

201405048A

厚生労働科学研究費補助金

厚生労働科学特別研究事業

健康医療分野における大規模データ研究のあり  
方に関する研究

平成 26 年度 総括研究報告書

研究代表者 黒川 清

平成 27 (2015) 年 3 月

## 目 次

### I. 総括研究報告

健康医療分野における大規模データ研究のあり方に関する研究	-----	1
黒川 清		

厚生労働科学研究費補助金（厚生労働科学特別研究事業）  
総括研究報告書  
健康医療分野における大規模データ研究のあり方に関する研究  
平成 27 年 3 月

研究代表者 黒川 清 政策研究大学院大学 政策研究科

【研究要旨】

戦略研究の課題については、これまで、原則として介入研究を実施してきたところであるが、今年度初めて、大規模データベースを用いた解析という研究手法の実施を試みている。本研究では、我が国におけるデータベース研究のあり方を提言した上で、平成 27 年度以降に実施予定の「健康医療分野のデータベースを用いた戦略研究（仮称）」の本格実施に向け、必要な研究基盤および評価方法等についての検討を行い、戦略研究の実施要綱の骨子を作成することを目的とした。

我が国において国や地方自治体、関連団体等が所有する、健康・医療・介護分野の主要な大規模データは、データの性質から大別すると、①リアルワールドデータ（NDB、介護保険総合 DB、KDB 等）、②レジストリ入力データ（地域がん登録、NCD 等）、③ゲノム系データ（東北メディカル・メガバンク等）、④二次利用データの 4 種類が存在するが、特に①については、それぞれのデータベースが個々に独立して整備されてきたこと等の理由で十分な利活用に至っていないのが現状であることがわかった。

一方、健康医療分野以外のビッグデータについてはウェブサービス分野で成長した企業での活用が目立っており、これらのデータ分析を担う人材は、健康医療分野の大規模データ研究を進めていく上で参考になる可能性があると考えられた。

ビッグデータの活用が進んでいる国では、個人情報保護等の課題を解消し、厳重なセキュリティのもと、オープン化が進んでいる点に特徴があった。健康医療分野以外でも政府主導で利活用や研究開発が進められており、ビッグデータの活用の動きは、先進国、途上国の別なく進められていた。

海外の医療分野の大規模データベースは、多くの研究者や民間事業者が利用できるよう整備も進められており、米国の ResDAC や HCUP は NDB の利活用拡大や DPC オンサイトリサーチセンター化を進めている我が国の今後の健康医療分野の大規模データベースの利活用においても特に参考になると考えられた。

以上のことから、我が国の健康医療分野における大規模データベース研究は緒に就いたばかりであり、他分野のいわゆるビッグデータを活用した成果や、海外におけるビッグデータ研究の仕組みに相当するものは確立されていない。ただし、大規模データ自体は各所で集積されつつあり、これらをいかにネットワーク化し、共同研究利用するかが今後の課題であることが改めて明らかとなった。

以上を踏まえ、健康医療分野の大規模データベース研究のあり方について検討を行

い、段階的な発展イメージを提言した。

最初のステップでは、政策形成に資する健康医療分野の大規模研究の日本社会における先導役となる研究を実践すべきである。この際、並行して大規模データの共同研究基盤の整備が必要である。さらに民間におけるビッグデータ活用関係者との意見交換を行い、健康医療分野の大規模データを研究していくために必要な人材、標準化すべき内容等について医学研究者とは異なる立場からも意見を集めることが望ましい。次のステップでは共同利用に基づく大規模データ研究の企画・実施を行う。ここでは大規模データの特性を十分に活かした施策や政策に直結する戦略研究を企画・実施して成果を社会に還元することを狙いとする。並行して大規模データ研究の標準化や教育の充実、人材育成をはかり、大規模データ研究の着実な成果を得るとともに普及拡大につなげることが望ましい。

前述のあり方検討を踏まえて、平成 27 年度から実施が予定されている「健康医療分野のデータベースを用いた戦略研究（仮称）」の研究マネジメント機能に関する要件整理を行い、実施要綱の骨子案として取りまとめた。

## A. 研究目的

平成 26 年度戦略研究「健康医療分野における大規模データの分析及び基盤整備に関する研究」は、既存のデータベースを用いた仮説を立案し、その分析を行うことにより、国民の健康に関する課題の解決に向けた各種エビデンスの創出を目的として実施され、平成 26 年 8 月時点で、6 つの課題が一次選考を通過したところである。平成 27 年度以降に戦略研究として本格実施される前のフィージビリティ・スタディとしての性格を有しており、平成 26 年度は研究実施計画書（フル・プロトコール）および関連書類を取りまとめる。平成 26 年度中に二次選考が行われ、最終的には 3 つの課題が平成 27 年度以降に戦略研究として実施される予定である。

戦略研究の課題については、これまで、原則として介入研究を実施してきたところであるが、今年度初めて、大規模データベースを用いた解析という研究手法の

実施を試みている。そのため、平成 27 年度以降の実施を予定されている戦略研究の成果を確実にするためには、平成 26 年度中に、研究のための基盤整備はもとより、当該研究手法の性質を踏まえたモニタリングや評価手法を合わせて構築して、適切かつ適時に評価等を行いながら研究を推進していくことが求められる。

そこで本研究では、我が国におけるデータベース研究のあり方を提言した上で、平成 27 年度以降に実施予定の「健康医療分野のデータベースを用いた戦略研究（仮称）」の本格実施に向け、必要な研究基盤および評価方法等についての検討を行い、戦略研究の実施要項の骨子を作成することを目的とした。

