (*28*). In an evaluation of approximately 4.8 million drug entries in VigiBase, Koda increased the automatic coding rate from 61% to 89% compared with a simple direct matching while maintaining 97% coding accuracy (*28*). Its three-tier design automatically selects a code when confidence is high, presents candidate codes when ambiguity remains, and withholds coding when expert judgment is required, making it a rational HITL implementation. The AI-DM Task Force of the Japan Pharmaceutical Manufacturers Association (JPMA) has examined natural language processing-based automatic coding of disease names entered as free text in CRF fields, including medical history and adverse events, to ICD-10, as well as the

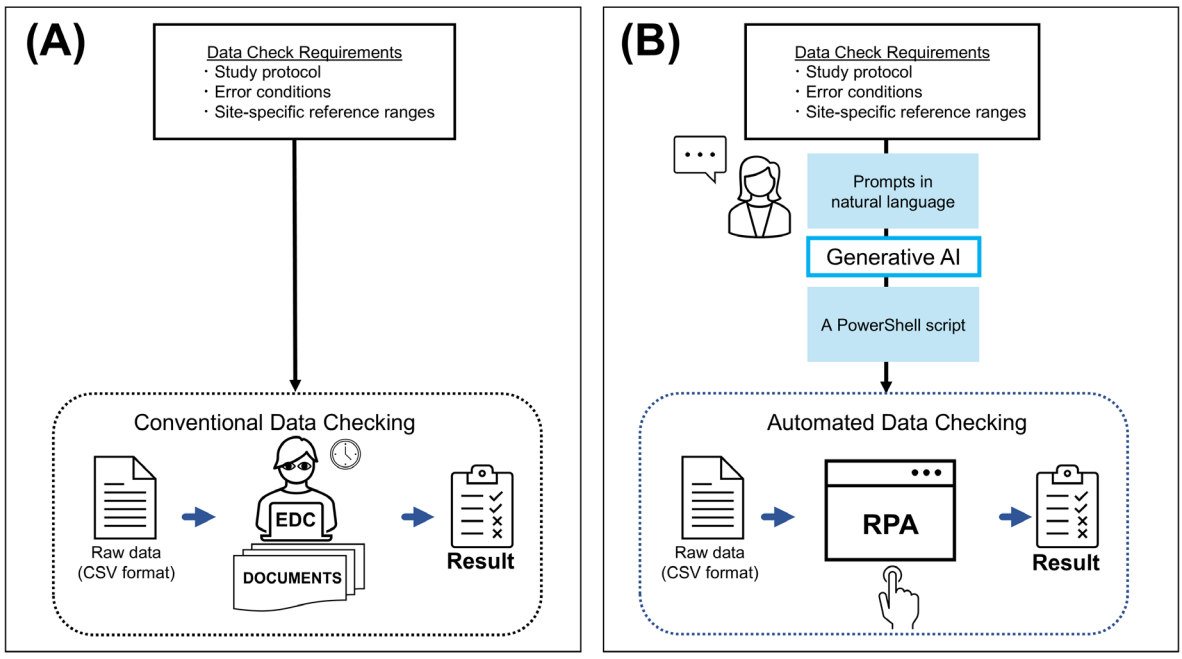

Figure 2. Conventional and AI-aided automated data-checking workflows. In the conventional approach **(A)**, data check requirements are manually translated into EDC-based review by a data manager. In the automated approach **(B)**, generative AI interprets the same requirements and generates executable scripts for robotic process automation (RPA), enabling automated error detection from raw CSV data.

human oversight needed for quality assurance (*29-31*). Similar initiatives have been reported for coding safety information, including adverse events, to MedDRA (*32*). Across drug and adverse event coding, the shared design principle is clear: AI processes high-confidence cases automatically, while human reviewers focus on cases requiring nuanced judgment. This approach helps concentrate limited specialist resources on high-value tasks.

3.3. Query generation

Query management is one of the most labor-intensive components of CDM, yet comprehensive query issuance appears to contribute only modestly to database quality. Stokman analyzed approximately 2 million queries from 20 Phase III trials at seven major pharmaceutical companies and found that the data-correction rate for conventional rule-based automated queries was approximately 1.4% (*33*). This suggests a shift from query quantity to query quality and prioritization, that is, toward risk-based query management. LLMs may offer a new way to address this challenge. In Japan, members of the JPMA task force evaluated two query-generation approaches: fine-tuning a base model and using an existing LLM without additional training. In both approaches, more than 80% of generated queries were judged suitable for direct operational use. The findings suggest that the choice of approach should depend on user-side requirements, including the availability of training data and the desired naturalness of query messages.

