

■ Topics | トピックス

「製薬協メディアフォーラム」を開催

テーマは「医療におけるビッグデータ：今後の展望と活用」

製薬協 広報委員会では、医療の分野においても今後活用が期待されている「ビッグデータ」を2015年度の製薬協メディアフォーラムのテーマとして取り上げ、2回にわたって紹介することにしました。第1回目のフォーラムは2015年8月18日に開催し、京都大学大学院 医学研究科 健康情報学分野 教授の中山健夫氏が「医療におけるビッグデータ」をテーマに講演しました。講演の概要は以下の通りです。



会場風景



「医療におけるビッグデータ」

京都大学大学院 医学研究科 健康情報学分野 教授 中山健夫 氏

1) ビッグデータの時代

2000年代以降、人間が生み出すデータ量は急激に増加しています。医学関係でビッグデータという言葉が初めて登場したのは2008年のNatureの特集号「ペタバイト時代の科学」ですが、その後、2014年には年間220編の論文がビッグデータをタイトルに冠するに至っています。医療におけるビッグデータおよび大規模データベースは、「ヘルスケア」と「ライフサイエンス」の2つの領域で大きな発展が見られます。

従来は、疫学研究でも研究対象として数千人から1万人前後の規模のデータを扱うことがほとんどでしたが、今では数万人から10万人規模の集団を対象とする研究へと

発展しています。しかし、データの扱い方を理解しないまま「ビッグデータ」を扱っても大きく間違った分析結果しか得られません。医療におけるビッグデータは生命にかかわることなので特に留意する必要があります。

ビッグデータには4つの特徴(4Vs)があります(図1)。

1つ目はVolume(容量)です。たとえばDVD映画1本は2ギガバイトです。

2つ目はVelocity(迅速性・経時性)です。次から次へと新しいデータが生成されることですが、たとえば、心電図や酸素飽和度などのモニターから間断なく継続的に測定可能で、時間とともに蓄積されていくデータはビッグデータとなるでしょう。なお、ゲノム情報は塩基対が30億という膨大な情報であっても塩基配列自体は変化しないためVelocityという要件は満たしていないこととなります。

3つ目はVariety(多様性)です。人間の生命活動、健康状態、医療や検診の受診などの人為的な事象、室内環境、ソーシャ

ルキャピタルや医療機関へのアクセスのしやすさといった地域的要因などの多様なデータを集積することにより、人間が病気になる原因を解析することが可能になり、予防や治療ができるようになります。

4つ目はVeracity(正確性)です。データが膨大になれば偏りが減りますし、データの抽出バイアス・選択バイアスも解消されます。ただし、ビッグデータが膨大であれば正確であるというわけではありません。ビッグデータが正確であるために真の値からのバイアスが少ないことが必要となりますし、測定バイアスにも注意が必要です。

これらの4Vsという特徴を持つビッグデータを賢く、慎重に活用できれば、私たちは新たなVision(視野)を得て、新たなValue(価値)を創り出していくことができるでしょう。

図1 ビッグデータ:その特性と意義



2) 医療ビッグデータと大規模データベース

単なる「ビッグデータ」とは未整備な膨大な素材の集まりと言えます。これに対して、「大規模データベース」とは、構造化されたビッグデータと言われています。わが国では、2002年に診断群分類(Diagnostic Procedure Combination、DPC)制度が導入されています。2011年までに878万件のDPCデータが蓄積されています。

また、2008年の「高齢者の医療の確保に関する法律」の成立で、都道府県ごとに医療費適正化計画が策定されるとともに、特定健診・特定保健指導が導入されて診療報酬明細書(レセプト)情報や特定健診等のデータの集積が急速に進みました。そして、2011年には遂に全医療機関の電子化への義務的移行がなされました。なお、特定健診のデータとレセプトデータを突合せることが当初の大きな眼目であったのですが、残念ながらまだ20%前後しか突合されていません。

薬害対策もビッグデータへの大きな推進力となりました。2010年4月に提示された薬害肝炎検証・検討委員会の最終提言「薬害再発防止のための医薬品行政等の見直し」においては、「電子レセプトなどのデータベースから得られた情報を活用して薬剤疫学的な評価基盤を整備することや、個人識別子等を用いて電子カルテ等のデータへの連携や高度な分析への活用を可能にすることが検討されるべき」と、指摘されています。かつては疫学研究をはじめするには、まず一次情報を各地の医療機関から収集するところからはじめなければなりませんでした。データベースにアクセスすれば解析できる世の中に向けて一歩ずつ進みはじめました。薬剤の副作用問題に適切に対処するためには、データベースを活用して迅速に知見を得ることが極めて重要です。

