

電子カルテデータを臨床研究や医薬品開発に活用するための取り組みと課題

武田理宏 (大阪大学大学院医学系研究科 医療情報学)
杉本賢人 (大阪大学大学院医学系研究科 医療情報学)
坂井亜紀子 (大阪大学医学部附属病院 医療情報部)
朝田委津子 (名古屋大学医学部附属病院 メディカル IT センター)
吉田直樹 (大阪大学医学部附属病院)
山本征司 (大阪大学医学部附属病院 医療情報部)
真鍋史朗 (大阪大学大学院医学系研究科 医療情報学)
白鳥義宗 (名古屋大学医学部附属病院 メディカル IT センター)
松村泰志 (国立病院機構大阪医療センター)

Initiatives and Identified Issues for Using Electronic Medical Record Data for Clinical Research and Drug Development

Toshihiro Takeda
Department of Medical Informatics, Osaka University Graduate School of Medicine
Kento Sugimoto
Department of Medical Informatics, Osaka University Graduate School of Medicine
Akiko Sakai
Department of Medical Informatics, Osaka University Hospital
Itsuko Asada
Nagoya University Hospital Medical IT Center
Naoki Yoshida
Department of Pharmacy, Osaka University Hospital
Masashi Yamamoto
Department of Medical Informatics, Osaka University Hospital
Shirou Manabe
Department of Medical Informatics, Osaka University Graduate School of Medicine
Yoshimune Shiratori
Nagoya University Hospital Medical IT Center
Yasushi Matsumura
National Hospital Organization Osaka National Hospital

【要旨】 医療機関の Real World Data (RWD) として、レセプト・DPC データ、電子カルテデータ、レジストリデータが挙げられ、これらの臨床研究や医薬品開発に活用するニーズが高まっている。このためには、RWD の標準化、品質確保と利活用できるデータ範囲の拡大が必要となる。

レセプト・DPC データは標準化、品質確保が行われたデータであるが、結果データに欠ける。電子カルテデータは各医療機関においてローカルコードで運用され標準化や品質管理が課題となる。

我々は大阪臨床研究ネットワーク（OCR ネット）での臨床研究支援の取り組みや臨床研究中核病院による Real World Evidence 創出の取り組み（臨中ネット）を通じ、RWD を利活用できる仕組みの構築に取り組んでいる。本稿では特に標準化の観点での取り組みと、これらの取り組みを通じて見えてきた課題を提示する。

【キーワード】 Real World Data 電子カルテ 標準化 品質管理

【Abstract】 Real World Data (RWD) stored in medical institutions include claims and DPC data, electronic medical record (EMR) data, and registry data. There is a growing need to utilize this RWD for clinical research and drug development. To achieve this goal, it is necessary to standardize and ensure the quality of RWD and expand the range of data that can be used. While claims and DPC data are standardized and quality-assured, they lack outcome data. Since EMR data is managed using local codes in each medical institution, standardization and quality control have become issues. We are working to establish a system to utilize RWD through the Osaka Clinical Research Network (OCR-Net) to support clinical research and the Real World Evidence Initiative by Clinical Research Core Hospitals in Japan (Rincho-Net). In this paper, we introduce our efforts, particularly focusing on standardization, and present the issues that have been identified through these efforts.

【Keywords】 Real World Data electronic medical record standardization quality control

1. はじめに

保険医療福祉情報システム工業会（JAHIS）による 2022 年度調査では、病院のオーダエントリシステム導入率は 53.6%、電子カルテ導入率は 49.7%、400 床以上の病院では、オーダエントリシステム導入率は 88.4%、電子カルテ導入率は 86.5% となっている。電子カルテには日々の診療情報が蓄積されるため、電子カルテに蓄積されるリアルワールドデータ（RWD）の臨床研究への利活用が期待される。

医薬品の承認申請を目的とした臨床治験は莫大な費用と時間を要する。一方、厳密な参加基準に基づき治験が実施されるため、多様な背景を持つ患者に投与を行った場合、治験で見られなかった有害事象が発生する可能性があるため、市販後調査が必要となる。このため、RWD を活用した医薬品の承認申請や市販後のエビデンスづくりが期待されている。海外では、米国食品医薬品局（FDA）から Framework for FDA's Real-World Evidence Program、欧州医薬品庁（EMA）から Use of patient disease registries for regulatory purpose — methodological and operational considerations が発出されるなど、RWD を活用できる環境が整えられてきた。

RWD を臨床研究や薬剤承認に用いる場合、対象の医療情報データベース（DB）が原資料となる電子カルテ DB との整合性が保たれていることの検証ができていなければならない。また、多施設の RWD 利

活用を考えるとデータ標準化が必要となる。さらに、品質担保、標準化されたデータ範囲が、臨床研究や薬剤承認に必要なデータをカバーできているかの課題がある。

日本では、2011 年度より独立行政法人医薬品医療機器総合機構（PMDA）と協力医療機関により医療情報データベース基盤整備事業（MID-NET プロジェクト）が開始された。本事業では協力医療機関が保有する電子カルテやレセプト（保険診療請求明細書）の電子情報をデータベース化して、それらを解析することができる。MID-NET が整備する医療情報データベースは協力医療機関の電子カルテデータを標準化し、品質を担保したものとなっている。

MID-NET は PMDA がその管理運営を行っている。医療機関のみで同等のことを行うことを考えた場合、どの医療機関が各医療機関の標準化や品質保証を担保するかなどの課題がある。本稿では RWD 利活用に向けた課題、大阪臨床研究ネットワーク（OCR ネット）や臨床研究中核病院による Real World Evidence 創出のための取組み（通称、臨中ネット）の取り組みを紹介する。