AI could support preferential detection of clinically meaningful inconsistencies and generation of corresponding queries by leveraging both structured clinical data and unstructured text, including query messages. Using AI to improve query quality and prioritization may help address a long-standing CDM challenge: maintaining data quality while reducing site burden from large query volumes. Further evidence is expected as real-world validation and quantitative effectiveness evaluations accumulate.

4. Regulations and guidelines for AI-aided CDM

AI use in CDM raises regulatory challenges that must be addressed alongside technical implementation. The current regulatory landscape is fragmented. Relevant documents include established clinical research regulations, AI-specific frameworks with varying degrees of legal force, and medical AI guidelines that often do not directly address clinical research. This section first characterizes the regulatory landscape for AI-aided CDM and then examines expectations across three issues most directly relevant to CDM practice: the integrity of computerized processes, human oversight and accountability, and the privacy and security of clinical research data. Supplementary Table S2 (*https://www.globalhealthmedicine.com/site/supplementaldata.html?ID=123*) summarizes each instrument's coverage of these issues.

4.1. The regulatory landscape

At the international level, three documents are particularly relevant, but they differ in legal force, scope, and direct applicability to CDM. ICH E6 (R3) (January 2025) provides the GCP framework for CDM and becomes binding in each jurisdiction when incorporated into domestic regulations (*3*). Through its standalone data-governance chapter and media-neutral coverage of computerized systems, it extends established CDM expectations to AI-aided workflows. However, it does not specifically target AI. The FDA-EMA Guiding Principles of Good AI Practice in Drug Development (January 2026) provide the most direct regulatory discussion of AI in drug development, including human-AI interactions and traceable documentation, but they are non-binding (*34*). The EU AI Act (Regulation (EU) 2024/1689, in force August 2024) is a legally binding AI-specific framework and classifies AI systems used as medical devices as high-risk, with potential extraterritorial reach where outputs are used in the EU; however, it is not specific to CDM, and AI systems developed solely for research and development are explicitly exempt (*35*). Thus, no single international instrument provides binding requirements and directly targets AI use in CDM.

In Japan, alignment between ICH E6 (R3) and domestic ministerial ordinances under the Pharmaceuticals and Medical Devices Act is underway, but current domestic implementation extends only to E6 (R2) (*36*). In contrast, AI-specific domestic instruments remain voluntary. The MHLW Guidelines for the Utilization of Digital Data in AI Research and Development focus on personal information protection but explicitly exclude clinical research data (*37*). The AI Business Guidelines issued by the Ministry of Economy, Trade and Industry and the Ministry of Internal Affairs and Communications (Version 1.1, METI/MIC, 2025) establish four cross-sector pillars: safety, fairness, transparency, and privacy (*38*). Japan AI Safety Institute (J-AISI/IPA) provides cross-sector references for AI evaluation and data-quality management (*39,40*). The Healthcare AI Platform Collaborative Innovation Partnership (HAIP-CIP) Guidelines for the Use of Generative AI in the Medical and Healthcare Fields (2nd Edition) are the most practically relevant domestic reference for CDM, providing detailed risk classifications across eight use cases, including research data processing (*41*). Thus, in Japan, binding requirements for AI-aided CDM currently come from clinical research regulations for computerized systems, whereas AI-specific domestic documents remain voluntary supplementary guidance and are not tailored to CDM.

4.2. Process integrity: validation, audit trails, and documentation

For AI-aided CDM, the central question is whether the data-management process remains verifiable, reproducible, and auditable when AI is used to support data review, data cleaning, or query generation. Because such outputs may influence how clinical research data are corrected, documented, and judged to be reliable for analysis, AI-aided workflows should be treated as part of the computerized processes subject to CDM quality requirements.

This interpretation is grounded primarily in ICH E6 (R3), which requires risk-proportionate validation and audit-trail maintenance for computerized systems regardless of technology, placing AI-aided data cleaning and query generation within the same quality expectations as conventional EDC tools (*3*). AI-specific frameworks reinforce this direction. The FDA-EMA Guiding Principles emphasize traceable documentation of data provenance, processing steps, and human-AI interactions, while the EU AI Act adds automated logging and lifecycle quality-management requirements for high-risk AI systems (*34,35*).

