

統合解析、レガシーデータ変換に関する UseCase報告

2021/2/15

日本CRO協会 CDISC検討チーム

株式会社インテリム

小林 俊木

発表の目的

- CRO各社に業務経験が蓄積されている。そこで統合解析およびレガシーデータ変換を中心に申請電子データ作成時に直面した問題事例をCRO協会加盟各社で収集し、事例集を作成している。
- 本発表では、作成中の事例集のうち、一部の事例に対し、CRO協会 データサイエンスWG CDISC検討チームで解決策を検討したので、各社業務時の参考にして頂きたい。

免責事項

- 本資料の内容は、日本CRO協会データサイエンスWG CDISC検討チームの見解に基づくものであり、協会に加盟している全CROの公式見解を示すものではありません。
- 業務委託内容の詳細・対応方針については、各CROと相談し、ご確認ください。

Agenda

- 背景
- 事例集の作成
- 一部事例に対する解決策
 - レガシーデータ変換に関する事例
 - 統合解析に関する事例
 - 申請全般に関する事例
- まとめ

背景

- 経過措置期間を終え、昨年から申請電子データの提出義務化が始まった。そのためレガシーデータのCDISC変換業務や統合解析(ISS/ISE)が今後増えると考えられる。
- CRO協会加盟各社はそれら業務に関する経験を積んできており、各社の経験事例をまとめる事にした。

事例集の作成

- CRO協会加盟各社に対し、下記2つのテーマで各社が業務で苦労した事例を収集している。また、収集した事例に対しての解決策を検討中である。
 - 1、レガシーデータ変換
 - 2、統合解析
- 現時点で収集している事例は、後程紹介する『CDISC業務の円滑な委受託関係構築に向けたチェックリスト 更新版』の別添資料として公開します。将来的には、解決策も記載して、正式公開予定。

一部事例に対する解決策

※提唱する解決策はCRO協会データサイエンスWG CDISC検討チームが提案するものであり、CRO各社およびPMDAの見解ではございませんので、ご留意ください。

レガシーデータ変換に関する事例 - ①

事例

- CDISCが考慮されていない時期の試験データをCDISC変換する際、Terminologyの記載用語に完全一致しないデータの取り扱いについて協議に時間を要した
- 自由記載で収集された項目もTerminology記載用語に読替えるべきか関係者間で協議をした

CRFで『AEの発現部位』を自由記載で収集

症例番号	有害事象名	発現部位
01-001	Cataract	Eye
01-002	Ophthalmia	Augen



SDTM変換で『EYE』に読替えるか？

Augen = 眼球

以下、SDTM Terminologyより抜粋

Codelist Name	CDISC Submission Value	CDISC Synonym(s)
Anatomical Location	EYE	Eyeball

レガシーデータ変換に関する事例 - ①

背景に応じて、対応を考えた方が良い。例えば、

- 当時の解析結果の再現性をとる場合、読替える事で再現性に影響が出るようであれば、読替えない方法もある。ただし、Pinnacle21によるバリデーションがERRORになる場合はPMDAとの相談を行う。
- RawデータからCDISC変換するのであれば、読替える事を推奨する。(早期に読み替え対象データを特定し、読替え内容について関係者間で協議すべきである。)

レガシーデータ変換に関する事例 - ②

事例

Rawデータが入手できない場合、Pinnacle21バリデーションでの各Issueの詳細な原因が特定できない事があり、Reviewer's Guideの説明が困難になった。



レガシー変換



Pinnacleバリ
デーション

非SDTM形式でまとめられたデータセット

CDISC標準データセット

Pinnacleバリデーションにより、Issueが発生

例1: Issueの内容的にRawデータを非SDTM形式に加工した際の不備だと思われるが、Rawデータがないため、特定ができない

例2: 根拠データがなく、またプロトコルからも判断できず、Req変数に値を格納できないことでIssueが発生

レガシーデータ変換に関する事例 - ②

Rawデータが入手できない事により生じる申請データ作成時の懸念事項や申請時の想定対応についてはPMDAと事前に相談をすることが望ましい。

懸念事項や想定対応の具体例

- ・根拠となるデータがなく、Req変数に値を格納できない
- ・Reviewer's GuideでのError, Warningの詳細説明の記載内容

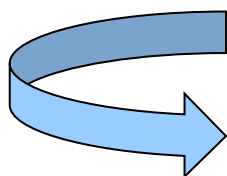
その相談結果に従った対応をした上で、Rawデータが入手できない旨は、Reviewer's Guideにも記載すべきである。

