PharmaSUG 2023 - Paper DS-157

Hard Coding: What is it? Why Does It Matter? What Should You Do When Faced With Such A Request?

Michael Nessly, ICON PLC

ABSTRACT

A clear operating definition of what constitutes hard coding is not readily available in the public domain. We define hard coding as when data, which are not present in the collected source data, or if present are incorrect, are inserted or corrected programmatically into the dataflow. Hard coding is the software development practice of embedding data directly into the source code of a program or other executable object, as opposed to obtaining the data from external sources. Outside of regulated activities such as analysis and reporting of clinical trials, it appears that this is a common and non-controversial practice.

To understand why hardcoding is very poor practice not to be taken lightly in clinical trials, one must be familiar with the concept of traceability. Traceability is, in essence, the provenance of data. Traceability is the property that enables the understanding of the data's lineage and/or the relationship between an element and its predecessor(s). Traceability facilitates transparency, an essential component in building confidence in results or conclusions. Traceability is not just good practice, it is required. Hard coding breaks traceability. This is risky in present use and can be disastrous in future use of data subjected to hard coding.

ICON opposes hard coding in any program that generates analysis datasets or TFLs used for analysis or reporting of clinical studies. While many statistical groups have strong policy against hard coding, receipt of requests for hard coding remain common today. This presentation aims to increase knowledge of the issues, foster engagement in seeking alternative approaches and provide instruction on how to effectively document when there are cases where hard coding is determined to be done.

INTRODUCTION

It is the expectation of the global health authorities that, for a clinical trial, there exists an electronic, immutable, auditable data stream from data capture through data reporting.

Note that this is a frequent audit line of inquiry by Sponsors and by health authorities.

This invokes the concept of traceability.

Hard coding is defined as when data, which are not present in the collected source data, or if present are incorrect, are inserted or corrected programmatically into the dataflow.

Hard coding breaks traceability.

This presentation aims to increase knowledge of the issues, foster engagement in seeking alternative approaches and provide instruction on how to effectively document when there are cases where hard coding is determined to be done.

TRACEABILITY

Traceability is a Clinical Data Interchange Standards Consortium (CDISC) Analysis Data Model (ADaM) principal and is defined in the CDISC ADaM Implementation Guide[2]:

Traceability – The property that enables the understanding of the data's lineage and/or the relationship between an element and its predecessor(s). Traceability facilitates transparency, which is an essential component in building confidence in a result or conclusion. Ultimately traceability in ADaM permits the understanding of the relationship between the analysis results, the analysis datasets, and the SDTM domains. Traceability is built by clearly establishing the path between an element and its immediate predecessor. The full path is traced by going from one element to its predecessors, then on to their predecessors, and so on, back to the SDTM domains, and ultimately to the data collection instrument.

Traceability in context of ADaM data sets means providing the method followed to derive an analysis

endpoint from source SDTM data. CDISC ADaM IG 1.0 strongly recommends the incorporation of traceability feature in ADaM data sets submitted to FDA.

Features exist in the ADaM standard that allow for traceability of analysis reporting results to ADaM to SDTM. Creating both SDTM and ADaM from the raw data is incorrect (especially when submitting only SDTM and Analysis data). Raw data should be the source for SDTM, SDTM should be the source for ADaM and ADaM is the source for the clinical study TFLs. This concept is so important that the CDISC ADaM team has authored the ADaM Examples of Traceability[1].

HARD CODING IS ADDRESSED IN STANDARD OPERATING PROCEDURES (SOPS)

ICON opposes hard coding in any program that produces analysis datasets or tables/figures/listings used for analysis or reporting of clinical studies data.

Text from ICON Standard Operating Procedures:

Preparation of Analysis Datasets, Tables, Listings, Figures and Other Output for the Analysis and Reporting of Clinical Studies Data

ICON opposes “hard coding” in any program that produces analysis datasets or tables/figures/listings used for analysis or reporting of clinical studies data. The intent or request by a Sponsor/other to “hard code” shall be escalated, approved by Senior Management (head of biostatistics and head of statistical programming), documented and stored in TMF.

ICON B&P opposes hard coding in any program that generates analysis datasets or TFLs used for analysis or reporting of clinical studies data and hence recommends re-opening of the database and correcting the data there.

The SOP has an accompanying Hard Coding Approval Template.

Outline of documentation of hard coding request

- Requestors: Lead Biostatistician and Lead Programmer
- Describe the hard coding request.
- Describe the rationale for the request including all options and why the ones not involving hard coding are not under consideration.
- Describe in detail how the hard coding request would be implemented.
- Describe future mitigating actions
- External and Internal approval signatures
- Sign off by the appropriate representative of the requesting Sponsor is required prior to ICON sign.
- There should be a Warning in any code which includes hard coding to flag the hard coding in the output log.
- It is also helpful to put an expiration date on the hard coding if it will apply for a limited period of time or a limited set of deliverables.

Sign off by the appropriate representative of the requesting Sponsor is required prior to ICON sign off.

Sign-off of the template is done jointly by the ICON heads of Biostatistics and Statistical Programming. The default position is to not sign off unless a compelling case can be made why there are no other viable or practicable alternative approaches.

Similar documentation is required for any programming code provided by Sponsor or other external sources which includes hard coding.

Additionally, as output deliverables might be distributed apart from the supporting data, metadata, and programs, consideration should be given to documentation of hard coding in the output. Footnotes to impacted tables may be needed for such documentation.

CASE EXAMPLES

Case example 1:

The study has been prematurely cancelled. To close out the study an abbreviated CSR is needed based upon partially screened and cleaned data. The study only has raw data. SDTM was contracted to vendor but since the study was cancelled, vendor would not provide SDTM. All programming work needs to be done from raw data directly. While programming for the abbreviated CSR delivery, team encountered obvious data entry errors against the study design. As the data were not cleaned and the sites were closed, source data could not be updated. The hard coding request has been raised to handle programmatically for the abbreviated CSR delivery.

The abbreviated CSR reports will not be submitted for regulatory submission and will be utilized for internal purpose.

Case example 2:

Analyses for an interim presentation of an ongoing oncology trial is requested based upon a data cut snapshot. A participant is known to have lost response by study criteria in the accumulating data not in the snapshot. The chief medical officer feels that to not include a known treatment failure in the analysis would not be transparent and complete reporting and requests that this loss of response be hard coded into the analysis. The biostatistical lead points out that the best approach would be a recut of the data with a programmatic search for any other participants who have lost response to avoid ascertainment bias but the CMO is adamant that there is no time for this. The data will be cleaned and updated for future analysis and reporting.

CONCLUSION

Summary

- Be aware of these issues.
- Be willing to engage in discussions with the Sponsor. The regulatory compliance and audit readiness risks are frequently not appreciated.
- Escalate such issues quickly with clear documentation of situation, perceived need and justifications.
- Do not regard hard coding as an easy or acceptable solution to data deficiencies.
- If it is decided to do hard coding, thoroughly document this decision and data consequences.
- This documentation resides in the study Trial Master File
- If it is decided to do hard coding, keep this to a minimum.
- It is prudent not to hard code for submission deliverables.
- If there are future deliveries, remedy the data situation in advance of these.

Note that the process of completing documentation by itself frequently results in the request for hard coding being withdrawn.

REFERENCES

1) ADaM Examples of Traceability | CDISC
2) ADaMIG v1.2 | CDISC

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Michael Nessly
Michael.nessly@iconplc.com