In Japan, J-AISI/IPA documents provide complementary perspectives on data quality and AI evaluation, although they are not specific to CDM (*39, 40*).

Several CDM-specific issues therefore remain unresolved. Current guidance does not clearly define how prompts, model versions, and model updates should be documented for reproducibility; whether AI-aided data cleaning qualifies as a validated process under existing GCP definitions; or how AI-generated review results should be retained in the Trial Master File (*42-44*). These gaps are important because AI outputs may influence data review, query generation, and final judgments about whether data are sufficiently reliable for analysis.

4.3. Human oversight and accountability

The second practical question is how human oversight should be designed when AI contributes to CDM decisions. AI does not remove accountability from qualified personnel; rather, it changes where human judgment should be placed within the workflow. The key issue is therefore not simply whether a human is present, but whether human review occurs at points where AI outputs may affect data quality, regulatory documentation, or final data interpretation.

Human oversight is consistently treated as a core safeguard across the relevant documents. ICH E6 (R3) requires risk-proportionate human oversight of computerized systems and retains ultimate decision-making with human personnel in clinical trials (*3*). The FDA-EMA Guiding Principles further position HITL as a lifecycle design principle, while the EU AI Act requires high-risk AI systems to allow humans to disregard outputs or intervene in operation (*34,35*).

Japanese guidance follows the same direction: the AI Business Guidelines emphasize a human-centered principle, the MHLW Guidelines for the Utilization of Digital Data in AI Research and Development require human oversight throughout the data lifecycle, and the HAIP-CIP Guidelines require final review by qualified personnel in relevant use cases (*37,38,41*).

Together, these instruments establish HITL as a regulatory expectation rather than an optional design preference. However, they do not specify how much human verification is sufficient for AI outputs to be considered authoritative under GCP (*43,44*). For AI-aided CDM, human review should therefore be defined as an accountable decision point, not as a superficial confirmation of AI-generated results.

4.4. Privacy and secure handling of clinical research data

The third practical question is how patient-level clinical research data can be protected when AI tools are introduced into CDM workflows. AI-aided CDM may involve external vendors, cloud-based processing environments, model improvement processes, or secondary use of submitted data. These features make it necessary to clarify what data are transferred, where they are processed, whether they are retained, and whether they can be used for model training.

Existing clinical research regulations already require protection of participant data and confidentiality, and AI-related frameworks extend this concern to lifecycle governance, cybersecurity, and control over data use. ICH E6 (R3) provides the GCP basis for patient data protection, while the FDA-EMA Guiding Principles and the EU AI Act broaden the focus to AI lifecycle governance and cybersecurity requirements for higher-risk systems (*3,34,35*). In Japan, privacy and secure handling are grounded in the Act on the Protection of Personal Information (*45*). The MHLW Digital Data Guidelines, the AI Business Guidelines, J-AISI/IPA documents, and the HAIP-CIP Guidelines add practical considerations such as privacy-by-design, safety, evaluation of privacy risks, and restrictions on vendor use of submitted data for retraining (*37-41*).

Thus, privacy in AI-aided CDM should not be limited to anonymization alone. It should also include vendor governance, data retention policies, model-training restrictions, and cybersecurity controls.

5. Risks associated with AI utilization

5.1. Risks associated with AI output

Beyond regulatory compliance, four AI-related risks require particular attention: hallucination, bias, privacy and security, and explainability and transparency. These

risks are relevant across domains, but their implications are especially important in clinical research because data accuracy is directly linked to patient safety and regulatory decision-making.

Hallucination, or generation of plausible but factually incorrect outputs, may be the most consequential risk for CDM (*43,46*). Such errors are difficult to detect precisely because they appear credible. In CDM, AI may generate structured data from clinical narratives that appear valid but are wrong, or may misinterpret data relationships when generating queries. These errors differ from random transcription mistakes because they can evade conventional quality-control processes (*47*).

Bias is also a major concern. LLMs trained predominantly on data from specific demographic groups, clinical settings, or disease areas may perform worse when applied to underrepresented populations or rare diseases (*43,48*). These are precisely the settings in which data quality is most critical and often most difficult to ensure. Clinical research involving underrepresented populations or rare diseases therefore requires particularly careful evaluation for AI bias.

Privacy and security add further complexity. LLMs can memorize and reproduce personally identifiable information in their training data, including names and contact details (*49*). Processing patient-level clinical data with AI therefore requires robust anonymization, secure computational environments, and clear policies on data retention and model training (*44,50*).

