Paper SA05

Approach of Programming Production Harmonized with Submission

Chengxin Li, Luye Pharma Group, Princeton, New Jersey, USA

ABSTRACT

This paper introduces one programming practice in CDISC studies: production harmonized with submission. The approach of programming production harmonized with submission can achieve submission ready once programming delivered, avoiding challenges in any discrepancies between submission and production with post processing model. The scope of this paper initiates from introduction of programming framework with end-to-end streamlined implementations, then describes common programming documents in production verses submission, and further illustrates the harmonized method. The method introduced here has been successfully piloted and applied in our production.

INTRODUCTION

The Clinical Data Interchange Standards Consortium (CDISC)[1] has defined a series of data models from data collection to submission. Clinical Data Acquisition Standards Harmonization (CDASH) is a data collection standard and harmonized with Study Data Tabulation Model (SDTM). SDTM standardizes the collected data in tabulations. SDTM should fully reflect the collected data (e.g., mapping for any collected data and deriving a limited number of variables, but no imputation for missing data). Furthermore, Analysis Data Model (ADaM) should only be derived from SDTM. The key endpoint analyses, inferential analyses, and complicated analyses should be designed with ADaM datasets.

The CDISC standards are widely implemented. With CDISC, it's aiming to achieve end-to-end streamlined data processing from data collection to analysis with traceability facilitated correspondingly. From programming perspective, figure 1, with self-explanations, depicts the end-to-end streamlined data processing architecture. The processes are streamlined with sequential layers, from data collection (layer 1), to data tabulation (layer 2), then to analysis dataset (layer 3), and further to displays (layer 4). Each layer is supported by respective CDISC standard(s) or study documents. Under this architecture, traceability is ensured reversely from any displays, e.g., a p-value in a table, back to the ADaM analysis dataset, to the tabulated SDTM dataset, and further back to the collected data. Associated with streamlined programming, figure 1 also lists corresponding output documents linked with submission.

The "end-in-mind" philosophy shall be another principle in CDISC implementation. Here the "end-in-mind" refers to the submission orientation with regulatory agency requirements in mind. The regulatory agency is the customer. Therefore, the design shall meet the agency requirements and facilitate agency reviews. With the end-in-mind philosophy, the best solution can be often determined among multiple options.

The agency requirements are transparent and available in public domain[2]. The Pinnacle 21 Community validator fully reflects the agency compliance rules. The Study Data Technical Conformance Guide[3] provides detail recommendations on CDISC practices as well.

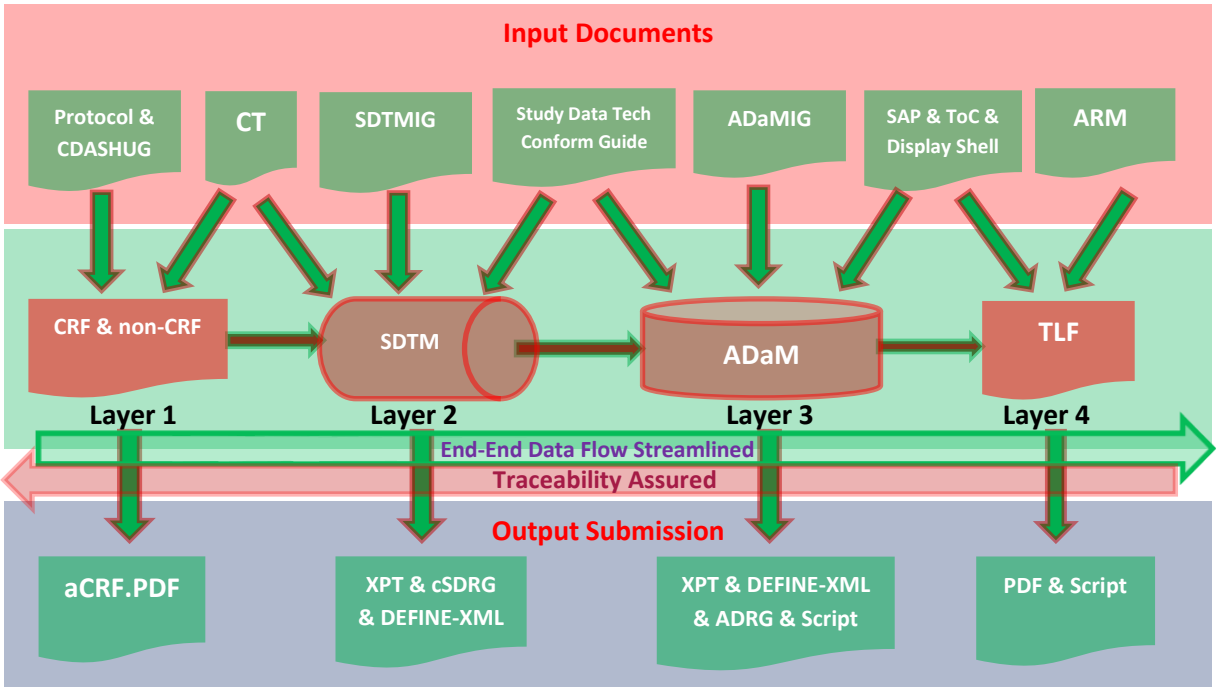
The scope of this paper is limited to statistical programming perspective for clinical data processing.

