

公募シンポジウム | 第40回医療情報学連合大会（第21回日本医療情報学会学術大会） | 公募シンポジウム

公募シンポジウム5

リアルワールドエビデンスを創出するための臨床中核指定病院ネットワークの取り組み

2020年11月20日(金) 09:00 ~ 11:00 F会場 (イベントホール・特設会場1)

[3-F-1-06] 臨床研究用データリソースの品質管理

*関 倫久¹、永島 里美¹、大江 和彦¹（1. 東京大学医学部附属病院企画情報運営部）

*Tomohisa Seki¹, Satomi Nagashima¹, Kazuhiko Ohe¹（1. 東京大学医学部附属病院企画情報運営部）

キーワード：Electronic Medical Record, Datawarehouse, Big Data, Database, Data Quality Assessment

電子カルテのデータベースにもとづく臨床研究を行う際には、リアルタイムで稼働している電子カルテデータベースを対象とするのではなく、二次利用目的に再構成されたデータリソース（DWH等）を構築し、それを対象にデータ抽出及び解析を行う。その理由は、リアルタイム業務処理にレスポンス上の影響を与えたくないこと、また、研究目的にデータ抽出するためにはテーブル構成を再編成し、個々のデータ形式を必要に応じて変換したりコードを標準化したりする必要があるからである。このプロセスでは単にデータ形式やコードの変換だけでなく、研究に使用するデータ項目のもつ意味を理解した上で、必要な情報を意味的に正しく構成できるように、元の複数データ項目からデータを取り出し変換や結合をして新しいデータを作成する必要もある。

このようなデータベースの再構築は、電子カルテベンダーやDWHベンダーにより病院ごとに開発されたソフトウェアにより処理されることが多いが、この過程は非常に複雑であり、結果として構築された研究用のデータベースが、元のデータのもつ値や意味的な情報を正しく保持しているかどうかを検証する必要がある。しかし、その検証手法は確立していないだけでなく、検証結果としてのデータベースの品質をどのような指標で評価すべきかについても指針が存在していない。

SWG4ではこの観点から、以下の4つのテーマについて検討を実施している。

- 1) 汎用的リソースとしてのSS-MIX2標準化ストレージのバリデーション
- 2) 2つのデータリソースからのFHIR準拠オブジェクト同士の意味的な検証
- 3) 臨床研究に使えるデータベースの品質指標の検討
- 4) 電子カルテ関連ベンダーとの二次利用データ品質管理手順の作成

本SWGは本年7月に設置され、一部のテーマを引き継いで活動を始めたばかりである。発表では上記の4つについて進捗状況を報告したい。

臨床研究用データリソースの品質管理

関倫久^{*1}、永島里美^{*1}、大江和彦^{*1, 2}

*1 東京大学医学部附属病院 企画情報運営部、*2 東京大学大学院医学系研究科 医療情報学分野

Quality control of data resources for clinical research

Tomohisa, Seki^{*1}, Satomi, Nagashima^{*1}, Ohe, Kazuhiko^{*1, 2}

*1 Department of Healthcare Information Management, The University of Tokyo Hospital

*2 Department of Biomedical Informatics, The Graduate School of Medicine, The University of Tokyo

Using electronic medical records for clinical studies requires data resources reconstructed for secondary use. Data for clinical studies should not be directly extracted from active clinical-use databases to avoid the risk of affecting system response. Additionally, appropriate data processing such as reorganization of the table structure, transformation of data formats, standardization of codes, and recreation of data are required to obtain necessary information based on the purpose of the research. Because of the complexity of this process, the accuracy of values and the preservability of semantic information in reconstructed databases need to be examined before use in clinical studies. However, a verification method for reconstructed databases has not yet been established. Moreover, there has been no indicator of the quality of the reconstructed databases. To address this issue, the following four themes have been set as the subjects of sub-working group 4: (1) validation of SS-MIX2 standardized storage as a general-purpose resource, (2) semantic verification between Fast Healthcare Interoperability Resources-compliant objects from two data sources, (3) examination of quality indicators for databases used in clinical research, and (4) development of a quality control procedure for secondary use of electronic medical records with related vendors. We report on the progress achieved in addressing the above four themes since July 2020.

