

医薬品等のベネフィット・リスク評価のための医療情報データベース システムの品質管理及び標準化手法に関する研究

業務主任者 鈴木 隆弘 千葉大学医学部附属病院 准教授

研究要旨：医療情報を集積したデータベースシステムは医薬品等の安全対策のさらなる向上等、今後の医療行政において重要なツールになることが期待されているが、データベースシステムを安全対策等に用いるには、解析手法の高度化を図ることはもちろんのこと、その前提として、解析に供するデータベースが高い信頼性を有していることが担保されている必要がある。本研究では、医療情報データベース基盤整備事業において構築した医療情報データベース（以下「MID-NET®」という。）をモデルとして、データベースの品質管理のための手法の検討及び課題整理を行い、医療情報のデータベースの品質管理に資する知見を得ることを目的とする。

研究方法：本研究では、MID-NET の品質管理手法検討の事前準備として、病院情報システム（以下「HIS」という。）及びMID-NET の統合データソース（以下「DS」という。）からのデータ抽出方法の検討を行った。その後、両者のデータ件数、内容をデータ種別ごとに比較検討を行うとともに、当該比較の際の課題等を整理した。併せて各作業の工数等についても検討した。

結果：平成25年8月分のデータを対象として、データ項目毎にHISのデータウェアハウス（以下「DWH」という。）からの抽出を行い、ID変換及び送信対象スライスを抽出したデータと、DSから抽出したデータを突合した。来院等情報や放射線実施はほとんど一致したが、データ種別によっては一致率が低いものが認められた。原因と対応方法を検討中であり、一部は既に解決することが出来た。

まとめ：1ヶ月分のデータを用いてHISとDSの内容のバリデーションを行った。基本的な情報は一致していたが、詳細なデータでは多くの差異が認められた。移行中のデータを用いたため、原因は様々だがいくつかのパターンを見つけることが出来た。今後は検証期間を拡げてデータベース全体のバリデーションを行う必要がある。

担当責任者

鈴木 隆弘 千葉大学医学部附属病院

A . 研究目的

医療情報を集積したデータベースシステムの利活用は、医薬品等の安全対策のさらなる向上等のために重要なツールとなることが期待されているが、データベースシステムを医薬品等の安全対策等に適切に利活用するためには、解析手法の高度化を図ることはもちろんのこと、解析に用いるデータベースシステムが高い信頼性を有していることが求められる。

信頼性の高いデータベースシステムを確立するためには、複数の病院情報システム等からデータベースシステムに集積されるデータの標準化やデータベースシステムの品質管理を確実に実施するための手法を確立する必要がある。

近年はがん登録をはじめとして、様々な学術団体でデータを収集し、臨床データベースを作る動きが広まっている。しかし、これらのデータベースは各々が独自に構築しているため相互に連携がなく、項目の標準化もされていないので、同じようなデータを何度も入力する必要があるなど臨床側の負担となっている。

そこで、本研究においては、厚生労働省及び独立行政法人医薬品医療機器総合機構（以下「PMDA」という。）が実施する「医療情報データベース基盤整備事業」の協力医療機関の一つとして千葉大学医学部附属病院（以下「当院」という。）に構

築した MID-NET を用いて、データベースの信頼性を確保するためのデータの標準化と品質管理手法及び他のデータベースとの連携について検討するとともに、迅速かつ低コストでデータベースを連携する上で必要な要件、課題等について調査し、データベースシステムの品質管理に資する基礎情報とする。

B . 研究方法

データ抽出・比較等の処理については、全て千葉大学医学部附属病院内で実施する。

1.MID-NET を用いた品質管理手法検討の事前準備（病院情報システム及び統合データソースからのデータ抽出等）

MID-NET は、図 1 に示すように、当院の病院情報システム（HIS）の電子カルテ、レセプト等の医療情報の各種データについて、インターフェースサーバを介し SS-MIX 標準化ストレージに移行させ、さらに匿名化等の加工処理を加えて統合データソースに蓄積する仕組みとなっている。医療情報データベース基盤整備事業において医薬品等の安全対策のためにデータを利活用するのに際しては、この統合データソースに蓄積されたデータが解析対象となる。

