

レセプト情報・特定健診等情報データベースの申出者対応部門の充実

研究分担者 満武 巨裕

一般財団法人 医療経済研究・社会保険福祉協会 医療経済研究機構、副部長

研究要旨

本報告書は、今後の日本におけるレセプト情報・特定健診等情報データベース(以下、NDB)の情報提供機能について、諸外国の先進的事例を参考にして、今後の充実について検討する。

今年度は、日本と類似の国民皆保険制度およびレセプト審査・支払い方式を導入し、一昨年からNational Patient Sampleという患者サンプルデータの試行提供を開始した韓国を調査対象とした。韓国の患者サンプルデータは、既に台湾において被保険者ファイル(ID)から、100万人をランダムサンプリングし、抽出された被保険者の入院、外来、調剤レセプトデータを提供している例を参考にしている。

韓国は、台湾を参考にランダム抽出した患者の入院、外来、調剤レセプトデータ(HIRA-NPS (HIRA NationalPatient Sample)の試行提供を一昨年から開始した。HIRA-NPSは、韓国国内において1年間に医療機関を利用した全患者対象を母集団として、性別・年齢(5歳単位)区間による患者単位の層化系統抽出を行ったデータセットである。現在、韓国のHIRA-NPSは5種類のテーブルで構成されている。また、患者サンプルデータは、5つの学会との覚書(MOU)を交わして検証が行われた。例えば、主要な検証の一つに、糖尿病およびジペプチジルペプチダーゼ4阻害剤の使用に関する有病率について、患者サンプルデータを用いた推定値が既存研究と整合性があることが証明された。一方、患者サンプルデータの期間が1年間単位であり、個人の縦断突合ができないために、有病期間が長い慢性疾患などの分析にも適していないことが指摘されている。

また米国CMS(Center for Medicare and Medicare Services)は、VRDC(Virtual Research Data center:バーチャル研究データセンター)というバーチャルアクセス機能を提供して、利用者に効率的かつ対費用効果の高い方法でメディケアとメディケイドプログラムデータへのアクセス環境を提供している。VRDCを利用する研究者は、承認されたデータファイルへ直接アクセスができ、CMSのセキュアな環境の中で研究が実行できる。また、研究者本人のローカルマシン(自身の研究室のワークステーションやPC)に、集約されたレポートと結果をダウンロードすることができる。

日本のレセプト情報等データベース(以下、NDB)から提供されるデータセットは、厚生労働省側で複雑な構造の電子レセプトを研究者の要望に応じて、ある程度分析し易

い形式に加工して提供する特別抽出と一月分のサンプリングデータの提供サービスが存在している。加えて基本データセットの設計と作成が検討され、今年度から試行提供が始まった。日本のNDBは、提供開始（2011年11月）から平成28年3月までの承諾件数は合計94件となったが、台湾と韓国の平成25年の提供件数は、韓国は115件/年、台湾は270件/年である。したがって平成28年中に公開されるNDBオープンデータをはじめとする集計情報も含めて、量と質の増加が必要である。その方向性の一つとして、韓国HIRAの患者サンプルデータ、米国CMSのVRDCは今後の日本の**申出者対応部門を充実**する上で有益な先行事例である。

A. 研究目的

本報告書は、今後の日本におけるレセプト情報・特定健診等情報データベース(以下、NDB)の情報提供機能について、諸外国の先進的事例を参考にして、今後の充実について検討する。

今年度は、日本と類似の国民皆保険制度およびレセプト審査・支払い方式を導入し、一昨年からNational Patient Sampleという患者サンプルデータの試行提供を開始した韓国を調査対象とした。韓国の患者サンプルデータは、既に台湾において被保険者ファイル(ID)から、100万人をランダムサンプリングし、抽出された被保険者の入院、外来、調剤レセプトデータを提供している例を参考にしている。