Explainability and transparency also pose important challenges, particularly for GCP audit-trail requirements and model reproducibility. LLMs may produce different outputs for identical prompts because of model-version updates or infrastructure changes. Standards have not yet been established for recording and preserving the rationale behind AI-generated judgments in audit trails, leaving a gap against fundamental GCP expectations (*51*). Rigorous documentation of prompts and model versions can partially mitigate this issue, but it does not fully solve it.

Addressing these challenges requires multilayered safeguards. Technical safeguards include confidence scoring and anomaly detection (*50*). Procedural safeguards include HITL verification at critical decision points (see Section 6.1) (*51*). Organizational safeguards include governance frameworks, training programs, and audit mechanisms (*52*). Validation protocols designed specifically for AI-aided CDM are an important target for future standardization.

5.2. Risks associated with AI infrastructure

The risks of introducing AI into CDM are not limited to model-intrinsic issues. Supply chain and geopolitical risks must also be considered in real-world operations.

External AI vendors may create continuity risks related to policy, contractual, and geopolitical factors. The 2024 cyberattack on Change Healthcare Inc. exposed the vulnerability of critical infrastructure configurations that depend heavily on a single vendor (*53*). In March 2026, Anthropic was designated as a supply chain risk by the United States government (*54*), illustrating that service continuity can be threatened by non-technical factors. Such service interruptions are a realistic risk for CDM operations that rely heavily on specific vendors.

In Japan, external-vendor risks are being discussed through the lens of data sovereignty in government cloud environments, the MHLW Guidelines for the Secure Management of Medical Information Systems (*55*), and the AI Business Guidelines issued by METI and MIC (*38*). Similar considerations apply to AI introduction in CDM. Institutions should decide from the design stage which operations can be entrusted to which vendors, what alternatives exist if service is interrupted, and how data can be retrieved and migrated.

For CDM, these risks are relevant because interruption, migration failure, or loss of access to AI-supported tools could affect data review timelines, auditability, and continuity of trial operations.

6. Discussion and future perspectives

6.1. Human-in-the-loop: Human-AI collaborative workflows

Most cases reviewed here share the same structural principle: AI performs primary processing and detection, while human personnel retain final judgment. This design, known as human-in-the-loop (HITL) (*7*), may serve as the central implementation framework for integrating AI into CDM.

HITL refers to architectures in which human judgment is embedded in cycles of AI proposal, detection, and generation. Depending on the degree of automation, human involvement can range from "human in the loop", where humans make individual decisions, to "human on the loop", where humans supervise AI (*56*). The appropriate degree of involvement depends on task risk and accuracy requirements (*56*). In CDM, tasks directly linked to patient safety, regulatory decision-making, and final data approval require greater human involvement, whereas routine primary checks can reasonably be delegated to AI. From a quality-assurance perspective, AI-generated confidence scores and supporting rationales can help responsible personnel make decisions. Such risk-based allocation of human oversight is especially practical in AROs, where high data quality must be maintained with limited CDM human resources.

The regulatory basis for HITL is increasingly clear: ICH E6 (R3) requires risk-proportionate human oversight for AI as a computerized system (*3*), and the EU AI Act mandates human oversight for high-risk AI systems (*35*). These provisions provide not only qualitative justification

for human involvement, but also a regulatory basis for incorporating HITL.

HITL design also has challenges. These include human bottlenecks as workload increases, rubber-stamping during verification, and cognitive deskilling caused by over-reliance on AI. Bottlenecks can be reduced through threshold-based designs that automatically process high-confidence cases and reserve human review for cases that genuinely require it (*57*). Rubber-stamping can be mitigated by explainability designs that present AI outputs with supporting rationale (*58*). Cognitive deskilling requires regular training so that CDM personnel maintain an accurate understanding of AI limitations (*59*).

6.2. Advanced data cleaning and quality control

In the near term, integration of EDC and AI may enable real-time quality checking. EDC products vary in their logical-check system, and human personnel currently translate check rules and conditions into system-specific syntax. AI support for logical-check construction could therefore reduce workload. In multisite and multinational trials, adaptive quality management is also emerging: AI can detect site-to-site differences in data quality and language interpretation in real time and feed those findings into monitoring strategies. As shown by the international precedents discussed above (*26,27*), hybrid approaches combining AI with existing databases are already approaching practical implementation, and early adoption in Japanese ARO settings is expected.

