

臨中ネットにおける SS-MIX2 データのバリデーションの有用性に関する検討

中村 直毅^{*1}、伊藤 詩歩^{*1}、宋 チュウ^{*1}、諸橋 朱美^{*2}、山下 暁士^{*2}、今泉 貴広^{*2}、大田 英揮^{*1}

^{*1} 東北大学病院メディカル IT センター、^{*2} 名古屋大学付属病院

An Examination of the Usefulness of SS-MIX2 Data Validation in the Clinical Research Core Hospital Clinical Information Network

Naoki Nakamura ^{*1}, Shiho Ito ^{*1}, Song Chong ^{*1}, Akemi Morohashi ^{*2}, Akira Yamashita ^{*2},

Takahiro Imaizumi ^{*2}, Hideki Ota ^{*1}

^{*1} Medical Information Technology Center, Tohoku University Hospital,

^{*2} Nagoya University Hospital

In order to utilize data stored in electronic medical records as real-world data, it is very important to work on maintaining and managing data quality. In the Rinchu Net project, SS-MIX2 data quality control work is being conducted using a Validation Tool developed by Nagoya University Hospital. The Validation Tool compares the data on SS-MIX2 with the data in the electronic medical record, and checks for data matches and inconsistencies. As a result of the validation, if there is a need for improvement, it is expected that the quality of SS-MIX2 data will be improved by repeatedly resolving defects. On the other hand, there is a difference in judgment as to whether it is necessary to perform validation for the entire period. In this paper, a system for validation for the entire period is established and implemented on a trial basis.

Keywords: SS-MIX2, Real World Data, Data Validation

1. はじめに

電子カルテに蓄積されているデータをリアルワールドデータとして活用するためには、データの品質を維持管理する取り組みは非常に重要である。臨中ネット事業では、名古屋大学病院が開発しているバリデーションツールを活用して、SS-MIX2 データの品質管理作業を実施している¹⁾。バリデーションツールは、SS-MIX2 上のデータと電子カルテ上のデータを比較し、データの一致・不一致を確認することができる。確認の結果、改善の必要な不具合である場合には、不具合を解消する取り組みを繰り返し行うことで、SS-MIX2 のデータ品質の向上が期待される。これまで、バリデーションツールの制約や運用の負荷を勘案して、バリデーションを行うデータの期間を約1か月に限定して、バリデーション作業を実施してきた。大量のデータがある中で、一か月間のデータではなく、全期間のデータを網羅的にバリデーションする必要性があるかについては、判断の分かれる所である。

2. 目的

臨中ネットで利用しているバリデーションツールを利用するためには、トランザクションストレージが必要である。電子カルテから出力されるトランザクションストレージは、全患者の診療情報が時系列順で出力されている。全期間を対象に特定の患者のデータのバリデーションを実施しようとする場合、膨大な量のトランザクションストレージのデータを取り込む必要があるため、バリデーションするのは困難な状況である。

そこで、SS-MIX2 ストレージから、特定の患者に限定して、全期間を対象とするトランザクションストレージを生成し、これを用いてバリデーションを実施するための仕組みの確立を目的とする。

3. 方法

3.1 バリデーションツール

SS-MIX2 ストレージのデータ品質の維持管理を円滑に行えるようにするため、名古屋大学病院が中心となって、SS-MIX2 と電子カルテ上のデータを確認・検証することができる



図1:バリデーションツールの画面

バリデーションツールを開発している。

バリデーションツールでは、最初に SS-MIX2 のデータを CSV 化して取り込むとともに、電子カルテのデータも、電子カルテのデータベースから ODBC 接続でデータを抽出し、そのデータを取り込む。取り込み後、データの分析工程を経て、双方のデータ件数、データの一致率、不一致となっている箇所の確認や分析をすることが可能となっている(図1)。

運用の手軽さを重視し、Microsoft 社製 Access 上にバリデーションツールを実装している。ただし、Access の機能上の制約によって、各々の Access のファイルは、2GB 以下になるように調整しながら運用している。データ容量が大きくなるテーブルは、本体のファイルとは分離して、異なるファイルに分