また米国CMS(Center for Medicare and Medicare Services)は、VRDC(Virtual Research Data center:バーチャル研究データセンター)というバーチャルアクセス機能を提供して、利用者に効率的かつ対費用効果の高い方法でメディケアとメディケイドプログラムデータへのアクセス環境を提供している。VRDCを利用する研究者は、承認されたデータファイルへ直接アクセスができ、CMSのセキュアな環境の中で研究が実行

できる。また、研究者本人のローカルマシン(自身の研究室のワークステーションやPC)に、集約されたレポートと結果をダウンロードすることができる。

日本のNDBから研究者に提供されるデータ件数も近年増加傾向にあるが、先行事例の米国、台湾、韓国にはおよばない。したがって平成28年中に公開されるNDBオープンデータをはじめとする集計情報も含めて、量と質の増加が必要である。その方向性の一つとして、韓国HIRAの患者サンプルデータ、米国CMSのVRDCは今後の日本の**申出者対応部門を充実**する上で有益な先行事例である。

B. 研究方法

韓国は、HIRA から患者サンプルデータについての資料提供を基にしている。(内容は、[Kim L](#)らの“A guide for the utilization of Health Insurance Review and Assessment Service National Patient Samples” Epidemiol Health.Vol;36,ArticleID:e20140008,2014に要約されている)

米国は、CMS のデータ提供に関するサポート業務を行っている ResDAC(Research Data Assistance Center)がインターネット

トで提供している Introduction to the Virtual Research Data Center (VRDC) (URL: <https://www.resdac.org/cms-data/request/cms-virtual-research-data-center>) を基にしている。

C. 研究結果

韓国は、日本における審査支払機関に該当する HIRA(The Health Insurance Review and Assessment Service (健康保険審査評価院))が、HIRA-NPS (HIRA National Patient Sample)として、韓国国内において 1 年間に医療機関を利用した全患者対象(約 4600 万名)を母集団として、性別・年齢(5 歳単位)区間による患者単位の層化系統抽出を行ったデータセットである。HIRA の患者サンプルデータは、患者の診断、治療、処置、手術歴、および医療サービス研究のための貴重な情報源である処方薬情報を含んでいる。しかし、入院患者の 10%および外来患者の 90%で構成されている国会の国家患者サンプル (NPS) は、重症化した入院患者を調査するのに十分な数を確保していない可能性がある。

そこで昨年韓国から韓国の HIRA-NPS は 5 種類のテーブルで構成されるようになった。具体的には、国家患者サンプル (HIRA-NPS) に加えて、国家入院サンプル (HIRA-NIS)、国家高齢者 (65 歳以上) サンプル (HIRA-APS)、および小児患者サンプル (HIRA-PPS) が追加された。追加は、NPS データに確保されていないグループの研究をサポートするために、利用可能とした別々のサンプルデータであ

る。

しかし、これらの患者サンプルデータは、一年間分のレセプト請求データをソースとして作成されている断面調査である。患者らのプライバシーを保護するために、毎年サンプルデータは作成され、特定の個人または医療サービス提供者も患者サンプルデータであって横断的な調査はできない。つまり、複数年の患者サンプルデータを使っても患者の長期間の観察研究を行うことができないようになっている。だが、また、医療援助プログラム、政府支出、および退役軍人患者のデータも請求データに含まれている。

しかし、レセプトの複雑な構造とレセプト請求データの膨大な量は、研究者に一定以上の負担を課すことになる。また、膨大なデータ量は、研究を行う上で非効率性をもたらす可能性がある。これらの制限事項を解決し、レセプトデータの利用向上と研究者へのアクセシビリティを向上させるために、HIRA は 5 つの異なる機関によって行われた検証を経た患者サンプルデータを開発した。

患者サンプルデータは、それぞれ 5 つのテーブルで構成されている。すべてのテーブルは、キーID を使用してリンク可能となっている。基本的属性テーブルは、このような性別・年齢および医療援助プログラムといった社会人口学的特性、主要診断名、二次診断名、診療開始日や実日数、患者の自己負担額などのから構成されている。医療サービステーブルは、入院患者のための処置、治療、薬剤情報など、患者に提供され入院と外来医療サービス情報から構成されている。診断情報テーブルは、