## B. 研究方法、手順

### 1. 我が国における大規模データベース等の活用状況の現状把握

我が国において国や地方自治体、関連団体等が所有する、健康・医療・介護分

野の主要なデータベースの活用状況について文献調査、ウェブ調査によって把握・整理した。また、健康医療分野以外のビッグデータ利活用事例についても、文献調査、ウェブ調査によって収集した。この際、健康医療分野以外のビッグデータが、当該分野にもたらす影響に留意して取りまとめた。

## 2. 海外におけるビッグデータ研究の仕組みに関する知見の整理

諸外国におけるビッグデータ研究の仕組みについて文献調査、ウェブ調査によって情報を収集し、我が国の健康医療分野のデータベースを用いた研究を取り巻く課題の観点から整理した。

## 3. 「健康医療分野のデータベースを用いた戦略研究（仮称）」の実施体制の検討

1. および 2. を踏まえ、平成 27 年度から実施が予定されている「健康医療分野のデータベースを用いた戦略研究（仮称）」の実施体制に関する検討を行った。まず、我が国の健康医療分野における大規模データベース研究のあり方について提言した上で、平成 27 年度から行う予定である戦略研究の実施体制の検討を行った。ここで取りまとめた要件をもとに、「健康医療分野のデータベースを用いた戦略研究（仮称）」の実施要綱の骨子を作成した。

### （倫理面への配慮）

本研究は、個人情報やヒト生体資料を扱うことなく、特定の個人を観察や介入の対象とするものでもない。したがって、倫理面の特段の配慮には該当しない。

## C. 研究結果

### 1. 我が国における大規模データベース等の活用状況の現状把握

(1) 我が国における健康医療分野の大規模データベース等の活用状況の現状

我が国では、厚生労働省を中心として、情報通信技術を活用した医療情報の電子化を行い、医薬品・医療機器の安全対策や研究開発への利活用の在り方を検討してきた。健康情報や医療情報を網羅的、経時的に把握することができれば、医療実態の把握が可能となり、医学の発展や健康増進、医療費適正化にもつなげることができる。しかし、個人情報保護の問題やいわゆるマイナンバー制度の課題があり、欧米での医療分野のオープンデータ化からは遅れを取っているのが現状である。

現在、大規模データベースとして活用されている、もしくは今後更なる活用可能性がある国内医療関連データベースの例を表 1 に示す。

健康・医療分野の大規模データは、後述する他分野の大規模データと性質や扱いが異なる。データの性質から大別すると、①リアルワールドデータ（NDB、介護保険総合 DB、KDB 等）、②レジストリ入力データ（地域がん登録、NCD）、③ゲノム系データ（東北メディカル・メガバンク等）、④二次利用データ（JMDC Claims Data Base、MDV）の 4 種類が存在する。研究を目的として本格的に利用されているのは④であり、特に①については、それぞれのデータベースが個々に独立して整備されてきたこと等の理由で十分な利活用に

至っていないのが現状である。

今後オープンデータ化して研究等への活用を推進していくためには、データの統合やセキュリティ管理（匿名化、暗号化）、法整備、二次利用への合意形成などへの対応が求められる。さらに、ライフサイエンス分野に知見のある統計学、情報工学人材の育成も急務である。

## (2) 我が国における健康分野以外の大規模データベース等の活用状況の現状

健康・医療分野以外におけるデータベースは、ICTの進展に伴い、各所で大規模化されるとともに、活用が模索されてきた。とりわけ、グーグル、アマゾン、フェイスブックといったウェブサービス分野で成長した企業での活用が目立っているが、自動車産業や通信業でも積極的な試みや取り組みが行われている。これらの分野ではデータを活用することで、生産性を抜本的に高めることができるとともに、新たな付加価値を創出することも可能となる。表2に事例を示す。

これらの事例はビッグデータの特性を活かしており、工夫をこらしたサービスが目立つが健康医療分野のデータとは分析の目的が異なる。一般的には、最終的な自社の利益につながるものであることが目的であるがデータ解析による社会問題の分析に取り組んでいる例もある。厚生労働科学研究が保健・医療に関する社会的課題の解決や施策の立案に資する科学的な臨床エビデンスの創出に資するものであることを鑑みれば全く異質なものであるということにはならない。なお、これらのデータ分析の担い手として、いわゆるデータサイエンティストという専門性が

ある。近年、急速に注目されつつありその定義や資質も含め途上段階ではあるが、健康医療分野の大規模データ研究を進めていく上で今後、参考になる可能性がある。

医学研究とデータサイエンスとを同等に扱うことは適切ではないが、医学研究における仮説構築、研究手法の選定や研究遂行、統計解析など課題解決のアプローチにおいては、同様の専門性が求められると考えられる。

## 2. 海外におけるビッグデータ研究の仕組みに関する知見の整理

諸外国では健康医療分野において、ビッグデータを利活用した先進事例が多数見られる。これを表3に示す。

ビッグデータの活用が進んでいる国では、我が国とは異なり個人情報保護等の課題が解消されており、厳重なセキュリティのもと、オープン化が進んでいる点に特徴がある。ただし課題解消までには各種検討や法制度制定がなされてきた経緯があり、我が国の研究利活用の方向性も見据えながらそれぞれの詳細を理解する必要がある。また、健康医療分野以外でも政府主導で利活用や研究開発が進められている国もある。主な取り組みを表4に示す。ビッグデータの活用の動きは、先進国、途上国のいずれにおいても進められており、革新的なサービスやビジネスモデルの創出を目的とした動きが活発になっている。

海外の医療分野の大規模データベースは、多くの研究者や民間事業者が利用できるよう整備も進められている。この取り組みは、NDBの利活用拡大やDPCオン

サイトリサーチセンター化を進めている我が国の今後の健康医療分野の大規模データベースの利活用においても特に参考になると考えられる。そこで本説では、米国の2つのデータ支援センターの具体的事例を取りまとめた。

