

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-1-02] データ基盤整備のための SWG1 における活動について

*山下 暁士¹、船越 公太²、山下 貴範³、竹村 亮⁴、永島 里美⁵、諸橋 朱美⁶、中村 泰三²、宮原 冬佳³、高田 敦史³、戸高 浩司²、馬場 英司²、白鳥 義宗¹、中島 直樹³、中山 雅晴⁷（1. 名古屋大学医学部附属病院 メディカルITセンター, 2. 九州大学病院 ARO次世代医療センター, 3. 九州大学病院 メディカルインフォメーションセンター, 4. 慶應義塾大学病院 臨床研究推進センター 生物統計部門, 5. 東京大学医学部附属病院 企画情報運営部, 6. 名古屋大学医学部付属病院 先端医療開発部 データセンター, 7. 東北大学大学院 医学系研究科 医学情報学）

*Satoshi Yamashita¹, Kouta Funakoshi², Takanori Yamashita³, Ryo Takemura⁴, Satomi Nagashima⁵, Akemi Morohashi⁶, Taizo Nakamura², Fuyuka Miyahara³, Atsushi Takada³, Koji Todaka², Eishi Baba², Yoshimune Shiratori¹, Naoki Nakashima³, Masaharu Nakayama⁷（1. 名古屋大学医学部附属病院 メディカルITセンター, 2. 九州大学病院 ARO次世代医療センター, 3. 九州大学病院 メディカルインフォメーションセンター, 4. 慶應義塾大学病院 臨床研究推進センター 生物統計部門, 5. 東京大学医学部附属病院 企画情報運営部, 6. 名古屋大学医学部付属病院 先端医療開発部 データセンター, 7. 東北大学大学院 医学系研究科 医学情報学）

キーワード：Data Quality Management, Real World Data, Data Validation Checklist, SS-MIX2

SWG1は病院情報システムから出力するデータ項目の選定やその品質管理、および臨床研究への応用を中心に検討しており、プロジェクト（PJ）ごとチームを構成して進めている。現在ある5つのPJのうち、本発表ではデータの品質管理を検討するPJ1と糖尿病患者管理をテーマとするPJ2について報告する。

PJ1では、現在多くの事業に用いられている Standardized Structured Medical Information eXchange 2（SS-MIX2）の品質管理方法の標準化について検討した。最初に品質管理を行うべきデータ種別およびデータ項目を選定、次に、どのような品質評価を行うべきかを議論の上整理し、多施設で応用できるチェックリストを作成した。並行して SS-MIX2データの整合性確認を容易にするためのツールを開発している。今後必要な作業は、品質管理手順の確立やデータ項目が拡張された際の対応方法、そして各施設間による差異の解消などを考えており、新たに SWG4 へ移行して活動を継続する予定である。

一方、リアルワールドエビデンス創出のための研究に必要なデータ項目を明確するためには、具体的な研究課題を複数実施し、検証する必要がある。PJ2はその一つとして開始された。2017年に出版された高齢者糖尿病診療ガイドラインは、それ以前と比して、重症低血糖リスクのある薬剤の使用において血糖コントロールの下限を定めた点が新しいが、それを受けて、診療行動パターンが2017年以降どう変化してきたかの検討を行った。九州大学病院で先行して入院歴のある高齢者糖尿病患者の診療情報を網羅的に抽出し解析を行い、2020年8月より順次多施設に拡大したところである。シンポジウムでは、九州大学病院での解析結果、並びに多施設の準備または解析状況について報告する。

Sub Working Group 1-Project 1 におけるデータ品質管理の検討について

山下 暁士^{*1}、山下 貴範^{*2}、竹村 亮^{*3}、永島 里美^{*4}、諸橋 朱美^{*5}、白鳥 義宗^{*1}、中島 直樹^{*2}、中山 雅晴^{*6}