患者の診断情報がすべて含まれている。このテーブルは、患者の合併症または全ての病名の履歴が必要と判断された場合に使用される。外来処方テーブルは、成分、投与量と供給日といった、外来患者のための処方薬剤の情報から構成されている。プロバイダテーブルは、患者の受診した医療機関の種類（プライマリケア、二次ケア、専門治療）、位置、病床規模、運営（経営）母体のタイプといった情報から構成されている。

患者サンプルは、韓国の患者全体の代表性を有しており、5つの学会として韓国予防医学会(Korean Society for Preventive Medicine)、韓国医療経済学会(Korean Association of Health Economics and Policy)、韓国医療情報・医療統計学会(Korean Society of Health Information and Health Statistics)、韓国医療政策・管理学会(Korean Academy of Health Policy and Management)、韓国疫学学会(Korean Society of Epidemiology)との覚書(MOU)を交わして、検証されている。したがって、研究を行う際に利用したデータが母集団の特性を有しているかについて検討するための説明を省くことができることが証明されている。主要な検証結果として、糖尿病およびジペプチジルペプチダーゼ 4 阻害剤の使用の評価の韓国有病率について、推定値は人口全体と整合のある患者サンプルであることが証明されている。また、血糖降下薬利用の処方の推定値についても検証が成功した。さらに、それぞれの血糖降下剤の外来処方率はすべて 95%信頼区間内であった。「視力低下や失明に関連する疾患の社会的コスト」¹ 患

者サンプルおよび人口の試験」においても、主要な眼疾患（白内障、緑内障、黄斑変性症、糖尿病性網膜変化）については女性患者においてより高い医療サービスの利用を示した。

一方、米国の CMS の VRDC は、研究目的のために CMS のデータにアクセスし、分析するための新しいソリューション（ツール）である。これまで CMS は、外部メディアに保存した暗号化データファイルを研究者に提供してきたが、VRDC は研究者がアクセスし、事実上、研究者のワークステーションや PC から CMS データの独自の操作・分析を行うことができる。VRDC は、より効率的で費用対効果の高い方法でタイムリーなデータにアクセスするための安全なメカニズムを提供している。

ただし、ユーザー要件としては、SAS プログラミング言語、ブロードバンドインターネット接続、Java6 以上のローカルマシンへのインストール、MS の Internet Explorer または Mozilla Firefox、Windows XP またはそれ以降の Windows オペレーティングシステムでなければならない。

複数の研究者が CMS VRDC 内の単一のプロジェクトに取り組むことも可能になっており、彼らの SAS ライブラリ内の仮想デスクトップ内で共同作業することができる。しかし、CMS VRDC のオンライン・セキュリティトレーニングを受け、完了した証拠を提供する必要がある。研究者全員が全てのセキュリティ要件を満たした後、利用者のアクセス権が付与され、CMS VRDC 環境に接続するために必要

なソフトウェアを備えたパッケージが提供される。ただし、複数の研究者が同じプロジェクトで作業している場合 VRDC の利用にあたっては、それぞれの申請および権利を得なければならない。人数分だけシートと呼ばれる利用権利を購入する必要がある。また、シートを共有することもできない。

このような条件の基、利用者は VRDC により次のファイルにアクセスすることができる。

- ・マスター受益者ファイル
- ・メディケア・パート A、B、D のレセプトデータ
- ・メディケアプロバイダー分析ファイルと MedPAR ファイル
- ・メディケイド (MAX) ファイル等である。

利用者は、SAS による分析が終了した後、ダウンロードできるのは集計情報や統計情報に限られている。個人を特定できるような情報または保護された健康情報は、VRDC から取り出すことはできない。CMS VRDC からデータをダウンロードするためのすべての要求は、個人を特定できる情報または保護された健康情報をスクリーニングするために、出力審査を通過する必要がある。

ダウンロードファイルは、研究目的にもよるが、地理的な単位 (州、郡等) 診断グループレベルに集計しなければならない。出力形式は、Excel テーブル、集約された SAS データセット、SAS 出力ファイル、ワード文書、PDF 文書である。CMS のダウンロードデータに関する審査プロセスは、一般的には 2 営業日 (48 時間)

