

□第9回国際医療福祉大学学会学術大会 記念講演□

医療福祉データの活用へ向けての本学の取り組みと未来

平松 達雄¹

抄 録

日本においても、全レセプトを集めたNDBや大部分の手術情報を網羅するNCDをはじめ、医療ビッグデータの活用が進みつつある。国際医療福祉大学においても附属病院のデータを集積して様々な学内活用するためのデータ活用基盤の整備を開始した。整備する基盤には、個人情報である医療データを外部提供せずに外部と連携する仕組みがあり、これにより国際共同研究を含む大きな研究成果に将来寄与することが期待できる。

キーワード：リアルワールドデータ、医療データベース、医療ビッグデータ

I. はじめに

世界で最も価値のある資源はもはや石油ではなくデータとなったといわれる¹⁾。石油石炭利用の歴史は単に燃焼させて光や熱を得ていた時代から、機械技術と組み合わせることで蒸気機関・内燃機関といった動力源となって産業革命・移動革命をもたらし、20世紀には石油から化学合成して作られるプラスチックや化学繊維が登場して私達の生活全体を変革してきた。データの活用も単に集計する時代から、20世紀になり推計統計の手法が徐々に開発され、20世紀後半にはコンピュータ利用と組み合わせることで高度な分析手法が発達した。さらにコンピュータそのものを前提とした機械学習が登場し、近年はより高度なディープラーニング手法が席卷してAIブームにつながるなど、データ利用の高度化が私達の社会に大きな影響をもたらす時代になってきている。これは「データ駆動型社会」と表現され、社会のすべての分野がデータを集めて解析することによって動いていく社会になるといわれている。

医療に目を向けると「医療ビッグデータ」という言葉もNHKスペシャルや日経新聞の社説でも取り上げられるなど一般に登場するようになってきている。このようなデータ駆動型医療福祉とでもいふべき時代に本学も積極的に活躍できるための取り組みが必要であり、本稿では現在取り組みを進めているデータ活用基盤整

備と、それによる研究活動の将来像ともいえる具体例を紹介する。

II. 国内外での医療データ活用の状況

1. 国内の状況

国内の医療福祉分野においても、下記に紹介するNDBやNCD、そしてMID-NETに代表される医療ビッグデータがあり、それぞれ活用されている。

まずはNDB、ナショナルレセプトデータベースは、国内の電子レセプトすべてが匿名化されて厚生労働省に集められたものである。現在は例外的なものを除き紙レセプトでの提出はないため、事実上日本のすべての保険診療のデータが入っている。150億件以上のデータが蓄積され、日本最大の診療情報データベースとなっている。本来は国や自治体の医療施策立案のために集められているものであるが、研究利用のために審査を経て抽出データを第三者提供することが行われてきた。改正法が本年成立し2020年度の施行後は介護レセプトデータベースと結合ができ、さらに民間企業の利用も可能になる予定である。NDBを研究利用するためには、まずは厚生労働省の「レセプト情報・特定健診等情報の提供に関するホームページ」²⁾に掲載されている情報を熟読したのち、事前相談窓口にコンタクトする。最終的には、条件を満たす申請を提出して審査に通れば、データを利用することができる。

¹ 国際医療福祉大学 未来研究支援センター 教授

なお NDB には汎用性の高い基礎的な集計を済ませた「NDB オープンデータ」という集計表も存在する。こちらは誰でも審査なしに厚生労働省のサイトからダウンロードして利用できる公開データとなっている。

NCD (National Clinical Database) は、主に外科領域で周術期情報を登録して集積するレジストリー型のデータベースである³⁾。外科系の多くの学会が協力して設立した一般社団法人により運営されている。情報の登録が外科の専門医取得と密接に結びついているため、全国の手術の95%以上が登録され、手術に関してほぼ全数データといってよいデータベースとなっている。

MID-NET は、厚生労働省と PMDA (独立行政法人医薬品医療機器総合機構) が整備する、主に医薬品の安全対策に役立つためのデータネットワークである⁴⁾。全国10拠点の協力医療機関(23病院、現在10病院追加中)にデータベースを整備し、必要に応じて10拠点であるいは各拠点のデータをあわせて調査を実施する。医薬品安全対策ではなくても公的研究費による公益性が高い研究には利用が可能となっている。

