

科学研究費助成事業 研究成果報告書

平成 28 年 10 月 13 日現在

機関番号：82696

研究種目：若手研究(A)

研究期間：2012～2015

課題番号：24689025

研究課題名(和文)大規模DPCデータセットを利用した意志決定支援システムの開発に関する研究

研究課題名(英文)development of decision support system using a large-scale DPC data set

研究代表者

堀口 裕正(Horiguchi, Hiromasa)

独立行政法人国立病院機構本部(総合研究センター)・その他部局等・研究員

研究者番号：50401104

交付決定額(研究期間全体)：(直接経費) 10,500,000円

研究成果の概要(和文)：本研究においては一般社団法人診断群分類研究支援機構を通じてDPC調査参加医療機関に本研究の趣旨を説明した上で、約1000医療機関から1500退院分のデータ収集を実施することが出来た。そこで収集したデータを活用して診療行為から主たる病名が判定できるかのシステムについて開発と検討を行い、その後外来の重症度を判定するシステムの開発を行い、検討を行った。

研究成果の概要(英文)：On in this study that explains the purpose of the present study to DPC hospitals in Japan throughout the DPC research support institute, it was able to carry out the 1500 discharge of data collected from about 1000 medical institutions. Using those data, we developed and evaluated the system which determined is the main disease name from intervention to leverage and develop the study, and then we developed a system for determining the severity of outpatient.

研究分野：医療情報

キーワード：機械学習

1. 研究開始当初の背景

現在、大量データから、有意な情報を抜き出して、分析するシステムの開発が盛んに行われている。特に Web でコンシューマーサービスを行っている企業等がログや入力された情報を使って非常に精度の高いレコメンドシステムの構築を行っている。また、その分析手法も一般的に入手可能なレベルまで一般化してきている。

また、Hadoop をはじめとした大規模データの分散処理も一般化してきており、その機械学習のためのモジュールである Mahout も非常に速いスピードで進化を遂げているなど、大量のデータ分析を行う環境が整っている。

医療分野においては、その意思決定に対してある治療法やその量・順序等を推奨するものとして、クリティカルパスの存在が挙げられる。これは、入院時の状態像に応じて退院までの治療計画を提示するもので、そのバリエーション分析とともに、医療界ではメジャーな手法になっている。

ただ、この方法では、入院時にパスが決まり、そこから外れる状況になった時には差異あるいは逸脱として計測されてしまう。その問題を解決する 1 つの手法として飯塚らが提唱したのが「患者状態適応型パス」であった。これは固定のパスを提示するのではなく、その場その場の患者像に応じて次のパスを提示していくもので、一種のレコメンドシステムのようなものであると考えられる。このシステムの開発・精度向上にあたっては、医療界の診療知識を構造化するためにプロフェッショナルオピニオンを集め、システムを開発し、一定数の医療機関で運用してそのフィードバックを得て改良していくという手法である。

一方で、もっと大量な診療データを投入してデータオリエンテッドにレコメンドシステムを開発することも可能であるが、日本においてその開発に必要なデータ収集し、分析・開発を実施し、成果を上げているプロジェクトはない。それは、データ収集を行うハードルが非常に高いことが挙げられ、データがあれば、現在の他業種の様々なシステム開発の知見等が、医療業界でも利用可能ではないかと考えている

2. 研究の目的

本研究では大量・多病院の医療行為データ (DPC データ等) を利用して、臨床現場で次のような医療行為を実施すべきかを判断する際に行為案を推奨するシステムの開発を、プロフェッショナルオピニオンをあまり投入せずデータオリエンテッドに行うことによって、利用者のプロとしての裁量を生かし、日本の医療における意思決定の質の向

上に貢献する事を目的とする。

3. 研究の方法

本研究においては、まず分析作業に必要なデータセットの作成を行った。

そこで収集したデータを活用して診療行為から主たる病名が判定できるかのシステムについて開発と検討を行い、その後外来の重症度を判定するシステムの開発を行い、検討を行った。

