

臨床試験データと CDISC 標準メタデータとの整合性を 検証するためのオープンソースソフトウェアの開発

岡田 昌史^{*1}

^{*1} 東京大学医学部附属病院 大学病院医療情報ネットワーク研究センター

Developing an open source tool to validate consistency of clinical trial data and its metadata definition

Masafumi Okada ^{*1}

^{*1} University Hospital Medical Information Network Research Center,
University of Tokyo Hospital

抄録:

臨床試験において収集されるデータのトレーサビリティおよび再利用性の向上のために、国際標準としてメタデータの記述方法、および典型的なメタデータの集合を定義した規格として、Clinical Data Interchange Standards Consortium (CDISC) による標準規格がある。しかし、CDISC に準拠した形式で記述したメタデータと実際に収集されたデータとの整合性を検証することができるソフトウェアが非常に少ないこともあり、日本のアカデミアにおける臨床試験、臨床研究においては、標準規格の導入による業務効率化が普及しているとは言い難い状況である。本研究では CDISC による標準規格を利用した臨床試験データの品質向上活動を普及させることを目的として、(1) CDISC Define-XML 形式で記述したメタデータと実際に収集されたデータとの整合性を検証する機能 (2) CDISC SDTM Conformance Rule のうち自動検査が可能なルールと実際に作成されたデータセットとの整合性を検証する機能 の2つを備えたソフトウェアをオープンソースソフトウェアとして開発したので報告する。

キーワード CDISC, 臨床試験, R, メタデータ, データマネジメント

1. はじめに

臨床試験において収集されるデータのトレーサビリティおよび再利用性の向上のために、国際標準としてメタデータの記述方法、および典型的なメタデータの集合を定義した規格として、Clinical Data Interchange Standards Consortium (CDISC) による標準規格がある[1]。個々のデータに紐づく単位や桁数、値の辞書などのメタデータについて、その記述方法を定めた規格(Define-XML[2])、データ自体の XML 表現規格(Dataset-XML)が定義されているが、加えて典型的な臨床試験における具体的な項目セットとメタデータ(Study Data Tabulation Model Implementation Guide: SDTMIG[4])、電子症例報告書における具体的な質問文、解析途中で作成される表の記述、これらで使用される標準語彙セットなど、臨床試験の企画段階からデータ収集、クリーニングと固定、解析、結果報告までの全ての過程で利用される規格群が網羅的に定義されている。

臨床試験では研究計画書の文中に、正しいデータが従うべきルールがメタデータとしてある程度指定されている。規制当局は Define-XML の形で、このメタデータを臨床試験データに付随させることを要求する。しかし、現実には収集されるデータはしばしばこれを逸脱し、精度や単位系が研究計画書による指定と異なった値が入力されるなどするため、データモニタリングが必須である。現状では、モニタリング業務は担当組織ごとに異なる内部手順に基づいて実施される場合が多く、メタデータをあらかじめ電子的に規定することなく試験を開始し、試験の最終段階になって、確定されたデータを見た上で、それに合致するように Define-XML を作成して検査に通るようにしたものを規制当局に提出する、というワークフローが支配的である。

しかし、試験開始時に、研究計画書をもとにメタデータを Define-XML で記述し、これを基準に自動的なモニタリングを実施していくというワークフローが実施できれば、従来は不透明であったデー

タ検査の基準が可視化され、どの段階でどのような基準に基づいて疑義データが発見されたのか、それがどのような経緯で修正されたのか、といったトレーサビリティの向上が期待できる。

現時点で、Define-XML により記述されたルールをベースとしてデータを検査する作業を支援する機能を持つソフトウェアは Pinnacle21 社による製品、XML4Pharma による実装があるが、標準規格に対応するソフトウェアとしては選択肢が少なく、実際に Define-XML をデータ検査の目的で利用している研究現場はほとんどない。そこで、新しいワークフローの導入を推進するために、臨床試験・臨床研究で収集されたデータセットが、Define-XML で定義したメタデータ定義と合致するかどうかを容易に検証することができるソフトウェア「Define2Validate」(以下 D2V と表記)を開発した。また、データセットが SDTMIG で定義される典型的なメタデータセットに適合しているかどうかを検証するツール「ConformanceRules2Validate」(以下 C2V と表記)を試作した。

2. 方法

1) 機能の定義

D2V については、Dataset-XML version 1.0 により記述されたデータ、および Define-XML version 2.0 により記述されたメタデータを入力として、その整合性を確認し、データセットごとに検証結果を集計して出力する機能と定義した。C2V については、Dataset-XML で記述されたデータ、および SDTMIG への準拠検証ルールを記述した文書である SDTMIG v3.2 Conformance Rules version 1.0[3]を入力として、検証結果を集計して出力する機能と定義した。Conformance Rules 文書自体の記述の厳密性に問題があるため、実装が可能な一部のルールに限定して試作した。

2) 実装

R 言語スクリプトとして実装し、R の対話的インターフェースから関数として呼び出す実装とした。

3. 結果

D2V について、Define-XML で定義される、データの値に関する制約は6種類あるが、そのうち4種類を実装した。C2V について、データの検査ルールは SDTM で規定される表の単位であるドメ

インにより異なるが、ドメインごとに実装できたルールの割合の範囲は、最小は TA ドメイン(24/81 ルール, 29.6%)、最大 AE ドメイン(52/106 ルール, 49.1%)であった。

4. 考察

1) 実現できた機能と想定される応用例

D2V は R 言語の関数として実装されているので、データマネジメント業務に本プログラムをそのまま組み込むことで、Define-XML を基準としたデータクリーニングと異常検知をワークフローに組み込むことができる。D2V で可能なデータ検査は Define-XML で可能な範囲に限定されるが、SDTM 準拠のデータセットの場合にはより詳細な検査が可能であり、C2V はそれを可能とする。

現状ではほとんどの臨床試験の現場において、データ検査ルールの標準化はなされていないが、D2V, C2V の導入によりデータの入力と同時に標準化された検査を通すことが可能となるので、検査プログラムの再利用性が向上するとともに、モニタリング過程の悉皆的な記録が可能となることで、データマネジメントの過程の効率化と可視化が進むことが期待できる。

5. 結語

本研究で開発したツールは臨床研究データセンターに新しいワークフローを導入することを可能にするものであり、日本の臨床研究の現場に広く標準規格の導入によるメリットを波及させるきっかけとなると思われる。

本研究は JSPS 科研費 JP15K15218 の助成を受けたものである。

参考文献

[1] CDISC. (2018). CDISC. Retrieved 7 February 2018, from <https://www.cdisc.org/>

[2] Define-XML. (2018). CDISC. Retrieved 7 February 2018, from <https://www.cdisc.org/standards/transport/define-xml>

[3] Study Data Tabulation Model Implementation Guide (SDTMIG). (2018). CDISC. Retrieved 7 February 2018, from <https://www.cdisc.org/standards/foundational/sdtmig>