民間企業による大型の匿名加工情報データベースも存在する。JMDC は保険者(企業健保)からレセプトを集めてデータベース化しており560万人(2018年)のデータがある⁵⁾。保険者のデータであるため、受診した医療機関にかかわらず情報があり、転職しなければ同じ保険であるため患者追跡性が良い。一方、企業健保であるため退職後の高齢者の情報はない。MDV は医療機関からデータを集めてデータベース化しており3000万人近い(2019年12月)データがある⁶⁾。医療機関からのデータであるため、一部の病院からのデータには検査結果値もある一方、対象外の医療機関を受診したときの情報は入っていない。上記どちらも民間企業のサービスであり、研究であるかどうかにかかわらず利用できる。

2. 海外の状況、特にオデッセイ(OHDSI)について

海外においても当然ながら各国で整備が進んでいる。日本より進んでいる地域も多い。米国において

65歳以上の高齢者および低所得者に公的保険を提供するCMS⁷⁾のデータ、英国のかかりつけ医制度の診療データ CPRD⁸⁾、それに北欧やその近隣諸国における全国的な医療情報制度のデータベース等が有名である。それらのほとんどは各国個別に進められているものだが、国際的な活動として最近急速に広まってきたオデッセイ(OHDSI)というものがある⁹⁾。今まで日本ではあまり紹介されていなかったこと、そして本学のデータ活用基盤はオデッセイの国際的な規格に準拠していることもあり、ここで詳しく紹介したい。

オデッセイは一連の活動を推進するコミュニティの名称で、古代ギリシャ叙事詩オデュッセイアの英語名であるOdysseyにちなんでいる。綴りは異なるがOdysseyと英語で同じ発音ができるようにObservational Health Data Sciences and Informaticsの略称としてOHDSIという表記をする。英語発音は人や場面によりオデシー、オウデシー等で、日本語ではOdysseyの慣習読みどおりにOHDSIも「オデッセイ」と読む。Odysseyはトロイア戦争の英雄オデュッセウスが戦のあと故郷へたどり着くまでの何年にもわたる苦難の物語であり、オデッセイコミュニティが成立するまでにあった紆余曲折を表しているとのことである。

オデッセイ(OHDSI)は米国が発祥でその中心ではあるが、米国以外にも広がっている。欧州では1億人以上のデータ蓄積を目指すEU政府と欧州製薬業界団体の共同プロジェクトが2018年からはじまった¹⁰⁾。アジアでは中国と韓国に地域コミュニティがあり、特に韓国は政府資金による三次医療機関の70%をカバーする大規模な事業が進行中である¹¹⁾。それ以外に地域コミュニティとして成立してはいないが、グローバルなオデッセイ活動の一環として世界の多くの国で活動がある(図1)。

ところが日本については地域コミュニティはおろか、オデッセイ活動をしている研究者も不在な状況であった。かつてアジア地域での薬剤疫学プロジェクトのためにその後オデッセイが引き継ぐことになるOMOP(オモップ)規格にデータ変換が行われた以外では、認識している限り日本では筆者が唯一OMOP

