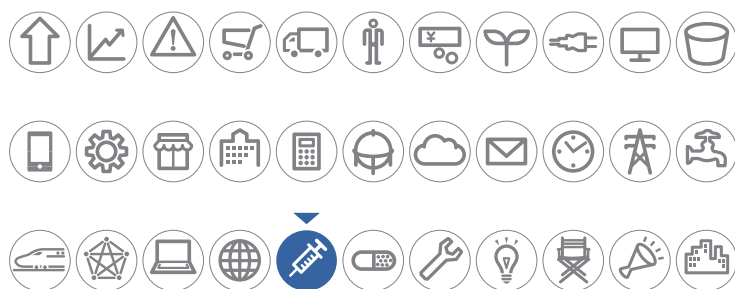


医療ビッグデータの 活用と今後の方向性



わが国の医療ビッグデータの活用と課題

一般財団法人医療情報システム開発センター理事長
東京大学大学院医学系研究科特任准教授 医学博士・医師
山本 隆一

医療分野のデータベースの現状と今後に向けた期待

株式会社 日立コンサルティング 公共本部 マネージャー
早川 和寿



Healthcare

わが国の医療ビッグデータの活用と課題

著者 山本 隆一

一般財団法人医療情報システム開発センター理事長
東京大学大学院医学系研究科特任准教授 医学博士・医師

はじめに

医療分野に限らず、情報を収集する場合、活用する目的があるからこそ電子化し、データベース化するのであって、利用できない、あるいは利用する必要のない個人情報は、プライバシー保護の観点からは、そもそも収集さえすべきではない。わが国の医療への情報化の波は当初、事務処理の合理化を目的として1960年代から現れた。レセコン、医事コンと呼ばれる比較的単純なシステムで、最初は計算機付きの印刷装置程度ではあったが、出来高支払い制を基本とした診療報酬請求には非常に有用なツールであった。さらに1980年代に入ると、医療費の高騰から、大規模病院における経費削減が大きな目的となり、オーダエントリーシステムが普及を始めた。これは病院内の伝票を削減することが目的であり、紙の運搬や記載された情報を医事システムに再入力するという手間が減り、事務経費の削減に一定の効果を示した。事務処理の時間が短縮されることで患者の診察後待ち時間も短縮され、単なる事務経費の削減だけではなく、サービスの向上にも効果があったが、医療自体のサービスの向上に直接つながるものではなかった。ここまでは主に経済的理由で導入が進められたために、普及速度もかなり速く、2000年ごろのわが国は世界でももっとも医療のIT化が進んだ国であった。

しかし医療というサービス全体から考えれば、医療機関の経済的課題の解決は中心テーマではない。患者とのコミュニケーションの向上、医療従事者間の情報共有の効率化、医療従事者の業務の合理化、医学・医療技術の発展、医薬品をはじめ医療の安全性の向上などはいずれも情報をうまく処理し活用することで改善することが期待されるが、これらを目的とした情報化がレセコンや医事コンほどスムーズに発展しているかと言えばそうではない。関係者間で精力的に努力はされているものの、まだ道半ばと言わざるを得ない。なぜスムーズには進まなかったのかを考察すると同時にわが国でもようやく始まった医療情報大規模データベース化と、これらのデータベースを二次利用する場合の課題について述べたい。

「処理」の時代から「分析」の時代へ

レセコンや医事コンは基本的にはレセプトを作成するという情報処理が使命であった。オーダエントリーシステムも基本は変わらない。仕事の単位はレセプトを月次に提出することが基本で、返戻への対応を終ればその単位の仕事は終わる。年報を作成し、推移を分析し、あるいは点数改訂でシミュレーションを行う程度の機能はあったが、利用者(医療機関)が情報を蓄積し、深く分析することはほとんどなく、その情報を疫学的な分析のために提供することもほとんどなかった。システム導入の目的から見れば当然であり、そのためにある程度単純なコンピュータシステムでも実用的であったと言える。つまりわが国の医療のIT化は分析にはあまり向いていないシステムとして始まったと言える。そしてこのことが当初の普及を加速したとも、また近年の医療そのものに寄与するためのシステムの普及を遅らせたとも言える。事務処理を中心としたシステムは目的に応じたデータ構造を持ち、不幸にも普及早期にはコンピュータシステム自体が非力だったこともあり、限界まで目的に最適化したものであった。あとで別の目的で利用することは、そもそも困難であり、またデータ保持も目的を達成する期間に限定されていた。オーダエントリーシステムになり、コンピュータシステムの能力にも多少の余裕が出てきたために、検査結果や処方履歴表示なども可能にはなったが、患者ごとの処理が中心であり、横断的な分析を意識したものは少なかった。多彩な目的のためのいわゆる電子カルテシステムも基本的には医事システムやオーダエントリーシステムをベースに設計されることが多く、過去の制限をどうしても引きずりがちであった。最近になり、あらためて設計されるシステムも増えてきて、ようやく過去の呪縛から解放されつつあるというのが現状であろう。医療情報システムの発展過程の問題だけではなく、わが国では施設横断的なデータの収集があまり進まなかった。国によっては医療の情報化自体はあまり発展していない時期から大規模な横断型データベースが整備されていた。例えば米国のMEDICAREやMEDICAIDはわが国

