The adoption of risk-based CDM approaches

Version 1 (July 2022)

Authors

- Catherine Célingant, Executive Director, Data Monitoring & Management – Oncology, Pfizer
- Lynne Cesario, Global Risk Based Monitoring Program Lead, Pfizer
- Patrick Nadolny, Global head, Clinical Data Management, Sanofi

Reviewers and Contributors

- Sanjay Bhardwaj, Head of Clinical Technology Strategy & Operations, Abbvie
- Joanna Florek-Marwitz, Head Risk Management and Data Quality, UCB
- Jérémy Nadolny, Sr Business Associate, Trial Operations
- Peter Stokman, Business Lead Data Review & Visualization, Bayer
- Demetris Zambas, Global Head, Data Monitoring & Management, Pfizer

The Society for Clinical Data Management (SCDM) would like to acknowledge the many volunteers who have contributed to the development of the SCDM reflection papers used as the primary basis for this topic brief.

Methodology

The SCDM Innovation Committee seeks to provide Thought-Leadership to our industry and support the SCDM vision of "*leading innovative clinical data science to advance global health research and development*". To that end, the SCDM Innovation Committee strives to demystify Clinical Data Science (CDS) and support the development of all Clinical Data Management (CDM) professionals, from subject matter experts (SMEs) working on clinical studies to CDM leaders setting the direction of their organizations.

The SCDM Innovation Committee is publishing topic briefs intended to serve as orientation guides on specific areas which are contributing directly or indirectly to the evolution of CDM into CDS. The content of those topic briefs is primarily an extract from the previously published SCDM Reflection Papers[1,2,3] which collectively provide a cohesive and comprehensive overview of CDS from the point of view of industry leaders. Due to the recent emergence of the CDS discipline and the absence of a comprehensive literature base regarding CDS within the Drug Development industry, this content was gathered from industry leaders through a consensus-based methodology. As CDS matures and technology evolves, we anticipate that literature on this topic will blossom.

Introduction

Regulators have issued guidance documents advocating for the use of risk-based and fit for purpose approaches starting over a decade ago. The Good Clinical Practice (GCP) guidance (i.e., ICH E6[4]) was updated in 2016 to reinforce the direction and introduce Quality Tolerance Limits (QTLs). More recently, the long-awaited guidance on General Considerations for Clinical Studies (ICH E8[5]) is further clarifying the expectations by focusing on two foundational risk-based principles: Quality by Design (QbD) and Critical to Quality (CtQ) Factors. These were previously mentioned in our third reflection paper[3].

Our industry has already successfully implemented risk-based approaches in the site monitoring and system validation spaces for many years. As a result, our traditionally risk-averse industry has become more comfortable with strategies that match efforts with the risks to patient safety and data quality. Leveraging this momentum, we now have a legitimate opportunity to expand the use of Risk-Based Quality Management (RBQM) principles to transform all study design and execution aspects within CDM.

Topic brief

To seize this opportunity, we must redefine our CDM processes and roles to control risks to activities essential to ensuring human subject protection and the reliability of trial results[4]. Learning from the evolution of traditional Risk-Based (site) Monitoring (RBM) and the initial resistance to adopt it, we must realize that this is a fundamental company culture and role change which must be endorsed from the top. So, first and foremost, we need to clearly explain what adopting risk-based approaches means. **It does not mean taking risk neither promoting risk. It does not mean asking other functions to increase their data oversight to perform activities CDM is no longer planning to perform. It means de-risking clinical trials** by proactively redirecting focus on what matters most and avoiding disproportionate focus on activities which have no impact on patient's safety or marginal impact on the reliability of the trial results. ICH E8 is clear, "**Inflexible, one size fits all approaches should be discouraged".** Standardized operating procedures are necessary and beneficial for conducting good quality clinical studies, but study specific strategies and actions are also needed to effectively and efficiently support quality in a study"[5]. This means moving away from a one-size-fits all process based on inflexible operating standards and 100% Quality Control (QC). Moving forward, quality must be infused at the study design stage to proactively prevent potential risks to arise as much as possible and implementing strong monitoring during the study through pre-defined Key Risk Indicators (KRIs) and QTLs.