MID-NET を用いた品質管理手法検討の事前準備として、当院の HIS 及び MID-NET の統合データソースからのデータ抽出手法

の検討を行い、比較検討のために必要なデータの抽出及び ID 変換等のデータ処理を行った。HIS からのデータ抽出にあたっては当院の HIS の開発・保守等を行う東芝医療情報システムズと、また、統合データソースからのデータ抽出にあたっては、PMDA と抽出条件等について協議の上、実施した。

2. データ標準化・品質管理に係る手法の検討

HIS 及び統合データソースから抽出したデータの件数、内容を適切に比較するための手法の検討を行った。PMDA の担当者と連携のもと、抽出データの突合を行い、比較を行った。

さらに、抽出データの比較等により、HIS から統合データソースにデータが移行される過程でデータの標準化処理が適切になされているかについて確認した。

3. データ標準化・品質管理に必要な工数等の把握

上記 2 を当院で実施するために必要な作業に要した工数について、当院担当者及び連携して作業を行った PMDA 担当者それぞれについて調査した。

4. データベースの品質管理を適切に実施するための課題整理

当院にて MID-NET の統合データベースに蓄積するデータ等の品質管理を継続的に実施するために留意すべき事項等につい

て整理を行った。

5. 他のデータベースとの連携に必要な課題整理

当院は特定機能病院として高度な医療を提供するだけでなく、医療情報の電子化と利活用に取り組んでいる。そこで、図 2 に示すように、SS-MIX 標準化ストレージ及び統合データソースに蓄積したデータと当院における他の臨床研究用のデータベースとの連携の可能性について、千葉大学バイオバンクシステムを対象として検討を行った。

特に論点となったのは匿名化 ID(セカンド ID)の取り扱いと、データの取得先及びタイミングであった。

(倫理的配慮)

本研究において、医療情報は、原則的には個人特定情報は匿名化処理を実施して取り扱った。またそれらに近い情報も、情報の特性に応じて暗号化等、厳重な管理の下において取り扱った。また本研究においては診療等の結果たるデータのみを取り扱うことから、研究対象者に対する侵襲等は発生しないと考えるが、研究対象者の人権を侵害することがないよう十分に配慮して医療情報を取り扱った。

C . 研究結果

1. MID-NET を用いた品質管理手法検討の事前準備 (HIS 及び統合データソースからの

データ抽出等)

当院のデータは現在、統合データソースに移行中であるが、蓄積されているデータのうち、平成25年8月分のデータを本研究の検討対象とした。

HISのデータウェアハウス(DWH)から、データ種別ごとに以下の日付項目を基準としてデータの抽出を行った。

データ種別	抽出基準項目
来院等情報(外来)	来院日時 2013年8月1日～ 2013年8月31日
来院等情報(入院)	入院日 2013年8月1日～ 2013年8月31日
来院等情報(退院)	退院日 2013年8月1日～ 2013年8月31日
傷病情報 (病名オーダ)	開始日 2013年8月1日～ 2013年8月31日
処方情報(オーダ)	実施開始日 2013年8月1日 ～2013年8月31日
注射情報(オーダ)	投与開始予定日時 2013年8 月1日～2013年8月31日
注射情報(実施)	投与開始実施日時 2013年8 月1日～2013年8月31日
検体検査情報	検査日時 2013年8月1日～ 2013年8月31日
放射線検査情報	実施開始日時 2013年8月1 日～2013年8月31日
生理検査情報	実施開始日時 2013年8月1 日～2013年8月31日
細菌検査情報	検査日時 2013年8月1日～ 2013年8月31日

さらに、ID変換及び送信対象スライスを抽出した(以下「HIS抽出データ」という)。なお、スライスとは、連番で振っている患者番号の、その下2桁をキーとして論理的及び物理的に分割された100個の

