別紙３

厚生労働行政推進調査事業費補助金
地域医療基盤開発推進研究事業
分担研究報告書

医療等情報の二次利用基盤構築に向けた国際規格対応の研究
「OMOP Oncology Module 実装に向けた課題検討」

研究分担者　岡村浩史　大阪公立大学　臨床検査・医療情報医学　准教授

研究要旨
本研究では、がん領域のリアルワールドデータの二次利用基盤構築に向け、OMOP Oncology Module を日本の実臨床データから実装する際の課題を検討した。Cancer Disease/Treatment Model の table/field、データソース、変換ルール、マッピング可能性を整理した結果、疾患状態の判断、診断・治療修飾子、分子・バイオマーカー、レジメン・サイクル、国内コードから standard concept への変換、マスタ管理等が主要課題であった。収集項目のスコープ定義、各種項目の判定定義、標準・変換マスタ管理方法、自然言語処理や入力テンプレートを含むデータ収集手順の整備が必要である。

Ａ．研究目的

医療等情報の二次利用基盤構築に向けて、国際的な標準データモデルである OMOP Common Data Model（OMOP CDM）を国内医療データへ適用することが重要となっている。特にがん領域では、個別化治療の進展、治療選択肢および治療ラインの多様化、長期的な疾患経過の把握、希少がんを含む症例集積の必要性から、リアルワールドデータ（RWD）の活用ニーズが高い。

一方で、がん領域の疾患情報および治療情報は、ステージ、グレード、TNM 分類、分子・バイオマーカー等の診断修飾子、ならびに初発、寛解、再発、進行、転移、終末期といった病勢の経時的遷移を伴うため、一般的な疾患情報と同じ粒度で OMOP CDM へ変換することは困難である。

本研究では、日本の医療機関におけるがん領域データを OMOP Oncology Module へ実装する際の課題を体系的に整理し、将来的な国内実装に向けたデータ収集・変換方針を検討することを目的とした。

Ｂ．研究方法

OMOP Oncology Module の設計思想を踏まえ、Cancer Disease Model および Cancer Treatment Model を中心に、実装上重要となる table および field を整理した。具体的には、episode、episode_event、condition_occurrence、measurement、drug_exposure、procedure_occurrence 等を対象に、がん領域データを表現する際の使用方法、想定される院内データソース、変換ルールおよび課題を検討した。

各項目について、

A:一般的な院内システムの構造化データまたは ETL 処理により設定可能な項目

B：断片化されたデータをルール・アルゴリズムまたはマスタ整備によりマッピング可能な項目

C：非構造化データとして存在し構造化のための整備が必要な項目

に分類した。

さらに、疾患情報、治療情報、標準概念へのマッピング、マスタ管理、医療機関間の運用差異という観点から、国内実装時に発生し得る課題を抽出した。本検討では、患者個票データは扱わず、標準仕様および実装方針に関する検討として実施した。

Ｃ．研究結果

Cancer Disease Model では、がん疾患の経時的な状態変化を episode および episode_event で表現し、診断情報を condition_occurrence、診断修飾子を measurement 等により管理する。原発がんの初回診断については、院内がん登録等から病名、診断日、ステージ、TNM 分類を比較的収集しやすい一方、転移がん、再発、進行、増悪時の診断情報は診療録、放射線レポート、病理レポート等に記載されることが多く、診断日や診断修飾子の特定が不確かになりやすい。

疾患状態の判定については、OMOP CDM 上では episode により遷移を表現できるものの、転移、再発、進行、増悪の判定が明示的かつ構造化されたデータとして存在しない場合が多い。このため、どの記録または検査結果を根拠に病勢を判定するか、また使用した定義をどのように参照可能な形で管理するかが重要な課題である。

診断修飾子については、ステージ、TNM 分類、分子・バイオマーカー、グレード、病理学的特徴、術後残存腫瘍、治療後評価などが候補となる。院内が

ん登録で収集可能な項目がある一方、二回目以降の診断、分子・バイオマーカー、病理レポート由来の情報はシステムが分散し、自然文で記載されることも多いため、癌種ごとに研究上必要な修飾子とその定義を定める必要があると考えられた。