3) レセプトデータの活用

大規模データベースの活用により、解析に足る十分な症例数の患者データを効率的に収集し、医薬品安全性におけるエビデンス診療ギャップ、すなわち医療の質の一面を定量化することができます。では、エビデンスに基づいてガイドラインが推奨する治療法や避けた方が良いとされている治療法が臨床現場でどの程度実施されているのでしょうか。

骨粗鬆症の事例を挙げて説明しましょう。骨粗鬆症の発症原因として、副腎皮質ステロイド薬の使用が知られています。わが国でも、臨床研究から得られたエビデンスに基づき、「ステロイド性骨粗鬆症の管理と治療のガイドライン」が2004年に発表されました。2011年に、複数の健康保険組合のレセプトデータベースに含まれる53万人の被保険者から条件を満たす2400人の対象者を基にこのガイドライン推奨の実施状況を解析したところ、わずか23%の医師しかガイドラインが推奨する治療法を実施していないということが明らかになりました。大規模なデータの解析によって骨粗鬆症の管理と治療が十分と

はいえない現状が明らかになったといえます。

民間の医療データベースとしては、日本医療データセンター (Japan Medical Data Center Co, Ltd、JMDC) によるものが研究に使われています。このデータベースには、複数の健康保険組合からのレセプト情報(入院、外来、調剤)が匿名化の上で名寄せ作業を経た270万人規模のデータが現時点で蓄積されています。健康保険組合からのデータによるため後期高齢者のデータは含まれていない点が限界ですが、継続的に対象集団の全数データが収集できる点は研究目的に適しているといえます。





国のナショナル・データベース (National Data Base、NDB：レセプトと特定健診・保健指導のデータを集積させたもの)は、年間18億件のレセプトデータが蓄積され、民間データベースより圧倒的大規模ですが、使いやすさの点でまだ多くの課題があります。2014年に東京大学と京都大学にNDBのオンサイトセンターが設置されました。民間も含めた今後の活用の拡大を視野に入れて、地道なルール整備が進められているところです。

また、厚生省は2015・16年度にデータベースを用いた戦略研究を別表のように4件採択しています(図2)。私たちもその1つとして高齢者医療の質の定量化に取り組んでいます。この課題では、NDBやJMDCなどいろいろなデータベースを使用目的やデータベースの特性に応じて使い合わせ、使い分けをします。NDBは精度が低いですが規模は大きいですし、たとえば、京大病院の院内がん登録のデータは精度が高いですが規模は小さいといえます(図3)。それぞれのテーマごとにデータベースの特徴と課題を整理していく予定です。今後2～3年で医療ビッグデータのさまざまな基盤の整備がいろいろ進んでいくでしょう。

図2 平成27年度厚生労働科学研究 健康医療分野のデータベースを用いた戦略研究

- 大規模データを用いた運動器疾患・呼吸器疾患・がん・脳卒中等の臨床疫学・経済分析
(代表者:康永秀生東京大学大学院医学系研究科教授)
- 地域包括ケア実現のためのヘルスサービスリサーチ-二次データ活用システム構築による多角的エビデンス創出拠点-
(代表者:田宮菜奈子筑波大学医学医療系ヘルスサービスリサーチ分野教授分野長)
- 高齢者医療の適正化推進に向けたエビデンス診療ギャップの解明既存データベースを利用した、京都大学オンサイトセンターにおけるレセプト情報等データベース(NDB)の活用方策の検討
(代表者:中山健夫京都大学大学院医学研究科教授)
- レセプト情報・特定健診等情報データベースを利用した医療需要の把握・整理・予測分析および超高速レセプトビックデータ解析基盤の整備(代表者:満武巨裕医療経済研究機構研究副部長)

図3 本研究で使用する各種データの特徴について

使用目的	データ	情報の精度の高さ	情報の規模の大きさ
課題の解決 利活用促進	 レセプト情報・ 特定健診等情報 データベース (通称:NDB)	低い •情報の精度が不明 •いわゆるレセプト病名の存在 •病床数や診療科等の情報は、 精度検証が全くなされていない。 •データクリーニングがほとん ど行われていない。	大きい •極めて悉皆性が高い
NDBのバリデーション +独自活用	 各種民間 データベース 〔JMDC, MDV, 調剤レセプト等〕	•含まれている情報自体は、 NDBと同様レセプト情報 (一部DPC含む)が主体。 •データクリーニングが行われ ており、各種分析が容易。	•特定の機関のデータで構成 されており、データの偏りは免 れない。 •数十万~数千万単位の規模 を擁するデータベースもあ る。
	 P-retriever® 〔データベース 管理システム〕	•レセプトに含まれている情報 に加えて 検査情報 も含ま れており、幅広い分析が可 能。	•一部医療機関のデータ(京大 病院ならびに複数の連携医 療機関)のみで構成されてお り、データの偏りは免れない。
	 CyberOncology® CyberNephrology® 院内がん登録	•レセプトに含まれる情報に加 え 検査情報・カルテ情 報、臨床的アウトカム 情報 も含まれており精度の 高い分析が可能。	•一部医療機関(京大病院)の データのみで構成されてお り、データの偏りは免れない。
		高い	小さい