箱(スライス)にそれぞれ振り分けて格納している処理単位である(例えば、患者プロフィールを抽出するという場合には全データの1/100のデータが最小単位となる)。

また、統合データソースからは、診療日をもとに平成25年8月のデータを抽出した(以下「DS抽出データ」という)。HIS抽出データ及びDS抽出データは比較用IDへ変換した。

2.データ標準化・品質管理に係る手法の検討

当院のデータは、スライス毎に移行されている。そこでまず、データ比較のために基準とする項目(以下「基本キー」という。)に比較用IDを用いて突合を行った。さらに、基本キーで一致したレコードについて、追加で基準とした項目(以下「追加キー」という。)を加えて突合を行った。

各データ種別の突合の際に用いた基本キー及び追加キー並びに突合の結果を、別紙「件数比較結果シート」に示した。

来院等情報(外来)

・基本キー及び追加キーでの比較の結果、一致率は98.3%であった。HIS抽出データのみにあるデータの一部は、DS抽出データに土曜日及び日曜日のデータが存在しないことに起因することが判明した。

来院等情報(入院)

・HIS抽出データとDS抽出データで診療科の異なるレコードが1件のみ存在した。

来院等情報（退院）

- ・入院と同じく HIS 抽出データと DS 抽出データで診療科の異なるレコードが 1 件のみ存在した。
- ・HIS のみに存在する ID が 4 名（4 件）存在した。

傷病情報（退院サマリ）

- ・退院サマリでは表記をはじめ、病名の情報の取り扱いについて違いが認められた。

傷病情報（病名オーダ）

- ・病名オーダでは転帰と疑い病名フラグの転送漏れが判明した。

処方・注射情報

- ・処方・注射オーダでは一致率が 70% 前後と低かった、用法・用量の取り扱いで HIS 抽出データと DS 抽出データに違いが認められ、対処方法を検討中である。

検体検査情報

- ・検体検査では一致率が 80% 台であった。調査の結果、外注検査結果の一部が DWH から抽出されないことが判明した。他にも件数は少ないものの、コードのずれや結果値が送信されていないレコードなどが認められた。

放射線検査情報

- ・放射線検査情報については HIS 抽出データのみ ID が 6 名分（52 件）存在したのみで、ほぼ一致していた。

生理検査情報

- ・HIS 抽出データのみ ID が 45 名分（50 件）、DS 抽出データのみ ID が 1 名分（1 件）存在した。
- ・消化器系の検査で HIS 抽出データのみ

存在するデータが 29 件存在した

細菌検査情報

- ・細菌検査では菌名の齟齬や結果値の合わないレコードが少数認められた。
- ・HIS 抽出データのみ ID が 7 名分（80 件）、DS 抽出データのみ ID が 14 名（236 件）存在した。

3. データ標準化・品質管理に必要な工数等の把握

今回の作業において、病院側は延べ 32 人日、PMDA スタッフは約 30 人日の工数を必要とした。

5. 他のデータベースとの連携に必要な課題整理

千葉大学バイオバンクシステムとの連携試験のために HIS からのデータ取得ソフトウェアを作成し、匿名化 ID には HIS から MID-NET への情報通信で用いるセカンド ID を用いた。

当初は統合データソースサーバからのデータ取り込みを想定していたが、検体の取得から登録の流れを精査したところ、バイオバンクへの登録時に必要な情報を、MID-NET 用に開発したリソースを用いて HIS から取得し、検体番号とセカンド ID を紐付けて記録するのが適当と判断した。症例検索などの際には MID-NET の検索・統計機能を用いて連携を試みる。

バイオバンクシステムでは主に病理組織情報が必要とされるため、データ項目のうち、MID-NET に存在するものは 1 / 3

以下と少なかった。また、当院ではDPC及び病理システムが共にHISと独立しているために、DWHからのデータ取得が出来ないことから、コピ&ペーストを自動化するソフトウェアも併用し、ユーザーの負担軽減と転記ミスの軽減を図ることとした。今後は本運用を開始し、データ項目の追加と標準化及び連携手法についての調査を継続していく。

