

Automating SDTM Compliance: A Web-Based Quality Control Tool Enhancing Clinical Trial Data Integrity

Pre-Built Validation | Zero Programming | One-Click Execution

Indraneel Narisetty, Gilead Sciences Inc
Gopal Reddy Palvatla, Gilead Sciences Inc

BACKGROUND

- SDTM (Study Data Tabulation Model) is the CDISC standard for regulatory submissions.
- Manual validation requires 4-6 hours per study with programming expertise.
- Current challenges: inconsistency, limited coverage (20-30 checks), network latency.
- Need for automated, comprehensive, user-friendly validation solution.

OBJECTIVE

Develop a desktop application to automate SDTM quality checks across 15+ domains with interactive visualization, eliminating validation programming requirements for end users.

METHODS

Technology Stack:

- Python 3.13 with PyQt6 desktop framework.
- Dash/Plotly for interactive data visualization.
- Pandas/NumPy for data processing.
- Pyreadstat for SAS dataset integration.

Implementation:

- 130+ pre-written automated quality checks across 15+ SDTM domains.
- Four category checks: CDISC compliance, data consistency, data validation and SDTM Mapping specification checks.
- Two-stage navigation architecture for network optimization.
- AG Grid tables with advanced filtering capabilities.

Deployment Options:

- Desktop application (standalone executable).
- Web server deployment (subscription-based).
- Scheduled batch processing for daily validation.
- CI/CD pipeline integration.

KEY FEATURES

15+ SDTM Domains Supported
DM, AE, CM, VS, LB, LC, EG, EX, DS, SE, MH, PR, QS, TU, TR, RS, SC, SV with domain-specific validation rules.

Intelligent Data Review

- Separate tabs for each dataset
- Pass/Fail status cards
- Subject-level filtering across domains.
- Integrated specs viewer (no Excel required)
- Real-time data preview without opening SAS application.

Advanced Filtering & Export

- Multiple filter conditions
- Auto-filter failed checks
- Cross-domain trend analysis.
- Excel export with conditional formatting
- Color-coded results (Pass/Fail/Warning).
- Advanced Filtering and Subject-level Navigation

VALIDATION CHECKS (130+)

CDISC Compliance (30+)

- Required variables
- Variable types & lengths
- Controlled terminology
- Domain structure

Date Validation (20+)

- ISO 8601 format
- Partial date handling
- Date sequence logic
- Study day calculations

Data Consistency (40+)

- Subject-level consistency
- Cross-domain validation
- Duplicate record detection
- Referential integrity

Visual Summaries (10+)

- Outlier detection
- Distribution analysis
- Missing data patterns
- Reference range checks

PERFORMANCE OPTIMIZATION

Challenge: Network Folder Scanning

Initial implementation: The system was checking an extremely large number of folders, which caused long loading times.

Solution: Two-stage dropdown with LRU caching

- Stage 1: Top-level folders only (~20 items)
- Stage 2: Conditional deep scan of selected path

Results: 94-98% faster initial load (<2 sec), 99.9% faster re-selection (<100 ms)

SDTM Quality Checks

Select Project

Select Study

Choose Analysis Folder

Choose Version

Choose a project...

Choose a study...

Choose an analysis folder...

Choose a version...

Run Checks

SDTM Quality Checks

Select Project

Select Study

Choose Analysis Folder

Choose Version

cdisc

cdisc01

primary_analysis

version1

Run Checks

Full Summary Table				
CHECK	Message	Notes	Details	
This check looks for DM entries where ARM is not equal to ACTARM	Fail	12 patients where ARM is not equal to ACTARM	DM	
Check for death information consistency across DM, DS, and AE	Pass		AE, DS, DM	
Check for missing or suspicious age values in DM. Checking age <18 and >90. If study allow, ignore this check	Pass		DM	
Check for consistency in informed consent dates between PRF01C and DS01C	Pass		DM, DS	
Check for duplicate USUBID in DM	Pass		DM	
Check for consistency between DTML and DTCHD in DM	Pass		DM	
Check for duplicate USUBID in DM	Pass		DM	
Check for patients with COVID-19 AE indicating drug withdrawal but no Treatment Discon form including AE		Check not run, did not detect COVID-19 preferred terms in AEECDICD	AEDS	
AE entries where ASACNOTH = "MULTIPLE" and ASACNOT1 or ASACNOT2 is NA		Missing columns in AE: ASACNOT1, ASACNOT2	AE	
Check for AEs leading to Study Discontinuation but no corresponding record in DS		Missing columns in AE: ASACNOTH	AE, DS	
Page Size: 1 to 10 of 51 H < Page 0 of 0 > H				

