

特集 医療データベース活用の現状——ビッグデータを臨床研究に活かす

展望：医療ビッグデータのさらなる活用に向けて

一般財団法人医療情報システム開発センター

山本 隆一

「ビッグデータ」とは何かと言えば、絶対的な定義があるわけではない。Wikipediaによれば、「市販されているデータベース管理ツールや従来のデータ処理アプリケーションで処理することが困難なほど、巨大で複雑なデータ集合の集積物を表す用語」とされている。「市販されている」「従来の」という言葉は相対的なものであり、簡単に言えば「集積、管理、分析が困難なほど大きなデータベース」ということであろう。この定義に従えば、ハイスピードシーケンサで大量のゲノム分析を行う場合を除けば、医療・健康分野でわが国に存在するビッグデータは、厚生労働省が「高齢者の医療の確保に関する法律」に基づいて運用しているレセプト等データベース（National Claim insurance and health checkup database：NDB）が唯一相当する可能性がある。ただ、このデータベースもIT技術の進歩により、「集積、管理、分析が困難なほど」とは早晩言えなくなることは確実である。本稿のタイトルに「ビッグデータ」という言葉を用いたが、ここでは少し定義を緩めて、「母集団に対して一定程度の網羅性をもつ大規模な医療・健康データベース」として論を進めていくことにしたい。

わが国の医療の情報化は1960年代から始まった。レセプトコンピュータ（レセコン）、医事コンピュータ（医事コン）と呼ばれる比較的単純なシステムで、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、オーダエントリシステムが普及を始めた。単なる事務経費の削減だけではなく、待ち時間の短縮などの効果

もあったが、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年ころのわが国は世界でも最も医療の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、医療機関の経済的動機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た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、まだ十分ではないと言わざるを得ない。

情報の処理から情報の利活用へ

医療情報システムの発展過程が主に経済的動機で進んだために、というだけではなく、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。たとえば米国では、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た。さらにCMS

Prospects: For the Further Use of Health Big Data

Ryuichi Yamamoto: General Incorporated Foundation Medical Information System Development Center

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開放している。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もつ国は多い。それに比してわが国は、国などが行う限定的な調査はあったものの、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てきた。1つは前述のNDBであり、電子化された診療報酬請求明細（レセプト）と、特定健診・特定保健指導のデータが、すべて格納されている。わが国のレセプトの電子化率は請求件数ベースでは医科では95%を超え、薬科では100%、歯科も急速に増加しており、すでに80%を超えようとしている。つまり、わが国で行われる保険診療を網羅的にみることができるデータベースと言える。DPC (Diagnosis Procedure Combination) の様式1やEFファイルの情報（診療明細情報・行為明細情報）を格納するデータベースも存在し、また、2016年から全国がん登録が実稼働を始める。病院はすべて登録が義務化されており、診療所は手挙げ方式ではあるが、格納されるデータががんの初期診断初期治療が主体であり、わが国で発生する「がん」患者からみれば、かなり網羅性は高い。さらに厚労省とPMDA（医薬品医療機器総合機構）が進めている医療情報データベース整備事業はすでに構築自体はほぼ終了しており、数百万人の比較的详细な臨床情報を含むデータベースで、平成30年の本稼働に向けて整備を進めている。このようにさまざまなデータベースの整備が進む一方で、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。

NDB

前述のように、「高齢者の医療の確保に関する法律」のなか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ている。本原稿執筆時点で、レセプトで100億件、特定健診特定保健指導データで1億2000万件を超えている。

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等の活用に関する検討会」の報告が、2008年2月にとりまとめられた¹⁾。そのなかで、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、2010年から、NDBの活用に関する有識者会議が厚生労働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、医科では枚数ベースでは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70%程度であるが、急速に増加している。また、審査支払い期間の査定結果は反映されているが、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、公費受給者番号は削除されており、保険者番号・記号番号・生年月日・性別でハッシュ関数を用いて同一人のデータを連結可能としている。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。

さて、このようなデータベースがさまざまな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不当な損害や不快感を与えないことで、特に患者のプライバシーを侵害することは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・保健指導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。他の情報と照合し、あるいは長期にわたる行動記録とみれば、本人が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、医療機関などや健診機関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などや健診機関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。もっともここまで医療機関側に配慮しているのは、国際的にみれば日本特有の事情かもしれない。フランスでは医療機関の特定は問題にされない場合が多く、米国ではかなり以前から病院は特定されても問題にならない。また、最近までクリニックや個々の医師に関しては特定されてこなかったが、数年前に公益が優先されるという司法判断が出され、CMSのデータ利用でも医師個人まで特定が容認されるようになっていく。

NDBのデータ提供にあたっては、当然ではあるが、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ることになる。これ以外にも審査項目はあり、提供体制の充実も検討されているが、当面は慎重にならざるを得なく、抑制的ではあるが、提供が開始されている。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、たとえば未知の薬剤副作用のシグナル検出は不可能である。ただ、1カ月分のレセプトに関して、外来1/100、入院1/10のサンプリングデータセットがつくられ、探索的にも利用可能となっている。頻度の低い病名、医薬品、医療行為はダミー値に置き換えられているが、提供も早く、安全管理基準もゆるいために利用はしやすい。これまでは毎年10月分（調剤は翌月分も含む）だけであったが、2016年からは1月、4月、7月の