レガシーデータ変換に関する事例 - ③

事例

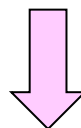
運用時のデータレビューが不十分だったのか日付データに実在しない日付 (e.g. 4/31) があった。
どのような対応が望ましいのか関係者間で協議が必要となった。

レガシー変換



Rawデータセット

作成したSDTMデータセットに
Pinnacleバリデーションを実行



AESTDTC	AESTDY
2000-04-31	

左記のような実在しない日付があったため、以下のWarningが発生し、不明日付の存在が発覚

Missing --STDY variable, when --STDTC variable is present

レガシーデータ変換に関する事例 - ③

背景やデータ値により対応を検討すべきである。

例1: 当時の解析等の計画書・仕様書で読替えを行う事が明記されている場合

2000-04-31  2000-05-01

解析計画書に明記

- ➔ CDISC変換時にその記載に従った読替えを行う。その際、Reviewer's Guideに経緯や格納アルゴリズムの記載が必須である。

例2: 当時の計画書等に読替えに関する記載がない場合

- ➔ 早期に関係者間で協議の上、対応方針を決定し、Reviewer's Guideに対応を記載する。

レガシーデータ変換に関する事例 - ④

事例

当時のCRFや解析に用いた仕様書類や解析プログラムがなく、**データセットのみ存在**している場合、どのように対応すべきか。

当時の解析結果の再現性確認ができない可能性があることを、事前に関係者間で合意すべきである。

その上で、CDISC変換開始前に、以下を行うべきである。

- 再現性をとれる **変換プロセスの検討**
- データ **トレーサビリティ確保** の検証

また、必要に応じてPMDAと相談し、再現性確認までのプロセスを明確にした上で、CDISC変換を行うべきである。

解決策

統合解析に関する事例 - ①

事例

- 統合元の試験毎にデータセットの仕様やコード値の定義が異なり、ミスを誘発した。
- 試験毎で統一されていない変数値の統一化や再割り当てに時間を要した。また、解析用データの構造が異なっているケースがあり、苦労した。

試験AのADLB

PARAM	PARAMN
Leukocytes	1
Hemoglobin	2
Hematocrit	3
Platelets	4

Hemoglobinの有無により、試験間でPARAMNが異なっていた



試験BのADLB

PARAM	PARAMN
Leukocytes	1
Hematocrit	2
Platelets	3

統合解析に関する事例 - ①

解決策

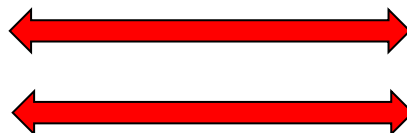
各試験の仕様を試験横断的に管理すべきである。
CTは個々の試験単位ではなく、全試験で統一する。
例えば、B試験ではPARAMN=2のデータを使わないとなったら、PARAMNを変えずに1,3,4のように歯抜けで定義する等、各試験で統合しやすい仕様を心掛けるべきである。

試験AのADLB

PARAM	PARAMN
Leukocytes	1
Hemoglobin	2
Hematocrit	3
Platelets	4

試験BのADLB

PARAM	PARAMN
Leukocytes	1
2は欠番	
Hematocrit	3
Platelets	4



統合解析に関する事例 - ②

免除相談にて、個別試験が非CDISC形式となっている



統合解析 (ISS) を実施するため、統合解析用ADaMを作成



統合解析用ADaMに対し、Pinnacleバリデーションを実施



日付の不整合 (開始日・終了日が逆転) のIssueを発見



個別試験でのデータクリーニング不備かと思われるが、
統合解析用ADaMにはどのような対応をすればよいか

事
例

統合解析に関する事例 - ②

状況に応じて、下記2通りの方法がある。

- 1、統合解析で使用しない変数であれば、統合解析用ADaMに含めないという選択もある。
- 2、統合解析で使用する変数の場合は、可能なかぎり原因を調査し、統合解析用ADaMで読替えるか検討すべきである。