で言うレセプト情報の大規模データベースをかなり以前から構築している。医療機関からの請求自体は紙ベースであっても合衆国政府のCenters for Medicare and Medicaid Service(CMS)が電子化しデータベース化している。さらにCMS自体が活用するだけでなく、研究者をはじめ外部の利用者にも、さまざまな条件はあるものの解放している。最近ではダッシュボードのような分析機能を付加し研究者にとっては非常に使いやすい形で、有償ではあるが、活発に提供されている。フランスでは公衆衛生のためのデータベースはかなり前から国が構築しているし、ほかにも大規模データベースを持つ国は多い。それに比してわが国は国が行う限定的な調査はあるものの、医療に関する大規模データベースは皆無であった。電子化自体の進行が早かったことを考えると不思議であり残念でもある。

しかし最近になっていくつかの重要なデータベースが登場した。その代表はナショナルデータベース(NDB)と呼ばれているもので、すでにデータベースとしては稼働して数年が過ぎている。その一方で今まで大きく取り上げられることが少なかったデータの公益目的の二次利用と個人のプライバシーのバランスが問題として顕在化してきた。

いわゆるナショナルデータベース(NDB)

よく知られているように「高齢者の医療の確保に関する法律」の中で医療費適正化計画の作成に資するために整備が決められたデータベースで、DPC(診断群別分類)を含むレセプトと特定健診・特定保健指導の結果が蓄積されている。年間にレセプト情報15億件以上、特定健診・特定保健指導情報2千数百万件が格納され、2015年夏ではレセプトが90億件、特定健診・特定保健指導のデータは1億2,000万件格納されている。

政策のために作成が決められたデータベースではあるが、その価値は高いもので、疫学などの医学研究や医療制度・医療経済の研究に有用であることは当然である。した

がってデータベース作成前からその情報の利活用について議論が行われ、故開原成允先生が座長として「医療サービスの質の向上等のためのレセプト情報の活用に関する検討会」の報告⁽¹⁾が2008年2月にとりまとめられた。その中で、医療サービスの向上のために利用は国や都道府県だけでなく広く認められるべきと答申されている。その答申を踏まえ、翌年からNDBの活用に関する有識者会議が厚労省保険局の検討会として設置され、検討が進められてきた。

ここでNDBの特徴を簡単に述べる。レセプト情報は電子化されたものを収集することになっており、紙で提出されたものは含まれないが、医科では95%を超えるレセプトがすでに電子的に提出されており、また薬科はほぼ100%電子化されている。歯科の電子化率は急速に上昇中で、現状70%程度であるが、今後急速に進むと予想されている。また審査支払機関の査定結果は反映されているが、査定の有無は含まれず、また保険者による査定は反映されていない。患者の氏名、生年月日の「日」、保険医療機関の所在地および名称、カルテ番号など、国民健康保険一部負担金減額・免除・徴収猶予証明書の証明書番号、被保険者証(手帳)などの記号・番号、公費受給者番号はハッシュ関数を用いて元に戻すことができない一意の識別子に変換されている。特定健診も同様で、受診者を識別できる情報は元に戻せない一意の識別子に変換されている。この変換は保険者・支払基金で行われ、さらにNDB格納前にもう一度変換される。したがってNDBに格納されたデータを保険者が見ても元には戻せない。しかしこの一意の識別子自体は同じ情報からは同じ値が生成されるため、レセプト同士あるいは特定健診同士であれば同一人のデータは同一人であることが、極めて低い確率の例外を除き、分かるように設計されている。つまり紐付け可能である。その一方で医療機関コードや健診機関コードはそのまま保持されている。

同一人の情報の紐付けで、レセプト同士、特定健診同士と述べたが、本来は特定健診とレセプトも紐付けできるべきではある。しかし現状はこの紐付けにやや難渋している。前述のハッシュは実際には2種類つくられており、一つは元の情報に名前を含み、被保険者証の記号・番号を

含まない。もう一つは元の情報に被保険者証の記号・番号を含み、名前は含まない。この二つのハッシュを用いることで、結婚などで名前が変わっても保険が変わらなければ、あるいは退職・転職などで保険が変わっても名前が変わらなければ紐付け可能である。しかしレセプトと特定健診では名前を含むハッシュはそもそも一致しない。レセプトに記載されている名前は漢字であり、特定健診はカナである。もう一つの被保険者証の記号・番号を含む値からつくられたハッシュを用いるしかない。このデータベースの設計当初はとにかく大量に生じるデータをスムーズに格納することが主課題であり、データの整合性の確保に関わる設計がやや甘く、表記の揺れなどへの対策が極めて不十分であった。レセプトを作成するソフトは歴史が古く、その仕様も安定しているが、特定健診関連の情報を取り扱うソフトは歴史が浅く、レセプトにおける審査支払い機関のような、横断的に形式をチェックする組織もないために、保険者内で処理が可能であれば、ほかのシステムと表記などに揺れがあっても実害はなかった。そのために例えば被保険者証の記号・番号に全角と半角の文字が混在したり、ダッシュ(横棒)のコードがばらばらであったりと、人が見れば、同じだと解釈できてもコンピュータ的にはまったく違うコードに見える表記が乱立することになった。本来であればこのような表記の揺れを正規化した上でハッシュ化しなければいけないが、その部分の設計が甘かったために、保険者によってはレセプトと特定健診の一致率が0%のものも存在した。現在この問題は修正中ではあるが、過去にさかのぼって修正することはかなり難しく、初期の設計の甘さは非難されるべきではあるが、現実的には今後のデータに期待するしかない。