サンプリングデータセットも提供が予定されている。また、同じくサンプリングではあるが、複数年次にわたり同一人を連結した、基本データセットも試作され、現在は試行中ではあるが、2016年には提供が開始される予定である。

さらに、オンサイトリサーチセンターも試験稼働が始まっている。これは厳重に安全管理を施されたリサーチセンター内で比較的自由にデータの検索や処理を研究者に許容する施設で、データを持ち出すためには再申請が必要であるが、利用の制限は少ない。研究者は研究環境のセキュリティ強化に腐心することなく、NDBを利用することが可能になる。実運用されれば利用の拡大につながることが期待される。また、現在はまだ準備中であるが、2016年前半にはNDBオープンデータが公開される。単純集計情報ではあるが、多数の都道府県別あるいは性年齢階層別の集計情報が公開され、たとえば医薬品の使用状況も含まれている。だれでも自由に利用でき、毎年更新され、また新たな集計のリクエストも受け付ける予定である。

大規模データベースの二次利用と法規制上の問題点

前述のように、わが国にもようやく医療・健康情報の大規模データベースが整備されてきた。全国がん登録は法律に基づいて構築されるために個人情報保護法制上の問題は生じない。NDBも法律に基づいて構築されているが、法に既定されている利用目的は限定的で、データベースの価値を発揮できるものではない。また、法律に基づいているとは言え、不用意な被害を患者等に与えては大きな問題になる。

一方で、大規模データベースを利用した研究は、介入研究と異なり、健康被害を与える恐れはない。考慮すべき問題の大部分はプライバシーの侵害や、プライバシーとは言えないまでも、想定外の影響を患者や情報提供機関に与えることである。また、利用しないことにより、得られるべき知見を失うことが、国民的な損失になりうる。いずれのデータベースでも利活用に関して国民的な合意が得られているという状況ではなく、議論が積み重ねられている。2015年9月に個人情報保護に関する法律が改正され、個人情報保護委員会は特定個人情報保護委員会を改組し、用意されたが、法自体が未実施であり、

政令や指針もまだ出揃っていない。したがって、あくまでも現時点での論点整理と最近の動向の紹介になる点はやむを得ない。

改正個人情報保護法

改正個人情報保護法の変更点は数多くあるが、大規模データベースの二次利用という観点からは匿名加工情報の導入と、「病歴」が要配慮情報の例として明記された点が大きな影響を与えると考えられる。

「匿名加工情報」は完全な匿名化とは言えないが、通常は特定が著しく難しい状態に加工された個人情報で、再特定を禁止し、安全管理に努めれば同意なく第三者提供が可能とされている。この状態にすることで医療機関などから情報を出す際に同意が不要になり、大規模データベースの二次利用という観点からみれば、ポジティブな要素である。ただ、匿名加工の手法や有効性は、個人情報保護委員会のもとで検討されることになる。法案の国会審議の経緯などをみると、かなり厳しい匿名加工になる可能性があり、時系列の複雑な情報を扱う医療では、匿名加工にあたらぬ情報を扱う必要がでることも予想される。たとえば、NDB や NCD (National Clinical Database) が匿名加工情報に相当しないとすると影響は大きい。一方で、匿名加工情報の定義があまりにゆるくなると、患者らが予期しない利活用が頻発する恐れもある。本来は利用目的の公益性や利用者の信頼性を考慮して、匿名加工の基準を決めるべきであろう。

「病歴」がどこまでの医療情報を示すのかはいまだに明らかではないが、要配慮情報は明確な同意を得ないかぎり収集できない情報であり、利用目的の変更なども厳しく制限される。大規模データベースは目的を明確にして作成されているわけではない。したがって、すべての学術上の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、あるいは目的の変更にあたる。大学病院などでは掲示されている利用目的に研究・教育が含まれていることが多く、個人を特定しない利用としては、いわゆる包括的同意は得られていると考えることもできるが、1980 年の 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、明確にすべきとしている利用目的の定義に合致してい

るとは言いがたい。まだ確定してはいないが、EU データ保護規則案でも、明確な同意とは扱われないように思われる。他国の規則はわが国とは関係ないようであるが、個人情報保護は国際的な調和が強く求められる分野であり、影響は受ける。