*1 名古屋大学医学部附属病院 メディカル IT センター、*2 九州大学病院 メディカルインフォメーションセンター、

*3 慶應義塾大学病院 臨床研究推進センター 生物統計部門、*4 東京大学医学部附属病院 企画情報運営部、

*5 名古屋大学医学部附属病院 先端医療開発部 データセンター、

*6 東北大学大学院 医学系研究科 医学情報学

Examination of Data Quality Management in Sub Working Group 1 – Project 1 of RinchuNet

Satoshi Yamashita^{*1}, Takanori Yamashita^{*2}, Ryo Takemura^{*3}, Satomi Nagashima^{*4}, Akemi Morohashi^{*5}, Yoshimune Shiratori^{*1}, Naoki Nakashima^{*2}, Masaharu Nakayama^{*6}

*1 Medical IT Center, Nagoya University Hospital, *2 Medical Information Center, Kyushu University Hospital,

*3 Division of Biostatistics, Keio University Hospital Clinical and Translational Research Center,

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

*5 Data Coordinating Center, Department of Advanced Medicine, Nagoya University Hospital

*6 Department of Medical Informatics, Tohoku University School of Medicine

The Sub Working Group1-Project1 (PJ1) was established to argue data quality control for conducting clinical research projects. In the first year, the discussion was focused on SS-MIX2 (Standardized Structured Medical Information eXchange 2) validation, as it was based on the precedent.

First, the validation of SS-MIX2 in this study was defined in four items. Next, the data types to be handled in the initial quality control activities were selected from the SS-MIX2 specification document (Version 1.2f) using the frequency of use in general clinical research as an indicator. Finally, data items were selected based on their importance in the clinical research, and the Checklist Version 1 was developed to identify which evaluation points should be applied to each item. In addition, we are developing a tool to assist in the implementation of validation based on the checklist.

Data quality control has to be continuously carried out in accordance with the established procedures and evidence retaining. In addition, audit of the operation is also necessary. In the future, we plan to steadily resolve the remaining issues, such as the establishment of quality control procedures, determination of quality control policies for the expansion of data items, and validation of data sources at each facility.

Keywords: Data Quality Management, Real World Data, Data Validation Checklist, SS-MIX2

1. 初めに

医療技術実用化総合推進事業「Real World Evidence 創出のための取組み」（以下、臨中ネット）は、臨床研究中核病院に Real World Data（以下、RWD）を用いた観察研究を行うための基盤を構築する目的で平成 31 年度後半より開始された。その VISION は「臨床的課題をばはじめとする広範な課題に、Real World Evidence 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」であり、4つの MISSION（1. 高い品質を確保するデータ管理、2. 将来的な拡大を見据えたシステム設計、3. 基盤を維持する自立的なエコシステムの検討・整備、4. これらを支える人材育成の実施）を掲げている。

VISION、MISSIONを実現するため、個々の課題を解決する目的で臨中ネットの Working Group（以下、WG）の下に Sub Working Group（以下、SWG）を設置することが令和元年度第1回 WGで決定した（図1）。SWG1はMISSIONを実現するための優先課題である「具体的な想定研究課題を設定し、必要なデータ項目の決定」を行うために設置された。SWG1では病院情報システムから出力するデータ項目の選定やその品質管理、および臨床研究への応用を中心に、非常に広範な項目を検討しており、それらを検討しやすくするためにまとめた課題についてはプロジェクト(PJ)ごとチームを構成して進めている。現在ある PJ は大きく分けて、まとま

た課題に対応するためのプロジェクト（PJ1:品質管理、PJ4:新規研究課題の検討、PJ5:データセット）と、課題を洗い出すために個別の臨床研究を実際に行うためのプロジェクト（PJ2:糖尿病、PJ3:SEAMAT）に分かれる。本発表では現在ある5つのPJのうち、データの品質管理を検討するPJ1について報告する。

