

日本医療研究開発機構 医薬品等規制調和・評価研究事業 事後評価報告書

公開

I 基本情報

研究開発課題名：（日本語）大規模医療情報の安全対策への活用拡大に向けたデータ品質管理に関する課題解決の研究
（英語）Study for Social Issues about Data Quality Management to Extend Medical Big Data Use in Regulatory Science

研究開発実施期間：令和元年7月1日～令和4年3月31日

研究開発代表者 氏名：（日本語）中島 直樹
（英語）Naoki Nakashima

研究開発代表者 所属機関・部署・役職：
（日本語）国立大学法人九州大学・大学病院メディカル・インフォメーションセンター・教授
（英語）Kyushu University Hospital・Medical Information Center・Professor

II 研究開発の概要

MID-NET®(Medical Information Database NETwork)は、本邦における医療ビッグデータを活用した医薬品等の安全対策の高度化の推進を目的として、電子カルテ等の医療情報を大規模に収集・解析を行う医療情報データベースであり、独立行政法人医薬品医療機器総合機構（以下、「PMDA」）により管理・運営されている。2008年に厚生労働省医薬食品局から発出された「薬害肝炎事件の検証及び再発防止のための医薬品行政のあり方検討委員会」の提言を受け、PMDAは電子診療情報の二次利用による医薬品の安全性に関する評価体制を構築するためにMIHARI Projectを開始した。これがMID-NET®の前身プロジェクトであり、「医薬品の安全対策等における医療関係情報データベースの活用方策に関する懇談会」より発出された「電子化された医療情報データベースの活用による医薬品等の安全・安心に関する提言（日本のセンチネル・プロジェクト）」に基づき、厚生労働省医薬食品局はPMDAと共に「医療情報データベース基盤整備事業」を2011年度より始め、2018年度の運用開始につながっている。

MID-NET®のデータは、現在10拠点23病院の協力医療機関における、電子カルテ、レセプトおよびDPCのデータをデータソースとして利用しており、各医療機関に設置したMID-NET®専用の統合データベースに格納されている。各拠点のデータをもとに医薬品行政の判断に活用するためには、評価可能な形式とするための「データ標準化」ならびに、データベースに格納されたデータがデータソースのものと同等であることを管理する「データ品質管理」を、PMDAならびに協力医療機関が協働して継続的に行うことが必須である。

本研究では、MID-NET®のデータを始めとした医療情報を、より広範囲に活用していくために必要な「データ標準化」「データ品質管理」を中心とした活動を中心に、各医療機関で抱える問題とその解決方法を検討した。

1) リアルタイムバリデーション手法の施設間比較

データを標準化するために、MID-NET®ではPMDAによる医薬品、検査、傷病名を中心としたコードの標準化提案を行っている。拠点となる各協力医療機関においては、その提案をもとに現場のデータとの整合性を確認した後、マッピングB表として標準マスタのマッピングテーブルを作成している。この手法によってMID-NET®におけるデータの標準化は達成されており、質の高いデータベースの作成の一助となっている。しかしながら、各医療機関における病院情報システムのデータの標準化はこの手法では達成できていないことが危惧された。また、PMDAによる標準化提案はその作業量の膨大さといった側面から、各協力医療機関宛に年に2回程度のマッピングテーブル作成に留まり、臨床現場の問題を迅速に解決するためには、リアルタイムでのマスタ標準化、すなわちリアルタイムバリデーションが望まれる。

そこで、本研究では、各医療機関における病院情報システムのマスタに注目し、医薬品、検査、傷病名とその関連マスタの差分を取得し、この差分に対してのバリデーションを行うことを企図した。この手法は、バリデーションの対象をマスタ全体ではなく差分とすることで、バリデーション作業自体の効率化に資するため、リアルタイムでのバリデーションを可能とするものである。

具体的には、電子カルテのマスタのうち、医薬品関連マスタ（薬剤、持参薬、投与経路、用法、薬剤単位）、検査関連マスタ（検体検査、薬物血中濃度、細菌検査）、傷病名関連マスタ（病名、病名転記、診療科）の11マスタを対象とする差分抽出ツールを開発した。協力医療機関が採用している電子カルテベンダーは4社に亘り、それぞれのベンダーにおいて同等の出力仕様のツールを開発することとなった。このツールはバッチファイルでファイル出力まで行う仕様であるため、協力医療機関における作業は特に発生せず、差分ファイル出力における負担を最小限のものとした。

最終的には、MID-NET®の協力医療機関10施設のうち、9施設へ本ツールを導入し、週1回稼動することで、ガバナンスセンターによるマスタガバナンスにつなげることができた。

2) MID-NET®本格稼働後に生じたシステム運用・データ品質課題の整理と改善策の検討

MID-NET®は本邦における大規模医療情報を用いたデータ分析の先駆けともいえ、本格運用までの構築過程で多くの課題を解決してきた。未踏の領域ということもあり、本格稼働後においてもシステム運用面およびデータ品質面において課題が発生するため、順次解決している現状がある。九州大学病院においても、本格稼働後の2019年以降に解決すべき課題が29件発生し、それぞれに対して解決してきた。これらの課題には病院情報システムのリプレイスに伴う課題を含むが、これは本事業を継続していく上で各協力医療機関が常に意識しておくべき課題である。

このように大規模医療情報を扱うデータ駆動型事業において各医療機関が行うべき項目や課題と、その解決法について、先行研究の内容を踏まえて「大規模医療情報の安全対策への活用拡大に向けたデータ品質管理に関する課題解決の研究」と題した教材にまとめた。