4. 研究成果

テーマ 1 分析用データセットの作成について

本研究において一般社団法人診断群分類研究支援機構を通じて、DPC 調査参加医療機関に本研究の趣旨を説明した上で協力を呼びかけ、承諾を得る活動を行った。まず、東京大学大学院医学系研究科・医学部倫理委員会の審査をうけ、データ収集及び本研究における分析に関して承認を受けた (平成 24 年度審査番号 3790)。その後、医療機関に協力依頼を行い本研究に対しては、一般社団法人診断群分類研究支援機構を通じて 1066 医療機関から承諾を受ける事が出来た。承諾をいただいた医療機関から発生する DPC データは概算で年間 300 万退院となり、本研究終了時点での 1500 万退院のデータ収集を行った。

また、併せて国立病院機構におけるレセプトデータの利用も行った

テーマ 2 傷病名推計システムの開発

1) 概要

傷病名の推定について、レセプトデータより自動的に作成するサポートモジュールをオープンソースの学習モジュールをベースに開発した。その後、このモジュールに、あらかじめ全国の NHO 病院から収集した約 10 万人分のレセプトデータをもとに学習データを作成し、機械学習モジュールに学習させる。そのうえで NHO の 6 病院を対象に平成 22 年 10 月から平成 22 年 12 月に退院した患者から無作為に 300 症例を抽出し、その症例のレセプトデータを使用して、オンライン学習器で分類を行った。また、抽出された症例について、機構内の診療情報管理士チームによるサポートを受け各医療機関内においてカルテレビューを実施し、各症例の主傷病名を調査・抽出を行った。今回の分析では、機械学習のソフトウェアとして Jubatus (<http://jubat.us>) というオンライン学習用のソフトを使い、分類の手法として Multi-Class Confidence Weighted Algorithm (CW) 及び Adaptive regularization of Weight Vectors (AROW) を使用した。また、機械学習における学習データは、国立病院機

構の他の医療機関のレセプトデータのうち、退院月の主病名が一意に定まっているデータのみを対象とし、その退院月の主病名データを持ってその入院の主病名と定義し、使用している。

2) 傷病名推計システムにおける結果

今回のモジュールを使用した機械学習による分類の結果は以下の通りであった。

(1) レセプトに記載されている主病名と機械学習結果が一致した割合は CW を使用した場合 45%、AROW を使用した場合 45.8%であった。

(2) カルテレビュー結果の主病名と機械学習結果が一致した割合は CW を使用した場合 28.7%、AROW を使用した場合 31.7%であった。

(3) そもそもレセプトに記載されている主病名とカルテレビュー結果の主病名が一致した割合は 43.3%であった。

ここまでで、本件において分類の手法としては CW より AROW のほうが精度が高いという結果が得られた。以降 AROW の結果のみを報告する。

(1) レセプトに記載されている主病名とカルテレビュー結果の主病名が一致した場合、機械学習における正解率は 63.5%であった。

(2) レセプトに記載されている主病名とカルテレビュー結果の主病名が一致しない場合、機械学習の分類結果とレセプトに記載されている主病名との一致率は 32.2%であった。

(3) 機械学習の結果をレセプトに記載のある病名に限って分類させたところ、レセプトに記載されている主病名と機械学習結果が一致した割合は 98.8%となった。

3) 傷病名検査システムに関する考察

病院の医師・事務職員等に多大な負担をかけることなく非 DPC 病院においても病院間共通ルールでの Minimum Data Set が作成できることが期待される。この Minimum Data Set が作成できることにより、現時点で DPC 病院以外の医療機関での診療機能分析で障害となっている患者の疾患名が正しくグループ化されて記載されていないなどの問題を解決することができ、医療の改善に資するものが大きいと考えられる。

ただ、本件の結果、通常の正解率では 45.8%と高くなく、さらなる改良が必要である。しかしながら、カルテレビューの結果とレセプトの主病名が一致している症例に限ると正解率は 3分の2程度まで上昇し、一致しない場合、レセプトの主病名に対する正解率は 3分の1を下回る。これはそもそも主病名決定の精度に疑問がある場合には機械判定がうまくできないことを示しており、まっとう