---

[1] http://www.cdisc.org

[2] https://www.cdisc.org/FDA-Final-Binding-Guidance-on-Standards

[3] http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf



CDASHUG=Clinical Data Acquisition Standards Harmonization User Guide; CT=Controlled Terminology; **SDTMIG**=Study Data Tabulation Model Implementation Guide; **ADaMIG**=Analysis Data Model Implementation Guide; **SAP**=Statistical Analysis Plan; ToC=Table of Content; **ARM**=ADaM Analysis Results Metadata; **aCRF**=Annotated Case Report Form; **TLF**=Table, Listing, and Figure; **XPT**=SAS Transport Version 5 Format File; c**SDRG**=clinical Study Data Reviewer's Guide; **ADRG**=Analysis Data Reviewer's Guide; **PDF**=Portable Document Format; **XML**=eXtensible Markup Language

Figure 1 End-to-end Streamlined Data Processing Architecture

PROGRAMING DOCUMENTS

Documentation is always the core component in data processing and submission. Conforming to programming SOP/guideline and fully supporting SAP, the production documents in programming commonly include display templates, table of contents (ToC), dataset specifications (SDTM, ADaM), and corresponding deliveries—tables, figures, and listings.

Regarding submission, as illustrated in submission outputs of figure 1, it's required all the datasets with XPT format, dataset specifications in DEFINE-XML format including analysis results metadata (ARM), the reviewer's guides (cSDRG, ADRG) in PDF format as complimentary documents to dataset specifications, the TLF in PDF formats, and the programming scripts in text format.

Figure 2 summaries programming documents in both production and submission.

Production	Submission
aCRF, providing instructions to SDTM mapping	aCRF.PDF
SDTM specifications; SDTM datasets as deliveries	SDTM.XPT; DEFINE-XML; cSDRG.PDF
ADaM specifications; ADaM datasets as deliveries	ADaM.XPT; DEFINE-XML(incl. ARM); ADRG.PDF
ToC; Display Templates; TLFs as deliveries	TLFs.PDF
SAS® Codes incl. Macros	Scripts.TXT with Readability, esp. for Core ADaM or Tables
QC Documents incl. QC codes	NA

Figure 2 Programming Documents in Production and Submission

Figure 2 shows that production documents and submission documents are aligned with minor differences.

Quality control (QC) is a critical component in programming process, commonly with independent double programming. However, the QC documents are in the scope of audit rather than in the submission.

The aCRF.PDF is a submission required document with SDTM-compliant annotations to facilitate reviews. In production, besides serving traceability from SDTM to data collection, the aCRF also additionally provides instructions to clinical database setup and SDTM transformations.

