



申請電子データのバリデーションエラー 説明事例の集計結果・事例紹介

PHUSE Japan working group / CJUG ADaM,SDTM

Presenters:

Shinichi Hotta (Sumitomo Dainippon Pharma)

Yasuhiro Iijima (Novartis Pharma K.K.)

Taku Sakaue (CHUGAI PHARMACEUTICAL CO., LTD.)



Disclaimer

本プレゼンテーションの内容は、発表者としての見解であり、所属する組織の意見や立場を示すものではありません。



Agenda

- Pinnacle 21のバリデーションエラー説明事例集とは
- 集計結果
- 個々の事例紹介
- Discussion



Pinnacle 21のバリデーションエラー説明事例集の作成

- 2020年7月から9月にかけて、CJUG ADaMおよびPHUSE Japan e-dataチームで、PMDAに申請した試験のxDRGに記載した、または、PMDAに確認相談を行ったバリデーション結果の説明の事例を収集した。
 - 合計20社から回答を得た。
- バリデーション結果は、以下の方法でSDTM、ADaM、defineの3種類に分類した。
 - define.xmlに関連するものを「define」。
 - その他のものでSDRGに記載するものは「SDTM」、ADRGに記載するものは「ADaM」。



SDTMのError（総数114個）のID毎の頻度の集計結果（上位5位、同順位含む）

RULE ID	MESSAGE	COUNT
SD2239	Inconsistent value for -- TPT	20
SD1212	STRESN does not equal --STRESC	12
SD1206	EXSTDTC date is after RFXENDTC	7
SD0035	Missing value for -- DOSU, when -- DOSE, -- DOSTXT or -- DOSTOT is provided	6
CT2001	<Variable Name> value not found in '<Codelist Name>' non-extensible codelist	5
CT2003	<Variable Name> and <DecodedVariable Name> values do not have the same Code in CDISC CT	5
SD0036	Missing value for -- STRESC, when -- ORRES is provided	5
SD1207	EXENDTC date is after RFXENDTC	5



CJUG ADaM Team

SDTMのError（総数114個）の対象のデータセット毎の頻度の集計結果（上位10位）

Datase ts	COUNT
LB	18
EX	16
TS	12
PP	9
CM	8
VS	8
PC	6
AE	5
CO	5
DM	5



ADaMのError（総数131個）のID毎の頻度の集計結果（上位5位、同順位含む）

RULE ID	MESSAGE	COUNT
AD0047	Required variable is not present	26
AD0198	Neither AVAL nor AVALC are present in dataset	20
AD0154	Multiple baseline records exist for a unique USUBJID, PARAMCD, BASETYPE	7
AD0196	Required Variable value is null	6
AD0059	*TM is not a numeric variable	4
AD0161	AWU is populated but both AWLO and AHWI are not populated	4
AD0172	Subject is on treatment (ONTRTFL), but analysis date (ADT) is after treatment end date (ADSL.TRTEDT)	4
CT2001	Variable value not found in non-extensible codelist	4



defineのError（総数33個）のID毎の頻度の集計結果（上位5位、同順位含む）

RULE ID	MESSAGE	COUNT
DD0039	Variable is in wrong order within Dataset <dataset>	12
DD0032	Missing NCI Code for Term in Codelist <codelist>	5
DD0033	Unknown NCI Code value for Codelist <codelist>	3
DD0074	Variable/Value Level Origin Type mismatch	3
DD0031	Missing NCI Code for Codelist <codelist>	2
DD0055	Invalid Class value	2
DD0059	Define.xml/CDISC dataset Description mismatch	2

Warningとなったバリデーション結果

- 今回の集計対象には含めていないが、アンケートではWarningに該当したバリデーション結果も収集された。
- PMDA申請では、Warningに関してxDRGに記載することは必須ではない。
- xDRGに記載するかどうかは、会社によって対応が異なっている。



CJUG ADaM Team

- 多かったエラーの紹介

Rule ID SD2239 : Message Inconsistent value for – TPT

Description Planned Time Point Name (--TPT) value must be consistent for all records with same Subject (USUBJID) and Assessment Date/Time (--DTC).

