



Validation Checks Performed by WebSDM™ on SDTM version 3.1.1 Datasets



Copyright © 2007 Phase Forward Incorporated. All rights reserved. Phase Forward, Lincoln Technologies, their respective logos, and WebSDM are trademarks or registered trademarks of Phase Forward Incorporated in the U.S. Patent and Trademark Office and other jurisdictions. Other marks may be the trademarks or registered trademarks of their respective owners.

Any use of or reliance on these materials is at the relying party's sole risk and will not create any liability or obligation for Phase Forward. The information contained in these materials: (i) is provided without warranty of any kind, express or implied, (ii) applies only to the version of the referenced software that is current at the time of publication, (iii) references only laws, regulations, rules, and technical standards as in effect at the time of publication, and (iv) is subject to change by Phase Forward at any time without notice. This document does not confer any license, implied or otherwise, to any Phase Forward software or other intellectual property, nor does it alter the terms and conditions of any license agreement governing the use of Phase Forward software. Nothing contained in these materials should be construed as the rendering of legal advice, and readers are advised to seek the appropriate advice for specific cases and uses.

UNDER NO CIRCUMSTANCES SHALL PHASE FORWARD BE LIABLE FOR DIRECT, INDIRECT, INCIDENTAL, CONSEQUENTIAL, SPECIAL, EXEMPLARY OR ANY OTHER DAMAGES OF ANY KIND ARISING FROM OR RELATING TO THESE MATERIALS. Some jurisdictions do not permit limitations of liability. In such jurisdictions, the foregoing limitations may not apply to you.

Validation Checks Performed by WebSDM™ on SDTM version 3.1.1 Datasets

The WebSDM™ product, offered by Phase Forward's Lincoln Technologies safety division, is a software product that was developed for use by the FDA under a Cooperative Research and Development Agreement to validate and review data in CDISC SDTM format. The WebSDM™ product may also be licensed for use by pharmaceutical and biotechnology sponsors who are submitting data to the FDA in CDISC SDTM format to support product marketing applications.

The table below lists the current set of validation checks performed by WebSDM™ Version 1.5 on data provided as SAS Transport files according to version 3.1.1 of the [CDISC Study Data Tabulation Model Implementation Guide: Human Clinical Trials](#) (SDTMIG) and the [FDA's Study Data Specifications](#). Additional errors may result from problems with a sponsor-created define.xml file that does not conform to the [CDISC Define.xml](#) schema or conflicts with the CDISC SDTMIG.

In the table below, the "ID" column records the control number applied to each check.

The "APPLIES" column identifies which SDTM object that may be affected by that particular check. A check may apply to all domains, a general class of domains (such as findings), an individual domain, or a set of variables in a given role (such as timing variables as indicated).

The "RULE DESCRIPTION" column explains in human readable form the general logic behind the check performed by the WebSDM™ product.

The "SEVERITY" column explains the relative significance of an anomaly detected by a check, rated LOW, MEDIUM or HIGH. Note there is no direct correlation between a severity value and a decision by FDA reviewers as to whether the data is acceptable for review or not. Instead, Severity is meant only as an indicator of potential problems or anomalies in the data and the error's potential to affect the interpretation or use of the data for specific purposes.

The "COMMENTS" column provides additional relevant information concerning a check. This may indicate whether a check is scheduled to be deprecated in the future.

It is important to note that the actual list of checks used by the FDA will evolve over time to reflect input received from FDA reviewers on the basis of the accumulating experience with SDTM data submissions to help identify potential data quality problems as early as possible in the review cycle to help ensure prompt and thorough reviews. Therefore, sponsors should be cautioned that additional or modified checks may be implemented in the WebSDM product prior to release of a new published version of these checks.