Additionally, it has been shown by TransCelerate that on average only 3.7% of the data entered in the eCRF is changed after initial entry with only 1.1% related to Source Data Verification (SDV) and 2.6% attributed to other data review processes[6]. A more recent article corroborated this finding by demonstrating that only 3.9% of data entered in EDC was ever queried leading to changes in only 1.7% of all entered data, critical or not[7]. It does not mean that those data validation processes are unnecessary or redundant. It indicates that the outcome does not proportionally reward the significant effort that many CDM organization place on data review and data validation strategies especially for non-critical data.

As a result, some industry leaders are questioning the efficiency and true impact of traditional data review and CDM processes. This perception was re-enforced by the release of regulations such as ICH E6 (R2)[4] and ICH E8 (R1)[5] recommending different, risk-based approaches to study monitoring and oversight.

But before diving into risk-based CDM approaches, let's first clarify what Data Quality is.

What is Data Quality?

Surprisingly, Data Quality is not universally understood within CDM (& beyond) and is often confused with Data Integrity. As stated in the 2018 MHRA guidance on 'GXP Data Integrity', data integrity is **not** data quality as "the controls required for integrity do not necessarily guarantee the quality of the data generated"[8].

It is somewhat easy to understand and demonstrate Data Integrity as many regulations such as 21 CFR Part 11 and the MHRA guidance have explicitly defined data integrity. Data integrity is often associated with ALCOA which is defined as **A**ttributable, **L**egible, **C**ontemporaneous, **O**riginal and **A**ccurate. But is ALCOA enough to reach data quality? The answer is unambiguously **NOT**.

Now, let's use an extreme case to differentiate Data Integrity from Data Quality. Let's assume that some data entered in the eCRF are:

- **A**ttributable as being entered by the site personnel and confirmed by the audit trail,
- **L**egible in the site source,
- **C**ontemporaneously collected at the time when the activity was performed at the site,
- **O**riginal to the source as confirmed by SDV and
- **A**ccurate (i.e., free from errors, complete and within ranges).

The data from the example above meets the ALCOA requirements demonstrating the core attributes of data integrity. But theoretically, it could have been collected from non-calibrated instruments or by non-medically qualified personnel. Data could have also been collected from sites not adhering to expected good research practices, such as propagating the same data from visit to visit (e.g., copying same vital signs data from one visit to another instead of collecting vital signs at each visit). These scenarios are rare but have happened. The corresponding data would clearly not be considered reliable nor quality data. They could lead to the exclusion of the site data in the Clinical Study Report, endangering the statistical power of the population and therefore negatively impact the study outcome. So, meeting the key criteria of Data Integrity (e.g., ALCOA) is not enough to ensure Data Quality.

For ICH E8, "quality of a clinical study […] is considered as fitness for purpose. The purpose of a clinical study is to generate reliable information to answer the research questions and support decision making while protecting study participants. The quality of the information generated should therefore be sufficient to support good decision making"[5].

In 2016, ICH E6 (R2) focused on activities essential to ensuring the reliability of trial results with capabilities to distinguish between reliable and potentially unreliable data. In 2018, MHRA defined data quality as "the assurance that data produced is exactly what was intended to be produced and fit for its intended purpose. This incorporates ALCOA"[8]. All of those suggests that Data Quality is reached when data is enabling the right decision-making (i.e., fit for purpose).

As a general guidance, we can define Data Quality vs. Data Integrity as follow:

> **Data Integrity** means that the **Data is managed the right way**
>
> **Data Quality** means that the **Data is credible and reliable**

Many CDM organizations have historically strived to achieve data integrity as a primary outcome. Processes have been designed to ensure that 100% of the expected data has been collected, all missing data identified, all inconsistent data queried, and all external data reconciled. While data integrity is a mandatory attribute of data quality, reaching data credibility and reliability through-fit-for-purpose clinical data reviews must become our priority.

