



Module 3.4a Transcript

Interpretation and Application of ICH E6(R3):

GOOD CLINICAL PRACTICE (GCP)

Welcome to the Interpretation and Application of ICH E6(R3): Good Clinical Practice.

Module 3 of this training discusses Data Governance in clinical trials.

Module 3 contains 4 sub-modules, which should be taken sequentially. Module 3.4, Data Lifecycle Management, is the final part of this data governance module.

This sub-module is split into two parts: Module 3.4a and Module 3.4b.

Key questions you will be able to answer after completing this module include:

- What are the general requirements for managing data across each step of the data lifecycle?
- What responsibilities do investigators and sponsors have throughout the data lifecycle? And,
- How can investigators and sponsors tailor their management and oversight throughout the data lifecycle based on risks that may impact critical to quality factors?

Section 4.2 of the E6(R3) guideline outlines the steps of the clinical trial data lifecycle. The data lifecycle is presented here as a linear process, reflecting the stepwise progression of data captured in a clinical trial. In practice, however, the stages of the data lifecycle are often revisited throughout the process and do not always occur in strict sequence.

Processes such as data review, data corrections, and data transfer and exchange may happen at various points, often overlapping with other steps in the data lifecycle. For this reason, the steps of the data lifecycle will be presented in a flexible non-linear manner in this training.

In this sub-module, we'll review how the data lifecycle is managed, focusing on the general requirements in Section 4.

We'll also examine the responsibilities of both the investigator and sponsor throughout the data lifecycle for ensuring participant protection, data integrity and ultimately, the reliability of trial results. These responsibilities are outlined in E6(R3) Sections 2.12 and 3.16.1 for the investigator and sponsor, respectively.

Data capture and data corrections are closely linked. Many of the requirements and considerations that apply when data are captured also apply when data are updated, corrected, or modified over time. To reflect how data are managed in practice, this sub-module will address data capture and data corrections together.

Click "Data Capture" to continue.

Data capture is the process of acquiring trial data into a structured system or record for documentation, reporting, and analysis.

In our Crohn's Disease case example, data capture happens in various ways. For example, when the participant enters data in the eDiary, when the clinician documents assessments directly into the eCRF during a trial visit, and when lab analysis result values are generated from the central lab analysis of biological samples.

At the point of data capture, automated data validation checks to raise data queries should be considered as required based on risk, and their implementation should be controlled and documented.

In this Crohn's Disease case, haematocrit represents data of higher criticality because it is one of the eight components of the Crohn's Disease Activity Index (CDAI) score, one of the co-primary efficacy endpoints in this trial. As such, any error in haematocrit directly affects the calculation of the CDAI score, which in turn may influence participant eligibility, disease activity classification, and assessment of response to the investigational intervention.

Automated validation checks within the eCRFs can flag haematocrit values that are clearly implausible or inconsistent with physiologic expectations. For example, this could occur if investigator site staff accidentally enter a haemoglobin value instead of a haematocrit value. These validation checks can help prevent data entry errors in this critical to quality factor.

We will talk more about data review later in this module.

There should be processes in place to correct data errors that could impact the reliability of the trial results.

Corrections should be traceable and supported by relevant metadata, including the date, time, the person or computerised system making the correction, and, where applicable, the reason for the change.

Corrections should also be justified and supported by source records around the time of original entry, and performed in a timely manner.

Let's look at an example in the Crohn's Disease case where a participant mistakenly entered an incorrect value in the eDiary. Because the entry could not be edited after submission, the participant immediately notified the investigator.

The investigator documented the issue and kept a record of the participant's communication. This created evidence captured at the time of the event, which is exactly what is needed to support a legitimate data correction.

The correction was then processed through the system's formal data change process, with the eCOA system audit trail capturing the original entry, the corrected value, and the reason for the change.

When data errors occur, corrections are necessary, but they should be traceable, timely, and backed by source records. This is what separates a valid correction from one that undermines the reliability of trial results.

The investigator or institution should maintain adequate source records that include pertinent observations on each of the trial participants under their responsibility.

Source records should be attributable, legible, contemporaneous, original, accurate and complete. Changes to source records and reported data should be traceable, should not obscure the original entry and should be explained if necessary (via an audit trail). Data reported to the sponsor should be consistent with the source records or the discrepancies explained.

The investigator should define what are considered to be source records, the methods of data capture and their location prior to starting the trial and should update this definition when needed.

In addition, unnecessary transcription steps between the source record and the data acquisition tool should be avoided.