#### (1) Research Data Assistance Center (ResDAC)

ミネソタ大学の研究データ支援センター (Research Data Assistance Center, ResDAC) は、研究のためにメディケアおよび/またはメディケイドデータを使用することに関心がある学術研究員、研究公務員および非営利研究者に無償の支援を提供しているCMS (連邦メディケア・メディケイドサービスセンター) 受託機関である。

##### a) ResDAC の概要

ResDAC のスタッフは、ミネソタ大学の疫学者、公衆衛生専門家、医療サービス研究者、生物統計学者および健康情報学専門家の共同体で構成されている。修士レベルのスペシャリストが常駐し、電話での相談に応じている。

##### b) データ利用のための教育機会の提供

ResDAC は、ワークショップや全米アウトリーチプログラムなど数多くの活動分野を通じて教育訓練の機会を提供している。ワークショップは、地域的にはResDAC が所在するミネソタ大学において、それ以外にも全米で開催している。

これらのワークショップでは、CMS データベースの長所や限界について、また、

支払請求に基づく研究で重要な医療問題をどのように探索できるかについて研究者らの認識が深まるように支援している。

疫学、公衆衛生、医療サービス、生物統計学および健康情報学の分野の専門家が進行役を務める会議、展示会およびプレゼンテーションで構成される全米アウトリーチプログラムにおいても教育訓練の機会を提供している。

さらに、ResDAC のスタッフが開発したオンライントレーニングおよびマニュアルも提供している。

##### c) データ利用契約

ResDAC のデータを利用するには、必ず、データ契約書を取り交わすことになる。公開されているデータ契約書の様式では、データ利用契約に関するオンライン・トレーニングコースを完了すること、関連する連邦規則にしたがってデータ保護と運用を行うこと、個人の識別にデータを利用しないこと、営利や競争目的に病院データを利用しないことなどが明記され、これらに従うことを求められている。

#### (2) Healthcare Cost and Utilization Project (HCUP)

Healthcare Cost and Utilization Project (HCUP) は、米国保健福祉省 (Department of Health and Human Services) の研究機関である Agency for Healthcare Research and Quality (AHRQ) が資金を提供して運営されているデータベースの管理組織である。

##### a) HCUP の概要

データは、学術研究機関、政府機関に

限らず、医療保険・投資・製薬・コンサルタント会社などの営利企業に対しても提供されている。NIS データベースは、1988年にスタートした当初は8州だったが、現在は全米44州が参加し、全米の約20%の退院患者データを有する巨大な診療報酬データベースに成長している。NISは集計データではなく、個票データ（マイクロデータ）で、診断名、主要処置・手術、入退院時の状態、患者特性（年齢・性・人種・居住地の所得中央値）、支払保険種類、総請求額、在院日数、病院特性（経営主体、規模、教育病院など）などの情報が含まれている。

#### b) 全米入院患者サンプル（NIS）の概要と利用

NISは、2012年にデータ収集が開始された。リハビリ病院および長期急性期病院を除き、HCUPに参加しているすべての地域中核病院からの退院患者の20%をサンプルとして抽出したものである。NISは、メディケア受給者、メディケイド受給者、民間保険加入者のみならず無保険者を含む全患者を対象としている。NISは、州入院患者データベース（SID）からサンプルを抽出しており、SIDはHCUPに現在提供されている全入院患者データで構成されている。

研究者らや政策立案者らは1988年～2012年のNISデータを入手することができる。NISのデータの特徴は以下のとおりである。

- ・ 米国人口の95パーセント超に相当するすべてのHCUP参加州から抽出。
- ・ リハビリ病院および長期急性期病

院を除く米国の地域中核病院からの退院患者のほぼ20パーセントを層別化サンプルとして抽出。

- ・ 新しい自己重み付け設計により、NISの旧バージョンと比較して推定値の誤差範囲が狭くなり、かつ、推定値の安定性と精度が向上。
- ・ 州識別子および病院識別子を使用しないため、患者の秘密保護が堅牢。
- ・ サンプルサイズが大きいため、希少疾患、稀な治療、および特別な患者集団の解析が可能。

#### c) HCUPデータトレーニングについて

HCUPデータの購入者およびユーザは全員、データを受領する前に、HCUPデータ利用契約（DUA）に関するトレーニングコースを完了し、HCUP DUAに署名する必要がある。

トレーニングの所要時間は約15分程度で、目的は以下のとおりである。

- ・ データ保護の重要性を強調すること
- ・ 不注意による違反のリスクを軽減すること
- ・ HCUPデータ使用に際してのユーザの個人的責任を説明すること

トレーニングコースを完了すると、HCUP DUA トレーニングコース認定コードが付与され、HCUPデータベースの購入申込時に、このコードが利用できる。

## D. 考察

### 1. 我が国のデータベース研究のあり方の検討

研究結果で取りまとめた通り、我が国



の健康医療分野における大規模データベース研究は緒に就いたばかりであり、他分野のいわゆるビッグデータを活用した成果や、海外におけるビッグデータ研究の仕組みに相当するものは確立されていない。ただし、大規模データ自体は各所で集積されつつあり、これらをいかにネットワーク化し、共同研究利用するかが今後の課題であることが改めて明らかとなった。

この現状を踏まえ、我が国における今後の健康医療分野のデータベース研究のあり方について検討を行った。

今後のあり方については、健康医療分野の大規模データベースをどのような研究に活用するかという観点からのあり方と、当該分野の研究をどのようにして成長・発展させるかという観点からのあり方という両面からの検討が必要である。

