

特集 : CKD Big Data

医療ビッグデータ活用の現状と展望

Current situation and future of practical use of big data in medical field

中島直樹

Naoki NAKASHIMA

はじめに

さまざまな社会活動によって蓄積された莫大なデータを real world data と言い、世界中の多くの分野でその活用が注目されている。医学においても「データ駆動型医学研究 (Data Driven Medical Study : DDMS)」と呼ばれる手法が臨床研究目的に用いられることが増えてきた。つまり、何らかの別目的で莫大に蓄積した電子データを二次利用するものである。例えば電子カルテデータの疫学利用である¹⁾。

従来の疫学手法は、その科学的信頼性や研究自由度の高さから、今後もますます重視されるべきものである(図1上)。しかしながら、その欠点である「解析に必要なデータ蓄積に時間とコストを要すること」は、医療安全の確保や、社会保障課題の改善などの多彩かつ喫緊の社会課題には必ずしも適さないことも多く、この欠点を補う研究手法として、DDMS が注目されているわけである(図1下)。

日本における国家規模の DDMS 基盤の代表例として、レセプト情報・特定健診等情報データベース(以下、NDB)²⁾

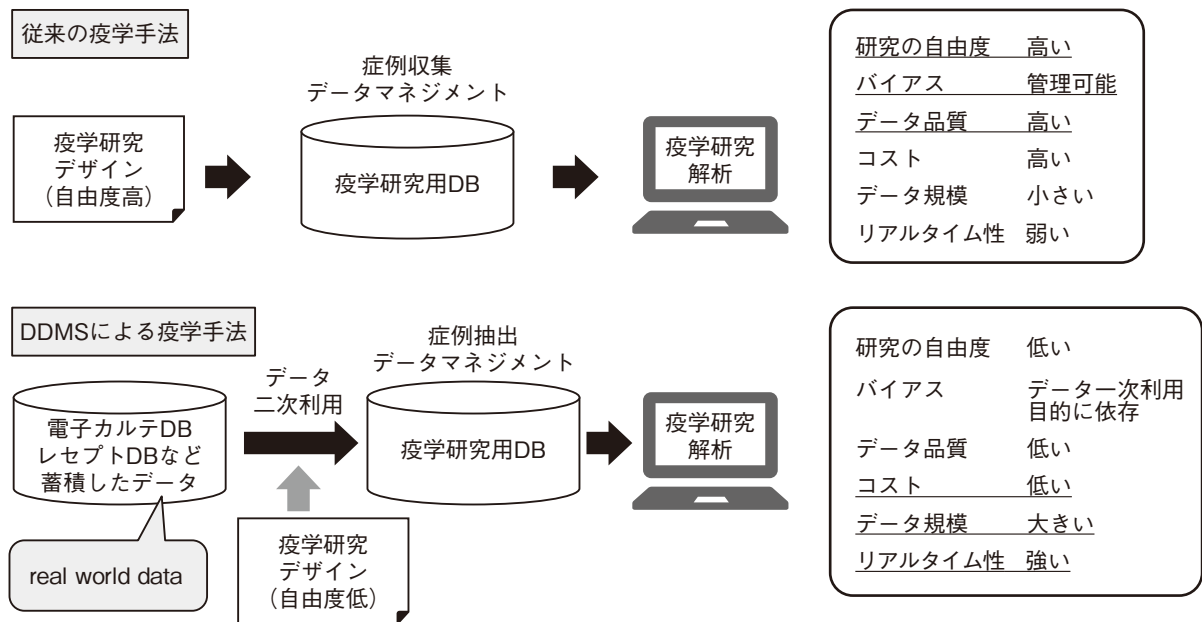


図1 従来の疫学手法とデータ駆動型医学研究手法(DDMS)による疫学研究
下線はそれぞれの長所

や医療情報データベース(以下、MID-NET)³⁾があげられる。

本稿では、医療情報の電子化の状況、DDMSの行政の戦略、DDMSの具体例、および将来の展望について述べたい。

日本における医療情報の現状

社会全体の情報化・電子化の進展を基礎として、健康・医療分野でも電子カルテの普及、レセプトオンライン化、特定健診制度の創設などを軸に、電子化がようやく進展してきた。その結果、保険者を中心に標準化された莫大な電子データが日々蓄積されている。