な結果であると考えられる。

また、レセプトデータの特性として、査定等の影響からレセプトの病名欄にはより多くの病名が記載されるが、逆に重要な病名の記載漏れは少ないことから、レセプトに記載されている病名の中で一番主病名になる可能性が高いと機械に分類された病名を解答としてみた場合、そのデータとレセプトの主病名との一致率は 98%を超え、ほぼ正解を出せる状況にあるといえる。

テーマ3 休日・夜間外に大病院を受診する軽症患者の識別モデルの検討と軽症患者推計

1) 緒言

救急外来は軽症患者のコンビニ受診により、重症患者の治療への支障や現場スタッフへの負担など救急医療提供体制において様々な問題が生じている。

国策としては、健康保険法の一部を改正する法律（平成 18 年法律第 83 号）により、時間外受診の軽症患者に対し選定療養費（時間外）を患者から徴収することが可能になり、軽症患者の受診抑制を図る動きがとられている。また、軽症患者の時間外受診を抑制するために、ホームページに時間外でも軽症患者が受診できる医療機関を掲載するなどの軽症患者の受診抑制のための啓発活動を行っている地域もある。選定療養費（時間外）の徴収により、一定の効果を示す報告もある一方で、消防庁によると平成 26 年度の救急車による搬送人員のうち 49.6%が軽症患者であるなど、未だ解決されていない問題も多い。

DPC データやレセプトデータなど医療費の支払いベースのビッグデータを用いた多くの研究において医療の質の評価や地域医療分析等に関する様々な手法が開発され、診療実態等について定量的に示されている。また、DPC データに加えて患者調査等の政府統計を利用した分析により地域医療の実態が明らかになり、地域保健医療計画等の保健行政にも活用されている。さらに、平成 24 年度より、レセプトデータから日付情報の取得できるようになり、レセプトを使った分析の可能性が飛躍的に拡大した。

夜間・休日・時間外の軽症患者による外来受診については、自治体や病院ごとの受療状況や患者の受療行動に関する報告があるものの研究は少ない。また、複数病院において、DPC およびレセプトデータにより、実際に提供された医療から軽症患者の診療実態を明らかにし、医療提供体制の観点も含めた分析はない。

本研究は、夜間・休日・時間外の軽症患者の外来受診（いわゆるコンビニ受診）の実態を明らかにするために、レセプトデータと診療録調査により軽症患者を識別するモデルを

考案し、そのモデルにより軽症患者を推計した。

2) 軽症患者識別モデルの考案

(1) 分析対象

分析対象施設は、国立病院機構に属する病院のうち、協力の得られた2施設で、いずれも一般病床 200 床以上の病院である。平成 24 年度に夜間・休日・時間外に当該施設に受診をした患者（「A000 初診料の時間外、夜間・早朝等、休日、深夜加算を算定した患者」、「A002 外来診療料の時間外、夜間・早朝等、休日、深夜加算を算定した患者」、「B001-2 小児科外来診療料の初診時もしくは再診時の時間外、夜間、休日、深夜加算を算定した患者」）について、各施設 150 人を分析対象とした。これらの分析対象者は、国立病院機構病院の夜間・休日・時間外に外来受診した患者の年齢構成に従って層化無作為抽出した。

(2) 診療録調査の内容

分析対象患者の診療録より軽症患者の判定、軽症患者における平日早期受診の必要性の有無の判定を行った。更に、主訴、受診時間、主訴発生から受診までの時間もしくは日数を診療録より抽出した。

主訴については、調査終了後に分析者が ICD10 コードの大分類「R00-R99 症状、徴候および異常臨床所見・異常検査所見で他に分類されないもの」からアフターコーディングした。

(3) 判定の定義

a. 軽症患者

診療録の記載内容から、その日の受診の必要性の有無を検討し、翌日以降の 平日受診で問題ないものを軽症と判定。この定義については、総務省消防庁の緊急度判定プロトコルを参考に定義した^{11,12)}。

b. 軽症患者における平日早期受診の必要性の有無

c. 診療録の記載内容から平日平常時間帯の早期受診の是非を判定

(4) 診療録調査の方法

抽出した各施設 150 名分の診療録を 2 名（医師、薬剤師）が独立で判定した。相違がある場合は協議し、最終的には医師の判断を優先した。判定精度を高めるため、各施設 10 名分の診療録を試験入力用とし、双方の意見のすり合わせを行った。