大規模データの健康医療分野への活用については、以下の観点などからの利活用に対する期待が高まっている。

- ・ 医療や診療の質向上
- ・ 個別化医療の実現（ゲノムデータ利用など）
- ・ 効率的な医療提供と医療費削減（創薬や新治療法開発促進、DNA シーケンサの低コスト・高速化など）
- ・ リアルデータを用いた将来予測

これらの期待は優先度に多少の違いはあるものの、どの国においても共通の期待であると言える。計算機精度やネットワークの進展等とも相まって具体的な課題解決の期待は更に高まる可能性がある。ただし、これらは成果が着実に現れてこそ堅実的な成長を遂げるものであり、結

果が現れない場合にはいわゆる「過度な期待」となり価値を下げていくことになる。したがって期待に応えるための堅実な成果を追究する必要がある。

当該分野の研究の成長や発展については、C.研究結果で示した我が国の実情を踏まえ、世界の先進的な動向も踏まえながら段階的に発展させていくことが望ましい。この際、我が国の大規模データの研究が直面している以下の課題を踏まえた検討が必要である。

- ①個人情報の取扱いなどデータ利活用のルール、法規制の問題
- ②医療情報や医学研究のインフラ整備が不十分
- ③大規模データの解析担当者の不足

①については現在、総務省を中心に検討が進められている個人情報保護法の改正やマイナンバー法制定により、研究基盤の整備が進むものと見込まれる。個人特定性低減データという考え方の導入や、マイナンバー法に基づく医療分野へのマイナンバーの活用等が大規模データ研究を促進させる可能性がある。このような法制度の改正に対応して医学研究の倫理指針も必要に応じた見直し等の議論が行われることになる。

②については、例えばNDBと介護総合DBを繋げられるようにする、自治体が保有する住民の健康関連データ等を結合できるようにする等といった、医療情報のデータベース同士を有機的に結合するインフラ整備が求められる。現在は部分的に結合されていたり、限定的に研究用に活用されていたりするが、研究を通して

この結合や利活用に関する有効性と課題を明らかにして着実な整備を行うべきであろう。戦略研究として、このような将来性を見据えた観点を含めた研究に挑戦することは時宜を得た取り組みとなると考えられる。

また、大規模データベースを用いた健康医療分野の研究は、従来の介入研究や観察研究と違い、歴史は浅く、研究の方法論が確立されているとはいいがたい。しかしながら医学研究である以上、「仮説（リサーチ・クエスチョン）」と「検証と妥当性確認（validation, verification）」は極めて重要である。研究を実践していく上で、仮説とその検証をどのように担保するのか、さらに質の高い大規模データベース研究とはどのようなものなのか、施策や政策に反映できるレベルの研究の質を求めるためには、このような観点からの研究実践と成果の検証が必要であろう。

③については、健康医療分野の大規模データ研究を実践できる研究者および研究チームの育成が必要である。民間企業を中心としたビッグデータの取り組みにおいては、いわゆる、データサイエンティストと呼ばれる専門家の育成が進んでいる。民間企業では CAO（Chief Analytics Officer：最高解析責任者）の役職を設ける企業もある。前述のとおり、これらの人材の定義や専門性については整備されつつある段階であることに留意する必要があるが、十分な社会的コンセンサスを、

将来的には、これらの人材を含めた研究チームの組成によって着実な成果を得るとともに、若手研究者の育成に繋げていくこともあり得るであろう。

段階的な研究発展のイメージを図 1 に示す。

最初のステップでは、先駆的大規模データ研究の実施が必要である。平成 27 年度から実施が予定されている「健康医療分野のデータベースを用いた戦略研究（仮称）」はここに位置づけられる。戦略研究では政策形成に資することに加えて人材育成も重視しているが、本研究においては、政策形成に資する健康医療分野の大規模研究の日本社会における先導役となることが期待される。研究の実践を通じて、同質の研究が広く展開されていくための先導を期待される。

並行して大規模データの共同研究基盤の整備が必要である。現在、厚生労働省で取り組まれている、NDB の研究利用の拡大や DPC オンサイトリサーチセンターの立ち上げの取り組みは、この基盤整備に含まれる。さらに民間におけるビッグデータ活用関係者との意見交換も提案する。健康医療分野の大規模データを研究していくために必要な人材、標準化すべき内容等について医学研究者とは異なる立場からも意見を集めることによって有益な示唆を得られると考えられる。

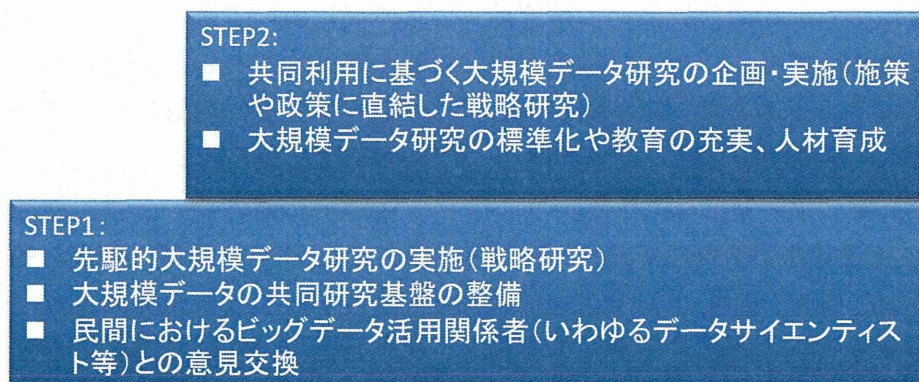


図1 今後の健康医療分野のデータベース研究の段階的な発展イメージ

ステップ1で形成した研究基盤等を用いて、次のステップでは共同利用に基づく大規模データ研究の企画・実施を行う。ここでは大規模データの特性を十分に活かした施策や政策に直結する戦略研究を企画・実施して成果を社会に還元することを狙いとする。並行して大規模データ研究の標準化や教育の充実、人材育成をはかり、大規模データ研究の着実な成果を得るとともに普及拡大につなげる。