To set the direction and help in distinguishing between reliable and potentially unreliable data, ICH E6 (R2) is advocating reviewing data differently in order to identify and evaluate:

- Data outliers and unexpected lack of variability,
- Data trends such as the range, consistency, and data variability within and across sites,
- Systematic or significant errors in data collection and reporting at a site or across sites,
- Potential data manipulation or data integrity problems

ICH E6 (R2) scope goes beyond traditional data cleaning processes and requires reviews combining patient data from all sources including but not limited to safety data, protocol deviations, audit trails, metadata and operational data. At the end of the day, the data need to reliably support the evaluation of the objectives set in the protocol. To achieve this, we must start on "the right foot" and ensure Quality by Design (QbD).

End-to-End Risk-Based Study Set-Up and Execution

The adoption of risk-based CDM approaches has a deep impact on our traditional CDM processes as shown in the process flow below. In this example, risk-based processes steps depicted in green have been added to the traditional CDM steps depicted in blue to illustrate the end-to-end nature of risk-based approaches.
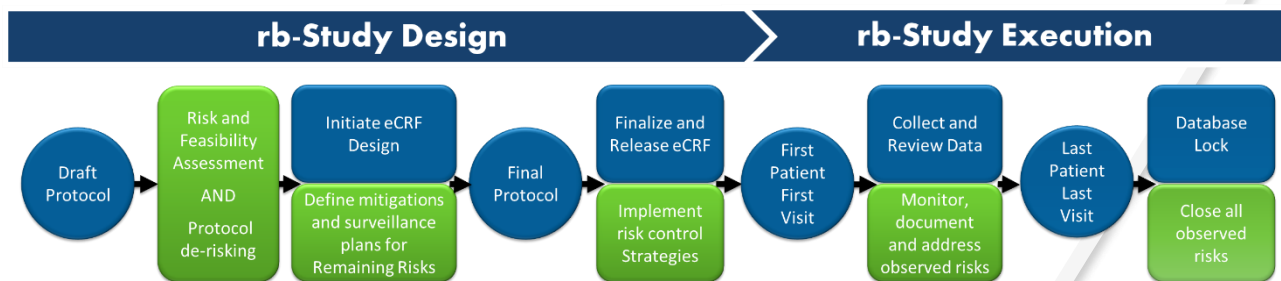
Study Quality by Design

ICH E6 (R2) dedicated a section on quality management (section 5.0) focusing on risks lifecycle management. It covers the quality controls used from the identification to the reporting of the risks. But rather than only controlling the risks by implementing mitigation and monitoring strategies, we should first and foremost use QbD to proactively avoid them.

ICH E8 (R1), states that "QbD in clinical research sets out to ensure that the quality of a study is driven proactively by designing quality into the study protocol and processes. This involves the use of a prospective, multidisciplinary approach to promote the quality of protocol and process design in a manner proportionate to the risks involved, and clear communication of how this will be achieved." [5]

QbD stands on the assumptions that quality should be planned proactively and not expected to result from retrospective QCs. First, QbD must rely on the appropriate foundation including the company "culture that values and rewards critical thinking and open, proactive dialogue about what is critical to quality for a particular study" [5], policies, systems, processes and people.

Next, QbD also relies on optimal protocol design with appropriate level of complexity. Protocol design is a critical step with an outsized influence on the ultimate success or failure of the trial. Consistent with the QbD principles, risks and mitigations should be identified prior to the protocol finalization. If possible, the protocol should be adjusted to eliminate unnecessary scientific and operational complexities and prevent their negative impact on data availability and reliability.

If de-risking the study protocol is not possible, the study team must implement timely mitigation strategies to manage risks and prevent them from materializing in form of issues. Of particular importance as part of this proactive risk mitigation process is the need to ensure that "all aspects of the trial are operationally feasible" and "avoid unnecessary complexity, procedures, and data collection" [4]. The Clinical Data Scientist must steer the study team to only perform procedures that are essential to the outcome of the clinical trial. Ultimately, the study team must proactively confirm that the planned protocol is operationally acceptable for all sites and countries involved, as increased protocol complexity can lead to poorer study data quality.