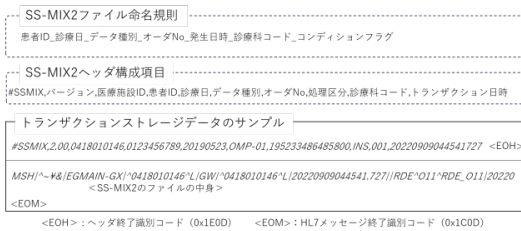


図 2: SS-MIX2 ファイルおよびトランザクションストレージフォーマット

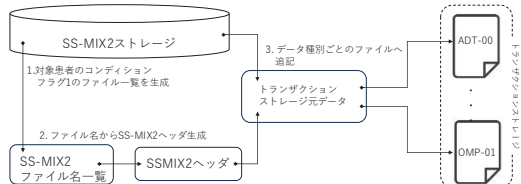


図3: トランザクションストレージの生成の概要

割し、各々のファイルが 2GB 以内に収まるようにしている。項目数の多い検体検査結果などのデータ種別では、単一のテーブルで 2GB 以上になる場合がある。この場合には、バリデーションする対象期間を短く設定し、複数回に分けてバリデーションツールを実行することで、Access のファイルが 2GB 以内に収まるように調整している。

3.2 トランザクションストレージ生成

SS-MIX2 ストレージからトランザクションストレージを生成するにあたって、対象患者は、臨中ネットにおいて行われている臨床研究²⁾の患者のうち、東北大学病院を受診している患者とした。トランザクションストレージは、図 2 に示すように、ヘッダ部およびデータ部から構成される。ヘッダ部は、SS-MIX2 のファイル名に含まれる情報を用いて生成する。その際、ファイル名に含まれない医療施設 ID および処理区分は、それぞれ、東北大学病院の医療機関コードと「INS」を記載する。データ部は、SS-MIX2 のファイルの内容をそのまま用いる。トランザクションストレージを生成するための概要を図 3 に示す。最初に、Linux 標準の find コマンドを用いて SS-MIX2 ストレージ中から、バリデーションの対象患者で、コンディションフラグが 1 (ファイル名の末尾が「1」) の SS-MIX2 のファイル一覧を取得する。その後、取得したファイル名を用いて、トランザクションストレージに書き込むための元データを生成する。その後、生成したデータを、データ種別ごとに異なるトランザクションストレージ用のファイルに追加書き込みすることで、トランザクションストレージが生成される。生成されたトランザクションストレージのファイルを用いて、実際にバリデーションツールを動かしてバリデーションを実施する。

4. 結果・考察

上述のように東北大学病院の SS-MIX2 ストレージを用いて特定の患者の全期間のデータを含むトランザクションストレージを生成する仕組みを構成した。対象患者は 7,300 人、実際の期間は 2019 年 1 月 1 日から 2023 年 9 月 11 日とし、コンディションフラグ「1」の SS-MIX2 のファイルは、1,814,170 個あり、このデータからトランザクションストレージを生成した。データの生成には、約 9 時間の時間を要した。

生成したトランザクションストレージを用いて、バリデーションツールを実行した。実行の概要を表 1 に示す。生成したトランザクションストレージを用いてバリデーションツールを実行した所、食事オーダ、検体検査結果、処方オーダ、注射実施、

表1: バリデーションの実行の概要

データ種別	患者基本		外来	入院実施	転科・転棟実施	退院実施	アレルギー情報	
	ADT-00	ADT-12	ADT-22	ADT-42	ADT-52	ADT-61		
対象患者	7300名	7300名	7300名	7300名	7300名	7300名		
対象件数	7300件	305722件	8127件	7553件	7426件	2479件		
※対象件数 (対象期間1か月)	32610件	56505件	1800件	1743件	1758件	497件		
実行時間	0時間14分3秒	0時間40分28秒	0時間29分33秒	0時間25分50秒	0時間31分2秒	0時間7分24秒		
Accessファイル数	2	1	1	1	1	1		
患者の分割	0	0	0	0	0	0		

