

CDISC SDTM を活用した REDCap2SDTM による 複数疾患レジストリデータ統合の検証

北山 恵^{*1}, 山本 景一^{*1}, 松崎 慶一^{*2}

^{*1} 和歌山県立医科大学, ^{*2} 京都大学

Verification of disease registry data integration by REDCap2SDTM using CDISC SDTM

Megumi Kitayama^{*1}, Keiichi Yamamoto^{*1}, Keiichi Matsuzaki^{*2}

^{*1} Wakayama Medical University

^{*2} Kyoto University

抄録:

同一疾患領域で複数のデータベース(DB)を統合することにより, より多くの症例数を対象としたデータ解析が可能となる. しかし DB ごとに評価項目やコードが異なるため統合作業は容易ではない. 本研究の目的は実際の臨床現場で行われたコホート研究のデータを CDISC SDTM を用いて統合を試み, 課題を検証することである. 独立した3つのDBをREDCap上に再構築し, CDISC SDTM に基づくメタ情報を入力した上で, SDTM データの生成を行なった. 各DBからCDISC SDTM のドメインに基づくSDTMデータが生成され, 一つのDBとして統合された. 実際の臨床データにおいても適切にメタ情報を定義することでDB統合が可能であり, 一つのデータセットとして解析可能であることが示唆された. 本研究が我が国の医療ビッグデータ活用推進の一助となることを期待している.

キーワード 臨床試験, 疾患レジストリ, データシェアリング, データベース, 医療ビッグデータ

1. はじめに

多くの医学系学会や団体では複数の疾患レジストリを運用しており, 同一の疾患領域で運用しているこれらのデータベース(DB)を統合することで, より多くの項目や症例数を対象としたデータ解析が可能となる.

近年では医学雑誌編集者国際委員会の声明により臨床試験に関する論文投稿時にデータ共有の方法を明記することが義務付けられていることや¹⁾, 乳がんを対象とした15のコホート研究における約90万人のデータをプール解析の手法を用いて統合した結果²⁾が発表されるなど, データ共有とその活用方法に注目が高まっている. しかし, 同一の疾患を対象としていても, 異なる目的で作成されたDBはデータ形式, 変数名, 変数コードなどが統一されていないことが多く, 観察項目や収集時期が標準化されていないなど, 複数DBの統合作業には多くの課題がある. 特に統計解析に用いるためには変数名や変数コードを統一する必要があるが, これらを手作業で実施することは工程数も多く, ヒューマンエラーを起こしやすいなど統合作業は容易ではない.

Clinical Data Interchanges Standards Consortium (CDISC)はプラットフォームに依存しないグローバルなデータ標準を開発・支援している. CDISC 標準の一つである Study Data Tabulation Model (SDTM)は臨床試験データ表形式モデルであり, 各DBの変数をCDISC SDTMドメインモデルを基にマッピングすることにより, 異なるDBの統合が可能となる.

REDCap は米国ヴァンダービルト大学が開発した Electronic Data Capture(EDC)システムであり, REDCap2SDTM^{3,4)}はCDISC SDTM のメタ情報を元

に臨床研究データをSDTM形式に変換するツールである. 山本らが仮想データに基づき検証を行っているが, 実際の臨床データを用いた検証はいまだ行われていない.

我々は, 実臨床で行われた複数のコホート研究のDBを, REDCap2SDTM を用いて統合することを試みた. 本研究の目的は, REDCap2SDTM で実際のコホート研究データを用いたDB統合を行い, CDISC SDTM 生成時の課題について検証することである. CDISC SDTM による既存の複数データベース統合時の課題について報告を行う.

2. 方法

1) REDCap への既存 DB の再構築

実臨床で実施された3つのコホート研究(DB:A(702症例), DB:B(635症例), DB:C(1,174症例))を対象とし, それらのDBをREDCap上に再構築した.

2) CDISC SDTM に基づくメタ情報の入力

DB 統合時に利用する評価項目について当該疾患領域の専門医の助言を元に精査し, CDISC SDTM のドメイン・変数・テストコードにマッピングしたメタ情報をREDCapに入力した.

3) REDCap2SDTM による SDTM データ生成

先行研究^{3,4)}ではREDCapからのCSV出力に基づきSDTMデータを生成していた. しかしこの方法はREDCapのCSV出力仕様に依存するため汎用性が低い.