As a key member of the study team, the Clinical Data Scientist must drive the conversation and proactively manage risks that matter most, mainly related to Critical to Quality (CtQ) factors associated to critical data and processes. The Clinical Trials Transformation Initiative (CTTI)[9] introduced the CtQ factors in 2015 and organized them around the six major categories below with strong emphasis on protocol design.

CtQ Categories	CtQ factors
Protocol Design	Eligibility Criteria Randomization Masking Types of Controls Data Quantity Endpoints Procedures Supporting Study Endpoints and Data Integrity Investigational Product (IP) Handling and Administration
Feasibility	Study and Site Feasibility Accrual (i.e., Enrollment Strategy)
Patient Safety	Informed Consent Withdrawal Criteria and Trial Participant Retention Signal Detection Safety Reporting Data Monitoring Committee (DMC) / Stopping Rules (if applicable)
Study Conduct	Training Data Recording and Reporting Data Monitoring and Management Statistical Analysis
Study Reporting	Dissemination of Study Results
Third-party Engagement	Delegation of Sponsor Responsibilities and Collaborations

Fig 1 CTTI CtQ categories and factors

Subsequently, they became a central theme in the ICH E8 guidance which defines these factors as "attributes of a study whose integrity is fundamental to the protection of study participants, the reliability and interpretability of the study results, and the decisions made based on the study results" [5]. They "are considered to be critical because, if their integrity were to be undermined by errors of design or conduct, the reliability or ethics of decision-making based on the results of the study would also be undermined" [5].

Lastly, the ICH E8 guidance is reemphasizing the need to ensure the scientific feasibility of the protocol and fit-for-purpose processes considering the diversity of data sources. "Operational criteria are also important, such as ensuring a clear understanding of the feasibility of the study, selection of suitable investigator sites, quality of specialized analytical and testing facilities and procedures, and processes that ensure data integrity"[5]. Last but not least, "The critical to quality factors should be clear and should not be cluttered with minor issues" [5].

Considering this frame of reference and to ensure QbD, the "identification of quality factors critical to ensuring the protection of study participants, the integrity of the data and the reliability of results" [5] is a critical responsibility of the Clinical Data Scientist

Study Risk Assessment

Additionally, the Clinical Data Scientist must also contribute, if not lead, the risk assessment and the mitigation of some of the risks associated with CtQ factors. The Risk Assessment must go beyond the typical risks impacting study timelines, deviation from data standards or identifying critical data and processes. Risk Management must start with risk prevention (i.e., Quality by Design) by identifying threats prior to the first patient being enrolled into the study which have the potential of leading to errors that could negatively impact the credibility and reliability of the trial results.

There are many risk areas associated with the CtQ factors, including but are not limited to the:

- Complexity of protocol designs such as umbrella, basket, platform and adaptive
- Vulnerability of the patient population (e.g., elderly, pediatric)
- Complexity of enrollment procedures (e.g., consent, eligibility, stratification and randomization)
- Deviations from standard of care
- Characteristics of the participating countries (e.g., Standard of care, customs, dialects)
- Planned rate and distribution of enrollment
- Number, profile and experience of the study sites and countries
- Nature of the protocol-required procedures, with specific emphasis on the burden they may place on patients and sites
- Organization of the trial (e.g., site-centric vs. decentralized) with telemedicine and home nursing
- Planned technologies used to collect data, including when patients bring their own device (BYOD)
- Complexity of the data flow, including variety of the data sources
- Oversight of the capture and modification of the eSource data owned by the sites
- Number and experience of the data and operational vendors
- Any other study execution activities which may lead to data errors that could negatively impact the credibility and reliability of the trial results

While many risks would be evaluated and accounted for in the overall risk mitigation plans from the multidisciplinary study team, some risks such as the complexity of the data, data flows, third party vendors and planned technologies used for data collection would be the primary focus of CDS.

