

厚生労働省標準規格 SS-MIX2 標準化ストレージのデータ品質改善の試み

中山 雅晴^{*1}、永島 里美^{*2}、大江 和彦^{*2}、山下 暁士^{*3}、白鳥 義宗^{*3}、
山下 貴範^{*4}、中島 直樹^{*4}、堤 英樹^{*5}、東海林 晋^{*5}、窪田 成重^{*5}

^{*1} 東北大学大学院医学系研究科医学情報学、^{*2} 東京大学医学部附属病院、^{*3} 名古屋大学医学部附属病院、
^{*4} 九州大学病院、^{*5} 富士通 Japan 株式会社

Proposal to Improve Data Quality of SS-MIX2 Standardized Storage

Masaharu Nakayama^{*1}, Satomi Nagashima^{*2}, Kazuhiko Ohe^{*2}, Satoshi Yamashita^{*3}, Yoshimune Shiratori^{*3},
Takanori Yamashita^{*4}, Naoki Nakashima^{*4}, Hideki Tsutsumi^{*5}, Shin Toukairin^{*5}, Narushige Kubota^{*5}

^{*1} Tohoku University Graduate School of Medicine, ^{*2} The University of Tokyo Hospital, ^{*3} Nagoya University Hospital, ^{*4} Kyushu University Hospital, ^{*5} Fujitsu Japan Limited

Standardized Structured Medical Information eXchange version 2 (SS-MIX2), a standard format of storage authorized by the Ministry of Health, Labor and Welfare (MHLW), has been widely used in database projects and health information exchanges because it allows medical data from various electronic medical record vendors to be stored in a common format and facilitates secondary use. Considering the purpose of these projects, quality assurance that the data has been transferred correctly is critical, but it is an expensive and labor-intensive process. Particularly, we believe that despite the experience and achievements of several similar projects, the methods for quality assurance of SS-MIX2 data have not been established, and that verification also imposes a significant burden on hospitals. We decided to analyze the causes of this inconsistency and examine appropriate countermeasures. First, we performed the following quality assurance assessments: (1) grammatical (whether the data transferred to SS-MIX2 adhere to SS-MIX2 specifications), (2) quantitative (whether the data in the hospital information system are equal to the volume of data transferred to SS-MIX2), and (3) qualitative (whether the data in the hospital information system corresponds to the data transferred to SS-MIX2). We discovered that not all these assessments had been sufficiently conducted and that there were no standard tools for performing them. By re-examining these issues, we aim to improve efficiency in terms of labor and cost by clarifying issues such as standard quality assurance as a vendor and establishing procedures and tools so that they can be expanded to other vendors.

Keywords: SS-MIX2, Quality management, Real-world data

1. はじめに

厚生労働省標準規格である Standardized Structured Medical Information eXchange version 2 (SS-MIX2) は異なる電子カルテベンダー由来の診療データを共通の様式で保存でき、2次利用も容易となるため、データベース事業や地域連携システムなど広く活用されている。¹⁾²⁾ その目的を鑑みれば、データが正しく移行されているという品質保証は必須かつ重要であるが、その徹底にはコストや労力がかかることが実情である。特に、今まで複数の同様なプロジェクトの経験と実績を持ちながら、その品質が定まらず、また検証作業に相応のコストと病院側へ少なからぬ負担を強いていることは問題であると考え、同一ベンダーのカルテを用いている複数の大学病院において検証を行い、不整合の原因分析と対策について検討してみることにした。

2. 方法・結果

2.1 現状把握

電子カルテがパッケージ通りではなく、各施設の実情に合わせたカスタマイズが入ることは多い。そのため、施設ごとの設定や仕様を確認して SS-MIX2 ストレージにデータを出力することが必要であることから、単一なやり方による品質保証は困難である。しかしながら、同一のベンダーであればおよそ

同一のデータ形式で運用されていることが予想され、ユーザー側としては毎回出力の度に手直しが必要であることは理解に苦しむところである。そのため、まずはこういった方法をとっているのかをヒアリングしたところ、SS-MIX2 の品質検証方法として確立されている手順はないということが明らかになった。基本的には、各施設での設定等を確認し問題なく出力できると判断すればテストデータで確認して、実際に電子カルテから SS-MIX2 へ出力を行う。そしてデータにエラーが発見されるたびに原因究明とその修正を行うというものであった。ある程度のトライアンドエラーは仕方がないものの、単に繰り返すやり方では、コストが嵩み、また多くのエラーが新しく生じることが不可避なので、施設側としては、品質管理の方法の統一化をはかり、どこまでがパッケージ共通として単一の手法が確立でき、さらに施設ごとにどの点を事前に把握すればエラーを未然に防げるか明らかにすることを要望、今回ベンダー側とユーザー側とで協力して作業を行うことになった。

2.2 方針

まず、必須と考えられることとして、①文法チェック(SS-MIX2に移行したデータがSS-MIX2の仕様に正しく準じているか)、②定量チェック(病院情報システムにおけるデータ数とSS-MIX2に移行したデータ数が等しいか)、③定性チェック

(病院情報システムにおけるデータと SS-MIX2 に移行したデータの内容や値が等しいか)を挙げてみたところ、その観点から体系立てていなかったこと、現テストパターン網羅性に限界があること、先行事業のノウハウが活かされていないことがわかった。そのため、検証ツールの作成や適切なサンプルデータの選択と網羅性の確保、先行施設を踏襲したマスタ設定のチェック項目一覧を作ることなどが決まり、ベンダーとしての標準品質保証、各病院によるカスタマイズや設定などの問題などを詳らかにし、労力面においても費用面においても効率化をはかることを目的とした。

2.3 文法チェック

文法チェックは、SS-MIX メッセージファイルが、SSMIX の仕様則に則っているか確認することを目的とした。文法チェックツールを用いて、1か月分の SS-MIX メッセージファイルを対象にフォルダ構造、階層構造、ファイル名、ファイル内の構造、セグメント名、セグメント構造、フィールドの型、Coded with Exceptions (CWE) 型の格納成分、Coding System (CS) 名ととりうるコード、名称の関係性、必須入力フィールド、フィールド長などを確認した。現時点で、共通的に出現する課題は11件、固有の設定・データに依存する課題が8件、統一診療科マスタ設定のされていないことによる課題が4件と明らかになっており、修正情報適用により課題をクリアできる状態と考えられた。

詳細を挙げると各施設共通で発生のものであり、OMP-01(RXE-7)の「JHSP 表 0001 依頼者の処方指示」マスタが1.2f 対応でないこと、ZPR-5 はオプション項目のため未編集であること、ADT-00(IN1-15)保険分類が「PE」(公費)の場合に県番号を出力することなど、SS-MIX アプリケーションの対応が必要となった。一方で、SS-MIX アプリケーション以外のものとして、ADT-00(PID-5)は電子カルテで患者氏名(漢字氏名)姓名の出力が、電子カルテの入力不足のため出力されないことや ADT-00(IN1-2)に廃止された保険が出力されることなどが指摘された。また施設固有に発生したものとして、ADT-00(OBX-5)身長・体重に数値でない値が入っていることで関連する調剤量が出力されていなかったこと、ADT-00(PID-33)患者情報の移行データの最終更新日時は電子カルテ上 Null 登録されると出力されないことなどがあった。要検討項目として、統一診療科マスタが設定されておらず、エラーとなること、ADT-22, ADT-12, OMP-01, OMP-02 の PV1-10(診療部門)、外来時の PV1-3(患者所在場所)、RXE-42(患者への配達場所)に対し、使用者定義表における「0069-診療部門」や「0302-ケアの場所」を定めるべきであることが挙げられた。

2.4 定量チェック

量的監査として、作成されていない、または不要に存在している SS-MIX メッセージがないか確認する。結果として、保有データと SS-MIX のデータ件数を比較する件数チェックツールを作成した。しかしながら、分析に時間を要し、今の所1施設のみにて結果分析を行っている状況である。方法として、月次処理にて結果を出力、直近の1か月分の SS-MIX トランザクションストレージと、電子カルテのオーダー件数等を双方向にチェック(メッセージ単位の比較)、日次処理で SS-MIX のトランザクションストレージと SS-MIX エラーファイルを取得し、月次処理にて HIS のレプリカ情報取得と比較処理・繰り越し処理を実施するなどして結果を比較した。

SS-MIX と電子カルテの件数差分となった21種類の課題

のうち、7種類にパッケージ不具合が確認され、修正資源リソースを調整中である。ツール上比較困難なケースについては「差異なし」と取り扱っていたが、その後の話し合いでユーザーの方から、それは「差異あり、ただし〇〇のケースのため課題調査対象としない」というように、「差異は挙げておいて、どう扱うべきかを論じるよう」取り扱いを変更することの意見があり、表示を変える方針とした。これらから、作成したツールはメッセージ件数の過不足が保証可能なツールであると判断できたが、実際に差異があった場合の調査のやりとり時間に時間を要してしまうため、今後全施設展開する際の課題となる。また、同時に新規の SS-MIX 課題が検出されたため、修正情報のリリース時期・適用時期を調整中である。実施済他施設の内容分析と、今後の展開時の分析手法も検討中である。

SS-MIX アプリケーションを修正する必要があるものとして、OMG-01, OMG-02 で、生体系オーダを特定の日付から日未定に変更した際に、修正前の日付の SS-MIX メッセージが残る点を、元メッセージの取り消しを実施するよう修正した。OMP-0 では注射オーダを実施中止(未実施確認)した際、出力処理がスキップされていた点、OMP-02 の注射オーダの入力後 SS-MIX の出力処理前に実施中止(未実施確認)がされた場合に、SS-MIX への出力がスキップされる点は出力ソフトの修正が必要であった。さらに、SS-MIX システムマスタ設定の修正として、移動実施系ではシステムマスタの不備でメッセージ作成の対象となっておらず、メッセージが作成されていなかった。また、ADT-21 では入院予定削除が処理対象となっておらず、取り消されたオーダの SS-MIX メッセージが残っていた。SS-MIX アプリケーションの改善検討項目として、ADT-61 では処理遅延による先月分比較時に SS-MIX 側情報が十分に取得できていない点があった。他にも、統一診療科マスタ設定があり、かつ医科・歯科の医療機関コードを別コードを出力する設定の場合、MSH-4(送信施設)に出力している医療機関コードと診療科の医科・歯科の組み合わせを誤る場合があった。

2.5 定性チェック

質的監査は、データの値(数字、文字列、コード)について、電子カルテの内容と SS-MIX の内容に誤りがないかを確認することを指す。方法として、詳細チェックツールを作成し、今までのところ、1次構築先行分を1施設に適用し、一定の結果が出たがツール不具合も多く、課題対応と残分対応をリスケジュールしている。

件数チェックツールにて一致の確認された SS-MIX メッセージファイルを対象に、各フィールドの値を確認する必要がある。特定の1か月に来院または入院している患者リストを用い、その患者の期間内のオーダー情報に対してチェックし、ツール実行時に SS-MIX トランザクションストレージから作成した csv ファイルと HIS レプリカデータベースから値を取得し、固有の変換ロジックを介して各フィールドの内容について文字列比較を実施する方針である。不具合と考えられるものについては、それぞれに開発元へフィードバックし修正情報を管理する。また、ツールの都合で自動確認の困難なものをピックアップ、データ要因の回避不能と考えられるものはその旨を明確にしてユーザーと相談する案件となる予定であり、ノウハウ蓄積も行う。

1施設での結果として、323 フィールド(全体の 80%)が比較実施でき、うち 138 フィールドが課題なく 100%一致が確認された。これも差異があった場合には修正作業により全施設展開には時間がかかりそうである。発見された課題には、以

下がある。OMP-12(注射実施)の RXA-4(注射実施終了日時)の値が空白となる。PPR-01(病名)の ZPR-5(病名交換用コード)の値が空白となる。OMP-01(処方オーダー)の RXE-10(調剂量)が特定の用法の場合に空白となる。他にも、ADT-12(外来受付)の PV1-44(来院時間)の時間部分の取得する値に不備がある。OMP-01(処方)の不均衡投与コメントが前システムの記載方法のままのオーダーがあり、RXE-3(与薬量-最小)、RXE-3(与薬量-最大)の値に誤りがある。

2.6 マスタ品質向上

マスタ管理に関しても、明細チェックツールを作成し、導入手順における組み込みの検討を進めていくこととなった。目的は、先行事例のノウハウ等から、マスタ設定におけるチェック項目一覧を作成し、設定不備を未然に防ぐことである。そして、システムマスタチェック項目一覧を元に、各施設の SS-MIX システム設定マスタ(HL7 マスタ)の設定内容を確認する予定である。

3. 考察

今回、SS-MIX2 の品質管理へベンダーとユーザーが協力して、効率的かつ普遍的な解決を目指して取り組んでいる内容を報告した。まだ途中の段階であり、未だ断片的であるが、文法、量的監査、質的監査、マスタ品質管理など進むべき大きな方向性は定めた。実際にツール作成という形で共通の手順へつなげている。本取り組みにより、今後ベンダー、ユーザーともにお互いの負担を減らしながら正確なデータ移行を担保し、各施設がリアルワールドデータを安心して活用できるよう進めて行きたい。まずは同一ベンダーによる手順の確立を目標とし、最終的には他ベンダーに展開していければと考えている。

参考文献

- 1) Yamaguchi M, Inomata S, Harada S et al. Establishment of the MID-NETR medical information database network as a reliable and valuable database for drug safety assessments in Japan. *Pharmacoepidemiol Drug Saf.* 2019; 28: 1395-1404.
- 2) Ido K, Nakamura N, Nakayama M. Miyagi Medical and Welfare Information Network: a backup system for patient clinical information after the Great East Japan Earthquake and Tsunami. *Tohoku J Exp Med* 2019; 248: 19-25.

