Data Quality Findings from JumpStart

DeYett Law[1], Crystal Allard[1], Eileen Navarro[1], Mary Doi[1],Timothy Kropp[1], Lilliam Rosario[1], Dave Epstein[2], Kathryn Matto[3], Austin Taylor[3], Jeno Pizarro[3], Margo Cohen[3]

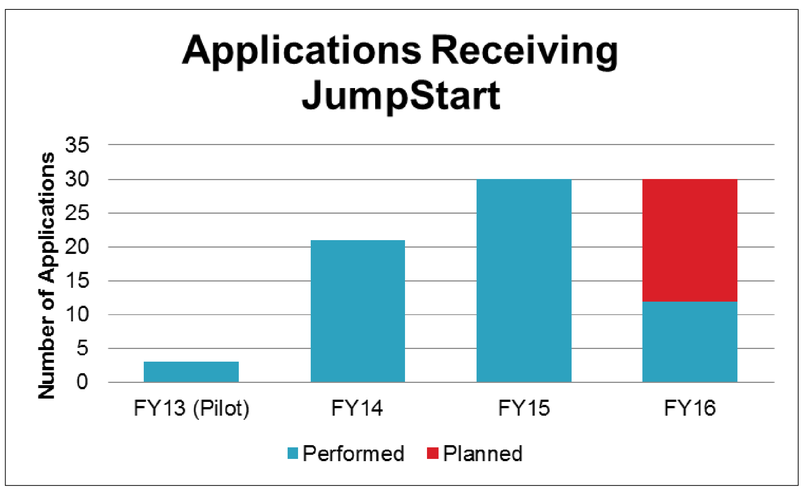
1 Office of Computational Science, 2 Chickasaw, 3 IBM

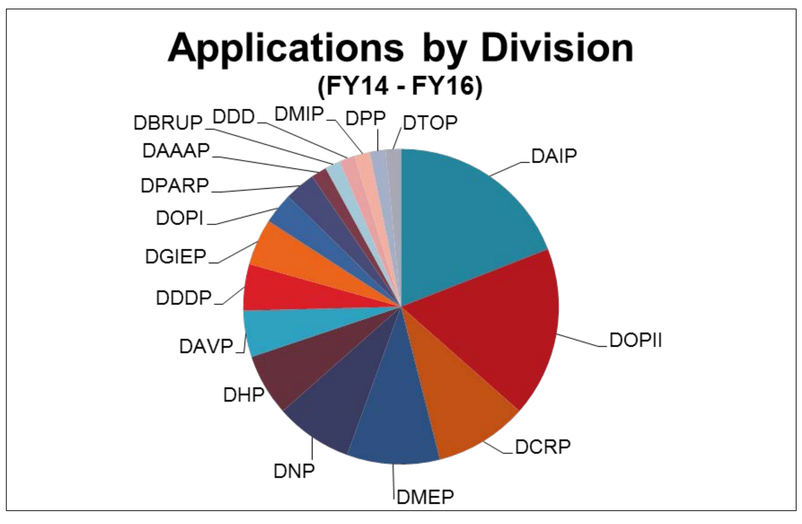
ABSTRACT: The JumpStart service provides CDER Review Teams with an assessment of the composition and quality of the clinical trial data submitted within an application. From inception, the JumpStart service has provided an assessment of 66 applications. Through this analysis, the service has tracked 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 2015 (% of applications)

Missing baseline tests in Laboratory Results, Vital Signs, or Microbiology domains	29%	Study Data Reviewer's Guide (SDRG) not provided or missing key sections	38%
Duplicate records found with potential contradictory information	67%	EPOCH variable not provided for all events, interventions, and findings domains	75%
Controlled terminology issues featuring ambiguous values which lead to confusion	63%	Adverse event seriousness criteria missing for adverse events designated as serious	46%
Treatment Emergent flag for Adverse Events not provided	63%	Trial summary domain not provided in the study data package	8%

JumpStart has analyzed 66 applications since FY13





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 SUPPQUAL 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 even 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
- Provide AE Treatment Emergent Flag within the SUPPAE domain
- Include Seriousness Criteria for all Serious Adverse Events
- Provide Trial Design domains that are complete and accurate