データ種別	食事オーダ		検体検査結果	処方オーダ	注射実施	病名情報
	OMD	OML-11	OMP-01	OMP-12	PPR-01	
対象患者	7300名	7300名	7300名	7300名	7300名	
対象件数	56378件	182321件	285741件	332686件	6119件	
※対象件数 (対象期間1か月)	12317件	29942件	40606件	62835件	10196件	
実行時間	2時間37分24秒	9時間28分43秒	16時間15分01秒	13時間45分15秒	1時間56分23秒	
Accessファイル数	2	7	5	3	4	
患者の分割	4	8	4	4	4	

表 2: 注射実施 (OMP-12) におけるデータ一致率

	診療科	PDB-3	PDB-7	PDB-8	PDB-12	ORC-1	ORC-4	ORC-170	ORC-170	ORC-170	ORC-260	ORC-260	RKA-3	RKA-4	RKA-50	RKA-50
今回	100.0	100.0	100.0	100.0	100.0	100.0	97.8	96.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
従来	100.0	100.0	100.0	100.0	100.0	99.8	85.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

	RKA-50	RKA-50	RKA-50	RKA-50	RKA-50	RKA-50	RKA-50	RKA-50	RKA-50	RKA-50	RKA-50	RKA-50	RKA-50	RKA-50	RKA-50	RKA-50
今回	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.8	96.4	99.9	99.9	100.0	100.0	100.0	100.0
従来	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.4	96.4	100.0	100.0	100.0	100.0	100.0	100.0

病名情報において、Access ファイルのサイズが 2GB を超えて、プログラムが停止する事象が発生した。そのため、対象患者を 4 分割もしくは 8 分割にして、複数回に分けてバリデーションツールを実行するようにした。検体検査結果(OML-11)、処方オーダ(OMP-01)、注射実施(OMP-12)では、データ件数やメッセージ内の項目数が多く、実行完了するまでに長時間を要し、検体検査結果においては、一回あたりの実行時間が10時間以上に及んだ。

注射実施 (OMP-12) におけるデータ一致率を表 2 に示す。「今回」の列は、今回の結果を示しており、「従来」の列は、昨年度のバリデーション作業で、1か月の期間に絞って実施した結果を示している。両結果ともデータ一致率は高く、これまでの期間を絞って行ってきたバリデーションの結果とおおむね似通った結果となっており、これまでのバリデーション作業を通して、データの品質管理がある程度機能していたのではないと思われる。

今後、臨中ネット内でリアルワールドデータを用いた臨床研究を行うための基盤整備が進む中で、データの品質を網羅的に確認できる仕組みを備えていることも重要である。そのため、バリデーションを網羅的に実施できるような仕組みについても並行して検討を進めたい。

5. まとめ

本稿では、SS-MIX2 のストレージからトランザクションストレージを生成し、生成したデータを用いてバリデーションツールを実行した。従来のバリデーション結果とおおむね似通った結果を示しており、全期間に渡ってデータの品質がある程度維持されているのではないかと考えられた。さらなるデータの品質の向上を目指して、今回のバリデーションで不一致となった項目に対して、SS-MIX2 のデータに依るものなのか、バリデーションツールに依るものなのか、両データを付き合わせながら確認作業を進める予定である。

参考文献

- 諸橋朱美, 山下暁士, 今泉貴広, 宮城英毅, 白鳥義宗, SS-MIX2 標準化ストレージ内に蓄積された RWD の品質管理方法の検討と信頼性検証, 第 41 回医療情報学連合大会 pp.978-982, 2021
- 今泉貴広, 他 慢性腎臓病患者における降圧薬と電解質異常: リアルワールドエビデンスを創出する臨床研究中核病院ネットワークの取り組み, 日本医療情報学会春季学術大会プログラム・抄録集, p.56-57, 2022