

病院情報システムに格納されている患者個票データとDPCデータを用いた疫学調査および医療評価を行う上での交絡因子調整に関する検討

国立病院機構東京医療センター 臨床疫学研究室/総合内科 尾藤誠司

<要旨>

東京医療センター病院情報システムに格納されているデータと同院DPCデータを統合させ、その統合データから抽出された二次データセットを解析することで、臨床評価指標の算出や臨床疫学研究を実施した。その上で、病院情報システムからの直接のデータ抽出が持つ利点や限界などについての検討を行った。今年度は、3つの具体的な研究仮説と共にデータ抽出を行い解析を行った。その結果、臨床疫学研究を行う上で病院情報システムが持つ多くの可能性について示唆を得ることができた。今後、さらに多くの事例を集積することによって、臨床のエビデンスを創出するとともに、病院情報システムの有効利用について更なる検討を行いたい。

<背景と研究目的>

ランダム化比較試験に代表される、臨床試験型の研究デザインによる臨床エビデンスは、質の高いエビデンスを生み出すことができるが、一方で、前向きにデータを収集するという手順が持つ、データ入力の際のバイアスや、研究事業にかかる経費、さらには倫理上の問題など、いくつかの懸念事項があげられる。一方、ここ数年、いわゆる「ビッグ・データ」と呼ばれる、研究や分析を前提としていない電子化され蓄積された大量のデータを解析に二次利用することで、新たな知見を生む可能性が注目されている。病院において、この「ビッグ・データ」にあたるものは、電算化レセプトデータおよびDPCデータであり、ここ数年これらのデータの二次利用が活発に行われつつある。

電子化され、施設内のデータベースに格納されている病院情報はDPCデータの他

にもさまざまな種類が存在する。例えば、検査部門データベースは、臨床検査値に関して信頼性の高いデータが格納されており、これらは患者のアウトカムを直接的もしくは間接的に表現するデータ源として重要性が高い。また、看護師が主に記録するバイタルサイン（呼吸数、血圧、脈拍数など）のデータなども、患者のアウトカムを表すデータ源として有用性が高いと考えられる。今回の申請者は、平成22-23年度 厚生労働省科学研究費補助金 地域医療基盤開発推進研究事業 「保存された診療データの二次利用適用レベルに準じた、医療提供プロセスおよびアウトカムの病院横断比較、年次縦断比較に関する多施設共同研究」において、DPCデータと、病院情報システムの他のデータベースに格納されている多数のデータベースのデータを連結させることによって、より詳細に臨床を反映するような臨床評価が可能であることについて明ら

かにしたとともに、電子情報の二次利用を行う上で、文字情報など、患者プライバシー情報の取り扱いに対して大きな配慮の必要性があることについて提示した。具体的には、研究や経営改善等を目的として電子化された病院情報の二次利用を試みる際、実際の電子カルテを個別に閲覧する場合には、データ収集の労力や、患者プライバシー保護に関する倫理的な問題点などについての劇的な改善は難しい、という点が一つである。また、実際の電子カルテを個別に閲覧することなくデータを抽出した場合、文書データにおいては、どうしても固有名詞などの情報が混入するため、特殊なソフトを開発した上で人工的な匿名化を行うなどの特殊な手続きが必要であったという点である。

一方、実際に患者個別の情報を閲覧することなく、データ抽出のためのプログラミングを行うことで、DPCその他の格納されているデータベースから直接匿名化された患者情報や臨床情報をCSVファイル等で抽出することができれば、データ収集にかかる労力を大幅に削減するとともに、解析の自由度は大きく広がり、さらには患者のプライバシー保護の観点からも大きな利点となる。今回申請者の目的は、特に疫学調査を目的としたうえで、病院内のデータベースから直接患者個票単位でのデータセットを抽出し、そのデータセットを用いてあらかじめ設定した疫学研究仮説の検証を行う解析の有用性を検証することにある。

