



Module 3.4b 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. This is Module 3.4b.

In this module, we'll continue to answer:

- 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?
- How can investigators and sponsors tailor their management and oversight throughout the data lifecycle based on risks that may impact critical to quality factors?

Now let's look at data transfer, exchange, and migration.

Click to continue.

During a trial, there may be a need to transfer, exchange, or migrate data for various reasons, for example when importing data from external parties to the sponsor's clinical database, when decommissioning a system, or when transferring data from the clinical database to a statistical environment for data analysis purposes.

As we see in the Crohn's Disease case, data often move between systems, for example, from a central lab to the eCRF, to the sponsor database and then to a statistical environment such as the Analysis Data Model, or "ADaM," datasets.

Validated processes or other appropriate processes such as reconciliation should be in place to ensure that electronic data, including relevant metadata, transferred between computerised systems retain their integrity and preserve confidentiality.

The data exchange or transfer process or system migration should be documented to ensure traceability, and data reconciliation should be implemented as appropriate to avoid data loss and unintended modifications.

For example, data may be lost during transfer if the transfer process is not adequately specified and validated. Examples of this include situations where the field lengths in the receiving database are different from those in the source database, resulting in data being truncated, such as adverse event descriptions or where character data are attempted to be loaded into a numeric field.

Eventually, the datasets should be finalised, the database should be locked and the datasets made ready for statistical analysis.

Click to continue.

Before unblinding and statistical analysis, the clinical trial database is “locked.”

Data of sufficient quality for interim and final analysis should be defined.

Such data are achieved through timely and reliable processes for data capture, data verification and validation, review, and rectification of errors and, where possible, omissions that have a meaningful impact on the safety of trial participants and/or the reliability of the trial results.

Several data management activities should be performed to finalise data sets prior to analysis.

These may include activities such as reconciling entered data and datasets, as well as reconciling relevant databases.

They may also involve correcting data errors and, where possible, addressing any missing information.

Additional activities may include medical coding, along with compiling noncompliance issues, including protocol deviations, and assessing their potential impact.

Finally, data extraction and determination of data analysis sets should take place in accordance with the pre-specified statistical analysis plan, or SAP, and should be documented.

These activities should be risk-based and documented in accordance with pre-specified procedures. Returning to the Crohn’s Disease trial, we can see what this looks like in practice. In this case, examples of data management activities identified as being of higher criticality include:

- Reconciliation of eDiary data to confirm that all entries have been transferred and stored in the eCOA database, and that any missing data have been identified and handled according to the protocol and data management plan
- Reconciliation of laboratory results from the central laboratory system against data recorded in the case report form, and
- Reconciliation of serious adverse events reported in the case report form with those logged in the sponsor’s safety database.

The sponsor is responsible for determining the data management steps needed before analysis to ensure the data are of sufficient quality.

For planned interim analyses, the sponsor should also manage access to and the ability to change data, based on what is needed to achieve adequate data quality for analysis.

In addition, before providing data for final analysis—and, where applicable, prior to unblinding the trial—the sponsor should restrict edit access to the data acquisition tools.

As previously explained, the investigator should confirm the accuracy, completeness, legibility and timeliness of data before database lock and provide their endorsement of the data.

Next, trial data are analysed according to the protocol, the statistical analysis plan, and principles of ICH E9, and then reported to applicable regulatory authorities.

Click to continue.

Statistical programming, data analysis and reporting are not covered in Section 4. Rather, as a distinct sponsor responsibility, activities related to statistical analysis are covered in Section 3.16.2.

Section 3.16.2 should be read in conjunction with ICH E9: Statistical Principles for Clinical Trials, and ICH E9(R1): Addendum on Estimands and Sensitivity Analysis in Clinical Trials, which provide detailed guidance on statistical principles for clinical development, trial design, conduct, analysis and reporting.

As stated in Section 3.16.2 of the guideline, the sponsor is responsible for several key activities related to statistical programming, analysis and reporting:

The first is that the sponsor should develop a statistical analysis plan that is consistent with the trial protocol and clearly describes how the data will be analysed.

And secondly, the sponsor should define, in advance, the criteria for including or excluding participants from each analysis.

To see how these sponsor responsibilities play out in practice, let's look at our Crohn's Disease case example. Here, a statistical analysis plan was produced that included the co-primary endpoint definitions and their analysis methods. The analysis was based on an Intention to Treat population, with the criteria for inclusion in that primary analysis clearly defined within the statistical analysis plan.

As introduced in Module 3.1, the two co-primary endpoints in this trial are CDAI Clinical Remission at Week 12, defined as the proportion of participants achieving a CDAI score of less than 150; and, Endoscopic Response at Week 12 defined as a greater than 50% decrease in the Simple Endoscopic Score for Crohn's Disease (or SES-CD) from baseline. For the purposes of this example, we'll be focusing on only the SES-CD endpoint.