REDCap2SDTM を, CDISC Operational Data Model (ODM)出力に基づきSDTMに変換するように改良した. 再構築した各DBのデータをODMで出力し,

REDCap2SDTM で SDTM データの生成を行なった。

REDCap2SDTM を用いた DB 統合のフローチャートを Figure1 に示す。

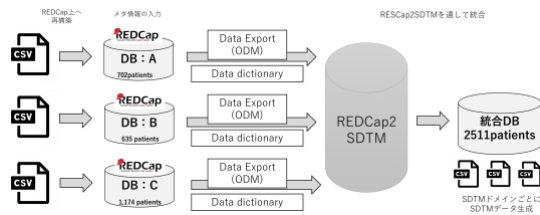


Figure1:DB 統合フローチャート

3. 結果

独立して構築された3つのDBをCDISC SDTM形式で1つのDBとして統合した。SDTMデータはメタ情報で定義したドメインごとにCSVファイルとして生成された。

マッピングされた評価項目のうち各DBで共通する評価項目は、DB:A及びDB:Bでは58項目中22項目、DB:Cでは102項目中22項目であった。各DBに紐づく研究実施計画により全体の評価項目数が異なるが、本疾患領域において収集する患者背景や臨床検査値などの基本情報は全DBにおける共通の収集項目としてマッピングに成功した。

本研究で生成されたドメインは“DM”（DEMOGRAPHICS, 被験者背景）, “CM”（CONCOMITANT MEDICATIONS, 併用薬）, “LB”（LABORATORY TEST RESULTS, 臨床検査結果）, “VS”（VITAL SIGNS, バイタルサイン）, “SUPPCM”, “SUPPDM”, “SUPPMH”の7つであった。SUPP-はドメイン用の補助修飾子であり、各ドメインで格納できない疾患特異的な評価項目等が格納される。

DBごとに個別の変数名で定義されていた評価項目は、CDISC SDTMによるメタ情報を元に一つの評価項目に集約された。性別の統合結果をTable. 1に示す。

Table.1 性別の統合結果

	Median(25%q-75%q) or counts(%)			
	A (N=702)	B (N=635)	C (N=1174)	integrate (N=2511)
Male	295 (42.0%)	300 (47.2%)	580 (49%)	1175 (47%)
Female	407 (58.0%)	335 (52.8%)	594 (51%)	1336 (53%)
Missing	0 (0%)	0 (0%)	0 (0%)	0 (0%)

4. 考察

各DBで変数名が異なる場合でも、適切にCDISC SDTMのメタ情報を定義することでデータの統合が可能となり、複数のDBを一つのデータセットとして解析が可能であることが示唆された。これまで仮想データにおける検証実績はあったが、実際の臨床データに

おいても本ツールを用いることでSDTMデータを生成することが可能であった。またDBにより定義されたコードはSDTM Terminologyあるいはコードリストに基づき自動的に再コーディングを行うことで、工数削減が可能となった。

適切なメタ情報を定義するためには、各DBにおいて収集した評価項目の意義を適切に把握することが必要である。そのためには当該疾患領域の専門医との協働により、各評価項目の意味を確認し、マッピングを行うことが重要である。同様に、疾患領域に特異的な用語など、SDTM Terminologyにおいて定義されていない用語については、近隣領域のTherapeutic Area Standards (TAS)を参考に、専門医と十分に議論を行いながらコードリストを作成する必要がある。

また、同一の評価項目が各DBにおいて縦断的に収集されている場合、メタ情報は各DBで同一でなければならない。適切にマッピングを行うために、DB間の整合性を取る必要がある。本研究ではDBのメタ情報の入力を手作業で行ったが、メタ情報を自動的に入力できる支援ツールを開発することにより、より効率的にDBの統合ができると考えられる。

5. 結語

CDISC SDTMは変数名やコードが異なる複数のDBの統合に有益であることが示唆された。本研究が我が国の医療ビッグデータ活用推進の一助となることを期待している。

参考文献

- [1] Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials: A requirement of the International Committee of Medical Journal Editors. PLoS Med. 14 (6):e1002315 2017.
- [2] Hazel B. Nichols, Minouk J. Schoemaker, Jianwen Cai, et al. Breast Cancer Risk After Recent Childhood : A Pooled Analysis of 15 Prospective Studies. Annals of Internal Medicine 170(1), 22-31, 2019
- [3] Keiichi Yamamoto, Keiko Ota, Ippei Akiya et al, A pragmatic method for transforming clinical research data from the research electronic data capture ‘ ‘ REDCap ’ ’ to Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM): Development and evaluation of REDCap2SDTM. Journal of Biomedical Informatics 70(3), 65-76, 2017.
- [4] 山本景一. REDCapとREDCap2SDTM: 多施設臨床研究データ収集システムと研究データ共有のための変換ツール. BRAIN and NERVE 69(7) 848-855 2017