## 2. 「健康医療分野のデータベースを用いた戦略研究(仮称)」の実施体制の検討

前述のあり方検討を踏まえて、平成27年度から実施が予定されている「健康医療分野のデータベースを用いた戦略研究(仮称)」の研究マネジメント機能に関する要件整理を行った。特に、これまで戦略研究の必須の要素としてあげられてきた要素との違いを整理し、介入研究を伴わない平成27年度の戦略研究における要件を明確化した。データ利用の際の倫理的な問題や、組織としてのデータベース管理体制等へ留意して要件整理を行った。

### (1) 研究計画書に求める要件の検討

観察研究であるビックデータ解析については、従来の介入的手法による臨床・疫学研究の延長線上で考えてしまうと、本来、大規模データが有するマクロなリサーチクエスションに対する有用性を見落としてしまう可能性がある。したがって、ビックデータの特色である「多数のサンプルを用いることによる相関関係の証明」が、「因果関係解明手法ではかえって難しい」ような、どのような政策形成に利活用可能であるかを、検討され目的設定が考慮されているかが、重要な採択方針のポイントになると考えられる。また、現在議論されているRCTとデータベースを用いた観察研究との有用性の比較結果によっては、将来的に、大規模データベースを用いた戦略研究がRCTを用いることが難しいテーマに対する一手法となることも考えられる。

以上を踏まえ、研究計画書の採択方針案を作成した。これを表4に示す。

### (2) 戦略研究の評価方法の検討

これまでの戦略研究は図2に示すとおり、研究計画を2年間、本研究の実施を5

年間と設定し、その節目において研究評価を行う流れで運用してきた。「健康医療分野のデータベースを用いた戦略研究」についても、平成 25 年度から 2 年間をかけて研究計画の策定を行ってきた。

しかし、「健康医療分野のデータベースを用いた戦略研究（仮称）」は介入を伴わない観察研究であるため、研究基盤整備や被験者リクルート等の手続きが不要である。そこで、研究実施期間は 2 年間と設定された。これまでの研究実施期間は 5 年間としていたため、今後は、中間評価時期が前倒しされ、本研究実施期間中の初年度の年度末に行われることになる。

### (3) 戦略研究の実施体制の検討

これまでの戦略研究では図 3 に示すとおり、研究遂行の際の実施体制を求めてきた。また、戦略研究プロジェクト推進室、研究データを取り扱うデータマネジメントセンターについては、表 5 に示す機能への対応を求めてきた。研究の諸課題を外部評価する委員会組織の設置を必須としていた。

「健康医療分野のデータベースを用いた戦略研究（仮称）」は、研究グループが比較的小規模であり、介入フィールドとの調整を伴わないため、戦略研究推進室の設置は必須ではないと考えられる。ただし、研究を推進するため、推進室に代わる機能は残しておくことが望ましい。

データマネジメント機能についても基本的には不要であるが、研究グループが自身でデータベースを保有している場合と、データセットを外部から入手する場合でマネジメントの方法が異なるため、注意が必要である。データベースを用い

た研究においては、研究の再現性と一般性を担保することが重要となるため、取り扱うデータの性質に応じて適切な対応を求める必要がある。

委員会組織については、倫理委員会の設置は必須であるが、介入研究のような危険性を伴わないため効果安全性検討委員会の設置は不要である。ただし、運営委員会のような外部有識者を交えた諮問機関は設置するのが望ましい。そこで、複数の研究テーマを採択する場合には、健康医療分野のデータベースを用いた戦略研究連絡協議会の新設を提案する。この連絡協議会においては、個々の戦略研究の進捗および課題に関することを協議するとともに、我が国における健康医療分野のデータベース研究の発展に資するため、本研究成果の公表や当該研究の標準化に関することを協議する場としての活用が期待される。

研究の主催・企画体制については、従来通り、厚生科学課、所管課、戦略研究企画・調査専門検討会の 3 者が連携するとともに、節目においては科学技術部会へ報告する仕組みを継続すべきである。

### (4) 「健康医療分野のデータベースを用いた戦略研究（仮称）」の実施要綱の骨子案の作成

ここまでの検討結果を踏まえて、以下のとおり、「健康医療分野のデータベースを用いた戦略研究（仮称）」の実施要綱の骨子案を作成した。この要綱は通常の厚生労働科学研究の規定に加えて、戦略研究に対する規定として発出するものである。研究代表者が戦略研究を実施するにあたっては、この規定に則った研究体制

で研究に取り組まなければならない。

骨子案では、従来の戦略研究のように戦略研究推進室の設置を求めている点がない点  
が大きな特徴である。また、運営委員会  
に代えて健康医療分野のデータベースを用いた戦略研究連絡協議会の設置を求めている点が新たな試みとなる。当該協議会は、個別テーマの進捗管理を行うとともに、研究終了後の成果の均てん化や当該研究領域の発展に向けた協議を行う場としての活用が期待される。

戦略研究（健康医療分野のデータベースを用いた戦略研究）  
実施要綱（案）

第1 目的

厚生労働科学研究費補助金における「戦略研究（健康医療分野のデータベースを用いた戦略研究）」（以下、「戦略研究」という。）は、研究実施計画書に基づく2年間（平成27年度から平成28年度）の研究目標を着実に達成することにより、政策科学の一層の推進を図ることを目的とする。