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

諸橋朱美^{*1}、山下暁士^{*2}、今泉貴広^{*1}、宮城英毅^{*2}、白鳥義孝^{*2}

^{*1} 東海国立大学機構 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、

^{*2} 東海国立大学機構 名古屋大学医学部附属病院メディカル IT センター

Exploration of quality assurance methods for RWD in SS-MIX2 and implementation of reliability validation

Akemi Morohashi ^{*1}, Satoshi Yamashita ^{*2}, Takahiro Imaizumi ^{*1}, Hideki Miyagi ^{*2}, Yoshimune Shiratori ^{*2}

^{*1} Data Coordinating Center, Department of Advanced Medicine, Nagoya University Hospital, Tokai National Higher Education and Research System, ^{*2} Medical IT Center, Nagoya University Hospital, Tokai National Higher Education and Research System

Abstract

Background: There are increasing expectations for the secondary use of electronic medical records as real-world data (RWD) stored in hospital information systems (HIS). However, there are challenges in data quality and standardization to promote their utilization for clinical research. The purpose of this study is to quantitatively investigate the quality of RWD stored in the Standardized Structured Medical Information eXchange2 (SS-MIX2) along the evaluation axes. Methods: Using the quality validation tool we developed, we compared the data from HIS and SS-MIX2 to check for consistency. The target data consisted of 337 items commonly used in clinical research out of approximately one month's worth of RWD from Nagoya University Hospital. We conducted two cycles for each of the two data transmission methods (four cycles in total). Each cycle consisted of measuring the data concordance rate, investigating the causes of data discrepancies, sorting out the problems, and solving the problems. Results: There were discrepancies in the RWD stored in the two systems. The data concordance rate was 88% on the first cycle and 92% on the second cycle for data-setup transmission, and 84% on the first cycle, and 91% on the second cycle for the real-time data transmission. Conclusion: The validation tool enabled us to quantify the "consistency" of the quality of the RWD stored in the SS-MIX2 storage.

Keywords: SS-MIX2, Real World Data, Electrical health records, Data Quality, Reliability

1. 背景

米国では、2016年に“21st Century Cures Act”が可決され、高騰する医薬品の開発コスト削減等に向け、Real-world Data (RWD) を活用する取組みが行われている。日本国内でも2018年に次世代医療基盤法が施行された他、疾患登録システムを活用した Clinical Innovation Network¹⁾の推進など、RWD 利活用による医療への貢献が期待されている。一方、各医療機関に蓄積された実診療データである RWD は、表現形式が統一されていないことや、データの品質が保証されていないこと等の課題を包含する。これは、電子カルテを含む日本の病院情報システム(Hospital Information System: HIS)が保険償還と臨床診療を目的に独自の発展を遂げ、研究等への 2 次利用目的で設計や運用が行われていないことが一因である。

臨床研究中核病院は、現状の課題を克服し、RWD を用いた新たな臨床研究手法による医療への貢献、安価で効率的な創薬の牽引役となることが求められている。しかしながら、このようなデータベース研究に関するデータ品質の管理方法や品質指標は具体的には確立されていない。一方、独自の品質管理方法を実施することで、信頼性の高い RWD の 2 次利用を実現しているシステムも存在する。独立行政法人医薬品医療機器総合機構(PMDA)が運営する MID-NET²⁾では、システムの本格運用前に協力医療機関の HIS データと、Standardized Structured Medical Information eXchange2 (SS-MIX2) 標準化ストレージを介して格納された統合データソースの間で信頼性検証が行われた。検証初期にデータの件数や内容に複数の齟齬を確認したが、その後の品質管理活動実施によって一致率が上昇したことを報告し

ている³⁾。

2018 年から名古屋大学医学部附属病院を含む臨床研究中核病院でも、MID-NET²⁾システム相当の高品質な RWD の利活用を目標に、国立研究開発法人日本医療研究開発機構の「医療技術実用化総合促進事業」(以下、本事業)における臨床研究中核病院機能実装の取組みの 1 つとして、SS-MIX2 標準化ストレージなどを用いた RWD の 2 次利用に向けた基盤構築を行っている。先行事例である MID-NET²⁾での信頼性検証結果が示すように、データ品質管理未実施の SS-MIX2 標準化ストレージに格納された RWD の品質は不透明であり、まずはそれを明らかにすることを目的に、本事業の取組みの一環として品質管理方法を検討することとなった。その中で、名古屋大学医学部附属病院は、SS-MIX2 標準化ストレージ品質検証ツールの開発を担った。

本研究では、本事業の取組みにて検討された SS-MIX2 標準化ストレージに格納された RWD の品質管理方法と開発した品質検証ツールの説明、名古屋大学医学部附属病院の SS-MIX2 標準化ストレージに格納された RWD の信頼性検証の実施結果を報告する。

2. 目的

本研究では、RWD の品質を検証するために開発したツールを用いて、名古屋大学医学部附属病院の SS-MIX2 標準化ストレージへ格納された RWD の品質を、「一貫性」を評価軸として明らかにする。また、検証結果から RWD が包含する課題を明確にすることで RWD 2 次利用の推進に繋げる。

3. 品質管理方法の検討

臨床研究中核病院が参画する「医療技術実用化総合促進

事業」内に発足された品質管理サブワーキンググループにて決定した方針を以下に示す。

3.1 品質管理対象の範囲

SS-MIX2 標準化ストレージに格納された全 36 データ種別中、特に臨床研究に必要とされる 11 データ種別(患者基本情報の更新、外来診療受付、入院実施、転科・転棟実施、退院実施、アレルギー情報の登録・更新、病名情報の登録・更新、食事オーダー、処方オーダー、注射実施通知、検体検査結果通知)とした。詳細項目は、SS-MIX2 ヘッダ 5 項目とメッセージに含まれるのべ 159 フィールド(266 成分)である。

3.2 品質評価軸の検討

データ品質とは、高品質なデータに関する特性とデータ品質を測定し改善するために利用されるプロセスを指す⁴⁾。前者は、データ利活用者の要求に応じて決定される多次元性を有しており、一元的に品質を評価できるものではない。評価軸を設定し、それに基づいて検証することが一般的である。

本事業における品質評価軸の選定では、RWD を用いた臨床研究を行う研究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 の信頼性を評価し、各研究課題に応じたデータ利用の可否を判断できる評価材料となり得ることを前提条件とした。

検討の結果、第一の目標としてデータの見え方や意味合いなど表現に関するデータ品質評価軸となり得る「一貫性」を評価軸として採用した。一貫性とは、ISO/IEC 25012⁵⁾、DMBOK 第 2 版⁴⁾ともほぼ同様の定義であり、「同じ実体を表す 2 つ以上のデータが一貫して表現されていること」とされている。その他の評価軸については、本事業において引き続き検討を行う。

3.3 評価の見える化

送信元となる HIS のデータが「SS-MIX2 標準化ストレージ仕様書 Ver.1.2f」の定義どおりに SS-MIX2 標準化ストレージに格納されている「一貫性」の度合いを客観的に評価できる形式で示すことが、RWD の 2 次利用に必要であると考えた。そこで、2 つのデータの一致率を算出することに加え、各臨床研究中核病院での HIS 運用を考慮した見解を「問題の有無」で示すこととした。また、多施設データの品質状況を一定程度等しく評価する必要があることから、各施設が評価を同じ対象・同じ評価項目で行えるよう、バリデーションの対象範囲・項目を記載した臨床研究中核病院共通の「データバリデーションチェックリスト(SS-MIX2 使用時)」を作成した。

4. 品質検証ツールの開発

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で RWD を保存する仕様であり、複数の患者 ID を含むデータや患者 ID を検索キーとしないデータの検索性は十分でない。そのため、SS-MIX2 標準化ストレージに格納されたデータ(以下、SS-MIX2 データ)

と HIS に格納されたデータ(以下、HIS データ)の一貫性の確認を容易にするツールが必要であると考え開発に至った。ツール開発は、MID-NET[®]でのデータ品質管理方法を参考に行い、データバリデーションツールと命名した(Figure.1)。

データバリデーションツールの仕様を以下に示す。

4.1 HIS データの抽出機能

電子カルテのレプリケーション DB から対象データを抽出し、比較実施を行うための temporary テーブルを作成する。

4.2 HIS マスタ取得・コード付与機能

院内独自に設定された薬剤や検体検査のローカルコードや HOT13 や JLAC10 などの標準コードが格納されたマスタを HIS から取得する。4.1.1 で抽出した HIS データにコードを付与する。

4.3 SS-MIX2 標準化ストレージ仕様書 ver.1.2f コードへの変換機能

4.2 を実施した HIS データを、SS-MIX2 標準化ストレージ仕様書 ver.1.2f コードに沿って変換し、SS-MIX2 データと突合比較が可能な形式にする。

4.4 SS-MIX2 データ抽出機能

SS-MIX2 標準化ストレージサーバ内のトランザクションストレージに格納されたデータを relational database に保存し、データ検索・抽出を可能にする。SS-MIX2 ヘッダは全 10 項目(SS-MIX 識別、バージョン、医療施設 ID、患者 ID、診療日、データ種別、オーダー No.、処理区分、診療科コード、トランザクション日時)から構成される。このうち、トランザクション日時が最新かつ、処理区分が「INS」のデータを抽出対象とする。

4.5 HIS データと SS-MIX2 データを突合比較機能

データに含まれる患者 ID やオーダー No.、診療日等を外部キー情報(以下、比較キー情報)として、突合比較検証フローに沿ってデータの検証を実施する(Figure.2)。第一段階目は、SS-MIX2 メッセージヘッダの患者 ID、オーダー No. を比較キー情報に、HIS データと SS-MIX2 データをリレーションし、データの件数を比較する。第二段階目では、SS-MIX2 ヘッダの比較キー情報が一致した SS-MIX2 ヘッダに含まれる項目の内容と、そのヘッダに附随する SS-MIX2 メッセージのセグメント件数を比較する。第三段階目では、セグメント件数が一致した SS-MIX2 メッセージフィールドの内容を比較する。各段階で不一致となった SS-MIX2 データは、次の段階の突合比較