To foster study QbD, Clinical Data Scientists and study teams must understand the advanced concepts introduced in Part 1[1], 2[2] and 3[3] of the SCDM Reflection Papers. This may require changes in both the composition of the protocol review team as well as the process for developing protocols. To ease this

evolution requiring deep critical thinking, CDS organizations could pre-define guidance (i.e., decision trees) for the determination of risks mitigations associated of with standard CtQ factors by leveraging historical information on process and data issues.

As a one-off example in figure 2, the Clinical Data Scientists may need to plan for a robust risk monitoring strategy leveraging analytics tools including KRIs and QTLs to proactively identify and quantifying trends resulting from known risks such as enrollment speed. Those data monitoring strategies could be documented in the sponsor's Integrated Quality Management Plan (IQMP) or in the Data Management Plan (DMP) and connected with other functional plans as appropriate, since the overall study risk management is a cross-functional responsibility.

	Scenarios for enrollment speed Very fast or very slow enrollment impacting data review strategies	
Potential Data Quality Risks	**Fast Enrollment (Scenario #1)** Unable to match the speed/volume of data collection with speed / frequency of data reviews	**Slow Enrollment (Scenario #2)** Higher risks of sites closure prior to DB Lock impacting data review and reconciliation strategies
Mitigations	• Leverage technologies (e.g., EDC, IRT, eCOA) to identify error as fast and as close as possible from the data source • Focus review on critical data points impacting eligibility • Prioritize the readiness of data review tools that matter most (Incl. KRIs & QTLs)	• Trend Screen Failure metrics • Monitor site closures • Fine-tune data review frequency considering early sites closure risks

Fig 2 Example of CDS risk mitigation guidance

Moreover, those mitigation measures should be pre-identified, prioritized and implemented to the extent possible prior to study start to allow for optimum quality controls during study conduct.

Risk-based study execution (i.e., the Quality Control stage)

With the adoption of risk-based principles, CDM is at the center of fit for purpose study execution. As examples, Clinical Data Scientists will need to:

- **Adjust data validation and review strategies to the risk assessment:** Target data monitoring on critical risks, data and processes. Evaluate the risk of eliminating non-critical data validation if other safety nets exist (e.g., aggregated data trending or statistical monitoring of non-critical data)

- **Define targeted CDM Specific QTL to demonstrate reliability of trials results (**e.g., rate of missing data for primary end point)

- **Identify systematic or process driven data issues** including those stemming from trial design and study conduct factors such as rate of enrollment, technologies used, etc. The key will be to efficiently and reliably monitor such risks through the holistic review of **all** clinical and operational data (i.e., finding data patterns and anomalies across studies, countries, sites, and patients).

- **Identify risk associated with rather complex data flows** resulting from disparate data sources.

- Understand **risk-based approaches** when conducting clinical studies in a **patient centric and decentralized environment**.

Thus, the Clinical Data Scientist should continuously monitor the data-related risks during study conduct by:

- Monitoring the risks identified during the study design phase using the defined KRIs and QTLs
- Performing holistic data reviews including review of the various data audit trails
- Monitoring for the possible emergence of any new risks, including but not limited to:
  - Risk to database availability which could delay study start
  - Risk to study timelines and data flow delays which could negatively impact the availability of study results for safety reviews, the potential submission and product approval
  - Impact of protocol amendments
- Assessing the effectiveness of the implemented risk mitigations
- Adjusting or augmenting risk mitigations as necessary

Risk Management and Mitigation

The identification of risks by assessing outliers and atypical data patterns frequently relies on KRIs, QTLs and tools based on statistical methodology. This means the resolution of "issues" is no longer as simple as using SDV and queries from edit checks to verify the accuracy of the data. Instead, the study team must understand the signal generated and apply critical reasoning to analyze the likely root causes. Once a signal is determined to be an issue, the underlying process or data issue needs to be addressed. Lastly, to close the loop, teams must follow up to make sure the issue has been fully resolved.