国内の医療データベースは急速に発展していますが、それぞれの運営母体も原資料からの収集方法も異なっているため、データの妥当性の評価(バリデーション：ある方法が正しいことを確認する過程)が大きな課題となります。

レセプトデータベースは、あくまで保険金支払い請求を目的として構築されており、研究目的としてデータ入力されているわけではありません。たとえば、レセプト情報では「心筋梗塞」とされていても、診療情報(カルテ)と突合すると実際の病名は異なることがよくあります。このため、本当に心筋梗塞ならば行われていたであろう治療(心臓バイパス手術や心臓カテーテル手術など)が実施されていることを確認して、心筋梗塞と定義し直します。このような地道な作業が、データベースを利用した研究には必要になります。

この数年間で、日本の医療ビッグデータを巡る環境は大きく変わり、数年前は世界から3~4周遅れだったのがいまや1周遅れまで取り戻し、一部の領域では世界の先頭を走っていると感じています。たとえばNDBの潜在的価値は超高齢者を含む高齢者に国民皆保険制度の下でどのような医療が行われているのか、一億人規模の人口を擁する国で解明できる現時点で世界唯一かつ最大のデータベースです。DPCのデータベース、そして日本外科学会などによる数百万件の手術例が登録されたナショナルクリニカルデータベースなど、単品のデータベースとしては世界最大規模です。これらがほかのデータベースとつながれば、その価値は飛躍的に高まるでしょう。

4) 展望と課題

ビッグデータにはいくつかの意義があります。

1つ目は、限られたデータでは見えない関係を見出すことです。タバコとがんの関係はわかりやすい事例でしょう。個々人を見るだけでは、非喫煙者でもがんになるし、喫煙者でもがんにならないというバラつきが目立ってしまいます。しかし、喫煙者100人と非喫煙者100人を比較してがんの発症率を比べれば一般的な関係が見えてきます。データの規模が大きくなれば、多くのリスク因子を精密に解明していくことができるでしょう。

2つ目に、データの母集団が大きいことによって、性別、年齢、地域、さまざまな背景に基づいたサブグループ(セグメンテーション)の設定が可能となり、属性別により詳細な分析をして、きめ細やかな基準値の設定や治療の方針の検討に役立てることもできます。

3つ目に、稀なものを集めることができます。たとえば、1万人に1人しか罹患しない稀少疾患の症例をいくつも集めることが可能となり、研究が成立しえます。

4つ目に、目の前の患者さんと似た症例の治療記録をデータベースから探すことによって治療の手がかりを得ることができます。敢えていえば、ゲノムや遺伝子とは別の視点からの個別化医療の可能性です。患者さんにとっても同じで、「自分と同じ病気になった他の人」を見つけることで、「同じ病気になった他の人はどうしているのか」という疑問に答えることも可能となるでしょう。

エビデンスベースの医療とは、「現在得られる最良の根拠を、良心的、明示的かつ思慮深く用いること」とされていますので、ビッグデータを誤って使ってははいけません。超高齢社会に直面するなか、患者の一生を軸としたシームレスな医療健康福祉の提供が望まれています。良質な医療と介護ケアを効率的に提供するためには、壮大で緻密なリアルワールドデータを適切に関係者間で共有できるよう、基盤整備が必要です。そのカギが医療健康情報の一元化であり、医療用の個人識別番号の導入だと考えます。

終わりに

健康・医療ビッグデータを巡る状況はめまぐるしく変わっています。個人情報保護が重要だと叫ばれている時代に、個人の医療情報をまったく出さないというのでは、個人も「多くのデータを持ち寄ってわかる」さまざまな情報の恩恵を受けることはありません。医療情報は、適切に守り、そしてお互いに持ち寄って共有・活用することによって新たな価値が生まれてくるものです。さまざまな立場の人々が専門的知識や知恵を持ち寄り、十分に議論することで、より良い医療と社会の実現に向けた、医療ビッグデータの構築・活用に必要な社会制度・法律の整備が進むことを願っています。

(広報委員会 政策PR部会 河上 崇陽)