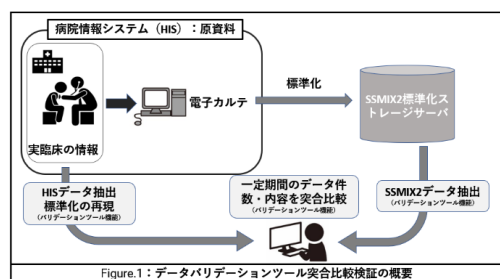


Figure.1: データバリデーションツール突合比較検証の概要

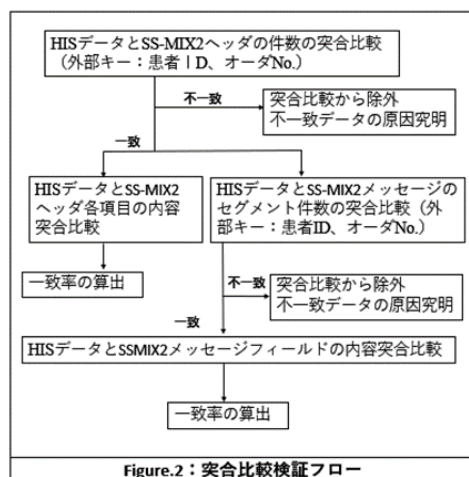


Figure.2: 突合比較検証フロー

対象としない。

5.RWD 品質検証

本研究では、3.1 で示した SS-MIX2 項目のうち、名古屋大学医学部附属病院にて定めた検証条件に該当するデータと、対応する電子カルテデータを、開発したデータバリデーションツールを用いて突合比較した。

1 サイクルをデータバリデーションツールの実行、一致率の確認、データ不整合の原因究明、「一貫性」を評価軸とする突合比較項目ごとの「問題有無の検討」、課題整理、可能な範囲の課題解消とした。そして、SS-MIX2 標準化ストレージへのロジックの異なる 2 通りデータ送信方法ごとに各 2 サイクル実施し、第 1 サイクルおよび第 2 サイクルでのデータ一致率の変化、SS-MIX2 データ送信方法の違いによる不一致原因の検証を行った。本研究における検証条件を以下に示す。

5.1 SS-MIX2 標準化ストレージサーバ

品質検証の対象となる SS-MIX2 標準化ストレージサーバは、Ver.1.2d から Ver. 1.2f へバージョンアップした上で、MID-NET®で見つかった課題を解消するためのプログラムが適用された仕様とした。このトランザクションログから SS-MIX2 データを作成し、比較対象データとして用いた。

5.2 病院情報システム

名古屋大学医学部附属病院独自のカスタマイズが施された電子カルテシステムである富士通社の HOPE EGMAIN-GX を使用した。システム内にあるレプリケーション DB から検証用 HIS データを作成し、比較対象データとして用いた。

5.3 データ比較検証ツール

開発した「データバリデーションツール Ver3.0.3」を用いた。

5.4 比較対象データ

富士通標準パッケージ仕様で SS-MIX2 標準化ストレージサーバへ送信対象のデータ種別やセグメント、フィールドとした。

「データバリデーションチェックリスト(SS-MIX2 使用時)」に記載された比較検証項目のうち、SS-MIX2 ヘッダ件数、ヘッダ内の 5 項目(外部キー含む)、メッセージに含まれる 44 セグメント中 41 セグメントの件数と、266 成分中 230 成分の内容の計 337 項目が該当した。

SS-MIX2 メッセージ作成には、Data-setup 送信(過去の指定した期間の HIS データをまとめて SS-MIX2 データに変換する方式)と、Real-time-data 送信(HIS のデータが生成されるたびに SS-MIX2 データに変換する方式)の 2 つの異なるロジックによるデータ送信方法が存在する。本研究では、SS-MIX2 標準化ストレージへ各送信方法で格納された SS-MIX2 メッセージを対象とした。テスト患者データや更新差分データ(SS-MIX2 メッセージ作成と比較用 HIS データ抽出を行っている間に変更や削除などの更新が行われたデータ)はこの限りではない。Data-setup 送信では、第 1 サイクル、第 2 サイクル共に 2019 年 1 月 1 日～2019 年 1 月 31 日(外来受診の受付のみ 2020 年 3 月 1 日～2020 年 3 月 31 日)、Real-time-data 送信では、1 サイクル目を 2021 年 2 月 13 日～2021 年 3 月 12 日、2 サイクル目を 2021 年 8 月 5 日～2021 年 8 月 24 日を対象の日付とした。比較対象データは診療日を基本とし、上記の条件を踏まえて抽出された。

5.5 問題の有無の判断基準

データバリデーションツールにて、不一致と判断されたデータについて、検証項目ごとに問題の有無の検討を実施した。例えば、データ表示が異なる場合はツール上、不一致となるが、大文字・小文字の違いなど意味的に同一なものは「一貫性」があると判断し、「問題なし」とした。また、バリデーションツールの検証ロジック不備により、結果が不一致となったデータに関しては、目視によって一貫性が確認できた場合、「問題なし」とした。

6. 検証結果

比較検証の結果、Data-setup 送信の第 1 サイクル、第 2 サイクル、Real-time-data 送信の第 1 サイクル、第 2 サイクルにおいて、それぞれ 377 項目中 296 項目、310 項目、284 項目 307 項目が問題なしと判断され、割合は 88%、92%、84%、

	Data-setup送信		Real-time-data送信	
	1サイクル	2サイクル	1サイクル	2サイクル
ADT-00	92%	96%	92%	92%
ADT-12	71%	71%	71%	71%
ADT-22	100%	100%	93%	100%
ADT-42	100%	100%	87%	100%
ADT-52	100%	100%	100%	100%
ADT-61	94%	94%	97%	100%
PPR-01	95%	95%	92%	92%
OMD	81%	81%	81%	81%
OMP-01	79%	95%	82%	95%
OMP-12	87%	91%	73%	89%
OML-11	86%	90%	76%	88%
合計	88%	92%	84%	91%

Figure.3 比較検証項目 “問題なし”の割合

91%であった(Figure.3)。

SS-MIX2 ヘッダの件数検証では、課題解消に向けた取り組みにより「問題なし」と判断できるデータ種別が増加した。一致率については、外来診察受付(ADT-12)で改善がみられた(Figure.4)。これは SS-MIX2 標準化ストレージへ送信しない設定である、受付状態が未受付の電子カルテデータが、比較対象の HIS データに含まれていたことが原因あり、データバリデーションツールの設定変更を実施した。食事オーダー(OMD)については、SS-MIX2 標準化ストレージへ送信する HIS データ日付が本来の仕様とは異なっていることが判明した。

SS-MIX2 ヘッダの内容検証では、送信方法に関わらず、外来診察受付(ADT-12)、食事オーダー(OMD)の項目「診療日」にて不整合を確認した。原因究明の結果、SS-MIX2 送受信プログラムが原因であり、2021 年度内に修正プログラムを適用することとなった。

SS-MIX2 セグメント件数検証では、Real-time-data 送信の第 2 サイクルにおける注射実施報告(OMP-12)にて不整合が解消された(Figure.5)。不一致が確認された HIS データの薬品名には共通して成分セパレータ「^」が含まれおり、データバリデーションツールが正常に作動しなかったことが原因であったため、ツールの改修を実施した。Data-setup 送信・Real-time-data 送信共に不整合が確認された患者基本情報の更新(ADT-00)・アレルギー情報の登録・更新(ADT-61)においては、データバリデーションツールの突合ロジック不備が原因であることが判明した。病名情報の登録・更新(PPR-01)

での不整合は、歯科病名などで1病名の中で病名コードが複数使用される病名で起きていることがわかった。この場合、電子カルテ側では複数病名を1行に、SS-MIX2 標準化ストレー

を行わない。しかし、緊急で診療が必要な場合、事前に準備している患者 ID(年齢、性別は既定値をセット済み)を利用してオーダを行い、後にその患者 ID を修正することで対応している。この際に、Real-time-data 送信であるとオーダが発生し

Data-setup送信		1サイクル		2サイクル	
		一致率 (%)	問題の有無	一致率 (%)	問題の有無
ADT-00	ヘッダ件数	49.1	あり	100.0	なし
ADT-12	ヘッダ件数	94.2	あり	99.6	あり
ADT-22	ヘッダ件数	100.0	なし	100.0	なし
ADT-42	ヘッダ件数	100.0	なし	100.0	なし
ADT-52	ヘッダ件数	100.0	なし	100.0	なし
ADT-61	ヘッダ件数	96.4	あり	96.9	あり
PPR-01	ヘッダ件数	76.3	あり	76.8	あり
OMD	ヘッダ件数	96.1	あり	97.6	あり
OMP-01	ヘッダ件数	97.9	あり	97.9	あり
OMP-12	ヘッダ件数	100.0	なし	100.0	なし
OML-11	ヘッダ件数	100.0	なし	100.0	なし

Real-time-data送信		1サイクル		2サイクル	
		一致率 (%)	問題の有無	一致率 (%)	問題の有無
ADT-00	ヘッダ件数	90.0	なし	98.2	なし
ADT-12	ヘッダ件数	93.1	あり	99.4	あり
ADT-22	ヘッダ件数	99.9	あり	100.0	なし
ADT-42	ヘッダ件数	99.9	あり	100.0	なし
ADT-52	ヘッダ件数	100.0	なし	100.0	なし
ADT-61	ヘッダ件数	74.4	あり	100.0	なし
PPR-01	ヘッダ件数	82.3	あり	56.4	あり
OMD	ヘッダ件数	78.0	あり	82.3	あり
OMP-01	ヘッダ件数	96.6	あり	95.2	あり
OMP-12	ヘッダ件数	99.2	あり	98.6	あり
OML-11	ヘッダ件数	99.2	あり	99.2	あり

Figure.4 SS-MIX2ヘッダ件数 検証結果

Data-setup送信				
データ種別	セグメント	1サイクル一致率 (%)	問題の有無	2サイクル一致率 (%)
ADT-00	INI件数	63.4	あり	70.4

Real-time-data送信				
データ種別	セグメント	1サイクル一致率 (%)	問題の有無	2サイクル一致率 (%)
ADT-00	OBX件数	99.9	あり	99.9
ADT-00	INI件数	57.2	あり	56.1
OMP-12	ORC件数	99.9	あり	100
OMP-12	RXA件数	99.9	あり	100
OMP-12	RXR件数	99.9	あり	100

Figure.5 SS-MIX2セグメント件数 検証結果

ジ側では、病名コードごとに2行のPRBセグメントを出力する仕様であることから件数が不一致となる。不一致データの目視による確認を実施した結果、該当データは全て同様のケースであり、意味合いとして同等であったことから「問題なし」と判断した。

SS-MIX2 メッセージの内容検証では、データに「一貫性」があり「問題なし」とした項目が全体で 90%を占めた(Figure.6)。一方で、日時情報、年齢、性別、薬剤名、数量など複数の項目で不整合を確認した。原因究明により、SS-MIX2 送受信プログラムや各種システムの不備、本学特有の電子カルテ運用が起因となることが判明した。また、Data-setup 送信では発生せず Real-time-data 送信のみで起こり得る不整合が確認された。これは、名古屋大学医学部附属病院独自の電子カルテ運用による「年齢」、「性別」の不整合である。名古屋大学医学部附属病院では、院内で生まれた新生児は通常患者登録

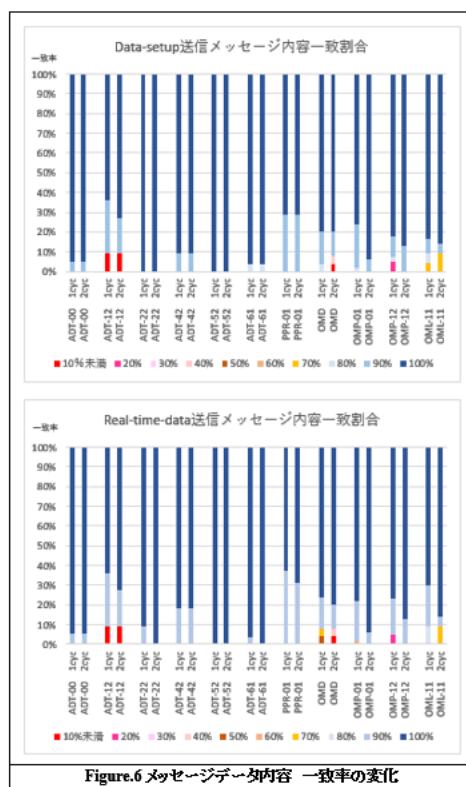


Figure.6 メッセージデータ内容 一致率の変化

た時点で SS-MIX2 メッセージが作成され、そこには事前に準備した患者 ID と患者情報が格納される。その後、電子カルテでは、患者 ID や年齢や性別等が修正され、そのデータで保存される。この場合、患者プロフィールのみの修正となるため、SS-MIX2 のオーダ情報は更新されないことから不整合となる。