今回我々は、東京医療センターが患者サービスを目的として病院情報システムを通じて記録し、蓄積した過去の診療データを具体的な仮説を設定した上で二次解析し、

その有用性や新たな可能性についての検討を行った。

<方法>

研究デザイン：既存データベースからのデータ抽出を行った上での二次解析

セッティング：東京医療センター病院情報システムに格納されているデータ、およびDPCデータ

研究仮説：今回は、以下の3つの仮説を設定しデータの抽出と解析を行った。

- ・ 一般病棟に緊急入院した内科系患者のうち、アルブミン値の最低値が2.5未満となった患者の退院時予後分布。さらに、アルブミンレベルと退院時死亡割合の相関関係
- ・ 入院中の高齢者に対する身体抑制が、退院時の当該患者のADL低下にもたらす影響
- ・ 患者特性および重症度で調整を行った上、肺炎で入院した65歳以上の患者における、第一選択とした抗菌薬の種類と、全死亡割合、平均在院日数、およびC.Difficile陽性者発生割合との関連性

データ抽出の方法：以上の3つの仮説を検証する上で必要となる患者の個票データをそれぞれ定義し、病院情報システムからの抽出についてはその情報管理担当者に、DPCデータに関しては医事課職員にデータ抽出に関する手順書を作成した上で依頼を行った。個票データはそれぞれの部署内で連結可能匿名化処理を行い、研究者は患者IDを知らされない状態で個票データを

受け取った。抽出されたデータを合体させ、解析を行った。

患者への説明の機会と方法（集団／個人／情報公開、書面／口頭／広報媒体）は厚生労働省疫学指針 3-1-(2)-イ、すなわち、「既存資料等のみを用いる観察研究」に準拠して行った。本研究の実施に対しては、病院ホームページおよび院内掲示による工事を行った。本解析は、東京医療センター倫理委員会での承認後実施された。

< 結果 >

3つの仮説に関するデータ抽出について、問題なく抽出が行われた。最初の仮説については、65才の入院患者において、血清アルブミン値が入院中に 2.5mg/dl 以下の値を撮った患者を電子的にスクリーニングを行ったが、問題なく抽出条件は完遂された。退院時死亡および退院時ADLについてはDPCデータより抽出され統合された。第二の仮説においては、「身体抑制を行ったかどうか？」については看護記録上の抑制カンファレンスの開催をフラグとして捕捉することが可能であった。また、退院時の身体機能については、看護ケアフロー上の看護必要度B項目が定量化されており、その積算値をもって測定することが可能であった。3つ目の仮説であるC.Difficileの検査実施もしくは陽性所見については、オーダリングシステムおよび細菌検査部門システムにおいて、CDトキシン検査の実施がなされたかどうか、さらにはその結果、培養結果などから捕捉することが可能であった。

< 考察と結語 >

我々は、具体的な統合データの二次利用方法として今回3つの具体的な解析事例を実施した。これらの結果はDPCデータのみから抽出することは全く不可能な患者アウトカムに関するデータを含んでおり、それらの変数定義を病院情報システムのデータを視野においたうえで行うことが可能であった。また、実際にそれらのデータの抽出が可能であった。このことは、現在診療情報としてひび蓄積されつつある病院情報システムのデータを疫学研究等に活用することのできる大きな可能性を示唆するものである。実際、これらの解析結果を国際学会で発表予定である。今後、病院で働く医師等の医療専門職が臨床研究を行う上で、これらのデータ活用は研究活動推進に向けて大きな支援手段となることが示唆された。

< 参考文献 >

- [1] 藤森研二、伏見清秀 他。「医療の質向上に迫るDPCデータの臨床指標・病院指標への活用」じほう(2011/3/30)
- [2] 尾藤誠司 他「保存された診療データの二次利用適用レベルに準じた、医療提供プロセスおよびアウトカムの病院横断比較、年次縦断比較に関する多施設共同研究」総括報告書 2011年4月

< 研究発表 >

学会発表
ISQua Conference, Edinburgh,
13th-16th October 2013.