

Data Quality Findings from JumpStart

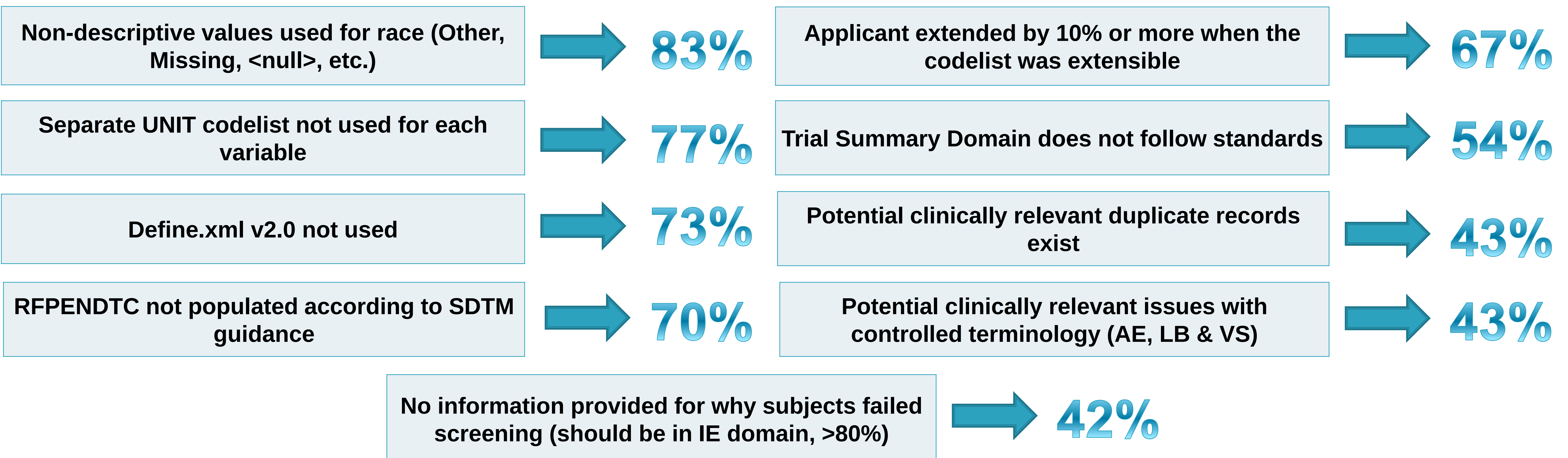


DeYett Law¹, Crystal Allard¹, Mary Doi¹, Lilliam Rosario¹, Barbara Witczak¹, Jesse Anderson¹,
Kathryn Matto², Austin Taylor², Jeno Pizarro², Margo Cohen²