さて、NDBがさまざまな目的の分析に有用であることは容易に想像されるが、重要な点は患者や医療機関に損害を与えることや、プライバシーを侵害することが許されないことである。万が一にもそのようなことが起これば、最悪の場合、そもそもの情報提供自体が行われなくなり、データベース自体が崩壊する。前述のように患者を直接識別する情報は注意深く追跡不可能な形式に変えられている。しかしレセプト情報や特定健診情報は複雑な情報

であり経年的に蓄積されればさらに複雑さを増す。ほかの情報と照合すれば特定できる可能性は排除できず、また、長期にわたる行動記録と見れば本人をよく知る人なら特定できる可能性は0ではない。したがって患者や受診者に関しては完全な匿名化情報とは言えず、前述の有識者会議も個人特定性に関しては分析の結果に至るまで慎重に審査し、特定性が排除されるものだけ利用が許される方向にある。また医療機関や健診機関などは直接特定できるコードがNDBには格納されている。これらは原則提供されないが、地理的要素が分析目的に含まれる場合、所在地情報は必要になる。このような目的には特定コードあるいは特定コードを地理情報に変換した情報が提供されなければならないが、その場合も特定の医療機関や健診機関などに影響がないことを、分析結果に至るまで慎重に審査され、また監査されることを条件に提供される。公共性の高い大規模医療機関では特定されること自体はその医療機関にとっては問題とは言えず、仮にほかの医療機関と比較されるような結果になったとしても、結果に客観性がある場合は比較に甘んじるべきではあるが、小規模医療機関では医療機関構成員の個人情報につながることもあり、医療機関が特定できることが患者の特定性を上げる可能性もある。さらに、分析・研究自体の公益性は重要であり、データ自体を安全に管理する能力も厳格に審査されている。全体として見れば、提供はかなり抑制的と言える。抑制的な側面の代表的な事例は探索的研究には原則提供されていないことで、例えば未知の薬剤副作用のシグナル検出のような探索的研究は現状不可能である。これに対して厚労省保険局と有識者会議は、より広い範囲に用いることができる、サンプリングデータセットを作成することを決め、2012年度に比較的単純なデータセットの提供を開始した。また、より利用性を高めた目的別データセットも厚生労働科学研究班を組織し、開発を進めている。さらに、オンサイトセンターを構築し、研究者の最終目的が安全でプライバシーの侵害の怖れがない結果の導出であれば、研究経過において探索的手法がとれることも認めようとしている。

大規模データベースの二次利用の問題点

NDBに関しては問題点も含めて説明をしてきたが、いわゆるビッグデータの場合、利用目的という観点からは目的を明確にして作成されているわけではない。二次利用は厳密な意味では個人情報保護法制上の目的外利用にあたる。1980年の経済開発協力機構(OECD)のプライバシーガイドライン⁽²⁾でも、2005年のわが国の個人情報保護3法案でも、さらには2005年のデータ保護プライバシー国際会議のモントルー宣言⁽³⁾でも、明確にすべきとしている利用目的の定義には合致しているとは言いがたい。これは2015年9月に成立した改正個人情報保護法でも同じである。改正個人情報保護法では匿名加工情報という新しいカテゴリを定義し、一定の制限下で同意なく提供可能とされているが、一方で病歴情報が、扱いにかなり強い制限のある要配慮情報に分類され、政令などの整備を待たなければ、利用条件の詳細は明らかにならない。NDBは高齢者の医療の確保に関する法律で作成が定められたものであり、この法で規定されている利用に関しては個人情報保護法制の適用外であるが、法で規定されていない利用に関して問題は解決されない。したがって現状は改正個人情報保護法でいう匿名加工情報として扱い、さらに厳しい条件を付けてはいるが、法的整合性はグレーと言わざるを得ない。あえて言えば公益性を厳しく審査しているので、公共の福祉優先の憲法の理念に沿っていると言わざるを得ないが、公益性が高いとはいえ、研究レベルのものをこの憲法原則に帰着させることには無理がある。第三者利用に関する有識者会議ではデータベースの利活用自体の法的根拠の必要性を指摘しており、法的整備が行われることにはなっているが、現時点では実現していない。端的に言えば、現状では非個人情報としない限り利用はできないと言え、実際に第三者提供においては匿名性に最大限の注意が払われている。前述のようにNDB自体は匿名化情報とは言えず、公益二次利用において実際に抽出されるデータセットで匿名性が確認されなければならない。NDBだけであれば、独自に法的根拠を与えることも不可能ではない、しかし、今年から構築される全国がん登録

のデータベースや介護保険、介護認定のデータベースなど、制度として収集されるビッグデータはほかにもあり、今後も増加することが予想される。また制度として収集されるものだけではなく、情報の本人への提供が進めば、いくつかの企業が電子お薬手帳サービスを開始しているが、コンシューマベースで収集される医療情報データベースも構築されるであろうし、いわゆるライフログに健康関連情報を組み合わせたデータベースが構築される可能性も高い。