本研究にて確認された課題は 5 つに分類分けすることができた(Figure.7)。不一致の原因分類ごとの課題件数は「マスタの不整備や不整合による問題に起因するもの」(9 件→2 件)、「SS-MIX2 送信プログラムの不備に起因するもの」(6 件→4 件)、「突合ロジックの課題・作業手順等、バリデーションツールの仕組みに起因するもの」(20 件→13 件)、「HIS データに問題が潜んでいたことによるもの(電子カルテのイレギュラーデータ・運用など)」(2 件→2 件)となり、第 1 サイクルから第 2 サイクルにかけてマスタ修正を中心に課題解決に取り組んだ

Data-setup送信 Real-time-data送信	1サイ クル	2サイ クル	1サイクル 実施後 の解決数	2021年度 対応完了 予定
課題分類	課題数	課題数		
マスタ起因	9 件	2 件	7 件	2 件
送受信PG起因	6 件	4 件	2 件	3 件
ツール・作業起因	20 件	13 件	7 件	13 件
電子カルテ起因	2 件	2 件	0 件	0 件
原因不明	3 件	4 件	0 件	—

Figure.7 課題分類

内容が結果として現れる形となった。残りの課題についても対応計画を立て、本年度内に対応予定である。

7.考察

品質の検証では、本事業で決定した SS-MIX2 の品質管理対象項目において、指定期間を対象とした「一貫性」の度合いを数値化し示すことができた。また、不整合のあった項目から RWD が包含する問題を抽出することができた。これは、研修者が RWD の利用可否を決定する際の重要な判断材料になると考える。

ただし、今回明らかにできたのは、SS-MIX2 仕様書に定義された一部の項目の「一貫性」のみである。前述したように RWD に対する評価軸や要望は利活用を実施する研究者の要求に応じて複数存在し得るものであり、本研究だけでそれを満たすとは考えにくい。研究者の要求を捉え、品質評価軸や評価項目を拡充していくことが今後の課題の一つである。

不一致データの原因究明においては、システム全体の詳細仕様を把握していなければ解明できないことも多く、システムベンダの協力が必要不可欠であることがわかった。本学の電子カルテ運用によるデータの不整合については、システムベンダが実施するサンプルデータを用いた SS-MIX2 標準化ストレージサーバのシステムバリデーションのみでは発見できない事象であり、RWD を用いてデータバリデーションを実施することの意義の1つであると考ええる。また、SS-MIX2 送受信プログラムの不備も見つかっており、改修されたプログラムは他施設へも適用される予定である。本学のみならず、他施設の RWD の品質が向上されることは、本研究の大きな成果といえる。

本研究では、Data-setup 送信と Real-time-data 送信、異なる2つの SS-MIX2 送受信方法で、各 2 サイクルずつ検証を行った。これは一般的な品質管理業務で実施されている PDCA サイクルの方法を再現するためであった。結果として、SS-MIX2 送受信方法で異なる傾向が得られ、各第 2 サイクルが「一貫性」の度合いが高かったが、第 2 サイクルにて新しく得られた課題もあった。データ品質を向上させていくためには、継続した RWD の品質検証・管理作業が必要であると言える。

8.結論

本研究により、開発したバリデーションツールを用いて SS-MIX2 標準化ストレージに格納された RWD の品質を、「一貫性」を軸にみえる化することができた。今後は、ツールを改良し他施設へ普及させること、評価軸を増やし RWD の品質をより詳細に評価できる基盤を整えることが課題である。

謝辞

本研究は、国立研究開発法人日本医療研究開発機構 2020 年度「医療技術実用化総合促進事業」の助成を受けたものです

参考文献

- 1) クリニカル・イノベーション・ネットワーク (CIN) の背景概要
<https://www.mhlw.go.jp/file/05-Shingikai-10801000-Iseikyoku-Soumuka/0000095014.pdf>. Accessed at August 19, 2021.
- 2) MID-NET.
<https://www.pmda.go.jp/safety/mid-net/0001.html>. Accessed at August 19, 2021.
- 3) Satomi INOMATA, Maori ITO, Kaori HIRATA, et al.
MID-NET® 本格運用に向けた取り組み The latest Activities toward MID-NET® Implementation in FY2018. RSMP Sep

2017 : vol.7 no.3, 215-224,

4) DAMA International (Data Management Association).

DAMA-DMBOOK Second Edition. Technics Publication. 2018

5) ISO/IEC 25012

<https://iso25000.com/index.php/en/iso-25000-standards/iso-25012>. Accessed at August 30, 2021.

降圧薬の種類による腎保護効果

－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－

今泉 貴広^{*1}、諸橋 朱美^{*1}、山下 暁士^{*2}、楯塚 八千代^{*1}、安藤 昌彦^{*1}、白鳥 義宗^{*2}

^{*1} 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

^{*2} 名古屋大学医学部附属病院メディカル IT センター

Renoprotective Effect of Antihypertensive Agents for Patients with CKD - Real-World Evidence Project among Core Clinical Research Hospitals -

Takahiro Imaizumi^{*}, Satoshi Yamashita^{*2}, Yachiyo Kuwatsuka^{*1}, Masahiko Ando^{*1}, Yoshimune Shiratori^{*2}

^{*1} Data Coordinating Center, Department of Advanced Medicine, Nagoya University Hospital

^{*2} Medical Information Technology Center, Nagoya University Hospital

Chronic kidney disease (CKD) is a risk factor for end-stage kidney disease (ESKD) requiring renal replacement therapy (RRT). Hypertensive nephrosclerosis is one of the major causes of ESKD, therefore, control of hypertension is essential to prevent the progression of CKD. This study aimed to describe the prescribing patterns of antihypertensive agents in patients with CKD and to investigate the association between the type of antihypertensive agents and CKD progression, which was defined as annual changes in eGFR, also known as eGFR slopes, estimated by ordinary least squares, using real-world data stored in each core clinical research hospital. Eligible participants were those who were measured eGFR at least once between April 1st, 2018 and September 30th, 2018, and whose initial eGFR was between 15 and 60 ml/min/1.73m². In five facilities, 22,821 participants were enrolled. The mean age was 68 years and the mean baseline eGFR was 48.8 ml/min/1.73m². Angiotensin receptor blockers (ARB) and calcium channel blockers (CCB) were the most common agents, followed by beta-blockers. The prescription patterns of beta-blockers varied across facilities. The mean eGFR slope was -0.56 ± 3.95 overall. Further investigation of comorbidities and background factors that may be confounding by indications for antihypertensive medication is needed to clarify the association between type of antihypertensive medication and progression of CKD.

Keywords: Chronic Kidney Disease, Clinical Epidemiology, eGFR Slope

1. 緒言

慢性腎臓病は血液透析などの腎代替療法を要する末期腎不全の高リスク集団であり、その発症および進展の予防が重要である。特に人口の高齢化を反映して腎硬化症による末期腎不全が増加しており¹、血圧管理は慢性腎臓病の進展予防において重要な役割を果たす^{2,3}。降圧薬の種類は、患者の背景疾患や病状に合わせて使用される⁴。レニン・アンジオテンシン系阻害薬は慢性腎臓病に伴う高血圧に対する高い降圧効果だけでなく、特にアルブミン尿を有する場合に腎臓病進展や死亡抑制などに関する質の高いエビデンスを有する^{5,6}。その他 Ca 拮抗薬やサイアザイド系利尿薬も推奨されるが、降圧薬の違いによる腎臓病がどのように進展するかは十分に明らかにはなっていない。

近年、リアルワールドデータ (Real world data, RWD) を用いた大規模な降圧薬の比較研究が行われた⁷。これは降圧薬新規処方患者に対する、各種降圧薬による心筋梗塞、心不全入院、脳卒中の予防効果を検証したものであった。同様の内容ではあるが約 20 年前に従来型の介入研究として実施された ALLHAT trial の追試としての位置づけであったが、短期間かつ低コストで実現した研究であった。我が国でも、RWD を用いた研究基盤の構築を推進するため、日本医療研究開発機構が支援する「医療技術実用化総合促進事業」の「リアルワールドエビデンス創出のための取組み(以下、臨中ネット)」が 2018 年に開始され、参画する臨床研究中核病院は、病院情報を研究に利活用するための基盤整備や体制構築を進めている。令和 2 年度より具体的な研究課題

である「降圧薬の種類による腎保護効果」(以下、本研究テーマ)を立案し、参加機関を募って実施することになった。令和 3 年 8 月現在で、参加予定機関のうち 5 機関からデータ収集を開始している。

今回収集したデータを用いて、臨床研究中核病院に通院中の保存期慢性腎臓病における降圧薬処方実態を記述し、データが得られている期間で検討可能な腎アウトカムである推算糸球体濾過量 (estimated glomerular filtration rate, eGFR) の年間低下率 (eGFR slope) について検討した^{9,10,11}。

2. 目的

本研究テーマにおけるデータ収集からデータ解析に至る一連のプロセスを概説し、令和 3 年 8 月時点でデータ収集を開始した 5 施設における中間結果として、降圧薬の処方パターンと eGFR slope の記述データについて報告する。

3. 方法

3.1 対象患者

2018 年 4 月 1 日から 2018 年 9 月 30 日の間 (以下、ベースライン期) に少なくとも 1 度 eGFR が測定され、その後 2 年間 (730±90 日) で 3 回以上測定された 18 歳以上の男女を対象とした。ベースライン期の最初の eGFR 測定日を index 日とし、15 以上 60 ml/min/1.73m² 以下であり、Index 日から 730±90 日後に最後の eGFR の測定が行われている者を対象とした。電子カルテや透析部門システムから抽出した情報に基づいて腎代替療法の実施症例を除外した。

3.3 降圧薬処方

Index 日の前後 90 日以内に 1 回以上処方された薬剤をベースライン処方と定義した。降圧薬の分類は、ATC 分類 (C02, C03, C05, C07, C08, C09) および「今日の治療薬」を参照し、最終的に臨床的観点から整理した。降圧薬の分類は、アンジオテンシン変換酵素阻害薬 (angiotensin converting enzyme inhibitor, ACEi)、アンジオテンシン受容体拮抗薬 (angiotensin receptor blocker, ARB)、カルシウム拮抗薬 (Calcium channel blocker, CCB)、 α 遮断薬、 β 遮断薬、利尿薬 (サイアザイド系利尿薬、ループ利尿薬)、ミネラルコルチコイド受容体拮抗薬 (mineral corticoid receptor antagonist, MRA)、その他の降圧薬とした。

3.4 eGFR slope の算出方法

Index 日から最終 eGFR 測定日までのすべての時点での eGFR を用いて、最小二乗法を用いて年間 eGFR 変化率を算出した。同日に複数の測定を行っている場合には最小値を用いた。年間変化率として適切な値として、 ± 20 ml/min/1.73m² per year の範囲に収まらないものについては除外した。次に、年齢、性別で調整した降圧薬種別ごとの予測 eGFR slope を算出して 95%信頼区間とともにフォレストプロットで示した。

3.5 共変量

年齢、性別、ベースライン時の eGFR を抽出した。その他の検査データ、病名、処方については抄録作成時点では収集が完了していないため、併存疾患、検査所見等に関する情報は含めていない。

4. 結果

4.1 患者背景記述

ベースライン eGFR が 15 から 60 ml/min/1.73m² であった対象者 23,589 名のうち、腎代替療法既往患者 768 名を除外し、5 施設合計 22,821 名が選択基準を満たした。患者背景を表 1 にまとめる。平均年齢は 69 歳であり、施設間でほぼ同等であった。男女比はほぼ 1:1 で、平均 eGFR は 48.8 ± 10 ml/min/1.73m² であった。また、降圧薬処方を受けた患者は 43% であったが、施設 #5 だけ 52% と多かった。降圧薬の種類は降圧薬処方有りの患者群において中央値で 2 種類であった。期間中の eGFR 測定回数は 2 年で中央値 12 回 (interquartile range, 7-21) であった。