In the medium to long term, CDM will need to accommodate diverse data sources, including unstructured data such as images, audio recordings, and free-text narratives. As DCTs spread and wearable devices and electronic patient diaries become new data sources, AI will play a larger role in data-source integration and quality assessment. Automatic generation of check rules is another promising direction. Today, protocol specifications are largely converted into check rules by hand; a workflow in which LLMs interpret protocols and generate draft check rules for CDM personnel to review and refine could reduce trial start-up workload. "Predictive CDM", in which models trained on historical trial data and query records proactively detect potential data issues in new trials, is also attracting interest. Temporal data variation is an important future target for AI. Current checks mainly compare values with reference ranges, for example, whether a laboratory value is within the normal range. Rapid or substantial changes over short intervals can therefore be missed even when values remain within normal limits. Such context-dependent anomalies may signal an adverse event prodrome or measurement errors, but they are difficult to detect with static, rule-based checks. AI-aided time-series analysis could detect clinically meaningful patterns by comparing trends within subjects, within sites, and across the trial as a whole. This would complement the current review processes that rely on CDM personnel's knowledge and experience and could move the field toward quality management that detects meaningful signals earlier. If realized, AI would move beyond retrospective data verification to become a proactive tool for anticipating and managing quality from the trial design stage.

6.3. Interdependent development of AI and data

This article has examined the potential of AI in CDM, including data collection and data cleaning in clinical research. Because AI systems learn from data, improving data quality is essential for improving AI performance. If Japanese medical data are underrepresented in global AI training datasets, AI performance for Japanese patients may be diminished (*60*). Preparing large volumes of high-quality data currently depends on human effort. If AI improves the quality of large datasets, those improved data can then support more capable AI, which can again be applied to improve data quality. AI and data are therefore mutually interdependent. This perspective is recognized internationally. A data-centric challenge led by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) proposed systematic methods for transforming raw clinical research data into AI-ready formats, showing that high-quality clinical datasets form the training foundation for next-generation AI and ML models (*61*). The J-AISI/IPA Data Quality Management Guidebook likewise emphasizes the "garbage in, garbage out" principle, arguing that data quality is the source of AI excellence and that high-quality outputs can fuel a virtuous cycle of further data-quality improvements (*39*). Realizing this interdependent development will require sustained use of AI to improve data quality at each stage of CDM.

7. Conclusions

This review examined AI-aided applications in CDM, with a focus on workflow support, human-in-the-loop implementation, and regulatory and risk-management considerations. Examples in data cleaning, medical coding, and query generation show that institutions are exploring and implementing AI-aided approaches at multiple stages of CDM workflows. In the near term, AI is most likely to contribute to repetitive, rule-based, or text-processing CDM tasks, such as data cleaning support, medical coding assistance, and query generation.

Important challenges remain. The regulatory framework for AI use under Good Clinical Practice continues to evolve, and institutions must remain vigilant about hallucination, bias, privacy and security, and explainability and transparency. Validation of AI-aided processes to a standard comparable to established computerized system validation is a prerequisite for

broader adoption. High-risk decisions and final data judgments should remain under qualified human oversight. In this context, HITL should be regarded as a safeguard that places human judgment at critical points where AI outputs may affect data quality or regulatory accountability. Combining AI-based technical support with human verification and organizational governance will be essential for managing AI-specific risks in CDM.

With policy support and technological maturation, Japan can contribute meaningfully to global discussions on AI-integrated CDM, particularly from the perspective of AROs that must maintain high data quality with limited CDM resources. The transition from predominantly manual, labor-intensive data management to human-AI collaborative workflows is no longer a question of whether, but of when and how.

Funding: None.