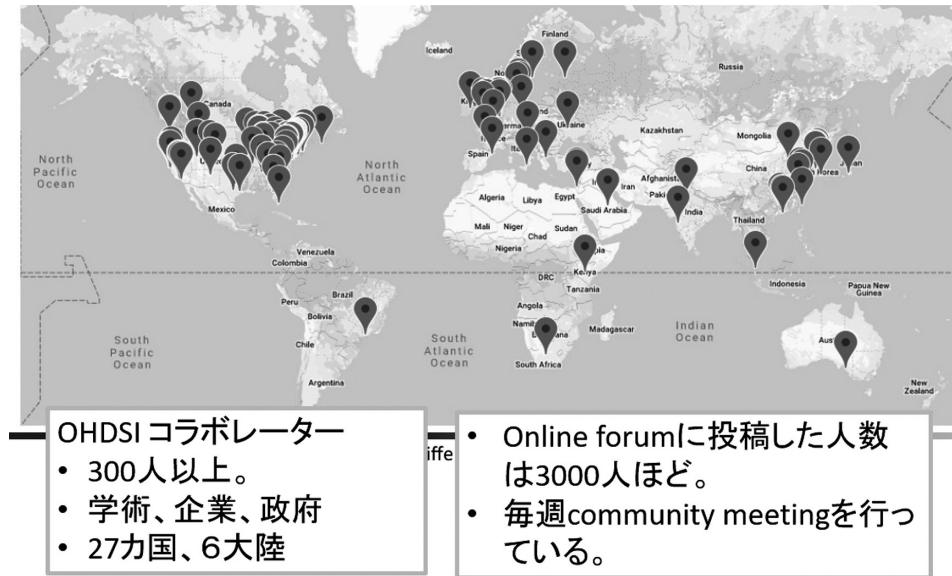


図1 グローバルなオデッセイの活動

規格を使って研究を進めていたが、やはりオデッセイ活動は行っていなかった。そのような中オデッセイは欧州でも上記の大規模事業が開始され、研究成果も世界的有数のジャーナルに掲載されるなどグローバルに大きく発展するようになってきた。このままでは日本は世界に大きく遅れをとる状況であったため、日本でもコミュニティを作ろうと、2019年6月から筆者と共同研究者・協力者が呼びかけを開始した。数度の会合ののち、2019年11月に無事オデッセイジャパン(OHDSI Japan)が成立し¹²⁾、本学東京赤坂キャンパスを主な拠点として活動している。国立・私立の大学病院、ナショナルセンター病院、大手製薬企業、IT企業、データ企業などの研究者・実務者がバランスよく参加するコミュニティとなっている。

オデッセイにはOMOP規格で形成されたデータネットワークの情報が登録された連携データネットワークがある。そこには現在17カ国から152のデータベースが登録されており¹³⁾、それらを総合すると重複を除いて推定6億人以上が含まれる。世界人口の10%が見えてくる水準まで来ている。

オデッセイは診療情報の分析に関するオープンサイエンスコミュニティであり、自身のミッションとしては実診療データに基づく医療エビデンスの生成をす

ずめる。オープンで自由な風土をモットーとしていて強制がなく、参加者の所属セクターを問わない風土と、個人情報のかたまりである医療データを組織外に出さなくて良いという方法により、とても参加しやすい特徴を持っている。さらに、オデッセイ自身のミッションはエビデンスの生成だと明確化していることで、各参加者組織の本業を邪魔せず、柔軟に同居しつつ推進できる存在となっている。つまり研究組織であれば共同研究として参加して一緒にエビデンスを生み出せばよく、非研究組織であれば本業と競合しないためオデッセイを活用してそれぞれの立場から本業を加速することに利用できる。

III. 本学での整備

1. 医療福祉臨床情報データベース(HWCI-DB)構想
本学には成田病院を入れて6病院、グループだと12の病院がある。これらの病院データを集めて、様々な学内グループ内活用するための基盤HWCI-DB(Health and Welfare Clinical Information Database, ヒューシーディービー)の整備する構想を進めている。そのイメージを図2に示す。ただし病院を持つ法人はグループに3つあるため、実際に1個にまとめるのかあるいは3個にまとめるのかは今後の検討課題である。