表 1. 患者背景

	Total	#1	#2	#3	#4	#5
症例数	22,821	3,418	5,912	5,807	4,195	3,489
年齢	69 (12)	68 (12)	67 (12)	70 (11)	69 (11)	68 (12)
性別	12,115 (53)	1,633 (48)	3,043 (51)	3,392 (58)	2,200 (52)	1,847 (53)
降圧薬 処方者	9,879 (43)	1,521 (44)	2,605 (44)	2,267 (39)	1,662 (40)	1,824 (52)
降圧薬数	2 [1-3]	2 [1-3]	2 [1-3]	2 [1-3]	2 [1-3]	2 [2-3]
eGFR	48.8 (10.0)	48.0 (10.2)	50.0 (9.8)	48.1 (9.9)	49.5 (9.8)	47.9 (10.3)
期間中 測定回数	12 [8-21]	12 [9-21]	13 [8-23]	11 [6-19]	12 [7-20]	14 [8-23]
eGFR slope	-0.56 (3.95)	-1.66 (3.43)	0.39 (4.27)	-0.48 (3.65)	-0.68 (3.72)	-1.09 (4.20)

4.2 降圧薬の処方パターン

表 2 に対象集団全体の処方パターンを示す。CCB が 61%、ARB が 59% の患者に対して処方されており、頻度の高い薬剤であった。次の処方の多い β 遮断薬であるが、施設ごとに 20~40% とばらつきを認めた (図 1)。特に #5 においては ACE 阻害薬が ARB と比較して相対的に処方割合が多く、 β 遮断薬の処方割合も高い。ARB 及び CCB の処方割合は施設間で大きな違いはなく、一貫して高かった。

表 2. 降圧薬の処方割合

降圧薬またはその組み合わせの種類	N (%)
各降圧薬の処方割合	
ACE 阻害薬	1,187 (12)
ARB	5,821 (59)
α 遮断薬	520 (5)
β 遮断薬	3,238 (33)
Ca 拮抗薬	5,990 (61)
サイアザイド利尿薬	927 (9)
ループ利尿薬	2,118 (21)
直接レニン阻害薬	39 (0)
ミネラルコルチコイド受容体拮抗薬	1,565 (16)
組み合わせ	
単剤治療	
RAS 阻害薬	1,442 (15)
Ca 拮抗薬	1,210 (12)
その他の単剤治療	707 (7)
2 剤治療	
RAS 阻害薬 + カルシウム拮抗薬	2,063 (21)
RAS 阻害薬 + 利尿薬	168 (2)
Ca 拮抗薬 + 利尿薬	118 (1)
RAS 阻害薬 + β 遮断薬	354 (4)
Ca 拮抗薬 + β 遮断薬	244 (2)
その他の 2 剤治療	468 (5)
3 剤治療	
RAS 阻害薬 + カルシウム拮抗薬 + 利尿薬	504 (5)
RAS 阻害薬 + カルシウム拮抗薬 + β 遮断薬	575 (6)
その他の 3 剤治療	844 (9)
4 剤治療	
RAS 阻害薬 + カルシウム拮抗薬 + 利尿薬 + β 遮断薬	486 (5)
その他の組み合わせによる 4 剤治療	696 (7)

ACE, angiotensin converting enzyme; ARB, angiotensin receptor blocker; Ca, calcium; RAS, renin angiotensin system

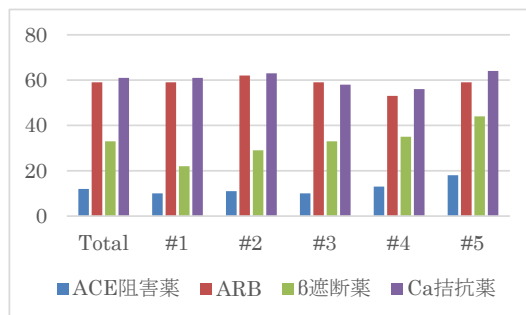


図 1 機関ごとの降圧薬処方パターン

4.3 ベースライン eGFR と eGFR slope

ベースライン eGFR と eGFR slope の分布はそれぞれ図 2、図 3 に示す通りである。#1 と #5 は全体として低下率が大きく、#2 は緩やかである。

4.4 降圧薬ごとの予測 eGFR slope

年齢、性別で調整し、多変量線形回帰分析によって算出された予測 eGFR slope を図 4 に示す。RAS 阻害薬のほうが Ca 拮抗薬に比して eGFR slope の低下率が高い傾向にあり、さらに RAS 阻害薬と Ca 拮抗薬の併用者においてはさらに低下率が高かった。

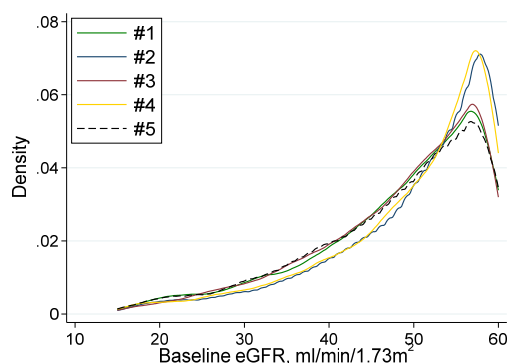


図 2. ベースライン eGFR の分布

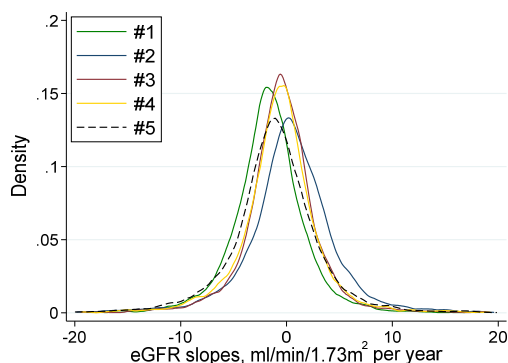


図 3. eGFR slope の分布

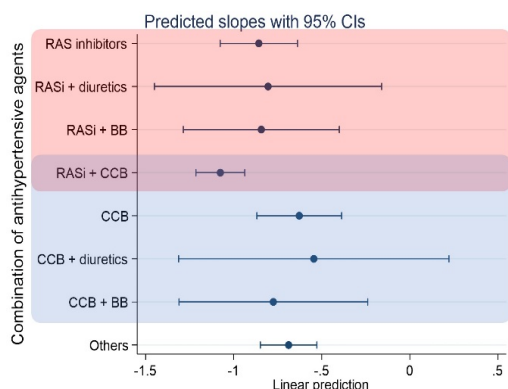


図 4. 年齢・性別で調整した eGFR slope

5. 考察

臨床研究中核病院に通院中の患者における降圧薬の処方パターンおよび eGFR 低下率の関係を、RWD を用いて記述した。RWD を用いた研究として我が国では前例のない試みであり、大規模コホート研究と同等あるいはそれ以上のデータを短期間に、低コストで収集することができた。

本研究では腎アウトカムとして eGFR slope を用いた。これにより短期間のデータ収集でも一定の結果を得ることができた。今後末期腎不全や心血管疾患の発症といったハードアウトカムとの関連を検討するための基礎データを収集することができた点で貴重な報告である。今回の報告では最小二乗法により求めた eGFR slope をアウトカムとしたが、混合効果モデルを用いてよりばらつきを抑え、精緻な解析を追加する。

本研究における CKD 患者に対する降圧薬の処方パターンは CCB と ARB がメインであった。これは過去の慢性腎臓病コホートの結果とも合致する¹²。一方、β 遮断薬の処方パターンについては同コホートにおいて 20% 程度だったが、今回の検討においては 20~45% と施設間で大きなばらつきが生じていた。β 遮断薬の処方割合の多い機関では ACE 阻害薬の処方割合も相対的に高いことから、心疾患の患者層が多いことが推測される。今後処方、病名、その他の検査所見を追加した解析を行って心血管疾患や悪性腫瘍などの併存疾患による交絡の調整を行って処方パターンのばらつきの背景を明らかにしていく予定である。これにより多変量モデルをベースにした予測 eGFR slope の予測精度は向上することが見込まれ、より正確な評価を行うことができると期待される。

本研究にはいくつかの限界がある。1 つは既存の診療情報から限られた情報のみで交絡因子を検討せねばならない点である。今回、降圧薬の適応交絡を調整するための情報として、電子カルテのテキスト情報は用いていない。このため病名、処方、検査データから交絡の調整を試みたとしてもさらなる残余交絡が存在する可能性がある。次に、透析や腎移植の既往のある患者の抽出において誤分類が生じる点である。今回情報収集を行った医療機関だけではなく、別の医療機関で血液透析を受けていたり、腎移植を受けていたりする可能性がある。そして共同研究機関の間で腎代替療法に関する情報収集体制が標準化されていないことが改めて明らかとなった。しかし、今回の降圧薬と腎保護に関して、腎代替療法に関する誤分類は差別的誤分類であると考えられる強い根拠はない。最後に、患者はいくつかの機関を受診している可能性がある。実際、処方割合は前向きコホート研究において 90% であったのに対し、本研究では 50% 程度であった。これは別の医療機関での処方による誤分類が生じている可能性を示唆している。Index 期間中に入院した患者の場合は持参薬報告の情報を収集できるが、入院歴のない患者ではこのような情報は得られない。これは患者の重症度などの病状により同一医療機関での処方となるか、他機関での処方となるかが変化しうるため、差別的誤分類である可能性は高い。この情報バイアスに対処するため、今後は薬剤情報の収集期間を様々に変化させて感度分析を行う予定である。

6. 結論

臨床研究中核病院に通院中の慢性腎臓病患者における降圧薬の処方実態と腎臓病進展の関係について記述・分析を行った。残余交絡の問題や標準化されていない方法論上の問題点を解決して研究を完遂させることが今後の課題である。

謝辞

データを提供いただいた本プロジェクト「降圧薬の種類による腎保護効果」の参加機関のメンバーの皆様へ厚く御礼申し上げます。

参考文献

- 1) 新田孝作、他. 日本透析医学会統計調査委員会. 我が国の慢性透析療法の現況(2019年12月31日現在). 透析会誌. 2020; 53: 579-632.
- 2) Yamagata K, Ishida K, Sairenchi T. et al. Risk factors for chronic kidney disease in a community-based population: a 10-year follow-up study. *Kidney Int.* 2007; 71: 159-166
- 3) Hirayama A, Konta T, Kamei K, et al. Blood pressure, proteinuria, and renal function decline: associations in a large community-based population. *Am J Hypertens.* 2015; 28: 1150-1156.
- 4) Brenner BM, Cooper ME, de Zeeuw D, et al. RENAAL Study Investigators. Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. *N Engl J Med* 2001;345:861-869.
- 5) Lewis EJ, Hunsicker LG, Clarke WR, et al. Collaborative Study Group. Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. *N Engl J Med* 2001;345:851-860.
- 6) 日本高血圧学会高血圧治療ガイドライン作成委員会. 高血圧治療ガイドライン 2019 (JSH2019).
https://www.jpnsh.jp/data/jsh2019/JSH2019_hp.pdf
- 7) Suchard M, Schuemie M, Krumholz H. et al. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *Lancet.* 2019; 394: 1816-1826.
- 8) The ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). *JAMA* 2002; 288: 2981-97.
- 9) Grams M, Sang Y, Ballew S, et al. Evaluating glomerular filtration rate slope as a surrogate end point for ESKD in clinical trials: an individual participant meta-analysis of observational data. *J Am Soc Nephrol.* 2019; 30: 1746-1755.
- 10) Inker L, Heerspink H, Tighiouart H, et al. GFR Slope as a Surrogate End Point for Kidney Disease Progression in Clinical Trials: A Meta-Analysis of Treatment Effects of Randomized Controlled Trials. *J Am Soc Nephrol.* 2019; 30: 1735-1745.
- 11) 濱野高行. 早期 CKD のエンドポイント. 日腎会誌 2018; 60: 577-580.
- 12) Imaizumi T, Hamano T, Fujii N, et al. Cardiovascular disease history and β -blocker prescription patterns among Japanese and American patients with CKD: a cross-sectional study of the CRIC and CKD-JAC studies. *Hypertens Res.* 2021; 44: 700-710.