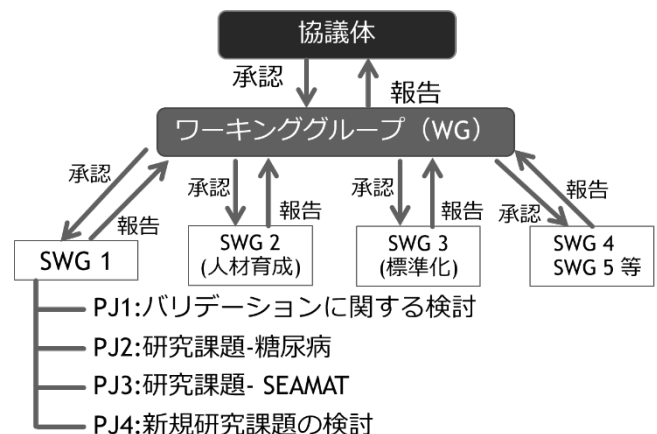


図1. 臨中ネットの組織

2. PJ1

SWG1-PJ1(以下、PJ1)は、RWDを用いた臨床研究課題に対応できるよう臨床研究中核病院内に品質管理体制を構築する取組みの一環として、データ品質管理手法を検討するために発足したプロジェクトである。これは、MISSIONの1つである「品質を確保するデータ管理」に該当し、質の担保されたRWDを用いたReal World Evidence創出には欠かせないものと考えられる。データ品質管理を行う範囲は、電子カルテなどを含む病院情報システム(HIS)とRWDを用いた臨床研究のデータソースとなるSS-MIX2(Standardized Structured Medical Information eXchange:厚生労働省電子的診療情報交換推進事業)標準化ストレージや医療用DWH内のデータである。メンバーは取りまとめを名古屋大学、スーパーバイザーを九州大学が担当し、東北大学、慶応義塾大学、東京大学、京都大学、岡山大学が参加した。

RWDを用いた臨床研究では、品質管理方法や品質に関する指標が明確には示されておらず、先行事例であるMID-NET¹⁾(Medical Information Database NETwork; 独立行政法人医薬品医療機器総合機構(以下、PMDA)が構築した大規模医療情報データベース)のデータ品質管理におけるノウハウを参考にした。そのため、初年度はMID-NETで使用されているSS-MIX2標準化ストレージに関するデータ品質の検証に絞った検討を行った。

3 PJ1で行われた取り組み

PJ1でこれまでにを行った取り組みとしては、1. バリデーションの定義、2. データバリデーションチェックリストの作成、3. バリデーションに関連したツールの開発、4. バリデーションそのものの実施、5. その他の品質管理に関する事項がある。標準コードの検討やマスター管理方法などについても当初は検討項目に含まれていたが、標準化を担当するSWG3が設置されたことに伴い、そちらに移管された。

3.1 バリデーションの定義

SWG1で議論を重ねる中で、「バリデーション」という語の多義性が問題になったため、議論を円滑に進める目的で、SS-MIX2標準化ストレージのバリデーションとは何かについて議論を重ね、最終的に以下の4つを行うことと定義した。

1. 病院情報システム(以下、HIS)データとSS-MIX2データの整合性の確認
2. SS-MIX2データに標準コードが適切に付与されていること
3. 別途定めるイベントと項目名との妥当性の確認
4. 不整合のある、または不正なSS-MIX2データを修正する

ここで言う整合性とはデータ件数が一致していることと各データ項目の内容が一致していることを指すことと定義した。一致が何を指すかについては、方針・手順を検討する際に決定することとした。標準コードについては、現時点ではMID-NETに準拠することとした。イベントとは入退院や外来受診、副作用発生などの出来事が発生したことを指すと定義し、イベントとSS-MIX2標準化ストレージのデータ項目との対応付けが妥当かの議論を行うことが3の具体的な内容であると定義した。そして、1-3の過程で問題が起こった場合には、4の修正を実施し、再度、評価・確認を行うこととした。