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. National Institutes of Health (NIH). Clinical research and trials. *https://orwh.od.nih.gov/clinical-research-and-trials* (accessed June 5, 2026).
2. Krishnankutty B, Bellary S, Kumar NB, Moodahadu LS. Data management in clinical research: An overview. Indian J Pharmacol. 2012; 44:168-172.
3. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guideline for good clinical practice E6(R3). *https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf* (accessed March 24, 2026).
4. Mahadik S, Sen P, Shah EJ. Harnessing digital health technologies and real-world evidence to enhance clinical research and patient outcomes. Digit Health. 2025; 11:20552076251362097.
5. Mitchell EJ, Goodman K, Wakefield N, *et al*. Clinical trial management: A profession in crisis? Trials. 2022; 23:357.
6. Musik S, Sasin-Kurowska J, Panczyk M. Bridging the past and future of clinical data management: The transformative impact of artificial intelligence. Open Access Journal of Clinical Trials. 2025; 17:15-33.
7. Cole S. What is human-in-the-loop? *https://www.ibm.com/think/topics/human-in-the-loop* (accessed April 1, 2026).
8. Olawade DB, Plabon SB, Ojo A, Ogunbona MA, Makanjuola BD, Olasilola OR. Human in the loop artificial intelligence in healthcare: Applications, outcomes, and implementation challenges. Int J Med Inform. 2026; 213:106362.
9. National Institutes of Health (NIH). Data quality management in clinical research. *https://oir.nih.gov/system/files/media/file/2021-08/data_quality_management-2015_05_15.pdf* (accessed March 24, 2026).
10. National Cancer Institute (NCI). NCI guidelines for auditing clinical trials for the NCI national clinical trials network (NCTN) program and NCI community oncology research program (NCORP) including NCORP research bases. *https://dctd.cancer.gov/research/ctep-trials/for-sites/nctn-auditing.pdf* (accessed March 24, 2026).
11. U.S. Food and Drug Administration (FDA). Guidance for Industry - Computerized systems used in clinical trials. *https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/fda-bioresearch-monitoring-information/guidance-industry-computerized-systems-used-clinical-trials* (accessed March 24, 2026).
12. eClinical Forum (eCF), the Society for Clinical Data Management (SCDM). Audit trail review: A key tool to ensure data integrity. *https://scdm.org/wp-content/uploads/2024/07/2021-eCF_SCDM-ATR-Industry-Position-Paper-Version-PR1-2.pdf* (accessed May 7, 2026).
13. Prasanna B, Kothapalli P, Vasanthan M. The role of quality assurance in clinical trials: Safeguarding data integrity and compliance. Cureus. 2024; 16:e67573.
14. All-Japan Conference of Deans of Medical Schools and Hospital Directors (AJMC). Summary of questionnaire results regarding the current status of support and management of clinical research. *https://ajmc.jp/wp/wp-content/uploads/2021/04/20151119_1_press.pdf* (accessed May 7, 2026). (in Japanese)
15. Yamaguchi T, Miyaji T, Hayashi Y, Suganami H. Clinical data management in Japan: Past, present, and future. J Soc Clin Data Manag. 2021; 1:1-6.
16. Japan Pharmaceutical Manufacturers Association. The ideal clinical data manager and skill set for the future-Focusing on Vendor Oversight, QMS, and DCT. *https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202305_DM-evol.html* (accessed March 27, 2026). (in Japanese)
17. Harper B, Smith Z, Snowdon J, DiCicco R, Hekmat R, Willis V, Weeraratne D, Getz K. Characterizing pain points in clinical data management and assessing the impact of mid-study updates. Ther Innov Regul Sci. 2021; 55:1006-1012.
18. Ministry of Education, Culture, Sports, Science and Technology, Ministry of Health, Labor and Welfare, Ministry of Economy, Trade and Industry. Ethical guidelines for medical and biological research involving human subjects. *https://www.mext.go.jp/content/20250325-mxt_life-000035486-01.pdf* (accessed April 6, 2026).
19. Ministry of Health, Labor and Welfare. Enforcement regulations on the clinical research act. *https://laws.e-gov.go.jp/law/430M60000100017* (accessed March 31, 2026). (in Japanese)
20. Ministry of Health, Labor and Welfare. Ministerial ordinance concerning standards for conducting clinical trials of pharmaceuticals. *https://laws.e-gov.go.jp/law/409M50000100028?occasion_date=20250401#Mp-Ch_2-Se_1* (accessed March 31, 2026). (in Japanese)
21. Fukuyama M. Society 5.0: Aiming for a new human-centered society. Japan spotlight. 2018; 27:47-50.
22. Morino E, Tokita D. AI in clinical trials: Current status, challenges, and future directions for emergency infectious disease clinical trials-Insights from the 2025 iCROWN Symposium. Glob Health Med. 2026; 8:70-71.
23. Takemura S. The challenge to develop and implement artificial intelligence (AI) technologies in health and medical care in Japan. Journal of the National Institute of Public Health. 2023; 72:2-13.
24. Hartley J, Jolevski F, Melo V, Moore B. The labor market effects of generative artificial intelligence. *https://doi.*