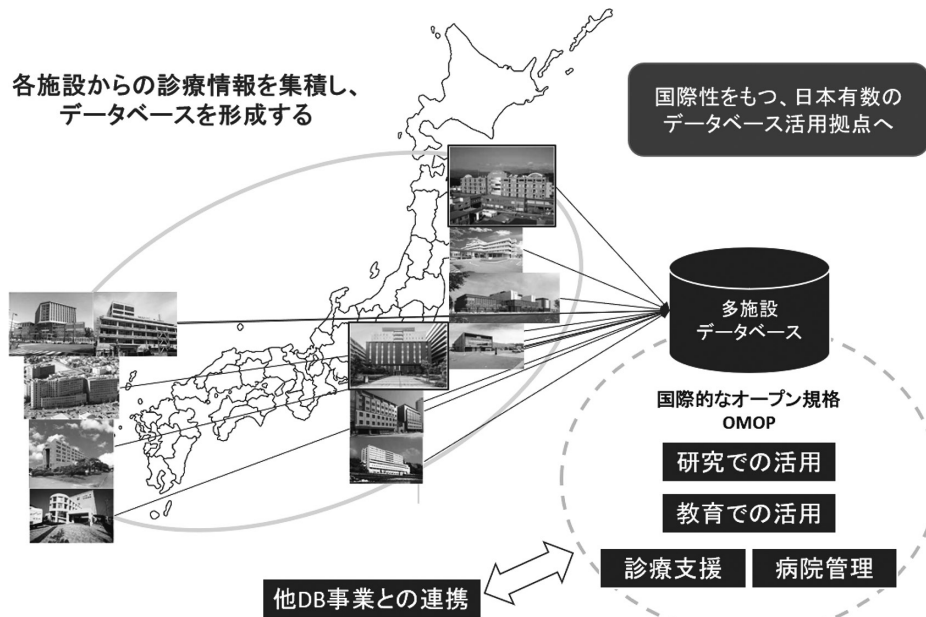


図2 IHW グループ 医療福祉臨床情報データベース (HWCI-DB) イメージ

現在はまず附属病院のデータを集積する基盤の整備を開始した。

データベース構築までのデータの流れを簡単に説明する。最初は容易に元データが準備できるように、ということから、どこの病院にもあるレセプトと、病院のシステムから取り出した検査値の一覧データを使用する。これらのデータを、すでに開発してあるツールを使ってデータベースに取り込み、様々な分析ソフトやツールを使って活用していく。データベース化するツールは、以前から筆者が手掛けていた文部科研で使う基盤ツールとして開発したものである。文部科研では共同研究する複数の大学病院でそれぞれ動作する必要があるため、仮想マシン上で動作するものとして開発して、ハードウェアやOSなどの環境の違いに柔軟に対応できるようにしている。

2. 取り込みツールの動作例

三田病院の許可を得てデータを早速取り込んでみた。図3のようなシンプルなブラウザ画面があり、レセプトの取り込み、検査値の取り込み、その後の検証等ができる。取り込み中には画面のバーグラフが順に伸びていき、取り込みの進行状況を表す。

画面の下の方には、取り込み動作を指示するためのメニューがある。取り込みが完了すると、「取込結果の確認」にて取り込みデータを示すための集計表が表示できる。年齢階層・男女別の人数、月別の入院患者数、外来患者数、開始月別の疾病数、薬剤の使用数、月別の処置と検査実施数といった一覧表があり、想定される範囲のデータが読み込まれたか、あるいは生年月日が1900年以前や2020年以降になっている等の異常な値があるかどうかの確認ができる。

3. データベース利用の簡易例

データベースに取り込んだあとは、汎用の市販分析ソフトからデータベースに接続して利用することができる。図4はTableau(タブロー)というソフトを使用した例で、男女別の生まれ年の分布図を示す。この例では横軸が年齢ではなく生まれ年なため、左に行くほど高齢者、右端は乳児になる。これを見ると基本的に男女とも高齢の方が多いのだが、左側の女性は二峰性になっていて、50歳ぐらいの人も多いことがわかり、病院の特徴が出ている。