この定義はSS-MIX2標準化ストレージに関してのみの定

義であり、その他のデータソースについては今後の検討課題である。

3.2 データバリデーションチェックリストの作成

臨中ネットでは、複数施設のデータを統合して解析することを目指していることから、全施設が一定程度の品質のデータを維持管理することを目的に、バリデーションの対象範囲・項目を記載した施設共通のデータバリデーションチェックリストを作成することとした。まずは、臨中ネット参加施設において多くの施設が二次利用データソースとして活用を検討しているSS-MIX2のデータバリデーション(HISデータとSS-MIX2標準化ストレージデータの一致性の確認)について検討した。

臨中ネットの目的が観察研究に用いるための高品質なRWDを事前に準備しておくことであること、個々の研究課題プロジェクトがSWG1で検討されているのと並行してPJ1の活動を実施していたことから、スモールスタートとしてバリデーションを行うデータ範囲・項目を決定することから開始した。その決定指標は、イベントを評価するのに必要になる可能性があるかどうか、臨床研究に用いる可能性が高いかどうかの二点とし、SS-MIX2標準化ストレージ仕様書(1.2f)のデータ種別、フィールド(以下、データ項目)から抽出した。検討の結果、11データ種別(病名、処方オーダ、注射実施、検査結果、患者基本情報、入院、退院、外来、アレルギー、転棟転科、給食)を選定し、この11データ種別内よりデータ項目を選定、データバリデーション実施項目とし、データバリデーションチェックリストVersion1を作成した。

本チェックリストは、以下の2点を考慮して作成した。

- ① 複数施設での観察研究を想定し、各施設に設置されたSS-MIX2標準化ストレージに格納されているデータの品質管理活動を一定のレベル以上かつバリデーション定義に沿って行えるように整備すること
- ② 観察研究を行う研究者(利活用者)が統合解析用データセットのデータ信頼性を確認する際、抽出元となるSS-MIX2標準化ストレージ内のRWDデータの信頼性を評価し各研究課題に応じてデータ利活用の可否を判断できる評価材料となり得ること

チェックリストの確認項目は、バリデーションの定義における「データの整合性の確認」として、検証条件におけるHIS及びSS-MIX2標準化ストレージのデータ件数(以下、データ件数)、及びデータ項目ごとの意味(以下、データ内容)とした。また、利活用者が、そのデータソースの信頼性を判断する客観的な指標として、HISデータとSS-MIX2データの一致率を記載することとした。なお、比較検証において不一致データが存在する場合、原因究明とともに、その経過についても追記する仕様とした。チェックリストVersion1は、施設毎に品質管理を進めるにあたり作成した土台となる資料であり、今後、実際に品質管理活動を行っていく過程で出てきた課題への対応や、個々の研究課題の必要に応じてデータ項目を追加するなど、適宜改定を行っていく予定である。

3.3 バリデーションに関連したツールの開発

上述したPJ1の活動と並行して、バリデーションに関連したツールの開発も実施した。SS-MIX2標準化ストレージは患者ID・診療日ごとのフォルダとデータファイルの組み合わせ

No	種別		メッセージ件数	データ件数	ヘッダ関連	データフィールド
1	病名	PPR-01	Ver.1	Ver.2	Ver.1	Ver.2 Ver.3
2	処方	OMP-01	Ver.1	Ver.2	Ver.1	Ver.2 Ver.3
3	注射実施	OMP-12	Ver.1	Ver.2	Ver.1	Ver.2 Ver.3
4	検体検査結果	OML-11	Ver.1	Ver.2	Ver.1	Ver.2 Ver.3
5	外来診察	ADT-12	Ver.3	Ver.3	Ver.3	Ver.3
6	患者基本	ADT-00	Ver.3	Ver.3	Ver.3	Ver.3
7	入院実施	ADT-22	Ver.3	Ver.3	Ver.3	Ver.3
8	退院実施	ADT-52	Ver.3	Ver.3	Ver.3	Ver.3
9	転棟転科	ADT-32	Ver.3	Ver.3	Ver.3	Ver.3
10	アレルギー	ADT-61	Ver.3	Ver.3	Ver.3	Ver.3
11	給食	OMD	Ver.3	Ver.3	Ver.3	Ver.3

