

研究分担者 東 尚弘 国立がん研究センターがん対策情報センターがん臨床情報部部长

研究協力者 奥山 絢子 国立がん研究センターがん対策情報センター院内がん登録室研究員

診療の質・実態を把握するためのビッグデータについては期待が高まっている一方で、そのデータの質に関する検証は少ない。疾患の数やスコアとアウトカムなどの全体的な傾向を示した事例は散見されるものの、本来入っているはずのデータが入っているのかについてのより直接的な検証は数少ないのが現状である。本研究においては、その一端を明らかにするために、221 施設の 2013 年診断例、主要 5 部位、20 歳以上の自施設初回治療開始例について欠測値や外れ値の頻度や、論理チェックを実施して検討を行った。矛盾や極端な外れ値は少ないものの、明白な不適切値が存在していることから、情報出力にあたって少なくとも現場に近いところで極端な値を検出して訂正する仕組みの必要性が示唆された。

A. 研究目的

診療の質を測定する方法としては、構造・過程・結果の 3 視点があるとされ、直感的なものは生存率等のアウトカムで評価する結果評価であるが、診療の結果は、診療の良し悪しを必ずしも反映しない、患者の基礎状態に影響される、診療から結果が表れるまでに時間がかかる、などの問題がある。過程の評価は、主に正しい診療行為が正しい患者に正しいタイミングで行われているか、が主眼であり、結果の評価の課題を補うことのできる視点となっているものの、何が「正しい診療行為」であるかについて明確にすることが伝統的に困難であるとされた。しかし、診療ガイドラインの普及に伴い「正しい診療」についても基準が立てられるようになり、過程評価にも道が開かれてきた。

過程評価のために次に必要なのはデータ源である。診療自体の記録は診療録にあるものの、その形は標準化されておらず、大規模に収集・解析するのは必ずしも容易ではない。そこで、診療報酬データのように他の目的で収集されたデータを流用することが行われるが、その妥

当性については検証が必要である。

利用のしやすい診療報酬と同等のデータとして DPC データがある。DPC の参加病院においては、DPC 導入の影響に係る調査への参加が求められており、この調査において生成されるデータは院外調剤の処方箋データや外来のデータも含まれるなどの貴重な情報源となっている。また、同時に入院患者については臨床情報が「様式 1」として付加的に入力されており、高い有用性が期待されている。一方でデータの質に関しては要検証とされている。本研究については、既存のデータにおいて一定の解析を行い品質管理に資する情報を提供することを目的とする。

B. 研究方法

都道府県がん診療連携拠点病院連絡協議会がん登録部会 QI 研究において収集された DPC データのうち、様式 1 のデータの提出のあった 221 施設の 2013 年診断例、主要 5 部位、20 歳以上の自施設初回治療開始例について、身長、体重、喫煙指数、高齢者情報、ADL における外

れ値と欠測値の分布を検討した。特に連続変量の代表としては、身長、体重の分布を性・年齢階級別に検討した。またほかの変量の分布や論理的整合性の検討等を行った。値の分布が著しく少ない(同じ値が入力されている)ことがないかなどの検証も行った。

C. 結果

221 施設から 51,876 件が該当 DPC として抽出された。身長、体重については、どちらも共に約 2% 程度の欠損があった。共に少数ではあるが、身長 100cm 未満、体重 500kg 以上と登録されている例があった。喫煙指数については、不明が約 10% あった。また喫煙指数が 10000 を超えて登録されていたのが約 10% あった。情報があった 51,760 件について高齢者情報についてみると、登録対象として 65 歳以上、または 40 歳以上の介護保険適用者とあるが 40 歳未満で登録されていた例が約 1% あった。入院時 ADL について、情報があった 51,760 件中、欠損が 2,475 件(約 4.8%)、不明は項目によって異なるが少なかったもので移乗が 340 件(約 0.7%)、多いもので階段 2,514 件(4.9%)であった。また、ADL について、平地歩行全介助と登録されているのに対し、階段が自立として登録されていたものが 20 件あった。

D. 考察

診療の質を評価するための貴重な情報源の一つとして、DPC データといった診療情報がある。今回の調査対象からは、施設で一律の値が入力されているといったことはなかったが、データ登録における入力ミスと思われる外れ値を示すものや論理矛盾がそのまま登録されていることがわかった。こうした入力ミスは、品質管理を行わない外れ値や論理矛盾のまま登録ができないようにシステムを組み込むことが、

今後 DPC データ等の診療情報を適切に管理する上で重要ではないかと考えられた。一方で、研究者らがこうした DPC データ等を用いる際には極端な外れ値や論理矛盾について確認した上で、研究解析を行なうことが重要と考えられた。

E. 結論

診療の質を評価するための貴重な情報源の一つとして、DPC データがあるが、DPC データを用いて診療実態を把握する際には、登録されたデータの質、外れ値、論理矛盾がないかといったデータクリーニングを行なった上で、データ利用することが重要である。また、診療情報の精度管理を行なう上で、入力ミスと思われる論理的矛盾を登録段階で許容しないような情報管理システムを構築することが今後の医療ビッグデータを利用する上で重要であると考えられた。

F. 健康危険情報

該当なし

G. 研究発表

1. 論文発表

特になし

2. 学会発表

特になし

H. 知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得 該当なし

2. 実用新案登録 該当なし

3. その他 特になし