図5は同じくTableauを用い、よく使われる医薬品を多い順に並べたものである。最上位には、当然のこ

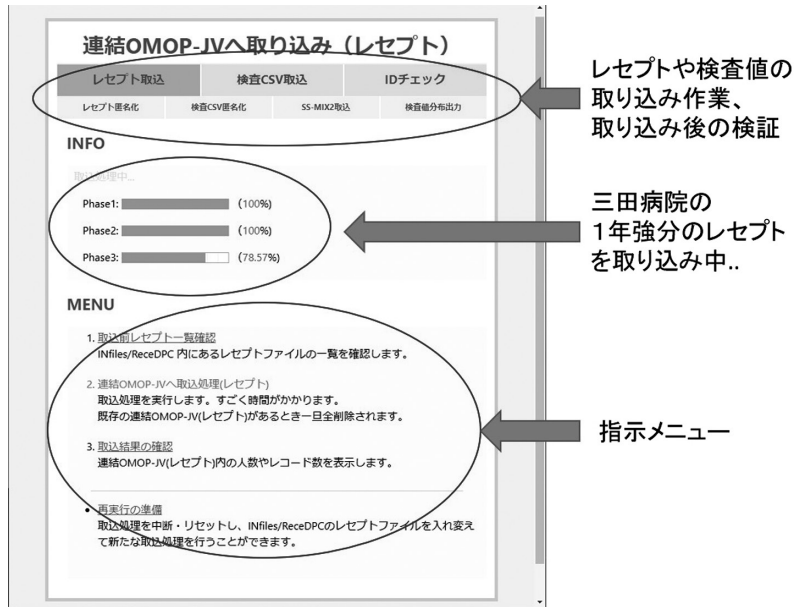


図3 現在のデータ取り込みツール操作画面

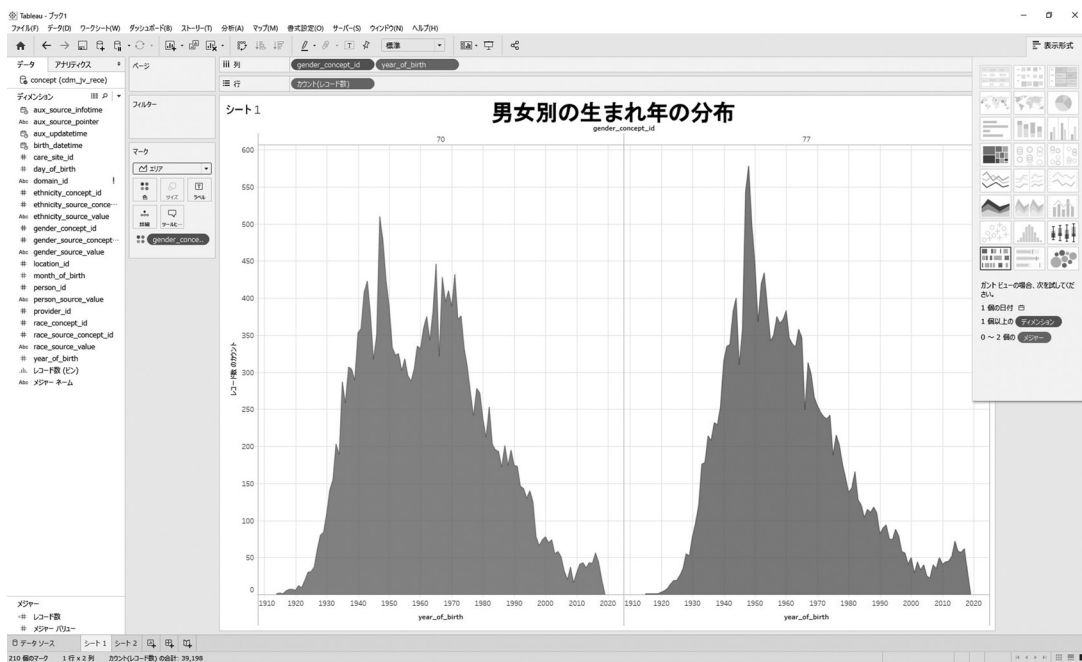


図4 男女別の生年の分布図

とではあるが、輸液や生理食塩水が並ぶ。次のグループは飲み薬で胃の保護剤、痛み止め、それから緩下剤となっていて、納得できる結果になっている。薬らしい薬で一番上はアムロジピン、高血圧の薬である。高血圧は厚労省の患者調査で最も多い疾患であり、妥当性が高い結果と考えられる。

次に別の分析ソフト、KNIME (ナイム)、というも

ので接続した例を示す(図6)。前述の Tableau は操作するとリアルタイムに表やグラフを作っていくソフトであるのに対し、この KNIME は、アイコンで示している機能を接続していき、プログラムのようなものを作成して、その処理結果を出すソフトである。誤解されがちだが、このソフトはアイコンでグラフィカルに操作できることが良いのではなく、一連の処理の途中

結果をすべて後から確認できる機能があり、すべての途中データをいつでも参照できるため、処理が各段階で想定通りに行われているかの確認が容易で処理結果に確信がもてる、ということが通常のプログラム言語にはないすばらしい特徴である。図6では診療科別の

疾患トップ3を抽出してみた。診療科が数字のままになっているのは、レセプトの診療科コードをそのまま表示しているからで、病院の診療科と必ずしも1対1に対応するものではないが、1が内科、2が精神科、となっている。内科では高血圧がトップで、精神科で

