

(令和) 5 年度厚生労働行政推進調査事業費補助金
(医薬品・医療機器等レギュラトリーサイエンス政策研究事業)
総括研究報告書

適切な医薬品安全性評価のための国際統合化を考慮した医療情報データベースの
品質管理・標準化に関する研究

研究代表者 横井英人 香川大学医学部附属病院 医療情報部 教授

研究要旨

本研究は、独立行政法人医薬品医療機器総合機構が管理・運営する医療情報データベース(MID-NET®)をはじめとした医療情報データベース(DB)の信頼性確保の手段について、その国際統合性を高め、医療情報 DB の医薬品安全性評価での利活用促進と医薬品安全対策の高度化に資する成果を創出することを目的とする。1 年度目については、海外の医療情報 DB の実態調査を行うこととした。研究班事務局が作成した調査項目案に対して分担研究者とともに医療情報管理の観点や MID-NET®の運用手順の観点から検討し、調査項目を設定した。この調査項目とインタビューフォームに基づいて調査を行い、その結果を踏まえ、あるべき医療情報 DB の性状について分担研究者と議論した。

中島 直樹

九州大学 大学院医学研究院医療情報学講座
教授

武田 理宏

大阪大学 大学院医学系研究科医療情報学講座
教授

安西 慶三

佐賀大学 医学部内科学講座 教授

原田 紗世子

独立行政法人医薬品医療機器総合機構 医療
情報科学部 調整役

A. 研究目的

本研究は、医薬品安全性評価に使用する医療情報データベース（診療録等の情報を電子計算機を用いて検索することができるように体系的に構成したもの。以下、「医療情報 DB」と略す。）に関する国際統合性を考慮した品質管理および標準化の指針の基盤となる施策を検討し、医薬品の安全性評価に利用する医療情報 DB の信頼性保証のための体制の構築と維持を的確に実行するための指針となる提言を行い、医療情報 DB の医薬品安全性評価での利活用促進と医薬品安全対策の高度化に資する成果を創出するものである。

本邦において様々な医療情報 DB が構築・運用されているが、その仕様等は様々である。医療情報 DB で集積されたデータを医薬品の安全性評価に活用するためには、医療情報 DB の信頼性の保証が重要となる*。そして、その信頼性は、医療情報 DB のデータ集積フローの観点から、『第一フェーズ：医療機関における病院情報システムでのデータの品質管理』、『第二フェーズ：病院情報システムから医療情報 DB へのデータ移行プロセスの品質管理』、および『第三フェーズ：医療情報 DB における集積データの品質管理』に依存すると考えられる。特に、第二フェーズは、医療機関における診療録等の情報を正確に DB 化するという点で医療情報 DB の信頼性に大きく影響する部分であると考えられる。例えば MID-NET®は、医療機関におけるオリジナルデータと MID-NET®集積データとの一致性等を保証する品質管理を継続的に実施することで、MID-NET®利活用者への高品質なデータ提供を実現している。従って、医療情報 DB の信頼性を保証するためには、医療機関における電子カルテシステム等の病院情報システムにおけるデータの品質管理とそのデータが正確に医療情報 DB に集積されていることを保証する体制整備が必要と考えられる。しかしながら、現状、その体制を構築し、維持するた

めの指標（いつ、何を、どこで、誰が、なぜ、どのように、という点）を明示したものはない。

本研究は3か年に亘り実施する計画で開始し、その初年度である令和5年度は、米国および欧州で医薬品安全性評価に活用されている医療情報DBの特徴、品質管理や標準化の考え方について情報収集し、それらを医療情報DB間で比較することを目的として実施した。

*厚生労働省、製造販売後の医薬品安全性監視における医療情報データベースの利用に関する基本的考え方について(平成29年6月9日、薬生薬審発0609第8号/薬生安発0609第4号)

B. 研究方法

医薬品の安全性評価に利用実績のある米国の Sentinel Initiative、英国の National Health Service (NHS)の基幹システム・Clinical Practice Research Datalink (CPRD)、および仏国の Système National des Données de Santé (SNDS)の医療情報DBの品質管理に関する情報収集を下記の手順で実施した。

1) 調査項目の設定

本邦の厚生労働省が発出している『医薬品の製造販売後データベース調査における信頼性担保に関する留意点について(平成30年2月21日、薬生薬審発0221第1号)』に示されている“製造販売業者等が製造販売後データベース調査を行う上で、再審査等の申請資料の信頼性を担保する観点から留意する必要がある事項”を参照して、本研究において対象とした医療情報DBに対する調査項目を設定した。