主な理由

- ✓ 同じ日時(--DTC)で違う項目の測定
- ✓ 時刻を収集していないのでユニークになっていない

その他

- ✓ データ入力ミス→解析に影響しないもの

VISIT	VISITNUM	--TPT	--TPTNUM	--TPTREF
PERIOD 1	3	PRE-DOSE	1	DAY 1, AM DOSE
		1H	2	
		4H	3	
		PRE-DOSE	1	DAY 1, PM DOSE
		1H	2	
		4H	3	
		PRE-DOSE	1	DAY 5, AM DOSE
		1H	2	
		4H	3	
		PRE-DOSE	1	DAY 5, PM DOSE
		1H	2	
		4H	3	

VSTPTNUM	VSTPT	VSELTMT	VSTPTREF	VSRFTDTC	VSDTC
1	30 MIN	PT30M	DOSE ADMINISTRATION	2006-08-01T08:00	2006-08-01T08:30
2	60 MIN	PT1H	DOSE ADMINISTRATION	2006-08-01T08:00	2006-08-01T09:01
3	90 MIN	PT1H30M	DOSE ADMINISTRATION	2006-08-01T08:00	2006-08-01T09:32

Rule ID SD1212 : Message --STRESN does not equal --STRESC

Description Standardized Result in Numeric Format (--STRESN) variable value should be equal Standardized Result in Character Format (--STRESC) variable value, when Standardized Result in Character Format (--STRESC) variable value represents a numeric value.

主な理由

- ✓ 丸め誤差
- 1.--STRESCには、丸めた値を格納しているため
- 2.XPTのフォーマットに起因するエラー

LBSTRESC	LBSTRESN
60.0	60.0
MODERATE	
12.1	12.1
TRACE	
1+	
BLQ	
33.8	33.8

Rule ID Message	CT2001 :<Variable Name> value not found in '<Codelist Name>' non-extensible codelist
Description	Variable must be populated with terms from its CDISC controlled terminology codelist. New terms cannot be added into non-extensible codelists.
主な理由	✓ CRFでCTでないコードで取得されてしまっている ✓ 解析でそのコードが使われている

Rule ID	SD1206 :
Message	EXSTDTC date is after RFXENDTC
	SD1207 :
	EXENDTC date is after RFXENDTC
Description	Start Date/Time of Treatment (EXSTDTC) variable value must be less than or equal to Date/Time of Last Study Treatment (RFXENDTC). End Date/Time of Treatment (EXENDTC) variable value must be less than or equal to Date/Time of Last Study Treatment (RFXENDTC).
主な理由	✓ 最終投与日(EXSTDTC or EXENDTC)がEXDOSE= 0 である (休薬を除いた最終投与日の日付をRFXENDTCに格納しているケース)

Rule ID Message	SD0035 : Missing value for --DOSU, when --DOSE, --DOSTXT or --DOSTOT is provided
Description	Dose Units (--DOSU) must be populated, when Dose per Administration (--DOSE), Dose Description (--DOSTXT) or Total Daily Dose (--DOSTOT) is provided.
主な理由	✓ 単位がない項目 ✓ CRFで単位を収集していない ✓ 単位はECDOSTXT(Dose Description)に入れている

Rule ID SD0036 : Missing value for --STRESC, Message when --ORRES is provided

Description Character Result/Finding in Std Units (--STRESC) must be populated, when Result or Finding in Original Units (--ORRES) is provided.