国民皆保険の下で、フリーアクセスが保証されている日本において、保険医療施設(病院、医科診療所、歯科診療所、調剤薬局)は合計約24万施設に上る一方、1億2,600万人の国民とn対1で紐づく医療保険者は3,000以上存在する。国民は、年間約20億回保険医療施設を利用し(月複数回は除く)、各施設で医療情報を電子カルテなどに記録されると同時に、レセプトを発生させる。さらに保険者が主体で実施する特定健診では、対象の40歳から74歳のほぼ50%に当たる2,600万人が年1回の特定健診をほぼ継続的に受診している。オンライン化が進んだレセプトは、診療報酬請求のために保険者へ個票データとして送られ、特定健診結果とも突合される。この仕組みは2015年度から「データヘルス計画」として各保険者が活用し、被保険者の健康改善や重症化予防を促している⁴⁾。また、このようにして保険者に集まるレセプトおよび健診データを匿名化し、厚生労働省で集約したものがDDMSの代表事例のNDBであり、世界でも有数の規模および複雑性を持った保健医療領域のビッグデータであることは疑いが無い。NDBについては後で詳細に述べる。

このように保険者の保有する健診情報やレセプト情報の標準化が進んだ一方で、各医療施設が保有する医療情報の標準化の遅れが指摘されてきた。潜在的な知識量としてはレセプトよりもはるかに多いデータが、電子化された後でさえも医療施設内に限局した可視化や解析にしか使えない不利な状況で長く存在してきた。この課題に対して、厚生労働省や日本医療情報学会などは取り組みを続け、種々の厚生労働省標準規格を提案・策定してきた⁵⁾。近年、これらの標準規格を組み合わせたSS-MIX2と呼ばれる電子カルテに外付けする標準的なデータのストレージによる運用指針が普及するに至り、レセプトや特定健診以外のDDMS環境が整備され始めたのである。SS-MIX2導入施設は年々増加し、2017年3月には大規模病院を中心に1,114施設に

至っている⁶⁾。現在は、蓄積された標準コードの付与が不十分であることなどによるデータ品質の低下という大きな課題があるものの、今後、DDMSの進展と相まって日本の医療情報基盤として整備されることが期待される。

政策的な戦略

特定健診制度が実施された2008年以降の約10年間、厚生労働省を中心に、データヘルス戦略が進められてきた。当初こそ「データヘルス」という言葉はなかったが、既にその骨格は組み込まれており、その後のNDB活用開始(2013年)、データヘルス計画(2015年)、データヘルス改革推進本部設置(2017年)と続いて、ますます加速している⁴⁾。

また、2015年には内閣官房を中心に次世代医療ICT基盤協議会が設置され、文部科学省、厚生労働省、経済産業省、総務省などの行政総力をあげた医療情報の基盤構築が開始された⁷⁾。DDMSに代表される医療データの二次利用には、大規模データの解析技術や情報セキュリティなどの技術的な課題のほかに、倫理的・法的課題など多くの課題が存在するが、同協議会にワーキンググループが設置されると同時に、日本医療研究開発機構(以下、AMED)を中心とした多くの研究事業により課題解決や事業化が試みられている。

さらには、従来の「日本再興戦略」から引き継がれた「未来投資戦略2017(2017年6月発表)」では、全分野の具体的施策の筆頭に「健康・医療・介護」があげられ、データ利活用基盤の構築をはじめとしてさまざまな推進することが明記された。特に在宅での日常データやロボット介護などの多くの情報のデータ化が予想され、現在の医療情報の範囲をはるかに越えた莫大なデータが将来蓄積・活用されることが期待される⁸⁾。