2) 情報収集

上記で設定した各調査項目に対する情報を以下の手順に基づいて収集した。

- ① 対象の医療情報DBに関するWeb上の公開情報を収集する。
- ② Web情報だけでは回答できなかった調査項目については、現地の運営スタッフ等から直接的に電子メールやインタビュー等によって情報収集する。

3) 調査結果のまとめ

収集した情報を元に各調査項目に対する回

答を記載した調査結果一覧を作成した。

4) 研究分担者との研究班会議

上述の調査項目の設定および調査結果一覧の作成に当たって、研究分担者との研究班会議を2回（調査項目の設定時、調査結果一覧の取り纏め時）開催し、内容の確認と精査を行った。

（倫理面への配慮）

本研究において患者の診療情報等の倫理面での配慮が必要となる情報は取り扱わなかった。

C. 研究結果

下記を調査項目として設定した。

- ① 全般的事項
 - 名称
 - 設立主体
 - 運用開始年
 - 設置目的
 - 年間運営予算
 - 利用金額/利用者
 - 格納済み医療データの種類、例数
 - データソース
 - データをアップロードする際の課題
 - 情報源となる医療機関数
 - データ集積フロー概要
 - 利用実績
 - 特徴 (Strength)
 - 欠点 (limitation)
- ② 事業者の組織体制
 - 責任者等の所属部署及び職責
- ③ 情報源となる医療機関の選定
 - 選定基準
 - 選定手順
 - 選定に関する規定
- ④ 医療データの授受（収集）の方法
 - 5W1H
 - 手順、頻度
 - 医療データが正しく入力されたこと、取り込まれたことを確認する方法
 - 情報源となる医療機関等の電子カルテ等に入力された医療データと医療情報DBに取り込まれた医療データの整合性確認

- 整合性確認プロセスの明確化（手順書化等）と確認結果の記録とその保存
 - 仮想的に作成したサンプルデータではなく、一定期間の実データに基づく確認
 - 継続性を持った定期的な確認
 - 情報源となる医療機関等から収集した医療データの品質管理に関する規定
 - 情報源となる医療機関等の電子カルテシステム等の運用方法、部門システム等が変更された場合の対応
 - 医療情報 DB の構築に関する規定や手順が変更された場合の対応
- ⑤ 個人情報保護
- 医療データの収集及び提供に当たって、同意取得、匿名化などが個人情報保護法等の法令やガイドラインを遵守した対応を行っているか。
- ⑥ データクリーニング
- 当該業務の責任者、管理者、担当者の所属部署及び職責
 - 実施する段階（情報源から収集した医療データを医療情報 DB に取り込む段階、または解析用データセットを作成する段階等）
 - 対象となるデータ項目
 - 具体的な内容と方法
 - 実施基準を変更する際の手順や過程
 - 一部を外部委託する際の業務委託先の管理方法（手順、計画）
 - その他のデータクリーニングに関する規程
- ⑦ コーディング
- 当該業務の責任者、管理者、担当者の所属部署及び職責
 - 実施する段階（情報源から収集した医療データを医療情報 DB に取り込む段階、または解析用データセットを作成する段階等）
 - 対象となるデータ項目
 - コードリスト等（マスタ）の整備方法（更新頻度及び手段）
 - 実施基準を変更する際の手順や過程
 - 一部を外部委託する際の業務委託先の管理方法（手順、計画）
- その他のコーディングに関する規程
- ⑧ セキュリティ
- 当該業務の責任者、管理者、担当者の所属部署及び職責
 - 情報セキュリティの規程
 - 医療情報 DB のログイン・ログアウトの管理（ログイン・ログアウト記録の作成・運用・保存）
 - 医療情報 DB で取り扱う医療データの内容
 - システム構成
 - 医療情報 DB 操作場所への利用者の入退室の管理
 - 利用者管理の規程
 - 利用者の範囲
 - 利用者へのアカウント設定等の管理
 - アクセス制御の規程
 - 医療データの重要度に応じてユーザのアクセス権限を付与できるような設定、制御
 - ネットワークセキュリティの規程
 - 事業継続計画書（BCP）
 - その他のセキュリティに関する規程
- ⑨ データバックアップ
- 当該業務の責任者、管理者、担当者の所属部署及び職責
 - バックアップの内容及び手順
 - バックアップの対象
 - 実施頻度（全体、差分）
 - バックアップデータを保管する世代数
 - バックアップデータの保存に用いるメディア
 - バックアップデータの保存場所
 - バックアップデータの保存期間
 - リカバリーの計画
 - リカバリーの具体的な内容及び手順
 - リカバリーのテスト結果とその記録
 - その他のデータバックアップ及びリカバリーに関する規程
- ⑩ 解析用データセット作成
- 当該業務の責任者、管理者、担当者の所属部署及び職責
 - 対象・方法・手順・期間
 - 結果報告
 - 一部を外部に委託する際の業務委託先

の管理方法（手順、計画）

- 解析データセットを利用させる方法
- ⑪ 医療情報 DB の Quality Control
 - 当該業務の責任者、管理者、担当者の所属部署及び職責
 - 計画と報告に関する規程
 - 品質管理記録
- ⑫ 医療情報 DB の Quality Assurance
 - 当該業務の責任者、管理者、担当者の所属部署及び職責
 - 計画と報告に関する規程（品質保証の在り方に関する規程を含む）
 - 品質保証記録
 - 品質マネジメントシステムの目的・考え方
 - 対象となるデータ量
 - 対象となるデータの更新頻度
 - 対応メンバーに必要な資格
- ⑬ 教育訓練
 - 実施時期、時間
 - 実施内容
 - 対象者
 - 教育担当者（どのような立場の人か）
 - 訓練結果の評価
- ⑭ 記録の保存
 - 当該業務の責任者、管理者、担当者の所属部署及び職責
 - 保存対象
 - 保存場所、保存手順、保存期間
 - その他のルール
 - 移管する場合の場所と手順に関する規程
 - 廃棄の手順
- ⑮ アウトカム定義のバリデーション
 - バリデーション研究の実績
- ⑯ 医療データの真正性、見読性および保存性の担保
 - 遵守法令
 - 具体的な対応内容
- ⑰ 現状の課題・問題点および今後の対応等
 - 全般
- ⑱ データメンテナンス
 - データメンテナンスの責任者

なお、本研究での調査対象とした医療情報 DB に関する調査結果は“調査結果一覧”（別添資料①）に示した通りであった。また、開催した班会議の内容は別添資料②の通りであり、各分担研究者からの意見に対して注釈を同議事録に追記した。

D. まとめ

今回の調査結果から本研究の調査対象とした各医療情報 DB の全体像とデータ集積フローの概要、品質保証の取り組み、データクリーニング、個人情報保護および特徴について別添資料③のとおり整理した。

1) Sentinel Initiative

Sentinel では医療情報 DB として患者データの集積を行っておらず、各データ提供者（DP: Data Partner）に集積してあるデータに対して、FDA がデータを利用するタイミングでクエリを行うことができる。その際、DP はデータ分散型の共通データモデル（Sentinel Common Data Model: SCDM）にデータを変換し、センチネルの運営組織（SOC: Sentinel Operations Center）に送付する。データ授受の前後で DP および SOC にて品質管理のプロセスが実施される。DP は医療機関、保険会社、医療情報システム関連会社等 50 機関程度存在しており、データの提供頻度は DP によって異なる。

データを DP から受け取る際、品質管理のプロセスとして、SOC は下記の 3 つにレベル分類されたデータの品質保証プログラムに従ってデータをチェックする。

- レベル 1: プログラムにより、自動でデータの完全性と有効性をチェックし、欠落している変数、値、データ項目等を確認する。
- レベル 2: プログラムにより、自動でデータテーブル間の一貫性をチェックし、データの整合性を確認する。
- レベル 3: SOC のスタッフにより、前回抽出したデータと比較確認する。

SOC の品質保証（Quality Assurance: QA）チームが各 DP によって提供されるデータ品質レ

ビューおよび特性評価プログラムからの出力を受け取った後、下記の手順が **SOC** によって実行される。

- ① データ品質アナリストは、データ品質の受け入れ基準が満たされていることを確認するために、出力の一次レビューを行う。
- ② 一次レビューを行うデータ品質アナリストは、データ品質調査報告書を作成する。
- ③ 別のデータ品質アナリストが出力と同報告書の二次レビューを行い、データ品質の受け入れ基準が満たされていることを確認する。
- ④ 二次レビューを実行するデータ品質アナリストは、追加の調査結果または修正を同報告書に注釈として記載する。
- ⑤ データマネージャは、出力と同報告書を確認・確定し、**Sentinel Secure Portal** またはその他の承認済みの安全な経路を使用して **DP** に同報告書を送信する。

データクリーニングは **SOC**、**DP** いずれにおいてもほとんど実施しておらず、ソースデータを可能な限り改変していない。実施するかは **DP** の裁量に任せられているが、クリーニングを実施するケースにおいても変数の整理など、限定的な用途での実施が挙げられている。また、コーディングに関して、**Claims** データについては予めコーディングされているため、**Sentinel** にデータを変換・アップロードする際に追加のコーディングは通常発生しないが、電子カルテ等その他由来のデータは明確なコーディングが定まっておらず、利用用途によってコーディングの方法を検討し、分析段階で変数を定義しているケースもある。