一方、現状の個人情報保護法制下で包括同意とよく似た同意の取り方として、厚生労働省が発出している「医療介護事業者における個人情報の適切な取り扱いのためのガイドライン」(厚労省ガイドライン)³⁾では、地域医療連携における紹介情報に関する問い合わせ、その問い合わせに答えること、専門医への相談、および家族への病状説明の 4 つの場合に限定して、黙示の同意を行うことを認めている。これはやや曖昧と言わざるを得ない利用目的をあらかじめ掲示などで周知し、詳しい説明を求めることができること、いつでも意見を変えることができること、および、何も表明しなければ同意をしたとみなすこと、を明示したうえで、何も表明がなければ同意を得たとみなしてよいとされている。このように何も言わなければ同意とみなされる場合に、任意の時点で非同意にすることを opt out と呼ぶことがあり、また非表明による同意を得ること自体を opt out で同意が得られたと呼ぶ。厚労省ガイドラインでは「黙示の同意」と呼んでいるが、opt out による同意と大きな違いはない。ガイドラインで明記されている 4 つの場合は医療現場では頻繁に行われることではあるが、最初の同意に関する意見表明(つまり何も言わない)の機会(通常は初来院時)では、実際にそのような行為が行われるかどうかは不明であり、情報が提供されうる相手も特定されていない。そのような行為が必要になった時点で初めて明確になる。このように同意表明機会には目的は曖昧にしか限定されず、利用機会が実際に生じた時点で明確になる事象への同意を「広義の同意(broad consent)」あるいは「未来への同意(future consent)」と呼ぶことがあるが、broad consent と opt out の組み合わせを黙示の同意として、個人情報保護法制上の同意の一形態ととらえていることになる。改正前の現行法制では厚労省ガイドラインに記載されている 4 つの場合では問題は生じていない。

ここで問題を 2 点に整理することができる。1 点目は改正個人情報保護法で、これまでの方針が適用

可能かどうか。2点目は厚労省ガイドライン上、4つの場合を明記して「黙示の同意」の適用を認めているが、これを大規模データベースの二次利用に拡張することが可能かどうかである。

さて1点目であるが、少なくとも厚労省ガイドライン上の4つの場合に関しては、これを明確な同意に置き換えることは現実には不可能で、現在進行中のいくつかの会議で、特段、変更なく運用できる方向性が出ている。よく知られているように、「病歴」が要配慮情報（機微情報）と既定され、法の23条2項でopt outによる収集は禁止されているが、政令などで適切に手当されるものと考えたい。2点目の「黙示の同意」を大規模データベースの二次利用に拡張可能かどうかは、現時点では明確にすることはできない。改正前の現行個人情報保護法に基づく厚労省ガイドラインで、わざわざ4つの場合に限定して「黙示の同意」を使うことを可としていることを考えると、そのまま適用することは難しいと思われる。

一方で、同意のあり方に関しては、少なくともわが国の個人情報保護法制に関しては詳細に議論されていない。わが国では法律レベルで細目にわたって厳しく既定すると、運用が難しくなる場面が増えるために、運用上の融通の余地を残しておくことが多いことに由来するのかもしれないが、解釈の幅が存在する。だからと言って、勝手な解釈をだれがしてもよいということにもならない。これから議論される指針等で明確に定義し、パブリックコメントなどで一応のコンセンサスを得ることが最低限必要であろう。

法規制上の問題点と外科領域の臨床研究の今後

大規模データベースの現状を紹介し、法規制上の問題点を考察したが、本質は医療・健康情報の本質的にもっている公益性を正面から取り上げた法体系がわが国に存在しないことにある。プライバシー保護は人権問題であり、少しもおろそかにできないことは言うまでもないが、医療・健康情報の公益利用

は、それなしに医学・医療の発展や、有効な施策の決定はできないものであり、研究者のプライバシー侵害の意図があるとは思えない。プライバシー保護を確保しつつ、社会の利益を目指す二次利用が合理的な手続きで進めることができる法体系の整備は医学にとって死活問題とも言える。また、ゲノム情報など遺伝に関する情報の取り扱いはさらに難しい。個人情報保護法制は本人、情報収集者にかかわる法規制であるが、遺伝する情報の場合、本人以外に血縁者の一定の確率のもとではあるが当事者である。しかし、個人情報保護法制では、本人と情報収集者以外は当事者ではないために、被害が生じても扱うことができない。そのため、いくつかの先進国では遺伝子情報による差別禁止を法制化している。幸い、個人情報保護法が改正され、実施は2017年度と予想される。そして、実施までの間に、関連政省令や、指針の整備で活発な議論が進められると予想されるし、内閣官房では新たに医療健康の利活用に関する制度整備も検討されている⁴⁾。これからの1年間で医療において大規模データベースの利活用を進めるうえで、正念場と考えられる。

文 献

- 1) 厚生労働省保険局「医療サービスの質向上のためのレセプト情報等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12月15日確認)
- 2) 22OECD 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12月15日確認)
- 3) 厚生労働省「医療・介護関係事業者における個人情報の適切な取り扱いのためのガイドライン」, 2010.
<http://www.mhlw.go.jp/topics/bukyoku/seisaku/kojin/dl/170805-11a.pdf> (2015年12月15日確認)
- 4) 内閣官房「情報通信技術 (IT) の利活用に関する制度整備検討会・中間整理」, 2015.
https://www.kantei.go.jp/jp/singi/it2/senmon_bunka/pubcom/chuukanseiri.pdf, (2015年12月15日確認)