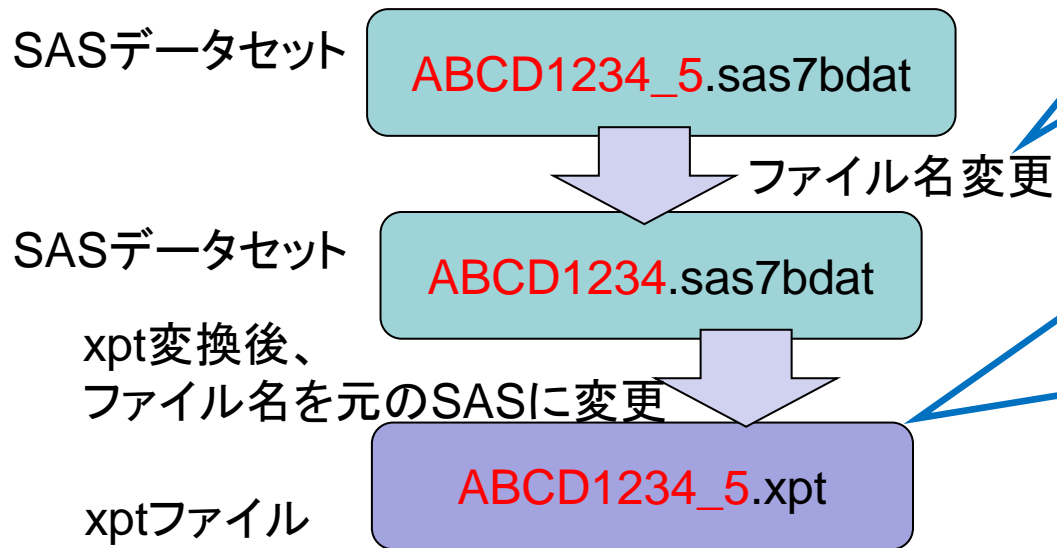
なお、読替えを実施する場合、経緯や内容についてはDefineもしくはReviewer's Guideにて定義する必要がある。

申請全般に関する事例 - ①

事例

臨床薬理領域においてときどき見られるが、XPT形式のファイル名が9文字以上の場合、変換したSASデータセット名が異なるため、Gateway提出後の照会事項となった。

ファイル名を8文字に変更し、変換先のXPTのファイル名を変更前のファイル名にして出力したケース



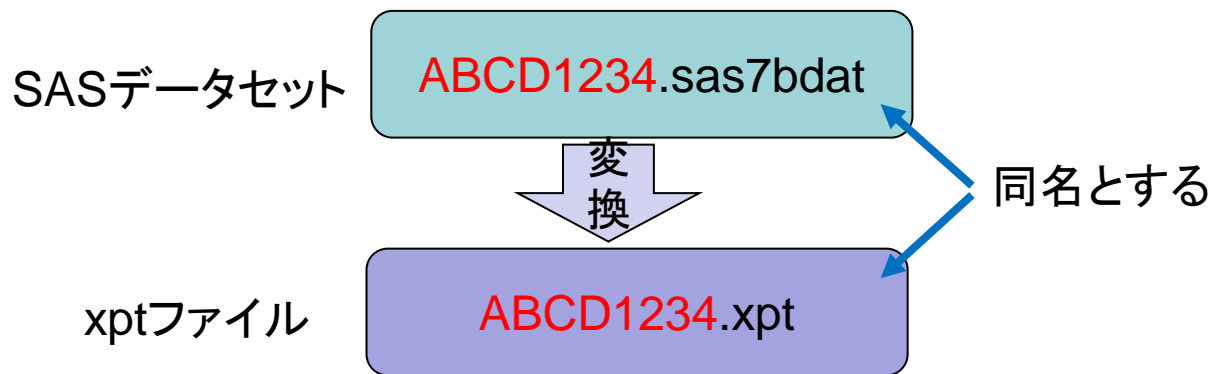
ファイル名が9文字以上の場合、xpt変換ができないので、xpt変換用にファイル名を変更

SAS形式のデータセットからxpt形式に変換するとき、別名での変換はできるが、電子データ提出ファイルとしては**NG**
(技術的ガイド4.1.1.4及び4.2.2参照)

申請全般に関する事例 - ①

解決策

XPTのファイル名を8文字にし、オリジナルファイル名との読み替え表を合わせて提出すべきである。



<読み替え表>

元のファイル名	変更したファイル名	変更理由
ABCD1234_5.xpt	ABCD1234.xpt	データセット名と異なるため
:	:	

申請全般に関する事例 - ②

事例

Gateway提出時にDescription欄に記載する場合は最大100文字であるが、100文字以内にした場合でもGateway提出時にエラーが出力される場合がある。

解決策

原因:

99文字など制限文字数に近い設定をした場合、何らかの理由(ファイル操作等)により、文字数が増え、結果100文字を超えるものと思われる。

対応策:

記載する文字数を90文字以内にするなど、余裕もった記載にすべきである。

まとめ

- 申請電子データの義務化に向けた経過措置期間を通じ、業界全体で、データの標準化が進んだが、レガシーデータ変換や統合解析等に関しては、各社業務時に想定外の問題が起きることが多くある。