Keywords: Electronic Medical Record, Datawarehouse, Big Data, Database, Data Quality Assessment

1. 諸論

臨床研究中核病院では、日本における医療リアルワールドデータの利活用促進を目指し、2018 年度より国立研究開発法人日本医療研究開発機構事業「医療技術実用化総合促進事業」の中で、リアルワールドエビデンスを創出するためのデータ基盤整備およびそれを下支えるための人材育成に取り組んでいる。本事業は、「クリニカルクエスチョンをはじめとする広範な課題にリアルワールドエビデンスとして回答できる持続可能な臨床研究基盤の構築を通じ最適な医療の実現に貢献する」をビジョンに掲げ、本事業の中のサブワーキンググループ4 (SWG4) では、臨床研究用データリソースの品質管理について検討を進めている。

電子カルテのデータベースに基づく臨床研究を行う際には、リアルタイムで稼働している電子カルテデータベースを対象とするのではなく、二次利用目的に再構成されたデータリソース(DWH 等)を構築し、それを対象にデータ抽出及び解析を行う。その理由は、リアルタイム業務処理にレスポンス上の影響を与えたくないこと、また、研究目的にデータ抽出するためにはテーブル構成を再編成し、個々のデータ形式を必要に応じて変換したりコードを標準化したりする必要があるからである。このプロセスでは単にデータ形式やコードの変換だけでなく、研究に使用するデータ項目のもつ意味を理解した上で、必要な情報を意味的に正しく構成できるように、元の複数データ項目からデータを取り出し変換や結合をして新しいデータを作成する必要がある。¹⁾

このようなデータベースの再構築は、電子カルテベンダーや DWH ベンダーにより病院ごとに開発されたソフトウェアにより処理されることが多いが、この過程は非常に複雑であり、結果として構築された研究用のデータベースが、元のデータのもつ値や意味的な情報を正しく保持しているかどうかを検証する必要がある。²⁾ しかし、その検証手法は確立していないだけでなく、検証結果としてのデータベースの品質をどのような指標で評価すべきかについても明確な指針が存在しない

現状があり、持続可能な臨床研究基盤の構築のためにはこれら検証方法や指針を整備する必要があると考えられる。

2. 目的

このような背景を踏まえ本 SWG4 では、二次利用データベース構築に際して、構築したデータベースが保持する値の正確性、意味的な情報の保存性を検証し、データベースの品質を評価する手法を確立することを目的とする。

3. 概要と評価

本 SWG4 では、二次利用データソースの品質管理に関し、以下の4つのテーマについて検討を行う。

- 1) 汎用的リソースとしての SS-MIX2 標準化ストレージのバリデーション
- 2) 2つのデータリソースからの FHIR 準拠オブジェクト同士の意味的検証システムの開発
- 3) 臨床研究に使えるデータベースの品質指標の検討
- 4) 電子カルテ関連ベンダーとの二次利用データ品質管理手順の作成

ここでは上記 4 テーマに沿ってその概要と進捗状況について報告をする。

3.1 汎用的リソースとしての SS-MIX2 標準化ストレージのバリデーション

SWG4 の前身である本事業の SWG1 のプロジェクト1では、医療情報の交換・共有による医療の質の向上を目的とした「厚生労働省電子的診療情報交換推進事業」において規格が定められ、現在我が国で広く用いられている SS-MIX2 標準化ストレージを基軸とした臨床研究基盤の構築に向けて、データバリデーションチェックリストの作成、バリデーションの範囲や項目の策定が進められてきた。SS-MIX2 標準化ストレージを対象に、施設共通のバリデーション手順書に従って、バリデーションを行うことにより、汎用的なデータリソースとして確立することが期待できると考えられ、SWG4 においても引き