For SDTM submission, currently it still requires XPT file format. The specification for SDTM transformation goes with define-xml. The corresponding SDTM review guide is named as Study Data Reviewer's Guide, specifically cSDRG for clinical. The DEFINE-XML and cSDRG are required for submission.

For ADaM submission, XPT file format is also required for the analysis datasets. The specification for ADaM goes with define-xml. The corresponding ADaM review guide is named as Analysis Data Reviewer's Guide (ADRG). The DEFINE-XML and ADRG are required for submission. The programming codes in text format are recommended to submit, esp. for core ADaM datasets and core tables.

To harmonize ARM with submission, there may have multiple options to manage ARM such as configured in a reporting system, consolidated in the Table of Contents (ToC), or explicitly embedded with macro parameters. To facilitate agency reviews, together with the corresponding programming codes, the ARM is recommended to submit with DEFINE-XML, esp. for key endpoints analyses.

TYPICAL PROGRAMMING APPROACH

Figure 3 illustrates the normal programming processes. It requires the programming processes conforming to programming SOP or guidelines from designs, implementations, and deliveries. In design stage, to develop the tables of contents, display templates, and dataset specifications, there require several primary sources as inputs such as the SAP, CRF, database structures, and team reviews. In implementations, the designs are improved and then finalized with further data learnings, independent programming findings, and blinded review comments. After the database lock, the validated datasets (SDTM&ADaM) and analyses (TLFs) are delivered and achieved along with programming documents.

It is a typical practice that the submission is isolated from production programming. During submission, the updates of programming and documentation may be unfavorably required, esp. for CDISC compliances. This always brings challenges to programming processes and resources. However, the challenges can be avoided by production harmonized with submission.