第2 研究内容等

(1) 研究課題

研究実施計画書のとおりとする。

(2) 研究期間

研究期間は、平成27年度から平成28年度まで（2年間）とする。

ただし、研究進捗状況等によっては、研究期間の途中であっても研究の終了または中止となる場合がある。

(3) 評価項目

主要評価項目及び副次評価項目は研究実施計画書のとおりとする。

(4) 研究方法

戦略研究は研究実施計画書に沿って実施するものとする。

なお、諸般の事情により研究実施計画書の変更等が必要な場合は、厚生労働省大臣官房厚生科学課長の諮問機関である戦略研究企画・調査専門検討会の承認後、所管課にて所要の手続きを経た後に変更するものとする。

第3 研究実施体制及び役割

(1) 実施主体

実施主体は、戦略研究にかかる研究費の受領及び執行管理を行うとともに、研究が円滑かつ適正に遂行されるよう、(2)の研究支援組織を必要に応じて整備する。

(2) 研究支援組織

(ア) 外部委員会

戦略研究の円滑な実施を支援するため、外部有識者からなる以下の委員会を必要に応じて設置し、運営を行う。

ア) 倫理委員会

①研究計画書の科学性及び倫理性を審査し、安全性・有効性の面から研究の継続が倫理的に問題となった場合は、研究計画の変更・中止の勧告を行う。

②戦略研究で活用するデータの管理等に関する倫理審査を行う。

イ) その他必要な委員会  
その他、研究の円滑な実施を支援するため、必要に応じて委員会を設置する。

(イ) データマネジメントセンター  
必要に応じて、戦略研究にかかるデータ収集・保管・管理等を行う。

### (3) 研究グループ

#### (ア) 研究リーダー

ア) 研究実施計画書に沿った研究の実施に関する総括を行う。

イ) 研究分担者、研究協力者の研究グループを構築する。その際、大規模医療データを取り扱う観察研究の専門家を参画させる。

ウ) 研究の実施に必要な研究費の配分、執行を行う。

エ) 研究についての実績報告書及び論文を作成する。

#### (イ) 研究分担者

ア) 研究リーダーと研究項目を分担して研究を実施し、分担した研究項目について実績報告書を作成する。

イ) 分担した研究項目の遂行に必要な経費の配分を受けた場合は、その適正な執行について責任を負う。

#### (ウ) 研究協力者

ア) 研究リーダーの研究計画の遂行に協力する。

## 第4 厚生労働省の事務担当

本研究は、(所管課) が所管する〇〇

〇〇研究事業の戦略型研究として行われる。

## 第5 その他

### (1) 戦略研究に必要な規程類等の作成

戦略研究の実施に当たり、あらかじめ所管課と協議の上、必要に応じて実施主体内の所定手続きを経て所要の規程等を定め、それに基づき実行するものとする。

### (2) 戦略研究についての報告及び公表

本研究においては、研究仮説の明確化と、成果の導出に至るプロセスの再現性の担保が重要である。そこで、戦略研究の進捗状況及び成果等については、戦略研究企画・調査専門検討会または所管課からの求めに応じて、定期的に報告あるいは公表するものとする。

### (3) 「健康医療分野のデータベースを用いた戦略研究連絡協議会」への参画

戦略研究の進捗状況及び成果等を定期的に把握するとともに、健康医療分野のデータベース研究のあり方等について議論することを目的として設置される、健康医療分野のデータベースを用いた戦略研究連絡協議会に参画すること。

①協議会は、以下に関することについて協議する。

- ・戦略研究の連絡調整に関すること
- ・個々の戦略研究の進捗および課題に関すること
- ・その他健康医療分野のデータベース研究の標準化等に関すること

②協議会における協議の結果は、戦略研究企画・調査専門検討会に報告

される。

(4) 厚生労働科学研究費補助金の不正使用及び研究上の不正行為に対する措置

厚生労働科学研究費補助金については「厚生労働科学研究費補助金取扱規程（平成10年4月9日厚生省告示第130号）」に基づき、適切な取扱いを行うこと。また、「厚生労働分野の研究活動における不正行為への対応等に関するガイドライン」（平成27年1月16日付科発0116第1号 厚生労働省大臣官房厚生科学課長決定）、「研究機関における公的研究費の管理・監査のガイドライン（実施基準）」（平成26年3月31日付科発0331第1号 厚生科学課長決定）、「厚生労働科学研究における利益相反（Conflict of Interest：COI）の管理に関する指針について」（平成20年3月31日科発科発第0331003号 厚生科学課長通知）等を遵守すること。

(5) その他

この要綱に定めのない事項については、あらかじめ所管課と協議の上決定し、その定めるところによるものとする。

## E. 結論

1. 我が国において国や地方自治体、関連団体等が所有する、健康・医療・介護分野の主要なデータベースの活用状況について把握した。国や地方自治体、関連団体等が所有する、健康・医療・介護分野の主要な大規模データは、データの性質から大別すると、①リアル

ワールドデータ、②レジストリ入力データ、③ゲノム系データ、④二次利用データの4種類が存在するが、研究を目的として本格的に利用されているのは④であり、特に①については、それぞれのデータベースが個々に独立して整備されてきたこと等の理由で十分な利活用に至っていないのが現状であることがわかった。

2. 一方、健康医療分野以外のビッグデータについてはウェブサービス分野で成長した企業を中心とした利活用が進んでおり、これらのデータ分析を担う人材は、健康医療分野の大規模データ研究を進めていく上で参考になる可能性があると考えられた。
3. 諸外国におけるビッグデータ研究の仕組みについて文献調査、ウェブ調査によって情報を収集し、我が国の健康医療分野のデータベースを用いた研究を取り巻く課題の観点から整理した。ビッグデータの活用が進んでいる国では、個人情報保護等の課題を解消し、厳重なセキュリティのもと、オープン化が進んでいる点に特徴が見られた。健康医療分野以外でも政府主導で利活用や研究開発が進められており、ビッグデータの活用の動きは、先進国、途上国の別なく進められていた。
4. 海外の医療分野の大規模データベースは、多くの研究者や民間事業者が研究利用できるよう整備が進められているものもあり、米国の ResDAC や HCUP は NDB の利活用拡大や DPC オンサイトリサーチセンター化を進めている我が国の今後の健康医療分野の大規模データベースの利活用にお