標準的な項目セット

特に糖尿病や慢性腎臓病(以下、CKD)など長い病歴を持つ疾患のデータベースを構築し、患者個別のデータの活用を推進するためには、あらかじめ項目の標準化を図っておくことが重要である。

糖尿病、高血圧症、脂質異常症、CKDの標準的な項目セットについては、日本医療情報学会、日本糖尿病学会が2011年度以降に策定を開始し、その後、日本腎臓学会、日本高血圧学会、日本動脈硬化学会、日本臨床検査医学会の計6つの臨床学会が、協調して2つの項目セットを策定した。まず、4疾患のいずれの用途のデータベースにも不可

欠な「ミニマム項目セット」を項目名、粒度、単位などの表現を標準化して策定し、それを前提に、4疾患の自己管理項目セットを策定したのである⁹⁾(表)。

疾患自己管理項目セット集の活用事例としては、日本腎臓学会が開始した「包括的CKD データベース構築事業(J-CKD-CB)」や「慢性腎臓病進行例の実態把握と透析導入回避のための有効な指針の作成に関する研究(REACH-J)」、国立国際医療研究センターと日本糖尿病学会が開始した「診療録直結型全国糖尿病データベース事業(J-DREAMS)」、日本医師会が進める「診療所糖尿病症例データベース(J-DOME)」などの疾患コホート研究がみられる。これらも DDMS の好例であろう。

また6臨床学会は、2016年度からは個人管理型健康記録(Personal Health Record: PHR)の推奨設定を策定した。これは、スマートフォンなどで実現したPHRに疾患自己管理項目セットを導入することを前提として、個人の健康リスク状態で階層化する、緊急性の高い数値によってアラートを発する、誤入力を検知する、あるいは入力を促す、などのための閾値や適正な入力間隔を設定したものである。今後、PHR利用者は、携帯電話のプロバイダーのようにPHR事業者を変えることが想定されるが、PHRのように長期のデータを保存・活用するサービスにとっては、データの保存性や相互運用性は特に重要である。項目セットのような標準規格が実装されることにより、PHRの保存性・相互運用性の確保が高まる。また、今後はPHRと他のデータベースとの連携やPHRがDDMSのデータ源ともなることを含め、PHRを越えて効率的にデータが蓄積あるいは活用されることが期待される。AMEDのPHR研究事業のなかでも6臨床学会が協力して医療情報システム開発センターが開発運用をしているPHRや、日本腎臓学会、日本糖尿病学会、日本医療情報学会などの研究グループが開発・運用している糖尿病PHRのDialBetics[®]などでは、疾患自己管理項目セットおよびその推奨設定を実装している¹⁰⁾。

NDB や MID-NET の実際

NDBは、前述したように世界最大規模のDDMS基盤と言ってよい。合計で年間約20億件のレセプトが3,000を超える保険者から送られてきている。1つのレセプト当たり多くの場合10件を超える診療行為データが記録されている。2009年度のレセプトデータから使用可能であるため、現在では最大で数千億件のデータを用いた経年的な解析が可能である。さらには、年間2,600万件の特定健診データ

と個票レベルでの突合が可能のため、検査結果などのアウトカム情報を解析に取り入れることも可能なのである(ただし、突合可能なのは2015年度以降のデータ)。欠点は、保険病名などのデータ品質が高いとはいえない、医療扶助などの公費データが含まれない、死亡などの保険資格喪失データがない、抽出に時間がかかる、などであり、今後の改善が望まれる。研究者が使用するためには、審査承認を受けなければならない。運用を円滑にする目的で、厚生労働省、東京大学、京都大学にオンサイトセンターが設置されている²⁾。