2. RWD のデータソースと利活用に向けた課題

RWD のデータソースとしてレセプト・DPC データ、電子カルテデータ、レジストリデータ、患者報告データ（PRO: Patient Reported Outcome）、健康

状態が把握できる IoT (Internet-of Things) データが想定される。臨床研究においては PRO や IoT データは医療機関で蓄積されるレセプトデータ、電子カルテデータ、レジストリデータと結合して解析をされることが多いと想定される。

2.1. レセプト・DPC データ

医事会計システムは電子カルテ基幹システム、部門システムからの実施情報を受け、診療報酬請求に必要な医事会計情報を管理するシステムである。医事会計システムからは毎月、レセプトが発行される。レセプトは、入院、外来別に作成され、患者の傷病名、診療行為、処方された薬、実施された処置・手術、使用された特殊材料、実施された検査項目など情報が記載される。2010 年度からはレセプトを電子媒体あるいはオンラインで提出するレセプト電算処理が原則義務化された。レセプト電算処理では、診療行為コード、傷病名コード、医薬品コードを標準コードであるレセ電算コードで表記する。

DPC (Diagnosis Procedure Combination) 導入病院では、最も医療資源を投入した病名、その治療に影響する副傷病名、手術や処置の内容、患者の状態や重症度をもとに医療費が算定される。これらの病院では診療録情報を記載する様式 1、診療明細情報を記載する E ファイル、行為明細情報を記載する F ファイル (併せて EF 統合ファイル) が作成される。様式 1 には患者情報 (生年月日、性別、住所地域の郵便番号)、入院情報 (入院年月日、入院経路、救急搬送の有無等)、退院情報 (退院年月日、退院時転帰、在宅医療の有無等)、診断情報 (傷病名、ICD-10 コード等)、手術情報 (K コード、麻酔方法、手術名等)、その他診療情報 (褥瘡の有無、ADL スコア、がんの TNM 分類、JCS、肺炎の重症度等) が記述される。EF 統合ファイルには、医科点数表に基づく出来高による診療報酬の算定情報として、一連の診療行為や一連の診療行為で使用された医薬品等の名称や使用量が記述される。

レセプト、DPC は診療報酬請求に用いることが目的であるため、実施データはあるが結果データを持たない。また、レセプトの傷病名は診療報酬請求を目的に登録されるため、その患者が本当に罹患して

いる傷病名を把握することが難しい。DPC では、入院に治して、主傷病名、入院の契機となった傷病名、医療資源を最も投入した傷病名、入院時併存症 (4 疾患以内)、入院後発症疾患 (4 疾患以内) が登録される。このため、傷病名は正しく取得できるが、その対象は DPC 導入病院、入院患者に限定される問題がある。

レセプトデータ、DPC データは日本で最も標準化されたデータであり、診療報酬請求に用いている点で品質保証がされていると捉えることができるため、多施設でデータ解析を行うことが容易である。レセプト、DPC データの特徴を正しく捉えることができれば、多施設で行う臨床研究等での良いデータとなる。厚生労働省保険局では 2009 年度より、電子レセプト及び特定健診データの突合可能な匿名化によるデータの収集を開始し、レセプトデータベース (NDB) として都道府県や研究者の使用が可能となっている。一方、結果データと合わせて解析を行う必要がある場合、NDB は利用できないため、各施設のレセプト、DPC データを電子カルテデータと突合して、データ出力を行う必要がある。

2.2. 電子カルテデータ

病院情報システムは、病院の業務を支援するシステムの総称である。病院情報システムは主に、基幹システムと部門システムで構成される。基幹システムは主に病院内の部門間で情報伝達するオーダエントリシステムと診療情報を記録、保管するシステムで構成される。

病院情報システムデータの二次利用に向けて業務用に使うデータベースと分析用のデータベースを別にし、業務用データベースで発生するデータは定期的に分析用データベースに移行される (データウェアハウス、DWH)。厚生労働省電子的診療情報交換推進指針 (SS-MIX, Standard Structured-Medical Information eXchange) に従って構築した SS-MIX2 ストレージに電子カルテデータを出力し、SS-MIX2 ストレージからデータ出力したデータを二次利用している医療機関もある。

2.2.1. オーダエントリシステム

オーダエントリシステムは、検体検査オーダ、画

像検査オーダー、病理検査オーダー、処方・注射オーダーなど、医師のオーダー情報を各部門システムに伝達するシステムである。各部門システムはオーダー情報に基づいた実施情報、結果情報を基幹システムに伝達するとともに、実施行為情報を医事会計システムに伝達する。基幹システムに伝達された実施情報、結果情報は診療情報として基幹システムに保存される。オーダーエントリシステムで管理されるオーダー、結果情報は構造化され、DWHに蓄積することで二次利用が可能である。一方のオーダー項目は各医療機関独自のハウスコードで管理されることが多く、多施設でデータの二次利用を行う場合、コードの標準化や単位の標準化を行うことが課題となる。

2.2.2. 診療情報を記録、保管するシステム

診療情報を記録、保管するシステムは主に、患者氏名、生年月日、性別、血液型、禁忌・アレルギー情報、障害情報、生活習慣情報などを管理する患者基本情報（患者プロフィール）、医療者の日々の記録を管理するプログレスノートで構築される。プログレスノートには二次利用が期待される情報が記載されるが、フリーテキストで記載されることが多く、二次利用は容易ではない。身長や体重、血圧は入院時に看護師が記録する場合は、構造化データとして記録されることが多く、利活用が可能である。一方、外来では、患者が問診で手書きする場合や、医師がプログラスノートに記載する場合など二次利用困難なケースが想定される。このように、入院患者のみ二次利用が可能なデータも少なくない。