制度として収集されるビッグデータの場合は公的資金が投入されているために、相対的に公益性の低い商用利用は単純には不可能で、集計結果などをオープンデータ化しなければ商用利用が難しいが、コンシューマベースの場合は契約などの同意を得れば商用利用も可能である。しかしクレジットカードの契約書を見ても分かるように、一般市民にとって契約書は理解しやすいものではなく、さらに医療情報の場合、本人がすべての情報の重要性を理解できるわけでもない。このような状況を考えると、契約ベースで何でも可と考えるわけにもいかない。その一方でこのような情報の利活用は、新たな産業の発展や、医薬品、医療機器をはじめ、健康医療関連商品の開発を健全に促進することも大いに期待される。

問題点の今後

大規模データベースができてきたことを述べてきたが、解決すべき問題も多い。制度として収集される公的データベースの場合は、十分な安全性を確保できるように集計処理をすれば、できる限りオープンデータ化することが重要で、この面の動きはないわけではないが、さらに促進する必要がある。また、公的データベースだけでなく、コンシューマベースのデータベースも含めて、包括的な個人情報保護法制だけではなく、健全な利用を促進し、プライバシーの侵害を実質的に防止できるデータ活用法のような制度整備を真剣に考えるべきではないだろうか。少し別の観点を加えれば、遺伝子に関わる科学

が進めば進むほど、遺伝子を直接観察しなくても、通常の医療・健康情報で遺伝形質が類推される場合が増えるが、このような遺伝に関わる情報の場合、本人のプライバシーだけが問題ではなく、親や子あるいはその子孫のプライバシーにも問題が生じる。このような第三者に影響する情報は現在の個人情報保護法制では考慮されておらず、また過剰な安全対策を求める極端な規制は科学研究や産業発展を阻害しかねない。不当な差別やプライバシー侵害は起こさないという前提で利活用は推進されなければならない。このように医療健康情報の利活用促進法のような特殊性を踏まえた法整備が望まれる。

(2015年9月 山本 隆一)

- (1) 厚生労働省保険局「医療サービスの質向上等のためのレセプト情報の活用に関する検討会」報告書, 2009
<http://www.mhlw.go.jp/shingi/2008/01/dl/s0130-16a.pdf>
 (2015年9月10日確認)
- (2) OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data
http://www.oecd.org/document/18/0,2340,en_2649_34255_1815186_1_1_1_1,00.html
 (2015年9月10日確認)
- (3) Montreux Declaration “The protection of personal data and privacy in a globalized world: a universal right respecting diversity”, 2005
http://www.privacyconference2005.org/fileadmin/PDF/montreux_declaration_e.pdf
 (2015年9月10日確認)



山本 隆一

一般財団法人医療情報システム開発センター理事長・東京大学大学院医学系研究科特任准教授 医学博士・医師

1952年 大阪市生まれ。大阪医科大学を卒業後、大阪医科大学第一内科、松下記念病院、

聖路加国際病院を経て大阪医科大学病院医療情報部助教授。2003年3月より東京大学大学院情報学環准教授、2013年5月より現職。研究内容は、医療情報の安全管理、医療におけるプライバシー保護の在り方、医療における公開鍵基盤の応用、医療従事者の権限管理モデル、医療コミュニケーション論など。著書(分担執筆)には『医療の個人情報保護とセキュリティ』(有斐閣)、『マイナンバー制度の全貌』(日経BP)など。

<役職等>

- ・日本医療情報学会 会長・理事長(2007~2010)
- ・日本医療情報学会第35回医療情報学連合大会 大会長(2015)
- ・(財)医療情報システム開発センター 理事長(2012~)
- ・内閣官房IT推進本部 評価専門調査会医療評価委員会 座長代理(2006~2008)、同座長(2009~2010)
- ・厚生労働省医療情報ネットワーク基盤検討会 委員(2003~)、同作業班主査(2003~)
- ・厚生労働省保健医療福祉分野標準化会議 委員(2005~)
- ・厚生労働省医薬品の安全対策等における医療関係データベースの活用方策に関する懇談会 副座長(2009~2010)
- ・日本医師会IT委員会 委員(2010~)
- ・日本医師会医療分野等ID導入に関する検討委員会 委員長(2015~)
- ・厚生労働省レセプト情報等の提供に関する有識者会議 座長(2010~)
- ・社会保障・税共通番号社会保障サブWG 副座長(2011~)
- ・日本学術振興会 バイオサイエンスデータベースセンター運営委員会データ共有分科会 委員(2012~)
- ・厚生労働省医療情報DBあり方検討会 座長代理(2013~2014)
- ・内閣官房IT総合戦略本部パーソナルデータの活用に関する検討会 委員(2013~)
- ・内閣官房IT総合戦略本部新戦略専門調査会医療・健康分委会 委員(2013~)
- ・内閣府医療介護情報専門調査会 委員(2014~)
- ・内閣府次世代医療IT推進協議会 委員(2015~)
- ・厚生労働省情報政策担当参与(2014~)
- ・厚生労働省医療等分野における番号制度の活用等に関する研究会 副座長(2014~)
- ・厚生労働省審議会がん登録部会 委員(2014~)



Healthcare

医療分野のデータベースの 現状と今後に向けた期待

著者 早川 和寿

株式会社日立コンサルティング 公共本部 マネージャー

はじめに

政府は、日本再興戦略や世界最先端IT国家創造宣言において、医療機関での診察情報や健康診断のデータ、日々の体重・血圧などの医療分野のデータ活用を重視し、医療費の適正化、臨床研究の促進、ヘルスケア分野の新たな産業の創出などを成長分野と位置付けている。

