

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

総論：医療ビッグデータの利活用

東京大学大学院医学系研究科臨床疫学研究システム学
小出 大介

はじめに

最近、医療にかかわらずビッグデータという言葉がよく使われるようになった。特に興味深いのはNEJM¹⁾やJAMA²⁾などの一流雑誌でも取り上げられるようになったことである。ただビッグデータに共通した概念や定義が必ずしもあるとは言えない状況であるが、ひと昔前では取り扱いが困難であったような大量データが、近年の情報技術の発達により処理できるようになったことから注目されてきていると考えられる。したがって、ここでもこの程度のゆるい概念に基づいて述べることにする。

また、『日本臨床試験学会雑誌』において特集されることについて、これら医療ビッグデータは主に観察研究など疫学研究に用いられることが多いことから違和感を持たれる読者もいるかもしれないが、本誌の投稿規定に掲載論文として「臨床試験・疫学研究に関する」とあることからご理解いただきたいと考える。また、これら観察研究に関しては、安全性にかかわる検討など臨床試験としては倫理上や現実的な問題として実施困難なりサーチクエスチョンに答えることができ臨床試験を補完しうることから、臨床試験分野にとっても役立つものと思われる。

現状でも臨床試験でEDC (electronic data capture) が活用されることが増えており、今後は医療ビッグデータを用いた臨床試験ということもありうるのかもしれない²⁾が、今回は既存の医療ビッグデータを中心にすでに実績のある方々にご執筆いただいた。

日本における医療ビッグデータ

国際的には医療分野のデータベースの所在と概要

を集めてWeb上で提供している組織として、ISPOR (International Society for Pharmacoeconomics and Outcomes Research)³⁾やEMA (European Medicines Agency) によるENCePP (European Network of Centres for Pharmacoepidemiology and Pharmacovigilance)⁴⁾などがあるが、あまり頻繁に更新されているわけではなく、アジア、ことさら日本に関する紹介は極めて少ない。しかし、日本薬剤疫学会の薬剤疫学とデータベースタスクフォース (TF) では、「日本における臨床疫学・薬剤疫学に応用可能なデータベース調査」を実施して、毎年情報を更新して公開しており、その数は増えている⁵⁾。また、国としても2015年6月に閣議決定された『日本再興戦略』改訂2015⁶⁾においても「匿名化した医療等のビッグデータの活用も進めることにより、医療や介護とも密接に連携した健康・予防関連サービスが活性化する」であるとか、「世界に冠たる医療等分野でのデータベースの構築により医薬品等の安全対策の充実や研究開発の飛躍的な促進も期待される」など、目玉のひとつとなっている。

そこで日本の医療ビッグデータの利活用の紹介として、まずDPC (Diagnosis Procedure Combination) を用いた研究を東京大学の康永秀生氏にご執筆いただいた。これはDPCデータ調査研究班が各施設から個別に収集したビッグデータで、日本のすべての急性期入院患者の約50%を占める。このデータは、評価しやすいように現実とは異なる理想環境下で実施されるランダム化比較試験 (RCT) とは異なり、リアルワールドの臨床データという強みがある。さらには、RCTでは倫理的な面から実施困難なりサーチ

Introduction: The Most Use of Big Health Care Data

Daisuke Koide: The Department of Clinical Epidemiology and Systems, The University of Tokyo Graduate School of Medicine

クエスチョンにも答えることや真のエンドポイントを用いた評価も、このようなビッグデータを用いて可能となる。しかしその一方、リアルワールドデータであることから、交絡やバイアスの影響が入り込みやすく、それらを可能なかぎり調整する必要や実施するための力など課題も指摘されている。

続いて、医薬品医療機器総合機構 (PMDA) の石黒智恵子氏と宇山佳明氏に、「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」と題してまとめていただいた。そのなかで現状の課題として、①人材育成、②データ標準化、③個人情報の取扱ルールの明確化、④バリデーション試験の実施、⑤信頼性保証基準の5点を端的に指摘していただいた。

さらに上記のデータベースとは特徴が異なり、がん登録などと同様なレジストリタイプとして National Clinical Database (NCD) の説明およびそれを用いた研究について、東京大学の隈丸拓氏らから紹介していただいた。これは外科領域の専門家集団が自らの医療の質を評価し高めていくことを目的に設置され、医療機関から地方厚生局へ提出される手術件数との照合で95%以上の網羅性を有し、また、データの質管理では監査なども併用するなど工夫されている。さらに、risk calculator による死亡率の予測値を術前患者へのインフォームドコンセントに用いたり、performance 指標による自院成績の全国成績との比較をしたりするなど、フィードバック機能も優れており、また臨床試験におけるヒストリカルコントロールデータを提供した実績など、興味深い実績に富んでいる。

そして、日本医療データセンターの木村真也氏から「民間の健康保険データベース」についてご紹介いただき、その特徴として「追跡性」「網羅性」「母集団(分母)の把握」の3点が示され、課題として高齢者などが少ない点や症病名の正確性、検査結果など把握困難なデータがある点なども記述していただいた。

利活用上の法的課題

このように医療ビッグデータの利活用が期待される一方で、PMDA からの指摘にもあるように個人情報

報の取扱ルールの明確化の問題もある。特に2015年9月に「個人情報の保護に関する法律」の改正法が公布され、全面施行に向けて準備が進められているが、病歴が要配慮個人情報となり、オプトアウトでの収集とデータベース化ができなくなる可能性があり、医療分野の個別法やガイドラインが求められるところである。この点に関しては、一般財団法人医療情報システム開発センターの山本隆一氏から、要点をわかりやすく展望において解説していただいた。

おわりに

医療の情報システム化に伴い発展してきた医療ビッグデータには、自ずとデータが蓄積していくことや、リアルワールドデータであるという魅力もあるが、一方で法的課題のみならず、交絡やバイアスが入りやすいという解析や結果の解釈上の課題もある。これら交絡やバイアスの話に関しては、今回は詳細を省くが、執筆者たちの記述中にも紹介されているように、プロペンシティスコアマッチングや sequence symmetry analysis (SSA) の手法を用いたりすることで、ある程度制御もできる技術が編み出されてきており、今後もさらに新たな手法も開発されていくものと思われる。

利益相反

筆者である小出大介の所属する臨床疫学研究システム学講座は MSD 株式会社による寄附講座である。

文 献

- 1) Schneeweiss S. Learning from big health care data. N Engl J Med 2014; 370: 2161-3.
- 2) Angus DC. Fusing Randomized Trials With Big Data: The Key to Self-learning Health Care Systems? JAMA 2015; 314: 767-8.
- 3) ISPOR International Digest of Databases.
<http://www.ispor.org/DigestOfIntDB/CountryList.aspx>
- 4) ENCePP Resources Database.
<http://www.encepp.eu/encepp/resourcesDatabase.jsp>
- 5) 日本薬剤疫学会 薬剤疫学とデータベース TF.
<http://www.jspe.jp/committee/020/0210/>
- 6) 首相官邸「日本再興戦略」改訂2015.
<https://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/dailjp.pdf>