プログレスノートに記載される診療情報を構造化データとして取得する方法として、あらかじめ取得したいデータを登録するテンプレートを設計し、医療者がそのテンプレートを用いてプログレスノートに記事記載する方法がある。テンプレートを用いてデータを取得できるのは、前向きデータ収集に限定される。また、各医療機関で独自に設計されたテンプレートでは、データの標準化は困難であり、多施設での二次利用は容易ではない。診療録直結型全国糖尿病データベース事業（J-DREAMS）では、糖尿病標準診療テンプレートと呼ばれる入力画面を電子カルテに導入し、各医師が糖尿病患者を診察する際に診療録の一部として入力し、SS-MIX2拡張スト

レージにデータを出すことでデータ収集を行っている。各専門領域の学会が主導し、疾患ごとにデータ収集項目を定め、各施設でテンプレートを用いてデータ収集する事例は、今後、増えていくと考えられる。

2.2.3. 部門システム

部門システムには紹介状、診断書、退院時要約（退院サマリ）など医療文書を作成する文書作成システム、他院からの紹介状や患者問診票、同意書など紙で発生する医療文書を管理するスキャンシステム、病理システム、重症病棟で使用するシステム、眼科など各診療科独自で使用するシステムなど、多くのシステムがある。部門システムには構造化データが蓄積されるものもあるが、費用面などの問題で部門システムのデータをDWHに移行していないことも多い。このような場合、メーカのSE作業で有償でないとデータ出力できない場合がある。

退院時要約や診療情報提供書は厚生労働省が進める医療DXの3文書6情報のうちの2文書であり、HL7 FHIR記述仕様が公表され、令和4年度の診療録管理体制加算では同文書のHL7FHIRへの対応予定の定例報告を求めるなど、標準化に向けた取り組みが進んでいる。

循環器領域で二次利用が期待される心電図、心臓超音波、心臓カテーテルの結果は部門システムに格納されることが多い。日本循環器学会はJCSデータ出力標準フォーマットガイドライン（SEAMAT: Standard Export data format）を定め、SS-MIX2拡張ストレージにデータを出すことでデータ収集を行っている。

2.3. レジストリデータ

レジストリとは、特定の疾患、疾患群、治療等の医療情報の収集を目的として構築したデータベースのことで、あらかじめデータ収集項目を決め、症例報告書を用い、前向きにデータ収集を行う。近年ではWeb方式のシステムでデータを収集するEDC（Electronic Data Capture）が主流となっている。EDCのセンターサーバは、通常はインターネット上にあり、海外にある場合もある。通常、電子カルテシステムはインターネットと繋がっておらず、電

子カルテとは完全に切り離されたシステムとなる。このため、医師やCRC（Clinical Research Coordinator）が電子カルテ画面を見ながら、インターネットに接続した端末からEDCにデータを転記する必要がある。データ転記は医師の労力やCRCの人件費が必要となるだけでなく、データ転記が発生することで転記ミスが生じる可能性がある。このため、データの信頼性を確保するため、電子カルテデータとEDC登録データを比較するSDV（Source Data Validation）が必要となる。また、EDCに登録したデータは医療機関には残らないため、医療機関独自の解析が行うことができない可能性がある。

3. 医療データの標準化

オーダエントリシステムは1990年代に多くの医療機関で導入された。この頃、多くの医療機関で医療データの標準化に対する考えは持っておらず、各医療機関がそれぞれのローカルコードを使用し、その後もローカルコードを維持管理していた。医療機関で電子カルテの安定稼働を第一に考えるため、医療機関がローカルコードを標準コードに置き換えてコードを管理することは大変難しい。

厚生労働省電子的診療情報交換推進指針であるSS-MIXでは、傷病名は国際疾病分類（ICD-10）、処方情報にはHOTコード、検体検査結果にはJLAC10コードの使用が推奨されている。しかし、SS-MIX2ストレージを設置する医療機関においても、標準コードを出力していない医療機関は少なくない。傷病名はDPCの傷病名登録がICD-10病名を使用することが義務付けられているため、多くの病院の病院情報システムでICD-10コードが使用されている。薬剤コードは物流で発展したコード体系であるJANコード、診療報酬計算で発展したレセプト電算処理用コード、治療薬選択支援を目的とした薬価基準収載医薬品コード、YJコードなど、様々なコードが用いられている。HOTコードはMEDIS-DCが整理した13桁の管理番号で、既存の4つの汎用コードとの対応表を持っているのが特徴である。JLAC10コードは日本臨床検査医学会が制定した臨床検査コードで、分析物コード5桁、識別コー

ド4桁、材料コード3桁、測定法コード3桁、結果識別コード2桁の17桁にて生成される。JLAC10コードは、相当の知識がないと付番ができない、付番者によって付番結果がばらつき、唯一性が確保できない、単位がないため二次利用によるデータ分析が不便である、といった問題点が指摘されていた。これらの問題点に対応するためJLAC11コードが提案された。JLAC11は、測定物コード（5桁）、識別コード（4桁）、材料コード（3桁）、測定法コード（3桁）、結果単位コード（2桁）の5つの要素区分でコーディングされる。