この動きに対し、わが国では、数年前からレセプト情報・特定健診等情報データベース（いわゆるナショナルデータベース、以下、NDB）をはじめとした医療分野の大規模データベースの構築が進んでおり、構築当初の目的以外のデータの活用や研究に向けたデータの第三者提供も、一部では視野に入ってきている。

これらの大規模データベースの発展的な活用に対しては、期待が大きい反面、現状ではプライバシー保護などについて、技術面、社会的合意形成面での課題も少なくない。本稿では、医療分野のデータベースの全体概況に触れたのち、わが国と諸外国のデータ収集・蓄積、提供状況を、請求系情報に対象を絞って紹介し、課題解決の方向性に主眼を置いて比較・考察を行う。

医療分野のデータベース構築の概況と検討対象領域の選定

本稿執筆にあたり対象分野を検討するため、医療分野のデータを大きく二つに分けた。一つは患者の診療時の医師所見や処置、検査結果、処方内容などの医療行為に関する直接的なデータである「診療データ」、もう一つは処置の内容・医療費などの請求系情報である診療報酬明細書（以下、レセプト）の「レセプトデータ」である。

まず「診療データ」は、医療機関で生成され、基本的にはそれぞれの医療機関で蓄積されている。一部の地域においては、診療の効率化を図るため、医療機関が診療データを共有する地域医療ネットワークを運用し、紹介状やサマリーを共有している。しかし、より発展的な活用をめざし、全国レベルで医療機関の診療データを一つのデー

タベースに蓄積することは、個々の医療機関から見れば業務面における必要性に乏しく、またデータ抽出の手間や利用している電子カルテなどに改修が発生する場合もあることから、現時点では実働ベースであり進んでいるとは言えない。ちなみに、現在、わが国に構築されているデータベース、あるいは構築予定のデータベースのうち、診療データを格納しているデータベースの代表例には、

- DPC⁽¹⁾レセプトに加えて検査データ、治療内容、施設情報などを含む「DPCデータベース（2017年に稼働予定）」
- 独立行政法人医薬品医療機器総合機構（以下、PMDA⁽²⁾）が構築し、10拠点の医療機関のデータを連携し、将来的に1,000万人規模のデータ収集をめざしている「医療情報データベース（提出されるのは集計表）」
- 外科を中心とした複数の学会が共同運営し、手術とその後の経過のデータが格納されているNCD⁽³⁾

などがある。

「レセプトデータ」は、社会保障の保険支払請求の仕組みの中で扱われる主要データである。データ量が膨大であることから効率化の必要性が極めて高く、以前より電子化が進められ、今では保険者や審査支払機関にそのデータが集約されている。「診療データ」と比較すると、少なくとも保険制度の確立している国においては、全国レベルで収集しやすいデータである。わが国においては、2015年時点で90億件のレセプトデータ（DPCレセプト含む）と1億2,000万件の特定健診・特定保健指導データがNDBに蓄積されており、現時点ですでに網羅的なデータベースと位置付けることができる。NDBを運用している厚生労働省にデータを提出する際に、被保険者番号や氏名などの本人情報を暗号化して識別子を生成する処理は発生するものの、保険者や審査支払機関がそもそも保有しているデータであるため、新たに収集を行う手間は発生しない。上記事情を勘案し、本稿では「レセプトデータ」を対象を絞って、収集・蓄積、第三者提供も含めた提供状況に着目して、わが国と諸外国を比較・検討することとした。

諸外国におけるレセプトデータの 収集・蓄積、提供状況

諸外国の事例については、人口規模、社会保障制度あるいは保険制度にわが国と共通点があることを考慮し、フランスと韓国を対象とした。フランスと韓国のレセプトデータの収集方法、蓄積状況、研究機関などに対する提供方法について、弊社による過去のヒアリング調査内容などを踏まえて記載する。

(1) SNIIRAM(フランス)

①概要

フランスの医療保障制度は、原則全国民が加入する疾病保険と、任意加入の補足的疾病保険の2階建てで構成されており、保険給付は償還払いが原則である。償還払いとは、受診時に診療に要した全費用を患者が医療機関に一旦支払い、その後、被保険者として請求を行い、保険者が払い戻す方式である。診療所からのレセプトは、保険者が出資して設立したレセプトオンラインシステムの運営会社であるSESAM Vitaleを経由して、保険者に送信される。

SNIIRAM⁽⁴⁾は保険者である全国被用者疾病保険金庫(以下、CNAMTS⁽⁵⁾)が運営するレセプトのデータベースであり、2002年から蓄積が進められてきた。CNAMTS、自営業者社会制度(以下、RSI⁽⁶⁾)、農業社会共済(以下、MSA⁽⁷⁾)の三つの疾病保険(保険者)に属する診療所で発生するレセプトが蓄積されている。この三つの疾病保険で、フランス在住者の96%をカバーするデータベースとなっている。また2007年からは、診断群別分類であるDRG⁽⁸⁾に基づく診療データを蓄積し、病院の評価に用いられているデータベースであるPMSI⁽⁹⁾のデータも取り込むようになり、統合的に活用されるようになった。