➡ 各社が経験した事例集をご活用頂けますと幸いです。

- レガシーデータ変換や統合解析等については、試験背景を考慮した臨機応変な対応が重要である。

本日紹介した事例集は、CRO協会ホームページにて公開予定ですので、ぜひご覧ください。

CRO協会からのお知らせ

CDISC業務の円滑な委受託関係構築に向けた チェックリスト 更新版

- 品質マネジメントの観点から考えた委受託業務の基本の観点到分類
 - 最終成果物(サービス)の特定と要求品質と納期の明確化
 - 要求品質を満たす最終成果物(サービス)を提供、調達できるための、効果的かつ効率的なプロセスの確立

GSCPP中項目	実施しないことによる影響	備考
業務委託契約締結前	■同一パッケージの場合など、他試験とのバージョンの差異などが原因となる場合がある。	要求品質の一側

近日公開いたします

URL: http://www.jcroa.or.jp/member/report/info_datascience.html

(日本CRO協会HPホーム > 協会について > 協会事業> データサイエンスWG)

最終OBL時、解析時	■固定条件が試験の場合、追加的な修正変更が発生する可能性がある。	サポートの期限について 申請時に万一対応事項が発生した場合を踏まえた契約になっていない場合、再契約が必要になり、迅速な対応ができず申請スケジュールに影響が出る可能性がある。
追加業務に関するスケジュールおよび費用の連絡経路、期限を合意する。	■追加作業発生連絡の遅れにより、試験全体のスケジュールに影響する可能性がある。	
作品後のサポートの期限を明確にする。 - 納品まで(CDISのデータ作成完了まで) - 申請まで - 照会事項対応まで	■ゲートウェイシステムを介した電子申請時の提出エラーや申請後の照会事項が発生した場合、契約内容によっては対応の要否・可否に影響が出る可能性がある。	固定タイミングについて 成果物の固定タイミングを予め明確にしないことと承認対応に必要なスケジュールが確保できない、また固定となるタイミングが明確になっていない場合、どの時点からの仕様変更が対応可能かの検討が必要となり、表裏社間で認識違いによる可能性がある。
成果物の固定タイミング(試験のマイルストーンに応じて)を合意する。	■成果物の固定タイミングが不明確となる。 - 承認プロセスのための作業スケジュールが十分に確保できない。 - 請求時期の予測が立てられない。	
業務開始時 担当者の役割を明確にする。(Who's whoの作成など)	■特にCDMでは、通常の窓口とSDTMの窓口を別々の担当者が担っている場合があり、問い合わせ先が分からなくなる場合がある。担当者が交代した際なども注意が必要。	レビュー時期およびレビュー観点について 同じ対応可能な現実的スケジュールを確保すること、レビュー頻度についても、当該マイルストーンに要求される品質を設定し、それを基にQCチェックの要否などを決定する。
試験のマイルストーンを共有し、各成果物のレビュー時期およびレビュー観点を合意する。追加や変更があった場合も速やかに連絡する。 外部データに関する情報を共有する。 - ベンダーとの契約内容の明確化 - ベンダーとの窓口(問い合わせ窓口)	■事前に合意した業務とは異なるマイルストーン(例えばOC)のためのデータ提出が見込まれる場合は、早期で連絡し、且対応できない可能性がある。 ■追加作業発生などによる、試験スケジュールの遅延が懸念される。	

謝辞

本発表を纏めるにあたり、多くの意見を頂いた
日本CRO協会データサイエンスWG CDISC検討チー
ムの皆様に感謝致します。

- ◆ 牧野 奈緒 (イーピーエス株式会社)
- ◆ 水上 七帆 (イーピーエス株式会社)
- ◆ 白石 友太朗 (エイツーヘルスケア株式会社)
- ◆ 湯川 智仁 (エイツーヘルスケア株式会社)
- ◆ 新城 博子 (エイツーヘルスケア株式会社)
- ◆ 田畑 一作 (シミック株式会社)
- ◆ 田野 貴洋 (パレクセル・インターナショナル株式会社)
- ◆ 阿知良 周 (PRAヘルスサイエンス株式会社)
- ◆ 小阪 英明 (株式会社CACクロア)
- ◆ 小野島 匠 (株式会社CACクロア)
- ◆ 薄 美有 (株式会社新日本科学PPD)

(敬称略、所属名順)

ご清聴ありがとうございました