臨床研究中核病院では、医療技術実用化総合促進事業（AMED）として、臨床研究中核病院における臨床研究の質を向上、臨床研究のさらなる安全性の確保を図る、病院情報システム内の医療情報データの標準化を図る、そのデータを研究等にも活用できる体制整備を行うことを目的として、Real World Evidence 創出のための取組み（通称、臨中ネット）を実施している。この取組みの中で、2020年に臨床研究中核病院12施設に対して、標準コード出力の可否について、アンケート調査を行った。診療科については、全施設がレセプト電算診療科コード、DPC様式1診療科コード、10施設がSS-MIX2統一診療科コードでの出力が可能であった。病名については全施設がICD10対応標準病名マスタ（MEDIS標準病名マスタ）、レセプト電算処理システム基本マスタでの出力が可能であった。薬剤についてはHOT9が8施設、厚生省コードが4施設、YJコードが全施設で出力可能であった。薬剤単位は3施設がMERIT9で出力が可能であった。検体検査結果についてはJLAC10コードに対応しているのが9施設だったが、全範囲にコーディングしているのは3施設で、必要性に応じてコーディングをする項目を決定している施設が多かった。

このように、臨床研究中核病院でもすべての医療機関で標準コードの出力はできない。このため、RWDの利活用に向けて標準コードを使用するには、病院情報システムのマスタ管理をローカルコードと標準コードの双方を用いるか、ローカルコードと標準コードの対応法を用意し、データ出力するどこかのタイミングでローカルコードを標準コードに変換

する方法が想定される。

4. 大阪臨床研究ネットワーク（OCR ネット）での臨床研究支援

OCR ネット（大阪臨床研究ネットワーク）とは、適正かつ効率的に多施設共同臨床研究を実施する体制を整備し、質の高い治験及び臨床研究の推進を図ることを目的に治験及び臨床研究に関する人材育成、教育研修、実施環境整備等を協力して実施する関西圏を中心としたネットワークで、大阪府下を中心に27の医療機関と協定をかわしている。筆者らは臨床研究実施支援、臨床研究インフラ整備を担当し、各病院の電子カルテシステムと阪大病院データセンターをセキュアなネットワークで接続し、レセプト、DPC、電子カルテデータ二次利用の仕組み（多施設共通データウェアハウス）、電子カルテと連携したレジストリデータ収集の仕組み（臨床データ収集システム、CDCS: Clinical Data Collection System）、臨床画像収集の仕組み（臨床画像収集システム）を構築している。

4.1. 多施設共通データウェアハウス

OCR-netに参加する6施設に共通のデータベース構造を持つ多施設共通DWHを構築した。大阪大学医学部附属病院の除く5施設は大阪府内の公的医療機関であり、医療情報担当部門を持たない医療機関も含まれる。このため、できるだけ各医療機関の手間をかけず、標準化されたデータを蓄積し、データ検索を行うことができる必要がある。

そこで、標準化と品質管理ができていないレセプト・DPCデータと、レセプト・DPCデータに欠ける結

果データを電子カルテDWHから蓄積する仕組みとした。多施設共通DWHには、レセプト・DPC情報として、①病名、②処方、③注射、④放射線治療、⑤手術・処置、⑥医療行為、⑦入院患者ファイル、⑧DPC様式1病名ファイル、結果データとして電子カルテDWHより①検体検査結果と②画像レポートが登録されている。レセプト・DPCデータは各医療機関で月次で生成されるファイルを人手により多施設共通DWHにアップロードを行う。検体検査結果は電子カルテデータベースにハウスコードで管理されるため、標準コード、標準単位に変換（例：赤血球数： $4.66 \times 10^6/\mu\text{L} \rightarrow 466 \times 10^4/\mu\text{L}$ ）を行う変換マスタを整備し、標準化したデータを多施設共通DWHに登録する仕組みを構築した。標準コードに変換する検体検査項目は、臨床研究に用いると想定される項目に限定し、各施設の協力を得ながら大阪大学医学部附属病院が整備を行った。

多施設共通DWHは共通のデータベース構造を持つため、共通のデータベース検索文でデータ抽出を行うことが可能である。データベース検索文はデータセンターである大阪大学医学部附属病院で作成し、自施設の多施設共通DWHで正しくデータ抽出できることを確認したのち、セキュアなネットワークを用いて、各施設の多施設共通DWHに配信を行う。配信を受けた各施設の多施設共通DWHはデータ抽出を行う。データ抽出結果は各施設の担当者がその内容を確認した上で、承認作業の後、データセンターに送信される。

利活用できるデータを広げるために、我々はフリーテキストで記載される画像診断レポートから深層学習を用いた自然言語処理技術を用いて情報抽出する取り組みを行っている。情報抽出は、固