②データの収集方法

SNIIRAMには、基本的に病院以外の診療所や専門医から公的保険に提出されるレセプトデータが収集されており、その件数は年間12億件程度となる。

レセプトデータは各保険者からSNIIRAMへ送信する際に匿名化処理が行われ、氏名と住所は削除される。また社会

保障関係のIDであるNIR⁽¹⁰⁾は、プライバシーコミッショナーであるCNIL⁽¹¹⁾のセキュリティガイドラインで示されているFOIN⁽¹²⁾というアルゴリズムを使って暗号化される。送信側(保険者)、受信側(SNIIRAM)で2回暗号化が実行されるが、名寄せはできるようである。データ送信のための通信回線も暗号化されている。SNIIRAMに送られるPMSIのデータについても同様の暗号化処理がされており、名寄せは可能である。

③データの蓄積状況

これまで蓄積されている総データ容量は450テラバイトであり、200億件のレセプトデータが格納されている。25テラバイトの三つを含む17のデータベースで構成されている。

④データの提供状況

SNIIRAMのデータは保険者や社会保障に関わる公的機関で利用が認められており、民間企業での利用は認められていない。SNIIRAMから提供可能なデータは三つに大別できる。一つ目は、SNIIRAMに格納されたデータを基に15の集計が行われたデータベース(DAMIR⁽¹³⁾)であり、71の組織がこれを利用することができる。二つ目は、すべての被保険者の100分の1程度に当たる66万人をサンプリングし、捕捉した時系列データであり、EGB⁽¹⁴⁾と呼ばれるものである。被保険者にはアンケート調査なども行われており、これらのデータも含まれる。EGBはPMSIと連携した形で提供されており、64の組織がこれを利用できる。

三つ目は、支払情報に限定し、個々のデータセットを提供するDCIR⁽¹⁵⁾というものであり、これもPMSIと連携した形で提供されている。DCIRについては、CNAMTS、CCMSA⁽¹⁶⁾、RSIのような医療保険者、介護負担金の運用を行っている全国自立連帯基金など、6種類の組織に利用者が限定されている。

上記以外に、個々の組織に対して医療データの研究の抽出を許可する場合がある。

(2) 健康保険審査評価院(HIRA)のデータウェアハウス(韓国)

①概要

韓国では、国民健康保険公団が唯一の保険者である点は

異なるものの、国民皆保険である点、医療機関の受診は基本的にはフリーアクセスである点において、わが国と類似している。韓国で発生したレセプトは、一次審査機関である健康保険審査評価院（以下、HIRA⁽¹⁷⁾）に集約されている。HIRAでは、審査業務を行う上で集約されるレセプトデータを、医療福祉政策の立案や評価、医療の質の評価や安全性向上に活用するため、過年度分をデータウェアハウス（以下、DWH）化し、データの二次利用を可能にしている。

このようなHIRAのDWHを研究者が利用できるサービスに、レセプトデータ分析サービスがある。このサービスは、研究者がHIRAの資料処理室に設置された専用の端末を用いてDWHにクエリを発行することで、統計値や集計値を取得できるものである。このような巨大なDWHに対しては、さまざまな学術研究分野への利活用のニーズが徐々に高まっていったが、データ量が膨大であり、保存・処理のために高性能のハードウェアが要求されること、資料処理室の人手が足りないことなどの課題があった。

このような状況を受けて、HIRAでは、DWHのレセプトデータを基にサンプリングデータセットである患者標本資料を作成し、より広く研究者がデータを活用できるようにした。

②データの収集方法

韓国では、医療保険においてレセプトのオンライン化が進んでおり、2005年には医療機関のオンライン請求率が99%を突破した。2013年にはオンラインによる請求件数は約13億件となっている。

データ収集の流れは次のようになっている。まず医療機関で被保険者を診療したのち、HIRAが定める様式に沿ったレセプトデータを医療機関内の業務システムで作成する。作成されたレセプトデータは、医療機関側のシステムで内部チェックにかけ、圧縮・暗号化し、電子レセプトのポータルサイトを通じてHIRAに送信する。レセプトデータを受け取ったHIRAは、HIRA内のシステムでレセプトを審査する。患者標本資料の基となるDWHは、この一連の流れの中でHIRAに集約されるレセプトデータで構成されている。レセプトによる請求行為は、韓国の医療の社会保障制度に組み込まれているものであり、データの収集に関しての患者からの同意は取得されていない。なお、保険支払請求では、韓国の国民IDであ

る住民登録番号が用いられており、DWHの中で患者の識別子としてもこの番号が用いられている。

③データの蓄積状況

患者標本資料の基となるDWHとしてHIRAで整理されているのは、2007年から2014年までの8年分、78億件のレセプトデータであり、その総量は33テラバイトに及ぶ。また、調剤レセプトについても116億件、6.5テラバイトが蓄積されている。