主な理由 ✓ (換算係数がないなどの理由で)算出できなかった

LBORRES	LBSTRESC
30	3.0
398	
350	350

Rule ID	AD0047 : Required variable is not present
Message	AD0198 : Neither AVAL nor AVALC are present in dataset
Description	Variables described in ADaM as Required must be included in the dataset. At least one analysis value (numeric or character) must be present in BDS.
主な理由	✓ BDSデータセットではない(OCCDS/OTHER)が、BDS データセットと認識されて出るエラー

Rule ID AD0196 : Required Variable value is null Message

Description Unlike SDTM, only some of ADaM Required variables (where Core attribute is 'Req') must be populated. CDISC ADaM rules affirm that PARAM and PARAMCD must be populated, while others are populated as appropriate. Known Exceptions: TRTP, TRTxxP, AGE, AGEU, RACE, SUBJID, AESER.

主な理由

- ✓ Screen failureのケースで、SDTM IGではなくFDAのTCGに従っているのでARM, ARMCD等が空欄
- ✓ 何らかの理由でコーディングできていない (AEBODSYS, AEDECOD)

ARMCD	ARM
A	Drug A
P	Placebo
P	Placebo

AETERM	AEDECOD	AEBODSYS
NAUSEA	Nausea	Gastrointestinal disorders
VOMITING	Vomiting	Gastrointestinal disorders
HEADACHE		

Rule ID AD0059 : *TM is not a numeric variable Message

Description The SAS TYPE for this variable does not equal Num.

主な理由 SDTMIGでは文字型となっているのでそれに従った

PCELTM	PCELTM	Planned Elapsed Time from Time Point Ref	Char
PT24H			
PT24H			
-PT15M			

Rule ID	AD0172 :
Message	Subject is on treatment (ONTRTFL), but analysis date (ADT) is after treatment end date (ADSL.TRTEDT)
Description	If ONTRTFL =Y then TRTSDT <=ADT <= TRTEDT should be true (analysis must be between start and end range). Ensure subject exists in both files.
主な理由	解析計画書でOn Treatmentの定義を最終投与日 (TRTEDT)+ xx後としているため

Rule ID Message	DD0039 : Variable is in wrong order within Dataset <dataset>
Description	Order of variables must follow the order specified by CDISC standard. See section 4.1.1.4 of the SDTM-IG for variable ordering requirements for the SDTM or section 3 of the ADaM-IG for information about variable ordering for ADaM datasets.
主な理由	✓ False Positiveで出ているケースが多い PMDA Validation Rules Version 2.0 では削除されているルールです



CJUG ADaM Team

- DISCUSSION

どのような説明内容が良いか？

技術的ガイド

(ア) 違反している場合には修正されるまで審査が開始されないルール

各臨床試験について必須であるデータセットの存在等の非常に基本的なルール

(イ) 違反している場合に、事前に説明がない場合には修正されるまで審査が開始されないルール

多くの場合、各標準及び実装ガイドに明確に定められているルールであり、違反している場合にはその理由及び修正が不可能な理由について申請前に PMDA に相談し、さらにデータガイドにおいて説明する必要のあるルール

(ウ) 違反した場合でも必ずしもその説明を求めないルール

(ウ)についても、臨床試験データの質の観点から、違反の理由について別途説明を求める場合がある。

FAQ

Q1-23 実務的通知 2(4)イに基づき、修正が不可能な Error に相当するバリデーションルール違反に関する必要な説明をデータガイドや「対面助言、証明確認調査等の実施要綱通知」の別紙 8「申請電子データ提出確認相談資料様式」の「電子データの CDISC 適合性に関する情報」欄に記入する際、留意点はありますか。

A: 修正が不可能な Error に相当するバリデーションルール違反に関して、PMDA では、相談者の説明内容に基づいて審査時に実施する解析への影響の程度を検討しています。例えば、「CRF で収集されたとおりにデータを格納した」や「(相談者側の)解析に影響しないため、修正は行わない」といった記載では、具体的なデータの状況が把握できず、審査時のデータ利用に影響があるかどうかの判断ができません。したがって、バリデーションルール違反の原因となったデータの実際の格納内容を記載してください。また、相談者側の解析に影響がないと説明する場合、その判断に至った理由を記載してください。なお、当該バリデーションルール違反の修正が不可能な理由についても漏れなく記入してください。

どのような説明内容が良いか？

- **SD1212 : --STRESN does not equal –STRESC**

既知のものは以下のような簡易的な説明でも良さそうだが…

- ✓ --STRESCには、丸めた値を格納しているため
 - ✓ XPTのフォーマットに起因するエラー
 - ✓ Pinnacle21には--STRESNの有効数字の桁数のチェックに制限があるため
- etc.