As a general example of this, investigator site staff initially record participants' observations in a temporary note pad to show the investigator. These notes are later entered into the electronic medical record and subsequently transcribed into the eCRFs a day later, rather than being documented directly in the electronic medical record and eCRFs at the time the observations are made.

The investigator should consider avoiding such extra transcription steps from a temporary note pad to the electronic medical record.

In the Crohn's Disease case, a similar efficiency can be achieved with laboratory data.

For example, instead of haematocrit results being manually transcribed in the eCRF by investigator site staff, results could be automatically transferred into the eCRF.

The data are automatically mapped into the correct fields and run through validation checks to make sure they are accurate and complete. Investigator site staff then simply review what has been transferred and confirm it, rather than re-entering the information manually.

This removes unnecessary manual entry, reduces the risk of transcription errors, and maintains a clear, traceable link from the central laboratory source to the eCRF.

The investigator should also ensure that data acquisition tools are used as specified in the protocol or trial-related instructions.

It is the sponsor's responsibility, among others, to:

- Pre-specify in the trial protocol the data to be collected and the method of collection
- Provide guidance to investigators and the institution, service providers and trial participants, where relevant, on the expectations for data capture, data changes, data retention and data disposal

In practice, the investigator delivers any guidance intended for trial participants, as the sponsor does not interact directly with participants.

When it comes to data corrections, the sponsor should allow correction of errors to data, including data entered by participants, where requested by the investigators or participants. Such data corrections should be justified and supported by source records around the time of original entry.

The sponsor should not make changes to data entered by the investigator or trial participants unless the changes are justified, agreed upon in advance and clearly documented and should not have exclusive control of data to prevent undetectable changes.

For example, in the Crohn's Disease case, exclusive control of the eCOA data by the sponsor is avoided by ensuring the investigators also have direct access to view these data and by using a third-party service provider to host the eCOA database.

Depending on the trial design, it may be appropriate to restrict investigator access to certain data identified in the protocol, particularly where such data could unblind the investigator. In these cases, the protocol should clearly specify the data to be withheld and confirm that they must not be shared with the investigator.

Next, we'll look at identification of relevant metadata, followed by the review of data and relevant metadata.

Click to continue.

Acquired data from any source, including data directly captured in a computerised system, should be accompanied by relevant metadata.

As we saw in Module 3.3, relevant metadata are those metadata needed to allow the appropriate evaluation of trial conduct.

We also learned in Module 3.3 that audit trails are a very important and specific type of metadata that should show activities, initial entry and changes to data fields or records, by whom, when and, where applicable, why.

For example, in the Crohn's Disease case, the participant is reporting their daily stool frequency, abdominal pain and well-being ratings in an eDiary. Examples of relevant metadata include typical audit trail data such as the unambiguous date and time when the entries were made or data were changed; what was entered or changed;

and, where applicable, why; the ID number of the participant who entered the data and the ID number of the device.

This information allows, for example, monitors, data managers, auditors or regulators to verify that the trial was conducted according to the protocol.

Next, we'll look at review of data and relevant metadata, an important step in the data lifecycle.

Click to continue.

Reviewing data isn't just about checking numbers. It's about understanding how, when, and by whom the data were captured and handled. When looking at both the data and the metadata together, a much clearer picture of data integrity and trial conduct can be gained.

For example, this review process can help identify missing data before it affects analysis.

It can highlight unusual patterns that may suggest data manipulation.

It allows identification of abnormal values, outliers, or data entered at unexpected times or inconsistent dates.

Also, it can detect errors in data processing, identify unauthorised system access, and recognise potential device or system malfunctions.

Importantly, metadata review may highlight when additional training is needed, whether for trial participants using devices or for investigator site staff entering or reviewing data.

And finally, it can help detect whether direct data capture has been implemented as defined in the protocol and protocol-related documents.

Together, these checks strengthen data reliability, participant protections, and overall trial quality.

To bring the importance of reviewing audit trails to life, let's return to the Crohn's Disease example.

In this trial, the audit trail review revealed that one participant was consistently entering symptom data for multiple days within a very short time window, rather than daily as required. This suggested the participant was completing entries retrospectively, rather than in real time at home, exactly the kind of pattern that is easy to miss without audit trail review.

Looking more broadly across the site, the audit trail also revealed identical timestamp patterns across multiple participants. This could be a more serious finding. Rather than reflecting independent daily entries made at home, this pattern could suggest

that data were entered in batches, for example, participants were making entries for multiple days at the time of site visits, or worse, that investigator site staff were entering participant data themselves, possibly through shared usernames and passwords. This kind of concern would be very difficult to detect through data review alone.