④データの提供状況

HIRAでは、DWHからサンプリングデータセットである患者標本資料を作成し、より広く研究者がデータを活用できるように提供している。患者標本資料の開始当初は、韓国国内で医療サービスを受けた全患者を対象としてサンプリングしたHIRA-NPS⁽¹⁸⁾のみであった。現在は、入院患者データセットであるHIRA-NIS⁽¹⁹⁾、高齢患者データセットであるHIRA-APS⁽²⁰⁾、未成年患者データセットであるHIRA-PPS⁽²¹⁾を加えた4種類の患者標本資料が提供されている。どの患者標本資料からも、氏名や住民登録番号は削除されている。サンプリングの結果、各年の患者標本資料に同一の患者が含まれることがあったとしても、年をまたいで同一患者を同定することはできない。レセプトに記載されている診療内容・処方内容に関しても、医療機関名は所在地の市に、薬品名は薬品の主成分名に一般化される。また、希少な疾病に関しては、病名に「希少疾病」とだけ表記し、処置の内容は削除される。患者標本資料は、韓国に居住している研究者であれば、必要事項を記入した申請様式を担当部署に電子メールまたはファックスで送ることで利用申請することができる。このとき、研究者は商用目的では利用しないことを誓約する必要があり、違反した場合には民事・刑事上の責任を問われる可能性がある。また、手数料が必要であり、どの標本資料であっても1年分のデータを取得するのに30万ウォン⁽²²⁾を支払う。患者標本資料は、1年ごとに最新版が追加されることになっており、利用できる患者標本資料は年を追うごとに増えていくことになる。

レセプトデータベースの比較

ここまで、請求情報であるレセプトデータを収集・蓄積、第三者に提供している事例として、フランス、韓国の状況について紹介してきたが、わが国のNDBも含めて、(1)収集・蓄積しているデータの比較、(2)データの提供状況の比較を行う。

(1) 収集・蓄積しているデータの比較

各データベースのデータの種類、データを名寄せする識別子、データの蓄積量について図表1に整理する。

① データの種類

フランスのSNIIRAMは、診断群別分類の診療データベースであるPMSIと連携している。わが国のNDBについても、医科レセプトに加えて、DPCレセプトを格納している。また、NDBは、特定健診・特定保健指導のデータを蓄積し、レセプトデータと名寄せできるようにしている点が特徴的である。これにより、疾病の予防として実施する特定保健指導の効果をレセプトの発生状況で確認することができる。

② データを名寄せする識別子

SNIIRAMとHIRAのDWHでは、国民に対して、^{しゅがい}悉皆性、唯一無二性のあるIDを用いて名寄せを行っているが、わが国には、唯一無二性のあるIDが存在しない。わが国の公的医療保険で用いられている被保険者番号は、所属する保険者が変わる際に変更となる。そのため、被保険者番号に加えて、

氏名などの本人情報を含めて識別子を生成し、名寄せをしている。

③ データの蓄積量

人口、医療機関の数などが異なるため、一概に評価はできないが、絶対数においてはSNIIRAMが多い。

(2) データの提供状況の比較

データの提供状況を比較するため、提供目的、提供先、提供データを図表2に示す。

① 提供目的、提供先

SNIIRAM、NDBは研究機関や医療政策に関わる政府機関などに限定されている。HIRAのDWHによる患者標本資料についても研究機関(研究者)への提供のみが明らかになっているが、政策目的における公的機関の利用も行われていると推察される。

② 提供するデータ

SNIIRAMは、データの粒度の違う提供データを用意しており、それぞれに利用者の範囲を限定している。また、サンプリングデータのEGB、個別データセットのDCIRの提供においては、診療データベースのPMSIと連携したデータを提供している。HIRAのDWHによる患者標本資料では、4種類のサンプリングデータセットが用意されている。NDBについては、研究者からの申請に基づき、有識者の厳密な審査を経て提供される。単純なサンプリングデータセットの提供も開始しており、現在、研究者の利用性を高めた目的別データセットの検討が

図表1 収集・蓄積しているデータの比較

	SNIIRAM (フランス)	HIRAのDWH (韓国)	NDB (日本)
蓄積主体	保険者 (CNAMTS)	一次審査機関 (HIRA)	政府 (厚生労働省)
データの種類	レセプトデータ (診断群別分類の診療データベースPMSIと連携有)	レセプトデータ	レセプトデータ (DPCレセプト含む) 特定健診・特定保健指導データ
データを名寄せする 識別子	社会保障関連のIDであるNIRを 基に識別子を生成	国民IDの住民登録番号を利用	被保険者番号や本人情報を基に 識別子を生成
データの蓄積量	200億件のレセプトデータ	76億件の医科レセプトデータ 116億件の調剤レセプトデータ	●90億件以上のレセプトデータ (DPCレセプト含む) ●1億2,000万件以上の特定健診・ 特定保健指導データ

出典:日立コンサルティングで作成

図表2 データの提供状況の比較

	SNIIRAM (フランス)	HIRAのDWH (韓国)	NDB (日本)
提供目的	医療保険分野の研究 (商用目的は禁止)、医療保険の経営改善、医療政策への貢献	研究 (商用目的は禁止)	学術研究、医療費の適正化に資する研究
提供先	CNIIなどが認めた公的機関、非営利研究機関など	非営利研究機関など	国の機関、大学、公的研究機関に対して、有識者の審査を経て提供
提供データ	概要	<ul style="list-style-type: none"> ●集計対象が異なる4種類のサンプリングデータ (患者標本資料) を提供 ●各患者標本資料は年単位で作成され、住民登録番号を削除していることから年をまたいでの名寄せは不可能 	<ul style="list-style-type: none"> ●国の機関、大学、公的研究機関に対して、有識者の厳密な審査を経て提供 ●オンサイトセンターの利用は、申請に基づいて審査
	データ種類	<ul style="list-style-type: none"> ① DAMIR (集計データ) ② EGB (サンプリングデータ) ③ DCIR (支払情報に限定した個別のデータセット) ④ 個別抽出 	<ul style="list-style-type: none"> ① HIRA-NPS (全患者) ② HIRA-NIS (入院患者) ③ HIRA-APS (高齢患者) ④ HIRA-PPS (未成年患者)