org/10.2139/ssrn.5136877 (accessed May 7, 2026).
25. Morgan Stanley. AI's impact accelerates. *https://www.morganstanley.com/insights/articles/ai-adoption-accelerates-survey-find* (accessed April 1, 2026).
26. Shi X, Prins C, Van Pottelbergh G, Mamouris P, Vaes B, De Moor B. An automated data cleaning method for Electronic Health Records by incorporating clinical knowledge. BMC Med Inform Decis Mak. 2021; 21:267.
27. Bönisch C, Schmidt C, Kesztyüs D, Kestler HA, Kesztyüs T. Proposal for using AI to assess clinical data integrity and generate metadata: Algorithm development and validation. JMIR Med Inform. 2025; 13:e60204.
28. Meldau EL, Bista S, Rofors E, Gattepaille LM. Automated drug coding using artificial intelligence: An evaluation of WHODrug Koda on adverse event reports. Drug Saf. 2022; 45:549-561.
29. Utsumi K. Let's consider using AI in data management. *https://www.aro.med.kyushu-u.ac.jp/wp-content/uploads/2024/10/seminar20241018_material-2.pdf* (accessed March 17, 2026). (in Japanese)
30. Japan Pharmaceutical Manufacturers Association. Model prototype for exploring the use of AI in data management operations. *https://www.jpma.or.jp/information/evaluation/results/allotment/q83i5d0000000mtx-att/DS_202406_AI_Prototype.pdf* (accessed March 30, 2026). (in Japanese)
31. Japan Pharmaceutical Manufacturers Association. Application of artificial intelligence in data management - AI to start from now on. *https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202305_2022TF1_1_AI_DM.html* (accessed March 23, 2026). (in Japanese)
32. Jane R. Autocoding adverse events to MedDRA – time to throw out the manual. *https://www.iqvia.com/blogs/2022/03/autocoding-adverse-events-to-meddra-time-to-throw-out-the-manual* (accessed April 1, 2026).
33. Stokman PG, Ensign L, Langeneckhardt D, Mörsch M, Nuyens K, Hochgräber G, Cassan V, Beineke P, Kwock R, Voortman A, Vogelgesang S, Boussetta S, Bitzer B. Risk-based quality management in CDM: An inquiry into the value of generalized query-based data cleaning. Journal of the Society for Clinical Data Management. 2021; 1.
34. U.S. Food & Drug Administration (FDA), European Medicines Agency. Guiding principles of good AI practice in drug development. *https://www.fda.gov/media/189581/download* (accessed March 26, 2026).
35. European Union. Regulation (EU) 2024/1689 of the European parliament and of the council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending regulations. *https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng* (accessed May 7, 2026).
36. Ministry of Health, Labor and Welfare. Regarding the revision of the "Guidance on the ministerial ordinance on standards for conducting clinical trials of pharmaceuticals". *https://www.mhlw.go.jp/web/t_doc?dataId=00tc8160&dataType=1&pageNo=1* (accessed April 7, 2026). (in Japanese)
37. Ministry of Health, Labor and Welfare. Guidelines for the utilization of digital data in AI research and development, *etc*. *https://www.mhlw.go.jp/content/001310044.pdf* (accessed April 7, 2026). (in Japanese)
38. Ministry of Internal Affairs and Communications (MIC), Ministry of Economy, Trade and Industry. AI business guidelines (Version 1.1). *https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20250328_1.pdf* (accessed April 7, 2026). (in Japanese)
39. Japan AI Safety Institute (J-AISI). Data quality management guidebook. *https://aisi.go.jp/assets/pdf/250331_Data_quality_management_guidebook.pdf* (accessed March 19, 2026).
40. Japan AI Safety Institute (J-AISI). Guide to evaluation perspectives on AI safety, Version 1.10. *https://aisi.go.jp/assets/pdf/ai_safety_eval_v1.10_en.pdf* (accessed March 19, 2026).
41. Healthcare AI Platform Collaborative Innovation Partnership (HAIP-CIP). Guidelines for the use of generative AI in the medical and healthcare fields (2nd Edition). *https://haip-cip.org/assets/documents/nr_20241002_02.pdf* (accessed May 7, 2026). (in Japanese)
42. Askin S, Burkhalter D, Calado G, El Dakrouni S. Artificial intelligence applied to clinical trials: Opportunities and challenges. Health Technol (Berl). 2023; 13:203-213.
43. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med. 2023; 6:120.
44. Olawade DB, Fidelis SC, Marinze S, Egbon E, Osunmakinde A, Osborne A. Artificial intelligence in clinical trials: A comprehensive review of opportunities, challenges, and future directions. Int J Med Inform. 2026; 206:106141.
45. Government of Japan. Act on the protection of personal information. *https://www.japaneselawtranslation.go.jp/en/laws/view/4241* (accessed June 8, 2026).
46. Hatem R, Simmons B, Thornton JE. A call to address AI "hallucinations" and how healthcare professionals can mitigate their risks. Cureus. 2023; 15:e44720.
47. Asgari E, Montaña-Brown N, Dubois M, Khalil S, Balloch J, Yeung JA, Pimenta D. A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. NPJ Digit Med. 2025; 8:274.
48. Badani A, de Moraes FY, Vollmuth P, Chung C, Mansouri A. AI and innovation in clinical trials. NPJ Digit Med. 2025; 8:683.
49. Carlini N, Tramer F, Wallace E, Jagielski M, Herbert-Voss A, Lee K, Roberts A, Brown T, Song D, Erlingsson U, Oprea A, Raffel C. Extracting training data from large language models. In: 30th USENIX Security Symposium (USENIX Security 21) Association, 2021; pp. 2633-2650.
50. Chen Y, Esmaeilzadeh P. Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges. Journal of Medical Internet Research. 2024; 26:e53008.
51. Zhou J, Li H, Chen S, Chen Z, Han Z, Gao X. Large language models in biomedicine and healthcare. npj Artificial Intelligence. 2025; 1:44.
52. Cao W, Zhang Q, Liu J, Liu S. From Agents to Governance: Essential AI Skills for Clinicians in the Large Language Model Era. J Med Internet Res. 2026; 28:e86550.
53. Arthur F, Daniel S. The cyberattack on change healthcare: Lessons for financial stability. *https://www.financialresearch.gov/briefs/files/OFRBrief-24-05-change-healthcare-cyberattack.pdf* (accessed May 7, 2026).
54. Anthropic. Where things stand with the department of war. *https://www.anthropic.com/news/where-stand-department-war?_bhlid=bfe019c8865f01ef699354be3d20cd037b6c8ff5&utm_source* (accessed March 27, 2026).
55. Ministry of Health, Labour and Welfare. Guidelines for the secure management of medical information systems, Version 6.0. *https://www.mhlw.go.jp/stf/*