D．考察

当院のHISからのデータ抽出については既存のDWHを用いることで効率よく行うことが出来、事前準備として今回の手法は有効と考えられた。

HISと統合データソースから抽出したデータの件数、内容の突合比較では多くの差異を検出することが出来たが、現在データ移行中であるため、それに起因した差異が認められている可能性もある。これらの差異はデータ移行完了後に確認するとともに、更に効率的な手法を追求する必要がある。また、統合データソースへのデータ移行過程での標準化処理については他院との比較も必要である。

今回、1ヶ月分のみを検証で約60人日の工数を要した。データベース全体のバリデーションを行うためには、単純に期間の倍数ではないとしても、相当の工数を必要とすると推定される。

本研究の限界として、実施期間が短いことから、品質管理を継続的に実施する

ために留意すべき事項については今後の課題となった。他データベースとの連携についても、バイオバンクそのものの実運用が来年度からなので、今後も新たな課題が発見されると思われる。

E．結論

1ヶ月分のデータを用いてHISと統合データソースの内容のバリデーションを行った。基本的な情報は一致していたが、詳細なデータでは数多くの問題点を見つけることが出来た。原因は様々であったがいくつかのパターンを見つけることが出来、一部は既に解決を得た。今後は検証期間を拡げてデータベース全体のバリデーションを行う必要がある。

F．健康危険情報

なし。

G．研究発表

なし。

H．知的財産権の出願・登録状況（予定も含む）

なし。

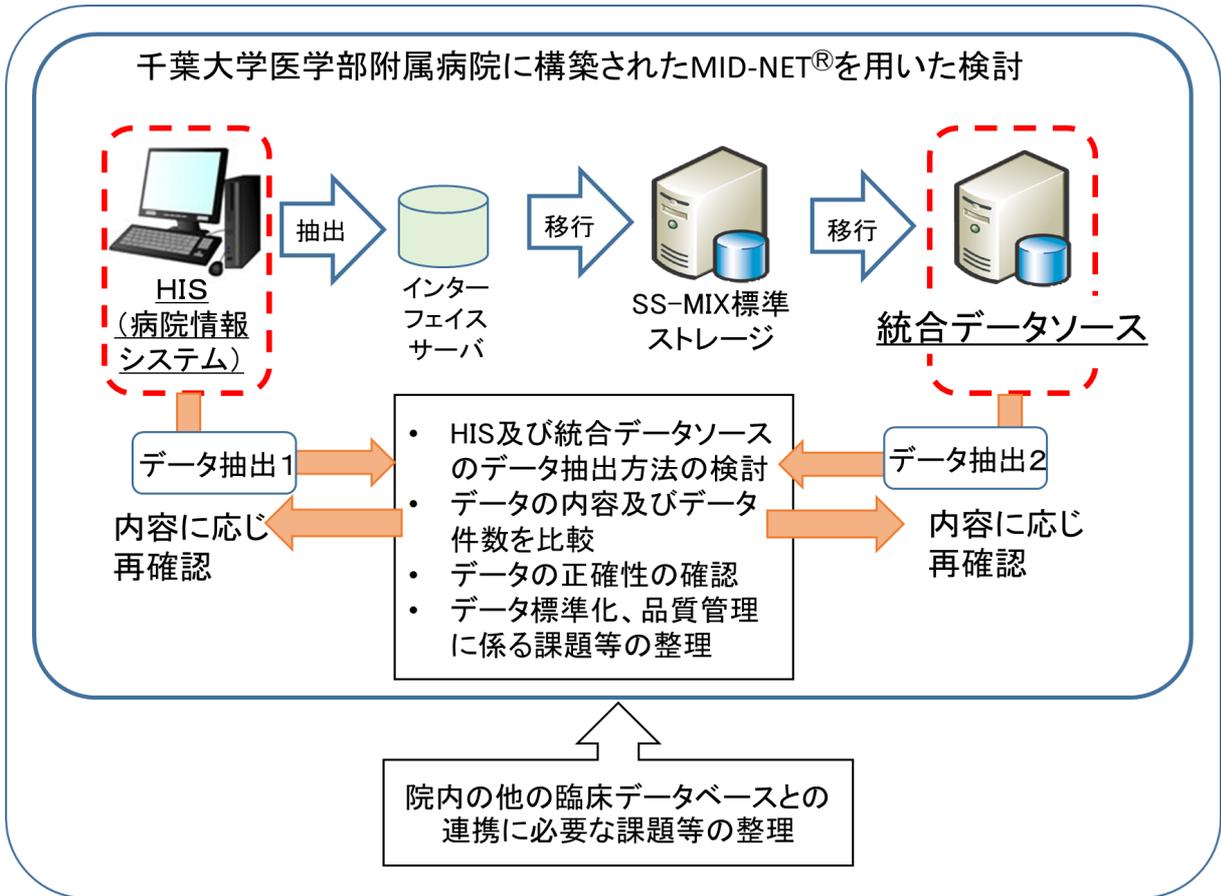


図1 . 千葉大学医学部附属病院に構築された MID-NET を用いた検討

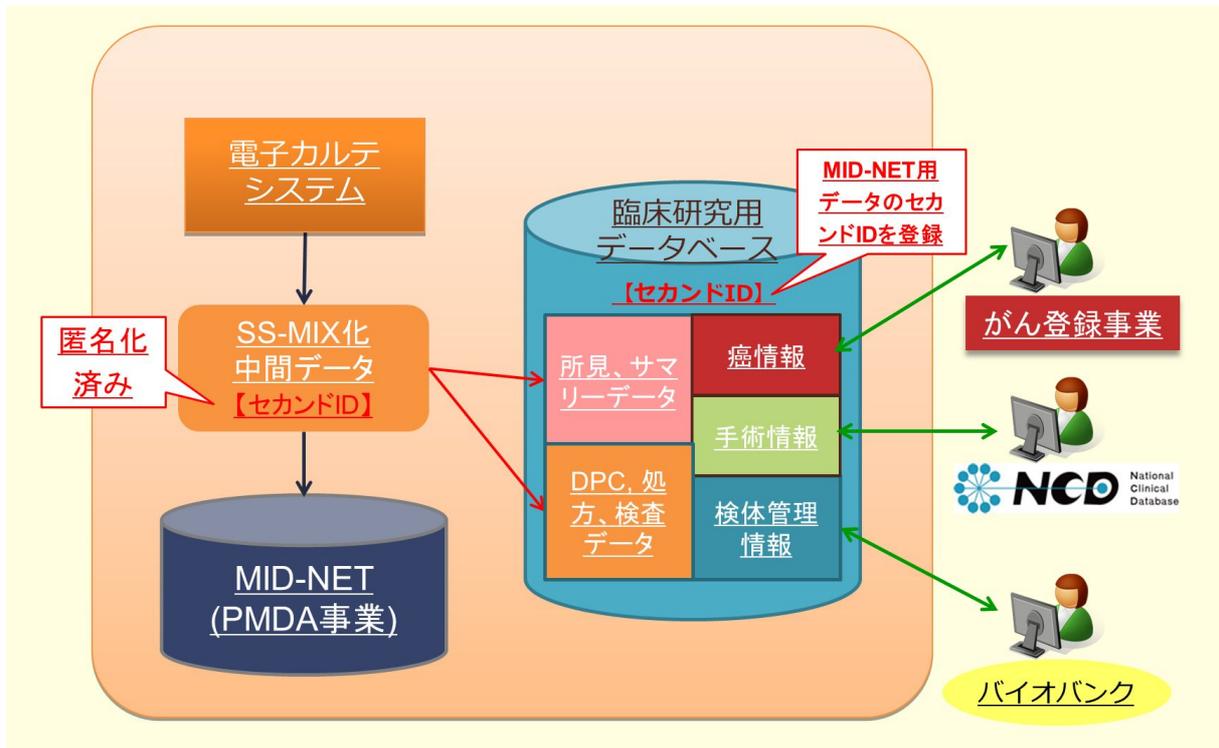


図2 . 医療情報データベースシステムと臨床研究用データベースの連携方法