(5) 軽症患者識別モデルの考案

診療録調査の結果と当該患者のレセプトデータを連結し以下の分析を行った。まず、分析対象者の記述統計および判定における係数を算出した。識別モデルは次の3つのモデルを考案し、診療録調査による軽症/軽症外に影響する要因についてロジスティック

ス回帰分析を行った。

- ・診療区分による識別（診療区分モデル）
- ・診察と投薬範囲内の治療と医療費による識別（診療内容-医療費モデル）
- ・主訴と医療費による識別（主訴-医療費モデル）

～ の識別モデル作成にあたっては、表1の変数を投入し、各変数の（偏回帰係数）の絶対値が0.4以上（OR 1.5, OR 0.7）を抽出、再度ロジスティック回帰分析を行った。この変数選択方法をとった理由は、p 値は N 数に依存することから¹³⁾効果（ ）による検討を行うためである。更に、軽症判定定義の特徴から、金・土曜日の判定が厳しくなるため、併せて曜日補正も行った。診療区分モデルで使用した「投薬」、「注射」、「処置」、「手術・麻酔」、「病理・検査」については電子レセプトの診療識別の項目を使用した。医療費については 1000 で除した値を変数で使用した。

次に、～ から得られた回帰式により各モデルにおける軽症/軽症外患者を判定し（カットオフ値=0.5）、診療録調査の判定と識別モデルの判定の感度、特異度、陽性的中度、陰性的中度を算出した。

3) レセプトデータのみで考案された識別モデルによる軽症患者の推計方法

(1) 分析対象

分析対象は、国立病院機構に属する病院のうち、協力の得られた一般病床 200 床以上の 84 施設の平成 24 年度の偶数月に夜間・休日・時間外に外来受診をした患者（「A000 初診料の時間外、夜間・早朝等、休日、深夜加算を算定した患者」、「A002 外来診療料の時間外、夜間・早朝等、休日、深夜加算を算定した患者」、「B001-2 小児科外来診療料の初診時もしくは再診時の時間外、夜間、休日、深夜加算を算定した患者」）のレセプトデータである。なお、2)の診療録調査対象とした2施設も分析対象に包含されている。

(2) 分析方法

上記の軽症患者識別モデルのうち、レセプトデータのみで識別可能なモデルである 診療区分モデルおよび 診療内容-医療費モデルにより夜間・休日・時間外の軽症患者を推計した。推計には2つのモデルにより患者ごとの期待を算出し、続いて各施設の受診患者数から病院ごとの軽症患者割合を算出し、病床規模別に多重比較(Dunnett t)を行った。

上記の分析については、統計解析ソフト SPSSVer23 を使用した。

本研究のこのパートについては、別途、国立病院機構中央倫理審査委員会の承認を得て実施した（承認番号 H27-0312001, 承認日 2015 年 3 月 25 日）。

4) 軽症患者識別モデルの結果

(1)分析対象者の概要

解析対象者は、判定者の入力試験入力用の10名と軽症/軽症外の判定が行えなかった者、特異な事情により受診した5名を解析対象から除外し、275名(A病院:140名、B病院:135名)とした。分析対象者の内訳は、初診178名(64.7%)、再診97名(35.3%)、軽症112名(40.7%)、軽症外163名(59.3%)であった。軽症/軽症外別にみると、初・再診($p=0.61$)、平日・休日($p=0.33$)については有意差を認めなかった。医療費($p=0.00$)、主訴発生から受診までの時間($p=0.00$)、早期平日受診の必要性($p=0.00$)については軽症/軽症外で有意差を認めた。主訴については、「R50.9_発熱、詳細不明」、「R10.4_その他のおよび詳細不明の腹痛」、「R05_咳」、「M79.6_(四)肢痛」による受診が多かった。