図1 多施設共通データウェアハウス

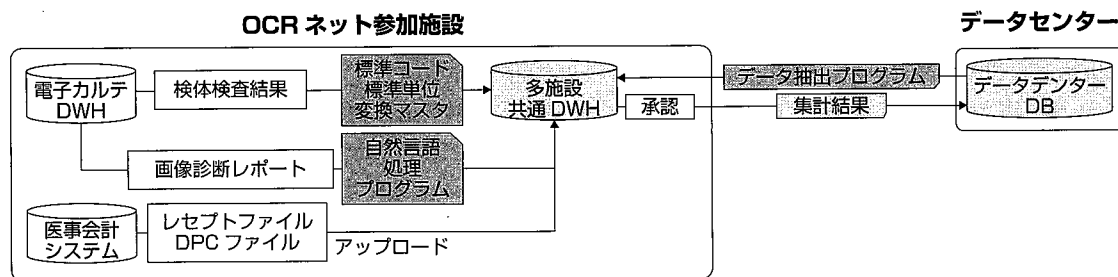
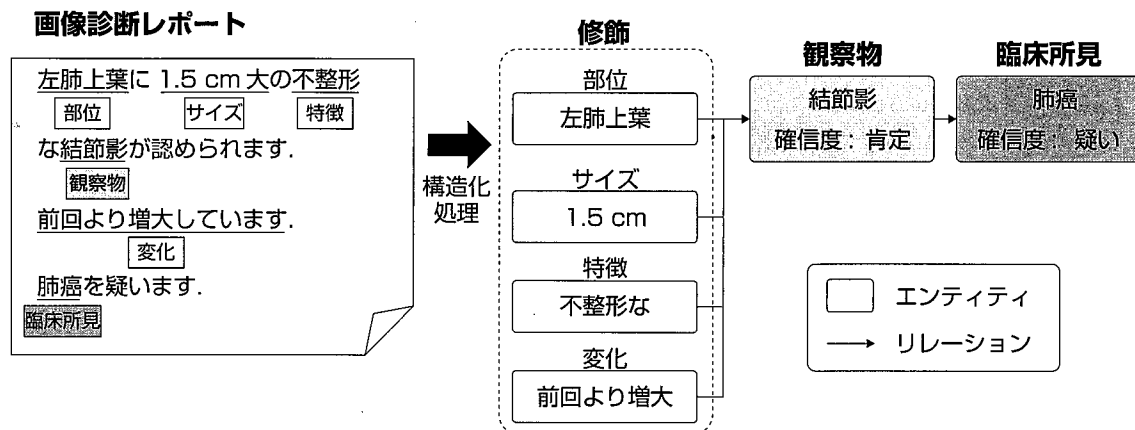


図2 画像診断レポートの情報抽出モデルの実例



有表現抽出、確信度判定、関係抽出の3つのタスクからなる。「上肺野に1.5 cm大の不整形な結節影が認められます。前回より増大しています。肺癌を疑います。」というレポートに対して、上肺野：部位、1.5 cm大：サイズ、不整形な：特徴、結節影：観察物、前回より増大しています：変化、肺癌：診断と言った形で固有表現抽出を行う。確信度判定では、観察物、診断に対する確信度を、肯定から否定の5段階で評価する。本例では、結節影は「認められる」であるので肯定、肺癌は「疑います」であるので疑いと判定される。最後に抽出された観察物の固有表現に対して、それぞれの固有表現との関係の有無を判定する。以上により、図2のような情報抽出モデルを構築することができる。画像診断レポート情報を構造化することで、臨床研究等で使用できるデータの幅が広がる。レセプトに登録される病名は、いわゆる保険病名の理由で正確でない可能性がある。肺癌の患者抽出は病名で行うよりも、画像診断レポートで抽出する方が正確かもしれない。また、社会問題となっている画像診断レポートの見落としに対し、重要所見を判定して注意喚起を行うようなシステムを構築することが期待される。

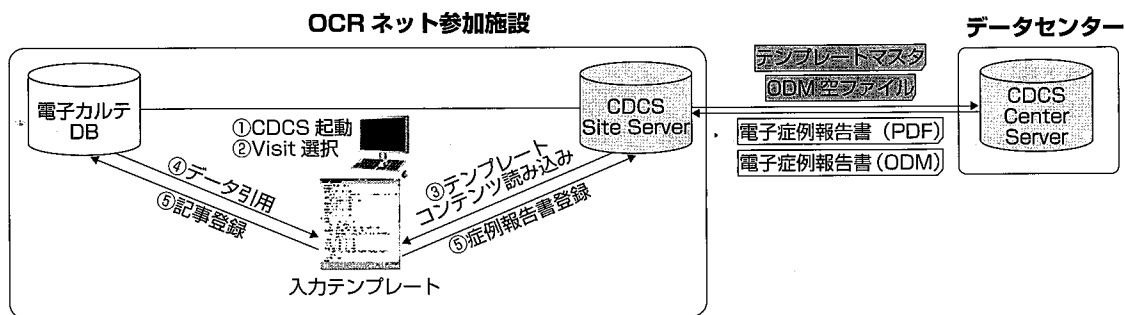
4.2. 臨床データ収集システム (CDCS: Clinical Data Collection System)

前述した通り、電子カルテとは別のネットワークで登録されるEDCには、電子カルテからデータ転

記が必要(転記の労力、転記ミス)、データが医療機関に残らないといった問題点が発生する。

この問題を解決するためOCRネットのセキュアなネットワークを活用し、電子カルテにEDCを組み込み、診療記録を記載しながら、電子症例報告書生成する臨床データ収集システム(CDCS: Clinical Data Collection System)を構築した(図3)。CDCSはOCRネットに参加する19施設に導入されている。CDCSは、患者番号を被検者番号に置き換えた上で、電子症例報告書を国際標準規格であるCDISC (Clinical Data Interchange Standards Consortium) のODM (Operational Data Model)でデータセンターに出力する。レジストリ情報を入力するテンプレートはデータセンターが作成し、各医療機関にテンプレートマスタとODMの空ファイルを送信する。各医療機関で医師またはCRC (Clinical Research Coordinator)は、電子カルテ上からCDCSを起動し、対象となる臨床研究を選択する。臨床研究を選択すると各visitのテンプレート一覧が表示されるので、Visitを選択して、データセンターから配布されたテンプレートを起動する。医師がテンプレートを入力すると、入力内容は電子カルテの経過記録に出力され、診療記録となる。テンプレート入力内容はEDCとして、データセンターに送信することで、データ登録が完了する。テンプレートは検体検査結果や処方オーダを引用することが可能となっている。このことは、医師の入力支援だけでなく、転記ミスをなくすることが可能となる。CDCSはEDCと診療記録

図3 臨床データ収集システム (CDCS: Clinical Data Collection System)