MID-NETは、医薬品の安全管理の科学性・信頼性を向上させるために、薬剤副作用の自発報告システムのみならず、処方全件サーベイランスを実施することを目的として、厚生労働省と医薬品医療機器総合機構(以下、PMDA)が2011年度に始めたDDMSの基盤的事業である。これまではPMDAおよび協力医療施設内での試行運用を行ってきたが、2018年度からは、協力医療施設以外の研究者や製薬企業などの有償での第三者活用を含めた本格稼働を開始する予定である。10の協力医療施設のレセプト情報とSS-MIX2を用いて、PMDAの複合施設統合データ処理センターを中心にデータ抽出・解析を行う(図2)³⁾。

現在400万人規模のデータベースであるが、リアルタイム性が強く、またデータバリデーションを重視して実施しており、データ品質が高いことが特徴である。製薬企業にとって販売開始後の副作用検証などを実施するには、前向きな比較対照試験などを組むよりも、DDMSであるMID-NETシステムを用いるほうがはるかに早く低コストで信頼性のある解析結果を入手することができるため、一定の需要が予測される。

次世代医療基盤へ

超少子高齢社会を迎えた日本では、医学の発展、費用対効果の向上、周辺産業の新興を促すためには、DDMSを活性化させる必要がある。そのDDMSの活性化は健康・医療・介護分野におけるデータ蓄積の質・量の向上が前提であるが、データ発生源であるPHRなどの情報取り扱い事業の運用コストをどのように捻出するか、つまりビジネスモデルをどのように成立させるか、は共通した課題である。

その一方では個人情報の保護による制約がある。電子化された個人情報は漏洩リスクが高く、特に医療情報は狙われやすい。そこで、2017年5月には個人情報保護法が改正された¹¹⁾。それに伴い、医療介護のガイダンスも公

表 疾患ミニマム項目セット集(グレーの項目)および疾患自己管理項目セット集(○がついた項目)

ID	項目	単位・表現	糖尿病 自己管理項目セット			高血圧 自己管理項目セット			脂質異常症 自己管理項目セット			CKD 自己管理項目セット		
			医療機 関から	健診な どから	家庭 から	医療機 関から	健診な どから	家庭 から	医療機 関から	健診な どから	家庭 から	医療機 関から	健診な どから	家庭 から
1	身長	cm	○	○		○	○		○	○		○	○	
2	体重	kg	○	○		○	○		○	○		○	○	
3	収縮期血圧	mmHg	○	○		○	○		○	○		○	○	
4	拡張期血圧	mmHg	○	○		○	○		○	○		○	○	
5	総コレステロール	mg/dL	○			○			○			○		
6	HDL コレステロール	mg/dL	○	○		○	○		○	○		○	○	
7	喫煙	あり, なし, 過去にあり	○	○		○	○		○	○		○	○	
8	血清クレアチニン	mg/dL	○			○			○			○		
9	尿蛋白	-, ±, +, 2+, 3+ 以上	○	○		○	○		○	○		○	○	
10	血糖	mg/dL	○	○		○	○		○	○				
11	糖尿病診断年齢	10 歳未満, 10 歳代, 以後 10 歳毎 80 歳代以上まで, 不明	○											
12	HbA1c(※ 1)	%	○	○								○	○	
13	ALT	IU/L	○	○					○	○				
14	網膜症	あり, なし, 不明	○											
15	高血圧診断年齢	10 歳未満, 10 歳代, 以後 10 歳毎 80 歳代以上まで, 不明				○								
16	血清カリウム	mEq/L				○						○		
17	心電図異常	あり, なし, 不明				○								
18	中性脂肪	mg/dL	○	○		○	○		○	○		○	○	
19	脂質異常症の診断年齢	10 歳未満, 10 歳代, 以後 10 歳毎 80 歳代以上まで, 不明							○					
20	冠動脈疾患の既往	あり(造影検査), あり(その他検査), なし, 不明							○					
21	CKD 診断年齢	10 歳未満, 10 歳代, 以後 10 歳毎 80 歳代以上まで, 不明										○		
22	血清アルブミン	g/dL										○	○	
23	血尿	-, ±, +, 2+, 3+ 以上(非肉眼的), 肉眼的										○	○	
24	尿アルブミン/クレアチニン	mg/gCr	○											
25	AST	IU/L	○	○										
26	腹囲	cm		○			○			○				
27	尿糖	-, ±, +, 2+ 以上	○	○										
28	γGTP	IU/L	○	○										
29	神経障害	あり, なし, 不明	○											
30	歯科定期受診(※ 2)	あり, なし, 不明	○											
31	尿酸	mg/dL				○						○	○	
32	家庭血圧(収縮期)	mmHg						○						
33	家庭血圧(拡張期)	mmHg						○						
34	腎不全家族歴(※ 3)	あり, なし, 不明										○		
35	尿蛋白/クレアチニン比	g/gCr										○	○	
36	尿蛋白(1 日量)	g/日										○	○	
37	血清総蛋白	g/dL										○	○	
38	尿素窒素	mg/dL										○		
39	Hb	g/dL										○	○	
40	シスタチン C	mg/L										○		