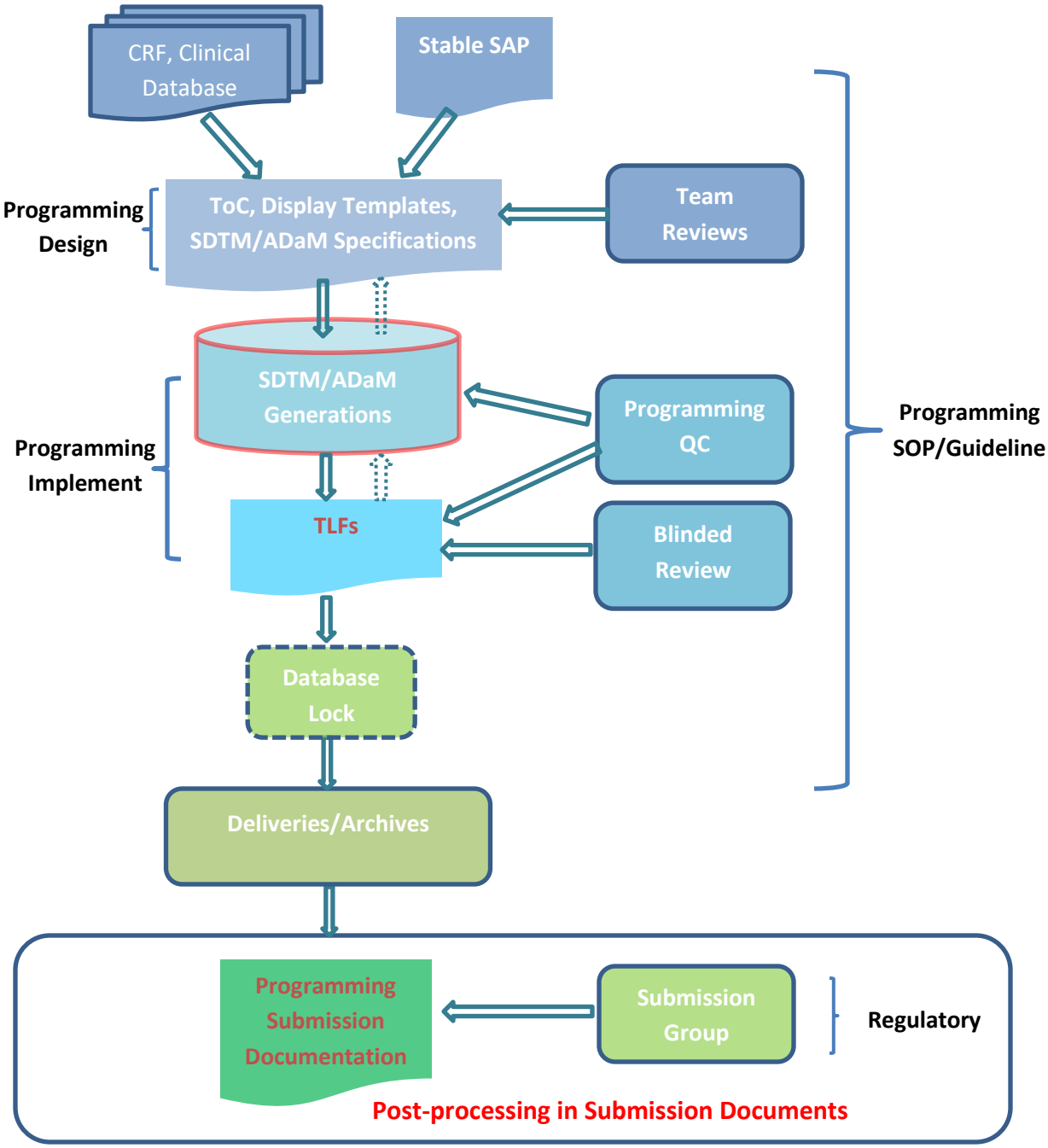


Figure 3 Typical Programming Processes

PRODUCTION ALIGNMENT WITH SUBMISSION

Figure 4 depicts the innovated programming process with production harmonized with submission. Compared with figure 3, the primary improvements are in applying end-in-mind philosophy to programming design stage, i.e., keeping submission requirements in mind during programming design. And the CDISC compliance checks become routine work at production during implementations. Correspondingly, the submission conformed documentation is switched from post processing to prior processing. During production, as live documents, the dataset specifications with DEFINE-

XML format and the review guides are also included in the team review package. The DEFINE-XMLs for datasets are evaluated and the review guides for datasets are finalized while datasets are finalized.