図 2. バリデーションツールのバージョンと対応しているデータ種別の範囲

で診療データを保存する仕組みであり、検索性に乏しく、HIS データと突合して内容の確認をするにはあまり適さない。そのため、SS-MIX2 データと HIS データの整合性の確認を容易にすることを目的にしたツールを開発する必要があった。

PJ1 が活動を開始する以前の平成 30 年度に、MID-NET で行われていることを参考にして、トランザクションストレージを利用して SS-MIX2 のデータを Relational database (以下、RDB) に保存し、検索できるようにするツール (バリデーションツール Ver.1) の開発を名古屋大学主導で実施した。この時は、開発期間に限られていたこと、PJ1 設置前であり、まだどのようにバリデーションを行っていくかについて議論が全くなされていなかったことより、明らかに臨床研究に用いるだろうと想定された 4 データ種別 (処方オーダー、注射実施、検体検査結果、病名オーダー) を対象とした。また、各データ項目について何を確認するかの議論がまだ始まっていなかったこともあり、メッセージ件数を HIS のオーダー数と比較するためにメッセージヘッダを取得する機能のみを有するものであった。

PJ1 が設置され、データの品質管理に関する議論が開始されると、バリデーションツール Ver.1 では機能が足りないことが明らかとなり、機能の追加が求められることとなった。そこで、令和元年 9 月より仕様の検討が開始された。仕様の検討は、システムの納入を年度内に行わなければならない都合から、前項のデータ種別の選定を検討していた時期よりわずかに早い時期に行わなければならない、前述の 4 データ種別のデータ項目をある程度網羅した形で RDB に保存する機能を持ったツール (バリデーションツール Ver.2) を構築する形となった。現在は、データバリデーションチェックリスト Version 1 に記載されている全データ項目を RDB に取り込めるだけで

はなく、SQL などのプログラムなどを記載しなくてもデータバリデーションチェックリストに記載されている評価項目を確認できるようにするツール (バリデーションツール Ver.3) の仕様の検討を行っているところである (図 2)。

3.4 バリデーション作業の実施

令和元年度、実際の SS-MIX2 標準化ストレージのバリデーション作業を慶応義塾大学、名古屋大学、岡山大学で行った。バリデーションツール Ver.1 を用いて、SS-MIX2 標準化ストレージの 4 データ種別 (処方オーダー、注射実施、検査結果、病名オーダー) のメッセージ件数と HIS データのオーダー件数を比較した。今回はその中から名古屋大学の状況を報告する。

令和元年度に名古屋大学ではバリデーション作業を 2 回実施した。1 回目は PJ1 が設置される前の 7 月より開始した。バリデーションツール Ver.1 を用いるために必要な準備を行った後、バリデーションツール Ver.1 を使用し、1 週間分の SS-MIX2 データを取得。データを MS ACCESS に格納。HIS データは参照系 DB と直接 ODBC 接続し、SS-MIX2 データと SQL を用いて比較した。1 回目は処方オーダーと注射実施の比較を名古屋大学のスタッフ 2 名で行った。不整合の内容確認とその修正は HIS ベンダーに依頼した。

処方オーダーの比較の結果を図 3 に示す。HIS のオーダー件数は 11085 件、SS-MIX2 メッセージ件数は 7595 件で、一致率は 68.5% であった。不一致の理由は 4 つに分類でき、HIS のカスタマイズに起因するものが大部分で、残りは SS-MIX2 データを抽出した際の日付設定ミスによるものであった。注射実施もほぼ同様の結果であった。

1 回目のバリデーションを通して得られた課題のうち重要なものは、

- SS-MIX2 と HIS の比較をする際に、HIS の DB 構造、各フィールドの内容に関して理解が深くないと SQL 自体を作成することができず、出た結果の評価もできない
- 不整合を見つけても、ベンダーの協力がなければ原因の特定が困難