(※ 1)HbA1c : NGSP 値, (※ 2)歯科定期受診 : 年 1 回以上, (※ 3)腎不全家族歴 : 2 親等以内の透析, 腎移植, 腎不全

□ 糖尿病のミニマム項目セット, □ 高血圧症のミニマム項目セット, □ 脂質異常症のミニマム項目セット, □ CKD のミニマム項目セット
(文献 9 より引用, 改変)

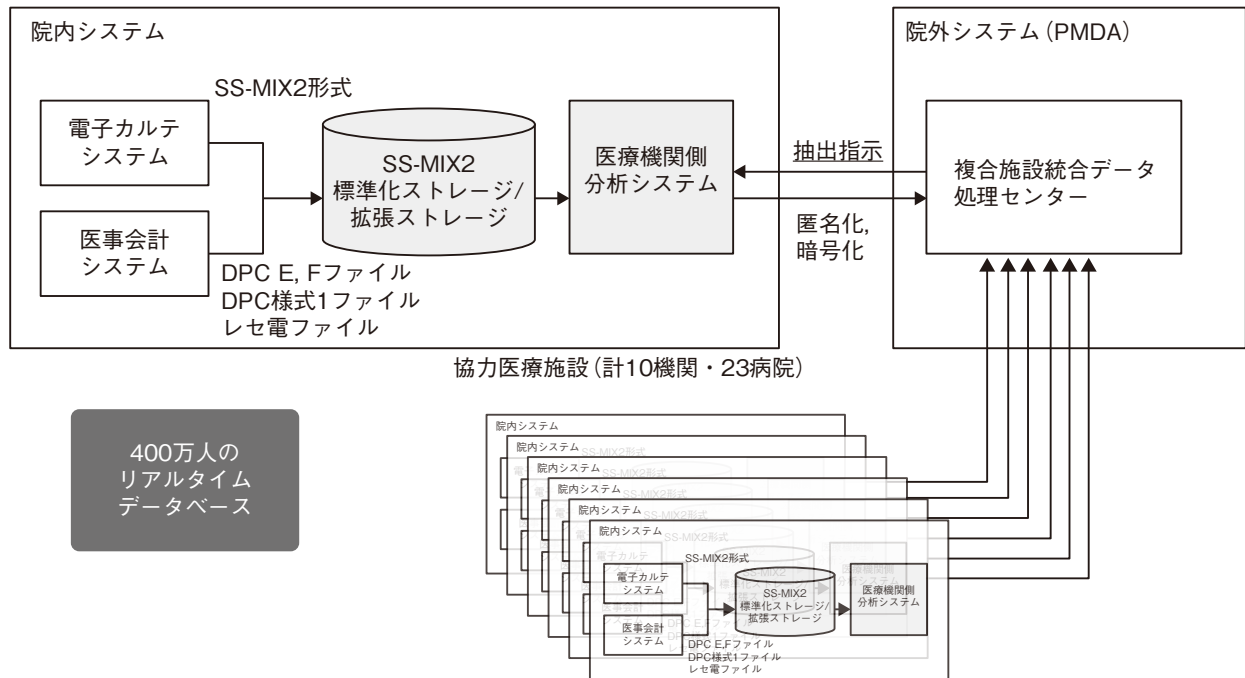


図2 MID-NET事業の概要

複合施設統合データ処理センターが協力医療施設のシステムへ抽出指示を出す。

表され¹²⁾、結果的にデータの二次利用に対しては従来よりも厳格化された。

そこで、改正個人情報保護法を前提としながらDDMSを推進するために、2017年4月には次世代医療基盤法が制定され、1年後には実施されることとなった¹³⁾。これは、データの取り扱い技術を保有する「認定機関」が、個人情報を定められた手法で匿名加工情報とし、有償で行政や研究者、企業などに提供できる仕組みの確立である。これによって、情報取扱事業者のビジネスモデルが成立しやすくなることが期待される。ただし、詳細な運用については未定であり、また認定機関の候補は数件あげられているが、実際にどの程度の規模で始まるかはまだ不明である。

あることを考慮した患者同意取得など倫理ガイドラインの遵守や、扱うデータの質や解析法の適正性などの正しい理解が必要である。同意取得を無視したり、質の低いデータや誤った解析法を使った結果を利用して都合の良い解釈を行うことなどは、社会的・科学的な信頼性を失墜させ、DDMSの進展や社会課題の解決に水を差すこととなることを肝に銘じなければならない。

利益相反自己申告：申告すべきものなし

文献

1. 武藤正樹. リアルワールド・データ活用の現状と課題. 月刊保険診療 2014; 69: 58-59.
2. 満武智裕. レセプトビッグデータ解析の現状と将来. 実験医学 2016; 34: 799-804.
3. 石黒智恵子, 宇山佳明. PMDAにおける医薬品の安全性評価を目的とした医療情報データベースの薬剤疫学的利活用: MIHARI Project と MID-NET. 薬理と治療 2016; 44(Suppl 1): 5012-5016.