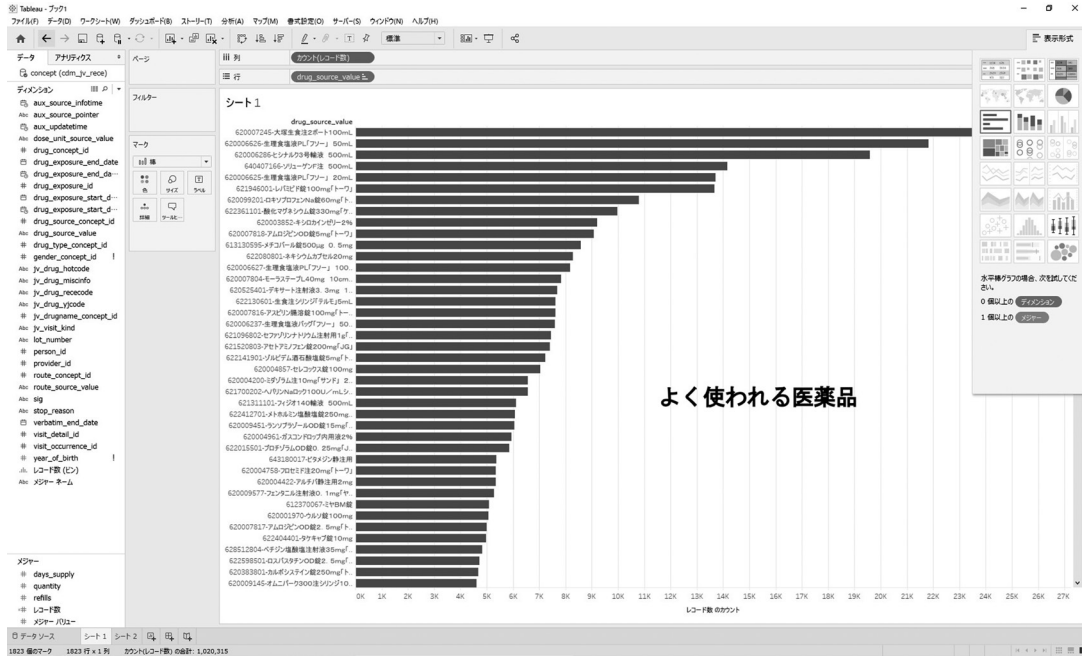


図5 よく使われる医薬品

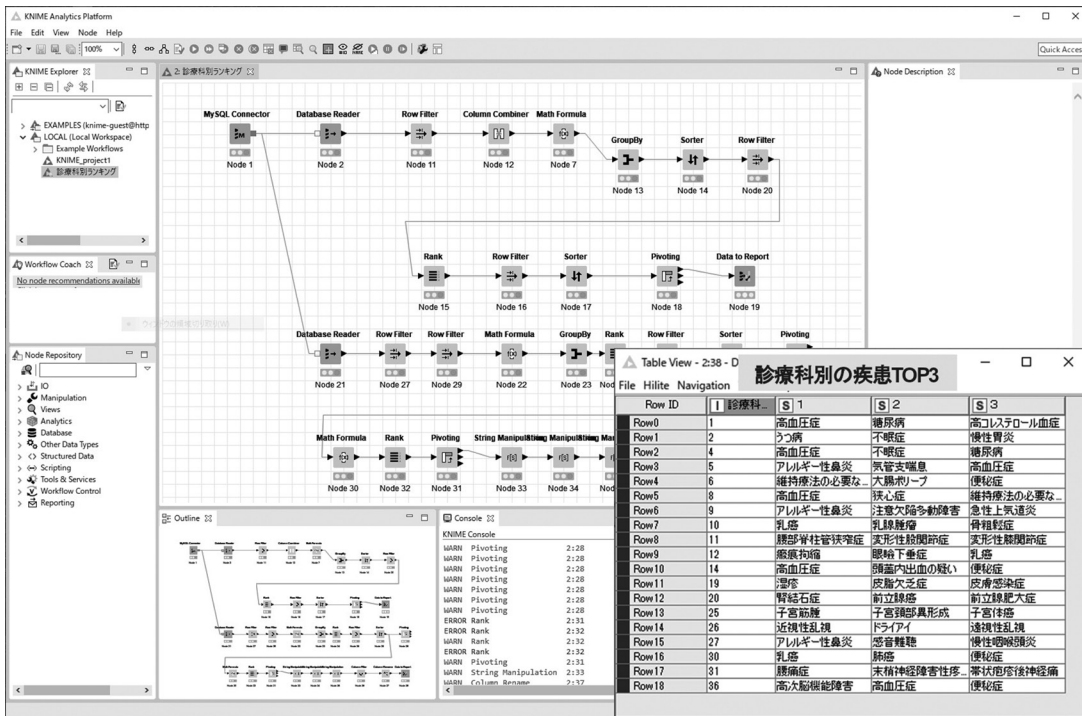


図6 診療科別の疾患トップ3

はうつ病、とこれも納得性が高い。

上記はデモンストレーション目的の簡易例を紹介したにすぎないが、データベースは様々な活用できる。たとえば、治験の対象になる候補患者の検索にはすぐに利用できる。そのほか、当然ながら研究で様々な分析して論文化することができる。大学院で教育に使うことも考えられ、AIを絡めて診療の高度化にも応用できる。そして病院管理にも使用してよりよく経営することにも貢献できる。