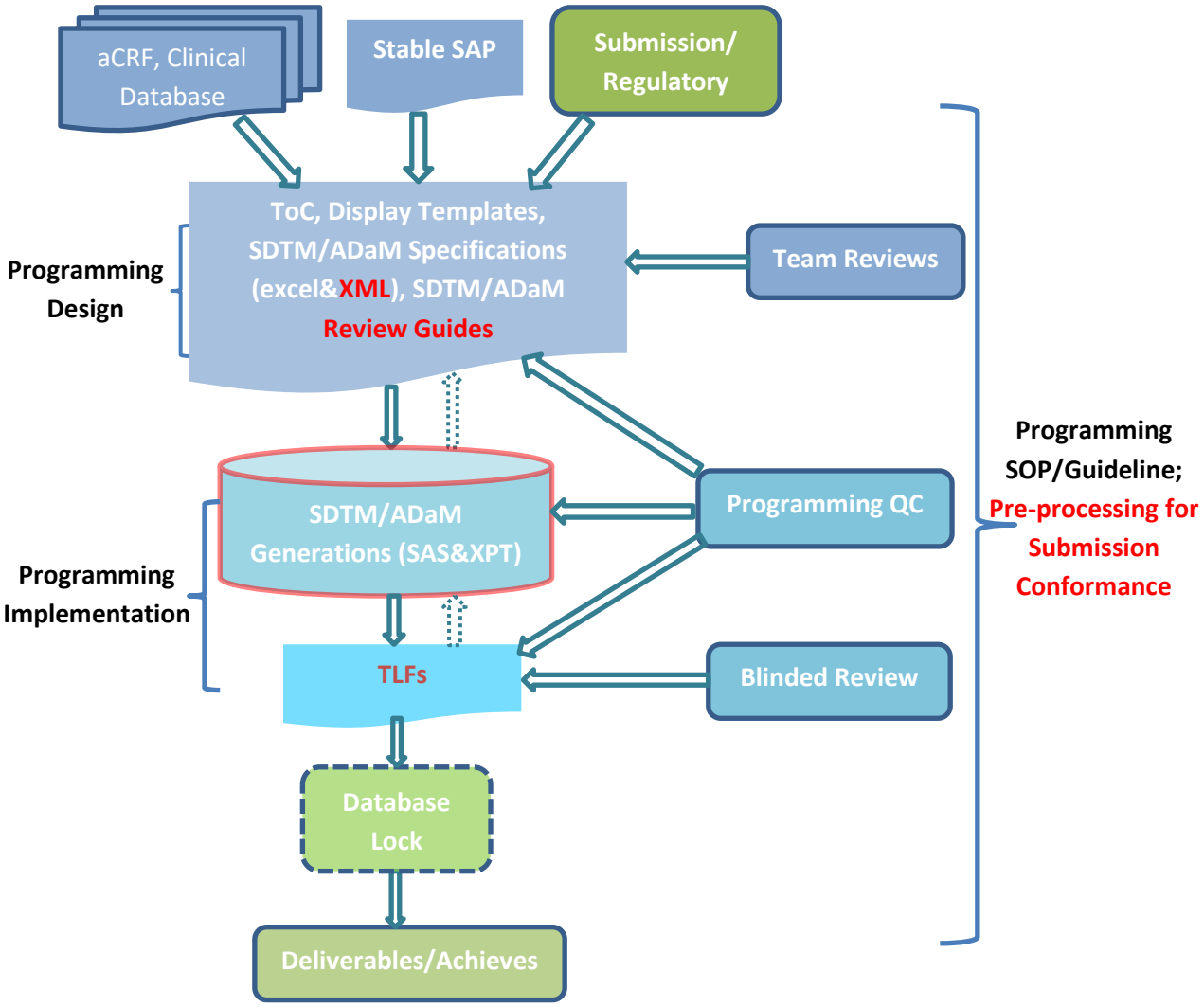


Figure 4 Production Aligned with Submission in Programming

With end-in-mind philosophy and end-to-end streamlined data processing, it becomes feasible to achieve the production harmonized with submission. By the end-to-end streamlined data processing, the aCRF shall be first produced to facilitate clinical database setup and SDTM transformations. Reversely, the SDTM can be generated with specification in XML format tracing back to the annotated CRF. The review guides are the complimentary documents to SDTM/ADaM specifications in CDISC studies. With company standard templates of cSDRG and ADRG, dataset review guides can be documented as rolling based during implementations and become part of final deliverables. The CDISC compliance checks shall be conducted as a routine programming practice. Any non-compliance issues can be caught up earlier and findings of data issues queried in time. Correspondingly, with the pre-processing practices in submission conformance, programming efficiency and quality can be improved.

Optionally, programming folder setup may be harmonized with Electronic Common Technical Documentation (eCTD) module 5 dataset structures. The file naming conventions, e.g., program name, shall comply with submission convention as well.

Production harmonized with submission in programming achieves submission ready for programming deliverables, and efficiently leads to submission documentation in high quality.

DISCUSSION

CDISC implementation presents challenges but also provides opportunities to improve programming processes for the deliverables with high quality and high efficiency. From statistical programming perspective, the paper summaries the best practices in CDISC implementations accompanying optimal programming approach.

From our practices in the programming approach of production harmonized with submission, it may take a time for programmers to switch and get used to a new programming style. However, the rewards are huge in terms of compliance, efficiency, and quality.

The programming standard operating procedure (SOP) and guidelines are developed from programming best practices and further provide guidance to programming productions. As needed, the programming SOP and guidelines shall be reexamined and updated correspondingly to reflect the best practices.