いても特に参考になると考えられた。

5. 以上を踏まえ、健康医療分野の大規模データベース研究のあり方について検討を行い、段階的な発展イメージとして、まず政策形成に資する健康医療分野の大規模研究の日本社会における先導役となる研究を実践した上で、次のステップで共同利用に基づく大規模データ研究の企画・実施を行うことを提言した。並行して大規模データ研究の標準化や教育の充実、人材育成をはかり、大規模データ研究の着実な成果を得るとともに普及拡大につなげることが望ましい。
6. このあり方検討を踏まえて、平成 27 年度から実施が予定されている「健康医療分野のデータベースを用いた戦略研究（仮称）」の研究マネジメント機能に関する要件整理を行い、実施要綱の骨子案として取りまとめた。

本研究の成果によって、「健康医療分野のデータベースを用いた戦略研究（仮称）」の確実な実施が期待される。また、今後の健康医療研究におけるビッグデータ活用のための研究の仕組みについての示唆が得られた。

#### **F. 健康危険情報**

該当なし

#### **G. 研究発表**

該当なし

#### **H. 知的財産権の出願・登録状況**

該当なし



表1 大規模データベースとして活用されている、もしくは今後更なる活用可能性がある国内医療関連データベースの例

データベース名称／管理者	データの内容・大きさや特徴
ナショナルデータベース (NDB) 【国 (厚生労働大臣)】	<ul style="list-style-type: none"> <li>・ レセプトデータ約 80 億 5200 万件 (2014 年 8 月時点)</li> <li>・ 特定健診・特定保健指導データ約 1 億 2,000 万件 (20 年度～24 年度分)</li> <li>・ 外部事業者にて維持管理を委託</li> <li>・ 第三者提供あり (匿名化済データ)</li> </ul>
介護保険総合データベース 【国 (厚生労働大臣)】	<ul style="list-style-type: none"> <li>・ 要介護認定データ：約 1,260 万件 (認定ソフト 2009 SP3 送信分、平成 22 年 1 月以降)</li> <li>・ 介護レセプトデータ：約 3,600 万件 (一か月あたりの給付明細データ件数)</li> <li>・ 介護保険総合 DB 全体のデータベース容量：約 1.7TB</li> <li>・ 第三者提供なし</li> </ul>
DPC データベース 【国 (厚生労働大臣)】	<ul style="list-style-type: none"> <li>・ 患者臨床情報 (傷病名等) と診療行為情報のデータセット、年間約 878 万件 (2011 年度) のデータが蓄積される</li> <li>・ 参加医療機関は 1,585 病院、49 万床 (2015 年 4 月時点)</li> <li>・ 医療機関ごとの集計表を公表</li> <li>・ 平成 27 年度よりレセプト情報等オンサイトリサーチセンター化により東京大学と京都大学にてデータを管理</li> </ul>
国保データベース (KDB) 【国保中央会】	<ul style="list-style-type: none"> <li>・ 大きさは不明</li> <li>・ レセプトデータ、特定健診等データ、介護レセプトデータ匿名化済 (※国保、後期高齢)</li> <li>・ 第三者提供あり</li> </ul>
健保連システム 【健保連】	<ul style="list-style-type: none"> <li>・ 大きさは不明 (健康保険組合数 1403 (2015 年 4 月 1 日時点)、加入者数約 3,000 万人)</li> <li>・ レセプトデータ、特定健診等データ (※被用者保険)</li> </ul>
地域がん登録 【各都道府県 (がん登録センター)】	<ul style="list-style-type: none"> <li>・ がんの罹患数・率、診断時の進行度、がん患者の受療状況、がん患者の生存率データ</li> <li>・ 大きさは不明 (2012 年度の登録数は 61 万例)</li> <li>・ 匿名化なし</li> <li>・ 第三者提供なし</li> </ul>
National Clinical Database (NCD)	<ul style="list-style-type: none"> <li>・ 手術情報 400 万件以上 (2014 年 3 月末時点)</li> <li>・ 日本で一般外科医が行っている手術の 95%以上をカバー</li> </ul>

データベース名称／管理者	データの内容・大きさや特徴
【一般社団法人 National Clinical Database】	<ul style="list-style-type: none"> <li>一する年間 120 数万件が入力されている</li> <li>・ 第三者提供あり</li> </ul>
バイオバンク【東北大学 東北メディカル・メガバンク機構】	<ul style="list-style-type: none"> <li>・ 宮城県、岩手県の住民総計 15 万人を対象としたコホートデータ（遺伝解析情報、血液・尿等の生体試料、検査値等の医療情報、生活習慣・住居環境等の健康情報）</li> <li>・ 匿名化済</li> <li>・ 第三者提供あり</li> </ul>
JMDC Claims Data Base【株式会社日本医療データセンター】	<ul style="list-style-type: none"> <li>・ JMDC が契約する複数の健康保険組合加入者のレセプトデータ（医科・調剤・DPC）</li> <li>・ 大きさは不明（総登録者数は約 275 万人）</li> <li>・ 第三者提供あり</li> </ul>
MDV【メディカル・データ・ビジョン株式会社】	<ul style="list-style-type: none"> <li>・ DPC データ、レセプトデータ（病院からの直接データ収集により全保険種類のデータを保有）</li> <li>・ 大きさは不明（総登録者数は約 750 万人、国内既存サービスでは最大規模の診療データを蓄積）</li> <li>・ 第三者提供あり</li> </ul>