ID	APPLIES	RULE DESCRIPTION	SEVERITY	COMMENTS
IR4000	All	Identifies domain table that has zero rows and therefore contains no data.	Medium	Null tables are normally not submitted.
IR4001	All	Identifies a null (empty) value found in a column where (Standard) Core attribute is 'Req'	High	Required variables may not be null.
IR4002	All	Identifies records where the value for a date doesn't conform to the ISO 8601 standard.	Medium	
IR4003	All	Identifies records where the value in the Domain Abbreviation column doesn't match the name of Domain	Low	
IR4004	All	Identifies records where non-unique values for Sequence Number variable exist within a subject	High	
IR4005	Findings	Identifies subjects where there are no records with a value of 'Y' in the baseline flag variable (Baseline Flag)	Low	
IR4006	LAB	Identifies Short Name of Measurement, Test or Examination values where standard units value (Standard Units) is not consistent across all records	Medium	
IR4007	AE	Identifies records where the value for the Preferred Term could not be found in the MedDRA dictionary	Medium	
IR4008	AE	Identifies records where Serious Event='Y' but none of Involves Cancer, Congenital Anomaly or Birth Defect, Persist or Signif Disability/Incapacity, Results in Death, Requires or Prolongs Hospitalization, Is Life Threatening, Other Medically Important Serious Event, or Occurred with Overdose equals 'Y'	Medium	
IR4009	All	Identifies records where Result or Finding in Original Units and Status both have a value, or where both are null	Medium	
IR4010	All	Identifies records where the value for Visit Number is formatted to more than two decimal places	Medium	
IR4011	DM	Identifies records that violate the condition [If Arm Code='SCRNFAIL' then Description of Arm must equal 'Screen Failure', and vice versa]	High	
IR4012	TA	Identifies records that violate the condition [If Arm Code='SCRNFAIL' then Description of Arm must equal 'Screen Failure', and vice versa]	High	Scheduled for Deprecation since 'Screen Failure' is not an actual ARM.

ID	APPLIES	RULE DESCRIPTION	SEVERITY	COMMENTS
IR4100	All (Timing)	Identifies records that violate the condition [(Study Day of Start of Observation less than or equal to Study Day of End of Observation)], limited to records where [Study Day of Start of Observation is not null and Study Day of End of Observation is not null]	High	
IR4101	All (Timing)	Identifies records that violate the condition [(Start Date/Time of Observation less than or equal to End Date/Time of Observation)], limited to records where [Start Date/Time of Observation is not null and End Date/Time of Observation is not null]	High	
IR4102	Findings	Identifies records that violate the condition [Baseline Flag either 'Y' or null]	Medium	
IR4103	Findings	Identifies records that violate the condition [Derived Flag either 'Y' or null]	Medium	
IR4104	Events, Interventions	Identifies records that violate the condition [End Relative to Reference Period in ('BEFORE', 'DURING', 'AFTER', 'DURING/AFTER', 'U')], limited to records where [End Relative to Reference Period is not null]	Medium	
IR4105	Findings	Identifies records that violate the condition [Fasting Status in ('Y', 'N', 'U')], limited to records where [Fasting Status is not null]	Medium	
IR4106	Events, Interventions	Identifies records that violate the condition [Occurrence in ('Y', 'N')], limited to records where [Occurrence is not null]	Medium	
IR4107	All	Identifies records that violate the condition [Status='NOT DONE'], limited to records where [Status is not null]	Medium	
IR4108	Events, Interventions	Identifies records that violate the condition [Start Relative to Reference Period in ('BEFORE', 'DURING', 'AFTER')], limited to records where [Start Relative to Reference Period is not null]	Medium	
IR4109	Interventions	Identifies records that violate the condition [Dose greater than or equal to 0], limited to records where [Dose is not null]	High	
IR4110	All (Timing)	Identifies records that violate the condition [Duration greater than or equal to 0], limited to records where [Duration is not null]	High	
IR4111	Findings	Identifies records that violate the condition [Result or Finding in Original Units is null], limited to records where [Derived Flag='Y']	Medium	Scheduled for deprecation, since the IG allows sponsors to populate this value if appropriate
IR4112	Findings	Identifies records that violate the condition [Result or Finding in Standard Format is not null], limited to records where [Derived Flag='Y']	Medium	