国立高度専門医療研究センター(NC)における多施設共同リアルワールド臨床データベース(JASPEHR)のデータ利活用環境の構築について

*星本 弘之^{*1}、美代 賢吾^{*1}、石井 雅通^{*1}、波多野 賢二^{*2}、平松 治彦^{*3}、
渡辺 浩^{*4}、三原 直樹^{*5}、野口 貴史^{*6}、北村 真吾^{*7}、櫻井 理紗^{*3}

*1. 国立国際医療研究センター医療情報基盤センター、*2. 国立精神・神経医療研究センター医療情報室、
*3. 国立循環器病研究センター情報統括部、*4. 国立長寿医療研究センター医療情報室、*5. 国立がん研究センター中央病院医療情報部、*6. 国立成育医療研究センター情報管理部、*7. 国立精神・神経医療研究センター睡眠・覚醒障害研究部

Development and evaluation for National center joint project of multi-facility real-world clinical database and its analytic environment

Hiroyuki Hoshimoto^{*1}, Kengo Miyo^{*1}, Masamichi Ishii^{*1}, Kenji Hatano^{*2}, Haruhiko Hiramatsu^{*3}, Hiroshi Watanabe^{*4},
Naoki Mihara^{*5}, Takashi Noguchi^{*6}, Shingo Kitamura^{*7}, Risa Sakurai^{*3}

*1 National Center for Global Health and Medicine, *2 Hospital Computer Center of National Center of Neurology and Psychiatry, *3 National Cerebral and Cardiovascular Center, *4 National Center for Geriatrics and Gerontology, *5 National Cancer Center, *5 National Center for Child Health and Development, *7 National Center of Neurology and Psychiatry

Abstract:

Six National Centers for Advanced Medical Sciences (NCs), including the National Center for Global Health and Medicine and the National Center of Neurology and Psychiatry, are promoting a project to build a multi-center clinical database to promote clinical research using real-world data (RWD). In addition to basic patient information, consultation history, tests, prescriptions, etc. obtained from SS-MIX2, this database will support data collection using templates defined by the HL7-FHIR-Questionnaire integrated into the electronic medical record. The input results of the template will be saved in FHIR-compliant JSON format. The collected data will be available for use at the data center (DC), where data cleaning processes such as verification of missing and abnormal values, mapping of local codes to standard codes, etc. will be performed. An on-site data analysis center (OAS) has been set up in the DC, where catalog data can be referred to and analyzed freely using analysis tools, and a remote access environment has been established to accommodate use from each NC. For the time being, the use of the data is limited to the staff belonging to the six NCs, and the JASPEHR data will be linked to the biobank data jointly operated by the six NCs in the future, and when completed, it will be a database of hundreds of thousands of records linking clinical information and genomic data derived from biobank samples. When completed, the database will contain hundreds of thousands of records linking clinical information and genomic data derived from biobank samples.

Keywords: Real World Data, HL7-FHIR, Patient Registry, Clinical Database.

1. 緒論

治験や臨床研究などにおける医療情報の利活用においては、従来、研究プロジェクトごとに電子カルテから人的資源を投入した上で手作業で必要項目を転記する、あるいは、専用のテンプレートを電子カルテシステムに組み込み、医師が診療中にテンプレートに必要事項を入力したものを収集する方法で研究用のデータベースを構築する方法がとられてきた。これらの方法では、せっかく電子化されている情報をオフラインかつ手作業で転記するという、非効率な方法であることに加え、事前に収集を予定していなかった項目が必要になった場合に、別途集めるのにも相応のコストがかかるなどの問題が以前より指摘されていた。この状況に対し、電子カルテに実際に格納されているリアルワールドデータ(RWD)が注目されており、次世代医療基盤法の施行などにより法的環

境が整備され、その積極的な活用が期待されている状況にある。これについて、我が国ではレセプト情報と特定健診・保健指導等情報データベース(NDB)が構築され、様々な研究に活用されているが、NDBには検査結果などが含まれていないことから、検査結果など診療行為の結果に関する情報も含むデータベースの構築が求められている。

国立高度専門医療研究センター(以下、NC)では、がん、循環器疾患、精神・神経疾患、難病・感染症、成育医療、老年医療などの専門分野に特化した診療と研究を実施しているが、これらの6施設の診療情報を統合したデータベースの構築は、様々な領域の疾患をカバーしており、臨床研究や創薬研究への利活用において非常に有益であると考えられる。われわれはこのような背景のもと、「電子カルテ情報を活用したリアルワールドデータ収集・提供基盤の構築事業(Japan Standard Platform for Electronic Health Records: JASPEHR)」

を国立高度専門医療研究センター医療研究連携推進本部 (Japan Health: JH)¹⁾ の補助を受けて実施しており、データベースの構築と利活用環境の整備を推進している、その概要を報告する。

2. データベースおよび分析システムの概要

2.1 データ収集

JASPEHR では、参加する 6NC8 病院について、現時点では以下の情報を収集している

- 1) 患者基本情報：性別、生年月
- 2) 病名
- 3) 検体検査結果
- 4) 処方・注射歴
- 5) 入退院歴
- 6) 外来受診歴

これらの情報は、各病院の SS-MIX2 標準化ストレージから取得し、患者氏名の除去や患者 ID の匿名化を実施した上で、暗号化したアーカイブファイルにまとめ、IPSec-VPN 経由で定期的に収集している。収集したデータは国立国際医療研究センター (NCGM) に設置したデータセンターにて、以下のデータ検証・クリーニングなどを経て、分析用データベースを構築して管理している。

データ収集対象は 2015 年 4 月以降に 6NC のいずれかの病院を受診した患者となっており、参加・非参加についてはオプトアウトとし、研究の概要を院内の掲示およびホームページに公示し、非参加の申し出があった場合には収集対象から除外している。

データの収集はケーアイエス株式会社の MCDRS という EDC システムを用いており、SS-MIX2 標準化ストレージから上記の 6 種類のデータを抽出し、施設ごとに匿名化した上で IP-Sec VPN 経由データセンターに送信している。

データ収集に当たり、データ中で使用するコード類については、診療科コードは SS-MIX2 に定められた診療科コード、検体検査項目は JIAC-10 コード、医薬品については HOT コードに加え、JASPEHR プロジェクトから要望して SS-MIX2 1.2g から追加された YJ コードを用いることとしている²⁾。その他についても SS-MIX2 仕様で定められたコードを採用している。

2.2 HL7-FHIR Questionnaire によるテンプレート実装

JASPEHR では、先行する J-Dreams³⁾ などの臨床データベース構築事業を参考に、SS-MIX2 による定型データの収集に加え、プロジェクト固有の詳細な情報収集のため、HL7-FHIR 仕様の Questionnaire によりテンプレートをベンダー非依存に定義し、その定義情報によって各ベンダーの電子カルテシステムに統合されたテンプレートを生成し、その入力テンプレートをを用いることにより、SS-MIX2 標準化ストレージには格納されていない情報の収集にも対応する予定である⁴⁾。この Questionnaire により生成されたテンプレートに入力された情報は、HL7-FHIR の Questionnaire-Response に準拠した JSON 形式のファイルとして SS-MIX2 拡張ストレージに保存され、データ収集の際にはファイルの状態で収集され、分析に使用される。

2.3 データ検証・クリーニング

収集したデータについては、復号化し、一次データベースに格納した上で検証とクリーニングを行っている。

検証内容については、データ項目の不備（欠落など）の確認、標準コードとローカルコードの確認、データの分布（年齢構成など）の確認など、予備的なデータ分析の結果見られた問題点について実施している。

データのクリーニングについては、上記項目について、スクリプト処理によりコード変換や不適切レコードの除外などの処理を実施しているほか、各 NC の SS-MIX2 出力仕様を検証し、必要に応じて修正を依頼するなどして、データの品質向上を図っている。

2.4 分析システム

データの分析環境は、病名検索や処方検索、検体検査結果検索などに対応し、利用者・プロジェクトごとのアクセス制限に対応した DWH を構築しているほか、高度な統計解析の実施のため、R などの統計処理ソフトを端末にインストールしている。これらの端末はすべてインターネットからは遮断された閉鎖系ネットワークとなっているほか、USB メモリなどの利用も制限し、データの持出を防止する設定となっている。

2.5 クリーンルーム

データ分析システムの端末は、許可された利用者のみ入室可能で、かつ、端末の操作状況をモニタリングする監視カメラが設置された居室内に設置されており、さらに端末の操作ログとデータベースのアクセスログが一定期間保存される環境となっている。この環境は NDB のオンサイトリサーチセンターのデータ利用環境などを参考に構築している⁵⁾⁶⁾。

2.6 リモートアクセスシステム

JASPEHR のデータベース及びデータ分析システムは、前述の通り NCGM に設置されており、他の NC の研究者にとっては、解析実施のために NCGM のクリーンルームを利用する必要があり、データの利用に不便が伴うことから、各 NC からデータ分析環境を利用可能なリモートアクセスシステムの整備を行った。

リモートアクセスシステムでは、接続元については各 NC の代表 IP アドレスからの接続のみアクセスを許可し、さらに、リモート端末についても端末認証用の証明書がインストールされた端末からのアクセスのみ接続を許可する設定となっている。その上で利用者ごとに発行した ID/パスワードにより認証を行っている。リモート接続によりデータ分析環境に接続すると、クリーンルームと同等の環境によるデータ分析の実施が可能であるが、データのダウンロードによる持出は許可しない設定となっている。

2.7 データ利活用手続き

現時点では JASPEHR のデータ利用は 6NC に所属する研究者のみに限定されている。

JASPEHR のデータについては、現状では分譲は行わず、データセンター内の分析システムで分析した結果を提供する形で運用している。

3. データベースおよび分析システムの概要

3.1 データ収集

JASPEHR データベースは 2021 年 4 月よりデータ収集を開始しており、概ね 2~3 ヶ月の間隔でデータ収集を行っている。2021 年 8 月 30 日の段階では、NCGM に加え、国立精神・神経医療研究センター (NCNP)、国立長寿医療研究センター (NCGG) の 3NC からのデータが登録されている。データの登録状況の内訳は表 1 に示す通りであるが、3NC で約 50 万 ID の患者に関する情報が集積されている。残る 3NC のうち、国立循環器病研究センター (NCVC) については現在データ登録準備中となっている。さらに、国立成育医療研究センター (NCCHD) については、VPN 接続作業中となっており、接続され次第データの登録を行う予定である。

表 1 JASPEHR データの登録状況

データ項目	NCGM	NCNP	NCGG
患者基本	318,971	44,717	133,961
病名	2,674,795	1,516,100	2,027,513
検体検査	70,149,643	6,586,044	2,027,513
処方	7,623,741	4,157,202	1,968,971
注射	5,704,692	392,869	1,098,018
入院歴	106,129	33,255	25,525
退院歴	106,307	33,288	25,482
外来受診	2,180,703	802,999	22,026

3.2 データクリーニング

収集したデータについて、内容の検証とのクリーニングを実施しているが、以下のような問題が見つかっている。

- 1) 特定の生年月月に不自然に患者 ID (匿名化 ID) が集中している。
- 2) 診療科コードが途中で変わっている。
- 3) 病名について、診断日が格納される PRB-7 に値が無く、プロブレム発生日が格納される PRB-16 に診断日と思われる日付データが格納されている。
- 4) JLAC-10 のマッピングが不適切で、ローカルコードが格納されている。
- 5) 定性値の表現形式が統一されていない。

これらについては、原因を調査の上で逐次対応しているが、その原因と対処の概要は以下の通りである。

3.2.1 特定の生年月月にピークがある

これは、ゲノム検査結果を管理するために設定したダミー患者 ID がそのまま出力されていたことが原因であり、データ抽出時に除外設定を行うことにより対処した。

3.2.2 診療科コードが途中で変化する

これは、2 つの原因が複合して生じた不具合である。一つは、診療科再編による臓器別診療科導入により、診療科が細分化されたことが原因であったことから、再編前後の診療科の対応表を作成し、分析時に必要に応じて変換することとした。もう一つは、データ収集期間の途中に発生した電子カルテシステム更新の際に診療科コードの変換設定にミスがあり、更新後のシステムから出力されるデータについては診療科コードがローカルコードとなっていた。これについては、設定を修正し、SS-MIX2 準拠の診療科コードが出力されるように修

正している。

3.2.3 診断日が格納されていない

これは、あるベンダーの電子カルテで、病名データを SS-MIX2 に出力する際に、医師が判断した病名の開始日 (電子カルテ上で病名登録時に任意に変更可能) をプロブレム開始日に設定しているが、医師が病名をシステムに登録した日 (≒ 診断日) については出力していない仕様とのことであった。これについてはデータクリーニング時には対応できないため、当該ベンダーと SS-MIX2 の出力仕様についての協議を行っている。