Below are some examples of signals that can be found with the potential responses made by teams.

1. All patients at a site in Puerto Rico are Hispanic: An atypical proportion of one ethnicity at the site may be statistically outlying compared to other study sites outside South America but not unexpected in this case. The team does not need to act on the signal but should follow up until the site has finished recruiting to see if the pattern evolves.

2. Many patients at a site have the same respiratory rate: Rather than questioning if the value was correctly entered into the source document, teams should think about how this lack of variability occurred. It is possible, but highly unlikely, that many patients at a site have the same respiratory rate. It is more likely that something was wrong with how the measurements were taken and/or recorded. Thus, the process for collecting and recording the rate should be reviewed. The importance of accurate data collection and recording reiterated to the site personnel. Since the existing data is not going to change, any issue with the process in taking measurements should be addressed, fixed, and monitored moving forward

3. Patients on an oncology trial have either no or a very low number of adverse events (AEs): This is statistically unlikely. The study team should ensure the site personnel understand how to collect AEs, and possibly use source data review (SDR) to check for unreported AEs. The site personnel may need retraining, and the study team must follow up to make sure the situation is resolved. Current data might not change, but the process must be fixed and then tracked for ongoing correctness.

To address the examples above, the Clinical Data Scientists and study team must dig deep into the data to understand the root cause of the issues. They need to perform detailed analysis and data review findings to resolve them. Occasionally, the team will need to go through multiple iterations of analysis and follow-up to fully understand the root cause. This requires a focus on details and strong communication skills as most findings will not result in queries, but rather in addressing systematic process issues and site behaviors.

*Continuous CDS process improvement*

Clinical Data Scientists and study teams must leverage the lessons learned during study execution and frontload the activities to adapt the processes. They must prevent further reoccurrence of the same issues in the study and in future trials. For systematic issues, the mitigation of a specific risk may involve a corrective and preventive action (CAPA).

Although the CAPA process is usually driven centrally by the quality organization, Clinical Data Scientists should be familiar with it. They, as SMEs in the risk management lifecycle, should also be comfortable contributing to the process through the characterization of the risk and suggesting pragmatic and robust remediations and preventive actions.

Conclusion

In summary and as stated by MHRA, *"it is recommended that the data validation activities (…) be focused on the data that is critical to the reliability of the trial results as identified by the risk assessment rather than excessive resource spent on raising data queries whose resolution makes little or no impact on the quality of the trial, the safety of the participants and reliability of the results. This is similar to the approach taken for proportionate source data verification (SDV)"* [10].

CDM must evolve substantially from the legacy QC-based strategies if it is to support QbD by 1) identifying CtQ Factors, 2) de-risking the protocol and 3) performing a CDM specific risk assessment. Only then CDM can implement safe and effective risk-based study execution strategies paired with robust continuous process improvements to deliver quality data sufficient to support good decision making. This will have a dramatic impact on the CDS roles by moving from catching mistakes to identifying problems that may jeopardize the trial. Overall, the end-to-end management of the operational and scientific risks shown below must be embedded throughout the entire CDM Framework, with connection to other functions involved in the process when necessary.
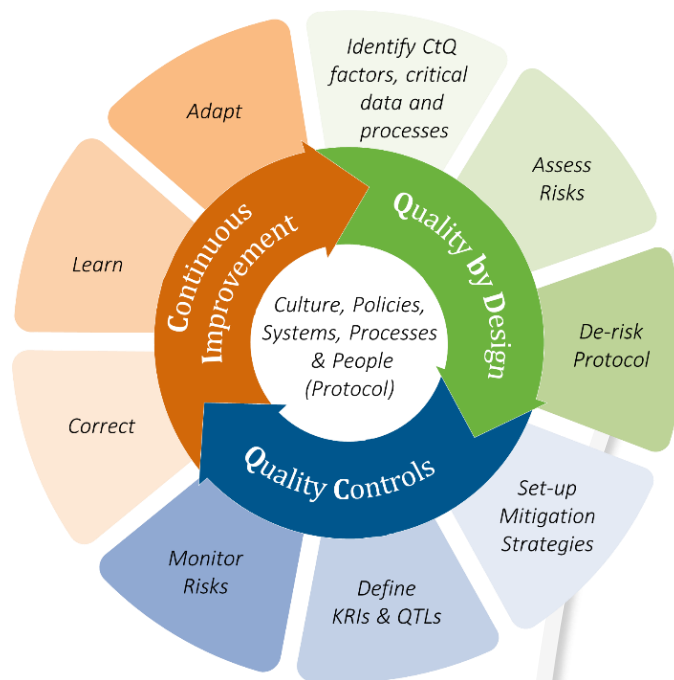
References