ID	APPLIES	RULE DESCRIPTION	SEVERITY	COMMENTS
IR4113	Findings	Identifies records that violate the condition [LENGTH(Name of Measurement, Test or Examination) less than or equal to 40 characters]	Low	
IR4114	Findings	Identifies records that violate the condition [LENGTH(Short Name of Measurement, Test or Examination) less than or equal to 8 chars, cannot start with a number or contain special chars]	Low	
IR4115	TS	Identifies records that violate the condition [LENGTH(Trial Summary Parameter) less than or equal to 40 chars]	Low	
IR4116	TS	Identifies records that violate the condition [LENGTH(Trial Summary Parameter Short Name) less than or equal to 8 chars, cannot start with a number or contain special chars]	Low	
IR4117	All (Timing)	Identifies records that violate the condition [End Relative to Reference Period is not null], limited to records where [End Date/Time of Observation is null]	Low	
IR4118	All (Timing)	Identifies records that violate the condition [Start Relative to Reference Period is not null], limited to records where [Start Date/Time of Observation is null]	Low	
IR4119	EX	Identifies records that violate the condition [Planned Elapsed Time from Reference Pt greater than or equal to 0], limited to records where [Planned Elapsed Time from Reference Pt is not null]	High	
IR4120	All (Timing)	Identifies records that violate the condition [Evaluation Interval greater than or equal to 0], limited to records where [Evaluation Interval is not null]	High	
IR4121	Events	Identifies records that violate the condition [Toxicity Grade is a valid number], limited to records where [Toxicity Grade is not null]	Medium	
IR4122	All	Identifies records that violate the condition [Reason Not Done is null], limited to records where [Status is null]	Medium	
IR4123	Findings	Identifies records that violate the condition [Date/Time of Collection is not null], limited to records where [End Date/Time of Observation is not null]	Low	
IR4124	Findings	Identifies records that violate the condition [Date/Time of Collection less than or equal to End Date/Time of Observation], limited to records where [Date/Time of Collection is not null and End Date/Time of Observation exists]	High	

ID	APPLIES	RULE DESCRIPTION	SEVERITY	COMMENTS
IR4125	Findings	Identifies records that violate the condition [Result or Finding in Original UnitsU is not null], limited to records where [Result or Finding in Original Units is not null]	Low	
IR4126	Findings	Identifies records that violate the condition [Result or Finding in Original UnitsU is null], limited to records where [Result or Finding in Original Units is null]	Low	
IR4127	Findings	Identifies records that violate the condition [Normal Range Upper Limit-Standard Units greater than or equal to Normal Range Lower Limit-Standard Units], limited to records where [Normal Range Upper Limit-Standard Units is not null and STNRHI is not null]	High	
IR4128	Findings	Identifies records that violate the condition [Standard Units is not null], limited to records where [Result or Finding in Standard Format is not null]	Low	
IR4129	Findings	Identifies records that violate the condition [Standard Units is null], limited to records where [Result or Finding in Standard Format is null]	Low	
IR4130	All (Timing)	Identifies records that violate the condition [Start Date/Time of Observation is not null], limited to records where [End Date/Time of Observation is not null]	Low	
IR4131	All (Timing)	Identifies records that violate the condition [Planned Time Point Name is not null], limited to records where [Planned Time Point Number is not null]	Low	
IR4132	All (Timing)	Identifies records that violate the condition [Planned Time Point Number is not null], limited to records where [Planned Time Point Name is not null]	Low	
IR4133	All (Timing)	Identifies records that violate the condition [Time Point Reference is not null], limited to records where [Elapsed Time from Reference Point is not null]	Medium	
IR4134	Interventions	Identifies records that violate the condition [Dose Units is not null], limited to records where [Dose is not null]	Low	
IR4135	Findings	Identifies records that violate the condition [Result or Finding in Standard Format is not null], limited to records where [Result or Finding in Original Units is not null]	High	
IR4136	All	Identifies records where values are not found in the study-specific codelist attached to a variable	Low	