IV. オープン規格であることを生かした連携活用・国際共同研究のすがた

HWCI-DBでは前述の国際的なコミュニティであるオデッセイ(OHDSI)が推進する国際的なオープン規格OMOP(オモップ)を採用している。この規格は、単に規格があるだけでなく、本規格のデータベース相互を個人情報のやり取りなしに連携する仕組みが、オデッセイにより維持されている。それにより国を超えた大規模な研究が可能で、その成果は下記に示すように世界のトップジャーナルにも掲載されるようになってきた。

オデッセイではOMOP規格専用の分析統合ソフトも開発している。このソフトでは特定の条件を満たす

患者群としての「コホート」がデータベース内に定義でき、複数のコホートを使って罹患率の計算や、COX比例ハザードモデル等を使った臨床疫学的な推定が可能で、さらに機械学習の機能もある。

このソフトを使って、オデッセイでは個人情報を移動させずにデータベースを連携することができる。国をまたいで医療データを提供するには特別な手続きを要する場合も多いため通常は国際連携は難しいが、オデッセイではデータを移動させないことで容易に国際連携ができる。一例として、図7の国際共同研究では、4カ国、11のデータベースが共同で研究を実施しており、対象の患者総数は2.5億人にもなる¹⁴⁾。世界レベルでの連携パワーを認識せざるおえない。

別の研究例では、JAMA network openに昨年掲載された3カ国での共同研究がある(図8)¹⁵⁾。糖尿病患者でメトホルミンを投与されていた患者が別の薬に切り替えることにしたとき、他の3種類の糖尿病薬ではどのタイプが良いのか、実臨床データを解析している。メインのアウトカムとしてはHbA1cの改善を見て、サブとして心筋梗塞、腎障害、眼障害の発生を見ている。これは実験的臨床研究であるランダム化比較試験(RCT)では実施できない種類の研究である。

そしてつい最近2019年10月には、オデッセイの研

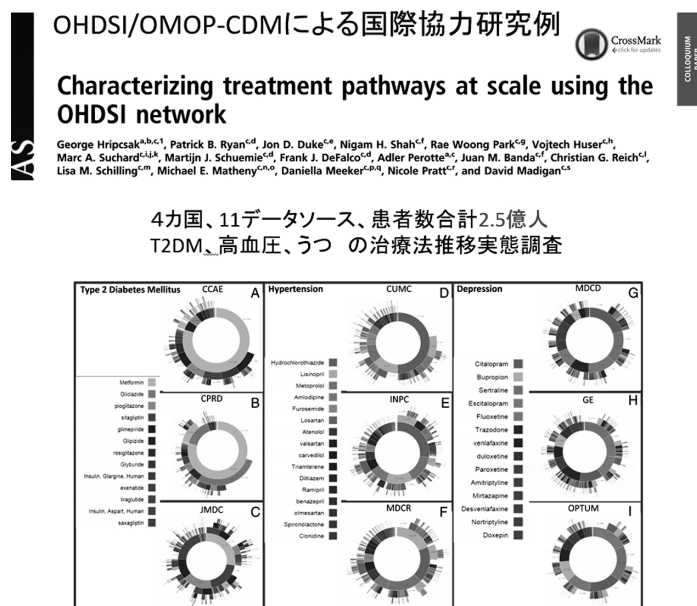


図7 大規模な国際共同研究例

究はLANCET誌に掲載された(図9)¹⁶⁾。高血圧症の薬では第1選択として多くの種類の薬が並列されるが、はたして実臨床ではどれが良いのかを、系統的に複数国で大規模な解析を行ったものである。

先進的な研究例をいくつか紹介したが、これらはすべて本学のHWCI-DBで採用するOMOP規格を推進するオデッセイの活動として行われたもので、HWCI-DBをすすめることで日本のオデッセイの拠点である