3.2.4 項目コードがローカルコードとなっている

これは、各施設で整備する変換テーブルの設定ミスであることから、各施設には都度検証結果をフィードバックし、変換テーブルの修正を依頼している。

3.2.5 定性値の表現形式が不統一

これは、各施設の定性値のテーブルを入手し、クリーニング時に変換することで対応予定である。

4. 考察

JASPEHR データベースの構築においては、データの収集は概ね順調であると評価している。また、データ利用環境の整備についても、COVID-19 による遅延はあるものの、おおむね初期に予定した環境の整備が完了している。

一方、研究での利用において重要となるデータの品質については、項目コードや定性値表現形式の不一致については事前に予想されていたが、診療科再編に伴うコード変更によりデータの連続性に影響が生じたり、ゲノム情報管理のためのダミー患者 ID の存在による生年月分布に不自然なピークが生じた件など、当初予期していない事項によるデータ品質の低下とそれらの問題への対処に関する知見は、今後 JASPEHR データベースを拡張する際や、他の臨床データベース構築においても有用であると考えられる。なお、診療科コードの変更や、検査方法や検査試薬の変更に伴う検査コードの変更などへの対処として、各施設のローカルコードの変更の履歴の収集は変換テーブルの作成などにおいて非常に重要であると考えられることから、今後はデータの収集に加えて、各施設のローカルコードテーブルについても収集を行う予定である。

HL7-FHIR-Questionnaire によるベンダーに依存しないテンプレートの定義とそれによるテンプレート入力画面の自動生成については、一部プロトタイプは稼働している者の、電子カルテと統合された環境については、本稿執筆時点では開発中である。このシステムが完成すると、J-Dreams など先行する臨床データベース構築事業において、多大なコストと工数をかけて展開していたデータ入力テンプレートについて、開発コストと配布コストの大幅な低下が実現されることから、SS-MIX2 標準化ストレージからは取得できない詳細な情報を容易に収集可能となり、将来的な臨床データベース構築に大きく貢献すると考えられる。

なお、JASPEHR データベースは将来的には、6NC のパイオバンク検体のゲノム情報と組み合わせることが検討されており、これが実現すると、難病や認知症など様々な疾患に関する研究や創薬研究における活用が期待される。

5. まとめ

JASPEHR データベースの概要とデータ利用環境について概説した。JASPEHR データベース構築における知見および、HL7-FHIR-Questionnaire によるベンダー非依存のテンプレート定義は、今後の臨床データベース構築において有用であると考えられる。

6. 参考文献

- 1) 国立高度専門医療研究センター医療研究連携推進本部 (JH) [https://www.japanhealth.jp/]. (2021/8/30 日閲覧)
- 2) SS-MIX2 仕様書・ガイドライン[http://www.jami.jp/jamistd/ssmix2.php] (2021/8/30 日閲覧)
- 3) 美代 賢吾.【電子カルテからの Real World Data の収集】J-DREAMS(診療録直結型全国糖尿病データベース事業).Precision Medicine. 2020 3 巻 9 号:804-807
- 4) 石井雅通、美代 賢吾. FHIR Questionnaire を用いたテンプレート情報の標準化の試みと試験実装. 第 40 回医療情報学連合大会 医療情報学. 2020 40(Suppl.):585-586.
- 5) レセプト情報・特定健診等情報の提供に関するホームページ [https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/iryohoken/resepto/index.html]. (2021/8/30 日閲覧)
- 6) オンサイトリサーチセンターにおけるレセプト情報・特定健診等情報の利用に関するガイドラインについて [https://www.mhlw.go.jp/stf/shingi2/0000135204_00001.html]. (2021/8/30 日閲覧)

がん診療統合臨床データベースの活用

－ レジメン詳細情報の院内共有とレジメン投与歴の後方視の確認 －

鈴木 一洋^{*1}、中野 泰寛^{*2}、箕輪 雄一^{*2}、横川 貴志^{*2}、根本 真記^{*2}、
望月 俊明^{*3}、笠原 あや菜^{*1}、高橋 輝^{*1}、清水 久範^{*2}、小口 正彦^{*1}

*1 公益財団法人がん研究会 有明病院 データベース&バイオバンクセンター、

*2 公益財団法人がん研究会 有明病院 薬剤部、

*3 公益財団法人がん研究会 有明病院 集中治療部

Utilization of Unified Integrated Cancer Clinical Database System

- sharing detailed chemotherapy regimen information and retrospective examination of dose histories -

Kazuhiro Suzuki ^{*1}, Yasuhiro Nakano ^{*2}, Yuichi Minowa ^{*2}, Takashi Yokokawa ^{*2}, Maki Nemoto ^{*2},
Toshiaki Mochizuki ^{*3}, Ayana Kasahara ^{*1}, Hikaru Takahashi ^{*1}, Hisanori Shimizu ^{*2}, Masahiko Oguchi ^{*1},

*1 Database and Biobank Center, Cancer Institute Hospital of JFCR,

*2 Pharmaceutical Department, Cancer Institute Hospital of JFCR,

*3 Department of Emergency medicine and intensive care medicine, Cancer Institute Hospital of JFCR

Since 2017, we have been developing an Integrated Cancer Clinical Database System (Cancer DB) which is now used in 7 surgical oncology sections. To extend to clinical oncology areas, we first developed a Chemotherapy Regimen Management System, which can be accessed throughout the hospital. The Regimen Management System is consisted of 27 features, which includes regimen number, cancer type, administration department type, clinical administration type, treatment protocol type, intra-course withdrawal period, drug reduction rules, drug withhold criteria, administration sequence, patient eligibility criteria, exclusion criteria, protocol explanation, main adverse effects and countermeasures. EHR cannot manage detail data like these, so regimen management data was printed out by the individual who felt the information was needed. When printed out, there is no way to check, who is referring to old data and who is not, but utilizing the Chemotherapy Regimen Management System, everyone accesses to the newest data.

Connecting the regimen management data with blood tests and pharmacist questionnaire results, we can retrospectively examine, not only liver or kidney function deterioration, but also adverse effect worsening. Furthermore, we are developing an ICI (Immune Checkpoint Inhibitor) Regimen Management function which can monitor steroid administration and ICU hospitalization.

We are integrating these features to the Cancer DB, and continuously discussing with medical staff to upgrade and improve the system.

Keywords: integrated cancer clinical database system, chemotherapy regimen, adverse event management

1. 諸論

当院は 2017 年から「がん診療に関する臨床情報を統合的に保存・管理するデータベース」(以降「統合 DB」)¹⁾の開発に着手し、外科系を中心に計 7 診療科で運用中である。内科系データベースを開発するにあたり、最初の一步として外来・病棟を問わず薬物療法のレジメン詳細マスタ情報を共有できるレジメン管理機能の開発を行った。

2. 開発目的

薬剤部では、2005 年頃からスタンドアローンのノート PC(1 台)内に Microsoft Access (以降「MSAccess」)で構築されたレジメン管理機能を利用してきた。当該端末はスタンドアローンであるため、電子カルテのネットワークには所属しておらず、日常業務のなかでレジメン管理機能の内容を参照するには、当該端末が設置されている場所(地下 1 階薬剤部エリア)に移動するか、あるいは、必要な情報を事前に印刷しておく必要があった。

外来業務中の医師や看護師が地下 1 階薬剤部エリアに移

動して MSAccess を直接確認することは現実的ではなく、院内 PHS で薬剤部へ問い合わせを行うことは薬剤部の負担となる。また印刷出力された情報も最新の内容ではない可能性があるため、レジメン管理機能を院内診療系の情報システムの一部として実装し、最新のレジメン詳細マスタ情報を電子カルテ端末からいつでも誰でも参照できることは、がんを専門とする医療機関において大変重要な事項である。

当院の電子カルテに搭載されているレジメンマスタ情報は、パスの機能がベースになっていることもあり、「休業規定・減量規定」や「主な副作用と対策」など、レジメンの実施確定あるいは投与する際に参照・確認する詳細情報を持つことができない。そこで、統合 DB にレジメン管理機能を新規に作成し、外来だけでなく病棟も含めた院内どこからでも参照可能なように、電子カルテから起動できる仕組みを構築した。

3. システム概要

レジメン管理機能では 27 のデータ項目があり、レジメン登

録番号やがん種をはじめ、実施部署区分(入院、外来点滴、外来処方等)、臨床使用分類(日常診療、単施設自主研究、多施設自主研究、治験等)、抗がん剤適応分類(進行再発癌、術前補助化学療法、術後補助化学療法、大量化学療法等)や、コース間での休薬規定、減量規定および中止基準、1日の投与スケジュール、患者条件、除外規定、プロトコル解説、主な副作用と対策等の項目を1画面で管理している。画面上部にはレジメン基本情報(レジメン登録番号、癌腫、実施部署区分、臨床使用分類、抗癌剤適応分類、薬剤等)を、画面下部にはレジメン詳細情報(総コース数、コース間での休薬規定、減量規定及び中止基準、1日の投与スケジュール等)を配置した。レジメン基本情報の画面例を図1に示す。

図1 レジメン基本情報

レジメン詳細情報の画面例を図2-4に示す。

図2 レジメン詳細情報(1)

図3 レジメン詳細情報(2)

図4 レジメン詳細情報(3)

レジメン管理担当薬剤師は、院内委員会で承認されたレジメン情報をレジメン管理画面に記載し、いったん仮保存する。仮保存の状態では、薬剤部レジメン管理担当ならびに薬剤部レジメン承認担当のみが、その内容を参照することができる。レジメン承認担当薬剤師が仮保存された内容をダブルチェックし、記載内容に問題がなければ[公開]ボタンを押下することにより、当該レジメン管理情報が院内全体に公開される。2021.08.19時点の公開レジメン数は1,396件である。

4. システム評価

2020.09 末に、これまで登録を行っていた MSAccess のデータを統合 DB のレジメン管理機能に移行し、2020.10 より院内全体での運用を開始した。2021.04.01 から 2021.07.31 の4ヶ月間におけるレジメン管理機能のアクセス状況を表1に示す。

す。

表1 レジメン管理機能のアクセス状況集計

年/月	アクセスなしの日数[日]	総アクセス回数[回]	薬剤師以外のアクセス回数[回]
2021/04	0	1279	117
2021/05	1	1023	112
2021/06	0	1063	102
2021/07	2	790	37

概ね1000回以上/1ヶ月のアクセスがあり、うち10%前後が薬剤師以外のアクセスであった。

レジメン管理機能について、レジメン承認担当薬剤師からのヒアリングにより、下記について良好なフィードバックを得た。

- 1) データのアクセスログが確認できること
- 2) 「登録」だけでなく、「使用中止」の管理ができること
- 3) 薬剤名称でレジメン検索ができること
- 4) 適応外使用のフラグがあること
- 5) レジメンに紐づいた情報を別ファイル(Excel や PDF)で格納できること

特に「使用中止」フラグの機能を備えていることで、途中で使用中止になったものが管理できること、また薬物療法レジメンの複雑化に伴い、すべての情報を画面項目に設けることが難しい状況において、レジメンに紐づいた情報を Excel や PDF ファイルごとアップロードして格納できることは、大変有用性の高い機能であるとのコメントも得ている。

5. 考察

薬物療法における安全管理は、様々な施設での試行錯誤やシステム開発²³⁾を通じて形成されてきた多くのノウハウが含まれていると考えられる。薬物療法の安全な実施において、5R (Right Patient, Right Drug, Right Dose, Right Route, Right Time) が重要である旨ことは言うまでもなく、当院の看護業務マニュアルにも Right Purpose を加えた 6R について記載されている。しかし、臨床試験等では確認することができなかった薬剤相互作用や患者背景に伴う有害反応を完全に防ぐことはできず、がん薬物療法における高リスク患者の投与後モニタリングは、診療における安全性の向上に寄与すると考えられる。

そこで、当院では前記レジメン管理機能とあわせ、レジメン投与実績情報(薬剤・投与量)、投与日の血液検査情報と薬剤師問診情報を連結(以降「レジメン投与実績管理情報」)することで、レジメン投与歴を後方視的に確認し、肝機能・腎機能の悪化、問診による有害事象 Grade の悪化などを網羅的に検索・確認できる機能も開発した。また、本機能の発展型として ICI(免疫チェックポイント阻害薬)レジメンに特化した投与実績管理情報画面も開発し、レジメン投与日以降のステロイド投与情報や ICU 入院履歴等を付与して、有害事象の発生状況などが後方視的に確認できるようになった。