ID	APPLIES	RULE DESCRIPTION	SEVERITY	COMMENTS
IR4137	All	Identifies records that violate the condition [Study Day of Visit/Collection/Exam doesn't equal 0]	Low	
IR4250	All	Identifies a column that was described in the domain description but not included in the SAS dataset for that domain	Low	
IR4251	All	Identifies a column where the expected datatype inferred from the description file does not match the actual datatype in the dataset	Medium	
IR4252	All	Identifies a column listed in the domain description as Required ('Req') but not included in the SAS dataset for that domain	High	
IR4253	All	Identifies a column listed in the domain description as Expected ('Exp') but not included in the SAS dataset for that domain	Medium	
IR4254	All	Identifies a column that appears in the SAS dataset but is not listed in the domain description	Medium	
IR4255	AE	Identifies a Sponsor-provided Flag Variable for 'Treatment emergent AE' where the derivation can't be executed	Medium	Relevant when using CTSD safety analysis tools only
IR4256	LB	Identifies a Sponsor-provided Flag Variable for 'Clinically Significant Lab' where the derivation can't be executed	Medium	Relevant when using CTSD safety analysis tools only
IR4257	VS	Identifies a Sponsor-provided Flag Variable for 'Clinically Significant Vital Sign' where the derivation can't be executed	Medium	Relevant when using CTSD safety analysis tools only
IR4258	Supplemental Qualifiers	Identifies a domain that appears to contain supplemental qualifier data but does not contain the Unique Subject Identifier variable	High	
IR4259	All	Identifies a variable where datatype in (study specific) description is not consistent with datatype implicit in SAS dataset	Medium	
IR4260	All	Identifies a variable present in SAS dataset but not present in (study specific) description file	Medium	
IR4500	All	Identifies non-Sequence Number domain subjects not found in the Demographics domain	High	
IR4501	All	Identifies Unique Subject Identifier+Visit Name+Visit Number combinations not found in the SV domain	Medium	
IR4502	DM	Identifies records where the value for Arm Code is not found in the TA domain	Medium	
IR4503	All	Identifies records where the value for Subject Element Code is not found in the TE domain	Medium	

ID	APPLIES	RULE DESCRIPTION	SEVERITY	COMMENTS
IR4504	IE	Identifies records where the value for Inclusion/Exclusion Criterion Short Name is not found in the TI domain	Medium	
IR4505	DM	Identifies Sequence Number subjects where no record for the subject is found in the Disposition domain	High	
IR4506	DM	Identifies Sequence Number subjects where no record for the subject is found in the Exposure domain	High	
IR4507	DM	Identifies Sequence Number treatment arms (Description of Arm+Arm Code combination) not found in the TA domain	Medium	
IR4508	CO	Identifies CO domain reference to an unknown related domain	High	
IR4509	RELREC	Identifies Related Records domain reference to an unknown related domain	High	
IR4510	Supplemental Qualifiers	Identifies Supplemental Qualifiers domain reference to an unknown related domain	High	
IR4511	RELREC	Identifies Related Records domain reference to a key variable that isn't defined in the target domain	High	
IR4512	Supplemental Qualifiers	Identifies Supplemental Qualifiers domain reference to a key variable that isn't defined in the target domain	High	
IR4513	RELREC	Identifies Related Records domain reference to a record that doesn't exist in the target domain	High	
IR4514	Supplemental Qualifiers	Identifies Supplemental Qualifiers domain reference to a record that doesn't exist in the target domain	High	
R4005	DM	Identifies records where values for Unique Subject ID variable(s) are not unique, limited to records where [Unique Subject ID is not null]	Medium	
R4006	DM	Identifies records that violate the condition [AGE greater than or equal to 0], limited to records where [AGE is not null]	High	
R4007	DM	Identifies records where value for [SEX] is not found in Codelist [SEX]	Medium	
R4008	DM	Identifies records where value for [COUNTRY] is not found in Codelist [COUNTRY]	Medium	
R4019	AE	Identifies records where value for [Serious Event] is not found in Codelist [YESNO]	Medium	
R4023	AE	Identifies records where value for [Congenital Anomaly or Birth Defect] is not found in Codelist [YESNO], limited to records where	Medium	