続き臨床研究中核病院内で共通の様式を元に品質管理作業を継続する方針である。

3.2 2つのデータリソースからの FHIR 準拠オブジェクト同士の意味的検証システムの開発

SS-MIX2 標準化ストレージや研究二次利用データベースなど、電子カルテデータベースから由来して生成された二次的データソースの相互間データ一貫性検証作業には客観性と透明性の確保、および作業の効率化が求められる。³⁾ そこで異なる2つのデータ・リソースからオブジェクトを取得した上で相互に自動的に比較し、修正が必要となる不整合を自動検出し、その原因を類推できるレポートを出力することで、人手によるバリデーション業務を支援もしくは将来的には置換するシステムの開発が求められる。

SWG4 ではこのような異なる2つのデータ・リソースの整合性の検証を自動化することを目的に、「意味的ペリフィケーションシステム」の設計、開発及びプロトタイプの実装を開始した。本システムでは、検証対象としたい対象データ項目の意味論的なデータオブジェクト表現を基本とし、そのオブジェクト表現と RDB へのマッピング定義、およびオブジェクト表現と SS-MIX2 形式へのマッピング定義を行うことにより、意味論的なデータオブジェクト表現を仲介して両者の直接比較を行う。現段階では SS-MIX2 ストレージ及び、電子カルテのデータベースを対象としてその差異を出力するシステムの開発を行っている。本システムの構成を図1に示す。

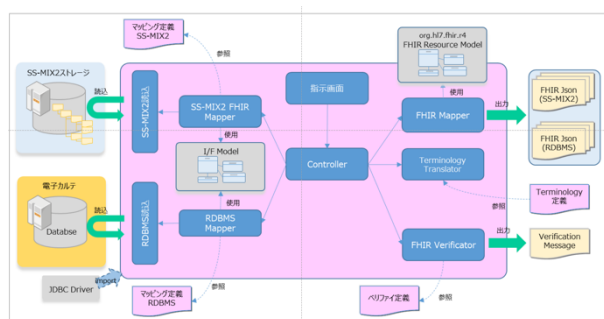


図1 意味的ペリフィケーションシステム

SS-MIX2 ストレージ及び、電子カルテデータベースから HL7 FHIR Resource ベースのオブジェクトへコンバートを行い、同リソースのデータを比較/検証し、その差異を出力する。

本システム開発は、最終的にデータリソースに対応したマッピング定義を作成することにより、電子カルテ、ベンダーを問わず動作するものとする 것을 目指す。今後は、臨床研究中核病院内の複数施設にて、実際の電子カルテデータでの評価を行う予定であり、評価にもとづいて改修すべき点を追加改修、開発する。

本システムの開発・導入が可能となれば、バリデーション業務の人的負担、時間的負担が軽減し、複数医療機関のデータの統合への域値が下がることが期待される。その結果、より多くの施設がバリデーションを行うことが可能となり、将来的なデータ統合事業への参画期間が増え、我が国における医療データの統合が進むことが期待される。

3.3 臨床研究に使えるデータベースの品質指標の検討

臨床研究に利用可能なデータベースの品質の検証手法及び評価指標が確立していない現状を踏まえ、臨床研究用

データリソースの品質管理について海外の事例を含めて文献の調査を行った。欧米ではデータベース研究が 1970 年代後半から実施され、同時にリアルワールドデータの「データの質」についてさまざまな評価指標が提唱されてきた。^{4) 5)} 近年、リアルワールドデータを医薬品開発や医薬品の安全性評価等に活用するための議論がなされ、FDA、PMDA から医療情報のデータベースを用いた研究についてのガイダンスも出されている。^{6) ~ 8)}