以内にとしている。しかし、審査内容が複雑である場合には追加の時間が必要とされる。

D. 考察

NDB データを利用する際、厚生労働省や関係省庁・自治体に属さない研究者等への第三者提供については、有識者会議 (レセプト情報などの提供に関する有識者会議) において医療サービスの質の向上を目的とする公益性の高い研究であることが前提で、有識者会議の承諾を得なければならない。承諾の敷居は高く、研究者等への第三者提供を検討した第一回は、43 件の申出に対して承諾件数は 6 件であった。提供開始 (平成 23 年 11 月) から平成 28 年 3 月までの承諾件数は 94 件となった。参考までに、日本と同様の社会保険方式でありレセプトも存在する台湾と韓国の 2013 年の提供件数は、韓国は 115 件 / 年、台湾は 270 件 / 年である。

この承諾件数が低い原因として、次の点を有識者会議は指摘している。(1) 申出者が求めるデータ項目が実際に格納されているデータでは実現困難であった申請が存在した、(2) データ提供にあたっての各種要件や必要な事項を申出者が十分に把握していない申請が存在した、(3) 提供側の情報提供が不十分であった等である。

しかし著者は、上記以外に NDB データの利用規約に原因があると考ええる。第一に、利用者の申請した範囲に分析方法が限定されてしまうことである。つまり、探索的にあれこれと自由に研究することができず、限定されたデータ項目及び期間しか提供されない。また、成果の公表前に、厚生

労働省の承認が必要であり、承認を得なければ発表することができない。加えて、データベースへの複写回数は原則一回、利用場所の施錠と入退室状況の管理、データの持ち出しは原則不可などの規約を守らなければならない。利用場所への外部検査官の立ち入り検査にも応じなければならない。実際に承諾を得た大学や研究所では、大半の研究機関では利用規定を満たすために新たな物理的場所の確保や入退室記録装置を導入しているケースが多く、予算等の問題もあって申請を見合わせる研究者が多い。

ここまで厳格な管理が求められるのも、NDBは医療機関から提供された医療関連情報だからであり、現時点では個人情報に準ずる取り扱いをするということになっているからである。ただし、レセプトに記載されている氏名や住所等の個人情報は全てハッシュ関数による暗号化が施されており、個人を特定することはまず不可能と言える。

NDBのオンサイトセンターは、東京大学と京都大学に拠点がおかれて開始される予定である。しかし、全研究者が二つの拠点に集まらなければならないのは、物理的な距離の問題、分析作業が終了するまでの滞在費用なども問題もある。

NDBから提供されるデータセットは、厚生労働省側で複雑な構造の電子レセプトを研究者の要望に応じて、ある程度分析し易い形式に加工して提供する特別抽出と一月分のサンプリングデータの提供サービスが存在している。加えて基本データセットの設計と作成が検討され、今年度か

ら試行提供が始まっている。平成28年中に公開されるNDBオープンデータをはじめとする集計情報も含めて、量と質の増加が必要である。その方向性の一つとして、韓国HIRAの患者サンプルデータ、米国CMSのVRDCは今後の日本の**申出者対応部門を充実**する上で有益な先行事例である。

E. 結論

NDBからのデータ提供は、特別抽出、サンプリングデータセット、基本データセット、NDBオープンデータをはじめとする集計情報も含めて、今後も量と質の増加が必要である。その方向性の一つとして、韓国HIRAの患者サンプルデータ、米国CMSのVRDCは今後の日本の**申出者対応部門を充実**する上で有益な先行事例である。

F. 研究発表

1) 「基本データセットの提供について」第29回レセプト情報等の提供に関する有識者会議(平成28年3月16日)
<http://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000117367.pdf>

2) 清武巨裕：レセプトビッグデータ解析の現状と将来．**実験医学 第34巻第5号**：799-804, 2016年

G. 知的所有権の取得状況

該当なし