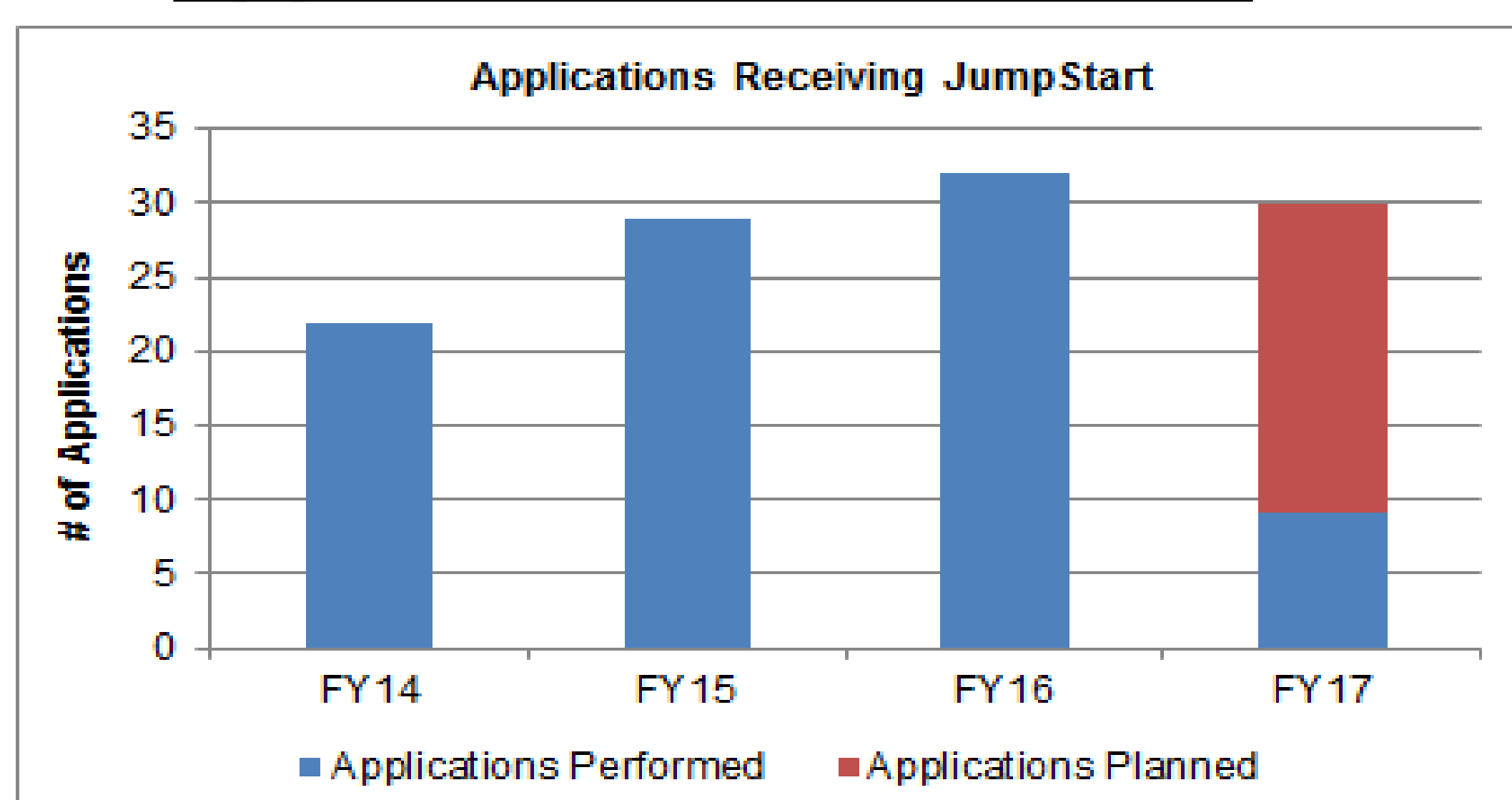
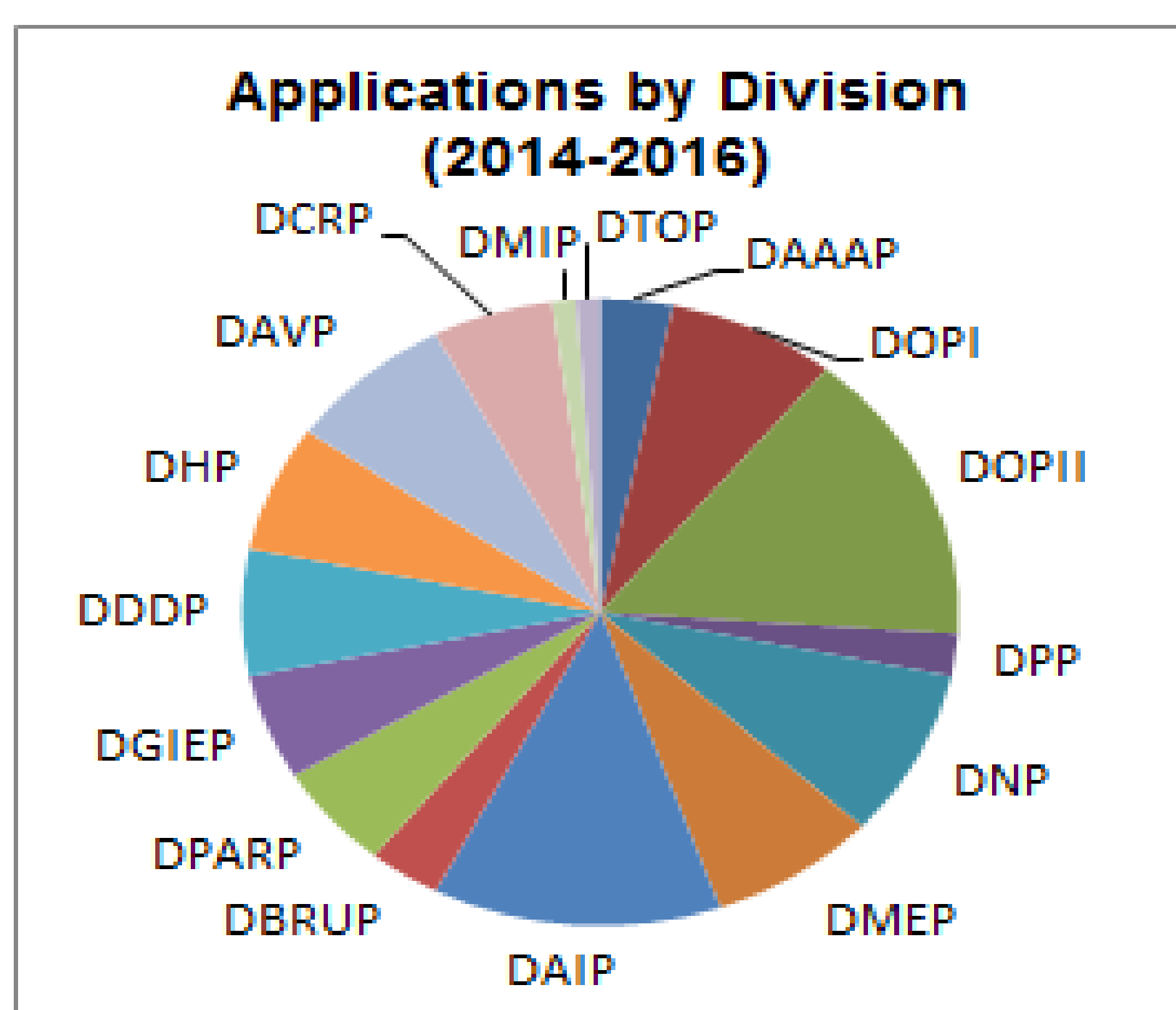
¹ Office of Computational Science, ² IBM