ID	APPLIES	RULE DESCRIPTION	SEVERITY	COMMENTS
		[Congenital Anomaly or Birth Defect is not null]		
R4024	AE	Identifies records where value for [Persist or Signif Disability/Incapacity] is not found in Codelist [YESNO], limited to records where [Persist or Signif Disability/Incapacity is not null]	Medium	
R4025	AE	Identifies records where value for [Results in Death] is not found in Codelist [YESNO], limited to records where [Results in Death is not null]	Medium	
R4026	AE	Identifies records where value for [Requires or Prolongs Hospitalization] is not found in Codelist [YESNO], limited to records where [Requires or Prolongs Hospitalization is not null]	Medium	
R4027	AE	Identifies records where value for [Is Life Threatening] is not found in Codelist [YESNO], limited to records where [Is Life Threatening is not null]	Medium	
R4031	IE	Identifies records where value for [Inclusion/Exclusion Category] is not found in Codelist [INCEX], limited to records where [Inclusion/Exclusion Category is not null]	Medium	
R4043	AE	Identifies records where value for [Concomitant or Additional Trtmnt Given] is not found in Codelist [YESNO]	Medium	
R4045	AE	Identifies records where value for [Involves Cancer] is not found in Codelist [YESNO], limited to records where [Involves Cancer is not null]	Medium	
R4046	AE	Identifies records where value for [Other Medically Important Serious Event] is not found in Codelist [YESNO], limited to records where [Other Medically Important Serious Event is not null]	Medium	
R4047	AE	Identifies records where value for [Occurred with Overdose] is not found in Codelist [YESNO], limited to records where [Occurred with Overdose is not null]	Medium	
R4062	DM	Identifies records where value for [AGEU] is not found in Codelist [AGEUNITS2], limited to records where [AGEU is not null]	Medium	
R4071	IE	Identifies records where value for [I/E Criterion Original Result] is not found in Codelist [YESNO], limited to records where [I/E Criterion Original Result is not null]	Medium	
R4072	IE	Identifies records where value for [I/E Criterion Result in Std Format] is not found in Codelist [YESNO], limited to records where [I/E Criterion Result in Std Format is not null]	Medium	

ID	APPLIES	RULE DESCRIPTION	SEVERITY	COMMENTS
R4073	IE	Identifies records that violate the condition [I/E Criterion Original Result = I/E Criterion Result in Std Format]	Low	
R4083	Supplemental Qualifiers	Identifies records where values for [Study Identifier, Unique Subject Identifier, Identifying Variable, Identifying Variable Value, Qualifier Variable Name] variable(s) are not unique	Medium	
R4096	DM	Identifies records that violate the condition [Subject Reference Start Date/Time is not null], limited to records where [Arm Code doesn't equal 'SCRNFAIL']	High	
R4097	DM	Identifies records that violate the condition [Subject Reference End Date/Time is not null], limited to records where [Arm Code doesn't equal 'SCRNFAIL']	High	
R4101	TE	Identifies records that violate the condition [Rule for End of Element is not null or Planned Duration of Element is not null]	Medium	
R4102	AE	Identifies records that violate the condition [Results in Death='Y'], limited to records where [Outcome of Adverse Event='FATAL']	Medium	
R4103	AE	Identifies records that violate the condition [Outcome of Adverse Event='FATAL'], limited to records where [Results in Death='Y']	Medium	
R4104	DV	Identifies records that violate the condition [count (distinct Protocol Deviation Term) = count(distinct Protocol Deviation Coded Term)], limited to records where [Protocol Deviation Term is not null]	Medium	
R4105	SE	Identifies records that violate the condition [Description of Unplanned Element is not null], limited to records where [Subject Element Code='UNPLAN']	Medium	
R4106	DM	Identifies records that violate the condition [AGEU is not null], limited to records where [AGE is not null]	Low	