3) SS-MIX2 拡張ストレージ有効性の調査

臨床現場から作成された大規模医療情報をより広範囲に活用していくためには、SS-MIX2 標準化ストレージにのみならず、SS-MIX2 拡張ストレージの活用も重要になる。SS-MIX2 拡張ストレージは、J-Dreams、J-Impact といった臨床学会のデータベース事業などでの活用が実用化されてきており、xml、CSV、PDF といった様々なファイル形式を扱うことができる柔軟性を特徴とすることから、今後の更なる活用も見込まれる。本研究では、SS-MIX2 推進コンソーシアムに協働を依頼して、全国規模でのSS-MIX2 拡張ストレージの調査を実施した。

2021年3月末調査では、全国1,652医療施設にSS-MIXストレージが導入されており、73.5%にあたる1,214施設で標準化ストレージを、68.5%にあたる1,131施設で拡張ストレージを導入していることがわか

った。標準化ストレージの利用目的としては、バックアップ（災害時、システムダウン時、リプレイス時）や地域連携、システム参照、臨床研究といった項目があげられた。拡張ストレージは、保管ファイルの種類として、文書と画像・動画の2種類に分けられた。文書としては、記録（医師経過記録、看護記録）、診断書（問診票、退院時サマリ）、レポート（放射線、病理、手術）、紙スキャン（臨床検査、処方）といった項目が、画像・動画としては、部門システム（放射線、眼科、耳鼻科）、エコー、心電図（画像、波形データ、構造化された自動解析結果）、心カテなどがあげられ、多岐に亘る活用がなされている現状が把握できた。特に地域医療ネットワークでの活用事例として、青洲リンクの所見、文書、画像など電子カルテ情報の一部情報や薬剤の調剤情報、MMWIN みんなのみやぎネットの調剤情報、診療情報、介護情報、そしてeClinicalBase (eCB) 医療イノベーション推進センターの糖尿病ミニマム項目セットや心血管系などのイベント情報や副作用有害情報などの活用事例が収集できた。

4) データ駆動型医学研究をテーマとした研究会の開催

臨床研究の分野としてのデータ駆動型の研究については、概念が確立されたのが最近であるということもあり、臨床現場の医療者にとってもまだまだ身近には感じられていない現状がある。日々の診療の結果を利用した大規模医療情報を用いた研究は、MID-NET®を代表としたレギュラトリーサイエンス分野においても注目されており、裾野を拡げてより多くの研究者による参画が望まれる。そのために、データ駆動型医学研究をテーマとした研究会を継続して開催してきた。

本研究結果として、新型コロナウイルスの蔓延といった社会情勢の変化により、予定通りの開催ができなかったが、2019年度7回、2020年度3回、2021年度12回の研究会を企画・開催した。

In Japan, Medical Information Database NETWORK (MID-NET®) is a database that collects and analyzes medical information, such as electronic medical records, on a large scale to promote advanced safety measures for pharmaceuticals and other products using medical big data. It is managed and operated by the Pharmaceuticals and Medical Devices Agency. In this study, we examined the problems encountered by each medical institution along with their solutions, focusing on activities related to “data standardization” and “data quality control,” which are needed for extensive usage of medical information, including MID-NET® data.

(1) Comparison of real-time validation methods among institutions

We focused on some masters of electronic medical record at each medical institution and determined differences in the names of drugs, laboratory tests, and injuries/diseases and their related masters. Next, we validated the masters against these differences. This method enables real-time validation because these differences, not the entire masters, are the targets of the validation process, which contributes to labor-saving in this validation process. In particular, a difference extraction tool was created for 11 electronic medical record masters, including drug-related masters (drugs, route of administration, dosage, and drug units), test-related masters (specimen tests, drug blood level, and bacterial tests), and masters related to names of injuries and diseases (names of diseases, posting of disease names, and medical departments). Four electronic medical record vendors were employed by the cooperating medical institutions, and each vendor developed a tool with equivalent output specifications. Because this tool was designed to provide output in a batch file, no additional work was required at the cooperating medical institutions, and the burden of difference file output was minimal.

Ultimately, the tool was introduced to 9 of 10 MID-NET® cooperating medical institutions and operated once a week, leading to master governance by the Governance Center.

(2) Organization of system operation and data quality issues arising after the full-scale launch of MID-NET® and study improvement measures

Each medical institution should focus on specific items for handling large-scale medical information; these items are summarized in a teaching material entitled “Research on Problem Solving for Data Quality Management for Expanding Utilization of Large-Scale Medical Information for Safety Measures,” which is based on the results of previous research.

(3) Research on the use of SS-MIX2 extended storage

To obtain large-scale medical information from clinical sites, it is crucial to use SS-MIX2 extended storage in addition to SS-MIX2 standardized storage. In this study, we requested the SS-MIX2 Promotion Consortium to collaborate with us to conduct a nationwide survey on SS-MIX2 extended storage.

Up to March 2021, it was found that 1,214 facilities (73.5% of the facilities with SS-MIX2) used standardized storage, and 1,131 facilities (68.5%) used extended storage. The data related to their use in several initiatives, such as regional medical networks (Seishu Link, MMWIN Minna no Miyagi-net) and eClinicalBase Medical Innovation Promotion Center, were also collected.

(4) Conducting workshops on data-driven medical research

Research using large-scale medical information that uses the results of daily medical treatment is attracting attention in the field of regulatory science, as represented by MID-NET®. After expanding its base, more researchers will participate in this field. To this end, we have continued to conduct research meetings on data-driven medical research.

Because of this research, we planned and conducted seven workshops in FY2019, three workshops in FY2020, and twelve workshops in FY2021, although we were unable to conduct these workshops as scheduled due to changes in social conditions, such as the spread of the novel coronavirus.