であった。両者とも、バリデーション実施にはベンダーの協力が不可欠であることを示しており、2 回目のバリデーションではその解決を図るため、ベンダーとバリデーション作業への支援についての契約を締結し、実施した。

2 回目のバリデーション作業では、処方オーダーと注射実施

HIS - SS-MIX2(TS) レコード比較 (処方)

- データの期間：1 週間分
- マッチング条件：患者ID、日付、データ種別

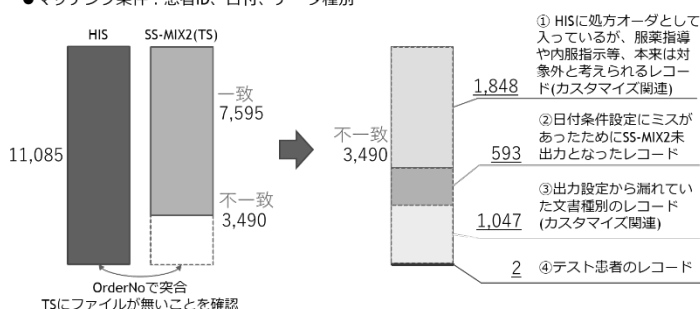


図 3. 処方のオーダーの比較結果

は不整合が修正されていたため、件数は一致していた。検査結果と病名オーダについては、1 度目の評価で出てきた不整合を修正し、2 度目の評価で SS-MIX2 メッセージ件数と HIS オーダ件数が一致していることを確認した。

3.5 その他の取り組み

前述した取り組み以外にも、臨中ネットに参加している 12 施設のデータソースとそのバリデーションの状況に関する調査を実施している。特に SS-MIX2 に関しては、サーバのバージョンも確認した。

また、日本 HL7 協会が行っている SS-MIX 適合性検証サービスの受審も九州大学と名古屋大学で実施した。その結果、検証サービスを受けた範囲内で、構造上は大きな不具合がないこと、出力データ形式についてはエラーの大部分が対応可能であること（現在、ほとんどの不具合は修正済）が確認できた。

4 PJ1 の今後の方向性

品質管理は定められた手順に従って実施し、必要なエビデンスを残し、後で適切に実施されたことを検証できることで保証されるとされている²⁾。そのため、品質管理の手順を定めそれを文書化すること、どのようにバリデーション実施の記録を残すかを決めることが今後必要となる。また、品質管理には継続性も重要であり、どのように日常のメンテナンスを実施していくかを決めることも必要となるであろう。

臨中ネットは今後、研究課題を実施していく予定であり、その過程で取り扱うデータ項目が拡張されることが想定されている。このため、データ項目の拡張時に品質管理対象をどのように定めるか、品質管理対象となった場合にどの評価項目を適用するかなどのポリシーを決めることも必要である。

また、各施設が臨中ネットで用いる予定のデータソース (SS-MIX2 であるかどうかに関わらず) 自体のバリデーション作業も同時に進めていかなければならないと考える。

PJ1 で今後行われる予定の取り組みは、データ品質管理について検討する目的で新たに設置された SWG4 で行われることになっており、引き続き着実に検討・作業を実施していく予定である。

5 PJ1 のまとめ

PJ1 は臨床研究課題に対応できるシステム構築のためのデータ品質管理の検討を目的として設置された。これまでに、バリデーションの定義、データバリデーションチェックリストの作成、バリデーションに関連したツールの開発、バリデーション作業の実施などを実施した。今後、データの品質管理に関する検討は SWG4 に引き継がれて行われていく予定である。

参考文献

- 1) Yamaguchi M, Inomata S, Harada S, et al. Establishment of the MID-NET® medical information database network as a reliable and valuable database for drug safety assessments in Japan. *Pharmacoepidemiol Drug Saf.* 2019 ; 28 : 1395-1404. doi:10.1002/pds.4879.
- 2) 山口拓洋. 品質管理・品質保証. *薬理と治療* 2020; 48 :13-16.