表 2 健康医療分野以外におけるデータベースの活用事例

企業名	サービス等の名称	取り組み概要
Google Inc.	ビッグデータ分析でみるインフルエンザ感染状況	インフルエンザ関連のトピックを検索するユーザ数と、実際にインフルエンザの症状を示す患者数との間に密接な関連性があることを発見。検索キーワードの検案件数を調べることで、インフルエンザの流行状況を国や地域ごとに予測できることを明らかにした。
ヤフー株式会社	ビッグデータ分析でみる感染状況	検索データからインフルエンザやノロウイルスの感染状況や終息状況を予測。2014 年 12 月の衆院選議席数についても予測（2013 年 7 月の参院選予測は的中率 92%）した。その他、経済やスポーツに関するビッグデータレポートも発行している。
本田技研工業株式会社	走行データを用いた渋滞回避システム	ナビゲーションシステム装着車の走行データを解析し、渋滞を回避することによりドライバーおよび地球環境へ貢献することを目的としたシステムを開発した。

企業名	サービス等の名称	取り組み概要
富士通株式会社	ビッグデータを用いた各分野での実験的取り組み	健康医療分野以外に、コンビニマーケティング、農業、移動体情報による利活用実験を実施した。
ゼンリンデータコム株式会社	天候データの予測	携帯電話基地局に設置した気象センサーによる天候データと地図情報を活用した、携帯電話向け測定・情報提供サービスを開発し、高精度かつ利便性の高いシステムを開発した。
パナソニック株式会社	SNS データ解析によるマーケティング	「車内」に関する SNS コメントを解析し、「悩み」「困りごと」を抽出、対策商品の自動車メーカーへの提案、事業機会の獲得につながった。
Visa Inc.	取引パターンデータ解析による不正使用判定の速度・精度向上の取組	従来1ヶ月を要していたカードの不正使用の判定について、1日平均1億5,000万件の取引の中から個々人の利用パターンを特定することで、13分での不正使用判定が可能となり不正利用率を史上最低レベルに抑制した。
損害保険ジャパン日本興亜株式会社 日産自動車株式会社	自動車の運転行動データ解析によるリスク分析と保険商品コンサルティング(「ドラログ」)	日産リーフの車載センサーを用いて運転行動データを収集し、リスク分析を実施。結果は保険料率に反映されるシステムを構築した(現在は走行距離のみだが、順次解析対象を拡大予定)。
株式会社 資生堂	サイトアクセスログデータを用いたターゲティング広告	資生堂自社サイトの顧客情報、Yahoo!DMP を活用した Yahoo のオーディエンスデータを組み合わせ、ターゲティング広告を実施した。CVR (成約率) が3倍に増加した。
和歌山県 紀美野町	参加者のコミュニケーションデータ分析による「婚活」イベントの精度向上	少子化対策の一環として商工会議所と開催している「婚活」イベント参加者の行動ログを解析し、カップル成立率を向上させる取組を実施した。今後、イベント企画への活用、参加者のソーシャルタイプによるアドバイス等を実施予定。

表3 海外の健康・医療分野に関する大規模データの活用事例

国	大規模データの活用事例	個人情報保護等への対応
デンマーク	<ul style="list-style-type: none"> <li>・ CPR (国民個人識別番号) : 行政と民間が CPR 番号をもとにデータベースを共有。病院での診察、納税、銀行口座の開設等で幅広く個人認証として利用。</li> <li>・ NPI : 患者の診療履歴、処方履歴、アレルギー等の情報統合、閲覧サービス</li> <li>・ FMK (共通医薬品カード) : 処方履歴の横断的一元化データベース</li> <li>・ SDN (健康データネット) : 患者医療情報等にアクセスできる医療関係者用のネットワーク基盤</li> <li>・ sundhed.dk : 一般的な健康・医療情報、患者の医療情報へのアクセス、カスタマイズされた市民個人の医療情報管理ツール</li> </ul>	<ul style="list-style-type: none"> <li>・ デジタルヘルス:異なる組織に存在する既存システムの横断的連携・統合化を実現するための組織</li> </ul>
スウェーデン	<ul style="list-style-type: none"> <li>・ NPO (全国患者サマリ) : 患者の医療情報一覧サービス</li> <li>・ PASCAL : 医薬品処方者が患者の過去の処方歴を閲覧</li> <li>・ UMO (若者向け) /1177 (一般市民向け) : 健康、医療等情報提供サービス (匿名 QA サービスなど)</li> </ul>	<ul style="list-style-type: none"> <li>・ 「全国 e ヘルス—医療及び介護 の分野における安全でアクセスが容易な情報のための戦略」(2010)</li> <li>・ 患者情報法 (2008 制定)</li> </ul>
英国	<ul style="list-style-type: none"> <li>・ Care.data : 医療関係者に対し研究利用に限定して患者の医療記録データを収集し、そのシステムに蓄積された医療データにアクセスできる仕組みを構築するプロジェクト。NHS が国民荷室の高いケアを提供するのが目的。</li> </ul>	<ul style="list-style-type: none"> <li>・ オプトアウトの方法、匿名化の手法について実施前に議論、対応がなされた。</li> </ul>
シンガポール	<ul style="list-style-type: none"> <li>・ EMRX : 公的・二次医療機関の電子カルテ情報の共有システム (診療・治療ごとの記録の共有)</li> <li>・ NEHR : 個人の長期的・横断的な健康</li> </ul>	<ul style="list-style-type: none"> <li>・ 二次医療を担う公立セクターのみで進められている</li> </ul>