本学もこれらのような研究に参加できるし、さらに本学が国際的な研究を主導することも可能になるであろう。

V. まとめ

本稿では、国内海外での医療データ活用の概況を紹介したのち、本学での取り組みについて説明し、その研究活用の将来像ともいえる例を提示した。本学で研

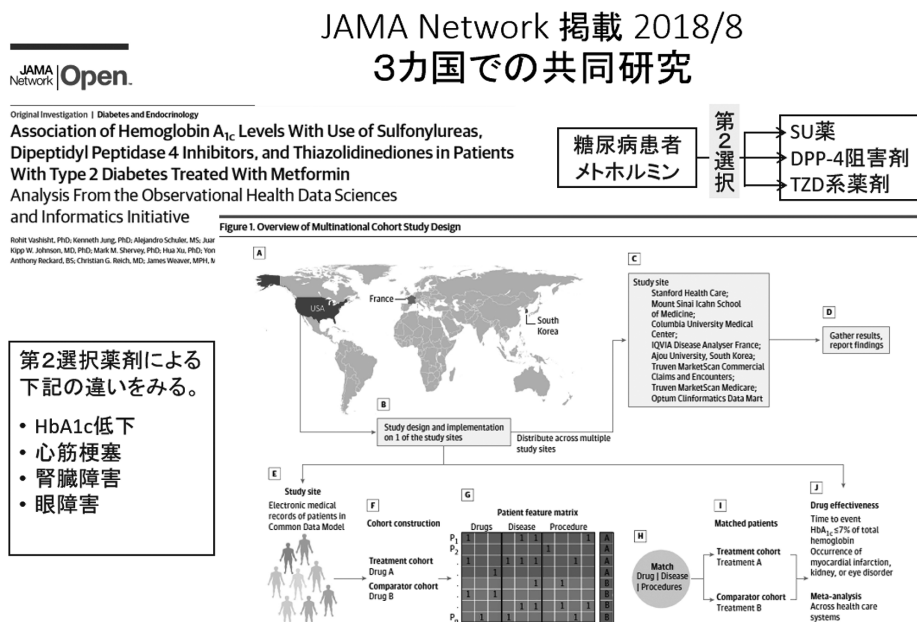


図8 医薬品の効果比較研究例



高血圧症の薬: 第1選択に並列される薬のどれが良いのか。系統的で複数国・大規模な(500万人)分析

Summary

Background

Uncertainty remains about the optimal monotherapy for hypertension, with current guidelines recommending any primary agent among the first-line drug classes thiazide or thiazide-like diuretics, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, dihydropyridine calcium channel blockers, and non-dihydropyridine calcium channel blockers, in the absence of comorbid indications. Randomised trials have not further refined this choice.

図9 LANCETに掲載されたオデッセイ研究

究に携わる方々も、このような研究活動へのご参加を
ぜひ検討されたい。

文献

- 1) The Economist, May 6, 2017
- 2) https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryoku/iryohoken/reseputo/index.html 2019.12.3
- 3) <http://www.ncd.or.jp/> 2019.12.3
- 4) <https://www.pmda.go.jp/safety/mid-net/0001.html> 2019.12.3
- 5) <https://www.jmdc.co.jp/jmdc-claims-database/> 2019.12.3
- 6) <https://www.mdv.co.jp/ebm/> 2019.12.3
- 7) <https://dnv.cms.gov/> 2019.12.3
- 8) <https://cprd.com/> 2019.12.3
- 9) <https://ohdsi.org/> 2019.12.3
- 10) <https://www.ehden.eu/> 2019.12.3
- 11) <http://www.feedernet.org/html/> 2019.12.3
- 12) <https://ohdsi-japan.org/> 2019.12.3
- 13) https://www.ohdsi.org/web/wiki/doku.php?id=resources:2019_data_network 2019.12.3
- 14) Hripcsak G, Ryan PB, Duke JD, et al. Characterizing treatment pathways at scale using the OHDSI network. *Proc. Natl. Acad. Sci. U.S.A.* 2016; 113(27): 7329-7336
- 15) Vashisht R, Jung K, Schuler A, et al. Association of hemoglobin A1c levels with use of sulfonylureas, dipeptidyl peptidase 4 inhibitors, and thiazolidinediones in patients with type 2 diabetes treated with metformin: analysis from the observational health data sciences and informatics initiative. *JAMA Netw. Open.* 2018; 1(4): e181755
- 16) Suchard MA, Schuemie MJ, Krumholz HM, et al. Comprehensive comparative effectiveness and safety of first-line anti-hypertensive drug classes: a systematic, multinational, large-scale analysis. *Lancet* 2019; 394(10211): 1816-1826