を同時に作成するため、臨床研究の根拠となったデータを確認する SDV (Source Data Verification) が簡便になる利点がある。

4.3. 臨床画像収集システム

臨床研究では、中央解析や AI 研究への利用を目的に臨床画像をデータセンターに集めるニーズがある。これに対して、各研究協力医療機関は、患者カルテで画像を選択し、匿名化処理を行った上で DVD/CD 出力を行い、被検者番号と匿名化番号の対応表を作成し、データセンターに郵送するのが通常である。

OCR ネットに参加する 15 施設に、画像収集ゲートウェイを設置した。画像収集ゲートウェイは各施設の画像サーバ (PACS) に Q/R (Query and Retrieve) 接続で目的の画像を収集し、患者情報を匿名化して、データセンターに送信する仕組みを持つ。

画像収集ゲートウェイは CDCS と連携を行っている。臨床研究ごとに収集する臨床画像のモダリティ、部位をあらかじめマスタ登録を行う。医師や CRC が CDCS テンプレートに、データセンターに送信する臨床画像の撮影日を入力すると、臨床画像のモダリティ、部位、撮影日と CDCS で発番する被検者番号は画像収集ゲートウェイに連携される。CDCS から与えられた情報で収集した臨床画像は、患者 ID を CDCS で発番した被検者番号に置き換え、データセンターに送信される。このことにより、データセンターは CDCS で収集した臨床情報と臨床画像収集システムで収集した臨床画像を、同一の被検者番号で管理することが可能となる。

臨床画像収集システムは、多施設共通 DWH から

収集した画像リストに基づき、臨床画像を一括で収集し、指定した被検者番号に匿名化してデータ収集することが可能である。多施設共通 DWH と臨床画像収集システムと組み合わせることで、構造化した画像診断レポート情報をアノテーションデータとして臨床画像 AI 研究を行うことが可能となる。

5. 臨中ネットでの標準化の取り組み

臨中ネットでは、臨床研究中核病院による Real World Evidence 創出を目的に、6 つのサブワーキング (SWG) で活動を行っている。SWG1 はデータ内容検討、SWG2 は人材育成、SWG3 は標準コード管理、SWG4 はバリデーション／品質管理、SWG5 はデータ出力、SWG6 はユースケース選定と臨床兼研究の実装の議論を行っている。大阪大学医学部附属病院は名古屋大学医学部附属病院とともに、SWG3 は標準コード管理の取りまとめを行っており、その取り組みを紹介する。

臨中ネットでは SS-MIX2 標準化ストレージ (Ver. 1.2f) に蓄積されるデータ対象にバリデーションを行い、データ出力を行うことを原則とするが、同等のデータを各医療機関の DWH からデータ出力することも認めている。前述の通り、全ての医療機関で標準コードを SS-MIX2 標準化ストレージに出力しているわけではない。そこで、医療機関ごとにローカルコードと標準コードの対応表を作成し、維持管理する方針となった。各診療科の標準コードの付与状況を調査し、全施設が標準コードを付与している診療科コード、病名コード、薬剤コードは対応表作成の対象外とした。

5.1. 検体検査結果の標準化

検体検査結果コードは、JLAC10は付番者によって付番結果がばらつき、唯一性が確保できない、2次利用によるデータ分析で必要な単位が統一されないとの理由でJLAC11を基本とし、測定物コードなどJLAC11が未採番の項目にはJLAC10コードを補完する方針とした。

臨床ネットでは、臨床研究に必要と考えられる検体検査項目に限定して標準コードを付与することとした。対象項目は国立大学病院遠隔バックアップシステム(The GEMINI Project)、J-CKD、J-DREAM、CTCAEより178項目を抽出した。今後、臨床研究で必要な検査項目が申請された場合は、標準コード対応表を更新する方針とした。

全ての施設にJLACコードに精通した担当者がいない等の理由により、対象となる検体検査項目に標準コードを中央で付与し、これにハウスコードをマッピングする方針とした。JLAC11コードと補完するJLAC10コードは先行するMID-Netプロジェクトと整合を取るため、九州大学病院内に設置するガバナンスセンターにコーディングを依頼した。JLAC11の測定法コードは検査機器や試薬により異なりため、「000: 特定しない」とした。JLAC10コードの測定物コードは「998: 方法問わず」とした。

各施設の結果値を標準結果単位への値に変換するために、単位変換係数を設定した。単位が同じ場合や「+」「-」、単位なしなどの場合、変換係数として「1」を記入した。単位が対象検査項目のJLAC11の結果単位と異なる場合は単位変換係数を記入した(ハウスコードの単位が「ng/dL」、JLAC11の結果単位が「ng/mL」の場合、単位変換係数は「/100」、 $1\text{ ng/dL} = 0.01\text{ ng/mL} = 1 \times 1/100\text{ ng/mL}$)。

対応表にはハウスコード、世代番号または有効開始日、施設での名称、材料、単位、単位変換係数、正常値、自施設で付番したJLAC10コードを記入した。

5.2. 内服薬の標準化

薬剤コードは全ての医療機関でYJコードが付番されていたため、このコードを標準コードとして使用した。臨床研究では薬剤使用量は必要と考えられる。そこで、薬剤単位はMERIT9コードを標準コードと

し、各施設のハウスコードとマッピングを行った。MERIT9コードに対応できない薬剤単位が複数の施設で確認された。他の単位に変換可能なものがほとんどであったが、バック(イミグランキット皮下注3mg、YJコード: 2160402G1026)やドーズ(ソリクア配合注ソロスター、YJコード: 3969501G1023)は新しい単位としてMERIT9コードに追加が必要と考えられた。