データベース運営者にとって当システムの強みは、柔軟性とある程度の裁量が行使できることである。すなわち、データ提供者である **DP** と直接コミュニケーションをし、必要なデータの収集および品質管理ができる。また、その結果として、徹底した品質管理を実施することができ、詳細なチェック、比較を行うことができる。一方、欠点としては、規制上の問題により、データ利用者が個々の患者情報を取得できな

いため、実施できる研究が限られることが挙げられる。また、品質管理に非常に工数がかかり、大量のデータを扱う **DP** にとって非効率なプロセスである。電子カルテデータが現在分析をするためには不十分であり、異なるコーディングシステムを利用しているためデータの収集が難しい上、データの縦断性を認識するのも困難である。

2) NHS の基幹システム/CPRD

CPRD への参加同意を得た **GP** (**General Practitioner**) から毎月特定形式の **EHR** データが集積されている。**NHS** (**National Health Service**) **England** は患者の **NHS** 番号、生年月日、郵便番号、性別を仮名化し、**CPRD** に送付し、**CPRD** で **GP** からの **EHR** データと **NHS** からのデータが紐づけられている。情報源である **GP** の医療データは、**GP** で **ICD-10** や **SNOMED** (**Systematized Nomenclature of Medicine-Clinical Terms**) を利用しコーディングを行い、特定ソフトウェアにて **EHR** を作成し **CPRD** に発信する際に **QRM** システムや専用ソフトウェア等を利用し独自に品質保証が行われている。**CPRD** に送付された医療データはデータの項目、構成要素、長さ、形式を確認し整合性がチェックされる。また、データベースの孤立レコード (他データと紐づいていないデータ) を特定することで、データ登録の不整合等の検知が行われる。さらにデータ内容はデータ項目ごとの値の「許容範囲」を設定することで異常値を検知している。そして、**CPRD** は、**NHS England** の提供する「データセキュリティ基準に対するパフォーマンスを測定できるオンライン自己評価ツールキット」を用いて年次の評価を実施し、データを安全に保持するために必要な基準を満たしていることを証明する必要がある。

CPRD の強みは、科学的根拠に基づいた医療への貢献が可能な充実したデータであり、患者の症例発見や全国的な診療ベンチマークを含む、診療レベルの処方と患者の安全性に関する質向上レポートが作成できることである。一方、欠点は、現行システムの多くが手作業で行われており、高度なスキルを持つ人材に依存しているため、拡張性がないことである。こうした欠点により、正確なリアルタイムデータを収集す

ることが難しい点や長期的なデータ分析は可能であってもデータにバイアスがかかっているため、実際のテキスト化においてうまく機能しない点が挙げられる。

3) SNDS

情報源となる医療機関の医療データは、医療機関がハッシュ関数を用いて患者データの仮名化を行った上で、各国営 DB にデータ送信され、その後、各国営 DB にて再度、ハッシュ関数を用いた仮名化を行い SNDS にデータを送信されることで集積される。一般的に、SNDS のデータは診療報酬を得るために登録されており、データの局所性が高い。病院が提供している診療サービスに該当する最高額の報酬を得ることが目的となっている。病院によっては、高額な診療報酬が得られる薬剤の情報だけを登録していることがある。安価な薬は特別な償還がないためである。2008 年にシステムが構築された時のデータ収集目的は償還ではなかった。当時はデータの二次利用を念頭に収集されており、現在では 20 年分のデータがあり、疫学や特定の研究に利用することもできる。データ利用料は無料であるが、全てのデータにアクセス可能であるため当該医療情報 DB のデータへのアクセスは厳しく制限されており、研究に利用する場合にはプロトコールを作成し、特定のセキュリティトークンを取得する必要がある。

SNDS に集積される医療データは、病院で医師主導の専属チームが ICD-10 等を利用し電子カルテのデータのコーディングを行い、上述のとおり各国営 DB に発信するが、入力内容は保険機関の監査により信頼性が確認されている。国営 DB の一つを管理する AITH（医療機関情報技術機構）では、医療機関から受領したデータの品質管理を行うツールを複数利用し、データの整合性やコーディングのエラーを検知し、課題のある医療機関を特定するなどしている。そして、トレーサビリティ、認証状況、データの整合性を監視により医療情報 DB の品質を確保している。なお、データクリーニングは、医療機関ではソフトウェアを用いて実施されることはあるが、データ発信後は実施していない。

SNDS の強みは、すべての病院の処置データ

が利用可能である。一方、欠点として医療機関が診療行為等のデータを登録するか否かは当該償還額と関連しているため、収集データには偏りが生じる。そのため、有効な分析を行うためにはデータ収集に関する背景等を理解する必要がある。

E. 健康危険情報
該当なし。

F. 研究発表
該当なし。

G. 知的財産権の出願・登録状況
該当なし。