The statistical analysis plan also sets out a clear process for imputation, that is, the statistical method used to estimate or substitute missing values so that if any

participant data is missing at 12 weeks, there's a defined approach for how to handle it.

So, let's now take a closer look at how the data analysis played out for the SES-CD endpoint in the Crohn's Disease trial.

To illustrate this process, we'll be looking at dataset extracts for two participants from a total of 380 enrolled in the trial. To make these easier to follow, the variable names have been replaced with descriptive labels.

A programmer begins by working with the extracted trial data, writing programs to create two SDTM datasets. The first is the SDTM Findings About domain, which captures the endoscopic scores across the trial visits for each participant.

The second is the SDTM Demographics domain, which includes each participant's allocated randomised treatment and the actual treatment they received, once the data were unblinded following database lock.

Looking at the data for our two example participants, you can see that participant 001 has SES-CD scores recorded at every visit, including Week 12. Participant 002, however, has a missing value at Week 12 which, as the statistical analysis plan anticipated might happen, is the result of a protocol deviation.

The Demographics domain reveals a further complication for this participant: despite being randomised and allocated to active treatment, they were mistakenly given placebo at the site.

These two SDTM datasets are then fed into another program, which creates the analysis or ADaM dataset named SES-CD. Looking at the ADaM table, you can see two new derived variables have been calculated from the baseline and Week 12 SES-CD values.

For participant 001, a baseline score of 26 and a Week 12 score of 3 give a change from baseline of 23, and a percentage change of 88.5%, a substantial reduction.

Because this exceeds the 50% threshold defined in the co-primary endpoint, the program codes this participant as a responder, with the SES-CD response parameter set to 1.

For participant 002, the missing Week 12 data have been handled according to the imputation method set out in the statistical analysis plan, resulting in the Week 12 value remaining missing.

The program has coded this participant as a non-responder, with the SES-CD response parameter set to 0.

You'll also notice that, in line with the Intention to Treat analysis defined in the statistical analysis plan, this participant's treatment has been coded as Active, the treatment they were randomised to, rather than the placebo they actually received.

Finally, the SES-CD response parameter, coded as 1 for yes and 0 for no, is used by another statistical analysis program to conduct hypothesis testing on the difference in the proportion of responders between treatment groups.

The output of that analysis is what you can see here in Table 3. Of the 182 participants in the active treatment group, 61 or 33.52%, were responders, compared to 18 out of 196, or 9.18%, in the placebo group.

The difference of 24.3% is statistically significant, with a p-value of less than 0.001 and a 95% confidence interval of 16.4 to 32.3.

You'll also notice that the file path for the source code of the analysis program is identified, and the runtime date and time stamp are located at the bottom of the table, which is important for traceability.

The process we've just walked through generates a trail of outputs, datasets, and programs. Getting that programming right is critical to quality. If a program were not functioning correctly, it could produce unreliable results, for example, a miscalculation of the percentage change in SES-CD score could lead to participants being incorrectly assigned their responder status, which would directly undermine the integrity of the results.

To guard against this, the sponsor should ensure that appropriate and documented quality control of statistical programming and data analysis are implemented.

Outputs should be traceable to the statistical software programs, dated and time stamped, and protected against changes, with access controls implemented.

The sponsor should retain all statistical programming records that relate to the outputs contained or used in the trial results, including any quality control and validation activities performed.

The sponsor should retain these records in a way that allows an evaluation of whether the statistical analysis plan was followed, how variables were calculated, and how any deviations were managed, ensuring full traceability of data transformations and derivations throughout the analysis process.

Finally, any deviations from the planned statistical analysis, or changes made to the data after the trial has been unblinded, should be clearly documented and justified in the clinical trial report, and should only occur in exceptional circumstances.

After all analysis and reporting activities have been completed, data must be retained and be accessible, if needed, for the entire retention period, according to applicable

regulatory requirements. Refer to sections 2.12.12 and 3.16.3 for the record-keeping responsibilities of investigators and sponsors, respectively.

Click to continue.

Trial data and relevant metadata should be archived in a way that allows for their retrieval and readability but should be protected from unauthorised access and alterations throughout the required retention period.

Retention ensures continued accessibility for audits, inspections, or future reference.

Further investigator and sponsor responsibilities for maintaining and retaining essential records and other trial data are covered at length in Module 5 of this training course. Please navigate to Module 5 for more information on record retention and access requirements.

Finally, at the end of the retention period, trial data and metadata may be destroyed.

Click to continue.

The trial data and metadata may be permanently destroyed when no longer required as determined by applicable regulatory requirements.

This brings us to the end of Module 3: Data Governance.

In this module, we looked at how to apply the key GCP principle of risk proportionality to data governance, and investigator and sponsor responsibilities for management and oversight of computerised systems and the data lifecycle.

Thank you for completing this module.

The Multi-Regional Clinical Trials (MRCT) Center of Brigham and Women's Hospital and Harvard is an ICH Training Associate. The MRCT Center designed and developed this training module in collaboration with ICH.