どのような説明内容が良いか？

• DD0039 : Variable is in wrong order within Dataset <dataset>

例1.丁寧な例

違反が検出された理由：

SVドメインの変数の順序が、Pinnacle21 Community V2.1.3のメタデータにて設定されているSVドメインの変数の順序と異なる。

<Pinnacle21 Community V2.1.3の設定>

..., VISIT, VISITDY, SVSTDTC, (TAETORD,) SVENDTC, EPOCH, ...

※カッコ()内は本試験で使用していない変数

<本試験のSVドメイン>

..., VISIT, VISITDY, EPOCH, SVSTDTC, SVENDTC, ...

修正しない理由：

本試験のSVドメインは、SDTM v1.4の以下の表における順序に準拠しているため。

・「Table 2.2.5: All Observation Classes — Timing Variables」において、VISIT, VISITDY, EPOCH, --STDTC, --ENDTCの順序で記載されている

・「Table 2.2.9: Subject Visits — One Record per Subject Visit, per Subject」において、VISIT, VISITDY, SVSTDTC,

SVENDTCの順序で記載されている

例2.簡易的な例

SDTM IGに従っているので、問題ないと判断した。

IGに従っている等、理由が明確なものは簡潔で良い？

どのような説明内容が良いか？

- **CT2001 :<Variable Name> value not found in '<Codelist Name>' non-extensible codelist**

例：拡張コード利用不可の変数であるが、"XXXX"の理由から"XXXX"を設定した

DVOUT value not found in 'Outcome of Event' non-extensible codelist /

Outcome of Event (OUT) のコードリストは、 non-extensible codelistですが、 Protocol

ViolationsのOutcomeとして、 CRFに以下のデータを収集していることから、 Errorとして検出されています。試験デザインに基づいて、 Outcome of Event (OUT) に以下のデータを追加しているため、 問題ないと判断しました。

- APPROVED TO CONTINUE
- ACKNOWLEDGED, NO ACTION TAKEN. SUBJECT PARTICIPATION COMPLETED
- PARTICIPATION TERMINATED

!

レガシーデータなどCSR
に使われている場合どの
ようにすべきか悩ましい

...

どのような説明内容が良いか？

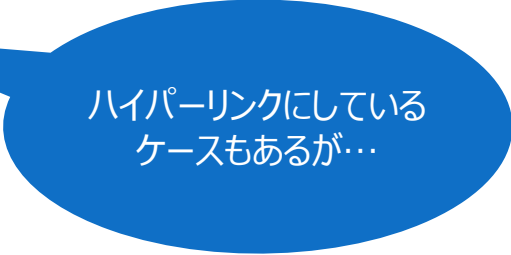
- Pinnacle21 のフォーラムを参照としたい場合

例 1. Pinnacle21のバグであると考えられる

Pinnacle Forum (posted on November 19, 2018)

<https://www.pinnacle21.com/forum/ct2001-sustrtpt-value-not-found-relation-reference-period-strtpt-non-extensible-codelist>

例 2. これは既知のエラーである



ハイパーリンクにしている
ケースもあるが…

どこまでデータを修正すべきか？

- Ongoingの試験では、各社CDISC対応が進んでいる一方で、レガシーデータの申請電子データ提出に関しては、どこまでCDISCに準拠するかは未だ各社様々で、課題であると考えている

まとめ

- よくある事例については、説明内容は簡潔にできることもある
- 同じ問題で悩むことがないようにするために、事例共有は有用

今後継続的に情報を共有していくことで、
より効率的な申請電子データ提出を目指しましょ
う！