(2)識別モデル

軽症患者か否かを従属変数としたロジスティック回帰分析の結果から得られたモデルをそれぞれ 診療区分モデル、診療内容-医療費モデル、主訴-医療費モデルとした。

診療区分モデルでは、「投薬」($OR=0.40$, $p<0.001$)、「注射」($OR=0.31$, $p<0.001$)、「手術・麻酔」($OR=0.06$, $p<0.001$)、「検査・病理」($OR=0.36$, $p<0.001$)が軽症判定との関連を認め、特に「投薬」と「手術・麻酔」を実施していないことが軽症に強く影響していた。診療内容-医療費モデルでは、「診察・投薬・注射の範囲内の診療」($OR=2.16$, $p=0.01$)、「医療費」($OR=0.68$, $p=0.01$)が、主訴-医療費モデルでは、「医療費」($OR=0.51$, $p<0.001$)、「R50.9_発熱、詳細不明」($OR=0.24$, $p<0.001$)が軽症に影響していた。

これら3つのモデルの精度を表7に示す。

診療区分モデルは感度が高く、診療内容-医療費モデル及び主訴-医療費モデルは特異度が高い。また、3つのモデルはいずれも陰性的中度が高かった。

5)レセプトデータのみで考案された識別モデルによる軽症患者の推計結果

分析対象者は、対象施設84病院に受診した110,199人の患者データを基に診療区分モデルと診療内容-医療費モデルにより推計した軽症患者(率)を推計した。診療区分モデルと診療内容-医療費モデルの相関係数は $R^2=0.99$ ($p<0.001$)であった。診療区分モデルの推計では、軽症患者43.8%(最小値=31.4%,最大値=61.2%)、診療内容-医療費モデルの推計では、軽症患者42.8%(最小値=31.5%,最大値=55.5%)であった(表8)。病床規模別にみると、いずれのモデルも、500床以上と200-299床で有意差を認めた(Dunnett t $p=0.01$, $p=0.00$)。

6)テーマ3の考察

本研究では、レセプトデータと診療録調査により夜間・休日・時間外外来受診における軽症の識別モデルを考案した。それを基に国立病院機構に属する200床以上の病院の軽症患者数を推計し、病床規模別の検討を行った。

(1)軽症患者識別モデルの考案

2施設の診療録調査を行った結果では、初診患者が60%であった。先行調査によると同法人における外来初診患者数割合が8.1%であり、夜間・休日・時間外受診患者における初診患者の受診多い。また、軽症/軽症外と初・再診との関連はなく初・再診による患者の重症度の違いを認めなかった。更に、軽症患者は、症状発症から24時間以降の受診が多く早期受診の必要性とも関連していた。

受療状況については、施設間差を認めた。先行研究では救急医療においては提供側、享受側いずれの視点でも地域差が生じていることは既知の事実であるが、本研究では、地域により患者層が異なるものの受診時の患者の状態に対し提供される医療に差はないと考えられることから、軽症患者の識別モデルの考案では両施設を併合して分析した。また、受診日に診断がつきにくい点を考慮し主訴での分析により、軽症患者識別モデルを考案した。

考案した3つのモデルをみると診療区分モデルでは、注射、手術麻酔、検査病理のいずれかを実施していないことが、診療内容-医療費モデルでは、診察投薬の範囲内の治療と医療費が、軽症患者の識別に影響していた。主訴-医療費モデルでは、医療費と「R50.9_発熱、詳細不明」が軽症患者の識別に影響していたが、ロジスティック回帰分析の結果、有意差を認めなかったものの「R10.4_その他のおよび詳細不明の腹痛」、「52.9_非感染性胃腸炎および非感染性大腸炎、詳細不明」、「R11_悪心および嘔吐」の効果()の値が高い。分析対象数を増やすことで他の症状による予測も可能であると考える。

更に、これらの3つのモデルの精度をみると、診療区分モデルは感度が比較的高く軽症患者を推計する上で予測率が高いと考えられた。一方で診療内容-医療費モデルでは特異度が高く軽症外患者の推計に適しており、軽症患者を過小評価する(軽症外患者を軽症と判定する率が低くなる)ため軽症外患者の見逃しを抑制できると考えられた。主訴-医療費モデルでは、当然のことながら、患者の主訴という情報量が増えることにより特異度および陰性的中度を上げることができ、より軽症外患者の判定に適するモデルといえる。