shingi/0000516275_00006.html (accessed April 7, 2026). (in Japanese)
56. Michael BC, Rick W. Is Human-on-the-loop the best answer for rapid relevant responses? *https://www.japcc.org/essays/is-human-on-the-loop-the-best-answer-for-rapid-relevant-responses/* (accessed May 7, 2026).
57. Chen M, Wang Y, Wang Q, Shi J, Wang H, Ye Z, Xue P, Qiao Y. Impact of human and artificial intelligence collaboration on workload reduction in medical image interpretation. NPJ Digit Med. 2024; 7:349.
58. Choudhury A, Chaudhry Z. Large language models and user trust: Consequence of self-referential learning loop and the deskilling of health care professionals. J Med Internet Res. 2024; 26:e56764.
59. Goddard K, Roudsari A, Wyatt JC. Automation bias: A systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc. 2012; 19:121-127.
60. Ueda D, Walston S, Takita H, Mitsuyama Y, Miki Y. The critical need for an open medical imaging database in Japan: Implications for global health and AI development. Jpn J Radiol. 2025; 43:537-541.
61. Domagalski MJ, Lu Y, Pilozzi A, Williamson A, Chilappagari P, Luker E, Shelley CD, Dabic A, Keller MA, Rodriguez RM, Lawlor S, Thangudu RR. Preparing clinical research data for artificial intelligence readiness: Insights from the National Institute of Diabetes and Digestive and Kidney Diseases data centric challenge. J Am Med Inform Assoc. 2025; 32:1609-1616.

Received May 7, 2026; Revised June 9, 2026; Accepted June 15, 2026.

Released online in J-STAGE as advance publication June 19, 2026.

**Address correspondence to*:
Hajime Ohyanagi, Department of Joint Center for Researchers, Associates and Clinicians (JCRAC), Center for Clinical Sciences, Japan Institute for Health Security, 1-21-1 Toyama Shinjuku-ku, Tokyo 162-8655, Japan.
E-mail: ohyanagi.h@jihs.go.jp