臨床研究では薬剤投与量が必要となる。1日投与量と投与日数をかけることで、薬剤投与量は求めることができる。一方、1日おき、1週間に1回といった服薬を行う場合、投与日数のみでは薬剤投与量は求めることができない。薬剤用法は、日本医療情報学会が定める処方・注射オーダ標準用法規格を用いることとしたが、標準用法は用法コードと補足用法コードで定義され、不規則投与は補足用法で記述される。参加14施設のうちハウスコードではあるが用法コードを出力していたのが13施設、補足用法コードを出力していたのは8施設であった。しかし、コード化できている補足用法にはばらつきがみられ、「2日に1回」(または「1日おき」)のような高頻度の補助用法はコード化率が高く、「週に2回」、「〇月〇日に内服」といった複雑なものはコード化率が低いことが分かった。このことから、各医療機関のコード化の状況から、不規則投与を行う薬剤の薬剤投与量を正確に把握することは困難な医療機関が多く存在することが明らかとなった。

標準用法は、JAMI標準用法規格の標準用法コード全表を利用して採番を行った。8施設で標準化が可能であった用法の割合は、12.4%から68.3%と大きくばらついた。標準用法コードが付与できなかった用法の内、コーディングルールでコード付与は可能だが、全表に含まれなかったものとして、時間指定の用法、「1日2回夕食後2時間と就寝前」といった特殊な服薬方法が挙げられた。例えば、「1日2回8時、20時」は全表に含まれるが、「1日2回9時、21時」は全表に含まれないが、当該施設の1年間の使用実績では「1日2回9時、21時」の使用頻度が高かった。コーディングルールでコード付与できなかったものは、「毎食事中」や「1日1回医師の指示通り」といったものが挙げられた。また、頓服

では「喉が痛い時」が「喉がゴロゴロする時」で良いか、読み替えが困難だったケースや、「疼痛時、発熱時」など複数の頓服条件が記載されるケースが認められた。標準用法コードは、日本病院薬剤師会、日本薬剤師会による内服薬、外用薬に関する「標準用法用語集」に収載されている用法と用語だけを対象としている。このため、例示したとおり、コーディングルールでコード付与できるものであっても、標準コードが付与できないケースが存在する。各医療機関で「標準用法用語集」に含まれない用法も処方実績が確認されたため、「標準用法用語集」の追加を日本病院薬剤師会、日本薬剤師会に依頼する必要があると考えられた。

6. まとめ

OCR ネットに参加する医療機関は大阪府下を中心とする医療機関で、医療情報部門を設置する医療機関もあれば、設置しない医療機関もある。標準化や品質管理について、十分な知識のある医療機関はほとんどない。このため、過大な標準化や品質管理を行うことは困難であるため、標準コードが整備されたレセプト・DPC データを中心に多施設共通 DWH を構築した。レセプト・DPC データは診療報酬請求を目的に確認されたデータで、品質管理ができていえることができる。レセプト・DPC データは結果データを持たない欠点がある。そこで、検体検査結果のうち、臨床研究に使用すると想定される項目に限ってコードの標準化と標準単位の変換を行った。標準化の作業は大阪大学医学部附属病院で取りまとめて実施している。このように、単施設がすべての医療機関の標準化を担当することで臨床研究に使用できるレベルの標準化は実現できるが、この DWH を、OCR ネット参加施設を超える医療機関の医療データと共に利活用するには、検討しなければならない事項が残っている。

臨中ネットでの取り組みでは、臨床研究中核病院に標準化担当者と品質保証担当者の人材育成を目指しており、標準コードの付与は各医療機関で実施している。標準コードにはコーディングのルールが作られているが、各医療機関で解釈などの相違が生じ

るため、中央でコーディングをコントロールする必要であった。臨中ネットは各施設が中央に組織を持たないため、今後、各施設で標準コード変換表の更新を行い、適切に維持することが課題となってくる。

薬剤用法の標準コードへの変換では、特に補足用法がローカルコードでもコーディングは行われていない医療機関を認めた。また、ローカルコードと標準コードをマッピングすることができないケースを認めた。これは、医療機関側が適切な用法となっていないケースと、適切なコードと考えらえるが対応する標準コードが定義されていないケースが認められた。今後は、国が進める電子処方箋、全国医療情報プラットフォーム、民間が進める Personal Health Record などに情報を提供する必要があり、このためには自施設のローカルコードの見直しも考える必要が出てくると予想される。また、各医療機関の利用実績に基づき標準コードを管理する団体にコードの追加を要望し、標準コードの整備を進めていく仕組みも構築していく必要がある。

今後、日本全国の医療機関で適切な標準コードが付与されるためには、標準コードコーディングの問い合わせに答え、必要であれば標準コードの更新を標準コード管理団体に提言するような、中央組織が必要ではないかと考える。

【謝辞】 臨中ネット事業はAMED 医療技術実用化総合促進事業「Real World Evidence 創出のための取組み」、OCR ネット事業は戦略的イノベーション創造プログラム (SIP) 第二期「AI (人工知能) ホスピタルによる高度診断・治療システム」、研究開発と Society 5.0 との橋渡しプログラム (BRIDGE) 「AI ホスピタルを実装化するための医療 AI プラットフォームの構築に必要な技術に関する研究開発」の研究支援を受けております。

臨中ネット事業でご指導を下された臨床研究中核病院の担当者様、OCR ネット事業にご協力いただいている OCR ネット参加医療機関の担当者様に感謝を申し上げます。

参考文献

- 高橋史峰 (2015) 「電子診療情報データベース (MID-NET) プロジェクトの現状と課題」『レギュラトリーサイエンス学会誌』5 巻 3 号, pp. 235-243.
- 藤森研司 (2016) 「レセプトデータベース (NDB) の現状とその活用に対する課題」『医療と社会』26 巻 1 号, pp. 15-24.
- Chen, Y., Matsumura, Y., Nakagawa, K., Ji, S., Nakano, H., Teratani, T., Zhang, Q., Mineno, T., Takeda, H. (2007) "Analysis of yearly variations in drug expenditure for one patient using data warehouse in a hospital." *J Med Syst*, Vol. 31,