Appropriate quality control should be applied, based on risk. When data captured on paper or in an electronic medical record are manually transcribed into a computerised system, the need for and the extent of data verification should take the criticality of the data into account.

The sponsor should apply risk-based quality control at relevant stages of data handling, focusing their quality assurance and quality control review activities on data of higher criticality and relevant metadata.

There should be procedures in place for reviewing trial data, audit trails, and other relevant metadata. These reviews should be a planned activity, and their nature and extent should be based on the criticality of the data in the specific trial. They should also be adjusted as needed based on experience during the trial.

Taking the criticality of data into account, it is the investigator's responsibility to:

- Review data in a timely manner, including relevant data from external sources that may impact participant eligibility, treatment decisions, and safety monitoring
- Ensure the accuracy, completeness, legibility and timeliness of the data reported to the sponsor in the data acquisition tools completed by the investigator site, and make corrections if needed. And,
- Endorse reported data at important, pre-defined milestones.

In order for the investigator to be able to perform their review, the sponsor should:

- Ensure the investigator has timely access to data, including relevant data from external sources during the trial
- Ensure the protocol clearly identifies data that should not be shared with blinded investigators and investigator site staff, and
- Ensure data that may unblind the investigator are not shared
- Finally, the sponsor should ensure that the investigator receives instructions on how to navigate systems, data and relevant metadata.

For example, in the Crohn's Disease case, the protocol required that the investigator or delegated investigator site staff review the eDiary data and then manually transcribe the data from the eCOA system into the eCRF.

As we saw in Module 3.1, this transcription process is particularly complex because investigator and investigator site staff must also apply protocol-specific rules when entering the data.

Specifically, the protocol requires that site staff enter data in the eCRF for the 7 days prior to the trial visit and exclude from the CDAI calculation the days when participants received bowel preparation medications, underwent endoscopy, or the two days following endoscopy.

In this scenario, the data being transcribed contribute directly to the co-primary efficacy endpoint. As such, they are considered data of higher criticality. When data sit at the higher end of the criticality continuum, the processes used to collect, review and transcribe them become critical to quality factors. Here, the transcription of eDiary data from the eCOA system into the eCRFs is one such critical process.

The transcription process may introduce risks, including manual transcription errors when investigator site staff transfer data between systems, incorrect identification of protocol-specified exclusion days, and entry of inaccurate values into the eCRF. Any of these errors could directly affect the integrity of the co-primary endpoint and, ultimately, the reliability of trial results.

Sponsor and investigator oversight should include data review, verification, and other quality control activities that are proportionate to the importance of these data.

Both the investigator and sponsor should put processes in place to ensure risk-proportionate oversight.

For example, in this case, the investigator and sponsor should ensure that investigator site staff review the eDiary data in the eCOA system, appropriately identify which days met exclusion criteria, and accurately transcribe data from appropriate days into the eCRF for CDAI score calculation.

In addition, the sponsor should ensure that monitoring incorporates targeted source data verification and source data and metadata review in a manner proportionate to the importance of the endpoint data. Sponsors should periodically reassess identified risks and adjust monitoring intensity as needed to ensure continued data integrity.

Other risk mitigation strategies include providing training to investigator site staff on how to identify excluded days in the eDiary data and implementing data validation rules in the eCOA system to help site staff identify days that should be excluded.

The E6(R3) guideline refers to “review of data” and “endorsement of data.” It’s important to understand the difference.

Data review and data endorsement serve distinct but complementary purposes in clinical trials. Review of data involves ongoing checks to verify that data are accurate, complete, consistent, and plausible or to identify important safety information or data inconsistencies.

On the other hand, endorsement of data is done by the investigator at important milestones, such as prior to interim or final analysis, to attest that data accurately reflect the source records and are complete and final.

While review of data is a quality control process, data endorsement represents the investigator's official confirmation of data integrity and readiness for analysis or submission.

Investigators are responsible for data entered into eCRFs and other data acquisition tools under their supervision.

The signature of the investigator signifies that, to the best of their knowledge and under their oversight, the data entered in the data acquisition tool and submitted to the sponsor retain their integrity.

The sponsor should seek investigator endorsement of reported data at important milestones.

And, finally, the acceptable timing and frequency for the endorsement needs to be defined and justified for each trial by the sponsor and should be determined in a risk-based manner.

Thank you for completing Module 3.4a. Next, in Module 3.4b, we'll continue to explore sponsor and investigator responsibilities related to the remainder of the data lifecycle steps from data transfer, exchange, and migration through statistical analysis to destruction.