ABSTRACT: The JumpStart service provides CDER Review Teams with an assessment of the composition and quality of the clinical trial data submitted within an application. From inception, the JumpStart service has provided an assessment of 92 applications. Through this analysis, the service has tracked common issues that impact review. JumpStart has also identified additional data-related items that could assist review, if implemented. JumpStart has developed suggestions based on this tracking and identification process.

Common Issues in 2016 (% of applications)



JumpStart has analyzed 92 applications since FY14



Data Package Suggestions from JumpStart Findings

Define File

- Use Define.xml v2.0
- Include detailed description of data elements, for example:
 - Code lists that describe categories, subcategories, and reference time-points
 - Detailed and reproducible computational algorithms for all derived variables
 - Applicable value level metadata and description of SUPQUAL domains
 - Explanations of sponsor-defined identifiers (i.e. – SPID, GRPID, etc)
- Provide separate unit code lists for each domain

Study Data Reviewer's Guide (SDRG)

- Provide SDRG for each data package with each section populated
- Fix all possible issues identified by FDA Validation Rules and include clear and detailed explanation for all “non-fixable” issues
- Provide Data Flow diagram that shows traceability between data capture, storage, and creation of datasets

Disposition Domain Suggestions

- Include time-point information (event start/end date) for the event in Disposition records, not just when the event was recorded
- Include records regarding subject study completion and last follow-up contact with subject in disposition domain
- Accurately code reasons why subject did not complete the study or study treatment

Generic Data Suggestions

- Harmonize MedDRA versions across studies within an application
- Ensure consistent subject death information across all datasets and records
- Remove duplicate records
- Use the most recent version of FDA Validation Rules and fix data issues identified
- Populate RFPENDTC according to SDTM guidance
- Properly use ACTARM variable
- Include study day variable (--DY (Study Day), -- STDY (Study Day of Start of Observation), -- ENDY (Study day of end of observation)) for all observational datasets
- Use CDISC controlled terminology variables when available
- Provide EPOCH variable in all appropriate domains
- Use Baseline flags in LB and VS domains
- Include Seriousness Criteria for all Serious Adverse Events
- Provide Trial Design domains that are complete and accurate