6. 結論

レジメン管理機能を、スタンドアローンのノート PC で動作する MSAccess から、院内全体で共有可能なデータベースへ移行し、電子カルテ端末からアクセスできる環境を構築した。平均1,000回/1ヶ月以上のアクセスがあり、その10%前後は薬剤師以外のアクセスであることから、薬物療法に関する情報共有において一定の役割を果たしていることが示唆された。また、レジメン投与歴データを後方視的に確認する機能については、臨床現場と定期的に相談しながら運用・改善を継続している。

7. 参考文献

- 1) 鈴木一洋, 笠原あや菜, 小口正彦. 集学的がん診療に特化した統合データベースシステムの開発. 医療情報学 2018; 38, 4:227-233.
- 2) 古川裕之. 薬剤情報と病院情報システム. 医療情報学 2001;21,3:195-204.
- 3) 佐藤篤郎, 坂田洋, 鈴木厚志, 加藤直紀, 黒田信行 ら. 病棟薬剤業務支援システム (Pharm-Support Ver.2) の開発と評価. 医療情報学 2014;34,2:75-80
- 4) 宝田 繁基, 福田 泰代, 香川 好美, 高口 恵子, 宮川 真澄 ら. 電子カルテと調剤支援システムを利用した注射抗がん剤の安全管理の取り組み. 医療情報学 2011;30,5:277-284

血液検査値の同時分布の可視化

古川 哲^{*1}、香川 璃奈^{*2}、興梠 貴英^{*3}、城 真範^{*1,3}、赤穂 昭太郎^{*1}

^{*1} 産業技術総合研究所、^{*2} 筑波大学、^{*3} 自治医科大学

A Visualization of Joint Distributions of Blood Test Values

Tetsu Furukawa^{*1}, Rina Kagawa^{*2}, Takahide Kohro^{*3}, Masanori Shiro^{*1,3}, Shotaro Akaho^{*1}

^{*1} National Institute of Advanced Industrial Science and Technology, ^{*2} Tsukuba University, ^{*3} Jichi Medical University

For personalized medicine based on assessment based on the quantitative context of the individual patient from the accumulated data, data based on clear conditions and accurate measurements and the accurate distribution of the population under these conditions are necessary. In order to extract and visualize items that are strongly related to each other among venous blood test items. We used Graphical Lasso, which is a method that can visualize a network with good visibility by adding a linear regularization term to the original data assuming the normal distribution and cutting off the nodes with weak relationships. We estimated the joint distribution of venous blood specimen test values obtained at two medical institutions and used graphical lasso on the variance-covariance matrix. We showed the relationship graph between about 30 to 50 pairs of test items, including the positive/negative correlation and the strength of the correlation. Examination of the results by physicians allowed reasonable interpretation.

Keywords: Blood test values, Joint distribution, Lognormal distribution, Graphical lasso, Pseudo data

1. 緒論

医療データの電子的な蓄積と、個別化医療の需要の高まりによって、蓄積されたデータからの患者個人のコンテキストを定量的に踏まえたアセスメントが期待されている¹⁾。そのためには条件が明確で正確な測定に基づくデータと、その条件下での母集団の正確な分布が必要である。

ここで我々は検体検査値に着目する。臨床現場では、診察の際に検査結果に基づくアセスメントが行われるが、その際には単純な基準範囲(RIs)に基づく異常値判定しかされないのが通常である。しかし実際には検査値がRIsのどの程度内側にあるか、どの程度外側にあるかによって、適切なアセスメントは異なるはずである。ところが検査値の分布は通常非対称であるため、RIsだけから最頻値、期待値、標準偏差などの統計的特徴を得ることが難しく、したがって得られた検査値がどの程度内側にあるか、外側にあるかを判断することも容易ではない。医療分野においては、ある特定のヒトの集団における検体検査値の分布の生データは、ごくわずかな集団における一部の検査値などの例外を除き、公開されていない。将来の疾患の発症の可能性などに基づく閾値や、一般的に健康な個人の95%をカバーする基準範囲²⁾など、少数の統計値のみが公開されると同時に、臨床現場で利用されている。

我々は既に、いくつかの検体検査項目について、実際に入院患者から得られた値の分布をパラメトリックな確率密度関数によって近似することを試み、対数正規分布と指数分布によって多くの検体検査項目が表現できることを明らかにした。パラメトリックな近似による表現は表計算ソフトウェアなどで扱いやすく、数理モデルを使った定量的な検討にも利用できる。この結果は基準血液検査値セットとして公開している³⁾。またその後の研究にて値のゼロ点に自由度をもたせた三つのパラメータを用いた対数正規分布が、静脈血の検体検査値の

分布を表現するにあたって、より適切であることも明らかになっている⁴⁾。

しかしながら、検体検査項目の中には相互に関連が深いものがあり、臨床現場では複数の検体検査項目の結果を同時に鑑別疾患に利用することがある。こうした場合、検査値のペアを同時分布で表現することが適切であるが、これまでそれがどのような分布になるかについては、ほとんど明らかになっていない。

同時分布に固有の問題として、 N 種類の検体検査項目に対してペアの種類は N^2 程度となり、データ数が不足して結果の信頼性が下がるという問題が考えられる。またペアの種類が多くなるため、それを人間が簡便に理解する表現方法にも課題がある。検体検査項目間の多くには医学的な知見からは関係がないだろうものが含まれる一方で、それをデータに基づいて合理的に削除する方法は自明ではない。こうした問題点を考慮し、十分にデータ数の多い検体検査項目間についてのみ同時分布を作成し、データに基づいて合理的に関係性を視覚化する必要がある。

こうしたデータ解析の方法の一つに、Graphical lasso(GLASSO)という方法がある⁵⁾。正規分布を仮定したデータに線形正則化項を加えて、関係性の弱いノード間を切断し、見通しの良いネットワークを可視化できる技術である。静脈血における検体検査項目の多くが対数正規分布によって近似可能であることを使い、対応する正規分布の同時分布をGLASSOにてスパース表現し、検体検査項目間のネットワークを可視化することができれば、患者個人のコンテキストを定量的に踏まえたアセスメントを実施するに当たり、基本的で重要な知見を提供することが想定される。

2. 目的

本研究では、多くの検査項目に対応するため、医療機関で得られた検体検査値を利用する。静脈血の検体検査値の同時分布を作成し、その分散共分散行列に GLASSO を使って検査項目間の関連の強さを視覚化して示す。

3. 方法

3.1 データ

二つの医療機関のデータを利用した。具体的には筑波大学附属病院全科で得られた入院患者 (2013 年から 2018 年、20 歳以上、102,830 名) の血液検査値 (TBA) と、自治医科大学附属病院循環器内科にて得られた入院患者及び外来患者 (2010 年 4 月から 2021 年 3 月、20 歳以上、25,058 名) の血液検査値 (JIC) から、日内変動が知られた検査項目、白血球分画の絶対数による測定値、有核赤血球、JDS 基準の HbA1c、各カットオフ値とコントロール値算出のためのデータ、男女毎に最低 2000 サンプルが得られてない検査項目は統計分布の推定に十分な数のサンプルが得られていないとして除外した。こうして TBA は男性 70 種、女性 77 種、JIC は男性 38 種、女性 29 種類の検査項目を選定した。どちらのデータも男女別にわけ、年齢区分を考慮せず、以下の解析を行った。解析においては倫理的配慮により、データを扱う担当者を限定した。

3.2 検査値の分布の近似

GLASSO を用いたネットワークを構成するためには、分布が正規分布であるという前提を満たす必要がある。このためまず、検査値の分布を正規分布に変換した。具体的には、まず患者によって検査回数が異なることによる影響を排するため、1 患者・1 検査項目についてデータ中の検体検査値を平均して 1 つの検査値として代表させた。構成したヒストグラムを Kernel density estimation にて連続点表現にし、3 パラメータの対数正規分布

$$f(x; \mu, \sigma, d) = \frac{1}{\sqrt{2\pi}\sigma(x-d)} \exp\left(-\frac{(\log(x-d) - \mu)^2}{2\sigma^2}\right)$$

を仮定してネルダー-ミード法⁶⁾による二乗和の最小化により、検体検査項目毎に d の値を決定した。検査値から d を減算し、対数をとることで検査値を適切に正規分布に写像することができる。

先行研究により指数分布が適切であった検査項目の一部 (TBA での CRP、D-Dimer) および白血球分画は対数正規分布での近似にて適切な結果を得られなかったため、今回の検討から除外した。各検査値は負の値を取りえないので、 d の値を最適化する際に d が正であるという制約をかけた。TBA の実装には Python(3.8.5)、Jupyter Notebook(6.4.0)、SciPy(1.6.3)、Seaborn(0.11.1)、scikit-learn(0.24.2)を利用した。JIC の実装には R(3.6.3)、Perl(5.3.0)を用いた。TBA については Windows10、JIC については Linux(4.19.0)環境にて実行した。TBA と JIC の両機関から得られた検査項目の中から MCV と HT と RBC を使い、対数正規分布の期待値 $\exp(\mu + \sigma^2/2 + d)$ を使って、これらの検査値間にある関係式 $HT/RBC \times 10$ の成立を確認した。

3.3 検査値ペアの共分散

元データから検体検査に対応する d を減算し対数をとった値のペアを、全検体検査項目毎に作成した。作成にあたっては検査値のペアを 1 患者について平均し、1 患者あたり 1 点の検体検査値ペアを構成した。こうして作られたペアのデータについて、サンプル数が 50 点未満のものは結果の信頼性に疑義があるとみなして削除した。残ったペア (TBA 男性 1326 ペア、TBA 女性 1711 ペア、JIC 男性 496 ペア、JIC 女性 253 ペア) について散布図を作成し目視確認の上で、標本分散を求め、分散共分散行列を構成した。

3.4 Graphical lasso の適用と可視化

すべてのデータについて Windows10 上の Python にて scikit-learn の graphical_lasso ライブラリ (0.20) を用いて GLASSO を適用し、ネットワークの枝数がおおよそ 30 から 50 になるように正則化パラメータを選定した。得られたネットワークを networkx ライブラリ(2.6.2)と matplotlib(3.4.2)にて図示した。GLASSO においてネットワークの枝の太さは分散共分散行列の逆行列 (精度行列) の成分を示している。すなわち、検体検査値のペア X_i と X_j からそれぞれに対応する d を引いて対数変換した値を Z_i と Z_j とし、同時分布 $p(Z)$ が、

$$p(Z) \propto \exp\left(-\sum_{i,j} A_{ij} Z_i Z_j\right)$$

にて示されるときの A_{ij} が枝の太さに対応する。得られた結果の妥当性と解釈を医師が検討した。

4. 結果

4.1 検査値の分布の例

図 1 は検査値の分布の例として、RBC (赤血球数) について、TBA と JIC の男女について Kernel density estimation によるヒストグラム (細線) と 3 パラメータの対数正規分布による近似 (太線) を示したものである。ヒストグラムは対数正規分布でよく表現され、医療機関による差異より性別による差異のほうが大きなことが分かる。共用基準範囲²⁾では男性が下限 4.35、上限 5.55 とされ、女性が下限 3.86、上限 4.92 とされているが、本研究の結果では最頻値は全てこの範囲内にある。

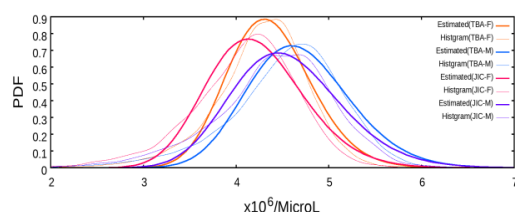


図 1 2つの医療機関での赤血球数のヒストグラムと対数正規分布による近似

表 1 は近似した対数正規分布の期待値を使って、平均赤血球容積 MCV を $HT/RBC \times 10$ の式によって計算し、測定された MCV を対数正規分布で近似したときの期待値と比較した結果である。差は最大で 2% 程度であることが分かった。

表 1 対数正規分布の期待値による平均赤血球容積の評価