SDTM Quality Checks

Full Dataset

AE - Full Dataset

CHECK	Message	Notes
Check for patients with COVID-19 AE indicating drug withdrawal but no Treatment Discon form including AE	Check not run, did not detect COVID-19 preferred terms in AEECDICD	AEECD
AE entries where ASASCTH = "MULTIPLE" and ASACNOT1 or ASACNOT2 is NA	Missing columns in AE: ASACNOT1, ASACNOT2	AE
Check for AEs leading to Study Discontinuation but no corresponding record in DS	Missing columns in AE: ASACNOT1	AE, DS
Check for duplicate USUBID in DM	Missing columns in AE: ASACNOT1	AE, DS
Check for patients with COVID-19 AE indicating drug withdrawal but no Treatment Discon form including AE	Check not run, did not detect COVID-19 preferred terms in AEECDICD	AEECD
AE entries where ASASCTH = "MULTIPLE" and ASACNOT1 or ASACNOT2 is NA	Missing columns in AE: ASACNOT1, ASACNOT2	AE
Check for AEs leading to Study Discontinuation but no corresponding record in DS	Missing columns in AE: ASACNOT1	AE, DS
Check for duplicate USUBID in DM	Missing columns in AE: ASACNOT1	AE, DS

SDTM Quality Checks

Full Dataset

AE - Full Dataset

CHECK	Message	Notes
Check for patients with COVID-19 AE indicating drug withdrawal but no Treatment Discon form including AE	Check not run, did not detect COVID-19 preferred terms in AEECDICD	AEECD
AE entries where ASASCTH = "MULTIPLE" and ASACNOT1 or ASACNOT2 is NA	Missing columns in AE: ASACNOT1, ASACNOT2	AE
Check for AEs leading to Study Discontinuation but no corresponding record in DS	Missing columns in AE: ASACNOT1	AE, DS
Check for duplicate USUBID in DM	Missing columns in AE: ASACNOT1	AE, DS
Check for patients with COVID-19 AE indicating drug withdrawal but no Treatment Discon form including AE	Check not run, did not detect COVID-19 preferred terms in AEECDICD	AEECD
AE entries where ASASCTH = "MULTIPLE" and ASACNOT1 or ASACNOT2 is NA	Missing columns in AE: ASACNOT1, ASACNOT2	AE
Check for AEs leading to Study Discontinuation but no corresponding record in DS	Missing columns in AE: ASACNOT1	AE, DS
Check for duplicate USUBID in DM	Missing columns in AE: ASACNOT1	AE, DS

SDTM Quality Checks

Full Dataset

AE - Full Dataset

CHECK	Message	Notes
Check for patients with COVID-19 AE indicating drug withdrawal but no Treatment Discon form including AE	Check not run, did not detect COVID-19 preferred terms in AEECDICD	AEECD
AE entries where ASASCTH = "MULTIPLE" and ASACNOT1 or ASACNOT2 is NA	Missing columns in AE: ASACNOT1, ASACNOT2	AE
Check for AEs leading to Study Discontinuation but no corresponding record in DS	Missing columns in AE: ASACNOT1	AE, DS
Check for duplicate USUBID in DM	Missing columns in AE: ASACNOT1	AE, DS
Check for patients with COVID-19 AE indicating drug withdrawal but no Treatment Discon form including AE	Check not run, did not detect COVID-19 preferred terms in AEECDICD	AEECD
AE entries where ASASCTH = "MULTIPLE" and ASACNOT1 or ASACNOT2 is NA	Missing columns in AE: ASACNOT1, ASACNOT2	AE
Check for AEs leading to Study Discontinuation but no corresponding record in DS	Missing columns in AE: ASACNOT1	AE, DS
Check for duplicate USUBID in DM	Missing columns in AE: ASACNOT1	AE, DS