(2)識別モデルによる軽症患者の推計の考察

診療区分モデルおよび診療内容-医療

費モデルは、レセプトデータのみで軽症患者を予測できることから患者推計を実施した。その結果、夜間・時間外・休日受診している患者の約4割が軽症患者であり、50%を超える施設もあることが明らかとなった。更に、病床規模別でみると500床以上の病院に対して200床以上300床未満の病院との間で有意差を認めていることから、病床規模が大きい病院であっても軽症患者が推定4割であるということは軽視できないものの、現在、国策として進められている病院機能分化が、夜間・休日時間外受診においても反映されているのではないかと考えられる。

先にも述べたように、患者の救急医療の依存度は上昇し、軽症でも救急車を利用する患者の増加、時間外受診の増加と一般外来受診患者の減少が問題となっている。既に多くの医療機関や自治体が救急医療を受けるにあたり患者に対する啓発活動を行っているが、症状発生後遅延して受診する患者に軽症が多いことを考慮すると、早期受診に対する患者の理解や協力の要請を求める必要があると言える。都市化が進むと医療への依存度は一般外来から救急外来にシフトさせ救急医療の需要を増加させるという報告もあり、コンビニ受診の増大が見込まれる地域もあることから、救急医療を真に必要とする患者に対する医療資源の確保の観点からも患者への協力の要請が一層必要になる。

7) 研究の限界

本研究にはいくつかの限界があり、今後の課題として以下を挙げる。

第1に診療録調査における軽症患者の判定基準である。軽症を明確に定義する先行研究がないため、総務省で議論された緊急度判定体系に関する検討会の考え方を基に適宜した。当該検討会の設置目的は、『増大する救急医療の需要に対し、救急医療に投入できる資源を有効に配分・活用し、緊急性の高い傷病者を優先して搬送することにより救命率の向上を図る等、「急ぐべきは急ぎ、待つべきは待つ」という緊急度判定の基本的な考え方が社会全体で共有されるよう推進すること』と記されている。本研究においてもこの考え方を踏襲し、あくまでも「平日まで待てる受診であるか」を重視する定義とした。判定にあたっては、判定者2名の精度管理を目的に各施設の診療録10名分のデータを試験入力用としたが、係数からもわかるように判定の定義については検討の余地がある。

第2に軽症患者の受診背景である。本研究では、必要性の低い夜間・休日・時間外受診患者と、結果的に軽症であった患者が混在しており、いわゆるコンビニ受診患者を明確に絞り込むことは困難な点である。本研究で軽症と判定された受診を否定するものではなく、それらの患者の中に夜間・休日・時間外受診の必要性が低い患者が存在し、受診を見

直す必要がある患者集団を絞り込んでいく。

以上のような限界を有するが、総務省消防庁の報告から)救急医療利用者の増大は今後も見込まれると考えられ、その中には一定数コンビニ受診も含まれることは容易に予測できる。コンビニ受診の今後が増加してくる背景には社会的な要因が大きく影響していると考えられ、これまで救急医療の過剰利用に関する行政の対応や多くの研究が報告されているものの夜間・休日・時間外を受診する軽症患者の推計を示す見解は見当たらない。本研究は、モデルの精度に検討の余地はあるものの、一定の手順により患者数を推計し約4割が軽症患者による受診であることが明らかになったことは有用であると考えられる。

8) 全体のまとめ

本研究においては当初病名推計システムのシステム開発及びデータセットの安定的な運用を行いつつ、本データセットを利用した分析として救急モデルの作成を行った。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 1件)
医療機関を多数保有する組織でのIT化と医療情報部門の今後 堀口裕正 新医療-2016年5月号 - 月刊新医療

〔学会発表〕(計 1件)
堀口裕正, 岡田 千春, 阿南 誠, 海野 裕也, 伏見 清秀, 村中 光: 機械学習を用いたレセプトデータの主傷病分類の精度検討 第32回医療情報学連合大会 2012

〔図書〕(計 0件)

〔産業財産権〕
出願状況(計 0件)

取得状況(計 0件)

〔その他〕
なし

6. 研究組織

(1) 研究代表者

堀口裕正 (Hiromasa Horiguchi)
国立病院機構本部総合研究センター
診療情報分析部 主席研究員

研究者番号: 50401104