4. 中島直樹. データヘルスのあゆみと対糖尿病戦略. プラクティス 2017. in press.
5. 本田憲業. 保健医療情報分野の厚生労働省標準規格の紹介.

おわりに

本稿では、主としてDDMSという形で、蓄積した健康・医療情報の二次活用について述べた。これらは、情報化の進展タイミング、社会課題の解決のための国民(患者)や行政からの期待、医療者や保険者の業務や経営改善の願い、製薬業界や新興健康産業などの業界の適正な活性化の必要性など多くの要望が一致しており、今後急激に進むことだろう。しかしながら、研究者には、データの目的外使用で

- 臨床核医学 2011；44：92-93.
6. 日本HL7協会. 第62回HL7セミナー資料. http://www.hl7.jp/docs/62seminar_1_HL7.pdf
 7. 健康医療戦略推進本部. 次世代医療 ICT 基盤協議会. http://www.kantei.go.jp/jp/singi/kenkouiryou/jisedai_kiban/kaisai.html
 8. 首相官邸：未来投資戦略 2017—Society 5.0 の実現に向けた改革—, 2017. http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/miraitousi2017_t.pdf
 9. 中島直樹, 田嶋尚子. 糖尿病を中心とした生活習慣病管理項目の策定「各臨床系学会の参画と標準化」. 内分泌・糖尿病・代謝内科 2014；38：573-580.
 10. 中島直樹. 遠隔医療の課題と解決. 医学のあゆみ 2017. in press.
 11. 個人情報保護委員会. 個人情報保護法について. <https://www.ppc.go.jp/personalinfo/>
 12. 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扱いのためのガイダンス. <http://www.mhlw.go.jp/file/06-Seisakujouhou-12600000-Seisakutoukatsukan/0000164242.pdf>
 13. 長谷悠太. 医療ビッグデータの利活用に向けた法整備 — 次世代医療基盤法の成立 —. 立法と調査 2017；391：3-16.