- No. 1, pp. 17-24.
- 稲岡則子・紀ノ定保臣・宇都由美子・石原謙・伏見清秀 (2007) 「データウェアハウスとデータ活用」『医療情報学』27 巻 3 号, pp. 261-268.
- Kimura, M., Nakayasu, K., Ohshima, Y., Fujita, N., Nakashima, N., Jozaki, H., Numano, T., Shimizu, T., Shimomura, M., Sasaki, F., Fujiki, T., Nakashima, T., Toyoda, K., Hoshi, H., Sakusabe, T., Naito, Y., Kawaguchi, K., Watanabe, H., Tani, S. (2011) "SS-MIX: a ministry project to promote standardized healthcare information exchange," *Methods Inf Med*, Vol. 50, No. 2, pp. 131-139. doi: 10.3414/ME10-01-0015.
- Manabe, S., Takeda, T., Hattori, A., Yamamoto, M., Shimai, Y., Namiuchi, Y., Yamaguchi, J., Yamada, T., Konishi, S., Matsumura, Y. (2021) "Practical use of a multicenter clinical research support system connected to electronic medical records," *Comput Methods Programs Biomed*, Vol. 210, pp. 106362. doi: 10.1016/j.cmpb.2021.106362.
- Takeda, T., Manabe, S., Hattori, A., Yamaguchi, J., Konishi, S., Yamamoto, Y., Takahashi, D., Matsumura, Y. (2020) "An Automatic Image Collection System for Multicenter Clinical Studies," *Stud Health Technol Inform*, Vol. 270, pp. 23-27. doi: 10.3233/SHTI200115.
- Sugimoto, K., Takeda, T., Oh, J.H., Wada, S., Konishi, S., Yamahata, A., Manabe, S., Tomiyama, N., Matsunaga, T., Nakanishi, K., Matsumura, Y. (2021) "Extracting clinical terms from radiology reports with deep learning," *J Biomed Inform*, Vol. 116, pp. 103729. doi: 10.1016/j.jbi.2021.103729.
- Matsumura, Y., Hattori, A., Manabe, S., Tsuda, T., Takeda, T., Okada, K., Murata, T., Mihara, N. (2016) "A Strategy for Re-using the Data of Electronic Medical Record Systems for Clinical Research," *Stud Health Technol Inform*, Vol. 228, pp. 297-301.
- Matsumura, Y., Hattori, A., Manabe, S., Takahashi, D., Yamamoto, Y., Murata, T., Nakagawa, A., Mihara, N., Takeda, T. (2017) "Case Report Form Reporter: A Key Component for the Integration of Electronic Medical Records and the Electronic Data Capture System," *Stud Health Technol Inform*, Vol. 245, pp. 516-520.
- JAHIS 医療情報システム (オーダエントリ・電子カルテシステム) 導入調査
- (https://www.jahis.jp/action/id=57?contents_type=23, 2022 年 9 月 24 日最終閲覧).
- U.S. Food & Drug Administration. Framework for FDA's Real World Evidence Program. 2018. (<https://www.fda.gov/media/120060/download>, 2022 年 9 月 24 日最終閲覧).
- European Medicines Agency. Use of patient disease registries for regulatory purpose — methodological and operational considerations. (<https://www.segci.org/ema-consultation-discussion-paper-use-of-patient-disease-registries-for-regulatory-purposes-methodological-and-operational-considerations/>, 2022 年 9 月 24 日最終閲覧).
- MID-NET (Medical Information Database Network) (<https://www.pmda.go.jp/safety/mid-net/0001.html>, 2022 年 9 月 24 日最終閲覧).
- 厚生労働省「令和 4 年度診療報酬改定の概要 入院 III (短期滞在手術等・DPC/PDPS)」. (<https://www.mhlw.go.jp/content/12400000/000920426.pdf>, 2022 年 9 月 24 日最終閲覧).
- 診療録直結型全国糖尿病データベース事業 (J-DREAMS) (<http://jdreams.jp/>, 2022 年 9 月 24 日最終閲覧).
- 退院時サマリー HL7 FHIR 記述仕様 第 1 版 (<https://www.hl7.jp/library/item/HL7J-FHIR-002.pdf>, 2022 年 9 月 24 日最終閲覧).
- 日本循環器学会「JCS データ出力標準フォーマットガイドライン」(https://www.j-circ.or.jp/itdata/jcs_standard.htm, 2022 年 9 月 24 日最終閲覧).
- 医療情報システム開発センター「医薬品 HOT コードマスター」(<http://www2.medis.or.jp/master/hcode/>, 2022 年 9 月 24 日最終閲覧).
- 日本臨床検査学会「臨床検査項目分類コード」(<https://www.jslm.org/committees/code/>, 2022 年 9 月 24 日最終閲覧).
- Clinical Data Interchange Standards Consortium. (<https://www.cdisc.org/>, 2022 年 9 月 24 日最終閲覧).
- 日本腎臓学会「J-CKD-Database」(<http://j-ckd-db.jp/>, 2022 年 9 月 24 日最終閲覧).
- Common Terminology Criteria for Adverse Events (CTCAE) version 5.0 (<https://jcog.jp/doctor/tool/ctcae5/>, 2022 年 9 月 24 日最終閲覧).