しかしながら、臨床研究用データリソースの品質管理について、どのような手法でどのような検証結果が出れば、そのデータリソースは臨床研究に必要なデータ品質を確実に確保できているといえるかについて、現時点で明確な指針は存在しない。リアルワールドデータは、日常診療から日々得られるデータであり、臨床試験のように試験計画に基づき収集されたデータでは無い。そのため、リアルワールドデータを利用する研究者、分析者は、そのデータの特性やデータの質を評価したうえで利用することが求められる。これまでの調査、検討を経て、SWG4 では今後さらに臨床研究に使えるデータベースの品質指標の検討を進める方針である。

3.4 電子カルテ関連ベンダーとの二次利用データ品質管理手順の作成

二次利用データリソースは、病院ごとに開発されたソフトウェアにより処理されることが多く、同じベンダーであっても、その構造や、検証方法は必ずしも病院ごとに同一では無く、また共通化・共有化されていないケースが多々存在するのが現状である。このような現状を踏まえ、SWG4 では、電子カルテ関連ベンダーと協議を行うことにより、二次利用データリソースの利用目的を理解したうえで、ベンダー側の品質管理手順についても検討を進める方針である。

4. 考察

本事業では、リアルワールドエビデンスを創出するための高品質のデータ基盤の整備を目標としており、データの品質管理及びそのための人材育成が当初より課題とされていたが、データの質の検証方法、評価指標についてより詳細な議論の余地があると考えられた。SWG4 はこれらの議論を進めるべく本年7月に設置され、一部のテーマを引き継いで活動を始めたばかりである。そのため、各テーマを設定し検討を開始した段階にあり、今後設定したそれぞれのテーマについてさらなる検証を進めていく方針である。また、各臨床研究中核病院の医療情報管理部門だけでなく、臨床研究の支援部門の連携も必要と考えられる。今後、各施設の担当者と共に議論を深め、取り組んでいくことが重要と考えられる。

5. 結論

SWG4 では、臨床研究に利用可能なデータベースの品質の検証手法及び評価指標を確立すべく、4つのテーマを設定し開発・検証を行っている。今後さらにこれらのテーマを軸にSWG4を進めていく方針である。

6. 謝辞

SWG4 は本年7月に設置され、それまでの SWG1-PJ1 の活動を引き継いだ部分があり、本抄録の3.1では、SWG1-PJ1をリードされた名古屋大学病院の山下暁士氏、九州大学病院の山下貴範氏を始めとする関係メンバーの検討内容が含まれている。また、3.2のテーマは、京都大学病院、大阪大学病院、岡山大学病院と共同で実施する計画となっている。ここに関係各位に感謝申し上げる。

参考文献

- 1) 武田理宏, 真鍋史朗, 松村泰志. 電子カルテデータ二次利用の現状と課題. 生体医工学 55(4):151-158,2017
- 2) Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013;20(1):144-151.
- 3) 高橋史峰. 電子診療情報データベース(MID-NET)プロジェクトの現状と課題. RSMP vol.5 no.3, 235-243, Sep 2015
- 4) Institute for Clinical Evaluation Sciences. Quality Assessment of Administrative Data (QuAAD): An Opportunity for Enhancing Ontario's Health Data, 2007.
- 5) NIH Collaboratory, Assessing Data Quality for Healthcare Systems Data Used in Clinical Research, 2014
[https://dcricollab.dcri.duke.edu/sites/NIHCR/KR/Assessing-data-quality_V1%200.pdf (cited 2020-Aug-20)]
- 6) U.S.Food and Drug Administration. Guidance: Use of Electronic Health Records in Clinical Investigations, 2018
[<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-electronic-health-record-data-clinical-investigations-guidance-industry> (cited 2020-Aug-20)]
- 7) U.S.Food and Drug Administration. Framework For FDAs Real World Evidence Program, 2018.
[<https://www.med.upenn.edu/cstr/assets/user-content/framework-FDAs-real-world-evidence-program.pdf> (cited 2020-Aug-20)]
- 8) 医薬品医療機器総合機構. 医療情報のデータベース等を用いた医薬品の安全性評価における薬剤疫学研究の実施に関するガイドライン, 2014
[<https://www.pmda.go.jp/files/000147250.pdf> (cited 2020-Aug-20)]