出典:日立コンサルティングで作成

開始されている。今後、患者標本資料のように複数のデータセットが準備されると想定される。

目的や粒度に応じて複数のデータを用意するなど、それぞれ柔軟に対応している点は共通している。一方、HIRAの患者標本資料では、SNIIRAM、NDBと異なり、年をまたいで名寄せすることはできない。プライバシー保護の観点からこのような措置が講じられていると考えられるが、長期的なデータ分析などにおいてデメリットも生じているものと推察される。

今後に向けた期待

ここまで「レセプトデータ」の収集・蓄積・提供に関して、フランスのSNIIRAM、韓国HIRAのDWH、わが国のNDBを比較してきたが、それぞれの状況を踏まえ、まとめとして医療分野のデータベースに対する今後に向けた期待を記述する。

(1) 医療分野の番号、連結の仕組みの確立

医療分野の研究や、データの活用をさらに発展させるには、地域や医療機関、保険者に分散している医療分野のデータを個人単位で連結できる必要があると考える。前述のように、フランス、韓国の例を見ると、^{しつぱい}悉皆性、唯一無二性のあるIDを加工した識別子、あるいはIDそのものを利用することでデータの名寄せを行っている。例えば、SNIIRAMは、診断群別分

類の診療データベースのPMSIと名寄せも、NIRを加工した識別子を用いて実現している。このように、複数のデータベースのデータを連結させて、新たな分析用のデータを作成するには、^{しつぱい}悉皆性、唯一無二性のある医療分野で利用できるIDや複数のデータベースで異なるIDが振られている場合に、連結する仕組みが必要である。

現在、わが国では、マイナンバーとは別に医療分野で利用できる番号制度について検討が進んでいる。医療分野の番号は、2018年から符番を開始し、2020年から本格運用を開始するというスケジュールが政府から示されており、今後の動きに注目が集まっている。

(2) データ提供に際しての諸課題の克服

わが国のNDBは、現状では政策立案や公衆衛生などを目的に全国民を対象にデータを収集している。そのためデータの提供は、目的外利用になる点も踏まえ、提供先の範囲を公的機関に限り、厳密な審査を行い、「患者のプライバシー保護」を最優先することが不可欠である。

一方で、医療費の適正化、臨床研究の促進、新たな産業の創出など、医療分野のデータ活用の本来の目的を実現するため、より早く詳細なデータを柔軟な形で研究に活用したいという研究者のニーズも大きい。これらの、一見相反する「患者のプライバシー保護」と「利用者の利便性向上」という課題が厳然として存在している。

この状況に対し、現時点では、例えばSNIIRAMのように提供するデータの内容や加工によってプライバシー保護の必要性（セキュリティレベル）が異なる複数のデータを整備することや、韓国の患者標本資料のようにカテゴリ化した複数のサンプリングデータを用意することも一つの方法として参考になろう。しかし、今後のわが国の大規模データベースの活用においては、社会的合意形成を前提に、匿名化や暗号化の各種技術の適切な応用と関連法・制度の整備されることで、より高い次元で上記課題が克服されるべきである。

このように、医療分野のデータの活用に向けて、データを名寄せする仕組みやプライバシーを保護しつつ、安心・安全にデータを簡便に活用できる環境が整備されることで、研究の発展を含めた、新たなイノベーションが創出されることが期待される。

- (1) Diagnosis Procedure Combination : 診断群別分類
- (2) Pharmaceuticals and Medical Devices Agency
- (3) National Clinical Database
- (4) Systeme national d'information inter-regimes de l'assurance maladie
- (5) Caisse nationale de l'assurance maladie des travailleurs salaries
- (6) Regime Social des Independants
- (7) La Mutualite sociale agricole
- (8) Diagnosis Related Goups
- (9) Programme de medicalisation des systemes d'information
- (10) Numero d'Inscription au Repertoire
- (11) Commission nationale de l'informatique et des libertes
- (12) Fonction d'Occultation des Informations Nominatives
- (13) depenses de l'assurance maladies inter-regimes
- (14) L'echantillon generaliste des beneficiaires
- (15) Le datamart de consommation inter-regimes
- (16) Caisse Centrale de la Mutualite Sociale Agricole
- (17) Health Insurance Review & Assessment service
- (18) National Patient Sample
- (19) National Inpatient Sample
- (20) Adult Patient Sample
- (21) Pediatric Patient Sample
- (22) 日本円で約 30,000 円 (100 韓国ウォン = 10 円換算)



早川 和寿


株式会社 日立コンサルティング
公共本部 マネージャー

立命館大学産業社会学部を卒業後、2004年に日立製作所に入社し、2006年から日立コンサルティングに出向、2015年6月に転籍。

日立コンサルティングでは、医療保険や介護保険、医療ICTなどの調査研究、実証実験案件を中心とした、ヘルスケア分野を担当。

記載内容(所属部署・役職を含む)は制作当時のものです。

HITACHI
Inspire the Next

 **株式会社 日立コンサルティング**

〒102-0083 東京都千代田区麹町二丁目4番地1 麹町大通りビル11階
TEL : 03-6779-5500 (代)
URL : <http://www.hitachiconsulting.co.jp/>
問い合わせ : info@hitachiconsulting.co.jp