1   SCDM, June 2019, The Evolution of CDM to Clinical Data Science – A Reflection Paper on the impact of the Clinical Research industry trends on Clinical Data Management. Available at https://scdm.org/wp-content/uploads/2019/09/SCDM-Reflection-Paper-Evolution-to-Clinical-to-Data-Science.pdf

2   SCDM, March 2020, The Evolution of CDM to Clinical Data Science (Part 2: The technology enablers) – A Reflection Paper on how technology will enable the evolution of Clinical Data Management to Clinical Data Science. Available at https://scdm.org/wp-content/uploads/2020/03/SCDM-Reflection-Paper-Evolution-to-Clinical-to-Data-Science-Part-2.pdf

3   SCDM, August 2021, The Evolution of CDM to Clinical Data Science (Part 3: Evolution of the CDM role) – A Reflection Paper on the evolution of CDM skillsets and competencies. Available at https://scdm.org/wp-content/uploads/2020/08/SCDM-Reflection-Paper-CDM-Role-Evolution-Part-3-.pdf

4   ICH, Integrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice E6(R2). Available at https://database.ich.org/sites/default/files/E6_R2_Addendum.pdf

5   ICH, ICH E8, General Considerations for Clinical Trials Draft Guidelines, Available at https://www.ema.europa.eu/documents/regulatory-procedural-guideline/ich-guideline-e8-r1-general-considerations-clinical-studies_en.pdf

6   TransCelerate, Sep 2014, Evaluating Source Data Verification as quality Control Measure in Clinical Trials. Available at https://journals.sagepub.com/doi/pdf/10.1177/2168479014554400

7   Stokman and al, 2021, 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, available at a https://www.jscdm.org/article/20/galley/40/download/

8   MHRA, March 2018, 'GXP' Data Integrity Guidance and Definition, Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/687246/MHRA_GxP_data_integrity_guide_March_edited_Final.pdf

9   CTTI, May 2015, Quality by Design (QbD) Project, Critical to Quality Factors Principles document, Available at https://ctti-clinicaltrials.org/wp-content/uploads/2021/07/CTTI_QbD_Workshop_Principles_Document.pdf

10 MHRA, January 2022, Oversight and monitoring, available at a https://www.gov.uk/government/publications/oversight-and-monitoring-of-investigational-medical-product-trials/oversight-and-monitoring-activities

Main abbreviations

AI	Artificial Intelligence
ATR	Audit Trail review
CAPA	Corrective Action and Preventive Action
CtQ	Critical to Quality
CTTI	Clinical Trials Transformation Initiative
CDM	Clinical Data Management
CDS	Clinical Data Science
DCT	Decentralized Clinical Trials
eCRF	Electronic Case Report Form
EDC	Electronic Data Capture
IQMP	Integrated Quality Management Plan
MHRA	Medicines and Healthcare products Regulatory Agency
QbD	Quality by Design
QC	Quality Control
RBM	Risk Based Monitoring
RBQM	Risk-Based Quality Management
SCDM	Society for Clinical Data Management
SDR	Source Data Review
SDV	Source Data Verification
SME	Subject Matter Expert