Cancer Treatment Model では、薬剤投与、手術、放射線治療、処置等を、episode、episode_event、drug_exposure 、 procedure_occurrence 、 measurement 等により管理する。特に、がん薬物療法では、単一薬剤ではなく、薬剤の組み合わせ、投与量、投与スケジュール、順序、間隔、期間をレジメンとして把握し、サイクルおよび治療ラインとして表現する必要がある。

薬剤情報については、日本で広く使用される YJ コードを OMOP CDM の standard concept へ変換する際、RxNorm 等へのマッピングが必要となる。さらに、個別薬剤をレジメンとしてまとめるためには、標準化されたレジメンマスタと変換ロジックが不可欠であり、同一レジメンのサイクル判定、episode_end_date の設定、休薬・中止時の終了日推定等についても定義とルール化が必要である。

手術、放射線治療、処置については、K コード、M コード、J コード等が院内情報システムから収集される候補となるが、これらは診療報酬の粒度に基づくため、部位、アプローチ、治療範囲等の臨床的意味を十分に表現できない場合がある。SNOMED 等の標準概念へ一意に変換することが難しいため、NCD 術式マスタ、STEM7、MEDIS マスタ、ICD-10-PCS、ICHI 等を含め、どのコード体系を基盤とするか検討を要する。

根本的課題として、研究目的に応じた収集項目のスコープと判定定義の明確化、医療機関ごとのシステム仕様・運用差異の吸収、standard concept への変換マスタとその更新管理体制の整備が挙げられた。

D．考察

OMOP Oncology Module は、がん領域に特徴的な疾患経過、診断修飾子、治療ライン、レジメン、サイクル等を構造的に表現するための有用な枠組みである。しかし、日本の実臨床データを用いて実装するためには、単に院内情報システムから取得可能なデータを OMOP CDM へ変換するだけでは不十分であり、研究プロジェクトの目的に応じて、どの項目を収集対象とし、どの定義で判定し、どのデータソースを採用するかを事前に定める必要がある。

特に、疾患状態の判断・遷移や診断・治療修飾子は、自由記載、レポート、検査結果のレビュー、画像評価等に分散して存在するため、自然言語処理等による抽出可能性を評価するとともに、不明なものを不明として扱う基準を含めた再現性のあるデータ収集手順が求められる。また、院内情報システムから自動収集できない項目については、従来型の手入力を単なる補完手段とするのではなく、研究上の再現性と品質を担保する入力テンプレートとして設計する必要がある。

さらに、標準概念へのマッピングは継続的なメンテナンスを伴う。変換元の国内コード、変換先の OMOP vocabulary、レジメンマスタ、手術・処置コード等はいずれも更新されるため、臨床医、研究者、データマネージャ、エンジニアが連携した横断的な管理体制が必要である。

今後、各種がん領域の疾患レジストリ等が Oncology Module を含む OMOP CDM の仕様に基づいて収集項目やソース、判定ロジックを定義し、国際間でデータの相互互換性が担保された、連合学習が可能なレジストリを構築していく将来像が期待される。

E．結論

がん領域データを OMOP CDM へ変換し、二次利用可能な RWD 基盤を構築するためには、OMOP Oncology Module の構造を理解したうえで、疾患情報・治療情報のスコープ、判定定義、データソース、変換ロジック、マスタ管理体制を明確化する必要がある。

今後は、国内の医療情報システムの実態を踏まえ、がん領域データを収集し OMOP CDM 化するためのデータマネジメント指針、標準マスタおよび変換マスタの管理方針、自然言語処理や入力テンプレートを含む実装手順を、臨床・研究・データマネジメントの観点から継続的に検討する必要がある。

F．健康危険情報

なし

G．研究発表

1. 論文発表

なし

2. 学会発表

なし

H．知的財産権の出願・登録状況

（予定を含む。）

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし